Understanding Norwegian youth online help seeking through bigdata analysis:

what do their digital traces tell us?

Jisu Lee

Master´s Thesis in Nordic Media

Department of Media and Communication

UNIVERSITY OF OSLO

December 1st, 2020

Abstract

As digital natives, young people are familiar with seeking help through online media. Conventional research has studied youth online help seeking through survey and interview, yet it is now available to approach their actual behaviors through the digital traces thanks to the bigdata environment. Therefore, this dissertation aims at understanding youth online help seeking through large amounts of automatically saved data when they are using youth support media.

For this, 285,447 questions from young Norwegians requesting advice from ung.no, a Norwegian public youth information website are used. Text mining on their questions is implemented along with additional analyses mixing data of timestamps and demographic information.

It is found that the Norwegian young help seekers have patterns of specific times, and differences in topics according to their age and gender. In addition, a total of 29 demographic groups can be clustered into four with their similar contents of concerns.

Studying online help seeking through big data and its analysis tools provides opportunities to supplement what previous studies have not been able to access and can be used as evidence of expansion of existing research boundary. However, concerning the quality and representativeness of bigdata, and validity in the research interpretation is still important as in any scientific research and certainly need to be considered more seriously.

Acknowledgements

Living in Norway, the first proverb that I learned as a foreigner was “ut på tur aldri sur” which means “out on a tour, never sour”. Honestly, I should admit that there were some sour moments while writing my master’s dissertation, most derived from worries towards an uncertain future and disappointments in myself. Nonetheless, I managed to finish this journey thanks to many great people, and now I feel like I became a more mature and responsible person who can finally understand the lesson of the proverb.

I would like to express my gratitude to my supervisor, Petter Bae Brandtzæg, who guided me with objective and realistic advices. I learnt a lot from him how to organize my thoughts and represent them in a scientific way. I also want to thank my co-supervisor Morten Goodwin from UiA. Despite of the distance, he always welcomed me to contact and discuss technical methodology. Furthermore, I appreciate professor, Trine Syvertsen, for stimulating my intellectual interest in the first year of the master’s program and giving us, the last generation of Nordic Media study, precious opportunities to experience Norwegian and Scandinavian media customs and culture. I will always remember how lucky I was to study at UiO receiving all the excellent academic support.

A very sincere thanks to my boyfriend, Jonas Øvrum, and my Norwegian family for standing by me during the whole time of my master and providing uncountable support. Also, I feel grateful to my Korean family to believe in me and let me follow my passion abroad. Finally, I would like to say how many inspiring friends that I met from IMK, IFI and the ones from language classes, and others outside of the university. While sharing so many interesting ideas, my horizon broadened, and my life became more colorful.

This two-and-a-half-year journey is over, and I’m excited to find what my next journey will be. I hope I remember this moment of me enjoying the joy of completion whenever I feel tired and regretful in the middle of work.

Table of Contents

1. Introduction 1

1.1 BACKGROUND AND PURPOSE OF THE RESEARCH 1
1.2 RESEARCH SCOPE AND METHOD 3

2. Literature Review 4

2.1 YOUTH HELP SEEKING 4
2.1.1 Definition of Help Seeking 4
2.1.2 Youth Help Seeking 6
2.1.3 Youth and Online Media 9
2.1.4 Literature Review of Youth Online Help Seeking 11
2.1.5 Use of Youth Help Seeking Research and Opportunity for Help Intervention 17
2.1.6 Summary of chapter 2.1 21
2.2. DISCUSSION OF BIGDATA APPROACH FOR YOUTH HELP SEEKING 23
2.2.1 Bigdata definition and characteristics 23
2.2.2 Bigdata use in Business 25
2.2.3 Bigdata Discussion in Social Science 31
2.2.4 Digital Media as Bigdata Source 38
2.2.5 Literature about youth online help seeking applying bigdata approach 41
2.2.6 Summary of chapter 2.2 47

3. Research Questions and Methodology 50

3.1 RESEARCH QUESTIONS 50
3.2 TEXT MINING 51
3.2.1 Text mining definition and the basic unit of analysis 51
3.2.2 Text mining process 52
3.3 RESEARCH PROCEDURE 58
3.3.1 Data collection 60
3.3.2 Preprocessing texts 62
3.3.3 Analyzing the data 67
3.3.4 Interpreting the results 75

4. Research Result 77

4.1 TIME TRENDS ANALYSIS IN THE NUMBER OF YOUTH ONLINE HELP SEEKING QUERIES 77
4.2 DEMOGRAPHIC ANALYSIS OF YOUTH ONLINE HELP SEEKERS 81
4.3 HOURLY TRENDS ANALYSIS IN THE NUMBER OF ONLINE HELP SEEKING QUERIES BY YOUTH DEMOGRAPHIC GROUPS 82
4.4 WORD FREQUENCY ANALYSIS IN THE CONTENT OF HELP SEEKING QUERIES 86
4.5 CO-OCCURRENCE WORDS ANALYSIS OF COMMUNICATION KEYWORDS 89
4.6 TOPIC MODELING OF HELP SEEKING QUERIES 93
4.7 CLUSTERING OF ONLINE YOUTH HELP SEEKERS 101

5. Discussion 107

5.1 DISCUSSION ABOUT BIGDATA ANALYSIS ON YOUTH ONLINE HELP SEEKING 107

5.2 Discussion about opportunities and challenge of bigdata research 110
5.3 Limitation of the research and future recommendation 113
6. Conclusion 114
6.1 Summary 114
6.2 Academic and social contribution of the research 116
Appendix 117
Reference 126

1. Introduction

1.1 Background and Purpose of the Research

Young people in a transitional period in which they experience turbulent physical, intellectual, emotional, and social changes has curiosity and worries about health, sexuality, career, relationship, etc. in their minds. The adaptive coping process, a communicative attempt to obtain external support to solve this problem, is called help seeking (D. Rickwood & Thomas, 2012, p. 180). Many governments around the world have implemented health interventions for young people such as contacting youth professionals so that they can access the right health information. However, it is reported that many of the young are reluctant to approach professionals and forming therapeutic relationships with them (Pretorius, Chambers, & Coyle, 2019). This is mainly due to negative preconceived notions related to seeking professional help, then young people prefer to contact informal sources such as parents and friends with whom have shared relationships and previous experiences (Gray, Klein, Noyce, Sesselberg, & Cantrill, 2005). Still, the youth challenge parental boundaries and have a strong strive for self-reliance (Arnett, 2000). Furthermore, they are unwilling to disclose their minds to even the closest ones when it comes to sensitive topics (Callahan & Inckle, 2012). In brief, facing embarrassment and the fear of stigma is inevitable through direct personal interactions.

Meanwhile, young people today are the generation who were born after the Internet, so-called “digital natives” (Prensky, 2001) or “born digital” (Palfrey & Gasser, 2011). They cultivate sophisticated knowledge and skills in information technology (Bennett, Maton, & Kervin, 2008) and use the Internet as a space for self-expression exploiting anonymity (Holloway & Valentine, 2003; Livingstone, 2009). In addition to this, they not only reinforce existing relationships and but also explore new ones through online communication (Lehdonvirta & Räsänen, 2011). Therefore, youth help seeking naturally spreads from offline to online, and it has been studied that it is more advantageous for the young to look for help through digital online media rather than direct personal interaction (D. Rickwood, Deane, Wilson, & Ciarrochi, 2005). This is considered to have significantly changed the nature of traditional help seeking. In other words, youth help seeking is implemented beyond direct interpersonal interaction through online media and information systems, and rather the latter is becoming more frequent. Therefore, there are studies that have narrowed the research scope of youth help seeking to youth online help seeking. Most of these studies focus on identifying psychological factors such as motivations and obstacles of online help seeking while surveys and interviews or focus group interviews are applied as methodologies.

Existing studies of youth help seeking have been used as evidence for the development of theories and experiments in health intervention (Costin et al., 2009; Joyce & Weibelzahl, 2011), but there are still several limitations that prevent them from becoming actual intervention policies. Among them, the WHO (G. Barker, 2007) noted that there is a lack of research that sufficiently reflects the viewpoints of the young. What they point out is that actual approach to understanding young people’s help seeking should be considered by observing how they decide when, where, and how to seek help and the reasons for why they seek help. Along with this critique, in 2007, WHO proposed therapeutic narratives and social marketing as new methodologies for future youth help seeking studies. Therapeutic narrative allows young people to describe their experiences and opinions related to help seeking, rather than answering questions from pre-scripted research tools. And social marketing is to actively use additional information from the youth and find their subdivided needs. Those suggestions might have seemed too ideal to be carried out at the time. However, in today’s bigdata environment, their ideas have the potential to be realized. It is possible to analyze actual youth help seeking behaviors and find patterns of subgroups of the young. In other words, by storing and extracting digital traces that are generated automatically when the youth seeking help, one can apply appropriate analytics to explore new perspective of youth help seeking.

In fact, bigdata is more commonly used in business, yet recently academia has also embraced this approach and enjoyed its empirical benefit. Business use bigdata as a decision-making means, in academia, however, the scientific framework should be still maintained. More precisely, rather than drawing a conclusion from bigdata at once, their analysis result can be acknowledged making new hypotheses and challenging existing theories. In this trend, there have been attempts to study youth help seeking using the bigdata of their digital traces, yet the number of literature is limited.

Considering that the trend of help seeking moves from face to face communication to online, there is an urgent need for more knowledge about youth online help seeking. Bigdata analysis can open hidden patterns that have been not allowed in the traditional studies confined in the theory-driven frame and limited research tools. Hence, this paper aims to explore youth online help seeking through a bigdata approach including its computational methods and find opportunities and limitations while conducing it.

1.2 Research scope and method

To explore characteristics regarding online help seeking among young people, 285,447 records of their help seeking from a question answering service of ung.no, a Norwegian public youth website are used. Research data is composed of time stamps of when questions were generated, demographic information of help seekers, and question texts. Statistical analysis and text mining method using lexicometrics and Machine Learning (ML) will be heuristically applied to analyze the data, and the results are interpreted with visualization.

This dissertation proceeds through the following chapters. Chapter 2 will start with dealing with the existing concept and theory of help seeking, research methods and limitations of existing literature. The second half chapter 2 is about discussing bigdata as a new approach to understand youth help seeking by reviewing business and academic research. In chapter 3, research questions will be suggested with implications from the previous chapter. Furthermore, the research methodology and procedure will be introduced. As the data includes the youth help seeking text queries, there is a subchapter fully describing text mining method. Chapter 4 contains visualization and results of planned analyses while in chapter 5 they will be interpreted along with discussion for answering the research questions. Finally, in chapter 6, the summary of whole dissertation together with its academic and social contributions will be described.

2. Literature Review

2.1 Youth Help Seeking

2.1.1 Definition of Help Seeking

According to Nelson-Le Gall (1985), traditional approaches to the study of help seeking were based on the values of Western individualistic cultures. Help seeking was understood as an act contradicting the dominant cultural background at that time where competition, self-reliance and independence were emphasized. To be more specific, the socialization process was characterized as making movements from innate reliance on others to self-sufficiency so that help seeking was often seen as an indicator of dependence, immaturity, passivity, and even incompetence in early studies of socialization and personality development (Beller, 1955; Gall, 1985; Murphy, 1962; Sears, Maccoby, & Levin, 1957). However, these negative perspectives on help seeking have been changed by social psychologists and sociologists who studied help seeking in the context of medical and social welfare (Gall, 1985). Today, the concept of help-seeking is defined in various fields, and the scope of its subjects has also been diversified to such as children, adolescents, adults and the elderly.

In the health and psychology field, Gourash (Gourash, 1978, p. 414) defined help seeking in a broad term which is “any communication about a problem or troublesome event that is directed toward obtaining support, advice, or assistance in time of distress”. Particularly in mental health field, Rickwood and Thomas (2012, p. 180) specified help-seeking as “an adaptive coping process that is the attempt to obtain external assistance to deal with mental health concerns”. As this form of the coping relies on other people so-called help provider(s) in nature, social relationships and interpersonal communication skills are emphasized in their notion.

Karabenick (1987) proposed “academic help seeking” in the education field. It was defined as a planned activity from a learner to improve the lack of information and skills necessary to achieve the academic goals. Academic help seeking is considered as constructive rather than dependent, focusing on a self-regulated learning strategy in which a help seeker oneself determines when help is needed, and how to ask and receive that help (e.g.Arbreton, 1994; Gall, 1985; R. S. Newman, 1991, 2000; Ryan & Pintrich, 1997).

In organizational contexts, help seeking is understood as a proactive request from a help seeker (van der Rijt et al., 2013). It includes not only the activity of a worker asking for consultation to the experienced to solve challenging problems at work, but also his or her voluntary information searching on work-related matters and reviewing superiors' feedback of certain labor performance. Lee (1997) also perceived help seekers as the beings who proactively define and frame their problems and have active role in problem solving and learning in organizations.

Help seeking has also become an important subject in the field of Information Commutation Technology (ICT). Puustinen and Rouet (2009) argued that the advent of the internet and digital technologies have facilitated much more sophisticated search activities, thus the boundary between help seeking and information searching has become blurred. Furthermore, using online media, one gets help from the people behind the computers through virtual interactions. Some stated that now it became inappropriate to differentiate help into human and non-human (Puustinen & Rouet, 2009; Zimmerman & Pons, 1986). Therefore, in ICT, help seeking is based on "the information system constituting an additional, intermediate stage in the communication" between help seeker and help provider (Puustinen & Rouet, 2009, p. 1017). Moreover, this discussion is extended beyond computer mediated interaction to the concept of help seeking through artificial intelligence (Karabenick, 1998).

As presented above, the definitions of help seeking varies in different fields. The following table 1 shows the summary of the above explanation. In this study, help seeking is understood from the viewpoints of health and ICT fields. In other words, help seeking is re-defined as one's communicative efforts to find solution about any problems of state of mind, and here the help is sought not only in face to face interactions, but also through information systems. Meanwhile, many research branches exist within help seeking studies especially in accordance with types of help seekers. Most commonly, help seekers are categorized by their life circle such as child, adolescent, adult and the elderly. In this study, the youth including the concepts of adolescents, teenagers and young people are the main subjects. Then, from the following part, it will be described about the youth help seeking in particular.

Table 1

The concepts of help seeking defined in different fields

Perspective	Concept	Author
Health	“any communication about a problem or mental health concerns that is directed toward obtaining support”	Gourash (1978), D. Rickwood and Thomas (2012)
Education	“planned activities to improve the lack of information and skills necessary to achieve the academic goals”	Karabenick (1987)
Organization	“a proactive request for help by consulting with someone to obtain specific information on work-related matters or to solve challenging problems at work”	van der Rijt et al. (2013)
ICT	“activity including information search and information system constituting intermediate stage in communication”	Puustinen and Rouet (2009)

2.1.2 Youth Help Seeking

Going through the period of adolescence, young people face physical, intellectual, psychological and social changes (Suzuki & Calzo, 2004). These changes come all the sudden when they are still living in the end of the childhood, being carried away to adulthood by the laws of nature. Curiosity and worries about health, sexuality, career, relationship, etc. occur in their minds and sometimes these can be developed to mental problems without being resolved. Many studies and reports have already shown that the young are particularly vulnerable for mental health difficulties (G. Barker, 2007; Gulliver, Griffiths, & Christensen, 2010; Slade, Johnston, Oakley Browne, Andrews, & Whiteford, 2009). Developing interventions for encouraging help seeking behavior in adolescence is important in order to reduce future risk behaviors and lead the young people to a higher quality of adult life (J. E. Anderson & Lowen, 2010; Brindis et al., 2007; Divin, Harper, Curran, Corry, & Leavey, 2018). For this purpose, how they seek help has been of a critical concern for many researchers.

Among them, Rickwood and her colleagues seem to have strived to lay the theoretical foundation for youth help seeking (e.g. D. Rickwood et al., 2005; D. Rickwood & Thomas, 2012). For instance, D. Rickwood et al. (2005) studied the sources that young people approach to get help, and broadly divided them into two categories. That are formal and informal sources. Formal help seeking is asking for help from any professional who has a recognized role and trained experience in providing supports. The examples are doctors, school nurses, teachers, youth workers, etc. On the other hand, informal help seeking sources are non-professional whom the youth may or may not share a personal relationship (Pretorius, Chambers, & Coyle, 2019). For instance, traditionally family and friends are the typical informal sources of the youth for getting supports (Gray et al., 2005).

The same authors also proposed a conceptual model of help seeking especially focusing on individual and psychological factors that facilitate for young people to attempt to communicate about their worries. More concretely, they explained the help seeking process as "a social transaction between the personal domain of the internal world of thoughts and feelings and the interpersonal domain of social relationships" (D. Rickwood et al., 2005, p. 8). The following figure 1 illustrates the corresponding model.

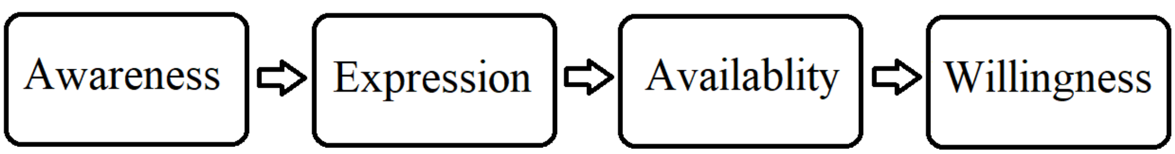


Figure1. D. Rickwood et al. (2005)'s Help-seeking model

The first stage suggested in this model is "awareness". It is that a help seeker recognizes personal domain in relation to problems such as feelings or symptoms, and makes the appraisal of them. The second is "expression", which is a declaration of need for help. The third one is "availability". It means that the seeker knows which sources are available and is able to access to them (Pretorius, Chambers, & Coyle, 2019). The last stage is "willingness" of the seeker to reach out and disclose their difficulties to help sources. These steps affect sequentially until the action of social transaction is taken. This help-seeking model is the most known for youth help seeking, still not many of existing literature employed model to their study designs (Pretorius, Chambers, & Coyle, 2019). Most literature were conducted with explorative purpose. As a result, diverse help seeking patterns among the youth are found. Still, some common tendencies among young people can be recognized as follows.

For instance, many studies have found that young people are reluctant to seek professional help (Andrews, Issakidis, & Carter, 2001; Gulliver et al., 2010; King, Bickman, Shochet, McDermott, & Bor, 2010; Pretorius, Chambers, & Coyle, 2019; D. Rickwood et al., 2005; Rowe et al., 2014; Zachrisson, Rödje, & Mykletun, 2006) and the ones who in need of psychological supports are those least likely to look for it (Aguirre, Silva, Billings, Jimenez, & Rowe, 2020; Davies, Lemer, Strelitz, & Weil, 2013; Divin et al., 2018; Gulliver et al., 2010; Kowalenko & Culjak, 2018; Michaud & Fombonne, 2005; D. Rickwood & Thomas, 2012; D. J. Rickwood, Deane, & Wilson, 2007; World Health Organization, 2012). Instead of contacting formal sources, the youth tend to cope with their problems by themselves or rely on informal sources first. When it comes to informal sources some stated that family is more important source (Ackard & Neumark-Sztainer, 2001; Chambers, Reid, McGrath, & Finley, 1997; Dickinson, 1978; A. W. Gould & Mazzeo, 1982; Gray et al., 2005; Schoen et al., 1997; van den Berg & Parry, 1983) while others found that friend is the first source and subsequently, they ask adult for help (Boldero & Fallon, 1995; A. W. Gould & Mazzeo, 1982; Raviv, Raviv, Vago-Gefen, & Fink, 2009; Raviv, Sills, Raviv, & Wilansky, 2000; Schonert-Reichl & Muller, 1996).

Regarding these tendencies, researchers have also presented the main obstacle factors of youth help seeking. For example, D. Rickwood et al. (2005) summarizing their 19 studies conducted over a number of years, reported that the main barriers of youth help seeking are high reliance on self to solve problems, lack of emotional competence, help-negation, and negative attitudes and beliefs related to seeking professional help. Gulliver et al. (2010) found the similar factors in their systematic literature review, but stigma and embarrassment is the most significant discouragement for young people seeking any helps, followed by having problem of recognizing symptoms, and preference for self-reliance. The fear of stigma here means being shamed not only by health professionals, but also by family and friends. Even though young people tend to believe that getting help from family and friends are more trustworthy due to the existing relationships and shared previous experiences (Camara, Bacigalupe, & Padilla, 2017), at the same time they are afraid that friends and family find out their mental health concerns and make negative judgment on them (Gulliver et al., 2010). This is more pronounced in youth who have a specific mental problem. For example, Rowe et al. (2014) who especially researched on youth help seeking for self-harm showed that the main hindrances are such fear of negative reactions from others including stigmatization, the fear of confidentiality being breached and being seen as "attention seeker". Lastly, the most recent systematic review conducted by Aguirre et al. (2020), also found that stigma and negative preconception about mental health services and professionals are the most important obstacles by comparing 90 related studies. Based on these studies, the most problematic factor for young people to seek help is seen as the concern about getting stigmatized, which has direct relation with the last stage of the help seeking model, "willingness". This can be critical because even a young person who have had no distractions in the previous three stages, would not take a help seeking action if he or she does not have willingness due to the fear of stigma.

However, there exist other ways for young to ask for help avoiding "willingness" stage being struck by stigma. This is when youth seek help online. D. Rickwood et al. (2005) already pointed out that youth increasingly look for help on the internet where direct personal interactions with others are not required. This is a noteworthy trend since their help seeking model has emphasized the importance of the social transactions implicit in traditional offline help seeking but the nature and need for these social transactions are different online (Pretorius, Chambers, & Coyle, 2019). So, in the next part, the characteristics of online media use of youth will be introduced first. Afterward, youth online help seeking which is the main topic of this research will be described.

2.1.3 Youth and Online Media

Youth are the most important demographic group of the digital generation who have grown with internet technology, and they live surrounded by online media both at home and at school. Their parents and teachers are "Digital Immigrants" who have received the Internet and online media at a certain point in their lives, while teenagers are naturally familiar with the digital environment because they were born after the Internet was established (Prensky, 2001). So, they are often referred to as “Digital Native” (Prensky, 2001) or “Bone Digital” (Palfrey & Gasser, 2011) or “Net Generation” (Tapscott, 1998). Many researchers have expected that the characteristics and attitudes of teenagers using the online media differ greatly from those of older generations. Some of common research results indicate as follows.

First of all, the youth of the digital native generation have sophisticated knowledge of and skills with information technologies (Bennett et al., 2008). According to Prensky (2005), young people live fully immersed in digital technology and are fluent in the digital languages of computers, video games and the Internet. As they have grown up surrounded by online tools, Frand (2000) argued that this immersion became so intense that young people no longer treat computers as technology. Accordingly, the youth are fundamentally different in thinking and processing information than the previous digital immigrant generation. Young people are used to receiving information very quickly, and they like parallel processes and multitasking. They prefer graphics over text, like random access like hypertext, and are characterized by functioning best when networked (Prensky, 2001; Prensky & Berry, 2001).

Second, young people use online media as a space for self-expression. Online media has given adolescents unprecedented access to the horizontal world where the boundary of adults and children are blurred through anonymity. This is a new opportunity for youth to escape the constraints of customs and traditions, and they seem motivated to seek out and experiment with identities and relationships in online space (Livingstone, 2009). In other words, in pursuing the reflective project of self, children and adolescents especially enjoy the Internet as a valuable place for self-expression (Holloway & Valentine, 2003; Livingstone, 2009). In regard with this, Bargh, McKenna, and Fitzsimons (2002) stated that one’s real self is easier to access through online. Here, the real self consists of personal traits that the individuals believe they have, yet are unable to express in the everyday social environment due to the normative, social and cultural constraints. However, the anonymity of online media offers a sense of freedom from them, so that the youth are encouraged to express a unique form of themselves in the online environment (Davis, 2012).

Third, young people are familiar with communicating through social media for reinforcing existing relationships and building new ones with others. According to a study by

Duggan and Brenner (2013), adolescents are a group that fully embraces online communication through social media technologies, and today's youth generation is recognized as the first group to grow with online social networking (D. M. Boyd & Ellison, 2007). Social media or social networking sites (SNS) are "websites which make it possible to form online communities and share user created content" (W. Kim, Jeong, & Lee, 2010, p. 215). There are studies reporting that the main motivation for young people to use social media is driven by the need to communicate with peer groups (V. Barker, 2009). In other words, the youth extend their interactions with existing friends from offline to online. On the other hand, young people are also used to socializing with strangers through social media. For some, the relationships built online can be more special than the existing relationships. For example, from online communities where people in common interest gather and exchange information or experience, the youth can make virtual networks. According to Lehdonvirta and Räsänen (2011) who conducted an international study of users in a teenage online community, the youth identify more strongly with the online community than with their neighborhood or offline hobby group.

As such, as digital natives, young people are more familiar with online media than anyone else, and even it is often advantageous to express themselves and making relationships online. Therefore, their help seeking behavior also naturally spread from offline to online, and there are many researchers who observed this with interest. In the next part, their literature will be introduced comprehensively to shed light on what have been important subjects in youth online help seeking so far.

2.1.4 Literature Review of Youth Online Help Seeking

Recently, online media have gained great attention as means for the youth to reach health and mental health related information and find a personal solution by asking and sharing their concerns (G. Barker, 2007). Growing use of computer-mediated technologies and online sources have significantly changed the nature of traditional help seeking by offering a new opportunity for young people to seek help without direct interpersonal interactions (Pretorius, Chambers, & Coyle, 2019; D. Rickwood et al., 2005). There are various literature dealing with youth online help seeking. They are descriptive and explorative types of research, which most of them applied quantitative methods using survey and others are qualitatively conducted with interview or focus group interview. Only a few of the literature employed theoretical frameworks such as Rickwood's help seeking model or other behavior theories for their study design (e.g., Best, Manktelow, & Taylor, 2014; Bradley, Robinson, & Brannen, 2012; Collin et al., 2011; Cunningham et al., 2014; S. Kauer, Buhagiar, & Sanci, 2017). Therefore, there were no criteria coincided among the studies, yet some meta-analyses have tried to integrate them and provide comprehensive understanding of youth online help seeking according to the common themes of studies. On the basis of the meta-studies, five common themes are categorized as follows; (1) motivation of youth for online help seeking, (2) types of online sources and youth preferences, (3) content of problem for seeking help (4) the youth experience and effect of online help seeking, (5) obstructive factors of online help seeking.

(1) Motivation for online help seeking

Studying on the motivation of the youth using online sources for help seeking is the most common starting theme in the current literature. This is usually portrayed by the comparison with traditional interpersonal interactions such as young people sharing their worries and asking for help from their surrounding people. As it is already inspected above, the biggest hindrance of young people not seeking help is due to the worry about being stigmatized. This psychological burden makes them reluctant to ask for help from not only professionals but also close acquaintances, because the moment they reveal their concerns, taking risk of being evaluated is inevitable. On the internet, however, anonymity and protection of personal privacy are guaranteed, thus these benefits attract the youth the most (Best, Gil-Rodriguez, Manktelow, & Taylor, 2016; Bradford & Rickwood, 2014; Ellis et al., 2013; Greidanus & Everall, 2010; Horgan & Sweeney, 2010; O'Dea & Campbell, 2011; Pretorius, Chambers, & Coyle, 2019). This can be understood in the same context as young people of digital natives perceive the Internet as a space for self-expression, freeing from normative, social and cultural restrictions.

Another significant facilitator for online help seeking is that the youth can easily access to online help sources (Bradley et al., 2012; Burns, Davenport, Durkin, Luscombe, & Hickie, 2010; Collin et al., 2011; Davis-McCabe & Winthrop, 2010; Horgan & Sweeney, 2010). As the internet becomes a part of daily activities for the youth, their help seeking has extended to online naturally (Pretorius, Chambers, & Coyle, 2019). The internet is free time and space constraints so that young people can approach online sources whenever they need via computer, laptop, tablet and mostly their mobile phone (Pretorius, Chambers, Cowan, & Coyle, 2019). Therefore, they have less difficulty accessing online support in a timely manner (M. S. Gould, Munfakh, Lubell, Kleinman, & Parker, 2002; Gray et al., 2005). In addition, as young people have sufficient technical experience in using online media, this make it easier for them to execute help-seeking behavior online. Furthermore, low monetary cost of using internet is mentioned together as an additional reason (Pretorius, Chambers, Cowan, et al., 2019).

In many research, the benefit of connecting with others and sharing personal experiences is also pointed out (J. Bell, Mok, Gardiner, & Pirkis, 2018; Birnbaum, Candan, Libby, Pascucci, & Kane, 2016; Horgan & Sweeney, 2010; Kummervold et al., 2002; Mar et al., 2014; Neal, Campbell, Williams, Liu, & Nussbaumer, 2011). While anonymity is guaranteed, the youth still have interactions with unknown help providers who understand their state of mind. Besides, they can find informal sources on online forums and communities on which there are many who are or have been in similar situation and get psychological comfort and indirect advice from them (J. Bell et al., 2018; Birnbaum et al., 2016; Mar et al., 2014). Considering the characteristics of adolescents as digital natives discussed earlier, the support obtained online community may act more importantly than that of offline.

Lastly, the youth can control the level of disclosure of their personal problems is also motivating factor for online help seeking (Best et al., 2016; Frost, Casey, & Rando, 2016). For instance, while interacting with professionals, there is certain power imbalance that may make the youth expose themselves beyond the range of where they feel comfortable. On the other hand, other internet, there is not that pressure and young people can flexibly reveal themselves and their trouble. This is related to giving certain self-determination to young people who long for being independent existence from adult's involvement.

(2) Types of online sources and youth preferences

Various types of online sources are found in existing literature, yet the most common one is seeking help through text-based queries using search engines (Best et al., 2016; Burns et al., 2010; Feng & Campbell, 2011; Mar et al., 2014; Mars et al., 2015; O'Dea & Campbell, 2011; Wetterlin, Mar, Neilson, Werker, & Krausz, 2014). According to Wetterlin et al. (2014), the main goals when searching online is information seeking about symptoms and ways of treatment such as how to stop the symptoms.

Following search engines, health information websites run by governments or certain organizations (Best et al., 2016; Burns et al., 2010; Feng & Campbell, 2011; Wetterlin et al., 2014), online discussion forums and communities (Eichenberg, 2008; Kummervold et al., 2002), social media (Best et al., 2014; Birnbaum, Rizvi, Correll, Kane, & Confino, 2017; Feng & Campbell, 2011), Internet-based self-help program (Bradley et al., 2012; Collin et al., 2011; Davis-McCabe & Winthrop, 2010; E.-H. Kim, Coumar, Lober, & Kim, 2011), Live Chat (Haner & Pepler, 2016), instant messaging (Frost et al., 2016) are where the youth look for help.

Preference regarding these resources varies depending on samples from each study. For instance, the study of Hansen, Derry, Resnick, and Richardson (2003) found that text-based query on search engine has benefit over information websites when it comes to personalizing information. In other studies, the young consider such online websites operated by government and professional organization as more trustworthy (Best et al., 2016; Birnbaum et al., 2017; Frost et al., 2016). The need for help seeking services run and recommended by professionals is a recurrent result throughout many of the studies (Best et al., 2016). Furthermore, the young who have high stress are likely to access immediate sources such as live chat and instant messaging (Frost et al., 2016; Haner & Pepler, 2016).

(3) Content of problem for seeking help

Each study tried to identify what kind of problems young people have for online help seeking, but the content varies widely from study to study. This is because some researchers approach overall youth health, while others study on some specific health issues, such as mental health or sex health. In addition, as privacy and ethical issues are taken into account, it seems that detailed investigations into the content of problems are not included and they tend to be generally and broadly portrayed. Therefore, it is difficult to synthesize content of problems, still some example can be shortly introduced.

For example, according to the study by Gray et al. (2005), that investigates a broad range of health problems of the youth who look for help online found that the top three subjects are specific diseases, sexual health and weight loss or gain. In the study conducted by Borzekowski and Rickert (2001), the topic most often explored through the internet by the young were sexually transmitted diseases, diet, fitness, exercise and sexual behaviors. Meanwhile, another online help seeking studies which in particular focus on mental health, figure out that the youth search for symptoms and treatment for anxiety, depression, insomnia etc. (Horgan & Sweeney, 2010; Mar et al., 2014; O'Dea & Campbell, 2011).

(4) The youth experience and effect of online help seeking

In the article of S. D. Kauer, Mangan, and Sanci (2014), the authors tried to put results of youth experience from 9 different research together. Even though there are various measures of experience, yet most of studies found that the youth have positive attitude on online help seeking overall. This includes the satisfaction of the youth using online sources, willingness to continue to reuse the sources, and interest in recommendation to friends. Interestingly, however, the authors found that young people's satisfaction with sought help was not very high. The studies included in their review show that only half of the youth were able to reach the information they were looking for and that considered the help programs helpful (S. D. Kauer et al., 2014).

Furthermore, more recent systematic review conducted from Pretorius, Chambers, and Coyle (2019), which include 7 studies that are not overlapped with those of (S. D. Kauer et al., 2014), also found the similar results. The youth answered that the online resources they had made use of were a little helpful and did not make things better or worse.

(5) Obstructive factors of online help seeking

Even though online help seeking has significant motivating factors that have changed the nature of social transactions with others emphasized in the traditional help seeking model, there are still some obstructive factors found. One thing is lack of internet-based health literacy. Internet-based health literacy means the ability of individuals to obtain and understand online health information and services for making right health decision (Maitz et al., 2020). Young people who lack internet-based health literacy would not know which resources to search for nor recognize which is helpful information or not (Best et al., 2014; Havas, de Nooijer, Crutzen, & Feron, 2011; Ruppel & McKinley, 2015). Maitz et al. (2020) conducted mixed method study which assessed health literacy of the 14 young students in advance and let them to perform an internet-based search on a health-related issue. The study found that the students judged their internet-based health literacy a lot higher than the actual value, yet they did not necessarily access high quality websites (Maitz et al., 2020). Along with this study, others also reveal that the youth are not used to identify online health information with objective standards. They tend to approach top-ranked results from search engines, and well-designed websites rather evaluating its quality themselves (Best et al., 2016; Druin et al., 2009; Gwizdka & Bilal, 2017; Park & Kwon, 2018; Subramaniam et al., 2015).

The other barrier discussed from existing literature is concerns about privacy and confidentiality of using online sources. The internet support anonymity a lot largely compared to traditional sources so that this motivates the youth to seek help online, yet there are still same concerns remain among young people (Best et al., 2016; Horgan & Sweeney, 2010; S. D. Kauer et al., 2014). For instance, Mar et al. (2014) found that young people worry that family and friends would find out that they are having trouble. Same has found from a cross-sectional survey from Pretorius, Chambers, Cowan, et al. (2019) and they suggested that it may be related to the result that many of their investigated youth would use their mobile phones for online help seeking instead of computer and laptop which can be shared with others more often.

Through this review, various aspects of the existing literature dealing with youth online help seeking behavior were investigated. It has seen that the traditional help seeking model suggests that social transactions between young people’s self-world and interpersonal relationships go beyond face-to-face interactions and take place in a virtual space, thereby solving the latent tensions to a large extent. In other words, online media encouraged the youth as digital natives to comfortably reveal themselves and communicate with others, so that to carry out such social transactions. However, the youth experience and satisfaction of online help seeking vary depending on preferred sources and so on, and there are still obstacles that hinder them to get appropriate solutions online. Thus, there is an effort to apply another theories and develop more sophisticated theoretical frameworks that integrate such observations. On the one hand, there are already some cases in which health interventions are attempted to be built based on the general results of previous youth help seeking or online help seeking studies. These two directions will be introduced in the next subsection.

2.1.5 Use of Youth Help Seeking Research and Opportunity for Help Intervention

To recapitulate the two directions mentioned above, the first is the development and expansion of frameworks in the field of youth online help seeking which have found its own specialty from general help seeking research area. This can be recognized as the future contribution to the academia when it comes to interdisciplinarity. The other direction is to provide a cornerstone for youth help interventions. Intervention programs based on existing results have been designed and tested. If its effectiveness is proven and allowed to be used for youth, it can make a huge contribution to society. The following starts with a detailed description of both directions.

(1) Specialization of the field

In the academic world, the existing studies have increased the value of this specific field that combines the topics of online media and the youth in help seeking. To be more specific, new theoretical frameworks are suggested and extension of D. Rickwood et al. (2005)'s Help-seeking model have attempted, which solely reflect characteristic of youth online help seeking, being independent from general help seeking. For instance, Pretorius, Chambers, and Coyle (2019) suggest that Self-Determination Theory (SDT) can be applied to conceptualize motivation in youth online help-seeking after analyzing recent 28 studies. SDT explains that there are three primary needs for well-being and motivation which are autonomy, competence and relatedness (Deci & Ryan, 2008). Pretorius, Chambers, and Coyle (2019) clustered the benefits and barriers identified from the existing studies in terms of their impact on autonomy, competence and relatedness as consistent with SDT. They noted that the young gave mixed responses about their satisfaction of online help seeking. For example, online help seeking by search engine satisfies their need of autonomy, yet is able to impede their competence due to the variety of results. Applying SDT, they argued that the online sources that are designed to meet these three needs will make positive effect on youth motivation on help seeking.

Meanwhile, Best et al. (2016) proposed a pathways-based extension of the existing help seeking model, especially reflecting analysis result of online help seeking behavior of male youth. To be more specific, the new model actively involves youth's perception of stigma and health literacy when they pass each stage of the existing help seeking model. Therefore, the pathways become more complicated including prediction of which online media such as social media, search engine, online communities, etc. that the youth would decide to get help. The figure 2 shows the description of their model.

(2) Evidence for health intervention

The other use of online help seeking studies is to design digital intervention for youth. The results of the studies have offered evidence for developing intervention tools as a part of health communication. For example, Costin et al. (2009) built an intervention employing e-cards which are personalized emails containing links to health information webpages. Their aim is to evaluate if the intervention may increase help seeking through a randomized controlled trial. Their active conditions are designed to promote help seeking behavior and intentions, and improve beliefs and knowledge related to help seeking. For this, they in advance reviewed the obstructive factors from existing studies and tried to modify them. One of factors they noted is the youth are lack of knowledge and understanding of which are the reliable sources, then they provided pre-selected health information webpages.

Another case is a study of text messaging intervention conducted by Joyce and Weibelzahl (2011). They emphasized the need for intervention that can overcome the barriers of online help seeking previously studied and designed an intervention tool and conduct experiments on young people. They not only considered obstructive factors but also motivating factors. More specifically, they noted that one of the benefits of online help seeking is that it gives youth to feel to solve problems on their own. Then, they made an initial contact technology based and there would be no human to human interactions, thereby maintaining the sense of solving alone. Regarding obstacle factors, one example is that they decided to use text messaging as a communication type for intervention to reinforce confidentiality and anonymity. What they explained is that texting can be done under the guise of normal behavior and avoid suspicions from friends or family.

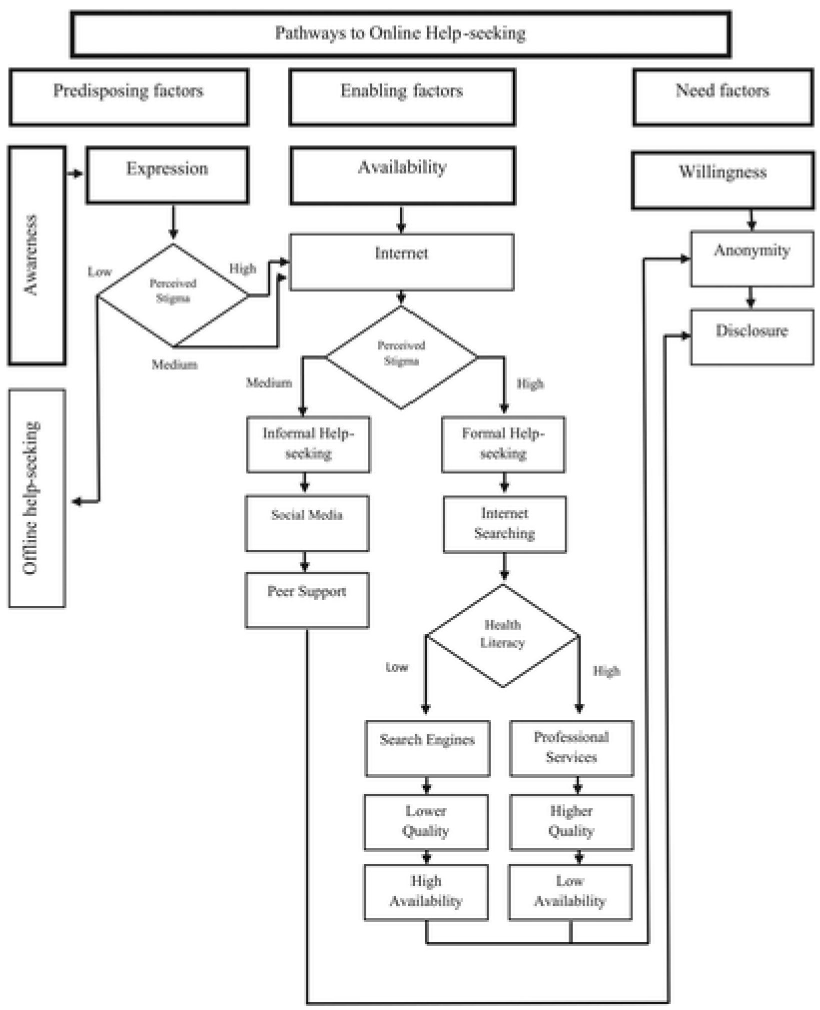


Figure 2. Best et al. (2016)'s conceptual model of pathways to online help seeking

(3) Limitation and new opportunity for intervention

Such experimental studies on interventions have been conducted, there has been also policy interests in developing intervention programs to loosen barriers to youth help seeking and youth online help seeking (Biddle, Donovan, Gunnell, & Sharp, 2006; Divin et al., 2018; Rothì & Leavey, 2006). However, despite the elaborate intervention designs, the effects of intervention on youth help seeking are reported to be insignificant. In a systematic review by Gulliver, Griffiths, Christensen, and Brewer (2012), it is found that some interventions improved intention of help seeking, yet there is no proven effect for help seeking behavior of youth. Related to this, the World Health Organization (WHO) also points out that the current youth seeking help research base is still flawed in achieving intervention development (G. Barker, 2007). The key reasons suggested here is that there is no consensus about "what help seeking is, how to measure and evaluate it, how to promote it and how to promote positive use of both informal and formal sources of support" (G. Barker, 2007, p. 29). It has already demonstrated that there is high variety in every aspect of youth help seeking studies in the literature review. Furthermore, it is also mentioned that there is lack of research on youth decision making. To be more specific, actual understanding of "how adolescents decide when, where and how to seek help" and "why young people seek help, and which help they seek" is required (G. Barker, 2007, p. 29).

As it is seen in the literature review, studies mostly focus on facilitators and barriers to help seeking behavior, yet the content of what the youth are worry about has received less attention and been investigated less clearly. What WHO indicates is that more observation of the youth and their needs should be take into account when researching help seeking. However, it is also not difficult to guess the reason that existing studies have not been able to directly reach young people's concerns and needs. It is because personal information and ethics issues are especially concerned due to the nature of the field. Interviews as a qualitative research method are supposed to help in-depth understanding of subjects, but there is no guarantee that youth would open to share their problems, and the number of samples would be too small. On the other hand, in quantitative studies, one can ask their problems through anonymous surveys, still the depth of response selection would not enough to reveal inner story of young people.

However, in terms of recommendations, WHO early came up with potential approaches in 2007. One of them is applying the concept of “therapeutic narratives” to youth help seeking. This methodology originates from mental health professionals listening to narratives of patients, which include “the story of health problem, the progression of the need, seeking help for the need” (G. Barker, 2007, p. 29), and “surrounding events which are both relevant and irrelevant” (Early, 1982, p. 1491). This is based on the idea that people have a variety of narratives that make up their understanding of who they are, and the issues they bring to a therapy are not restricted to the client themselves, but are influenced by cultural discourse about identity and power (Madigan, 2011). Applying therapeutic narratives to youth help seeking, one let the youth tell “the meanings, beliefs and self-reported behaviors related to their perceived need for help and their perceptions of available formal and informal social support” (G. Barker, 2007, p. 29). Another suggested approach is learning from social marketing thereby regarding young people as consumers and trying to understand where, when and which services they use (G. Barker, 2007; Brunetti et al., 2001). This a practice of customer analysis which can be advanced to predict potential customers. This is a practice of customer analytics that use customer data to understand different needs of customer. This can be developed to predict customer behaviors and finally change them in the direction company wants. Furthermore, customer segmentation that is dividing customers into several groups that are similar in specific ways can be applied to the youth help seeking so that better interventions can be recommended to each group perhaps categorized by their demographics, needs, preferences of sources, etc.

2.1.6 Summary of chapter 2.1

This section 2.1 began with narrowing the scope of this study into youth online help seeking among a wide range of areas of help seeking. D. Rickwood et al. (2005)’s renowned help seeking model indicates that help seeking behavior is a social transaction between the young people’s self-reliance and the benefits of disclosure to others. This transaction tends to be hampered by fears of being stigmatized in the case of face-to-face communication. Yet in the internet space where anonymity is guaranteed and easily accessed through digital media, the transaction is more flexibly taken place by young people who are digital natives in this era. Thus, the existing literature suggesting this uniqueness of online help seeking for youth was reviewed. They generally dealt with motivational factors for online help seeking, types of online source and youth preferences, the problems of asking for help, young people' experience and effects of online help seeking. Among them, the main focus is on motivational and obstructive factors, and relatively few studies on the content of problems or experiences. In addition, all of these studies were exploratory studies, and used consistent research methods, such as surveys, interviews, or focus group interviews.

These existing studies contribute to the consolidation of expertise in this field by bringing together three different topics; youth, seeking help and online. Meanwhile, based on the findings of these studies, health intervention programs have been built and tested, which are critical to encouraging young people to seek help and providing them right information. However, as any reliable effect of intervention has not yet been discovered, it is still difficult to make policy intervention. Related to this, WHO stated that consensus among high variety of research practices should be achieved first. Furthermore, they underscored that there are no efforts of studying how youth actually seeking help in real life, which raise the need of observational study and user centered study. Therefore, for the future recommendation, the new approaches such as therapeutic narratives and social marketing methodologies are suggested to apply to youth help seeking.

This study found the last two recommendations significantly interesting. Considering that they were suggested in 2007, these ideas might have seemed too ideal. When attempting a therapeutic narrative that induces the youth to tell their own stories, there would always be concerns about ethics and personal information, and it would take too much time and human resources to secure a sufficient number of interviewees. In addition, when applying social marketing-style customer analysis, there would have been limitations in obtaining data such as their access logs to online sources and usage information. However, these difficulties can be receded today in “the bigdata era”. This implies that a technical foundation has been laid for storing massive records of youth online help seeking, extracting the necessary parts as research data, analyzing them with state of art computational tools, and finally obtaining interesting patterns and insights. Being able to approach the actual behaviors of young people and their raw narrative stimulates the challenge of new research practices for youth online help seeking. So, the next section 2.2 is about understanding bigdata and exploring the background for applying bigdata to youth online help seeking. Specifically, it will be discussed starting with the concept of bigdata, examining how bigdata is being accepted in social marketing and social science research, and reviewing the cases where youth online data are used in help seeking research.

2.2. Discussion of Bigdata Approach for Youth Help Seeking

2.2.1 Bigdata definition and characteristics

In the McKinsey's first bigdata report published in 2011, bigdata was defined as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Manyika et al., 2011, p. 1). Related to the size, the authors explained that the concept of "big" is subjective, there is no established criterion for how large a data set must be to be considered bigdata, and as the technology advances over time, it was assumed that the size of the data set recognized as bigdata also increases. As it is seen here, at least in the early days of bigdata, the size used to be the main interest, which is now only considered as one of features of bigdata. For instance, the Gartner Report describes bigdata as "high volume, high velocity and high variety information assets that demand cost effective, innovative forms of information processing for enhanced insight and decision making" (Beyer & Laney, 2012; Gandomi & Haider, 2015, p. 138), focusing on the aspect of utilization, rather on the size. Furthermore, The International Data Corporation (IDC) also explained bigdata as "a new generation of technologies and architectures designed to extract value economically from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis" (Vesset et al., 2012, p. 1), emphasizing the value creation aspect of bigdata. Likewise, in a new McKinsey's report published in 2016, which revisited the previous one, their interest has certainly shifted to value creation through bigdata analysis while expressing their expectations for artificial intelligence technologies such as machine learning and deep learning. Taken together, it is evident that bigdata now does not only indicate data itself, and its qualitative meaning is closer to the essence of bigdata. So, bigdata can be defined here as large-scale data that can create value at low cost by using state of art processing and analysis methods.

Accordingly, the characteristics of bigdata which are known for 3V (Volume, Variety and Velocity) suggested by Laney (2001) have extended as 5V by adding 2 more V (Value and Veracity). “Volume” means the increased amount of generated and collected data and “Variety” refers to the different types of data. “Velocity” indicates the speed at which the data is collected and analyzed, which should be rapid and available in real time. To these 3V, “Veracity”, that is, the quality of data for accurate analysis, and the ability to convert the data into “Value” are also added. Following table 2 shows the detail explanation of 5V.

Table 2

Description of 5V

5V	Description
Volume	The amount of digital information exponentially increases every year, entering the era of Zetabyte (ZB).
Variety	There are multiple data forms (e.g., structured, semi-structured and unstructured data) from different sources.
Velocity	As data inflows at high rate (e.g. in real time), the speed of data processing and analysis is important.
Veracity	It is about understanding that there are integral discrepancies in bigdata and measuring the accuracy of data and its potential use for analysis (Sivarajah, Kamal, Irani, & Weerakkody, 2017; Vasarhelyi, Kogan, & Tuttle, 2015).
Value	The value of data is unlocked when extracting certain knowledge and insights from massive data.

These characteristics draw most attention in the industry, rather academia. In the industrial sectors, bigdata has been actively adapted as a tool of providing evidence for decision making and leveraging economic benefit in the future (Zillner et al., 2016). Furthermore, existing business intelligence and marketing intelligence that rely on data analytics to gain business insights have been reshaped since new tools and methods processing bigdata, including state of art artificial intelligence, have been developed together with bigdata (S. Fan, Lau, & Zhao, 2015).

2.2.2 Bigdata use in Business

(1) Bigdata analytics in business

Business intelligence broadly embraces strategies and technologies for analyzing business information to make better decision regarding operating company. When it comes to bigdata analytics in business, usually following three are introduced, which are descriptive, predictive and prescriptive analytics (Joseph & Johnson, 2013). Descriptive analytics is to investigate data and information to identify the current state of a business situation. Certain patterns and exceptions can be found, and this result will be presented as standard reports, ad hoc reports, and alerts (Sivarajah et al., 2017). Predictive analytics is to forecast unknown future events regarding business. As this use current data to make predictions about the future possibility, not only statistical modeling but also artificial intelligence techniques such as machine learning can be engaged (Waller & Fawcett, 2013). Lastly, prescriptive analytics is concerned with optimization and randomized testing to assess the best course of business action in a certain scenario (Joseph & Johnson, 2013). As more elaborated artificial intelligence-based decision techniques involved, it advances prediction and presents actionable suggestions and their potential outcomes. The following table 3 shows how different big analytics can be used in business.

Table 3

The scope and example of bigdata analytics in business

	Descriptive	Predictive	Prescriptive

Scope	What is has already happened?	What will happen in next future? What trends will continue?	How to achieve the best outcome for any given condition?
Example in practice	Netflix uses data mining to determine the correlations between various events.	ING uses it to analyze all their customer's data and enable them to also predicts their customer's behavior.	Amazon.com utilizes price optimization based on demand to increase the online shopping revenues.

Note. Adapted from Ghani, Hamid, Hashem, and Ahmed (2019)

Some scholars include diagnostic analytics between explanatory and predictive analytics, or more interestingly suggest pre-emptive analytics implemented by state-of-the-art artificial intelligence as a next level of prescriptive analytics. As such, today's interest towards business analytics in the era of bigdata lies not in understanding the past, but in proactively dealing with unknown future by analyzing data changing in real time. In other words, it has been already well acknowledged that bigdata is considered as one the most significant technological disruptions in business, which contain methods in which a machine can learn by data. Nonetheless, bigdata business analytics, mainly driven by practice, is still descriptive analytics (Reed & Dongarra, 2015; Sivarajah et al., 2017), and predictive analytics rather prescriptive analytics.

S. Fan et al. (2015), in their article, introduced some examples of bigdata analytics that can be used in the actual business workplace in more detail at the marketing level. In particular, they stated that marketing has traditionally relied on customer data and improved the 5P's of marketing mix which are product, people, promotion, price and place by analyzing them. Using this explanatory framework, examples and methodologies of bigdata analytics based on 5P are described in the following sections.

(2) Product reputation management and product ontology

In marketing, it is important to monitor the reputation of company's products and services. Conventionally, survey is a usual method to ask consumers to evaluate the quality of products and services. Morinaga, Yamanishi, Tateishi, and Fukushima (2002) pointed out that survey is pricey when it comes to getting responses of few customers' opinion about limited products and service. To alleviate this, a new framework of mining web content was presented in their article, which is collecting related posts and reviews written by customers on the web and analyzing what they imply. Text mining techniques such as word frequency, word co-occurrence analysis and sentiment analysis to discern whether a text is positive or negative are used to examine opinion of customer in relation to each product and service.

Furthermore, recently ontology mining based on latent topic modeling has been applied for product marketing. This is a work of extracting free discussion about a product from online social media and building an ontology of the product based on themes found in the social texts (Lau, Li, & Liao, 2014). One example is a study by Jeong, Yoon, and Lee (2019) who conducted ontology mining using user generated texts about a certain phone model written on online social forum, reddit.com. They perceived social media texts as an emerging voice of the customer, identified product ontology based on topic modeling. To be more specifically, from 23,614 text documents including posts and comments on the research product, they extracted 65 topics by applying Latent Dirichlet Allocation (LDA) algorithm, then manually defined the top six topics.

(3) Customer segmentation

Customer segmentation refers to the practice of dividing a customer base into homogenous subgroups that are internally similar yet differ from one another (Grunig, 1989; Rogers & Storey, 1987). The customers share similar preferences and respond to a specific marketing signal will be categorized in a same group. This allows the company to better understand different customers and plan strategies targeting each group. In one example case conducted by Brito, Soares, Almeida, Monte, and Byvoet (2015), they used customer data of a manufacturer of custom-made shirts and implemented customer segmentation based on five criterion. They are; (1) product characteristics (type of fabric, color, collar type, pattern), (2) demographic and biometric (gender, age, collar size, BMI), (3) geographic (Nationality), (4) psychographic (lifestyle, purpose) and (5) behavioral (price sensitivity). By clustering analysis using K-Medoids algorithm, individual 10,775 shirts orders were grouped into 6 clusters. Furthermore, they did subgroup discovery analysis applying CN2-SD algorithm and found 49 rules among variables.

There is another study (Greco & Polli, 2020) which made different approach to label customer types. In the study, 107,500 tweets and retweets data related to a specific sports clothing brand were scraped and the content were clustered based on the bisecting K-means algorithm. Total five clusters are found, and the authors then performed network analysis with Louvain method to see relations among the words in each cluster for meaningful labeling. In the end, the five clusters are explained as the combination of sport lovers and fashion lovers. The sport lovers are defined as people who seem to be mostly focused on the technological innovation and its use love. These sports lovers are again comprised of three clusters which are those who like the milestone model, those who like to be fit and those who like new releases. On the other hand, they defined fashion lovers are someone considered the brand's image more. Two clusters that are bargain hunters and the fashion customers belong to fashion lovers.

(4) Promotional marketing analysis and Recommender Systems (RS)

Promotional strategies are designed to increase their sales and revenue of a company (D. R. Bell, Chiang, & Padmanabhan, 1999). Promotional marketing analysis is to see how customers respond to the strategies or how different categories of products are related to the promotional effect (Pauwels, Hanssens, & Siddarth, 2002). In the bigdata environment, more and more log data can be utilized for promotion analysis (Lu, Ba, Huang, & Feng, 2013). One study by Bae and Park (2018) measured effect of advertising which is a major promotion tool. In this case, 78 hundred million online commercials log data were used as samples. Ad impressions and Click Through Ratio (CTR) on a mobile application were used as tools of measuring effect, while their forms (e.g. bar, full-page, and floating), material (e.g. text and image) and characteristic (e.g. hedonic and utilitarian) were designated as independent variables. Applying regression model to this massive data, they found that the advertising effect of the floating advertisement was higher than that of the bar-type advertisement. And the floating advertising represented in text had higher effect than that of in image. In addition, the advertising effect of the floating advertising was higher when the content has utilitarian characteristic than hedonic.

Another practice for promoting product and service awareness to potential customers is Recommender System (RS) that predict preference a customer would give to an item (Ricci, Rokach, & Shapira, 2011). One of the approaches for recommendations is based on the assumption that customers who had similar preferences in the past, would like similar items in the future. Wang, Chan, and Ngai (2012) used demographic information of customers and their ratings about tourist attractions scraped from online travel website for developing RS. Focusing on six cities that are New York City, Paris, London, Rome, Chicago and Berlin, they extracted 33,040 data which include the ratings and demographic information of tourists who have rated (five-point Likert scale) the attractions belonging to the category of "Museums, Zoo, and Aquariums" in the website. For prediction, they applied three different machine learning methods and 10-fold cross-validation for evaluation. As a result, the best version of RS identified around 90% tourists that belong to the class rated as 5. This shows that how customer demographic information can be utilized in regard with predicting their preferences of unknown products and services.

(5) Pricing strategy

While traditional empirical research on pricing strategies are based on survey data and regression methods, the digitization of stock and purchase records, and the growth of e-commerce have made price related information available for estimating demand and optimization of prices. For instance, major league baseball has adopted dynamic pricing based on bigdata analysis to improve revenue management (Steinbach, 2012). This means that they have implemented a flexible pricing strategy based on changing consumer demand during a season. This involves integration of many variables such as timing of ticket sales, weather, construction around the ball park, teams on the rise, the potential for a record-setting event, amount of chatter about a game in social media and what tickets are selling for on ticket marketplace (Erevelles, Fukawa, & Swayne, 2016).

Another study is about the use of bigdata for price optimization in retail conducted by Bradlow, Gangwar, Kopalle, and Voleti (2017). They used 308,460 Stock Keeping Unit (SKU) codes from 42 stores of national-scale retail chains as samples for implementing price optimization, and field experiment with A/B testing approach. For the price optimization they used predictive econometric model which includes a number of input variables such as price, feature, trends, seasonality, demand shifters, special events, reference, etc. Afterwards, the optimized prices were test in 21 different stores for 13 weeks. Analyzing the results with difference model, model-based price optimization based on bigdata improved gross margin dollars both managerially and statistically at the test stores compared to control stores (Bradlow et al., 2017).

(6) Place-based bigdata analysis

Traditionally, place marketing has focuses on where and how products and services are bought. With the widespread use of mobile technology, the interest of place-based marketing has placed on customer mobility and their location data (Dhar & Varshney, 2011). One example is that a study (Y. Chen, Zhang, Guo, & Cao, 2017) used location information from registered customers in shopping malls for the learning base of a real-time recommendation. The dataset consists of 89,794 check-ins from 39,038 customers on 211 shops for 7 months, and they were collected using an opt-in Wi-Fi network. The authors argued that the customer's preferences can be inferred from their check-in activities and time of stay. So, they proposed a time-aware recommendation model which suggests a customer a set of shops at a specific time slot by learning customer's preference from their check-in frequency and average check in time.

There is another example of using big amount of customers' location data is conducted by Sorensen et al. (2017), which attempted to find patterns of in-store shopper behavior. The used data were 654,000 transactions in 40 supermarkets, hypermarkets, convenience and specialty stores in the USA, UK, China, and Australia. The key behavior they focused on was how much of the store a customer covers on a single trip, basket size which indicates number of items a customer bought, and the amount of time a customer spent in a store. Applying various statistical models, they confirmed some findings that can be useful for marketing in retail and manufacture. Main findings are that "customer's store coverage, basket size and trip length of each show consistent patterns. And customers on small trips, of which there are many/most, typically cover a small proportion of the store, shop quickly, and purchase just a few items" (Sorensen et al., 2017, p. 188).

When bigdata reveals its intrinsic value beyond the characteristic of size, its meaning becomes intact. Bigdata has mainly been in the spotlight in business, so some examples of bigdata analytics at the marketing level firstly reviewed. Yet, the use of bigdata does not exist only within business. The field of public policy, health, medicine, science, engineering, etc. have acknowledged the value of bigdata. In addition to this, bigdata has become regarded as research materials in social sciences and even humanities as well as. In particular, in social sciences including such as communication study, bigdata has the advantage of reflecting the aspect of human society as it is. So, the following subsection will describe how bigdata has brought new trend to academia especially in social science research.

2.2.3 Bigdata Discussion in Social Science

Not only in industry, but also what bigdata has brought to academia is considered as paradigm shifts which challenge conventional research practices. As this dissertation belongs to social science research in the broad sense, and online media research in the narrower sense, the future explanation will first deal with how the field of social science has received bigdata. Afterwards, examples of youth help seeking literature using online social data are reviewed.

First of all, the following is what D. Boyd and Crawford (2012) claims about bigdata metanoia in research.

"Bigdata creates a radical shift in how we think about research …. Bigdata reframes key questions about the constitution of knowledge, the processes of research, how we should engage with information, and the nature and the categorization of reality … bigdata stakes out new terrains of objects, methods of knowing, and definitions of social life (D. Boyd & Crawford, 2012, p. 665)".

As a matter of course, social scientists make sense of empirical data using theory-driven approaches to explain social phenomena, rather than just describe them (Qiu, Chan, & Chan, 2018). Theory works as a lens to understand mechanisms of social phenomena so that either explicitly or implicitly its importance used to be emphasized more than that of data and methodology (e.g., Babbie, 1989). However, bigdata has raised doubts as to whether theory-based research is still most substantial. Bigdata becomes greater and greater, reconfiguring in many instances how research has been conducted. In this bigdata trend, data-driven approach is also applied in academia not exclusively remaining in the business industry, then social scientists have started to adopt them in the study of communication, politics, phycology, economics, sociology etc. (Kitchin, 2014; Yarkoni & Westfall, 2017). The following explains in more detail how this trend is being settled and how it is driving change in social science.

(1) Data spectrum and naturally occurring data

So far, researchers have strived to collect and control data for their studies. Not only finding samples that have representativeness, but in the case of quantitative studies, a sufficient number of samples should be ensured while being efficient in terms of time and cost (Miller, 2010). The cost and efficiency of using existing research methodologies, however, have been reconsidered in today's bigdata era which was established based on the wide availability of the Internet and online media. Bigdata are generated constantly and exhaustively so that researchers have become relatively free from pressure on the number of samples and their repressiveness (Kitchin, 2014). Unlike in the past when data could be captured only once at a high cost, it is now possible to collect the same type of data repeatedly, even very often at a fraction of the cost

(R. M. Chang, Kauffman, & Kwon, 2014). Therefore, in social science, “bigdata refers to datasets that are too big for researchers’ humans to code a representative sample of the entire dataset” (Riffe, Lacy, Fico, & Watson, 2019, p. 92). Meanwhile, the data spectrum has increased not only in width but also in depth. It has become easier for researchers to access to data at different analytical levels in between micro, meso and macro. To be more specific, bigdata encompasses the whole range of society, so through macro-level analysis, subtle population patterns and heterogeneity that cannot be found in small-scale data analysis can be discovered (J. Fan, Han, & Liu, 2014). On the other hand, bigdata also captures social life in great detail, it makes possible to narrow down the analytical level significantly for studying small subgroups or relatively rare events (J. Fan et al., 2014).

Examples of bigdata are countless. Still, to quote some from literature, there are online survey data, social media postings, online political discourse, digital journalism, user logs on the Internet and mobile application, consumer data from enterprise, real time video footage, measurements from sensors planted into objects or environments, etc. (Kitchin, 2014). Such data are differentiated from conventional samples which are created by systematic human intervention in that most of them are from “naturally occurring” or “raw” or “user-generated” social and digital media sources. To be specific, the traditional methods of social science studies such as surveys, interview and experiments by their nature involve researchers into the research process, making it inevitable to intervene in experimenter effects (Babbie, 1989; Shah, Cappella, & Neuman, 2015). For example, survey respondents are expected to choose the most approximate answer within the boundary of examples given by researchers. When it comes to experiment, researchers may unintentionally treat experimental and control groups differently during the research. Furthermore, it is difficult to overlook the possibility that survey respondents and interviewees may have social desirability bias in which overreport their good qualities and underreport bad or undesirable behavior to be viewed favorably by others (Babbie, 1989; Krumpal, 2013; Shah et al., 2015).

In contrast, in bigdata analysis researcher can collect the data reflecting the society and behaviors of people as they are, free from the artificiality of traditional methodology. To emphasize this characteristic, IBM researchers (2012) portrayed bigdata as the new natural resource in this era. Due to these advantages, R. M. Chang et al. (2014) stated that bigdata support greater realism in social science research in that researchers are now able to understand subjects from “the digital traces of their behavior without asking them questions” (p. 74) and no longer have to architect artificial settings which only mimicking the real world.

(2) Computational methods for empirical data

In traditional research, research data are extracted from “scarce, static, clean and poorly relational data sets” (Kitchin, 2014, p. 2) while bigdata has characteristics of “abundance, exhaustivity and variety, timeliness and dynamism, messiness and uncertainty, high relationality” (Kitchin, 2014, p. 2). Hence, the process of analysis is different from what has been done in traditional research. The type of bigdata is normally categorized as structured, semi-structured and unstructured (Gandomi & Haider, 2015). Structure data has standardized format, mostly presented in a table in spreadsheets or relational databases. It is relatively easy for machines to read and conduct analysis when a given data are structured. However, Cukier (2010) estimates that structured data are only a small subset of bigdata, approximately, around 5% of all existing data. Then, the others consist of semi-structured and unstructured data. Semi-structured data are in the middle ground between structured and unstructured data, which are not confined into a table, yet contain machine-readable marker and (Gandomi & Haider, 2015). Lastly, unstructured data refer to the data that are lack of pre-defined format so that they are the most tricky ones for machines to manipulate (Gandomi & Haider, 2015). Text, images, audio, video, etc. are the typical examples of unstructured data that need more sophisticated processing for analysis. Therefore, in using bigdata in social science research, cooperation with computer science that can handle different types of data is essential.

In fact, the meetings of social science and computer science have existed before the era of bigdata. Computational Social Science (CSS) is one of the examples. The history of CSS dates back to the late 1960s and its advent came while theoretical and methodological advances were demanded by frontiers across disciplines in academia (Cioffi-Revilla, 2017). Prominent social science theories such as Field Theory, Functionalist Theory, Conflict Theory, the Theory of Groups, Political Systems Theory, Decision-making Theory etc., required “new formalisms that could treat conceptual and theoretical complexity of human and social dynamics, beyond what could be accomplished through systems of mathematical equations solved in closed form" (Cioffi-Revilla, 2017, p. 36). CSS has developed with the introduction of statistical software and programming languages for several decades, and its importance had been recognized focusing on agent-based models and simulation of complex social systems until 2009 when it was significantly redirected by (Lazer et al., 2009). Lazer and his colleagues highlighted the situation at the time when large-scale data analyses had already been conducted in places like Google, Yahoo and the National Security Agency in the United States, and expressed their concerns that computational social science works could become the almost exclusive domain of private companies or government agencies. Hence, what they argued was that CSS should engage in empirically analyzing bigdata which reflects social phenomena in detail, rather merely focusing on simulation of theory. From then, massive amount of social data has been analyzed in different social science, such as computational communication science, computational sociology, computational political science, etc. applying various Artificial Intelligence (AI) techniques of computer science. Among them, machine learning is an indispensable technology in data analysis (N.-C. Chen, Drouhard, Kocielnik, Suh, & Aragon, 2018).

Machine Learning (ML) is a set of methods that makes a machine find rules in empirical data which are assumed to be drawn from an independent and identically distributed set (Rudin, 2015), and then applying these rules to unseen data for the purpose of prediction. Rules range from simple statistical models to ensembles of hundreds of algorithms (Siegel, 2013). As can be seen from the definition, in fact, the philosophies between social sciences and machine learning are quite different. ML methods aim to obtain knowledge from data, whereas social science research make hypotheses from already obtained theories for test data. Also, while ML works are focused on prediction, social science research are more about to understand why such social phenomena have occurred. Nevertheless, social science can benefit greatly from ML (e.g., N.-C. Chen et al., 2018; Hindman, 2015; Rudin, 2015; Zhang, Wang, Xia, Lin, & Tong, 2020). For example, in quantitative research, ML has been used to support the weakness of traditional regression analysis (Hindman, 2015). Such as decision trees, Support Vector Machines (SVMs), dimension reduction techniques, etc. allows social science researchers to distill bigdata with numerous variables to infer causal relationships or relationships between latent variables, or to predict unseen data (N.-C. Chen et al., 2018; Hindman, 2015). Meanwhile, in quantitative studies, the ML methods can be used for qualitative coding of text data such as media content or interview scripts. Coding refers to the process of assigning descriptive or inferential labels to chunks of data, and linking them to facilitate formulation of meaning and explanation (N.-C. Chen et al., 2018; Lawrence, 2003; Miles & Huberman, 1994; Tierney, 2012). Qualitative coding is a labor intensive and time-consuming task in that researchers examine data in detail and find relevant or potential points of interest for assigning labels (N.-C. Chen et al., 2018; Liew, McCracken, Zhou, & Crowston, 2014). Manually coding the entire bigdata is challenging for social scientists, yet ML methods can help machine to automatically code the data by learning from pre-coded sets and predicting the remaining unseen data. Even when there is no pre-coded data, machines can perform clustering analysis by finding the similarities between documents.

(3) Paradigm shift in social science

Machine Learning, which uses as many as hundreds of algorithms to determine the best model or explanation, is a radically differentiated approach from traditional social science research wherein a method is decided based on researchers' knowledge of technique and the data (Kitchin, 2014; Siegel, 2013). However, the trend of adopting bigdata and ML methods in academia has been gradually expanding, and scholars, accordingly, are discussing new epistemological approaches to understanding the world. According to Kitchin (2014), new data analytics seek to gain insights "born from the data", rather than testing a theory by analyzing relevant data. In some radical community, even "the end of theory (C. Anderson, 2008)" has declared and new forms of empiricism has come up as suggestion.

"There is now a better way. Petabytes allow us to say: 'Correlation is enough.' ... We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot ... Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all. There's no reason to cling to our old ways"

The above is what Anderson wrote in his article (2008, pp. 2-3). He insisted that the traditional scientific framework that woven around hypotheses and models based on theory is unnecessary in the bigdata era. According to his logic, rich data and cutting edge learning based approaches produce more accurate and actionable results than existing scientific methods of incrementally building knowledge without theoretical justification (Shah et al., 2015). In this way, analyzing causal relationship between variables is ignored and theoretical models are no longer required to be set. Instead, it is claimed that the research methodology should focus on better data and analytical tools (R. M. Chang et al., 2014). The radical bigdata advocates who share the view with him believe that bigdata has made it possible to pursue an extreme form of inductive empiricism (e.g., Dyche, 2012; Prensky, 2009; Siegel, 2013; Steadman, 2013). This perspective has attracted a lot of attention and already accepted in business in significant degree (Kitchin, 2014). Nonetheless, there have been many criticisms against them in academia.

In contrast, the importance of the scientific framework has been constantly emphasized in social science and still in computational social science. For social scientist, the bigdata era is perceived as a new paradigm, not where theory comes to an end (R. M. Chang et al., 2014; Kitchin, 2014; Shah et al., 2015; Zhang et al., 2020). In other words, researchers have fully recognized the impact and opportunities of bigdata, and discussed how social science can successfully involve them while sticking to the scientific framework. As a result, the efforts of understanding social dynamics by incorporating a mode of induction in the research design have been attempted (Kitchin, 2014). This differs from the traditional experimental deductive design, since “it is allowed to generate hypotheses and insights ‘born from the data’ rather than ‘born from the theory’” (Kelling et al., 2009; Kitchin, 2014, p. 6). Yet, it does not intend to make a conclusion through induction, but offers a new mode of hypothesis generation before a deductive approach is employed (Kitchin, 2014). In other words, this is a hybrid process in which existing theories are used to guide the knowledge discovery process rather than simply assuming that all relationships within the data are meaningful, and the theories can be challenged by new findings from bigdata. Scholars believe that this gives many opportunities to discover more interactive factors from social phenomena and investigate existing theories from a whole different perspective (González-Bailón, 2013; Qiu et al., 2018; Schwartz & Ungar, 2015; Zhang et al., 2020).

2.2.4 Digital Media as Bigdata Source

In the previous part, it has seen the efforts of social science to accept bigdata within the scientific framework. Bigdata research within the social sciences is encouraged to contribute to the creation of new hypotheses and re-examine theories rather than reach conclusions from the observation of data. Thus, exploratory studies using bigdata have been conducted in various fields of social science and it cannot be overlooked that the many sets of bigdata used for the studies originate from digital media. Therefore, in this part, digital media data as a source of bigdata will be discussed.

(1) Digitalization

Brennen and Kreiss (2016) defined digitalization as the way many domains of social life are restructured around digital communication and media infrastructures. Digitalization begun in the latter half of the 20th century and keep continuing to the present day (Schoenherr, 2004) while having laid the foundation for bigdata and their further use. Through digital media and technologies, digitalization has already been proceeded in work, production, healthcare, business, human identity, time and space, communication, education, etc., and, eventually in the totality of our everyday life (Roth, Dahms, Welz, & Cattacin, 2019; Wajcman, 2008). Accordingly, massive amounts of digital records and creations are generated, hence, data management architectures are required to store and process them.

With regard to storage, Relational Database (RDB), NoSQL Database, Distributed File System (DFS) are included depending on the type of data. RDB allows to store, modify and manage relational data while NoSQL is for non-relational data. Meanwhile, DFS refers to a set of server service that enables data files to be accessed and shared between different computers through a network (Pokorný, 2015). In addition to this, stored data are processed in a batch or in real time so that researchers can collect and analyze them. Real-time processing includes in-memory computing and data stream processing, while distributed processing includes cloud computing and Hadoop.

As such, large amounts of data derived from human behavior collected through digital media is being stored and processed in databases and cloud servers. And they are to be analyzed for revealing their values according to the intention of researchers.

(2) Digital traces

Howison, Wiggins, and Crowston (2011, p. 769) defined digital tracking data as “records of activity (trace data) undertaken through an online information system (thus, digital)”. This digital tracking data can be collected from various digital media such as websites, social media platforms, smart phone apps, sensors, etc. With these data, researchers can reveal fine-grained behavioral pattern of people over time (Stier, Breuer, Siegers, & Thorson, 2020).

Menchen-Trevino (2013) divided various digital trace data into two main categories which are participation trace data and transactional data. The former is generated from deliberate contributions to online spaces. For example, when people post their articles on online forums, update their status and profile on digital platforms, writing tweets and retweets and give “likes” on Facebook, those activities leave digital traces. This is more related to social media bigdata which will be introduced after. On the other hand, the latter, transactional data, consist of event logs which include a timestamp and metadata about the digital action performed. For instance, by retrieving webpages, clicking on hyperlinks, using search engines, accessing mobile apps, or simply carrying mobile phones with location setting, transactional data are generated and accumulated.

When it comes to participation trace data, they can be extracted online through relatively clear tool and methods. For instance, researchers can collect data via the Application Programming Interface (API) of platforms such as Facebook or Twitter or web crawlers (Stier et al., 2020). One social media API aggregation company, Gnip, Inc, was claimed to track three billion actions per day (Menchen-Trevino, 2013). However, transactional data often the exclusive property of the digital service provider so that researchers need to cooperate with the companies that allow use of such data (e.g., Bond et al., 2012; Hampton, Goulet, Marlow, & Rainie, 2012). Otherwise, researchers can gather transactional data by incentivizing study participants to install plug-ins on their digital media or handing out sensors such as GPS and movement trackers or wearable badges (Stier et al., 2020).

(3) Social media bigdata

Social media is a set of Internet-based applications to describe content that various users can continuously modify through participation and collaboration, rather than created and published by individuals alone (Kaplan & Haenlein, 2010). The content from social media have contributed to the creation of bigdata extensively across multiple online platforms or websites (Ghani et al., 2019; Schroeder, 2016).

As social media covers a wide range, it is difficult to list their types, yet D. M. Boyd and Ellison (2007) referred to them as "social network sites" in a narrow sense and suggested their three aspects. First, users can create public or semi-public profiles. Second, users are allowed to connect to others and form a network. Last, they can relate to other users and the activities publicized in their network. The typical examples of social network sites that generate bigdata are Facebook, Twitter, Instagram, LinkedIn, Wiki, YouTube etc. (Ghani et al., 2019).

Meanwhile, Aichner and Jacob (2015) deal with social media in a wider sense in their article, which not only encompass social network sites but also creation sharing sites, collaboration websites, discussion forums, microblogs, etc. The table 4 below shows their descriptions and examples of each type of social media.

Table 4

The descriptions and examples of each type of social media

Type of social media	Description	Examples
Blogs	A blog (from "web" and "log") is a chronological list of postings, which can be read and commented upon by visitors.	The Huffington Post (huffingtonpost.com)

Business networks	Individuals use business networks to establish and maintain professional contacts. Registered users create a personal profile and share personal details such as the type and duration of their education, professional experience and expert knowledge.	LinkedIn (linkedin.com)
Collaborative projects	Collaborative projects bring together internet users with a common interest and/or certain knowledge in order to plan, develop, improve, analyse and/ or test technological, academic, scientific or fun-oriented projects.	Wikipedia (wikipedia.org)
Forums	A forum is a virtual discussion platform where users can ask and/or answer other users' questions and exchange thoughts, opinions or experiences	Gaia Online (gaiaonline.com)
Microblogs	Microblogs restrict the length of postings to approximately 200 character that may also include pictures or weblinks.	Twitter (twitter.com)
Photo sharing	Photo-sharing websites offer services such as uploading, hosting, managing and sharing of photos.	Flickr (flickr.com)
Products/ services review	Product and service reviewing websites sell and provide information about products.	Amazon (amazon.com)
Social bookmarking	Social bookmarking describes the concept of saving and organizing internet bookmarks at a centralized platform in order to share them with friends and other users.	Pinterest (pinterest.com)
Social gaming	Social games are online games that allow or require social interaction between players.	World of Warcraft (warcarft.com)
Social networks	Social networks connect people that know one another, share common interests or would like to engage in similar activities.	Facebook (facebook.com)
Video sharing	Video-sharing platforms allow users to upload and share personal, business or royalty-free videos and to watch them legally.	YouTube (youtube.com)
Virtual worlds	Virtual worlds are populated by many users who can create a personal avatar, and simultaneously and independently explore the virtual world, participate in its activities or communicate with others.	Second Life (secondlife.com)

Note. Adapted from Aichner and Jacob (2015)

2.2.5 Literature about youth online help seeking applying bigdata approach

Previously in the chapter 2.2.4, the literature on youth online help seeking, which obtained their research data through questionnaires or interviews or focus group were presented. In this section, however, another type of studies that apply a bigdata approach will be introduced. For a rich review, some studies using bigdata of phone call and text messages are also included, considering that they still are mediated communication different from face to face interactions and the call logs and the text messages were stored in databases for further analysis. Through the review, interestingly, it is found that there are some characteristics of these bigdata studies that are consistent with what WHO proposed, that is studying youth help seeking in a way of therapeutic narratives and social marketing. The details will be described below in accordance with research data, research content and methods.

(1) Research data that show real narratives of the youth

The data collected in the literature are thousands to hundreds of thousands of online texts that are written by young people. They are not responses to specific surveys or interviews, but voluntary questions or requests posted when they need help. These data had been anonymously submitted to online media and were extracted for the purpose of research taking into account privacy and ethics issues. In the most of papers, the purity of data was emphasized, which means that they were not created by the researchers' measuring tools, so that contain the real narratives of the youth. For example Harvey, Locher, and Mullany (2013) pointed out that existing self-report methodologies used to get young people's knowledge and opinion may decline respondents to provide answers or replace earnest answers with some random replies. On the other hand, the authors emphasized the benefit of using naturally-occurring data, stating that they represent genuine contemporary issues of young people, arguably give more insights than are possible through artificially-constructed methods such as survey and interviews (Harvey et al., 2013).

The data from the reviewed literature were collected from online websites, online forum communities, Crisis Text Line (CTL), peer-supported youth hotline and social questioning and answering (Q&A) services. The texts are written with the clear purpose to obtain help assuming communication with virtual counselor(s) or peer(s). Unfortunately, there seems yet to be no studies using log data of youth health information seeking through search engines. Online websites are run by governments or nonprofit organizations and provide services for the youth to consult with professionals about health or mental health related issues. For instance, some of studies used the data from a British governmental adolescent health website, Teenage Health Freak[1]. The online forum communities are where people with common interests gather and share their opinions. One example of online forums in the literature is reddit.com which contains over 11,400 active user-created communities. Similarly, social Q&A is a service which allows people to ask and answer questions on any topic in everyday life such as Yahoo! Answers (Oh & Park, 2013). CTL is an international nonprofit organization that provide free 24/7 support via mobile text messages for people in crisis. When the young send text messages to given numbers, volunteer counselors respond in real time. The service is available in the United States, Canada, United Kingdom and Ireland. A hotline service is similar to CTL, and this has traditionally offered counseling service by phone call. Now, some of hotlines have also included text and email services.

To sum up, the data used in these literature consist of texts written by young people in person with a specific purpose to seek help. This is somewhat consistent with therapeutic narratives what WHO recommended for future youth help seeking studies. Just like therapeutic narratives method aims to listen to what patients describing about history of problem, the progression of the need and seeking help for the need, the online sources such as online websites, online forum communities, social Q&A, hotline and CTL services have same role of inviting the young to reveal innermost feelings.

(2) Research content focusing on youth issues and user segmentation

It has seen that previous studies have less interest in content of youth problems or difference between demographic groups of the young, rather focusing on motivating and obstructive factors for them to help seeking online. The ones applying bigdata, however,

---

[1] http://www.teenagehealthfreak.org

explored who are looking for which kinds of help. While there are studies that analyzed the overall help seeking content of young people (e.g., Gray, Harvey, Macfarlane, & McPherson, 2008; Harvey et al., 2008), there are also more specialized studies dealing with major mental health problems, that is, eating disorders (e.g., Moessner, Feldhege, Wolf, & Bauer, 2018), self-harm (e.g., Brookes & Harvey, 2016), anxiety (e.g., Thompson, Sugg, & Runkle, 2018), depression (e.g., Brookes & Harvey, 2016; Harvey, 2012; Thompson et al., 2018), suicidal thoughts (e.g., Szlyk, Roth, & García-Perdomo, 2020; Thompson et al., 2018) and sex health (e.g., Carr & Bednarek, 2019; Harvey et al., 2013; Oh & Park, 2013).

There are also studies that segment the young and see the differences among them in regards of help seeking. For instance, a study by Thompson et al. (2018) focused on demographic and socioeconomic characteristic of the young in the United States and examined the relationship with their help seeking especially on mental health. They used about 850,000 structured data from CTL records, which consist of phone numbers of the youth who sent the messages and 28 crisis issues from the pre-established list that counselors have classified at the end of each conversation. Since phone numbers include county codes, they attempted to explore the geographic patterns in help seeking between the young across the rural-urban continuum. Applying spatial error regression, they found that rurality was the strongest predictors for low help seeking rates, also suggested that young people living in rural areas were slow to adopt CTL, despite high suicide rates (Thompson et al., 2018). Another study (Kerner et al., 2020) used the data from a peer support hotline service in Los Angeles, and examined the number of contacts according to user demographics; gender and age, hotline contact types, and content of issues. As a result, at the demographic segment level, they found that most of the contacts were made by girls aged 15 and 16 years old. At the whole level, anxiety and stress turned out the main causes of contact, and found that more than 30% use text messages instead of calling to the hotline. Unfortunately, they did not analyze the relationships between the demographic factors and the contact types or the content of issues, and presented this as one of their limitations of the study. The other study conducted also in the United States conducted by Szlyk et al. (2020) used suicide related data from a US-based branch Crisis Text Line. From the data, firstly, the authors clustered the young into five subgroups based on their frequency of access to the service which is the total number of text conversations on the same phone number and conversation number for frequent text users. Afterwards, classifying psychosocial problems in their text, and using a

Latent Class Analysis (LCA), the authors identified the five groups into three classes who are in lower distress, anxious distress and relational distress.

These practices seem similar to how customer data are used in social marketing presented in section 2.2.2. For instance, companies actively make use of customer data including their demographics, locations, lifestyles, etc. for conducting customer segmentation. In addition to this, they analyze customer reviews about their products and services to grasp opinions and needs of customers. This trend has also found in such literature and this shows the realization of the other suggestion from WHO that is perceiving young people as consumers and applying social marketing strategies for youth help seeking.

(3) Corpus linguistics (Text mining) and Machine Learning as research methods

Since most of the research data in the literature were online text written by young people, the analyses often aimed to find meaningful patterns within the text through some specialized methods namely Corpus Linguistics or Text Mining. The language that humans use in everyday life is called "natural language" and there have been attempts to analyze natural languages with computers in different disciplines. The representative ones are Computational Linguistics in linguistics and Natural Language Processing (NLP) in computer engineering. The research areas they are dealing with have many similarities and are not thoroughly distinguished, but it is generally evaluated that Corpus Linguistics has contributed to the development of natural language processing. NLP is not only used to analyze huge amounts of written and spoken language data based on linguistic knowledge provided by computational linguistics by applying artificial intelligence methods such as machine learning and deep learning, but also develop conversational agents such as chatbot and voice assistants. Thus, in social sciences, Corpus Linguistics rather than NLP has been referred to as a discipline that can contribute insights into the analysis of text data (Pollach, 2012). Meanwhile, while Corpus Linguistics has a strong academic character, nowadays text mining is better known as an analysis tool. Text mining adopts both Corpus Linguistics and NLP technologies, and is used in a more practical aspect. This is for both quantitative and qualitative examinations of text data. To be specific, quantitative analyses of texts are used as a means of context-specific qualitative interpretation, so that it is possible to approach insights about how people conceptualize and discursively construct their subjective experiences and understandings of certain social phenomena (Brookes & Harvey, 2016).

Some of the literature on youth help seeking with bigdata approach are from a British academic community that combines Corpus Linguistics and Health communication. They often use keyword analysis, collocational analysis and concordance analyses that are looking for themes from keywords, finding relationship between them and patterns of use in situ (Harvey et al., 2008). For instance, Harvey et al. (2008), using anonymous emails sent by British young people from Teenage Health Freak, a website of nonprofit organization for youth helping, found words that occur together with important verbs and nouns related to help seeking such as “tell”, “ask”, “talk”, “answer”, “advice”, “question”, “advise” and “explain”. Based on this, they conducted further analysis and figured out that it is articulated in the emails that why adolescents having difficulty in sharing problems with health professionals. One of their findings is that young people expressed concerns over confidentiality especially on intimate, sensitive issues, particularly sexual health issues. The table 4 shows their content words that collocated with help seeking keywords.

Table 4

Harvey et al. (2008)’s Content words that collocate with keywords concerning help seeking

Keywords	Content words
Tell	Abortion, afraid, alone, boyfriend, dad, doc, doctor, doctors, eating, embarrassed, friends, family, gay, go, GP, leave, lose, mates, mum, need, nipples, normal, older, parent, parents, people, person, pill, police, raped, really, risk, said, scared, see, shy, stressed, teacher, things, thinks, trust, truth, try, wondering
Ask	Advice, afraid, boy, boys, courage, dad, doctor, embarrassed, find, friend, friends, girl, girls, help, mum, parents, people, question, really, scared, stupid, sorry, thought, wanted
Talk	Afraid, boyfriend, dad, doctor, family, feel, find, friend, friends, girls, hard, mates, mum, parents, phone, people, scared, school, shy, stressed, tell, time, told, wants
Answer	Dr, email, getting, give, know, need, really, please, question, questions, worried, wrong

Advice	Ask, get, give, good, like, need, please, really, talk, thanks
Question	Ask, asking, answered, doctor, Dr, health, help, know, name, need, really, reply, sent, stupid, worried
Advise	Give, help, need, please
Explain	Know, please, hard

Note. Reprinted from Harvey et al. (2008)

Furthermore, there are several studies using Machine Learning methods. For instance, Moessner et al. (2018) who studied peer to peer help seeking and giving behavior with the data that were written by teens who have eating disorder from online community, employed several unsupervised machine learning methods. First, through topic analysis with Latent Dirichlet Allocation (LDA) model, they found 9 topics related to social support and eating disorder specific content. Second, by conducting social network analysis, they attempted to describe the overall communication patterns and identify the most influential users. Lastly, a linear network autocorrelation model was applied to estimate associations in language among network neighbors (Moessner et al., 2018). Another example is conducted by Szlyk et al. (2020). It has already been introduced that they use LCA for clustering the young, but this is not Machine Learning methods. Still, their data were extracted by a Machine Learning model. To be precise, they analyzed the data which the model had identified as suicide related out of the total CTL text data. The model had been programmed to predict suicide risk by reading variability in text messages understanding context.

2.2.6 Summary of chapter 2.2

Section 2.2 was aimed at discussing how bigdata can be new an approach to supplement the limitation of youth help seeking research and development of a help intervention program. Thus, the concept and characteristics of bigdata were introduced first, and it was pointed out that bigdata should be defined not by quantity but by its value. As seen earlier, in business, the value of bigdata has been fully supported in terms of being a means of decision making. Various examples of bigdata analytics were investigated at the marketing level, and this was to grasp idea of applying social marketing-style methods to youth help seeking research which had been mentioned by WHO at the end of the section 2.1.

In academia, however, especially in social sciences, there exist disagreements on the acceptance and utilization of bigdata for research. This is because bigdata differs greatly from the existing philosophy of social science. Thus, the differences were described in the aspects of data characteristics, methodology and epistemology. While the extremes argued for conclusions based on data without concerning theoretical explanation as business does, social science scholars generally criticize it. They recognize the benefits that social science can gain from embracing bigdata, yet state that the research should stick to scientific procedures. As a result, a hybrid data-driven research process that combines induction and deduction was introduced. In other words, bigdata research is expected to establish new hypotheses and re-examine theories through inductive analysis.

It was noted that the source of bigdata used in research came from digital media. In other words, digital media can now play its role as the source of bigdata. Bigdata in social science studies reflect human behaviors, which are collected in various forms through digital media. Appropriate storage and processing techniques for different data types have been developed, and researchers are now able to access these digital traces according to their research purpose. The types of digital traces are broadly described in two ways. One is participation trace data, which are the by-product of certain activities such as creation, empathy, and sharing. The other is transactional data, which are automatically aggregated data without intention of record. They consist of event logs including time and location coordinate of the connection and activities. In recent years, the number of studies using digital traces of various social media to understand their content and interactions has increased.

Lastly, literature on youth seeking help applying bigdata approach were reviewed. This was implemented for comparison with the literature using traditional methodology such as surveys and interviews for youth online help seeking presented in 2.2.4. Despite the difficulty of elaborate contrast, it was seen that the bigdata literature were able to reach a variety of youth issues and differences among the young that previous studies could not include. Specifically, taking advantage of the "naturally occurring" characteristics of bigdata, youth issues were extracted from the texts written by actual young people themselves. Furthermore, the literature using bigdata showed the difference of help seeking behaviors between the young, by relating demographic data. This is in line with therapeutic narratives and social marketing methods recommended by the WHO when they pointed out the limitations of the existing youth help seeking research. Even though there was no comment on bigdata, their recommendations have realized in the bigdata era. This is also related to advances in corpus linguistics and text mining, which are used as methods for analyzing vast amounts of text data. These methods enable quantitative and qualitative analysis of texts, and help researchers analyze content in a short time and apply various statistical models and machine learning models.

It is believed that the theoretical backgrounds and literature reviews presented throughout the chapter 2 provided evidence for youth help seeking to be studied through bigdata. Thus, this dissertation will be described as one of the exploratory attempts, such as those done in a heuristic way in previous bigdata literature on youth help seeking. And it will present results obtained through bigdata analysis and discuss the limitations and opportunities in using bigdata for youth help seeking research. So, chapter 3 begins by introducing the research questions and explains methods and data used for the research.

3. Research Questions and Methodology

3.1 Research questions

This study aims to investigate how to approach youth help seeking with bigdata. To be more specific, by using youth digital traces related to help seeking from online media, and analyzing them with computational tools, it is attempted to supplement what traditional help seeking studies have not reached before. This dissertation belongs to Nordic media study, and accordingly, the subjects are Norwegian youth and bigdata from Norwegian media are used. To the best of my knowledge, there is no youth help seeking study conducted using bigdata in Norway. Yet, Lassemo, Sand, and Tøndel (2020) have recently published their article about mapping 6,823 questions from LGBT young from a Norwegian youth support website, to identify "how violations of norms related to sexual orientation, gender identity, gender expression and gender characteristics". This dissertation also uses the same online source, but includes text data generated by the entire youth, not a specific group. In addition to this, their time stamps and demographic information are also analyzed. The following is the official research questions for this dissertation.

<The 1st research question>	"How can bigdata be used to understand youth online help seeking?"
<The 2nd research question>	"What are the opportunities and challenges when using bigdata in media research?"

Conventionally, both quantitative and qualitative methods have been used in youth help seeking studies. As discussed throughout chapter 2, each method such as survey, interview, and focus group interview has its own advantages, but in terms of pre-designed measurement tools, they have difficulties in catching the real aspects of the youth, and to avoid social desirability bias. And the more a study is planned to take a deep look into one's perception and experience, the more limitations follow regarding resources and the number of samples. Meanwhile, text mining has been acknowledged as a method to cope with both quantitative and qualitative analysis and to supplement these limitations (e.g., Ampofo et al., 2015; Baker et al., 2008; Wiedemann & Wiedemann, 2016). In other words, this allows researchers to analyze massive amount of unstructured text data statistically by using meanings and relationships of words. In computational social science, text mining has been gradually adopted as a digital version of traditional research methods or developed as a brand-new research method for bigdata analysis (Ignatow & Mihalcea, 2016). In this dissertation text mining method will also be applied for analyzing Norwegian youth query documents. Therefore, in the next subsection, the general process of text mining will be firstly introduced.

3.2 Text mining

3.2.1 Text mining definition and the basic unit of analysis

Text mining also called as text analytics refers to a method that extract information from textual data by computational work. Textual data are typical types of unstructured data which are represented by human natural language that computer cannot read directly. Therefore, text mining naturally involves Computational Linguistics or Natural Language Processing (NLP), which facilitates for a computer to understand semantic representation from human language by using linguistic, statistical concepts, and Artificial Intelligence techniques such as Machine Learning (ML).

There are different textual data extracted from digital media, such as social network feeds, emails, blogs, online forums, survey responses, corporate documents, news, customer reviews etc. (Gandomi & Haider, 2015). From these texts, when a set of documents is collected for an analysis, this is called a corpus. Corpus includes several documents, and a document consists of paragraphs. Again, these paragraphs are made up of sentences, and a sentence contains words. Words are also made up of several morphemes. In other words, texts have a hierarchical structure of corpus, documents, paragraphs, sentences, words, and morphemes. Most text mining studies use word as the basic unit of analysis. Word can be also called as “term” or “token” in some cases.

3.2.2 Text mining process

Finding meaningful knowledge in a collection of text documents requires a series of steps. The following figure 3 shows the general process of text mining.

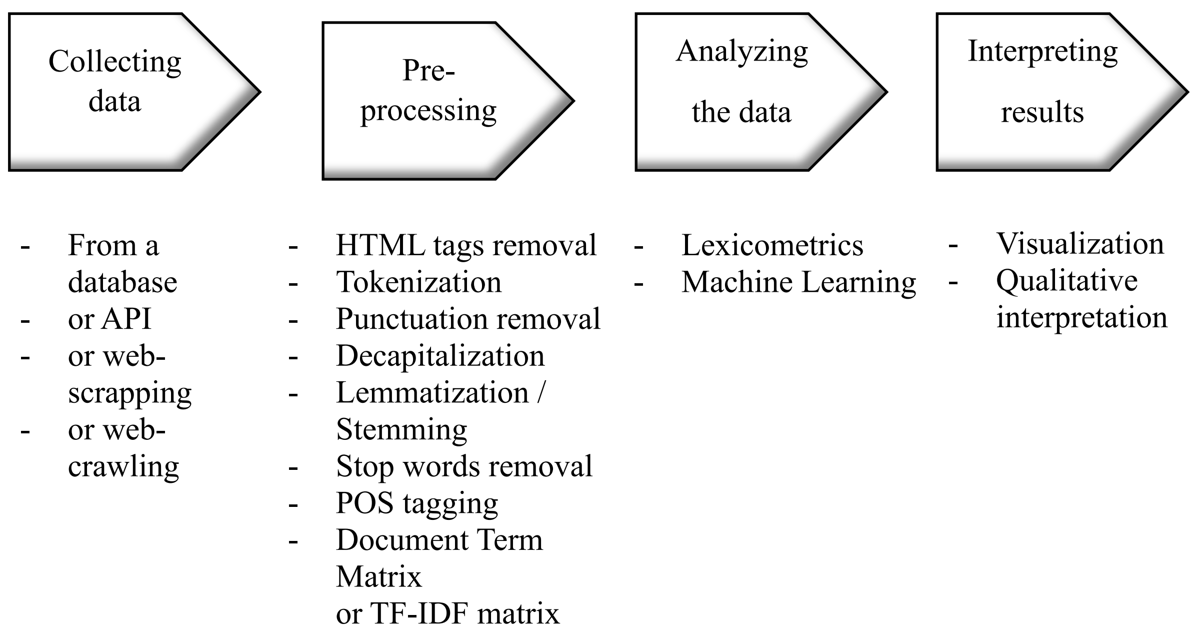


Figure 3. The process of text mining

(1) Collecting Data

In the first step, the documents that contain the desired information need to be identified and collected. Data can be obtained from internal database or from external sources. In the latter case, one might get an open Application Programming Interface (API) that domain makes public their internal data. Certain content from webpages can be also collected through web scrapers or web crawling.

(2) Preprocessing

As text documents themselves cannot be analyzed directly with computer, preprocessing should be preceded. To be more specific, preprocessing includes a wide range of tasks from getting rid of unnecessary tags from corpus to representing words in form of numbers so that mathematical, statistical and ML algorithms can be applied for analysis (Miner et al., 2012).

Collected data may include not only text but also some tags from such as HTML (Hypertext Markup Language) which is made to display documents in a web browser. They are not necessary for analysis hence it is better to be removed. As mentioned above, the basic unit of text mining is word. Therefore, it is required to break a document down to a word level. This task is called tokenization or word tokenization. Words are equivalent to tokens in this case, yet one could also use sentence or paragraph as token. Each document is planned to be converted into a matrix so called Document Term Matrix (DTM) by counting frequency of word tokens for facilitating calculation for analysis. Yet, before making a DTM, many techniques can be applied to volume down the document and extract most necessary words to be efficiently counted. For instance, one might get rid of all punctuation marks and reduce word variability by decapitalization, and lemmatization or stemming. By removing punctuation marks, only alphabet and numerical tokens will remain. Decapitalization can be implemented to alter the tokens with capital letters to be lower-cased. In this way, same words with different characters cannot be double counted. Lemmatization or stemming is finding a root of words. For instance, the words “connected” and “connects” have “connect” as their stem, and “am”, “are” and “is” have “be” as their lemma. Using stemming or lemmatization, different forms of words can be presented as the same one word while preserving their essence meaning. In addition to this, specific words which frequently appear but not carrying any meaning can be also removed. Those words are called stop words. Text mining software have their own lists of stop words for each language. In English such as “a”, “the”, “is”, etc., and In Norwegian “en”, “dette”, “er”, etc. treated as stop words. Researchers can add more words to the stop words list to better analyze. At this point, the word tokens are ready to be counted with their frequencies and a DTM can be made, yet one might identify their Part Of Speech (POS). POS is “a category to which a word is assigned in accordance with its syntactic functions” ("Part Of Speech," 2020). The main POS are noun, pronoun, adjective, verb, adverb, preposition, etc. By tagging POS of tokens, for instance, a matrix only having noun tokens or/and verb tokens can be created according to the purpose of the analysis.

Through the preceding process, each document is composed of necessary words (or tokens or terms), and it is to be changed into a numerical form. In other words, a DTM is generated, which describe the frequency of all the word terms that occur in a whole collection of documents. The following figure 4 simplifies way of creating a DTM.

Total terms from the documents

Documents

	it	is	puppy	cat	pen	a	this
it is a puppy	1	1	1	0	0	1	0
it is a kitten	1	1	0	0	0	1	0
it is a cat	1	1	0	1	0	1	0
that is a dog and this is a pen	0	2	0	0	1	2	1
it is a matrix	1	1	0	0	0	1	0

Figure 4. The principle of generating DTM. Given 6 examples documents, a DMT is created by taking each word token out from the documents and counting its frequency. This is adapted from (Zheng & Casari, 2018).

Meanwhile, there are methods to count term frequency more effectively for making a matrix. One of them is using TF-IDF statistics (Luhn, 1957). TF-IDF is the product of Term

Frequency and Inverted Document Frequency. The term frequency (TF) refers to the frequency of a word in a document, while the document frequency refers to the number of documents the word occurs in. Dividing document frequency by total number of documents in the collection then applying log function, inverse document frequency (IDF) is calculated. By multiplying TF with IDF, each word obtains different TF-IDF scores in given documents. The higher the score, the more important the word in a particular document is. In other words, TF-IDF algorithm is used to spot those words that discriminate documents from each other (Antons, Grünwald, Cichy, & Salge, 2020). In this way, one can create a matrix that reflects the importance of words beyond simply counting their frequency. The following figure 5 shows the formula of TF-IDF calculation.

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j
df_i = number of documents containing i
N = total number of documents

Figure 5. TF-IDF formula. Reprinted from Gopalakrishnan and Venkateswarlu (2018)

(3) Analyzing the data

There are various types of analysis for extracting meaningful information from textual data. The following table 5 summarizes them by categorizing into two different computational methods which are Lexicometric and Machine Learning (ML). Lexicometric techniques are related to the frequency with which words occur in text has been used from the beginning of computational text processing, and its algorithms and models have been extended from simple counting to statistical method (Wiedemann & Wiedemann, 2016). Meanwhile ML is the study of computer algorithms that allow computer programs to automatically improve through experience (Mitchell, 1997). It means that a computer itself improves its performance by updating model parameters with new units of observed data (Wiedemann & Wiedemann, 2016). Broadly, there are two types of learning methods in ML, that are supervised and unsupervised. In supervised ML, certain external knowledge related a document can be used to model the association between that knowledge and features of the documents. It is analogous to manual coding in a content analysis procedure (Wiedemann & Wiedemann, 2016). For instance, with supervised ML, unknown documents can be classified to a pre-defined category. Unsupervised ML, in contrast, is to identify hidden structures derived from data themselves. Since there is no external knowledge given, unsupervised ML uses statistical approaches (e.g., method of moments) that provide a clustering of data points satisfying certain familiarity criteria among data (Wiedemann & Wiedemann, 2016).

Table 5

Various types of text mining categorized with lexicometric and ML methods

Lexicometrics	Frequency analysis	• Counting specific terms or concepts terms in documents • Long-term comparison of trends in word frequency may reveal peaks and dips in term usage, and corresponding concepts in a domain (Wiedemann & Wiedemann, 2016).
	Key term extraction	• Identifying important terms that describe the subject of a document by applying statistical measures (Archer, 2009)
	Co-occurrence analysis	• Finding terms that frequently occur together from documents

<table>
<tr><td rowspan="6">Machine Learning</td><td colspan="2">Unsupervised</td></tr>
<tr><td>Document clustering</td><td>Grouping documents according to similarity of their content
“Clusters should have the property of optimal similarity of documents within the cluster and maximum difference of documents between clusters” (Wiedemann & Wiedemann, 2016).
K-means, hierarchical, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) are the most common algorithms</td></tr>
<tr><td>Topic models</td><td>Topic modeling refers to a set of “algorithms for discovering the main themes that pervade a large and otherwise unstructured collection of documents” (Blei, 2012).
Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF) are the common algorithms for topic modeling.</td></tr>
<tr><td colspan="2">Supervised</td></tr>
<tr><td>Classification</td><td>Using pre-defined categories provided from document external knowledge, assigning these categories to given data</td></tr>
<tr><td>Information Extraction</td><td>Identifying key phrases and relationships within text by pre-defined knowledge (Gaikwad, Chaugule, & Patil, 2014)</td></tr>
</table>

	Sentiment analysis	• Identifying positive, neutral, or negative attitudes from texts (Pang & Lee, 2004)

Note. Adapted from Wiedemann and Wiedemann (2016)

(4) Interpreting the results

As Kelle (1997, § 5.7) states, “none of these steps can be conducted with an algorithm alone. In other words, at each step the role of the computer remains restricted to an intelligent archiving system, the analysis itself is always done by a human interpreter”. Therefore, after analyzing text data, the results are to be visualized and interpreted. Visualization tools simplify discovered patterns and trends by using graphs, word cloud, network, colormaps, etc. These not only increases the readability for readers but also helps researchers to capture the important aspects from a huge amount of information and interpret them in accordance with their research purpose.

3.3 Research procedure

The research proceeds according by following the text mining process introduced in the section 3.2. There are timestamps and demographic information along with text and their use will be explained in analysis part after preprocessing text data. The entire research procedure is summarized in the following figure 6.

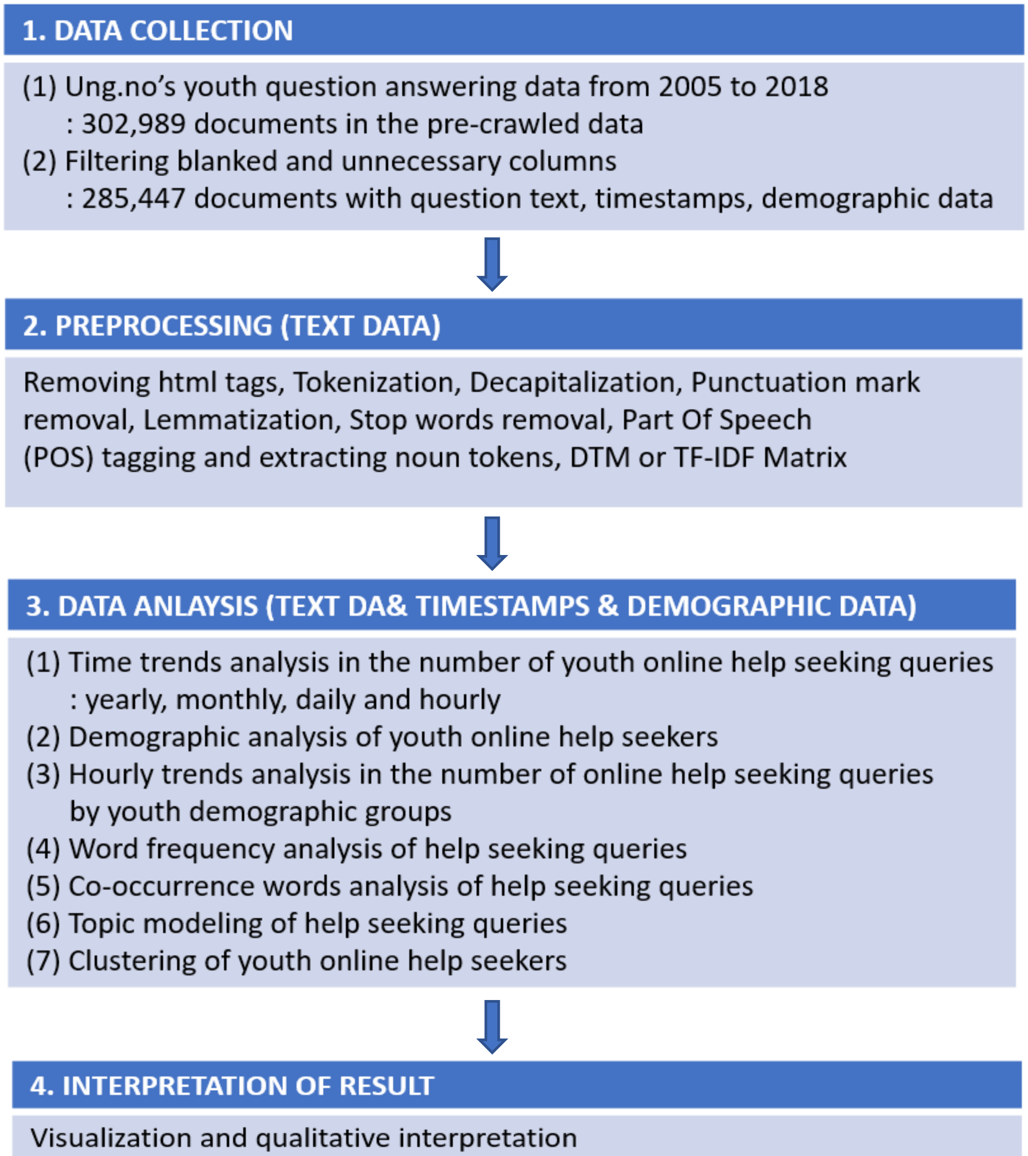


Figure 6. The summary of the entire research procedure

3.3.1 Data collection

Data used for this research had been collected from Ung.no, a Norwegian public help providing website for the young. Through the author's engagement in Social Health Bots[2] research project, it was possible to access pre-crawled data which had generated during 2005 to 2018 from Ung.no's question answering service. Any Norwegian young person can use this service to make queries regarding their state of mind and get answers from youth professionals. This help seeking and giving are anonymously proceeded, and youths do not need to provide any personal information, but there are selectable options of gender and age. The following figure 7 shows a screenshot of the webpage when young people making questions.

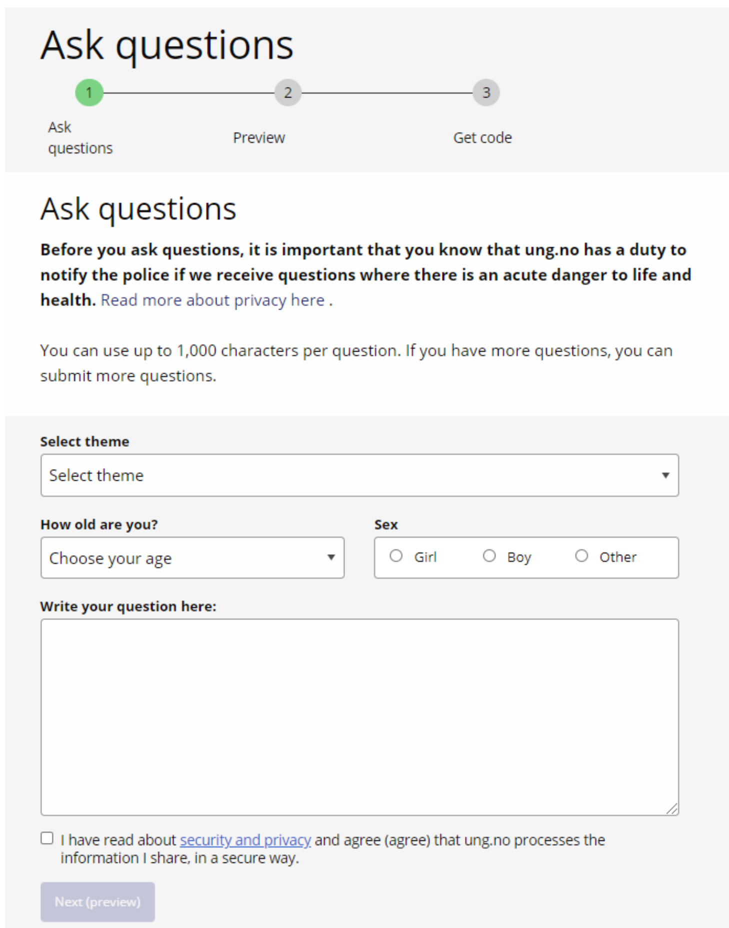


Figure 7. A webpage of help seeking service Ung.no(https://www.ung.no/still-sporsmal/).

[2] https://www.sintef.no/en/projects/socialhealthbots/

The page is translated from Norwegian to English by google translation.

Recently data protection has been actively discussed with the notion of the bigdata and its use. Governments around the world have adjusted existing regulations or established new ones. For example, as of May 2018, EU enforces a set of new data protection rules not only for restraining companies to abuse personal data but also safeguarding the right of citizens in the EU, called the General Data Protection Regulation. In Norway, the Norwegian Data Protection Authority (Datatilsynet) is in charge of this. The data source of this research, Ung.no runs as a public service under the supervision of Datatilsynet. At the same time, there is no personally identifiable information asked, furthermore, young people are suggested to read articles about security and privacy before submitting questions. Given the facts, private information security in this research is considered to be well guaranteed. In this reason, this research has proceeded without any reports or deliberations.

The given pre-crawled excel file includes total 302,989 documents with event logs, user generated texts and so on. The excel columns consist of “ID”, “Received time”, “Received date”, “Gender”, “Age”, “Question”, “Answer”, “Short version of question” and “Answer checked”. The first three “ID”, “Received time”, “Received date” and “Answer checked” are administrative information. “Gender” and “Age” are demographic information of the youth. “Questions” and “Answers” contain the content of what young users asked and received regarding their concerns. “Short version of question” is a summary of each question. Among them, only necessary columns are selected for this research, which are “Received time”, “Received date”, “Gender”, “Age” and “Question”. In addition to this, all documents including any blank space in such columns are deleted. Through this process, a total of 285,447 documents were collected for this study, and the period was reduced to about 10 years from 2008 to 2018. The following table 6 illustrates the composition of the selected data.

Table 6

The composition of the given data and selected parts

Digital trace	Item	Description	Selected
Administrative information	ID	• Automatically generated code when a question is asked. • This is used for the young to check answers	No
	Received time	• Recording of time when inquiry entered to the system. • This is presented in form of HH:MM:SS	Yes
	Received date	• Recording of date when inquiry entered to the system • This is presented in form of YYYY-MM-DD	Yes
	Answer checked	• Confirmation whether answers are checked or not	No
User information	Gender	• Gender of the youth • The category consists of boy, girl and other	Yes
	Age	• Age of the youth • The category consists of 0, 12 to 20	Yes
Counseling content	Questions	• Free text question written by the youth	Yes
	Answers	• Free text answer written by youth experts	No
Etc.	Short version of question	• Summary of questions	No

3.3.2 Preprocessing texts

In the given data, the content of "Questions" is written in free text. Then, it is necessary to preprocess them. To do this, a Natural Language Processing (NLP) software library, spaCy[3], applying Norwegian Bokmål language model[4].

[3] https://spacy.io/
[4] https://spacy.io/models/nb

A series of tasks for preprocessing from removing unnecessary parts to creating TF-IDF matrix is described below. To better explain, one example document will be presented, and it will show how it is changed every step of preprocessing.

(1) Removing html tags

Some of documents still include html tags which are the leftover from web scrapping or crawling. They are removed using regular expression (regex). In the figure 8, it is shown that how an example document including html tags such as “/r” or/and “/n” is changed.

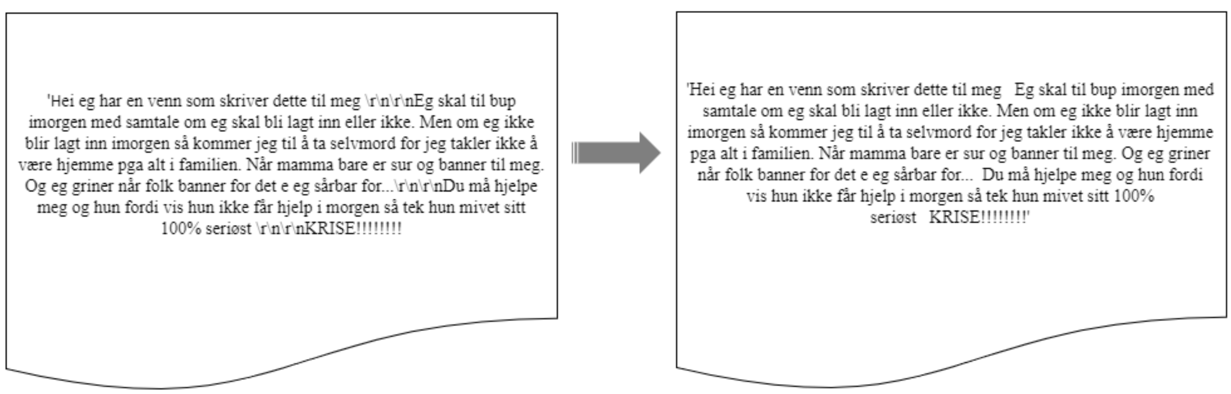


Figure 8. Change of the example document before and after getting rid of html tags

(2) Tokenization

Through tokenization, each document is split at the word level. The figure 9 below shows how the example document becomes word tokenized.

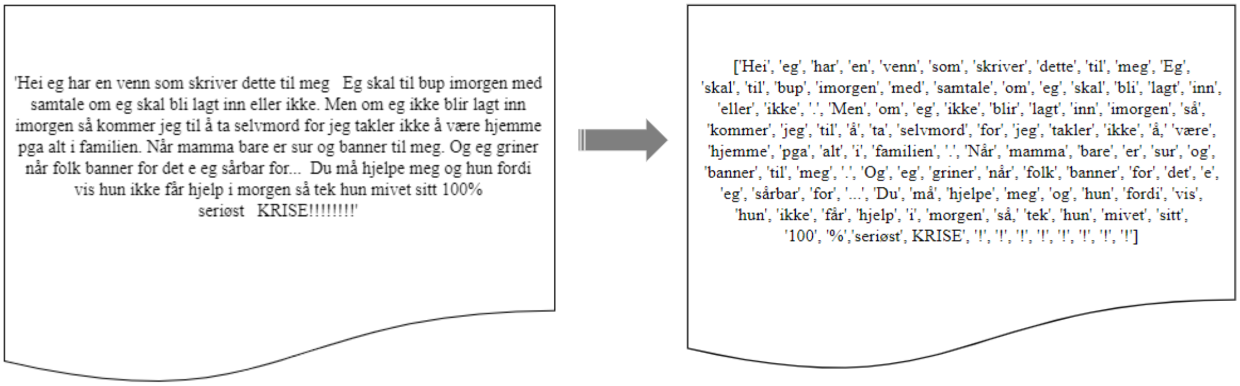


Figure 9. Change of the example document before and after tokenization

(3) Decapitalization and punctuation mark removal

Capital letter in documents are changed into their form in lower case through decapitalization. Furthermore, punctuation marks such as ".", "?", "!", ":", ";", ")" are also removed as it is seen from the figure 10.

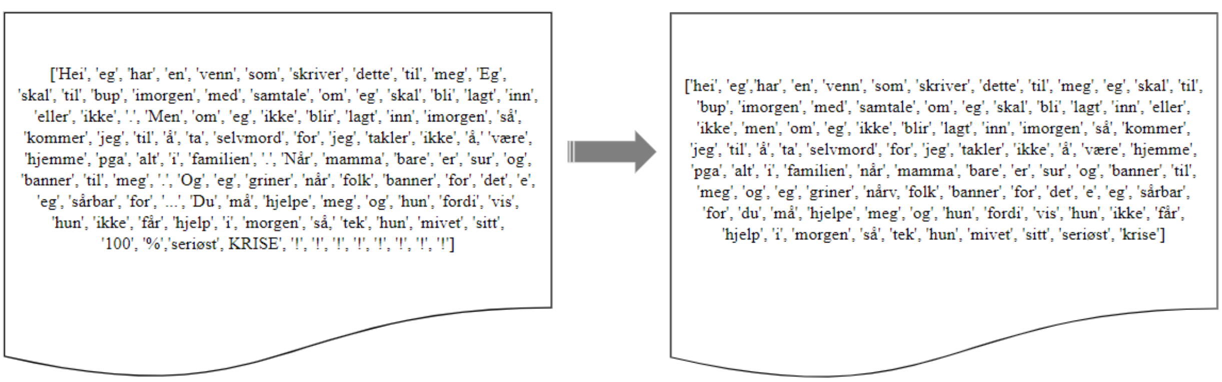


Figure 10. Change of the example document before and after decapitalization and punctuation mark removal

(4) Lemmatization

In this research, lemmatization is implemented instead of stemming. Then, words expressed in various ways considering its singular/plural types, definite/indefinite articles, past/present/future forms, etc. are returned to their essential forms. The figure 11 presents the change of example document after lemmatization.

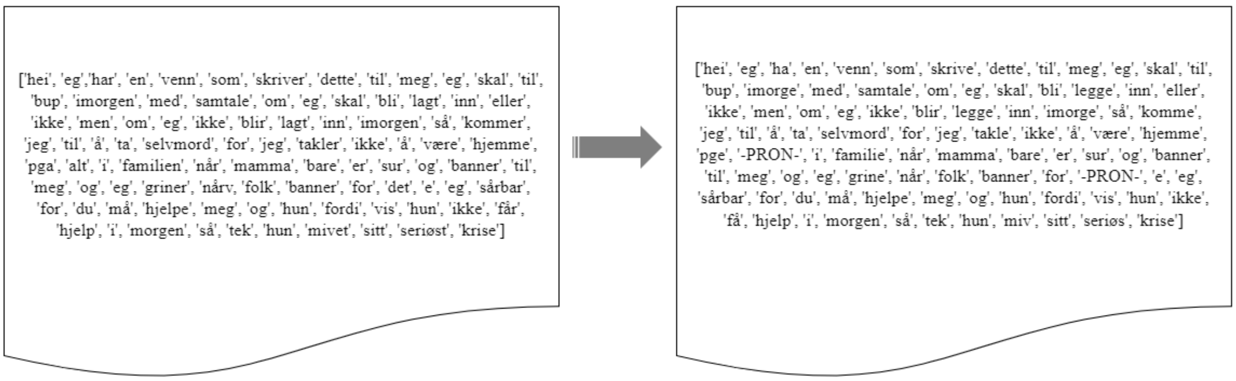


Figure 11. Change of the example document before and after lemmatization

(5) Stop words removal

The stop words originally included from a spaCy's Norwegian stop word list are filtered out. In addition to this, manually designated stop words by the author(myself) are also removed. For instance, "hilsen (Regards)" and "takk (Thank you)" are some of the most common words that appear at the end of questions, but they are not important enough to be involved for analysis. The figure 12 shows the example document after getting rid of stop words. The whole stop word list is attached in the appendix of the dissertation.

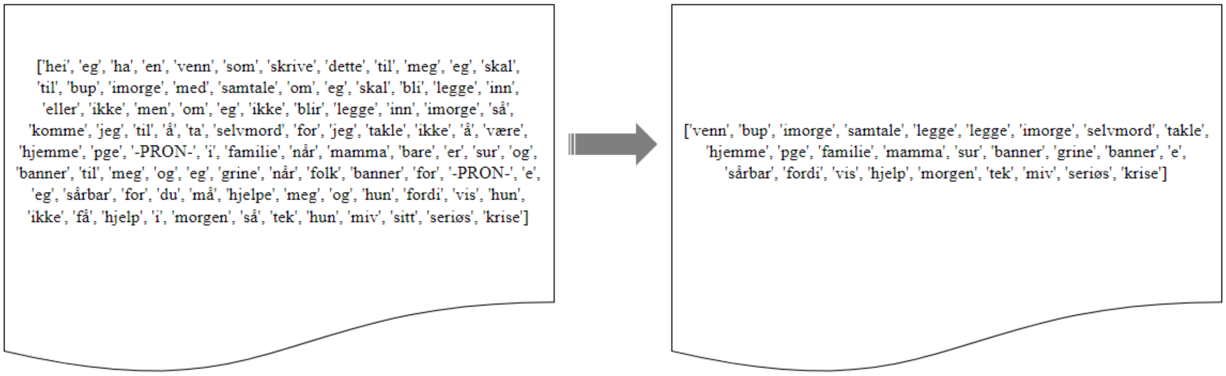


Figure 12. Change of the example document before and after stop words

(6) Part Of Speech (POS) identification and extracting noun tokens

Some analyzes include the entire POS such as noun, pronoun, adjective, verb, adverb, preposition etc., while for the other analyzes only tokens that identified as noun are extracted. In this way, the size of the documents becomes smaller, still containing main tokens which are good index for finding topics etc. In the figure 13, noun tokens from example document are presented. This shows the big difference between the original and the preprocessed document.

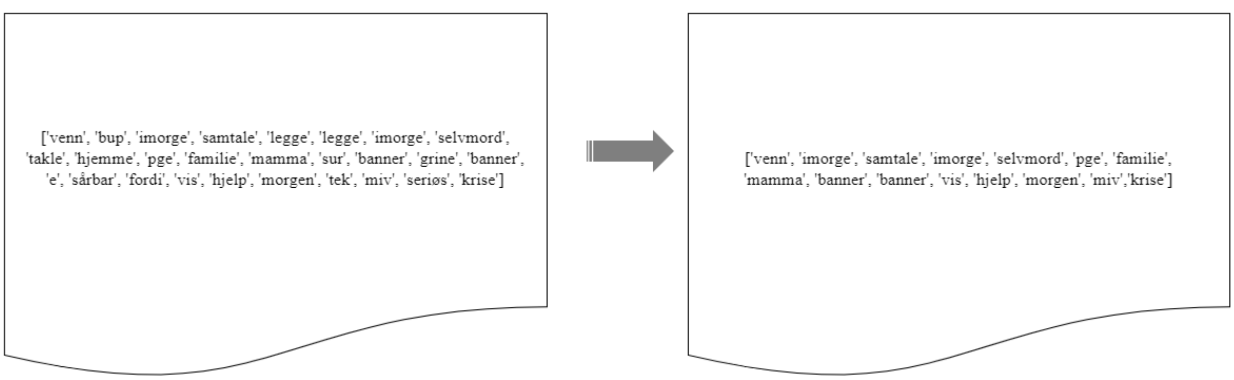


Figure 13. Change of the example document before and after extracting noun tokens

(7) DTM and TF-IDF Matrix

Depending on the analyses, in some cases a DTM is generated, in other cases a matrix with TF-IDF score is calculated. DTM expresses the frequency of each word, a TF-IDF matrix presents the importance of each word, facilitating more sophisticated semantic analysis. By representing words and documents into DTM and TF-IDF matrix, further statistical and machine learning models are to be applied.

3.3.3 Analyzing the data

This part describes analyzing the data which include not only the text, but also demographic information and timestamps. There are total seven analyses, and the table 7 below shows them. As a tool for analysis, python, a programming language, is used. Python allows researchers to analyze bigdata through a variety of packages regarding to statistic, NLP, ML and visualization. The applied packages are also presented in the same table.

Table 7

The summary of seven analyses and corresponding methods, data and applied analytical packages

	Analysis	Methods	Data	Analytical packages
1	Time trends analysis of help seeking behavior	statistics	Timestamps	• Pandas[5]
2	Demographic analysis of youth help seeker	statistics	Demographic	• Pandas
3	Hourly trends analysis of help seeking by demographics	statistics	Timestamps & Demographic	• Pandas

[5] https://pandas.pydata.org/

4	Keyword frequency analysis of help seeking content	Text mining (lexicometrics)	Question text	• Pandas • NLTK (nltk.FreqDist[6])
5	Co-occurrence words analysis of communication keywords	Text mining (lexicometrics)	Question text	• Pandas • NLTK (nltk.collocations[7])
6	Topic modeling of help seeking content	Text mining (Unsupservised ML)	Question text	• Pandas • Sklearn (TfidfVectorizer[8]) • Gensim (LdaMallet[9], CoherenceModel[10])
7	Youth clustering by help seeking content	Text mining (Unsupservised ML)	Question text & Demographic	• Pandas • Sklearn (TfidfVectorizer, Kmeans[11])

(1) Time trends analysis in the number of youth online help seeking queries: yearly, monthly, daily and hourly

In the first analysis, using timestamps generated when young people made questions, it will be analyzed how the number of help seeking queries changed. The largest unit is based on year and it is counted how many help seeking cases are occurred every year from 2008 to 2018. Furthermore, by gradually narrowing down the level of units to month and day it is to investigate if there are particular periods when the youth need more or less help. Lastly, the count of help seeking queries are analyzed on 24-hour basis to see its patterns for a day. Each result will be expressed in a compressed way through line and pie graphs.

[6] http://www.nltk.org/api/nltk.html?highlight=freqdist
[7] https://www.nltk.org/_modules/nltk/collocations.html
[8] https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html
[9] https://radimrehurek.com/gensim/models/wrappers/ldamallet.html
[10] https://radimrehurek.com/gensim/models/coherencemodel.html
[11] https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html

(2) Demographic analysis of youth online help seekers

From the total documents with 28,5447 questions, gender and age of help seekers will be identified. As seen from the screenshot of ung.no's question answering service in the chapter 3.1.1, there are 3 attributes in gender category which are boy, girl, and other. In the case of age, there are ten attributes consisting of the age of 12 to 20 and 0 who did not reveal their age. Through this second analysis, each proportion of gender and age groups are counted. This allows a relative comparison of which population groups are more active in seeking help through ung.no and show who are the main users of the help seeking service. The age and gender share are calculated as a percentage and visualized through pie graphs.

(3) Hourly trends analysis in the number of online help seeking queries by youth demographic

Mixing timestamps and demographic data, it will be analyzed whether there are differences in the time for seeking help by gender and age. The level in time is set to the smallest one which is 24-hour basis for a day. Through this, it is possible to approach hourly help seeking behavior of a relatively minority population groups, for example, including "other" gender and the ones who entered their age as 0. Results of each gender and age are presented together in same line graphs for better comparison.

(4) Keyword frequency analysis in help seeking content

From the fourth analysis, text mining methods will be applied. As a type of lexicometrics analysis, keyword frequency will be examined that is to find out which words appear the most in the entire help seeking query corpus. For this, the corpus will go through the whole preprocessing steps presented in the chapter 3.3.2 from getting rid of html tags to filtering stopwords and then only noun tokens will be selected to count their frequencies. To do this, Natural Language Toolkit (NLTK[12]) package from Python will be used which includes a class of frequency distribution computing the top N most frequently used words from text data.

Total top 50 words keywords will be extracted from the help seeking content from ung.no and they will be displayed in a table and visualized through WordCloud which provides an overview of corpus by distilling them down to several words and reflects their frequencies with font size. This is considered as one of the most straightforward and powerful visualization libraries in text analytics (Heimerl, Lohmann, Lange, & Ertl, 2014).

(5) Co-occurrence words analysis of communication keywords

As another type of lexicometrics analysis, co-occurrence terms will be found that used together with the certain words suggested in Harvey et al. (2008)'s research. In the section 2.2.5, as one of literature, their work was introduced. They designated some communication related terms such as "tell", "ask", "talk", "answer", "advice", "question", "advise" and "explain" to investigate sources, content and limitations of youth online help seeking. In this study, the same words will be used and also all the parts of speech (POS) will be included as Harvey et al. (2008) did. Yet, there will be more sophisticated manipulation required. For example, NLTK's collocations package is used to find a bigram which is "a two-word sequence of words (Jurafsky & Martin, 2018)" like "ask my[13]", "explain about", or "difficult talk". In given examples, a bigram of "ask" is "my", that of "explain" is "about", and that of "talk" is "difficult". However, the words such as "my" and "about" do not give any clues of help seeking sources or content, while the word "difficult" may denote difficulty of communication. To alleviate this problem, the size of window is to set as five. Window size is one of parameters when finding bigram, which indicates the number of words spanned by a collocation. The default value is set as 2 in NLTK's collocations package and this means that It is only can be detected consecutive word to a designated word. For instance, in a sentence of "ask my mom but difficult", when window size is two, the bigram of "ask" is only "my" which is right next to "ask". However, if window size is five, non-contiguous bigrams such as "mom", "but", "difficult" will be also counted. According to Church and Hanks (1990, p. 23), larger window sizes has benefits "to highlight semantic

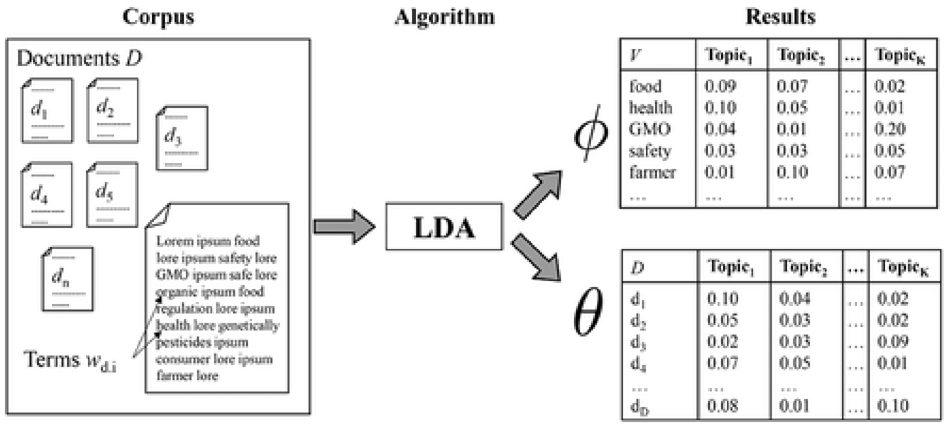
[12] https://www.nltk.org/
[13] It is assumed that there is a sentence, for example, "ask my mom".

concepts and other relationships that hold over larger scales". Accordingly, for this co-occurrence word analysis, the window size is set to five. Moreover, bigrams of eight terms will be presented in a table in order of their frequency.

(6) Topic modeling of help seeking queries with LDA

The sixth analysis is to identify topics of help seeking from the query data written by Norwegian young people. For this, an unsupervised ML method, topic modeling is applied. Topic modeling infer potential topics that penetrate the entire corpus in a way that forms groups of words with similar meaning based on their statistics. There exist several statistical algorithms that can be used for topic modeling such Latent Semantic Analysis (LSA), probabilistic Latent Semantic Analysis (pLSA) and Latent Dirichlet allocation (LDA), etc. Among them LDA is applied in this study, since it was introduced for outperforming LSA and pLSA (Alghamdi & Alfalqi, 2015) and as a result, it becomes standard in text mining research especially in field of communication (Hong & Davison, 2010; Maier et al., 2018). Following figure 14 shows the principle of LDA explained by Maier et al. (2018).

For applying LDA to the help seeking query corpus, pre-processing steps up to stopword removal will be conducted and only the noun words are extracted. Afterwards, they are converted into TF-IDF matrix to fit into LDA algorithm. Furthermore, LDA will be implemented using Python's "Gensim" library. Gensim provides APIs related to various topic modeling and among them, "wrappers.LdaMallet" is used for this study, that is wrapped to utilize Java-based MALLET (MAchine Learning for LanguagE Toolkit). LdaMallet is slower than Gensim's standard LDA model based on Variational Bayes sampling method, but yield more accurate result by using an optimized Gibbs sampling algorithm (Yao, Mimno, & McCallum, 2009).



We assume that each document, *d*, in a corpus can be described as a probability distribution over topics. This distribution, called θ_d (the topic distribution of document *d*), is drawn from a Dirichlet distribution with prior parameter α (which must be chosen by the researcher).
Thus, each topic can be defined as a probability distribution over the entire corpus vocabulary, i.e., all the different words that appear in the documents. More technically, for each topic k, we draw ϕk, a distribution over the V words of the vocabulary from a Dirichlet distribution with prior parameter β (which must be chosen by the researcher).
Within each document (d = 1, …, D) and for every word in that document (i = 1, …, Nd), in which i is the index count for each word in document d and Nd is the total length of d, we sample:

a) a topic (zd,i) from the respective topic distribution in the document (θd), and
b) a word (wd,i) from the respective topic's word distribution ϕk, in which k is zd,i, the topic we sampled in the previous step.

Figure 14. the principle of LDA. Reprinted from Maier et al. (2018)

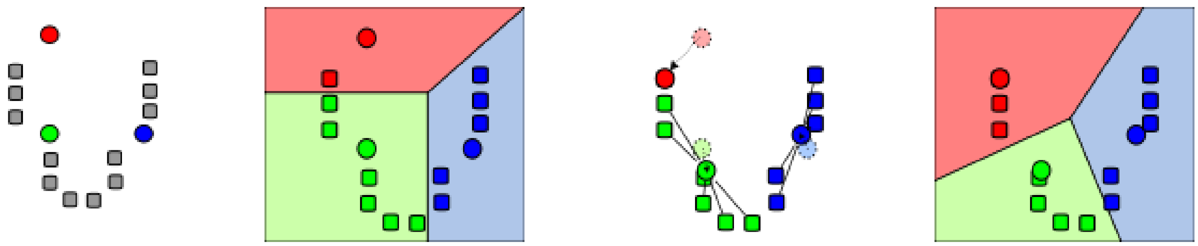
Meanwhile, in topic modeling, it is important to select an appropriate number of topics as the results vary depending on how many topics are assigned. LDA, however, cannot determine itself the number of topics so that researchers should set the number of topics and perform the analysis. In general, after researchers setting a random default number, the process of testing and adjusting is repetitive until a result that allows a reasonable interpretation for each topic is obtained. Yet, there are also measurement tools to set the optimal number of topics which are perplexity and topic coherence. Perplexity is a measurement of how well a probabilistic model describes a dataset (Zhao et al., 2015). By learning for each set number of topics, it is possible to select the optimized one when it is found the section with the lowest value[14]. It is, however, often not correlated with human judgements when it comes to verifying topic quality (J. Chang, Gerrish, Wang, Boyd-Graber, & Blei, 2009), so topic coherence will be used for this study which is known as an improved alternative (Mimno, Wallach, Talley, Leenders, & McCallum, 2011; D. Newman, Lau, Grieser, & Baldwin, 2010). Topic coherence score a single topic by measuring "the degree of semantic similarity between high scoring words in the topic" (Stevens, Kegelmeyer, Andrzejewski, & Buttler, 2012, p. 954). The optimal number of topics can be selected when coherence score is high, since a higher score indicates a more similar word pair in a topic (Stevens et al., 2012) and eventually, they describe topic better. Therefore, in this study, each topic coherence score assuming the number of topics from 1 to 20 will be calculated. By finding a section with the highest score, the topic modeling is proceeded. Afterwards, the author will name topics by comprehensively considering words and contexts constituting each topic. The results, finally, will be presented in a table. In addition, distances between topics are reduced and visualized in quadrants of intertopic distance map using Python's "pyLDAvis" library.

(7) Help seeking youth clustering with K-means

Lastly, an unsupervised ML clustering analysis will be conducted using demographic information of the youth and help seeking question data. To be more specific, after classifying the whole question corpus by gender and age, total 30 documents will be generated, since there are three types of gender and ten types of age. The 30 documents will be processed and only noun tokens will be taken and converted into a TF-IDF matrix. Then, K-means is applied for grouping those close to each other based on their similarity of content. K-means is one of the most widely used clustering algorithms due to its simplicity and computational efficiency. Its principle is to group data based on cluster center point (centroid) closest to data. Initially, the centroid is randomly set, but keeps changing until it minimizes the average distance between

[14] Perplex means "inability to deal with or understand something". Therefore, lower perplexity indicates better probabilistic model.

data and their centroids (Singh, Tiwari, & Garg, 2011). The figure 15 below shows a demonstration of how K-means organize clusters.



STEP 1

k initial "means" (in this case *k*=3) are randomly generated within the data domain (shown in color)

STEP 2

k clusters are created by associating every observation with the nearest mean.

STEP 3

The centroid of each of the *k* clusters becomes the new mean.

STEP 4

Steps 2 and 3 are repeated until convergence has been reached.

Figure15. The demonstration of K-means clustering. Reprinted from Reprinted from Demonstration of the standard algorithm, In Wikipedia, Weston.pace, Retrieved November 30, 2020, from https://en.wikipedia.org/wiki/K-means_clustering

However, similar to LDA model in topic modeling, K-means cannot determine itself the number of clusters. To alleviate this problem, researchers can use the elbow method to find the optimal number, K, of clusters. Elbow method is “a method which looks at the percentage of variance explained as a function of the number of clusters” (Bholowalia & Kumar, 2014, p. 18). By plotting a graph with certain range of K values calculated their Sum of Squared Error (SSE), if the graph looks like an arm, then the “elbow” which is the point of inflection on the curve, is where the SSE decreases the most. Accordingly, this indicates the optimal value of K ("Elbow Method," 2020). The following figure 16 shows the formula of SSE, where Y_i correspond to the centroid and $\hat{Y}_i$ correspond to the data point.

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Figure 16. The formula of Sum of Squared Error

In this study, SSE will be calculated within the range of K values up to 10, by confirming the elbow through the graph, clustering with an optimal number of k will be performed. Afterwards, the result will be visualized using "Principal Component Analysis (PCA)" from Python's "sklearn[15]" library, which reduces the distance between documents and clusters in two dimensions. Furthermore, the top 30 words of each cluster will be presented in a table with the interpretation of how each clustered youth can be different in terms of help seeking content.

3.3.4 Interpreting the results

The results for each analysis are presented in the next chapter, including visualization and interpretation. Meanwhile, applied visualization packages for the analyses are described in the table 8 below.

Table 8

The summary of visualization packages for each analysis

	Analysis	Representation	Visualization packages
1	Time trends analysis in the number of help seeking queries	• Graph • Pie chart	• Matplotlib (pyplot[16])
2	Demographic analysis of youth online help seekers	• Pie chart	• Matplotlib (pyplot)

[15] https://scikit-learn.org/stable/
[16] https://matplotlib.org/api/pyplot_api.html

3	Hourly trends analysis in the number of help seeking queries by demographic groups	• Graph	• Matplotlib (pyplot)
4	Word frequency analysis in the content of help seeking queries	• Table • Graph • WordCloud	• WordCloud[17]
5	Co-occurrence words analysis of communication keywords	• Table	N/A
6	Topic modeling of help seeking queries	• Intertopic distance map	• pyLDAvis (pyLDAvis_gensim[18])
7	Clustering of online youth help seekers	• Graph • Table	• Sklearn (PCA[19])

[17] https://amueller.github.io/word_cloud/index.html
[18] https://pyldavis.readthedocs.io/en/latest/modules/API.html
[19] https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

4. Research Result

4.1 Time trends analysis in the number of youth online help seeking queries

(1) Yearly

Using the whole dataset, it is counted how many help seeking queries were made by Norwegian youth during the October of 2008 to December of 2018. As it is seen in the figure 17, the number of queries gradually increased in general. There were exceptionally many requests in 2009, and in 2017, the number increased by about 2.25 times compared to the previous year.

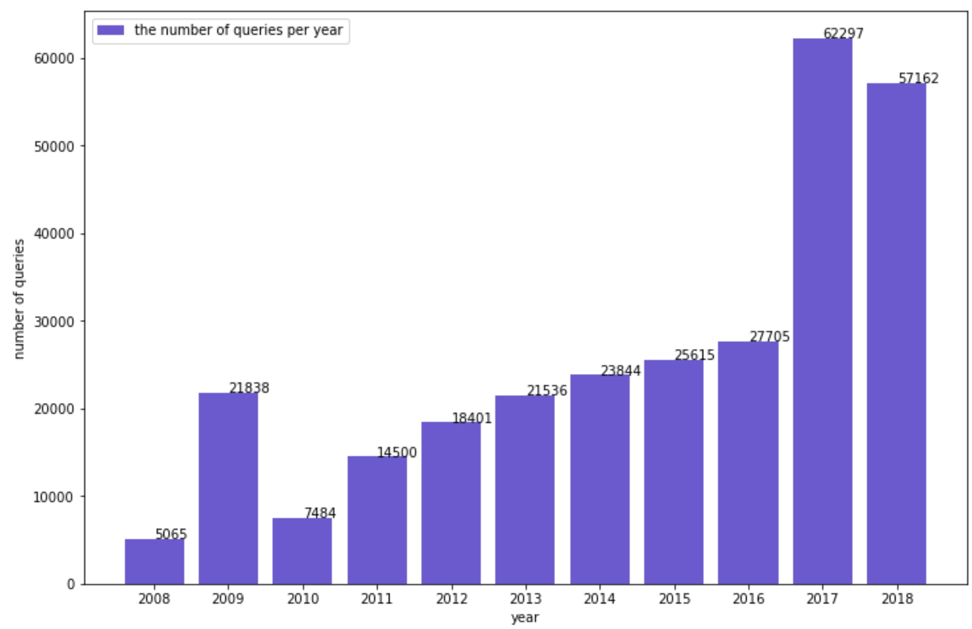


Figure 17. the number of queries per year during 2008-2018

(2) Monthly

Counting the number of youth online help seeking queries by month, it is found that July and December are the months that the Norwegian youth submitted help seeking queries least.

This can be related to the Norwegian yearly school plan in which having vacation during June to July, and Christmas holiday in December. The graph in the figure 18 shows the tendency and the number of queries aggregated by each month. In contrast with the vacation periods of June to July and December, the number of help seeking queries in the months before and after the vacation, such as January, May, and November are when young people asked for support most. In particular, January is the month of yearly highs.

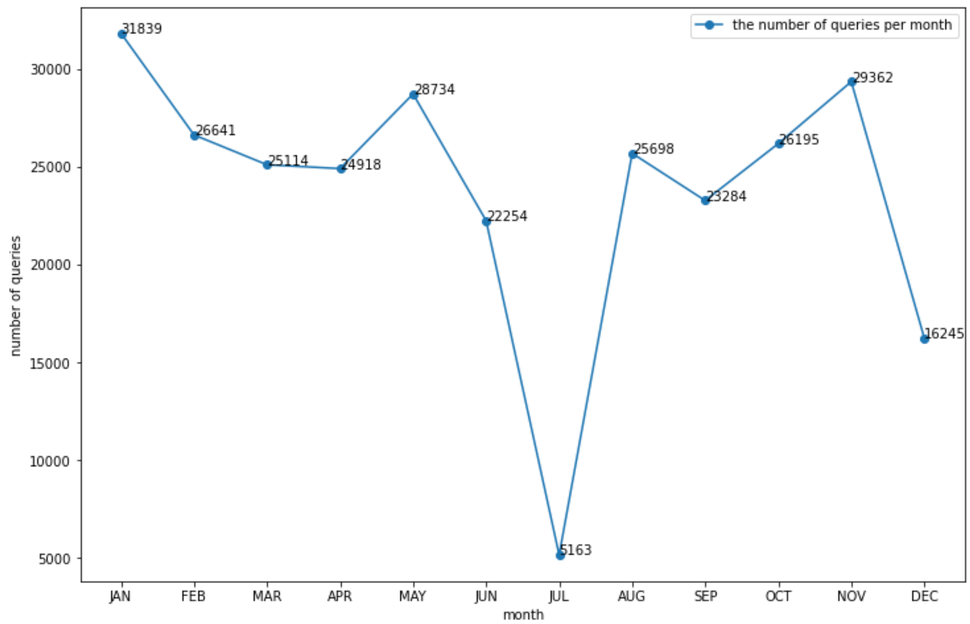


Figure 18. the number of queries per month

(3) Daily

Counting on a daily basis, Norwegian youth online help seeking through ung.no tends to be occurred slightly more in the first half than the the second half of month. When counting the number of queries from the 1st to the 15th and the 16th to the 31st, the former is about 1.08times higher than that of the latter. The number of help seeking dropped sharply by the end of the month.

During the week, meanwhile, the queries of help seeking were generated on Mondays the most and the number of them continually decreased until Saturdays, when the fewest are written.

On Sundays, however, Norwegian young people seek for help 1.37 times more than on Saturdays. Daily tendency is described in the figure 19 while the figure 20 illustrates the number of queries by day of the week.

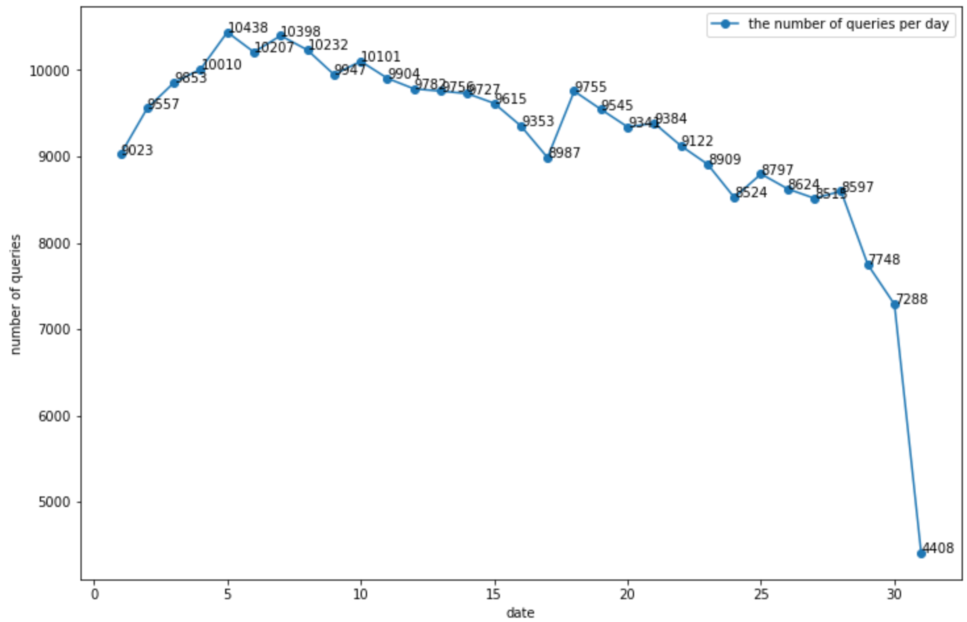


Figure 19. The number of queries per day

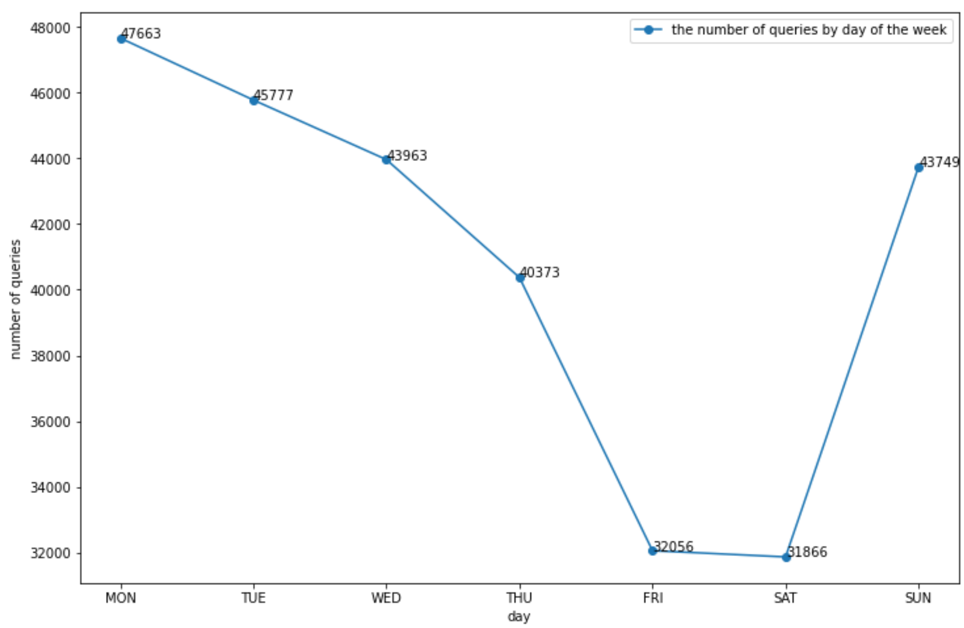


Figure 20. The number of queries by day of the week

(4) Hourly

For each hour of the day, the number of help seeking behavior through online queries were counted. As a result, it appears that help is most sought at night, with the peak around 22:00. In addition, there are three leaps are shown for a day; from 8:00 am to 9:00 am at the beginning of the day, from 2:00 pm to 3:00 pm after primary school(grunnskole[20]) are finished, and from 8:00 pm to 9:00 pm before the young finishing a day and going to bed. The figure 21 illustrates the information mentioned while the figure 22 shows that approximately 31% of the help seeking queries are made from 20:00 when the last leap begins, to 23:00, just before midnight when the graph draws a downward line.

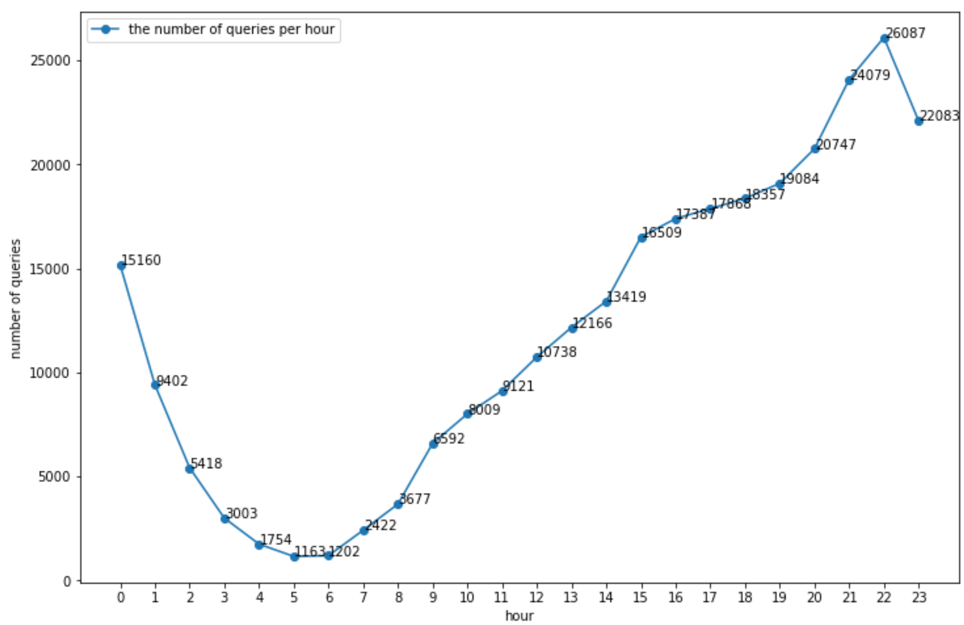


Figure 21. The number of queries per hour during a day

[20] Grunnskole is "a term for the 10-year schooling which in Norway is compulsory for children between 6 and 16 years. (…) The grunnskole comprises the primary school, which consists of 7 grades divided into the primary school level (1st-4th grade) and the intermediate level (5th-7th grade), as well as the junior high school which consists of 8th, 9th and 10th grade.

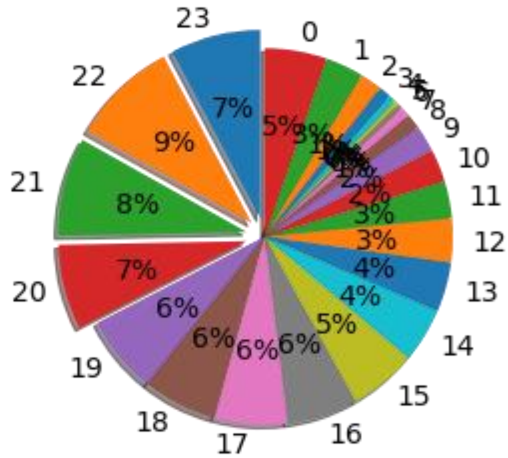


Figure 22. The percentage of queries per hour during a day

4.2 Demographic analysis of youth online help seekers

As explained earlier, the ung.no queries were written by Norwegian young people in various gender and age. The gender types are girl, boy and other while the age types vary from 12 to 20 and 0. Instead of entering a specific age, some young people wrote their age as 0. The following figure 23 shows the proportion of gender who sought help. More than 66% of the help seeking, which is 189,484 queries out of 285,447 total queries are requested by girls while about 33% are from boys. In other words, Norwegian girls used this support services twice more than boys. Furthermore, the young who selected their gender as other were less than one percent of the total help seeking.

Meanwhile, by counting queries by ages, the average age of help seekers of ung.no is found to be 15.64. To be more specific, the average age of girls is 15.48, that of boys is 15.98, and that of others is 15.59. More than 65% of help seekers using this query service is the ones between age of 13 to 16, which are equivalent to junior high school(ungdomsskole) students. Furthermore, the population who asked for help the most is 13-year-old girls with 36,775 queries while there were no queries generated by “other” gender in the age of 12. The figure 24 contains the age share of the help seeking queries and detailed numbers are attached to the appendix.

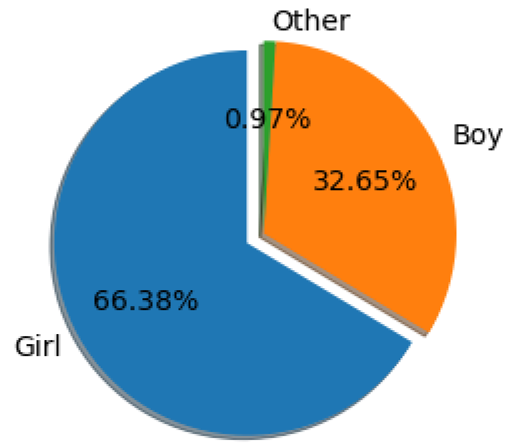


Figure 23. The gender share of the help seeking queries.

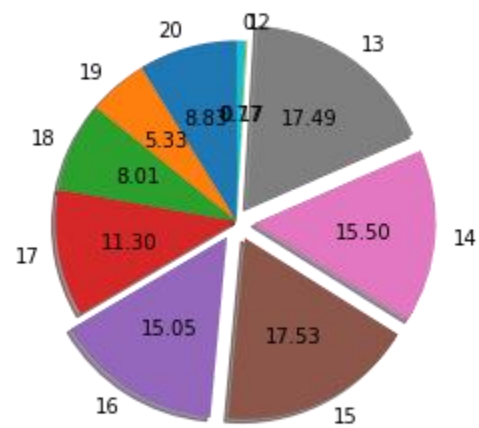


Figure 24. The age share of the help seeking queries

4.3 Hourly trends analysis in the number of online help seeking queries by youth demographic groups

The third analysis was planned to find whether there is a difference in help seeking times during a day depending on age or gender of youth by combining two data timestamps and demographic information. The figure 25 and 26 show the corresponding results.

As already seen in the first analysis, in the whole population level, it is at 22:00 when the most queries were written and there were three leaps where the number of queries drastically increases. However, this analysis found that there are some demographic groups who have a time trend that deviate from the general tendency.

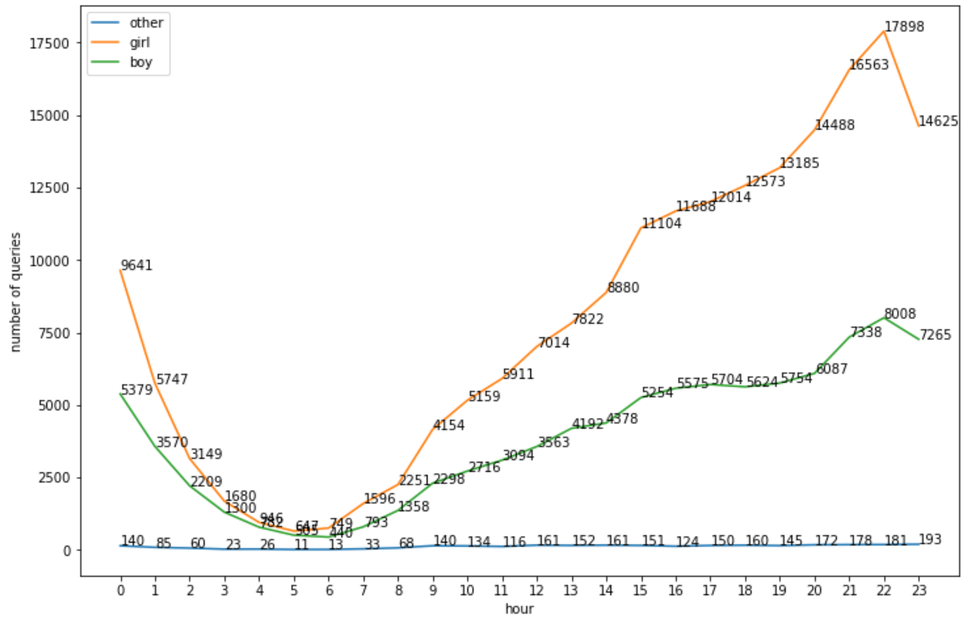


Figure 25. The number of queries by different gender types per hour

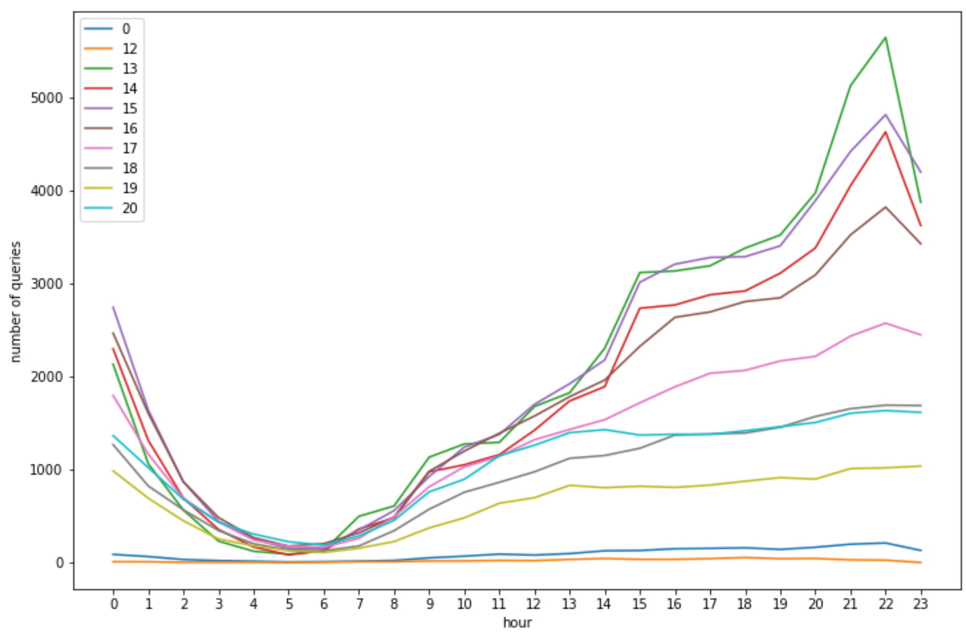


Figure 26. The number of queries by different ages per hour

One of them is the group who specified their gender as other. In the figure 27, the peak time in which queries are sent by the "other" gender the most is at 23:00, that is one hour later than girl and boy groups. In general, the number of queries written by girls and boys are gradually increased after 5:00 or 6:00 in the morning until 22:00, yet the graph of the "other" gender describes a sharp incline from 6:00 to 9:00 and has continuous irregularity afterwards until reaching the highest number of queries around 23:00.

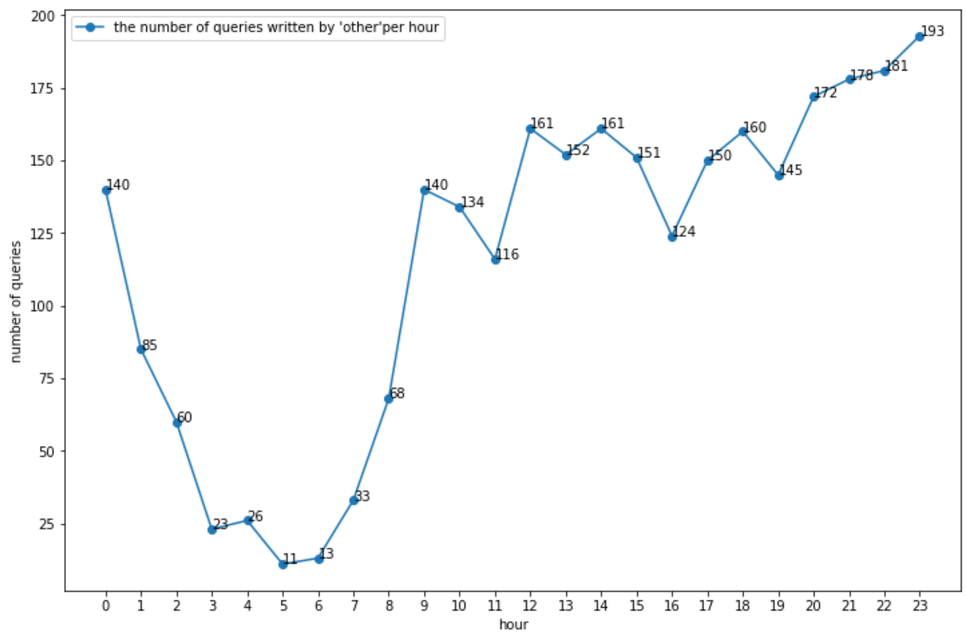


Figure 27. The number of queries by gender type other

Meanwhile, the age group of 12 and, 18 and 19 have their own distinctive time trends. The figure 28 presents the graph of age 12, where the peak time of help seeking is early in the evening around 18:00 that is four hours earlier than the average peak time. Since then, the overall number of queries is generally[21] declining until the next morning about 5:00. Also, after the first leap in the morning, the second leap starts around 12:00 which is also earlier than other groups.

[21] There is an exception, the hour of 23:00 when no queries are written.

This may relevant to the school time of intermediate level of primary schools where the 12 years old belong to.

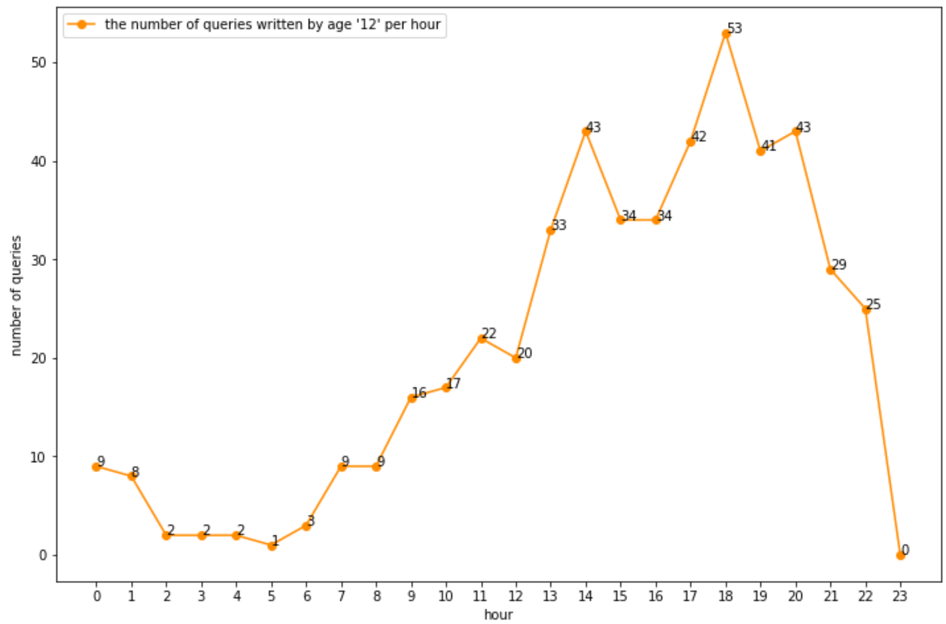


Figure 28. The number of queries by age 12

When it comes to the age group of 19 and 20, their number of help seeking queries rises gradually after 6:00 in the morning and its degree of increase after 13:00 is much smaller than other groups. In addition, a similar number of queries were still made at midnight in contrast to the other age groups which are less likely to request help after 23:00. This may represent lifestyle patterns of young adults, that is not restricted by school schedule anymore. The figure 29 shows the detailed graphs of these two age groups.

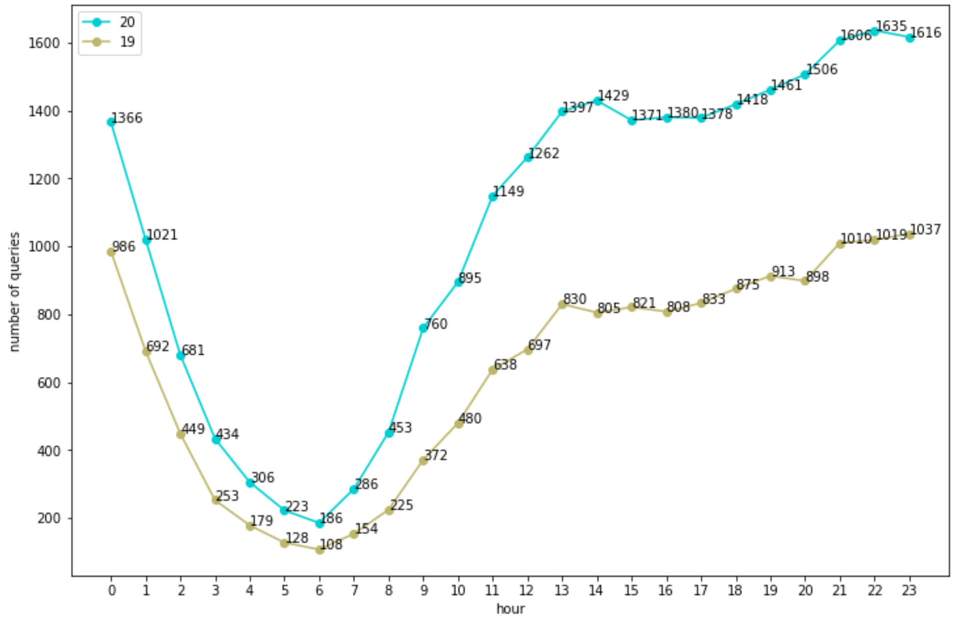


Figure 29. The number of queries by age 19 and 20

4.4 Word frequency analysis in the content of help seeking queries

From the whole query texts, the most frequent noun terms ranked from 1 to 50 were found and presented in a table 9. The most frequent word is “girl”, which appeared 86,293 times from the whole corpus, followed by “school”, “boy”, “friend” and “desire”. “Problem” is placed in the 6th, which may indicate the young acknowledged their situation as problems and made queries with the purpose of getting advice from experts. Terms related relationship including family, friend and boy/girlfriend, school life and sex are viewed in the highest rankings and continue to be placed across the entire list, while the ones regarding job, money, body and mental health are added in between.

The 50 key words are also suggested in a line graph with the actual frequency in the figure 30 and visualized in the figure 31 in a form of WorldCloud which applies different font sizes according to the frequency of the terms to facilitate intuitive understanding.

Table 9

The top 50 frequent words from help seeking queries

	English	Norwegian	Frequency		English	Norwegian	Frequency
1	girl	jente	86,131	26	body	kropp	12,267
2	school	skole	74,541	27	friend	venninne	12,202
3	boy	gutt	62,683	28	belly	mage	11,206
4	friend	venn	57,305	29	teacher	lærer	11,137
5	desire[22]	lyst	37,379	30	children	barn	10,314
6	problem	problem	36,097	31	high school	vgs	9,528
7	mom	mamma	35,209	32	possibility	mulighet	9,386
8	class	klasse	32,663	33	danger	fare	9,218
9	sex	sex	31,510	34	food	mat	9,137
10	boy/girlfriend	kjæreste	28,951	35	weight	vekt	8,950
11	parent	forelder	27,260	36	photo	bilde	8,565
12	menstruation	mens	23,309	37	advice	råd	8,419
13	dad	pappa	21,312	38	thought	tanke	8,400
14	permission	lov	19,906	39	education	utdanning	8,269
15	life	liv	19,379	40	anxiety	angst	8,212
16	job	jobb	19,215	41	nett	nett	8,124
17	mom	mor	16,526	42	summer	sommer	8,101
18	subject	fag	16,294	43	birth control pills	p-pille	7,566
19	tips	tips	15,770	44	pill	pille	7,317
20	family	familie	15,062	45	head	hode	7,242
21	doctor	lege	14,834	46	end	slutt	7,186
22	feeling	følelse	14,757	47	best friend	bestevenn	7,146
23	grade	karakter	14,590	48	math	matte	7,124
24	money	penger	13,418	49	psychologist	psykolog	7,066
25	relationship	forhold	12,417	50	exam	eksamen	7,059

[22] 'lyst' is interpreted as desire since it is counted as noun. However, in many contexts, it uses in a meaning of 'want'. For example, 'har lyst til' and 'har lyst på' are some of the most common idioms in Norwegian, which mean 'want to' and 'want'.

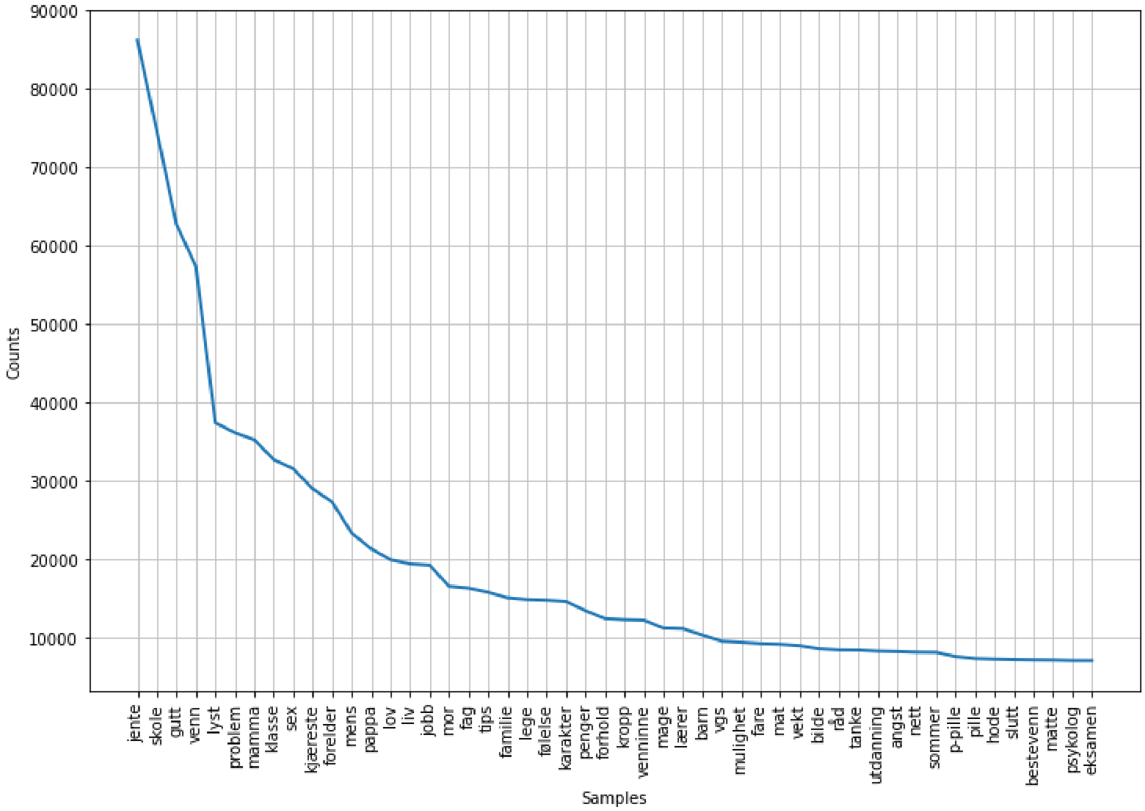


Figure 30. The graph of the most frequent 50 words from help seeking query texts

Figure 31. The word cloud of the most common 50 words from help seeking query texts

4.5 Co-occurrence words analysis of communication keywords

By applying NLTK bigram package with window size five, the words occur with communication keywords designated in the study of Havely at el (2018) are examined. Their keywords were "tell(fortelle)", "ask(spørre)", "talk(snakke)", "answer(svare)", "advice(råd)", "question(spørsmål)", "advise(rådgi)", "explain(forklare)", yet there are no co-occurrence words with advise(rådgi) found in the given help seeking queries. Therefore, the other seven keywords and words occurred together with them are described in table 10 with their number of contingencies. Among the whole found co-occurrence words, the ones appeared more than 100 times are listed in the table. However, the communication word "explain(forklare)" has small amount of co-occurrence words, then the words accompanied more than 10 times are included as an exception. Furthermore, the words which does not have important meaning[23] and ambiguous[24] are removed from the table while the intact version is attached to the appendix.

For better interpretation, co-occurrence words are identified into three categories which are communication targets, content and feelings. First of all, the communication targets that Norwegian youth reached are preferentially informal ones, namely parents, friends and boy/girlfriend rather formal ones who are teacher, school nurse, doctor, psychologist, etc. This tendency is consistent with existing help seeking literature. Interestingly, the communication words such as "tell", and "speak" that describe a common conversational situation have "parents" as co-occurrence words more often. However, the words more clearly imply the intention of help seeking such as "ask", "answer", "advice", "question", "explain" are likely to accompany "friend" and "you guys" which indicates ung.no's experts. Meanwhile, there several words are repetitively occurred regarding the help seeking content. For instance, the young people want to "know" about something they "wonder" and communicate how they "feel". In addition, help is also sought when they "begin" and "try" to do something. Looking into noun words, the most common themes of help seeking from every communication word are related to school and sex issues. However, this is too broad and complicated to detect details so that the next analysis, topic modeling, will give better interpretation of help seeking content. Lastly, this

[23] For example 'stille spørsmål(ask question)' is an idiomatic expression so that 'stille' is not added as co-occurrence words for 'spørsmål' in the table.

[24] For example, 'mens' means menstruation in noun, and while in adverb. For this analysis, it is impossible to identify which meaning 'mens' implies as whole POS was taken.

co-occurrence analysis infers how the Norwegian young feel when they ask for help. Examining adjective words among bigrams of communication words, it is found that there are words describing negative feelings in common such as “scared”, “difficult” and “struggled”.

Table 10

Co-occurrence words with communication keywords

Word	Category	Co-occurred words
tell (fortelle)	Subject related	parent(forelder), mom(mamma, mor), friend(venn), you guys (dere[25]), girl(jente), boy(gutt), dad(far, pappa), girl/boyfriend(kjæreste), girlfriend(venninne), family(familie), doctor(lege), teacher(lærer), psychologist(psykolog), each other(hverandre), who(hvem), best friend(bestvenn), helsesøster(school nurse), etc.
	Theme related	know(vite), feel(føle), talk(snakke), like, try/test(prøve), wonder(lure), say(si), happen(skje), school(skole), think(tenke), should(burde), want(lyst), begin(begynne), problem, ask(spørre), feeling(følelse), find(finne), class(klasse), sex, hurt(vond), help(hjelp), life(liv), truth(sannhet), history(historie), hit(slå), relationship(forhold), fall in love(forelske), permission(lov), pregnant(gravid,) secret(hemmelighet), child welfare(barnevern), etc.
	Feelings related	scared(redd), struggled(slite), dare(tørr), difficult(vanskelig), maybe(kanskje), glad(happy), sick of/bored(lei), right(rett), bad(dårlig), incredible(incredible), gjerne(would like to), cry(grate), insecure(usikker), angry(sur), etc.
ask (spørre)	Subject related	friend(venn), you guys (dere), girl(jente), boy(gutt), mom(mamma), girl/boyfriend(kjæreste), parents(forelder), girlfriend(venninne), dad(pappa), teacher(lærer), doctor(lege), best friend(bestevenn), each other(hverandre), you(deg[26]), buddy(kompis), family(famile), school nurse(helsesøste), etc.

[25] [26] ‘You guys(dere)’ or ‘you(deg)’ indicates experts of ung.no who are expected to answer to the young help seekers

	Theme related	know(vite), answer(svare), like, no(nei), find(finne), wonder(lure), school(skole), say(si), think(tenke), begin(begynne),feel(føle), help(hjelp), try/test(prøve), class(klasse), want(lyst), sex, melding(message), problem, job(jobb/jobbe), permission(lov), photo(bilde), advice(råd), tips(tips), fall in love(forelske/forelsket), feeling(følelse), possible(mulig), hit(slå), kiss(kysse), seek/search/apply(søke), drink(drikke), facebook, sleep(sove), life(liv), party(fest), danger(fare), nudes, snapchat(snap), decide(bestemme), pay(betale), sick(syk), interest(interessere), etc.
	Feelings related	dare(tørr), scared(redd), maybe(kanskje), sure(sikker), angry(sur), sick of/bored(lei), insecure(usikker), embarrased(flau), stupid(dum), difficult(vanskelig), would like to (gjerne), bad(dårlig), suddenly(plutselig utrolig), crazy(gale), properly(sikikkelig), struggled(slite), etc.
speak (snakke)	Subject related	friend(venn), boy(gutt), girl(jente), mom(mamma, mor) school nurse(helsesøster), who(hvem), dad(pappa), psychologist(psykolog), girl/boyfriend(kjæreste), each other(hverandre), parents(forelder), you guys(dere, deg), family(familie), girlfriend(venninne), doctor(lege), teacher(lærer), counselor(rådgiver), bestfriend(bestevenn), buddy(kompis), brother(bror), sister(søster), best girlfriend(bestevenninne), general doctor(fastlege), woman(dame), contact teacher(kontaktlærer), etc.
	Theme related	try/test(prøve), begin(begynne), like, feel(føle), school(skole), say(si), class(klasse), think(tenke), problem, wonder(lure), want(lyst), need(trenge), class(klasse), happen(skje), help(hjelp), send(sende), find(finne), sex, tips, hurt(vond), facebook, relationship(forhold), fall in love(forelske/forelsket), life(liv), message(melding), telephone(telefon), permission(lov), hit(slå), social(sosial), engelsk(english), angst(anxiety), kjempe(fight), snapchat, work(jobbe/jobb), internet(nett), advice(råd), sick(syk), danger(fare), suicide(selvmord), depression(depresjon), subject(fag), period(periode), child welfare(barnevern), miss(savne), interest(interessere), flirting(flørte), quarrel(krangle), thought(tanke), party(fest), photo(bilde), couple(par), self-confidence(selvtillit), situation(situasjon), possibiltiy(mulighet), eat(spise), kiss(kysse), sleep(sove), summer(sommer), body(kropp), BUP(mental health care service for children and young people) , msn, drink(drikke), night(natt), weight(vekk), safe(trygg), sucide thought(selvmordstanke), mobile(mobil), money(penger), etc.

	Feelings related	scared(redd), difficult(vanskelig), dare(tørr/tørre), maybe(kanskje), sick of/bored(lei), angry(sur), incredibly(utrolig), bad(dårlig), cry(gråte), weired(rar), sure(sikker), suddenly(plutselig), insecure(usikker), shy(sjenert), shit(dritt), properly(ordentlig), normal, stupid(dum), embarrassed(flau), depressed(deprimert), annoyed(irritere), crazy(gale), jealous(sjalu), nice(hyggelig), sad(trist), nervous(nervøs), serious(seriøs), funny(gøy), etc.
answer (svare)	Subject related	you guys(dere,deg), girl(jente), boy(gutt), friend(venn), girl/boyfriend(kjæreste), mom(mamma), etc.
	Theme related	ask(spørre), know(vite), question(spørsmål), wonder(lure), speak(snakke), like, send(sende), possible(mulig), message(melding), try/test(prøve), feel(føle), say(si), begin(begynne), think(tenke), school(skole), find(finne), need(trenge), want(lyst), help(hjelp), problem, sex, happen(skje), class(klasse), should(burde), fall in love(forelske), price(pris), etc.
	Feelings related	difficult(vanskelig), scared(redd), maybe(kanskje), sure(sikker), sick of/bored(lei), angry(sur), incredibly(utrolig), happy(glad), struggled(slite), rude(frekt), bother(gidde), etc.
advice (råd)	Subject related	you guys(dere) girl(jente), boy(gutt), friend(venn), mom(mamma), girl/boyfriend(kjæreste), parent(forelder), dad(pappa), etc.
	Theme related	need(trenger), know(vite), wonder(lure), ask(spørre), help(hjelp), speak(snakke), feel(føle), begin(begynne), say(si), like, want(lyst), money(penger), buy(kjøpe), live/stay(bo), possible(mulig), job(jobb), pay(betale), etc.
	Feelings related	bad(dårlig), maybe(kanskje), would like to(gjerne), difficult(vanskelig), etc.
question (spørsmål)	Subject related	you guys(dere), girl(jente), boy(gutt), friend(venn), etc.
	Theme related	know(vite), answer(svare), wonder(lure), find(finne), school(skole), possible(mulig), permission(lov) should(burde), think(tenke), right(rett), send(sende), categori(kategori), seek(søke), fee(føle), begin(begynne), need(trenger), speak(snakke), help(hjelp), subject(fag), problem, understand(forstå), want(lyst), sex, happen(skje), possible(mulig), possibility(mulighet), etc.
	Feelings related	maybe(kansje), insecure(usikker), scared(redd), normal, etc.
explain (forklare)	Subject related	you guys(dere), girl(jente),friend(venn), boy(gutt), etc.
	Theme related	know(vite), try/test(prøve), feel(føle), happen(skje), situation(situasjon), wonder(lure), speak(snakke), understand(forstå), say(si), begin(begynne), problem, etc.

	Feelings related	difficult(vanskelig), etc.

4.6 Topic modeling of help seeking queries

Topic modeling was implemented for the whole 285,447 help seeking queries generated from 2008 to 2018 by Norwegian young people. First of all, in order to find the optimal number of topics, topic coherence scores for when the number of topics is from 1 to 20 were calculated. As a result, as shown in the figure 32, when the number of topics is 19, the score is found to be the highest at 0.6992. Therefore, LDA model was applied by setting the number of topics to 19.

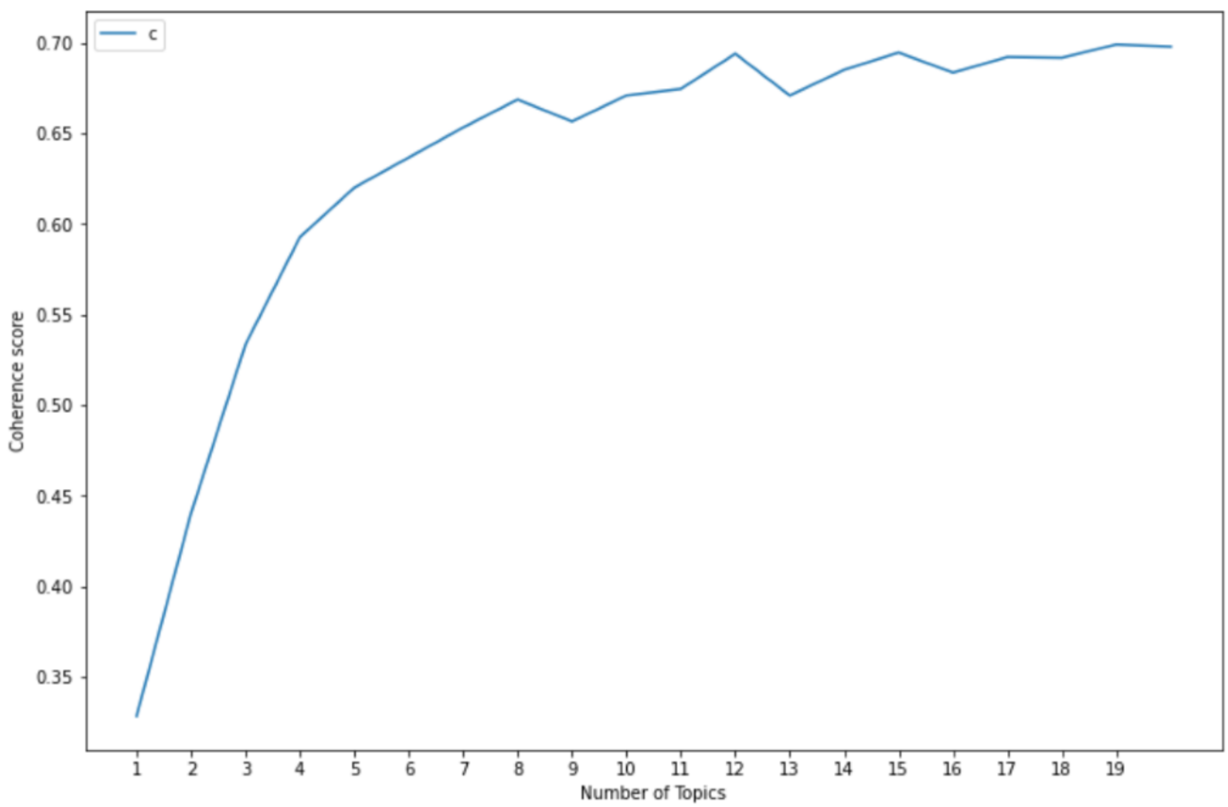


Figure 32. The graph of coherence score. The highest score is 0.6992 when the number of topics is 19, followed by when the number is 20(0.698) and 15(0.6947).

Meanwhile, name of topics derived from the LDA should be assigned by researchers themselves. For example, when it is found a topic with words such as “sex”, “menstruation”, “birth control pill”, “condom” and “pregnant”, etc., it can be named as “sex health” by considering their semantic relations. In this way, the total 19 topics of help seeking queries are named as follows; (1) family, (2) study, (3) everyday complaints, (4) weight, (5), school life, (6) mental health, (7) affairs at parties or on holidays, (8) friend, (9) money, (10) career, (11) couple relationship, (12) physical health, (13) illegality, (14) reputation, (15) falling in love, (16) permission and rights, (17) information searching, (18) body appearance, (19) sex health.

To be more specific, the first topic is family issues. This contains various types of family member, and words indicating their discords such as “child welfare”, “quarrel”, “violence”, etc. The second topic is about study, mainly consisting of words related to academic paths and subjects. The third topic contains everyday complaints and dissatisfaction. Words belong to this topic are negative expressions and skepticism about life and the world. The fourth topic is weight including words related to food, body, and exercise. The fifth topic is about school life, covering overall school activities such as classes, tests, and assignments. Furthermore, words such as “teacher”, “advisor” and “principal” also appeared in this topic. The sixth topic is mental health problems, which cover not only stress and symptoms caused by anxiety and depression, but also more serious actions expressed by words such as “self-harm” and “suicide”. Furthermore, words related to treatment and counseling are also included. The seventh topic is affairs that happened at parties or on holidays. Words like “alcohol”, “tobacco”, “abuse”, “rape” are seen in this topic. The eighth topic is friend-related matters, for instance, what might happen in a group of friends so called “gang(gjeng)”. Especially there are words about “russ” which is tradition of Norwegian high school students celebrating their final spring semester by traveling with bus are seen together. The ninth topic is money. Containing words are about financial assistance and rights to rent or buy such as “car” or “apartment”. The 10th topic consists of career-related words such as job and education. The 11th topic covers couple relationship, which is made up with words indicating problems between boyfriend and girlfriend, beginning and end of relationship, and their conditions such as age and distance. The 12th topic is physical health issues containing words refer to various parts of the body and medical supports such as doctor, symptoms and treatment. The 13th topic is illegality. Words consist of age restrictions and penalties for criminal activities. The 14th topic is reputation, which also includes being bullied among friends.

Interestingly, snapchat-related terms such as photos, nude photos, and profiles also appear. The 15th topic is about falling in love, with words about feelings and actions of affection. This span a variety of genders. The 16th topic consists of words about getting permission from parents and rights as youth. The 17th topic is about recommendation. It is seen that relevant words are mainly related to asking for suggestions regarding media, media contents and military service. The 18th topic is body appearance. Sexual body parts and puberty related terms are belonging to this topic. Lastly, the 19th topic is sexual health. Corresponding words are about sex, menstruation, pregnancy and contraception. The list of top 30 most relevant term in each topic are introduced in the table 11. Furthermore, the topics are also presented in a quadrant of intertopic distance map in the figure 33.

Table 11

The 19 topics of help seeking content and their top 30 most relevant terms

Topic	Name	Top 30 most relevant terms
1	Family issues	mom (mamma), dad(papa), mom(mor), child(barn), family(familie), danger(fare), brother(bror), child welfare(barnevern), sister(søster), house(hus), parents(forelder), quarrel(krangel), being scolded(kjeft), stepfather(stefar), sibling(søsken), foster home(fosterhjem), room(rom), little brother(lillebror), the young(unge), aunt(tante), daughter(datter), uncle(onkel), little sister(lillesøster), violence(vold), stepmother(stemor), son(sønn),quarrel(krangling), confirmation(konfirmasjon), grandmother(bestemor), visit(besøk)
2	Study issues	subject(fag), grade(karakter), high school(vgs), math(matte), exam(eksamen), average(snitt), the 1st year of high school(vg1), autumn(høst), study competence(studiekompetanse), the 2nd year of high school(vg2), english(engelsk), study(stadium), study choice(linje), study specialization(studiespesialisering), diploma(vitnemål), self-study(privatist), the 3rd year of high school(vg3), medium, university(universitet), abroad(utland), language(språk), chemistry(kjemi), study(studie), education(utdanning), physics(fysikk), communication(kommunikasjon), extension(påbygg), medicine(medisin), level(nivå), Spanish(spansk)

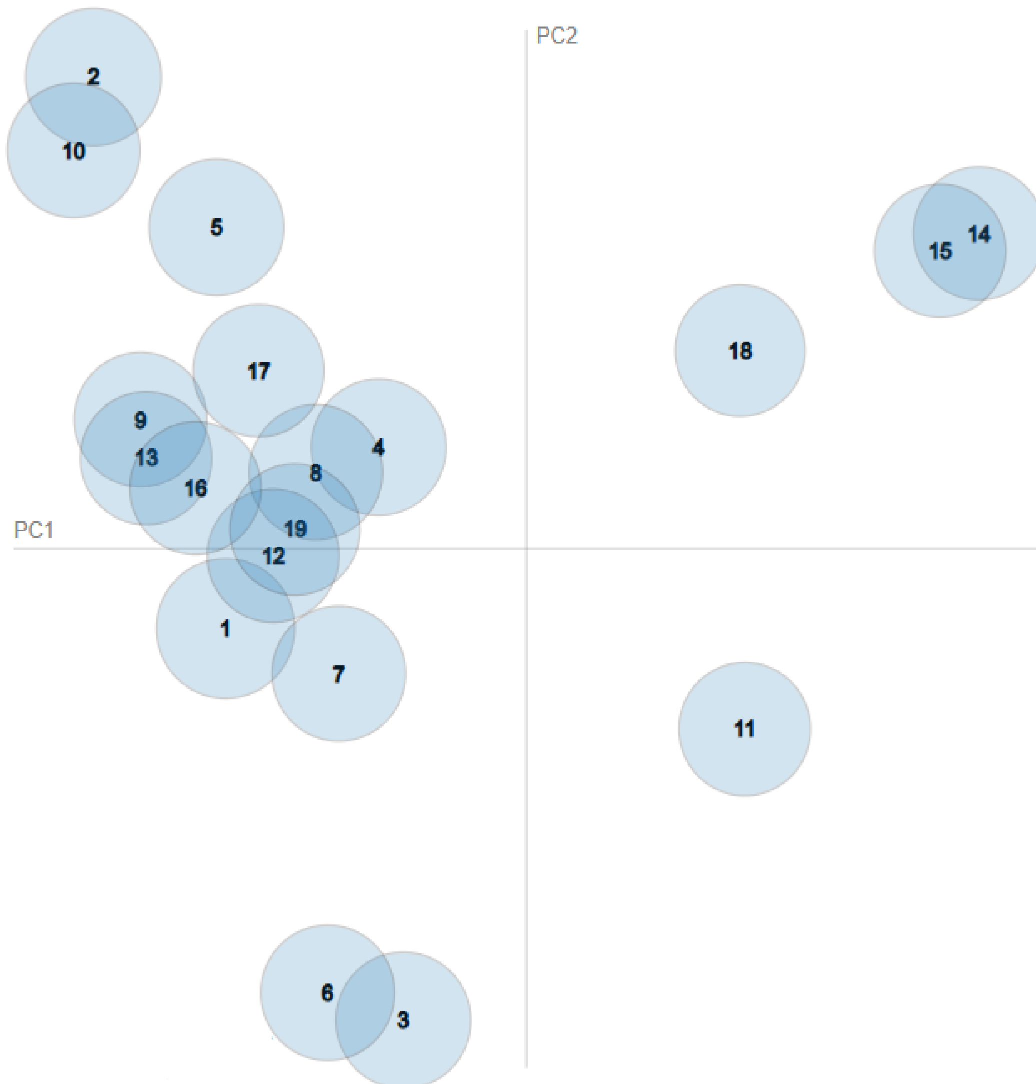
3	Everyday complaints	life(liv), problem, family(familie), mistake(feil), reason(grunne), word(ord), advice(råd), suicide(selvmord), meaning(mening), world(verden), rest, weather(vær), shit(dritt), situation(situasjon), way(vei), thought(tanke), solution(løsning), future(fremtid), fuck(fæn), guilt(skyld), hate(hat), hope(håp), conscience(samvittighet), how(åsse), human(menneske), way(måte), past(fortid), tear(tåre), every day(hverdag), sin(synd)
4	Weight issues	belly(mage), food(mat), body(kropp), weight(vekt), training(trening), cm, kilo, muscle(muskel), thighs(lår), tips, dinner(middag), eating disorder(spiseforstyrrelse), exercise(trener), fat(fett), water(vann), sports(idrett), thinner(tynner), shape(form), diet(kosthold), strength(styrke), kg, calories(kalori), activity(aktivitet), breakfast(frokost), handball(håndball), excercise(øvelse), candy(godteri), thick(tykk), result(resultat), height(høyde)
5	School issues	school(skole), teacher(lærer), teaecher(lære), pupil(elev), absence(fravær), test(prøve), homework(lekse), gym, grade(karakter), assignment(oppgave), junior high school(ungdomsskole), motivation(motivasjon), county(fylke), halfyear(halvår), school year(skoleår), remark(anmerkning), classroom(klasserom), principal(rektor), school day(skoledag), midterm(tentamen), presentation(presentasjon), advisor(rådgiver), teachers(lærere), school work(skolearbeid), contact teacher(kontaktlærer), high school(vgs), teaching(undervisning), school place(skoleplass), gym class(gymtime), junior high school(ungdomskol)
6	Mental health issues	problem, anxiety(angst), head(hode), psychologist(psykolog), thought(tanke), depression(depresjon), period(periode), school nurse(helsesøster), thoughts(tanker), nausea(kvalm), self-harm(selvskading), stress, illness(sykdom), situation(situasjon), sleep(søvn), everyday(hverdag), control(kontroll), suicidal thoguht(selvmordstanke), diagnosis(diagnose), energy(energi), hospital(sykehus), shape(form), panic(panikk), degree(grad), heart(hjerte), sound(lyd), voice(stem), treatment(behandling), medicine(medisiner), dizziness(svimmel)
7	Affairs at parties or on holidays	girlfriend(venninne),alcohol, night(natt), room(rom), party(fest), weekend(helg), buddy(kompis), night(kveld), bed(seng), morning(morgen), couple(par), eye(øye), way(vei), door(dør), visit(besøk), end(slutt), clock(klokke), smoke(røyk), abuse(overgrep), weekend(helgen), birthday(bursdag), guy(fyr), horse(hest), trip(tur), rape(voldtekt), wall(vegg), light(lys), beer(øl), second(sekund), neighbor(nabo)

8	Friend issues	friend(venn), best friend(bestevenn), wil(vill), talk(snakk), girld frined(vennine), best girlfrined(bestevenninne), free time(fritid), friendship(vennskap), veninne(girl frined, lag(team), russ, gang(gjenge), bus(busse), slow(sakt), time(tid), maybe(kansje-typo), supposed to(skule-typo), together(ilag), naughty(frekk), contact(kontakt), gang(gjeng), group of friends(vennegjeng), mean(slem), movie(kino), one(ein-typo), many(mange), soft(mykje),russ time(russetid), lose(miste), win(seier)
9	Money issues	money(penger), scholarship(stipend), note or receipt(lapp), car(bil),apartment(leilighet), support(støtte), cars(bile), Norwegian Kroner(kr),driver's license(førerkort), student loan(lånekasse), laon(lån), advice(råd), moped, a type of apartment(hybel), requirement(krav), right(rett), possibility(mulighet), message(beskjed), contract(kontrakt), way(vei), August(august), municipality(kommune), Norwegian Labor and Welfare Administration(nav), May(mai), right(rettighet), student January(januar),account(konto), finance(økonomi), month(mnd)
10	Carrer	job(jobb), education(utdanning), possibility(mulighet), health(helse), apprentice(lærling), study choice(linje), choice(valg), music(musikk), salary(lønn), abroad(utland), shop(butikk), profession(yrke), extension(påbygg), plan, certificate of apprenticeship(fagbrev), work(arbeid), boss(sjef), course(kurs), requirement(krav), possition(stilling), kinder garden(barnehage), education(utdannelse), service, company(bedrift), upbrining(oppvekst), nurse(sykepleier), dance(dans), offer(tilbud), application(søknad), jobs(jobber)
11	Couple relationship	boy/girlfriend(kjæreste), problem, relationship(forhold), girl(jente), end(slutt), type, age(alder), couple(par), tips, advice(råd), man(mann), woman(dame), manage(greie), start, chance(sjanse), month(mnd), month(månde), Christmas(jul), reason(grunne), question(spm[27]), begining(begynnelse), guy(fyr), page(side), tooth(tann), ex(eks), price(pris), dentist(tannlege), distance(avstand), gift(gave), number(nr)
12	Physical health	doctor(lege), pain(smerte), skin(hud), pimple(kvise), abdomen(underliv), face(ansikt), lump(klump), wound(sår), cancer(kreft), foreskin(forhud), throat(hals), dot(prikk), hole(hull), mouth(munn), narrow(trang), kind(slags), foot(fot), eye(øye), finger, back(rygg), injury(skade), operation(operasjon), area(område), nose(nese), knee(kne), ear(øre), labia(kjønnsleppe), tablet(tablett), fungus(sopp), penis head(penishod)

[27] Abbreviation of "spørsmål"

13	Illegality	name(navn), police(politi), age limit(aldersgrense), fine(bot), consequence(konsekvens), letter(brev), game(spill), punishment(straff), video, use(bruk), case(sak), online(nett), case(tilfelle), message(beskjed), prision(fengsel), punishment(straffe), hashish(hasj), record(rulleblad), information(informasjon), shop(butikk), paper(papir), evidence(bevis), lawyer(advokat), knife(kniv), drug(narkotikum), goods(vare), post, drug(stoff), mail, cannabis
14	Reputation	girl(jente), class(klasse), photo(bilde), confidence(selvtillit), tips, conversation(samtale), junior high school(ungdomsskole), group(gruppe), elementary school(barneskole), bullying(mobbing), step(trinn), comment(kommentar), free time(friminutt), snapchat, junior high school(ungdomskole), snap, mark(merke), reputation(rykte), summer vacation(sommerferie), glance(blikk), start, closeeness(nærhet), bully(mobbe), ps, boy(gutt), 8th grade(8.klasse), ball, appearance(utseende), nude photo(nakenbilde), profile(profil)
15	Falling in love	boy(gutt), girl(jente), feeling(følelse), message(melding), ex-boy/girlfrined(ekse), bisexual(bifil), gender(kjønn), falling in love(forelskelse), sing(tegn), falling in love(forelske), attention(oppmerksomhet), boy/girl friend(kjærste), homo, personality(personlighet), msn, hug(klem), love(kjærlighet), interest(interesse), closet(skap), closeness(nærhet), kiss(kyss), handsome(kjekk), jealous(sjalu), reality(virkelighet), step(steg), eye contact(øyekontakt), boyfriend(guttevenn), contact(kontakt), date, love(kjærlighets), grief(sorg)
16	Permission and rights	parents(forelder), permission(lov), summer(sommer), show(vis), holiday(ferie), online(nett), mobile(mobil), phone(telefon), rule(regle), right(rett), example(eksempel), trip(tur), dog(hund), border(grense), summer holiday(sommerferie), animal(dyr), rule(legel), place(sted), number(nummer), example(feks), fall, meet(møte), goods(vare), flight(fly), internet(internett), right(rettighet),muslim, pc, permission(tillatelse), pce
17	Recommendation	desire(lyst), kind(slags), film, youth(ungdom), online(nett), website(nettside), dream(drøm), difference(forskjell), defense(forsvar), tips, example(eksempel), suggestion(forslag), tema, fall, price(pris), page(sider), interest(interesse), miliatry(militær), plan, page(side), question(spørsmål), idea(ide), book(bok), session(sesjon), topic(emne), history(historie), place(sted), book(bøk), information(informasjon), actor(skuespiller)
18	Body appearance	hair(hår),penis, girl(jente), tits(puppe), hand(hånd), arm, breast(bryst), klær, body(kropp), snus, page(side), puberty(pubertet), butt(rumpe), age(alder), sexual organs or pee(tiss), cm, back(rygg), porno, leg(bein), tips, size(størrelse),

		pants(bukse), weather(vær), ps, top(topp), brand(merke), shower(dusj), horny(kåt), ass(rompe), shoulder(skulder)
19	Sex health	sex, menstruation(mens), birth control pill(p-pille), pill(pille), condom(kondom), intercourse(samleie), blood(blod), pregnant(gravid), orgasm(orgasme), test, fold(brett), bleeding(blødning), tampon(tampong), pause, pregnancy(graviditet), panty(truse), vagina(skjede), virgin(jomfru), semen(sæd), contraceptive(p-stav), chance(sjanse), sign(tegn), abortion(abort), emergency contraception(angrepille), bandage(bind), month(mnd), partner, side effect(bivirkning), menstruation(menstruasjon),vagina



1. Family, 2. Study, 3. Everyday complaints, 4. Weight, 5. School life,
6. Mental health, 7. Affairs at parties or on holidays, 8. Friend, 9. Money,
10. Career, 11. Couple relationship, 12. Physical health, 13. Illegality,
14. Reputation, 15. Falling in love, 16. Permission and rights,
17. Recommendation, 18. Body appearance, 19. Sex health

Figure 33. The intertopic distance map

From the figure 33, the structural relationship between topics can be read. For instance, there are several connected topics are found such as topic of study and career, everyday complaints and mental health, and reputation and falling in love. Furthermore, on the left side of the quadrant, many topics are partially overlapped each other. Topics of money, illegality, permission and rights are share common words together, while topics about physical health and sexual health are overlapped the most. In addition to this, when it comes to topic of friend, there are weight, physical health, and sexual health topics are largely engaged.

4.7 Clustering of online youth help seekers

The last analysis is to cluster demographic groups based on their help seeking content. To do this, firstly the whole query corpus generated by Norwegian young has been classified by their demographic features. Since there are three types of gender and ten types of age, it was expected to obtain 30 documents, yet as it has been found in the second analysis there are no data written by whom are 12 and marked their gender as “other”. Therefore, total 29 documents were prepared to be clustered by K-means algorithms.

To find the optimal number of clusters, K, the elbow method was applied. To be more specific, each SSE were calculated when K is one to ten. As it is seen in the figure 34, the inflection point is found when K is four. Therefore, the optimal number of clusters is set to four and K-means was implemented. As a result, 29 documents were grouped into 4 clusters, which are generally consistent with age and gender.

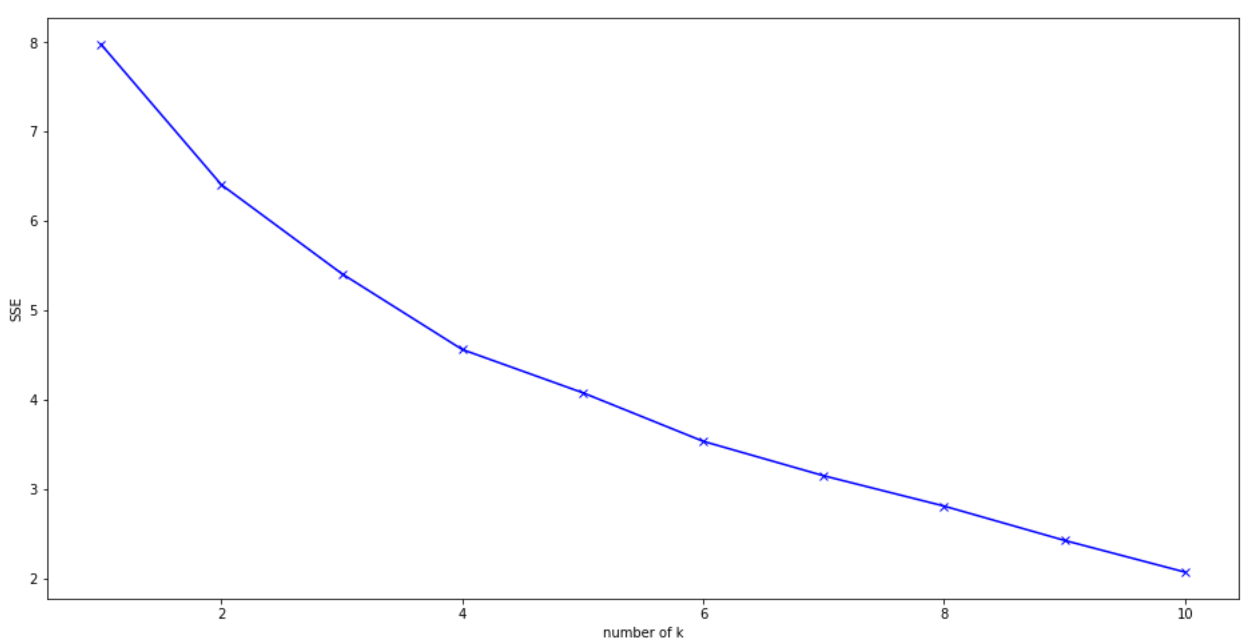


Figure 34. The graph of SSE based on the elbow method. The elbow is shown as when K is four. The degree of decrease gets smaller after passing four.

The first cluster consists of queries written by girl who are in 12 to 17 and did not reveal their age, and those who has "other" gender in their age of 13 to 17. This cluster is the biggest one which takes up more than half of help seeking. The second cluster represents the queries from young adult age of 18 to 20 in each gender group. They are 22 percent of the whole demographic. Furthermore, the third cluster contains documents written only by boys whose age ranges from 12 to 17 and the boys who marked their age as 0 are included in this cluster. This cluster is about 24 percent of the total. Lastly, the fourth cluster has documents made by a single demographic group whose gender is other and age was entered as 0. Only 0.01% of the whole help seeking cases belong to this cluster. The table 12 shows the summary of the clusters and their composition. In addition, the figure 35 visualizes distances between each document and how they are clustered in two-dimensional graph through dimensional reduction by PCA.

Table 12.

The summary of the clusters and its demographic groups and ratio

Cluster	Demographic groups	Ratio
1	girl 0, girl 12, girl 13, girl 14, girl 15, girl 16, girl 17 other 13, other 14, other 15, other 16, other 17	53.89%
2	girl 18, girl 19, girl 20, boy 18, boy 19, boy 20, other 18, other 19, other 20	22.17%
3	boy 0, boy 12, boy 13, boy 14, boy 15, boy 16, boy17	23.92%
4	other 0	0.01%

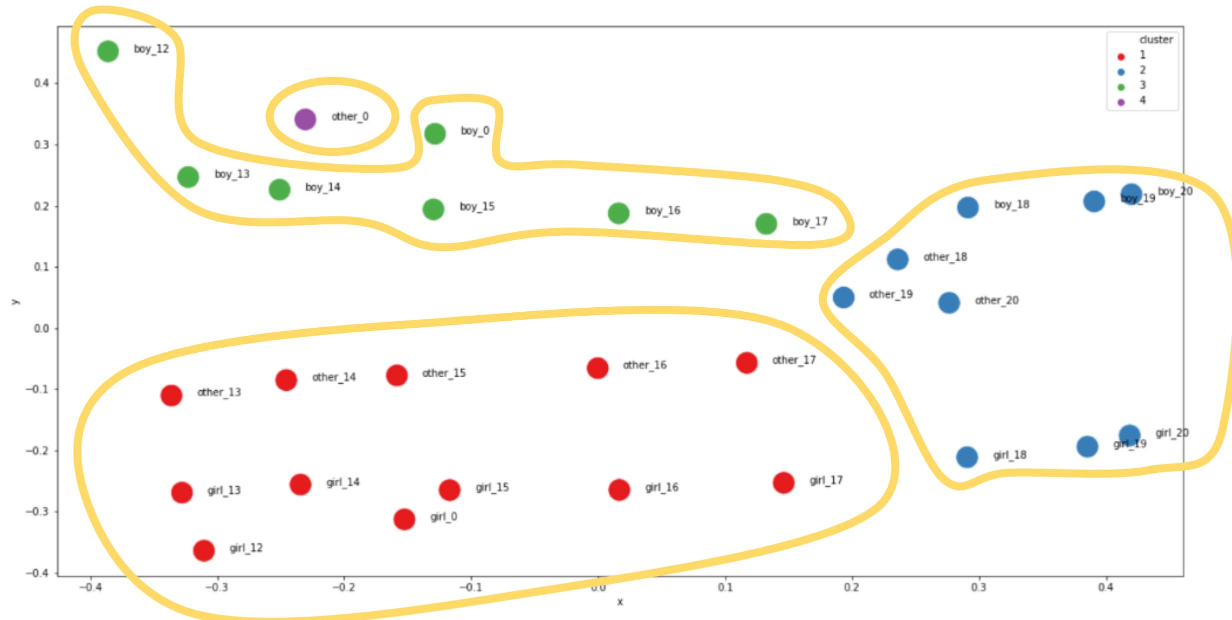


Figure 35. The visualization of clustering results by PCA

The result of clustering indicates that youth share similar demographics are more likely to have similar worries. Interestingly, the help seeking documents of "other" gender are located in the middle of those of boy and girl in the figure. According to k-means function, "other" gender under 18 years have questions in common with girls, while all the youth who are in 18 to 20 share similar issues regardless of their gender. Meanwhile, teenage boys under 18 formed their own cluster. Furthermore, the young people who are "other" gender and did not disclose their age are certainly separated from another demographic group. The top 30 words of each cluster are presented in the table 13 to find their characteristics in terms of help seeking content.

Table 13.

The top 30 words representing each cluster

Cluster 1: girl0,12-17, other 13-17

girl(jente), school(skole), friend(venn), boy(gutt), mom(mamma), desire(lyst), class(klasse), problem, parent(forelder), sex, mens(menstruation), boy/girlfriend(kjæreste), dad(pappa), life(liv), mother(mor), family(familie), permission(lov), body(kropp), tips, teacher(lærer), children(barn), feeling(følelse), fag(subject), pill(pille), grade(karakter), money(penger), **girlfriend(venninne)**, doctor(lege), belly(mage), danger(fare), hair(hår), job(jobb), **food(mat)**, high school(vgs), psychologist(psykolog), bestfrined(bestevenn), anxiety(angst), relationship(forhold), **weight(vekt)**, head(hode), **gender(kjønn)**, advice(råd), matte(math), nett(online), tanke(thought), **brother(bror)**, **bresat(bryst)**, **pupil(elev)**, photo(bilde), age(alder)	relationship & weight
Cluster 2: girl18-20, boy18-20, other18-20	
school(skole), girl(jente), boy(gutt), job(jobb), problem, sex, boy/girlfriend(kjæreste), friend(venn), subject(fag), desire(lyst), perission(lov), life(liv), money(penger), **exam(eksamen)**, education(utdanning), grade(karakter), doctor(lege), family(forelder), relationship(forhold), possibility(mulighet), pill(pille), child(barn), mom(mor), menstruation(mens), **lærling(apprentice)**, family(familie), advice(råd), feeling(følelse), high school(vgs), psychologist(psykolog), thought(tanke), anxiety(angst), class(klasse)**, month(mnd)**, **autumn(høst)**, **study(studium)**, **apartment(leilighet)**, photh(bilde), body(kropp), oneline(nett), **message(beskjed)**, math(matte), **diploma(vitnemål), couple(par)**, tips, **abroad(utland)**, situation(situasjon), mom(mamma), **loan(lån)**, **self-study(privatist)**	study & future
Cluster 3: boy0,12-17	
boy(gutt), school(skole), girl(jente), friend(venn), class(klasse), desire(lyst), sex, problem, permission(lov), **penis**, parents(forelder), mom(mamma), boy/girlfriend(kjæreste), tips, job(jobb), dad(pappa), life(liv), hair(hår), grade(karakter), money(penger), subject(fag), **cm**, mom(mor), age(alder), teacher(lærer), feeling(følelse), **foreskin(forhud)**, doctor(lege), **puberty(pubertet)**, body(kropp), **training(trening)**, family(familie), **show(vis)**, **game(spill)**, high school(vgs), **semen(sæd)**, photo(bilde), math(matte), education(utdanning), danger(fare), online(nett), head(hode), possiblity(mulighet), vg1, advice(råd), relationship(forhold), belly(mage), **page(side)**, **alcohol(alkohol)**, buddy(kompis)	sexual organs & amusement
Cluster 4: other 0	

boy(gutt), **table tennis ball(bordtenisball)**, desire(lyst), ex-boy/girlfriend(ekse), mom(mamma), child(barne), danger(fare), buddy(kompis), **martial arts(kampsportssevne)**, friend(ven), family(familie), **introduction(innføring), death(død)**, girl(jente), doctor(lege), **censor(sensor)**, **mole(føflekk)**, **caseworker(saksbehandler)**, **uncle(onkel)**, online(nett), **punishment(straff), mushroom drug(sopp)**, child(barn), **end(slutt)**, feeling(følelse), teacher(lærer), **computer page(dataside)**, **bronze age(bronsealder)**, tmeg[28], life(lliv)[29], child, **include(omfatt)**, **online blog(nettblogge)**, **banging**, **reviewer(anmelder)**, **tetrahydrocannabinol(thc)**, **birthday(bursdag)**, **security(sikerhets)**, **bomb attack(bombetrussel)**, duche[30], dite[31], **municipal council meeting(kommunestyremøte)**, **rectal penetration(endetarmsåpning)**, snus, **climate(klima)**, bestfriend(bestevenn), anxiety(angst), **please(plzz)**, **reader post(leserinlegg)**, **methamphetamine(meth)**	drug, etc.

Note. The words in bold are the ones appear in only corresponding cluster

Although the clusters contain various topics previously shown in the topic modeling analysis, there are differences between each cluster. For instance, the first cluster includes more words related to relationships such as “girlfriend”, “brother”, etc. which do not appear in the 30 top words of other clusters. So, girls and “other” gender who are under 18 may have more problems in relationships. Furthermore, “food”, “weight” and “breast” also appear alone in this cluster so that it can be also possible to interpret that they are concerned about their body and weight, especially considering that words such as “body” and “belly” are seen together.

Meanwhile the second cluster shows another type of unique words like “exam”, “studium”, “diploma”, “self-study” that are related to study and education. Moreover, words such as “apprentice”, “abroad”, “apartment” and “loan” are only shown in this cluster, so that the young adult who are above 18 may need help regarding their study and career and future plan.

By looking into the words from the third cluster, what Norwegian boys under the age of 18 ask through ung.no service can be inferred. “Penis”, “cm”, “foreskin”, “puberty” and “semen”

[28] [29] [30] [31] Examples of typo mistakes recognized as noun words

[29] It is written as ‘lliv’, but treated as ‘liv’ considering it as a common typo

are the words only shown in this cluster, which are related to sexual organs. Moreover, words indicating amusement are also included such as “training”, “game” and “alcohol”.

The last cluster representing the demographic group who marked their gender as other and age as 0 has distinct word composition from others. Some words such as “mushroom”, “tetrahydrocannabinol” and “methamphetamine” are drug related while sexual words such as “banging”, “rectal penetration” also appear. Furthermore, “online blog”, “reviewer”, “reader post” which are the words about online media, types of sports like “table tennis ball” and “martial arts”, and uncommon words such as “bronze age”, “bomb attack”, “municipal council meeting” are included. In addition to this, there are several typo mistakes that is difficult understand its original intention in its top 30 words. These observations show that this demographic group communicates special concerns which are differentiated from the other existing youth.

In chapter 6, there will be discussions on the findings from the seven analyses. By interpreting each outcome and comprehensively imparting meanings to it, this research will finally find answers to the two research questions.

5. Discussion

5.1 Discussion about bigdata analysis on youth online help seeking

When it comes to the first research question, “how can bigdata be used to understand youth online help seeking?”, the way of the seven analyses heuristically designed in this research can be the answer in a broad sense. To be more precise, data-driven characteristic in this research is powerful. Rather than testing only specific hypotheses reflecting existing research and theories of youth help seeking, new possibilities of finding patterns implied by bigdata itself was attempted to be drawn with exploratory purposes through various analyses. If this wasn't a big data research, the research would have mainly focused on online help seeking sources or motivations and obstacles of young people, which previous studies had already been interested in. Perhaps, by collecting data from different subjects, yet using the same traditional methodologies, their claims would have been tested. However, as bigdata is not created but found, this by-product of the help seeking behaviors of the youth includes more variables than those of existing studies, such as time record, demographic information and raw counseling texts. Thus, the researcher utilized them at the most for discovering their hidden values. Even so, it does not mean that this bigdata research only aimed to produce new results that are totally new from previous youth online help seeking research. Rather, taking the literature as a pathfinder, some analyses were to revisit their findings. As a result, the outcomes of the seven analyses support or supplement existing theoretical foundations.

For example, the fifth analysis was finding terms that occur along with specific communication keywords related to help seeking from youth online question texts. The sources of support and how young people feel when requesting help were captured which coincide with the existing discoveries, namely, they are more likely to contact informal sources than to formal ones, and feel difficulty and fear in communicating their state of mind already pointed out as a hindrance to will of help seeking. It is of great support for the academic community of youth online help seeking that these findings were reproduced not through common research tools, but in a huge amount of text written directly by research subjects.

Meanwhile, other analyses provided new findings that can contribute to expansion of the field of youth online help seeking by considering the influence of variables that existing research overlooked or could not approach and applying bigdata analytics. The first to third analyses used help seeking time and demographic information of the youth while the fourth and sixth analyses dealt with help seeking content that was difficult to make a thorough study of due to ethical and methodological limitations. The seventh analysis is implementing segmentation of help seekers, which originates from social marketing where difference between users are more emphasized than integration. The following explains the details provided by each analysis.

In the first analysis, it is found that Norwegian youth tend to ask for help more in the beginning of year and month, on Mondays during week, and mainly at night in daily basis. Relatively fewer questions were generated during vacation periods, the end of the months, on Fridays and Saturdays. This analysis of timestamps revealed that there are certain periods that the young make help seeking questions more or vice versa. This may imply that the existing research could have yielded various results depending on time. For instance, a research conducted in the beginning of the year may describe that young people are more active in help seeking compared to the one conducted before vacation. In this regard, this bigdata research analyzing 10 years of queries can suggest time as a latent variable of help seeking.

The second analysis which examined demographics of the online help seekers found that the queries written by girls, especially relatively younger ones aged 13 to 16 are accounted for 66 percent of the total. This can imply that they are more active in help seeking or the online help seeking media is more favorable type for young girls compared to other gender and age groups. The proportion of boys is about half of that of girls, and this gender imbalance is in line with the study of Best et al (2016), who expanded Rickwood et al (2005)'s traditional help seeking model for boys. However, not a lot of literature have mentioned demographic differences so far. Meanwhile, in the third analysis, the patterns of help seeking by hour were investigated based on demographic information of the young. By setting the unit of analysis at gender and age level, it is found that there are some subgroups who have different tendencies from others. In regards to the group who marked their gender as "other", the time when help is sought the most is about 23:00 which is one hour later than girl and boy groups. Furthermore, their number of queries are uneven compared to others having a gradual increase and decrease of the flow. This can be interpreted that the group of “other” gender asked questions whenever they want, rather than following a general schedule of students. Meanwhile, the group of 12 years old has their highest number of queries around 18:00 that is about four hours earlier than other age groups. And the number of queries by the young adults who are of 18 and 19 are gradually increasing during the day and stays at a certain level until midnight. The results show the trends of the internal minority groups that could not be approached through the whole size analysis. Furthermore, it is assumed that there are indirect time restraints and differences between the youth due to their various lifestyles and school schedules even though help can be sought any time if the Internet is accessible. These findings from the second and third analyses convey importance of demographic variables, in contrast to existing literature that have viewed youth as one intact group and have focused on generalizing their online help seeking.

The fourth analysis was to identify the words that appeared the most in the query text data while the sixth analysis was for extracting topics through a machine learning technique beyond simply counting frequent words. In previous literature, researchers were not able to observe the actual help seeking behavior of young people, instead they indirectly approached youth issues, through research tools such as questionnaires or interview questions. Due to ethical limitation, not every study contained content of help seeking, yet several such as disease, sexual health, weight, anxiety, depression, and insomnia were suggested. On the contrary, in this study, content of help seeking were inferred by directly analyzing query text written by young people. Thus, as a result, it is found that terms related to relations such as friend and family, and related to school were at the top rank of the frequency and 19 different topics are sorted. In contrast to the traditional literature which suggested only a few topics and described the young mainly seeking physical and sex health information, this bigdata research reveals that young have various issues, presumably need more psychological and social supports regarding relationships or studies.

Lastly, in the seventh analysis, a total of 29 demographic groups were clustered into four groups based on help seeking content. In general, each cluster consists of same gender or continuous age range. Yet, there is found one exclusive group who marked their gender as “other” and age as 0. The four clusters are represented by top 30 relevant words. Among them, by ignoring common words and focusing on the ones that independently appeared in each group, their relatively dominant words have found. This implies the possibility of specific concerns occurring in different demographic groups. Such segmentation is a new practice of help seeking research, rather more common in bigdata analytics in social marketing. However, this can provide evidence to further expand the scope of existing research by understanding differences and similarities within the youth and discovering minorities.

5.2 Discussion about opportunities and challenge of bigdata research

In the second research question, finding opportunities and challenges of using bigdata in media and communication research was proposed. Based on the procedure of this research, three opportunities and challenges are summarized in this subsection. First and foremost, bigdata excites researchers to widen their research perspectives and broaden their research scopes. Principally, researchers deduce hypotheses from previous theories, and repeat hypothesis testing in different contexts. In this way, they can strengthen or dispute the main statements, yet there are limitations of involving unseen variables assumedly related to a certain phenomenon and investigating new relationships. It is because theory-driven characteristic is a philosophy of scientific study so that getting out from the given frame is taboo for researchers. So far, exploratory studies have played a role to confront this limitation but at a slow pace. However, with bigdata, researchers can be more creative to challenge previous studies and can conduct various preliminary analyses. For instance, many media and communication researchers are interested in interactions of people made via media. Considering that the most of the bigdata is generated from online media, now they receive more opportunities to look deeply into the interactions and investigate hidden factors by accessing non-interventional digital traces. In this research, only the fifth analysis is directly related to the existing literature and the others are designed in a way to exhaustively use new features given by the bigdata of ung.no service. In this way, the research could suggest new findings that can be a starting point to expend the existing boundary of youth online help seeking. Bigdata research alone cannot lead a conclusion despite of their quantity, still academia can enjoy its benefits as seen in the literature review.

Another opportunity of bigdata research is to obtain more objectivity in the perspective of methodology. Such computational calculations and a well-designed machine learning model can better analyze data more impartially. For instance, when it comes to content analysis in media studies, coding meaning of the content can vary according to the subjectivity of each researcher. However, some of text mining methodology which have been introduced as a substitute of manual coding in the literature review, are based on a reliable criterion, that is the frequency of words in content. As if researchers prepare coding schemes for a content analysis, a pretrained model must be premised when it comes to supervised machine learning. Nonetheless, once a machine learns a classification model and proves its high-performance level, a consistent analysis will be automated regardless of researchers. Furthermore, such as the sixth and seventh analyses of this research, there are unsupervised machine learning ways to implement text mining. Based on statistical algorithms, in this case, topic modeling and clustering are implemented. The principles of LDA and K-means are static, and they can be applied equivalently in any data. Despite parameters affecting results, there are also ways of designating their values by mathematical optimization. In this research, by adopting a coherence score and elbow method, it was possible to avoid subjective decision of the researcher.

Being able to cross analytical levels of data is the other opportunity. As it is already pointed out in the literature review, bigdata has not only a horizontally but also a vertically large spectrum. Then, it is possible to approach research subjects in a macro level to micro level. For instance, in this research, some analyses were intended to understand help seeking behaviors by treating young people as a whole single unit, while in the others they were sorted by gender and age. By doing so, both general and demographic specific tendencies are found in one research. To be more specific, the first and fourth analyses are complementary in a way showing general time trends of help seeking and revealing differences between age and gender groups. In the fourth analysis, the most frequent words from the whole help seeking content were found, yet the seventh analysis emphasized there are dominant themes in each demographic group. This characteristic of bigdata may particularly be useful for Public Relations in the context of communication studies. Such as audience segmentation which aims to identify different internal groups across a whole public can be more elaborated by controlling analytical levels.

In contrast, there are also concerns and challenges using bigdata in research. First, finding good data is difficult. As words such as “infobesity” and “information overload” imply the negative side of the bigdata era, collecting data suitable for a research purpose and removing noise heavily influences the quality of research. In this study, initially 302,989 documents had been collected, but the actual number of the data used for the analyses was 285,447 decreased by 5.79% after filtering some data with errors such as blanks. While bigdata is continuously generated in online media, not all of them contain value as research. Furthermore, the good data can be also interpreted as ethically clean data. Fortunately, the used data from ung.no guarantee anonymity of the help seeker, yet online media ask people to fill in personal information to become users and have access authority on their digital activities. As traditional methodology has always emphasized the protection of informants, their online privacy should be also secured.

Second, the large amount of data does not guarantee the representativeness of a research subject. Representativeness of sample is one of the typical precautions in traditional research, yet this apply same in bigdata research regardless of the size of data. Suppose a media researcher tries to study a certain subject through of social media, bigdata only represents behavior or opinion of users from selected social media and those of who are willing to communicate. In this way, even in this research, the data are biased in two aspects. One is that the results only describes Norwegian young people who actually approached ung.no for the purpose of help seeking, and the other is that 66% of the data was written by girls. This will be mentioned later as one of the limitations of this research. Such like, more strict standards for generalizing the results of big data research is required.

The last challenge is that a direction and results of research can vary depending on researchers' knowledge and interpretation of bigdata analytics. As mentioned, machine learning methods in particular have a number of parameters that make a difference in results. Therefore, researchers should know much about statistics and the use of machine learning algorithms. Furthermore, there is a problem of subjectivity of researchers in qualitative interpretation in bigdata research. For instance, for the sixth and seventh analyses in this research, LDA and K-means algorithms were used applying coherence score measurements and the elbow method to obtain objectivity in methodology. Thus, the optimal numbers of topics and clusters were found by the machine, yet it was entirely up to the researcher to name and interpret outputs. In this regard, academia may develop some guidelines to assure validity in bigdata analysis.

5.3 Limitation of the research and future recommendation

The limitations of this research mostly stem from the inherent challenges of bigdata research that were previously introduced. First, the result does not represent the whole youth in Norway, but the ones who are willing to reach this particular online support service. Then, young people who use other online sources such as online communities and forums or web search engines may have different tendencies of help seeking. In addition, the number of data could not be collected equally for each demographic group so that some of the results might be biased to girls. However, it was also intended to see the composition of online help seekers without any manipulation, so that a meaningful outcome was discovered, that young girls are the most active in asking for a help among others. Furthermore, through subgroup analyses, some part of the limitation has been resolved. In future research, depending on the purpose of the research, it would be possible to include more online media and apply an appropriate sampling method.

Meanwhile, there is a lack of ways to avoid filtering typing mistakes in text data. However, they can be minimized by repetitive correction or better managing stopwords in the future research. When it comes to preprocessing in text mining, it is also recommended to create a thesaurus[32] which treats synonyms as one word, accordingly, allows more efficient calculation. Moreover, utilizing more advanced NLP techniques, one may try even detailed and more accurate result. For example, the structure of a sentence can be analyzed by distinguishing the subject, object, modifier, etc. with a parser[33].

In the next page, the final chapter will provide the summary of this research and suggest its academic and social contributions

[32] Thesaurus indicates a dictionary of synonyms where, for instance, the words both have meaning of ‘mom’ in Norwegian, ‘mor’ and ‘mamma’ are treated as one word.

[33] In natural language, “parser is a program that works out the grammatical structure of sentences, for instance, which groups of words go together (as "phrases") and which words are the subject or object of a verb” ("Parser," 2020).

6. Conclusion

6.1 Summary

Youth today as digital natives, it has been reported that they are more likely to use online media for help seeking rather than face to face interaction. According to previous literature, online help seeking offers the benefits of accessibility and self-expression for young people, and above all things, they feel free from being stigmatized compared to direct personal relationships. While scholars point out that online environment have changed the nature of help seeking, their themes and way of research sticks to traditional ones, notably focusing on motivational and obstructive factors for help seeking by using survey and interview. Throwing a doubt to this, this research claimed a necessity of research using bigdata accumulated online everyday reflecting actual social phenomena, not collected by human intervention. Such quantitative and qualitative values of bigdata are well acknowledged in business, yet their epistemological challenge is still significant in academia. This is because the philosophy of bigdata in which data itself can make a conclusion that is out of the frame of science studies where a theory and testing its hypothesis are prioritized. However, benefits of using bigdata have gradually gotten more attention in terms of its function of exploration. Then, a hybrid data-driven research that combines induction and deduction used to be introduced. In other words, bigdata research is expected to establish new hypotheses and re-examine theories through inductive analysis, before a deductive process. In this background, this research set two research questions which are; (1) "How can bigdata be used to understand youth online help seeking?" and (2) "What are the opportunities and challenges when using bigdata in media research?".

Afterwards, as research data, bigdata generated from the question answering service of ung.no, a Norwegian youth support website was taken. To be more specific, the bigdata consists of digital traces of the youth, that are timestamps automatically saved when asking for help, and demographic information and query text that young people voluntarily filled out. By applying statistical analysis, and text mining including lexicometrics and machine learning techniques, a total of seven analyses were implemented to find hidden patterns of help seeking behavior and content at the level of the whole youth population as well as at the subgroup level. These seven analyses were not designed to test only specific hypotheses reflecting existing research and theories of youth help seeking, but to attempt to find new patterns implied by bigdata itself. Therefore, some of results are consistent with what have been reported in the existing literature and others provide a broader understanding of online youth help seeking by including various factors that have not been observed before. Accordingly, an answer to the first research question can be suggested. That is, taking previous literature as a guide, while exploiting the abundance of big data itself as time, demographic and query text are heuristically used in this research.

For the second questions dealing with opportunities and challenges of bigdata in research and particularly in the field of media, the following are suggested. First, one of the main opportunities that bigdata provides for researchers is creativity in research perspective and scope. Especially, given that most of the big data is created in online media, communication researchers can now look into hidden variables that were not found in theories by accessing naturally occurring digital traces. In the perspective of methodology, bigdata analytics can offer better objectivity than traditional research tools. When it comes to coding in media content analysis, text mining methods based on statistics and machine learning may solve some of the problem of subjectivity and reliability researchers. Furthermore, by utilizing a vertical spectrum of bigdata, different analytical levels can be adopted in research. For instance, this facilitates performing segmentation of the public so that PR researchers make more elaborated comparisons between subgroups beyond understanding the entire public.

On the other hand, as one of challenges, difficulties in finding bigdata which is suitable for research and ethically clean. As a lot of online media contain noise data and user information for which the level of consent is uncertain, media researchers should be particularly cautious of the quality of bigdata. Furthermore, bigdata does not ensure representativeness of sample. Researchers should not be misled by the amount of bigdata but recognize that it represents data of people who use certain online media and are willing to communicate. Lastly, in terms of dealing with bigdata, researchers' knowledge about bigdata analytics can be significant. For interpreting results, some guidelines for validity may be required. Thus, not only efforts from individual media researchers but also support from academia can be necessary for mitigating this challenge.

Meanwhile, at the beginning of this research, it was seen that existing help seeking literature could be used as evidence for enhancing health intervention for young people.

However, in 2007, the WHO pointed out that they were still insufficient to support actual intervention policies due to the lack of understanding actual behavior and the perspective of the youth. On the other hand, with bigdata it was possible to realize what they suggested, which are therapeutic narrative and social marketing approach. Therefore, in the next subsection, how this study can contribute to society will be briefly mentioned together with its academic service.

6.2 Academic and social contribution of the research

This research can contribute to both academy and society. More specifically, there are results in accordance with existing literature while the research broadens insights into difference between demographic groups. This may lead more researchers to study online help seeking of a specific group, presumably actively using bigdata. Certain bigdata alone cannot establish a theory, but if related studies create sufficient theoretical basis, they can be eventually developed into a more sophisticated framework of youth help seeking and youth online help seeking.

At the same time, this research can propose new types of help intervention that are differentiated from existing ones. This is enabled by the characteristics of bigdata which have similar effects of therapeutic narrative and social marketing approach that WHO suggested in the past. For example, one can see the need of segmented intervention reflecting distinctions of youth by their age and gender as well as a youth support service that is available for 24 hours. These findings stimulate the development of a help intervention combining with artificial intelligence such as chatbot that makes prediction based on personal traits of help seekers found during conversation. This is in line with the viewpoint of the ICT field about help seeking introduced earlier in the beginning of the study, which does not divide the boundary of help from humans and non-humans.

Appendix

Appendix 1. A total stopword list. They are a combination of preset from Norwegian Bokmål language model of spCay library and the ones determined by the researcher.

```
'når', 'kontakt', 'annet', 'bak', 'viktig', 'and', 'måneder', 'snilt',
'hjelp', 'vant', 'mot', 'tirsdag', 'hvordan', 'han', 'runde', 'mye', 'mens',
'ett', 'annen', 'denne', 'langt', 'norsk', 'det', 'også', 'gang', 'gangene',
'beste', 'spørsmål', 'saken', 'først', 'opplyser', 'prosent', 'andre',
'ville', 'kom', 'gjør', 'tider', 'måten', 'politidistrikt', 'søndag', 'uten',
'svaret', 'for', 'hilse', 'måneden', 'hørte', 'funnet', 'person', 'ho',
'sine', 'land', 'både', 'nei', 'en', 'timene', 'ntb', 'fra', 'meg', 'fire',
'gikk', 'som', 'nå', 'laget', 'hei', 'kroner', 'kunne', 'mange', 'synes',
'tatt', 'folk', 'by', 'svært', 'måned', 'gå', 'mars', 'dermed', 'kommer',
'kort', 'åring', 'fredag', 'kamp', 'bris', 'hørt', 'nye', 'la', 'der',
'frankrike', 'tingen', 'fikk', 'disse', 'svar', 'året', 'stor', 'msci',
'tilbake', 'inn', 'etter', 'uken', 'hilsen', 'time', 'spørsmålet', 'dette',
'må', 'oslo', 'usa', 'klart', 'månedene', 'alt', 'videre', 'dei', 'ifølge',
'mål', 'hun', 'poeng', 'tingene', 'seg', 'snill', 'er', 'ny', 'september',
'skulle', 'litt', 'forhånd', 'ting', 'oss', 'kan', 'ønsker', 'ganger',
'igjen', 'leder', 'vår', 'og', 'om', 'del', 'gi', 'viser', 'tok', 'mandag',
'mer', 'alle', 'snille', 'gjorde', 'man', 'her', 'skal', 'til', 'landet',
'drept', 'millioner', 'hjelpen', 'rundt', 'vært', 'enn', 'like', 'personen',
'måtte', 'sier', 'd', 'mener', 'var', 'timer', 'kvinner', 'torsdag', 'tid',
'så', 'i', 'ja', 'vi', 'sted', 'et', 'hadde', 'noe', 'over', 'ingen', 'ble',
'samtidig', 'blant', 'uke', 'nok', 'utenfor', 'gangen', 'den', 'tiden',
'helt', 'dager', 'flere', '%', 'henne', 'god', 'opp', 'høre', 'fem', 'på',
'fått', 'tillegg', 'bedre', 'godt', 'veldig', 'satt', 'ser', 'se', 'ut',
'samme', 'eller', 'kveld', 'mellom', 'ham', 'president', 'blitt', 'få',
'hvor', 'dagene', 'noen', 'uka', 'fortsatt', 'gjort', 'tror', 'da', 'siden',
'blir', 'ha', 'stedet', 'dag', 'skriver', 'dagen', 'norge', 'av', 'hu', 'å',
'har', 'minutter', 'høres', 'mannen', 'sin', 'under', 'gjennom', 'ukene',
'regjeringen', 'sagt', 'heisann', 'allerede', 'heie', 'årene', 'tyskland',
'års', 'personer', 'selvom', 'hører', 'fjor', 'eg', 'bli', 'to', 'de', 'men',
'jeg', 'timen', 'hans', 'tak', 'fotball', 'hennes', 'kl', 'sa', 'menn',
'ligger', 'slik', 'vil', 'år', 'tidene', 'plass', 'være', 'første', 'gjøre',
'komme', 'ta', 'tidligere', 'du', 'russland', 'at', 'ned', 'ikke', 'kampen',
'bare', 'sverige', 'dem', 'står', 'personene', 'går', 'med', 'siste', 'sett',
'måte', 'onsdag', 'å', 'senere', 'uker', 'seks', 'får', 'fram', 'før',
'selv', 'sitt', 'hele', 'hjelpe', 'tre', 'ved', 'takk', 'norske', 'lørdag',
'neste', 'politiet', 'grunn', 'stund', 'mennesker', 'sammen', 'store',
'hvorfor', 'løpet', 'hva'
```

Appendix 2. Detailed numbers in demographic information of the help seekers from ung.no.

	Girl	Boy	Other	Age subtotal	(%)
0	1,531	638	38	2,207	1%
12	320	177	0	497	0%
13	36,775	12,606	555	49,936	17%
14	31,281	12,530	428	44,239	15%
15	33,837	15,790	425	50,052	18%
16	27,669	14,931	366	42,966	15%
17	20,350	11,613	294	32,257	11%
18	13,775	8,889	200	22,864	8%
19	9,351	5,716	143	15,210	5%
20	14,595	10,296	328	25,219	9%
Gender subtotal	189,484	93,186	2,777	total 285,447	

Appendix 3. Bigrams of each communication keyword from the fifth analysis.

Word	Co-occurred words(number)
fortelle	(('fortelle', 'min'), 1744), (('fortelle', 'mine'), 1572), (('vite', 'fortelle'), 1559), (('min', 'fortelle'), 1490), (('fortelle', 'vite'), 1314), (('fortelle', 'forelder'), 1124), (('fortelle', 'mamma'), 968), (('fortelle', 'føle'), 946), (('venn', 'fortelle'), 872), (('fortelle', 'venn'), 871), (('fortelle', 'fortelle'), 870), (('snakke', 'fortelle'), 860), (('fortelle', 'like'), 845), (('mine', 'fortelle'), 729), (('fortelle', 'fordi'), 664), (('dere', 'fortelle'), 649), (('fortelle', 'snakke'), 617), (('prøve', 'fortelle'), 607), (('føle', 'fortelle'), 601), (('hvis', 'fortelle'), 578), (('lure', 'fortelle'), 570), (('fortelle', 'jente'), 558), (('fortelle', 'redd'), 551), (('jente', 'fortelle'), 547), (('fortelle', 'si'), 516), (('fortelle', 'skje'), 514), (('redd', 'fortelle'), 513), (('skole', 'fortelle'), 506), (('fordi', 'fortelle'), 504), (('like', 'fortelle'), 497), (('mamma', 'fortelle'), 462), (('fortelle', 'gutt'), 447), (('aldri', 'fortelle'), 444), (('gutt', 'fortelle'), 439), (('fortelle', 'pappa'), 436), (('fortelle', 'kjæreste'), 429), (('tenke', 'fortelle'), 428), (('burde', 'fortelle'), 421), (('lyst', 'fortelle'), 419), (('si', 'fortelle'), 418), (('begynne', 'fortelle'), 405), (('fortelle', 'mor'), 400), (('fortelle', 'begynne'), 385), (('fortelle', 'lure'), 381), (('fortelle', 'problem'), 377), (('fortelle', 'hvis'), 368), (('kjæreste', 'fortelle'), 358), (('skje', 'fortelle'), 357), (('fortelle', 'prøve'), 357), (('fortelle', 'tenke'), 356), (('fortelle', 'skole'), 352), (('fortelle', 'aldri'), 350), (('forelder', 'fortelle'), 350), (('fortelle', 'venninne'), 337), (('problem', 'fortelle'), 336), (('dag', 'fortelle'), 332), (('venninne', 'fortelle'), 320), (('spørre', 'fortelle'), 316), (('fortelle', 'spørre'), 314), (('fortelle', 'følelse'), 289), (('mitt', 'fortelle'), 287), (('fortelle', 'dere'), 283), (('fortelle', 'burde'), 279), (('finne', 'fortelle'), 270), (('fortelle', 'familie'), 270), (('fortelle', 'ha'), 269), (('heller', 'fortelle'), 268), (('fortelle', 'sånn'), 268), (('fortelle', 'jo'), 268), (('klarere', 'fortelle'), 268), (('klasse', 'fortelle'), 262), (('tørre', 'fortelle'), 261), (('pappa', 'fortelle'), 252), (('egentlig', 'fortelle'), 252), (('jo', 'fortelle'), 249), (('fortelle', 'egentlig'), 241), (('fortelle', 'trenge'), 239), (('hver', 'fortelle'), 237), (('sånn', 'fortelle'), 237), (('fortelle', 'dag'), 234), (('tid', 'fortelle'), 233), (('lege', 'fortelle'), 233), (('fortelle', 'slite'), 231), (('sex', 'fortelle'), 231), (('fortelle', 'heller'), 230), (('alltid', 'fortelle'), 225), (('fortelle', 'finne'), 223), (('fortelle', 'mitt'), 223), (('ringe', 'fortelle'), 222), (('mor', 'fortelle'), 221), (('fortelle', 'lærer'), 221), (('holde', 'fortelle'), 219), (('ha', 'fortelle'), 218), (('fortelle', 'sex'), 215), (('fortelle', 'klarere'), 213), (('vanskelig', 'fortelle'), 212), (('mi', 'fortelle'), 209), (('fortelle', 'klasse'), 208), (('fortelle', 'lyst'), 207), (('fortelle', 'kanskje'), 206), (('fortelle', 'holde'), 205), (('følelse', 'fortelle'), 201), (('psykolog', 'fortelle'), 199), (('trenge', 'fortelle'), 195), (('slite', 'fortelle'), 194), (('fortelle', 'alltid'), 192), (('vond', 'fortelle'), 191), (('fortelle', 'tid'), 188), (('fortelle', 'bra'), 185), (('hjelp', 'fortelle'), 182), (('fortelle', 'hverandre'), 181), (('fortelle', 'glad'), 180), (('fortelle', 'sikker'), 179), (('fortelle', 'hvem'), 178), (('bra', 'fortelle'), 177), (('fortelle', 'fare'), 176), (('leng', 'fortelle'), 175), (('lærer', 'fortelle'), 175), (('ganske', 'fortelle'), 175), (('tørr', 'fortelle'), 174), (('stole', 'fortelle'), 173), (('fortelle', 'liv'), 173), (('fortelle', 'psykolog'), 173), (('kanskje', 'fortelle'), 171), (('fortelle', 'hjelp'), 170), (('fortelle', 'ganske'), 169), (('fortelle', 'hver'), 169), (('fortelle', 'lei'), 168), (('uke', 'fortelle'), 167), (('glad', 'fortelle'), 167), (('kjenne', 'fortelle'), 167), (('fortelle', 'lege'), 165), (('hos', 'fortelle'), 163), (('fortelle', 'vond'), 162), (('fortelle', 'tro'), 162), (('fortelle', 'rett'), 162), (('lei', 'fortelle'), 162), (('virkelig', 'fortelle'), 161), (('bestevenn', 'fortelle'), 161), (('fortelle', 'vanskelig'), 161), (('familie', 'fortelle'), 161), (('helsesøster', 'fortelle'), 160), (('sikker', 'fortelle'), 159), (('hjem', 'fortelle'), 157), (('fortelle', 'sannhet'), 155), (('sende', 'fortelle'), 155), (('oft', 'fortelle'), 155), (('møte', 'fortelle'), 155), (('fortelle', 'sende'), 154), (('liv', 'fortelle'), 152), (('forhold', 'fortelle'), 152), (('bør', 'fortelle'), 152), (('fortelle', 'historie'), 152), (('fortelle', 'dårlig'), 150), (('dra', 'fortelle'), 149), (('fortelle', 'mi'), 149), (('person', 'fortelle'), 147), (('hjemme', 'fortelle'), 146),

	(('fortelle', 'gammel'), 145), (('utrolig', 'fortelle'), 145), (('fortelle', 'helsesøster'), 145), (('fortelle', 'slå'), 144), (('fortelle', 'nesten'), 144), (('fortelle', 'slutte'), 144), (('fortelle', 'oft'), 143), (('liksom', 'fortelle'), 142), (('fortelle', 'leng'), 142), (('rett', 'fortelle'), 141), (('fortelle', 'vise'), 141), (('fortelle', 'kjenne'), 141), (('hverandre', 'fortelle'), 141), (('fortelle', 'utrolig'), 140), (('fortelle', 'person'), 139), (('nesten', 'fortelle'), 139), (('fortelle', 'virke'), 139), (('vise', 'fortelle'), 138), (('fortelle', 'bruke'), 137), (('fortelle', 'håpe'), 137), (('gjerne', 'fortelle'), 137), (('gråte', 'fortelle'), 135), (('fortelle', 'hjemme'), 135), (('måned', 'fortelle'), 135), (('fortelle', 'forhold'), 135), (('gammel', 'fortelle'), 134), (('enda', 'fortelle'), 134), (('fortelle', 'ta'), 134), (('fortelle', 'møte'), 133), (('fortelle', 'forstå'), 133), (('lenger', 'fortelle'), 132), (('dårlig', 'fortelle'), 132), (('fortelle', 'liksom'), 132), (('eneste', 'fortelle'), 132), (('håpe', 'fortelle'), 131), (('fortelle', 'forelske'), 131), (('fortelle', 'hos'), 130), (('fortelle', 'skjønne'), 129), (('fortelle', 'virkelig'), 129), (('fortelle', 'ei'), 128), (('snill', 'fortelle'), 127), (('hjelpe', 'fortelle'), 127), (('fortelle', 'bestevenn'), 126), (('fortelle', 'lov'), 124), (('fortelle', 'stole'), 123), (('hvem', 'fortelle'), 123), (('derfor', 'fortelle'), 123), (('skrive', 'fortelle'), 122), (('fortelle', 'elske'), 122), (('slå', 'fortelle'), 122), (('fortelle', 'skrive'), 122), (('tips', 'fortelle'), 122), (('lov', 'fortelle'), 121), (('usikker', 'fortelle'), 121), (('gravid', 'fortelle'), 119), (('god', 'fortelle'), 119), (('fortelle', 'grei'), 118), (('uansett', 'fortelle'), 116), (('fortelle', 'sur'), 116), (('fortelle', 'deg'), 116), (('fortelle', 'ho'), 116), (('forelske', 'fortelle'), 114), (('fortelle', 'eneste'), 114), (('fortelle', 'enda'), 113), (('fortelle', 'dra'), 113), (('forstå', 'fortelle'), 112), (('ingenting', 'fortelle'), 112), (('fortelle', 'masse'), 110), (('klare', 'fortelle'), 110), (('fortelle', 'flytte'), 110), (('fortelle', 'fin'), 109), (('fortelle', 'gå'), 109), (('osv', 'fortelle'), 108), (('fortelle', 'uke'), 108), (('fortelle', 'spise'), 108), (('fortelle', 'hjelpe'), 107), (('tro', 'fortelle'), 107), (('starte', 'fortelle'), 107), (('svare', 'fortelle'), 107), (('bruke', 'fortelle'), 107), (('fortelle', 'jobbe'), 106), (('fortelle', 'lære'), 106), (('fortelle', 'usikker'), 106), (('grei', 'fortelle'), 105), (('snar', 'fortelle'), 105), (('fortelle', 'faktisk'), 105), (('prate', 'fortelle'), 104), (('fortelle', 'sånt'), 104), (('fortelle', 'osv'), 103), (('skjønne', 'fortelle'), 103), (('fortelle', 'forelsket'), 103), (('fortelle', 'bør'), 103), (('fortelle', 'hemmelighet'), 102), (('slutte', 'fortelle'), 102), (('fortelle', 'barnevern'), 101), (('virke', 'fortelle'), 101), (('fortelle', 'ingenting'), 100), (('flytte', 'fortelle'), 100), (('ta', 'fortelle'), 100))
spørre	(('vite', 'spørre'), 2492), (('spørre', 'svare'), 1919), (('spørre', 'spørre'), 1888), (('spørre', 'vite'), 1876), (('min', 'spørre'), 1802), (('spørre', 'min'), 1670), (('spørre', 'like'), 1588), (('like', 'spørre'), 1577), (('spørre', 'nei'), 1505), (('venn', 'spørre'), 1410), (('snakke', 'spørre'), 1400), (('spørre', 'dere'), 1313), (('spørre', 'fordi'), 1304), (('spørre', 'finne'), 1224), (('spørre', 'jente'), 1188), (('jente', 'spørre'), 1171), (('gutt', 'spørre'), 1150), (('spørre', 'venn'), 1090), (('spørre', 'snakke'), 1036), (('spørre', 'gutt'), 968), (('mine', 'spørre'), 961), (('hvis', 'spørre'), 949), (('spørre', 'mamma'), 940), (('lure', 'spørre'), 887), (('skole', 'spørre'), 884), (('spørre', 'lure'), 871), (('spørre', 'si'), 861), (('tenke', 'spørre'), 822), (('begynne', 'spørre'), 803), (('spørre', 'føle'), 799), (('spørre', 'mine'), 770), (('spørre', 'hjelp'), 747), (('si', 'spørre'), 726), (('finne', 'spørre'), 724), (('fordi', 'spørre'), 712), (('prøve', 'spørre'), 706), (('svare', 'spørre'), 705), (('spørre', 'jo'), 702), (('spørre', 'skole'), 700), (('spørre', 'tenke'), 691), (('aldri', 'spørre'), 686), (('spørre', 'hvis'), 685), (('føle', 'spørre'), 671), (('klasse', 'spørre'), 667), (('spørre', 'prøve'), 645), (('spørre', 'begynne'), 643), (('spørre', 'aldri'), 641), (('lyst', 'spørre'), 632), (('mamma', 'spørre'), 622), (('spørre', 'alltid'), 615), (('dag', 'spørre'), 614), (('tørr', 'spørre'), 605), (('spørre', 'sex'), 593), (('tørre', 'spørre'), 581), (('spørre', 'sånn'), 580), (('dere', 'spørre'), 574), (('sende', 'spørre'), 563), (('alltid', 'spørre'), 554), (('spørre', 'redd'), 543), (('hver', 'spørre'), 528), (('redd', 'spørre'), 527), (('derfor', 'spørre'), 505), (('spørre', 'lyst'), 502), (('spørre', 'kjæreste'), 501), (('jo', 'spørre'), 488), (('spørre', 'sende'), 487), (('ringe', 'spørre'), 473), (('spørre', 'bra'), 468), (('spørre', 'forelder'), 458), (('spørre', 'lov'), 456), (('spørre', 'skje'), 454), (('spørre', 'hvem'), 450), (('egentlig', 'spørre'), 446), (('sånn', 'spørre'), 442), (('mitt', 'spørre'), 427), (('nei', 'spørre'), 423), (('tid', 'spørre'), 421), (('spørre', 'oft'), 419), (('spørre', 'dag'), 417), (('spørre', 'klasse'), 417), (('kjæreste', 'spørre'), 410), (('spørre', 'tid'), 405), (('spørre', 'egentlig'), 394), (('burde', 'spørre'), 393), (('spørre', 'trenge'), 387), (('sex', 'spørre'), 384), (('kanskje', 'spørre'), 384), (('oft', 'spørre'), 377), (('venninne', 'spørre'), 375), (('spørre', 'ha'), 374), (('heller', 'spørre'), 372), (('spørre', 'kanskje'), 370), (('liksom', 'spørre'), 367), (('spørre', 'pappa'), 362), (('spørre', 'ho'), 356), (('melding', 'spørre'), 351), (('spørre', 'problem'), 348), (('spørre', 'heller'), 342), (('spørre', 'mitt'), 341), (('rett', 'spørre'), 340), (('spørre', 'håpe'), 337), (('lærer', 'spørre'), 334), (('spørre', 'lærer'), 334), (('spørre', 'venninne'), 332), (('leng', 'spørre'), 331), (('skje', 'spørre'), 330), (('spørre', 'grei'), 327), (('mi', 'spørre'), 326), (('spørre', 'sikker'), 322), (('spørre', 'rett'), 322), (('bort', 'spørre'), 321), (('sikker', 'spørre'), 321), (('problem', 'spørre'), 320), (('spørre', 'fortelle'), 316), (('ganske', 'spørre'), 315), (('fortelle', 'spørre'), 314), (('hverandre', 'spørre'), 307), (('forelske', 'spørre'), 304), (('spørre', 'liksom'), 302), (('kjenne', 'spørre'), 301), (('spørre', 'burde'), 298), (('hjem', 'spørre'), 295), (('trenge', 'spørre'), 294), (('pappa', 'spørre'), 294), (('gammel', 'spørre'), 293), (('spørre', 'dra'), 293), (('spørre', 'kjenne'), 292), (('hjelp', 'spørre'), 291), (('hvem', 'spørre'), 287), (('bruke', 'spørre'), 287), (('spørre', 'penger'), 285), (('forelder', 'spørre'), 285), (('spørre', 'vise'), 283), (('spørre', 'hver'), 282), (('uke', 'spørre'), 281), (('hos', 'spørre'), 281), (('bra', 'spørre'), 278), (('spørre', 'sånt'), 278), (('spørre', 'hjem'), 277), (('dra', 'spørre'), 275), (('nesten', 'spørre'), 275), (('spørre', 'skrive'), 273), (('rar', 'spørre'), 272), (('spørre', 'ganske'), 270), (('spørre', 'mor'), 269), (('spørre', 'bruke'), 269), (('spørre', 'gammel'), 268), (('spørre', 'lege'), 262), (('spørre', 'holde'), 261), (('spørre', 'jobbe'), 261), (('lege', 'spørre'), 261), (('spørre', 'kjøpe'), 257), (('spørre', 'hos'), 255), (('lov', 'spørre'), 255), (('møte', 'spørre'), 254), (('skrive', 'spørre'), 251), (('spørre', 'bilde'), 251), (('spørre', 'spørsmål'), 251), (('spørre', 'deg'), 250), (('spørre', 'jobb'), 250), (('lei', 'spørre'), 246), (('spørre', 'sur'), 246), (('spørre', 'osv'), 245), (('spørre', 'råd'), 245), (('spørre', 'tips'), 244), (('spørre', 'ei'), 242), (('holde', 'spørre'), 240), (('spørre', 'møtes'), 239), (('jobb', 'spørre'), 239), (('spørre', 'lei'), 234), (('vise', 'spørre'), 234), (('spørre', 'mi'), 233), (('spørre', 'leng'), 232), (('usikker', 'spørre'), 231), (('spørre', 'virke'), 228), (('spørre', 'nesten'), 227), (('spørre', 'hu'), 227), (('sånt', 'spørre'), 226), (('osv', 'spørre'), 225), (('spørre', 'møte'), 222), (('spørre', 'sitte'), 221), (('bilde', 'spørre'), 220), (('flau', 'spørre'), 218), (('spørre', 'skjønne'), 217), (('penger', 'spørre'), 215), (('alene', 'spørre'), 214), (('ta', 'spørre'), 214), (('spørre', 'forelske'), 213), (('grei', 'spørre'), 212), (('skjønne', 'spørre'), 212), (('spørre', 'usikker'), 212), (('ha', 'spørre'), 211), (('spørre', 'gale'), 210), (('tør', 'spørre'), 208), (('spørsmål', 'spørre'), 206), (('dum', 'spørre'), 205), (('sur', 'spørre'), 205), (('mor', 'spørre'), 203), (('spørre', 'klarere'), 202), (('spørre', 'tørre'), 202), (('time', 'spørre'), 197), (('klarere', 'spørre'), 195), (('jobbe', 'spørre'), 195), (('glad', 'spørre'), 192), (('spørre', 'hvilken'), 192), (('vanskelig', 'spørre'), 192), (('spørre', 'mulig'), 191), (('slå', 'spørre'), 190), (('følelse', 'spørre'), 190), (('spørre', 'slite'), 189), (('virke', 'spørre'), 188), (('spørre', 'flytte'), 188), (('hjemme', 'spørre'), 187), (('spørre', 'snar'), 186), (('håpe', 'spørre'), 185), (('snar', 'spørre'), 184), (('hjelpe', 'spørre'), 184), (('spørre', 'fin'), 184), (('gjerne', 'spørre'), 183), (('spørre', 'hjemme'), 182), (('tips', 'spørre'), 180), (('spørre', 'ta'), 179), (('spørre', 'hilsen'), 177), (('bør', 'spørre'), 176), (('vond', 'spørre'), 174), (('spørre', 'ringe'), 174), (('grunne', 'spørre'), 174), (('slutte', 'spørre'), 173), (('spørre', 'bore'), 171), (('lenger', 'spørre'), 170), (('spørre', 'rar'), 170), (('små', 'spørre'), 169), (('ho', 'spørre'), 169), (('spørre', 'gi'), 168), (('spørre', 'kysse'),

	167), (('spørre', 'virkelig'), 165), (('spørre', 'derfor'), 164), (('bestevenn', 'spørre'), 164), (('spørre', 'spise'), 164), (('dårlig', 'spørre'), 163), (('måned', 'spørre'), 163), (('spørre', 'alene'), 162), (('spørre', 'hjelpe'), 161), (('spørre', 'bytte'), 159), (('legge', 'spørre'), 159), (('spørre', 'følelse'), 158), (('spørre', 'vente'), 158), (('plutselig', 'spørre'), 157), (('spørre', 'svar'), 157), (('bore', 'spørre'), 155), (('masse', 'spørre'), 155), (('spørre', 'slutte'), 155), (('merke', 'spørre'), 155), (('utrolig', 'spørre'), 154), (('spørre', 'hverandre'), 154), (('starte', 'spørre'), 153), (('person', 'spørre'), 153), (('spørre', 'forhold'), 153), (('ingenting', 'spørre'), 152), (('spørre', 'melding'), 152), (('mulig', 'spørre'), 152), (('deg', 'spørre'), 151), (('spørre', 'uke'), 151), (('kompis', 'spørre'), 150), (('spørre', 'vond'), 150), (('spørre', 'utrolig'), 150), (('ei', 'spørre'), 149), (('gange', 'spørre'), 149), (('spørre', 'enda'), 149), (('ikkje', 'spørre'), 149), (('virkelig', 'spørre'), 149), (('spørre', 'flau'), 149), (('butikk', 'spørre'), 148), (('kjøpe', 'spørre'), 148), (('prate', 'spørre'), 147), (('familie', 'spørre'), 147), (('spørre', 'person'), 147), (('spørre', 'sann'), 147), (('spørre', 'legge'), 145), (('spørre', 'bestevenn'), 145), (('fin', 'spørre'), 145), (('spørre', 'glad'), 144), (('spørre', 'synes'), 144), (('feil', 'spørre'), 144), (('vente', 'spørre'), 143), (('spørre', 'slå'), 142), (('forhold', 'spørre'), 142), (('spørre', 'masse'), 142), (('snill', 'spørre'), 142), (('spørre', 'tørr'), 141), (('forelsket', 'spørre'), 141), (('spørre', 'tro'), 140), (('sitte', 'spørre'), 140), (('spørre', 'søke'), 139), (('spørre', 'normal'), 139), (('spørre', 'familie'), 139), (('spørre', 'synse'), 139), (('slutt', 'spørre'), 138), (('spørre', 'akkurat'), 138), (('samtale', 'spørre'), 138), (('kysse', 'spørre'), 137), (('spørre', 'selvfølgelig'), 137), (('spørre', 'ingenting'), 137), (('spørre', 'dårlig'), 137), (('spørre', 'lære'), 136), (('tro', 'spørre'), 136), (('spørre', 'slutt'), 135), (('nett', 'spørre'), 135), (('spørre', 'dum'), 135), (('riktig', 'spørre'), 134), (('synes', 'spørre'), 134), (('spørre', 'fore'), 134), (('søke', 'spørre'), 133), (('stille', 'spørre'), 133), (('spørre', 'gjerne'), 133), (('spørre', 'helsesøster'), 132), (('begge', 'spørre'), 132), (('enda', 'spørre'), 132), (('spørre', 'gå'), 131), (('glemme', 'spørre'), 131), (('uansett', 'spørre'), 131), (('altså', 'spørre'), 130), (('spørre', 'kompis'), 130), (('spørre', 'god'), 130), (('eneste', 'spørre'), 129), (('akkurat', 'spørre'), 129), (('god', 'spørre'), 127), (('spørre', 'time'), 126), (('spørre', 'små'), 126), (('hu', 'spørre'), 126), (('spørre', 'bør'), 126), (('spørre', 'altså'), 25), (('spørre', 'drikke'), 124), (('gå', 'spørre'), 124), (('spørre', 'faktisk'), 123), (('facebook', 'spørre'), 123), (('forstå', 'spørre'), 123), (('spørre', 'mens'), 123), (('spørre', 'd'), 122), (('spørre', 'gange'), 122), (('ene', 'spørre'), 122), (('spørre', 'sove'), 121), (('normal', 'spørre'), 120), (('liv', 'spørre'), 120), (('spørre', 'ikkje'), 120), (('fest', 'spørre'), 120), (('spørre', 'helst'), 120), (('spørre', 'snill'), 119), (('spise', 'spørre'), 119), (('spørre', 'nekte'), 119), (('stoppe', 'spørre'), 119), (('spørre', 'vanskelig'), 118), (('spørre', 'bety'), 118), (('spørre', 'spille'), 118), (('spørre', 'starte'), 118), (('flytte', 'spørre'), 118), (('spørre', 'lenger'), 117), (('flørte', 'spørre'), 116), (('synse', 'spørre'), 116), (('gale', 'spørre'), 116), (('ca', 'spørre'), 116), (('spørre', 'drive'), 115), (('spørre', 'vanlig'), 115), (('lære', 'spørre'), 115), (('møtes', 'spørre'), 115), (('huske', 'spørre'), 113), (('spørre', 'fare'), 113), (('spørre', 'henge'), 113), (('gråte', 'spørre'), 113), (('faktisk', 'spørre'), 113), (('spørre', 'kino'), 113), (('spørre', 'uansett'), 113), (('spørre', 'mene'), 112), (('spørre', 'ligge'), 111), (('spørre', 'forelsket'), 111), (('spørre', 'eneste'), 111), (('spørre', 'bort'), 111), (('helsesøster', 'spørre'), 111), (('helst', 'spørre'), 111), (('kjempe', 'spørre'), 111), (('klein', 'spørre'), 111), (('par', 'spørre'), 11), (('spørre', 'plutselig'), 110), (('spørre', 'kjøre'), 110), (('gi', 'spørre'), 110), (('spørre', 'begge'), 109), (('stå', 'spørre'), 109), (('rom', 'spørre'), 109), (('spørre', 'navn'), 109), (('vis', 'spørre'), 109), (('elske', 'spørre'), 109), (('fore', 'spørre'), 108), (('mens', 'spørre'), 108), (('vanlig', spørre'), 107), (('spørre', 'følge'), 107), (('spørre', 'bo'), 107), (('spørre', 'nudes'), 105), (('spørre', 'feil'), 105), (('snap', 'spørre'), 105), (('be', 'spørre'), 105), (('spørre', 'lite'), 105), (('råd', 'spørre'), 104), (('spørre', 'liten'), 104), (('drive', 'spørre'), 104), (('vel', 'spørre'), 103), (('spørre', butikk'), 103), (('drikke', 'spørre'), 102), (('ute', 'spørre'), 102), (('bestemme', 'spørre'), 102), (('spørre', 'likevel'), 102), (('spørre', 'skikkelig'), 101), (('spørre', 'liv'), 101), (('spørre', 'betale'), 101), (('mye', 'spørre'), 101), (('pleie', 'spørre'), 101), (('spørre', 'syk'), 101), (('spørre', 'stå'), 100), (('interessere', 'spørre'), 100), (('snakk', 'spørre'), 100), (('liten', 'spørre'), 100), (('slite', 'spørre'), 100))
snakke	(('snakke', 'snakke'), 5826), (('vite', 'snakke'), 4275), (('snakke', 'vite'), 4138), (('min', 'snakke'), 3595), (('snakke', 'min'), 3539), (('prøve', 'snakke'), 3403), (('venn', 'snakke'), 3314), (('snakke', 'venn'), 3151), (('begynne', 'snakke'), 3055), (('like', 'snakke'), 3009), (('snakke', 'føle'), 2992), (('snakke', 'mine'), 2862), (('føle', 'snakke'), 2770), (('snakke', 'skole'), 2661), (('gutt', 'snakke'), 2513), (('snakke', 'jente'), 2508), (('skole', 'snakke'), 2397), (('snakke', 'gutt'), 2317), (('snakke', 'fordi'), 2303), (('jente', 'snakke'), 2297), (('snakke', 'mamma'), 2229), (('snakke', 'like'), 2169), (('mine', 'snakke'), 2030), (('aldri', 'snakke'), 1935), (('fordi', 'snakke'), 1874), (('snakke', 'helsesøster'), 1834), (('snakke', 'si'), 1733), (('snakke', 'forelder'), 1730), (('snakke', 'aldri'), 1669), (('snakke', 'prøve'), 1651), (('snakke', 'begynne'), 1627), (('klasse', 'snakke'), 1557), (('klarere', 'snakke'), 1531), (('snakke', 'problem'), 1475), (('mamma', 'snakke'), 1458), (('snakke', 'spørre'), 1400), (('si', 'snakke'), 1392), (('tenke', 'snakke'), 1364), (('problem', 'snakke'), 1332), (('snakke', 'sånn'), 1289), (('snakke', 'lure'), 1263), (('snakke', 'redd'), 1254), (('hver', 'snakke'), 1245), (('snakke', 'oft'), 1231), (('snakke', 'tenke'), 1224), (('snakke', 'hverandre'), 1221), (('redd', 'snakke'), 1189), (('lyst', 'snakke'), 1157), (('lure', 'snakke'), 1151), (('tid', 'snakke'), 1138), (('snakke', 'hver'), 1126), (('trenge', 'snakke'), 1103), (('hvis', 'snakke'), 1102), (('snakke', 'tid'), 1099), (('hvem', 'snakke'), 1098), (('sånn', 'snakke'), 1091), (('vanskelig', 'snakke'), 1090), (('snakke', 'alltid'), 1083), (('heller', 'snakke'), 1076), (('snakke', 'pappa'), 1064), (('snakke', 'klasse'), 1052), (('snakke', 'jo'), 1037), (('snakke', 'heller'), 1037), (('spørre', 'snakke'), 1036), (('alltid', 'snakke'), 1030), (('snakke', 'psykolog'), 1002), (('burde', 'snakke'), 994), (('snakke', 'hvis'), 986), (('snakke', 'nesten'), 957), (('kjenne', 'snakke'), 952), (('snakke', 'lærer'), 947), (('snakke', 'kjenne'), 945), (('oft', 'snakke'), 941), (('tørr', 'snakke'), 938), (('jo', 'snakke'), 936), (('kjæreste', 'snakke'), 924), (('snakke', 'hjelpe'), 923), (('snakke', 'klarere'), 923), (('slutte', 'snakke'), 911), (('nesten', 'snakke'), 910), (('hverandre', 'snakke'), 909), (('dag', 'snakke'), 891), (('snakke', 'følelse'), 885), (('snakke', 'ganske'), 882), (('snakke', 'dere'), 875), (('egentlig', 'snakke'), 861), (('snakke', 'fortelle'), 860), (('tørre', 'snakke'), 847), (('mitt', 'snakke'), 838), (('pappa', 'snakke'), 831), (('snakke', 'skje'), 825), (('forelder', 'snakke'), 799), (('hjelp', 'snakke'), 791), (('liksom', 'snakke'), 789), (('snakke', 'egentlig'), 785), (('snakke', 'finne'), 783), (('snakke', 'lyst'), 774), (('snakke', 'liksom'), 764), (('snakke', 'kjæreste'), 763), (('dere', 'snakke'), 760), (('snakke', 'dag'), 756), (('snakke', 'trenge'), 752), (('snakke', 'virke'), 746), (('slite', 'snakke'), 746), (('møte', 'snakke'), 731), (('ganske', 'snakke'), 728), (('snakke', 'familie'), 727), (('snakke', 'sende'), 727), (('finne', 'snakke'), 722), (('snakke', 'sex'), 715), (('skje', 'snakke'), 706), (('snakke', 'kanskje'), 705), (('person', 'snakke'), 705), (('lei', 'snakke'), 696), (('snakke', 'mor'), 694), (('snakke', 'venninne'), 677), (('snakke', 'hjelp'), 676), (('snakke', 'sånt'), 668), (('snakke', 'vanskelig'), 661), (('snakke', 'lege'), 657), (('følelse', 'snakke'), 645), (('snakke', 'slite'), 641), (('hjelpe', 'snakke'), 635), (('venninne', 'snakke'), 630), (('fortelle', 'snakke'), 617), (('snakke', 'møte'), 607), (('lærer', 'snakke'), 589), (('snakke', 'ha'), 586), (('snakke', 'lei'), 584), (('sur', 'snakke'), 582), (('snakke', 'svare'), 579), (('gammel', 'snakke'), 577), (('snakke', 'mitt'), 577), (('snakke', 'leng'), 576), (('snakke', 'forstå'), 576), (('kanskje', 'snakke'), 572), (('snakke', 'bra'), 572), (('snakke', 'burde'), 571), (('leng', 'snakke'), 570), (('forelske', 'snakke'), 569), (('holde', 'snakke'), 569), (('snakke', 'lenger'), 567), (('bra', 'snakke'), 567),

(('uke', 'snakke'), 564), (('snakke', 'masse'), 563), (('snakke', 'holde'), 559), (('ha', 'snakke'), 556),
(('snakke', 'osv'), 552), (('glad', 'snakke'), 551), (('helsesøster', 'snakke'), 545), (('alene', 'snakke'), 539),
(('virke', 'snakke'), 538), (('psykolog', 'snakke'), 536), (('snakke', 'person'), 531), (('eneste', 'snakke'), 530),
(('snakke', 'rådgiver'), 530), (('snakke', 'gammel'), 527), (('lenger', 'snakke'), 515), (('bort', 'snakke'), 514),
(('snakke', 'tips'), 512), (('vond', 'snakke'), 511), (('familie', 'snakke'), 509), (('snakke', 'utrolig'), 501),
(('sitte', 'snakke'), 499), (('virkelig', 'snakke'), 489), (('utrolig', 'snakke'), 485), (('snakke', 'ta'), 484),
(('snakke', 'facebook'), 484), (('snakke', 'time'), 482), (('snakke', 'bore'), 479), (('snakke', 'me'), 478),
(('snakke', 'samtale'), 478), (('dårlig', 'snakke'), 477), (('hjemme', 'snakke'), 473), (('bør', 'snakke'), 472),
(('mi', 'snakke'), 469), (('snakke', 'uke'), 468), (('sex', 'snakke'), 460), (('snakke', 'ei'), 457), (('snakke',
'sur'), 455), (('snakke', 'slutte'), 454), (('snakke', 'stygg'), 452), (('forhold', 'snakke'), 452), (('snakke',
'alene'), 452), (('snakke', 'glad'), 451), (('mor', 'snakke'), 444), (('tips', 'snakke'), 443), (('snakke',
'flytte'), 443), (('snakke', 'skjønne'), 442), (('snakke', 'ho'), 442), (('snakke', 'forelske'), 440), (('snakke',
'lære'), 439), (('snakke', 'vond'), 438), (('snakke', 'rett'), 436), (('liv', 'snakke'), 434), (('snakke',
'virkelig'), 434), (('snakke', 'begge'), 434), (('snakke', 'stole'), 432), (('svare', 'snakke'), 430), (('snakke',
'små'), 428), (('snakke', 'håpe'), 425), (('rett', 'snakke'), 422), (('sende', 'snakke'), 418), (('god', 'snakke'),
418), (('snakke', 'hu'), 418), (('ikkje', 'snakke'), 417), (('skjønne', 'snakke'), 416), (('snakke', 'eneste'), 416),
(('snakke', 'høy'), 409), (('rar', 'snakke'), 408), (('gråte', 'snakke'), 408), (('starte', 'snakke'), 408),
(('snakke', 'hjemme'), 406), (('hate', 'snakke'), 406), (('snakke', 'melding'), 404), (('hos', 'snakke'), 403),
(('hjem', 'snakke'), 401), (('sånt', 'snakke'), 400), (('time', 'snakke'), 400), (('snakke', 'starte'), 400),
(('bore', 'snakke'), 399), (('sikker', 'snakke'), 399), (('snakke', 'tørr'), 397), (('snakke', 'ingenting'), 394),
(('snakke', 'sosial'), 394), (('snakke', 'rar'), 394), (('snakke', 'hvem'), 390), (('snakke', 'sitte'), 390),
(('snakke', 'sikker'), 388), (('måned', 'snakke'), 386), (('dra', 'snakke'), 385), (('osv', 'snakke'), 384),
(('snakke', 'dårlig'), 383), (('snakke', 'nett'), 382), (('snakke', 'mye'), 380), (('ringe', 'snakke'), 380),
(('bestevenn', 'snakke'), 376), (('samtale', 'snakke'), 376), (('snakke', 'møtes'), 375), (('helst', 'snakke'), 375),
(('snakke', 'skrive'), 374), (('snar', 'snakke'), 373), (('møtes', 'snakke'), 371), (('forstå', 'snakke'), 371),
(('snakke', 'mi'), 369), (('gjerne', 'snakke'), 367), (('flau', 'snakke'), 366), (('ingenting', 'snakke'), 363),
(('snakke', 'dra'), 360), (('snakke', 'liv'), 355), (('snakke', 'god'), 355), (('grei', 'snakke'), 354), (('tør',
'snakke'), 354), (('snakke', 'vise'), 351), (('snakke', 'tørre'), 350), (('ta', 'snakke'), 350), (('snakke', 'bry'),
349), (('snakke', 'mulig'), 348), (('bruke', 'snakke'), 348), (('derfor', 'snakke'), 347), (('snakke', 'telefon'),
343), (('flytte', 'snakke'), 342), (('små', 'snakke'), 340), (('snakke', 'enda'), 340), (('greie', 'snakke'), 340),
(('snakke', 'gråte'), 339), (('stole', 'snakke'), 339), (('snakke', 'plutselig'), 338), (('lov', 'snakke'), 338),
(('bry', 'snakke'), 338), (('usikker', 'snakke'), 337), (('slå', 'snakke'), 335), (('snakke', 'fin'), 334),
(('snakke', 'bruke'), 334), (('snakke', 'forhold'), 331), (('snakke', 'snar'), 330), (('snakke', 'gå'), 330),
(('klare', 'snakke'), 329), (('enda', 'snakke'), 329), (('plutselig', 'snakke'), 329), (('snakke', 'bestevenn'),
329), (('gå', 'snakke'), 329), (('orke', 'snakke'), 328), (('sosial', 'snakke'), 326), (('snakke', 'helst'), 326),
(('sjenert', 'snakke'), 324), (('skrive', 'snakke'), 323), (('hu', 'snakke'), 323), (('snakke', 'fore'), 323),
(('begge', 'snakke'), 322), (('melding', 'snakke'), 322), (('pleie', 'snakke'), 322), (('snakke', 'voksen'), 322),
(('snakke', 'ikkje'), 321), (('elske', 'snakke'), 319), (('snakke', 'usikker'), 318), (('snakke', 'lite'), 317),
(('fin', 'snakke'), 312), (('snakke', 'derfor'), 312), (('ei', 'snakke'), 311), (('merke', 'snakke'), 311),
(('snakke', 'engelsk'), 311), (('angst', 'snakke'), 311), (('snill', 'snakke'), 310), (('lære', 'snakke'), 310),
(('lege', 'snakke'), 310), (('snakke', 'dritt'), 309), (('snakke', 'grei'), 306), (('vise', 'snakke'), 305),
(('snakke', 'måned'), 303), (('snakke', 'hos'), 302), (('feil', 'snakke'), 296), (('uansett', 'snakke'), 296),
(('snakke', 'barnevern'), 294), (('snakke', 'lov'), 291), (('snakke', 'uansett'), 291), (('fore', 'snakke'), 286),
(('akkurat', 'snakke'), 286), (('snakke', 'bort'), 286), (('snakke', 'legge'), 285), (('mulig', 'snakke'), 284),
(('masse', 'snakke'), 283), (('forelsket', 'snakke'), 283), (('snakke', 'gjerne'), 283), (('snakke', 'ordentlig'),
282), (('snakke', 'normal'), 282), (('snakke', 'dum'), 281), (('snakke', 'msn'), 281), (('ho', 'snakke'), 279),
(('snakke', 'akkurat'), 276), (('snakke', 'skikkelig'), 276), (('snakke', 'ringe'), 276), (('snakke', 'flau'), 275),
(('dum', 'snakke'), 273), (('faktisk', 'snakke'), 272), (('snakke', 'slå'), 269), (('skikkelig', 'snakke'), 269),
(('snakke', 'kjempe'), 269), (('snakke', 'hjem'), 268), (('snakke', 'fare'), 266), (('snakke', 'elske'), 266),
(('snakke', 'hate'), 266), (('snakke', 'engang'), 266), (('snakke', 'snapchat'), 265), (('snakke', 'vanlig'), 265),
(('snakke', 'deg'), 264), (('snakke', 'snill'), 263), (('snakke', 'spesiell'), 261), (('snakke', 'hilsen'), 260),
(('snakke', 'jobbe'), 259), (('unna', 'snakke'), 258), (('snakke', 'slutt'), 258), (('nett', 'snakke'), 257),
(('slutt', 'snakke'), 256), (('snakke', 'bør'), 255), (('snakke', 'merke'), 254), (('synes', 'snakke'), 253),
(('stille', 'snakke'), 253), (('snakke', 'faktisk'), 251), (('snakke', 'synes'), 250), (('stå', 'snakke'), 248),
(('nekte', 'snakke'), 246), (('normal', 'snakke'), 246), (('facebook', 'snakke'), 245), (('snakke', 'pleie'), 245),
(('snakke', 'synse'), 244), (('kjempe', 'snakke'), 244), (('snakke', 'gi'), 242), (('jobb', 'snakke'), 242),
(('legge', 'snakke'), 241), (('deprimert', 'snakke'), 241), (('snakke', 'stille'), 240), (('snakke', 'råd'), 239),
(('snakke', 'foran'), 239), (('klein', 'snakke'), 239), (('synse', 'snakke'), 238), (('jobbe', 'snakke'), 238),
(('snakke', 'snap'), 237), (('snakke', 'smile'), 237), (('syk', 'snakke'), 236), (('mye', 'snakke'), 235),
(('irritere', 'snakke'), 235), (('drive', 'snakke'), 234), (('snakke', 'd'), 234), (('ca', 'snakke'), 233),
(('snakke', 'klein'), 231), (('lite', 'snakke'), 230), (('gale', 'snakke'), 229), (('leite', 'snakke'), 229),
(('råd', 'snakke'), 229), (('snakke', 'greie'), 229), (('spesiell', 'snakke'), 228), (('snakke', 'feil'), 227),
(('snakke', 'ende'), 226), (('snakke', 'syk'), 226), (('nei', 'snakke'), 226), (('smile', 'snakke'), 226),
(('snakke', 'tro'), 224), (('tro', 'snakke'), 223), (('fortsette', 'snakke'), 223), (('ignorere', 'snakke'), 223),
(('snakke', 'ca'), 221), (('liten', 'snakke'), 220), (('snakke', 'kompis'), 219), (('fare', 'snakke'), 219),
(('depresjon', 'snakke'), 219), (('ane', 'snakke'), 218), (('barnevern', 'snakke'), 217), (('snakke', 'savne'), 217),
(('snakke', 'orke'), 217), (('interessere', 'snakke'), 215), (('snakke', 'liten'), 215), (('snakke', 'irritere'),
215), (('deg', 'snakke'), 214), (('måte', 'snakke'), 214), (('snakke', 'føles'), 213), (('snakke', 'face'), 213),
(('gange', 'snakke'), 212), (('kompis', 'snakke'), 212), (('snakke', 'ignorere'), 212), (('snakke', 'klare'), 211),
(('flørte', 'snakke'), 210), (('snakke', 'snakk'), 210), (('krangle', 'snakke'), 210), (('tanke', 'snakke'), 209),
(('savne', 'snakke'), 209), (('flink', 'snakke'), 207), (('sjalu', 'snakke'), 207), (('snakke', 'forelsket'), 206),
(('slett', 'snakke'), 205), (('rom', 'snakke'), 204), (('snakke', 'bytte'), 204), (('sint', 'snakke'), 203),
(('snakke', 'flørte'), 201), (('fest', 'snakke'), 201), (('snakke', 'bror'), 200), (('spise', 'snakke'), 199),
(('snakke', 'sånne'), 199), (('snakke', 'forskjellig'), 198), (('snakke', 'søster'), 198), (('kun', 'snakke'), 197),
(('snakke', 'jobb'), 197), (('snakke', 'gire'), 197), (('snakke', 'kun'), 197), (('snakke', 'bilde'), 196),
(('snakke', 'par'), 196), (('snakke', 'ansikt'), 196), (('snakke', 'krangle'), 196), (('snakke', 'tema'), 195),
(('snakke', 'gale'), 194), (('snakke', 'kalle'), 194), (('snakke', 'lengre'), 193), (('snakke', 'sint'), 193),
(('engang', 'snakke'), 193), (('snakke', 'mere'), 192), (('snakke', 'kjent'), 192), (('snakke', 'særlig'), 191),
(('sjenere', 'snakke'), 190), (('snakke', 'vær'), 189), (('snakke', 'nei'), 189), (('høy', 'snakke'), 189),
(('snakke', 'sann'), 189), (('snakke', 'ane'), 189), (('snakke', 'ord'), 188), (('føles', 'snakke'), 188),
(('snakke', 'miste'), 188), (('vill', 'snakke'), 187), (('vel', 'snakke'), 186), (('kjent', 'snakke'), 186),
(('selvtillit', 'snakke'), 186), (('snakke', 'leite'), 185), (('snakke', 'angst'), 185), (('ute', 'snakke'), 184),
(('bo', 'snakke'), 184), (('håpe', 'snakke'), 184), (('snakke', 'medium'), 183), (('glemme', 'snakke'), 182),
(('prate', 'snakke'), 182), (('altså', 'snakke'), 181), (('snakke', 'spørsmål'), 181), (('snakk', 'snakke'), 181),
(('lengre', 'snakke'), 180), (('stoppe', 'snakke'), 179), (('stygg', 'snakke'), 179), (('miste', 'snakke'), 179),

	(('snakke', 'prate'), 179), (('snakke', 'sjenert'), 178), (('snakke', 'be'), 177), (('bror', 'snakke'), 177), (('gidde', 'snakke'), 177), (('snakke', 'hyggelig'), 175), (('mobbe', 'snakke'), 175), (('snakke', 'fac'), 175), (('par', 'snakke'), 175), (('snakke', 'vennine'), 175), (('snakke', 'glemme'), 174), (('kalle', 'snakke'), 174), (('snakke', 'e'), 174), (('snakke', 'nekte'), 174), (('bekymre', 'snakke'), 173), (('stor', 'snakke'), 173), (('snakke', 'bo'), 173), (('snakke', 'ene'), 172), (('snakke', 'fest'), 172), (('kveld', 'snakke'), 172), (('situasjon', 'snakke'), 171), (('snakke', 'vente'), 171), (('snakke', 'altså'), 171), (('snakke', 'skype'), 171), (('plage', 'snakke'), 171), (('d', 'snakke'), 170), (('såre', 'snakke'), 170), (('henge', 'snakke'), 170), (('vente', 'snakke'), 169), (('ende', 'snakke'), 168), (('grunne', 'snakke'), 168), (('alder', 'snakke'), 168), (('halv', 'snakke'), 168), (('snakke', 'henge'), 167), (('kontakte', 'snakke'), 166), (('type', 'snakke'), 165), (('snakke', 'plage'), 164), (('trist', 'snakke'), 164), (('nervøs', 'snakke'), 163), (('vanlig', 'snakke'), 163), (('snakke', 'gange'), 161), (('mulighet', 'snakke'), 161), (('snakke', 'm'), 160), (('snakke', 'spise'), 160), (('snakke', 'forklare'), 160), (('sette', 'snakke'), 160), (('vær', 'snakke'), 160), (('snakke', 'enig'), 160), (('bilde', 'snakke'), 159), (('snakke', 'le'), 159), (('snakke', 'interessere'), 159), (('snakke', 'såre'), 159), (('snakke', 'tulle'), 159), (('tema', 'snakke'), 158), (('forskjellig', 'snakke'), 158), (('snakke', 'situasjon'), 158), (('bestemme', 'snakke'), 157), (('vekk', 'snakke'), 156), (('hode', 'snakke'), 156), (('snakke', 'tanke'), 156), (('snakke', 'bestemme'), 156), (('ene', 'snakke'), 156), (('snakke', 'barn'), 155), (('snakke', 'søke'), 155), (('snakke', 'trist'), 155), (('snakke', 'kysse'), 154), (('ord', 'snakke'), 153), (('sove', 'snakke'), 153), (('snakke', 'grunne'), 152), (('snakke', 'fortsette'), 152), (('pga', 'snakke'), 151), (('e', 'snakke'), 151), (('ødelegge', 'snakke'), 150), (('sommer', 'snakke'), 149), (('særlig', 'snakke'), 149), (('ønske', 'snakke'), 149), (('snakke', 'Facebook'), 149), (('snakke', 'mobbe'), 148), (('snakke', 'vel'), 148), (('foran', 'snakke'), 148), (('søster', 'snakke'), 147), (('telefon', 'snakke'), 147), (('snakke', 'deprimert'), 147), (('snakke', 'sette'), 147), (('be', 'snakke'), 146), (('snakke', 'bety'), 146), (('frem', 'snakke'), 146), (('snakke', 'lang'), 146), (('snakke', 'bestevenninne'), 145), (('hyggelig', 'snakke'), 145), (('snakke', 'drive'), 145), (('egt', 'snakke'), 144), (('finnes', 'snakke'), 144), (('snakke', 'bekymre'), 144), (('snakke', 'fastlege'), 144), (('spørsmål', 'snakke'), 143), (('barn', 'snakke'), 143), (('bestevenninne', 'snakke'), 143), (('snakke', 'stå'), 142), (('snakke', 'Snapchat'), 142), (('drikke', 'snakke'), 142), (('gruppe', 'snakke'), 142), (('snakke', 'åpen'), 142), (('inne', 'snakke'), 142), (('kutte', 'snakke'), 142), (('nær', 'snakke'), 141), (('snakke', 'kveld'), 141), (('øye', 'snakke'), 141), (('snakke', 'seriøs'), 141), (('snakke', 'via'), 140), (('bety', 'snakke'), 140), (('gi', 'snakke'), 139), (('snakke', 'ødelegge'), 139), (('me', 'snakke'), 138), (('komme', 'snakke'), 138), (('år', 'snakke'), 138), (('snakke', 'halv'), 138), (('dø', 'snakke'), 137), (('snakke', 'virkelighet'), 137), (('følge', 'snakke'), 136), (('snakke', 'type'), 136), (('inni', 'snakke'), 136), (('snakke', 'dame'), 136), (('snakke', 'kontaktlærer'), 136), (('klar', 'snakke'), 136), (('snakke', 'stor'), 136), (('psykisk', 'snakke'), 136), (('snakke', 'hvilken'), 135), (('snakke', 'rektor'), 135), (('søke', 'snakke'), 135), (('snakke', 'nevne'), 135), (('dame', 'snakke'), 135), (('snakke', 'etterpå'), 134), (('selvmord', 'snakke'), 134), (('le', 'snakke'), 133), (('leve', 'snakke'), 133), (('snakke', 'ute'), 133), (('snakke', 'depresjon'), 133), (('fag', 'snakke'), 132), (('snakke', 'stoppe'), 132), (('snakke', 'flytende'), 132), (('periode', 'snakke'), 131), (('seriøs', 'snakke'), 131), (('snakke', 'friminutt'), 131), (('engelsk', 'snakke'), 131), (('unngå', 'snakke'), 130), (('mat', 'snakke'), 130), (('kontakt', 'snakke'), 130), (('karakter', 'snakke'), 130), (('snakke', 'vei'), 129), (('snakke', 'sjalu'), 129), (('lang', 'snakke'), 129), (('tanker', 'snakke'), 129), (('friminutt', 'snakke'), 129), (('snakke', 'øye'), 129), (('sann', 'snakke'), 129), (('snakke', 'rom'), 129), (('vennine', 'snakke'), 129), (('snakke', 'oppleve'), 128), (('snakke', 'mene'), 128), (('vei', 'snakke'), 128), (('snakke', 'sjenere'), 128), (('snakke', 'kontakte'), 127), (('snakke', 'åpne'), 127), (('snakke', 'ilag'), 127), (('snakke', 'funke'), 127), (('snakke', 'slikt'), 127), (('velge', 'snakke'), 127), (('kropp', 'snakke'), 127), (('bytte', 'snakke'), 127), (('snakke', 'komme'), 126), (('snakke', 'ein'), 126), (('frekk', 'snakke'), 125), (('snakke', 'vill'), 125), (('mage', 'snakke'), 125), (('snakke', 'finnes'), 124), (('fyr', 'snakke'), 124), (('mens', 'snakke'), 124), (('etterpå', 'snakke'), 124), (('snakke', 'dei'), 124), (('snakke', 'fritid'), 124), (('gire', 'snakke'), 124), (('søt', 'snakke'), 124), (('ordentlig', 'snakke'), 123), (('voksen', 'snakke'), 123), (('snakke', 'huske'), 123), (('snakke', 'mulighet'), 123), (('snakke', 'BUP'), 123), (('snakke', 'klar'), 123), (('ifra', 'snakke'), 122), (('ungdomsskole', 'snakke'), 122), (('snakke', 'etterhvert'), 122), (('snakke', 'mens'), 122), (('snakke', 'måte'), 122), (('snakke', 'ting'), 122), (('hvert', 'snakke'), 122), (('snakke', 'velge'), 122), (('nærhet', 'snakke'), 122), (('huske', 'snakke'), 121), (('snap', 'snakke'), 121), (('gøy', 'snakke'), 121), (('snakke', 'mobil'), 121), (('snakke', 'spille'), 121), (('sist', 'snakke'), 120), (('msn', 'snakke'), 120), (('kysse', 'snakke'), 120), (('snakke', 'sjelden'), 119), (('snakke', 'drikke'), 119), (('mere', 'snakke'), 119), (('snakke', 'pge'), 119), (('støtte', 'snakke'), 119), (('snakke', 'nær'), 119), (('snakke', 'hvert'), 119), (('snakke', 'støtte'), 119), (('oppføre', 'snakke'), 118), (('snakke', 'slett'), 118), (('snakke', 'Hilsen'), 118), (('snakke', 'gidde'), 118), (('vurdere', 'snakke'), 118), (('snakke', 'frem'), 118), (('snakke', 'natt'), 117), (('snakke', 'følge'), 117), (('snakke', 'vekk'), 117), (('snakke', 'inni'), 117), (('snakke', 'pga'), 117), (('interesse', 'snakke'), 117), (('forklare', 'snakke'), 116), (('snakke', 'misse'), 116), (('snakke', 'morsom'), 116), (('åpne', 'snakke'), 116), (('snakke', 'kutte'), 116), (('fritid', 'snakke'), 116), (('snakke', 'personlig'), 116), (('misse', 'snakke'), 116), (('snakke', 'negativ'), 115), (('snakke', 'nettopp'), 115), (('snakke', 'far'), 115), (('snakke', 'unna'), 115), (('snakke', 'sommer'), 114), (('snakke', 'tusen'), 114), (('far', 'snakke'), 114), (('sånne', 'snakke'), 114), (('full', 'snakke'), 114), (('pge', 'snakke'), 114), (('dritt', 'snakke'), 113), (('snakke', 'rygg'), 113), (('penger', 'snakke'), 113), (('snakke', 'reise'), 113), (('snakke', 'gruppe'), 113), (('snakke', 'felles'), 113), (('snakke', 'gøy'), 112), (('besøk', 'snakke'), 112), (('treffe', 'snakke'), 112), (('mnd', 'snakke'), 112), (('absolutt', 'snakke'), 112), (('vis', 'snakke'), 111), (('snakke', 'vis'), 111), (('snakke', 'enten'), 111), (('snakke', 'likevel'), 111), (('snakke', 'ellers'), 111), (('snakke', 'kropp'), 110), (('snakke', 'kjøpe'), 110), (('oppleve', 'snakke'), 110), (('vid', 'snakke'), 109), (('stresse', 'snakke'), 109), (('natt', 'snakke'), 109), (('snakke', 'egt'), 109), (('takle', 'snakke'), 109), (('ekkel', 'snakke'), 109), (('ensom', 'snakke'), 109), (('snakke', 'nervøs'), 109), (('helg', 'snakke'), 109), (('snakke', 'fremmed'), 109), (('snakke', 'psykisk'), 109), (('snakke', 'mening'), 108), (('æ', 'snakke'), 108), (('trygg', 'snakke'), 108), (('likevel', 'snakke'), 108), (('angre', 'snakke'), 107), (('snakke', 'ettersom'), 107), (('sterk', 'snakke'), 107), (('snakke', 'språk'), 107), (('snakke', 'hode'), 107), (('sliten', 'snakke'), 107), (('populær', 'snakke'), 106), (('snakke', 'oppføre'), 106), (('åpen', 'snakke'), 106), (('snakke', 'lage'), 106), (('att', 'snakke'), 106), (('snakke', 'selvom'), 106), (('snakke', 'interesse'), 106), (('selvom', 'snakke'), 106), (('snakke', 'lee'), 105), (('snakke', 'inne'), 105), (('snakke', 'unngå'), 105), (('snakke', 'idag'), 104), (('snakke', 'vurdere'), 104), (('borte', 'snakke'), 104), (('idag', 'snakke'), 104), (('snakke', 'vert'), 104), (('side', 'snakke'), 104), (('snapchat', 'snakke'), 103), (('snakke', 'sove'), 103), (('snakke', 'lag'), 103), (('snakke', 'selvmord'), 103), (('snakke', 'elev'), 102), (('selvmordstanke', 'snakke'), 102), (('ekstrem', 'snakke'), 102), (('rådgiver', 'snakke'), 102), (('pen', 'snakke'), 102), (('snakke', 'fyr'), 101), (('snakke', 'alder'), 101), (('snakke', 'trå'), 101), (('ubehagelig', 'snakke'), 101), (('snakke', 'slik'), 101), (('hvilken', 'snakke'), 101), (('nevne', 'snakke'), 101), (('teit', 'snakke'), 101), (('klasserom', 'snakke'), 101), (('mobil', 'snakke'), 101), (('snakke', 'ifra'), 100), (('snakke', 'ønske'), 100), (('no', 'snakke'), 100), (('snakke', 'penger'), 100), (('forandre', 'snakke'), 100), (('gjelde', 'snakke'), 100), (('snakke', 'søt'), 100))

svare	(('dere', 'svare'), 3851), (('håpe', 'svare'), 2287), (('spørre', 'svare'), 1919), (('vite', 'svare'), 934), (('svare', 'vite'), 839), (('svare', 'jente'), 818), (('svare', 'spørsmål'), 729), (('svare', 'spørre'), 705), (('svare', 'dere'), 664), (('svare', 'fore'), 662), (('svare', 'nei'), 649), (('svare', 'lure'), 607), (('svare', 'svare'), 602), (('hvis', 'svare'), 598), (('snakke', 'svare'), 579), (('like', 'svare'), 561), (('svare', 'gutt'), 500), (('lure', 'svare'), 486), (('min', 'svare'), 479), (('sende', 'svare'), 471), (('svare', 'mulig'), 453), (('svare', 'melding'), 449), (('svare', 'min'), 447), (('svare', 'snakke'), 430), (('svare', 'fordi'), 427), (('svare', 'like'), 403), (('svare', 'mine'), 402), (('prøve', 'svare'), 390), (('melding', 'svare'), 377), (('svare', 'sende'), 371), (('svare', 'prøve'), 344), (('svare', 'hvis'), 343), (('svare', 'føle'), 341), (('svare', 'aldri'), 340), (('svare', 'si'), 332), (('svare', 'begynne'), 321), (('svare', 'tenke'), 316), (('svare', 'venn'), 304), (('svare', 'mitt'), 297), (('jente', 'svare'), 297), (('venn', 'svare'), 291), (('si', 'svare'), 289), (('svare', 'jo'), 285), (('skole', 'svare'), 275), (('svare', 'skole'), 271), (('fordi', 'svare'), 261), (('svare', 'alltid'), 255), (('finne', 'svare'), 250), (('gutt', 'svare'), 249), (('svare', 'trenge'), 247), (('svare', 'sånn'), 245), (('svare', 'hilsen'), 242), (('mine', 'svare'), 238), (('skrive', 'svare'), 238), (('deg', 'svare'), 234), (('sånn', 'svare'), 227), (('dag', 'svare'), 223), (('svare', 'dag'), 218), (('vanskelig', 'svare'), 215), (('svare', 'lyst'), 207), (('ta', 'svare'), 206), (('svare', 'redd'), 205), (('svare', 'snar'), 205), (('svare', 'finne'), 202), (('kanskje', 'svare'), 201), (('svare', 'skrive'), 201), (('føle', 'svare'), 200), (('svare', 'kanskje'), 191), (('aldri', 'svare'), 190), (('tenke', 'svare'), 189), (('jo', 'svare'), 188), (('tusen', 'svare'), 184), (('mitt', 'svare'), 182), (('redd', 'svare'), 182), (('svare', 'ei'), 178), (('svare', 'ha'), 176), (('skje', 'svare'), 175), (('svare', 'problem'), 173), (('begynne', 'svare'), 172), (('spørsmål', 'svare'), 172), (('fin', 'svare'), 172), (('snill', 'svare'), 172), (('svare', 'hjelp'), 166), (('slutte', 'svare'), 164), (('svare', 'kjæreste'), 162), (('sikker', 'svare'), 162), (('tid', 'svare'), 161), (('svare', 'bra'), 160), (('virkelig', 'svare'), 160), (('svare', 'håpe'), 158), (('ringe', 'svare'), 157), (('svare', 'egentlig'), 156), (('oft', 'svare'), 156), (('klarere', 'svare'), 156), (('svare', 'gammel'), 154), (('svare', 'tusen'), 154), (('svare', 'heller'), 152), (('hver', 'svare'), 151), (('stille', 'svare'), 151), (('hjelp', 'svare'), 149), (('svare', 'klasse'), 147), (('sex', 'svare'), 145), (('svare', 'ganske'), 145), (('svare', 'deg'), 145), (('bruke', 'svare'), 144), (('svare', 'nesten'), 143), (('alltid', 'svare'), 141), (('bra', 'svare'), 140), (('trenge', 'svare'), 139), (('svare', 'skje'), 139), (('svare', 'sikker'), 139), (('mamma', 'svare'), 135), (('svare', 'oft'), 135), (('klasse', 'svare'), 135), (('svare', 'ringe'), 132), (('svare', 'mamma'), 131), (('svare', 'lei'), 130), (('svare', 'burde'), 130), (('svare', 'sex'), 129), (('problem', 'svare'), 127), (('skjønne', 'svare'), 126), (('svare', 'sur'), 125), (('svare', 'bruke'), 125), (('sur', 'svare'), 124), (('svare', 'leng'), 121), (('nei', 'svare'), 120), (('svare', 'utrolig'), 119), (('svare', 'rett'), 118), (('egentlig', 'svare'), 118), (('svare', 'virke'), 116), (('svare', 'glad'), 115), (('svare', 'mvh'), 115), (('glad', 'svare'), 115), (('svare', 'slite'), 113), (('svare', 'feil'), 112), (('svare', 'frekt'), 112), (('osv', 'svare'), 111), (('svare', 'Hilsen'), 111), (('gidde', 'svare'), 110), (('lyst', 'svare'), 110), (('hvem', 'svare'), 110), (('svare', 'forelske'), 110), (('svare', 'virkelig'), 108), (('liksom', 'svare'), 108), (('pris', 'svare'), 107), (('svare', 'ta'), 107), (('svare', 'fortelle'), 107), (('svare', 'fin'), 107), (('ho', 'svare'), 107), (('svare', 'klarere'), 105), (('svare', 'hver'), 104), (('kjæreste', 'svare'), 103), (('svare', 'rask'), 103), (('svare', 'liksom'), 102), (('svare', 'kjenne'), 102), (('svare', 'stille'), 100), (('svare', 'usikker'), 100), (('svare', 'uke'), 100))
råd	(('dere', 'råd'), 1664), (('trenge', 'råd'), 1567), (('god', 'råd'), 821), (('råd', 'jente'), 782), (('vite', 'råd'), 629), (('håpe', 'råd'), 503), (('råd', 'min'), 478), (('råd', 'gutt'), 473), (('lure', 'råd'), 407), (('dårlig', 'råd'), 399), (('råd', 'lure'), 398), (('råd', 'vite'), 392), (('tips', 'råd'), 380), (('min', 'råd'), 378), (('råd', 'dere'), 357), (('råd', 'tips'), 285), (('virkelig', 'råd'), 272), (('mine', 'råd'), 265), (('råd', 'hilsen'), 250), (('spørre', 'råd'), 245), (('hjelp', 'råd'), 242), (('snakke', 'råd'), 239), (('råd', 'hjelp'), 234), (('råd', 'snakke'), 229), (('råd', 'føle'), 222), (('råd', 'betale'), 219), (('føle', 'råd'), 214), (('råd', 'skole'), 212), (('råd', 'trenge'), 211), (('råd', 'mine'), 209), (('mamma', 'råd'), 204), (('skole', 'råd'), 202), (('fordi', 'råd'), 201), (('råd', 'venn'), 201), (('råd', 'burde'), 199), (('råd', 'slite'), 193), (('råd', 'problem'), 187), (('råd', 'begynne'), 187), (('råd', 'si'), 186), (('tenke', 'råd'), 184), (('venn', 'råd'), 181), (('jente', 'råd'), 179), (('råd', 'tenke'), 179), (('råd', 'flytte'), 173), (('råd', 'mamma'), 172), (('si', 'råd'), 170), (('snill', 'råd'), 163), (('råd', 'like'), 163), (('råd', 'kjæreste'), 161), (('råd', 'ei'), 160), (('råd', 'fordi'), 158), (('heller', 'råd'), 158), (('prøve', 'råd'), 157), (('råd', 'lyst'), 154), (('hjelpe', 'råd'), 151), (('problem', 'råd'), 148), (('kanskje', 'råd'), 147), (('råd', 'gammel'), 145), (('råd', 'forelder'), 145), (('penger', 'råd'), 144), (('råd', 'kjøpe'), 143), (('råd', 'penger'), 142), (('råd', 'sånn'), 139), (('forelder', 'råd'), 138), (('råd', 'tusen'), 138), (('redd', 'råd'), 137), (('gjerne', 'råd'), 137), (('flytte', 'råd'), 135), (('derfor', 'råd'), 134), (('råd', 'hjelpe'), 134), (('mitt', 'råd'), 134), (('råd', 'råd'), 133), (('råd', 'mvh'), 133), (('lyst', 'råd'), 133), (('sånn', 'råd'), 132), (('råd', 'prøve'), 131), (('egentlig', 'råd'), 130), (('råd', 'mitt'), 130), (('familie', 'råd'), 128), (('råd', 'ha'), 125), (('like', 'råd'), 123), (('råd', 'hvis'), 123), (('hvis', 'råd'), 123), (('råd', 'heller'), 119), (('råd', 'snar'), 118), (('råd', 'bo'), 118), (('jo', 'råd'), 116), (('råd', 'mulig'), 114), (('råd', 'bør'), 113), (('råd', 'pappa'), 111), (('råd', 'bruke'), 110), (('råd', 'jobb'), 109), (('råd', 'kanskje'), 108), (('betale', 'råd'), 108), (('pappa', 'råd'), 107), (('råd', 'håpe'), 107), (('råd', 'finne'), 107), (('vanskelig', 'råd'), 106), (('råd', 'bore'), 106), (('råd', 'spørre'), 104), (('råd', 'jobbe'), 102), (('råd', 'virkelig'), 101), (('bra', 'råd'), 101), (('gire', 'råd'), 101), (('råd', 'Hilsen'), 100))
spørsmål	(('spørsmål', 'mitt'), 4284), (('vite', 'spørsmål'), 1083), (('dere', 'spørsmål'), 932), (('svare', 'spørsmål'), 729), (('spørsmål', 'dere'), 631), (('spørsmål', 'mine'), 609), (('håpe', 'spørsmål'), 590), (('stille', 'spørsmål'), 483), (('spørsmål', 'egentlig'), 468), (('spørsmål', 'lure'), 468), (('spørsmål', 'vite'), 427), (('spørsmål', 'min'), 385), (('spørsmål', 'jente'), 380), (('spørsmål', 'hvis'), 378), (('finne', 'spørsmål'), 335), (('skole', 'spørsmål'), 315), (('min', 'spørsmål'), 306), (('spørsmål', 'gutt'), 303), (('spørsmål', 'mulig'), 297), (('spørre', 'spørsmål'), 251), (('hvis', 'spørsmål'), 243), (('jente', 'spørsmål'), 239), (('spørsmål', 'skole'), 236), (('lure', 'spørsmål'), 235), (('spørsmål', 'lov'), 233), (('hvilken', 'spørsmål'), 226), (('spørsmål', 'burde'), 226), (('spørsmål', 'finne'), 225), (('spørsmål', 'hvilken'), 225), (('tenke', 'spørsmål'), 221), (('spørsmål', 'rett'), 217), (('sende', 'spørsmål'), 216), (('spørsmål', 'si'), 214), (('spørsmål', 'håpe'), 212), (('spørsmål', 'prøve'), 207), (('spørsmål', 'spørre'), 206), (('mitt', 'spørsmål'), 205), (('spørsmål', 'stille'), 198), (('si', 'spørsmål'), 196), (('mine', 'spørsmål'), 194), (('kategori', 'spørsmål'), 193), (('spørsmål', 'søke'), 192), (('føle', 'spørsmål'), 191), (('sette', 'spørsmål'), 188), (('egentlig', 'spørsmål'), 188), (('spørsmål', 'begynne'), 187), (('legge', 'spørsmål'), 187), (('sikker', 'spørsmål'), 187), (('fordi', 'spørsmål'), 185), (('spørsmål', 'tenke'), 183), (('prøve', 'spørsmål'), 182), (('snakke', 'spørsmål'), 181), (('spørsmål', 'like'), 173), (('spørsmål', 'svare'), 172), (('spørsmål', 'vel'), 164), (('spørsmål', 'venn'), 160), (('begynne', 'spørsmål'), 159), (('jo', 'spørsmål'), 156), (('spørsmål', 'føle'), 153), (('spørsmål', 'fordi'), 149), (('spørsmål', 'trenge'), 149), (('skjønne', 'spørsmål'), 148), (('spørsmål', 'sende'), 148), (('spørsmål', 'bruke'), 146), (('spørsmål', 'altså'), 145), (('gutt', 'spørsmål'), 144), (('spørsmål', 'snakke'), 143), (('like', 'spørsmål'), 142), (('spørsmål', 'skje'), 141), (('hjelp', 'spørsmål'), 140), (('spørsmål', 'ha'), 140), (('spørsmål', 'velge'), 138), (('spørsmål', 'kanskje'), 137), (('venn', 'spørsmål'), 135), (('trenge', 'spørsmål'), 135), (('kanskje', 'spørsmål'), 134), (('usikker', 'spørsmål'), 134), (('spørsmål', 'sånn'), 133), (('sånn', 'spørsmål'), 130), (('spørsmål', 'spørsmål'), 129), (('fag', 'spørsmål'), 129), (('skrive', 'spørsmål'), 127), (('spørsmål', 'problem'), 127), (('rett',

	'spørsmål'), 125), (('lyst', 'spørsmål'), 124), (('spørsmål', 'følge'), 124), (('spørsmål', 'finnes'), 124), (('forstå', 'spørsmål'), 123), (('aldri', 'spørsmål'), 122), (('spørsmål', 'lyst'), 122), (('bruke', 'spørsmål'), 122), (('spørsmål', 'jo'), 121), (('dag', 'spørsmål'), 121), (('problem', 'spørsmål'), 119), (('søke', 'spørsmål'), 118), (('spørsmål', 'sex'), 118), (('skje', 'spørsmål'), 116), (('velge', 'spørsmål'), 115), (('spørsmål', 'leng'), 115), (('spørsmål', 'kategori'), 114), (('ta', 'spørsmål'), 114), (('mulig', 'spørsmål'), 113), (('helst', 'spørsmål'), 113), (('derfor', 'spørsmål'), 113), (('spørsmål', 'legge'), 112), (('spørsmål', 'ta'), 112), (('snill', 'spørsmål'), 111), (('redd', 'spørsmål'), 111), (('sex', 'spørsmål'), 111), (('spørsmål', 'bør'), 110), (('spørsmål', 'skrive'), 110), (('heller', 'spørsmål'), 109), (('bra', 'spørsmål'), 109), (('spørsmål', 'normal'), 109), (('spørsmål', 'gammel'), 109), (('spørsmål', 'hjelp'), 108), (('ganske', 'spørsmål'), 106), (('gjerne', 'spørsmål'), 106), (('mens', 'spørsmål'), 105), (('hjelpe', 'spørsmål'), 104), (('karakter', 'spørsmål'), 104), (('svar', 'spørsmål'), 103), (('spørsmål', 'passere'), 103), (('spørsmål', 'ganske'), 103), (('spørsmål', 'handle'), 102), (('uke', 'spørsmål'), 102), (('spørsmål', 'mulighet'), 102), (('spørsmål', 'forelder'), 101), (('spørsmål', 'ei'), 100))
forklare	('forklare', 'eneste'), 20), (('stå', 'forklare'), 20), (('syk', 'forklare'), 20), (('forklare', 'forståelig'), 20), (('mor', 'forklare'), 20), (('forklare', 'd'), 20), (('forklare', 'rar'), 20), (('forklare', 'farlig'), 19), (('forklare', 'mene'), 19), (('glad', 'forklare'), 19), (('lov', 'forklare'), 19), (('e', 'forklare'), 19), (('forklare', 'bestevenn'), 19), (('sette', 'forklare'), 19), (('forklare', 'ikkje'), 19), (('inni', 'forklare'), 19), (('forklare', 'oppleve'), 19), (('forklare', 'små'), 19), (('forklare', 'ting'), 19), (('lese', 'forklare'), 19), (('mulighet', 'forklare'), 19), (('velge', 'forklare'), 19), (('forklare', 'smerte'), 19), (('karakter', 'forklare'), 19), (('lite', 'forklare'), 19), (('komplisere', 'forklare'), 19), (('forklare', 'slett'), 19), (('forklare', 'nett'), 19), (('forklare', 'oppgave'), 19), (('forklare', 'velge'), 19), (('spise', 'forklare'), 19), (('forklare', 'ca'), 19), (('synse', 'forklare'), 18), (('nettside', 'forklare'), 18), (('forklare', 'forhold'), 18), (('bore', 'forklare'), 18), (('ane', 'forklare'), 18), (('forklare', 'vanlig'), 18), (('vanlig', 'forklare'), 18), (('forklare', 'inni'), 18), (('forklare', 'time'), 18), (('forklare', 'gire'), 18), (('samtale', 'forklare'), 18), (('forklare', 'sette'), 18), (('forklare', 'eventuell'), 18), (('forklare', 'ene'), 18), (('forklare', 'hos'), 18), (('forklare', 'år'), 18), (('gire', 'forklare'), 18), (('flytte', 'forklare'), 17), (('forklare', 'bør'), 17), (('helsesøster', 'forklare'), 17), (('forklare', 'ane'), 17), (('forhold', 'forklare'), 17), (('slå', 'forklare'), 17), (('jobbe', 'forklare'), 17), (('vansklig', 'forklare'), 17), (('forklare', 'ende'), 17), (('forklare', 'gråte'), 17), (('stoppe', 'forklare'), 17), (('forklare', 'mens'), 17), (('helst', 'forklare'), 17), (('oppgave', 'forklare'), 17), (('deg', 'forklare'), 17), (('forklare', 'nærmere'), 17), (('forklare', 'utdanning'), 17), (('forklare', 'kalle'), 17), (('forklare', 'spise'), 17), (('forklare', 'måned'), 17), (('forklare', 'nekte'), 17), (('masse', 'forklare'), 17), (('mening', 'forklare'), 17), (('bør', 'forklare'), 17), (('mi', 'forklare'), 17), (('forklare', 'ferdig'), 17), (('slags', 'forklare'), 17), (('forklare', 'usikker'), 17), (('forklare', 'helst'), 16), (('forklare', 'enda'), 16), (('merke', 'forklare'), 16), (('forklare', 'kropp'), 16), (('psykisk', 'forklare'), 16), (('forklare', 'vel'), 16), (('finnes', 'forklare'), 16), (('plutselig', 'forklare'), 16), (('hete', 'forklare'), 16), (('fag', 'forklare'), 16), (('forklare', 'legge'), 16), (('forklare', 'hvem'), 16), (('forklare', 'kjøpe'), 16), (('forklare', 'bekymre'), 16), (('forklare', 'kun'), 16), (('forklare', 'lenger'), 16), (('forklare', 'huske'), 16), (('forklare', 'ein'), 16), (('forklare', 'side'), 16), (('enda', 'forklare'), 16), (('forklare', 'klare'), 16), (('år', 'forklare'), 16), (('forklare', 'derfor'), 16), (('leste', 'forklare'), 16), (('mene', 'forklare'), 16), (('føles', 'forklare'), 16), (('studere', 'forklare'), 16), (('forklare', 'stille'), 16), (('gidde', 'forklare'), 15), (('forklare', 'mvh'), 15), (('forklare', 'dersom'), 15), (('altså', 'forklare'), 15), (('ei', 'forklare'), 15), (('forklare', 'tanke'), 15), (('forklare', 'hjem'), 15), (('forklare', 'skade'), 15), (('forklare', 'foregå'), 15), (('snar', 'forklare'), 15), (('forklare', 'trekke'), 15), (('forklare', 'råd'), 15), (('forklare', 'fore'), 15), (('forklare', 'sliten'), 15), (('smerte', 'forklare'), 15), (('forklare', 'lett'), 15), (('forklare', 'fin'), 15), (('stille', 'forklare'), 15), (('forklare', 'mulighet'), 15), (('hilsen', 'forklare'), 15), (('anmelde', 'forklare'), 15), (('kropp', 'forklare'), 15), (('faktisk', 'forklare'), 15), (('forklare', 'trist'), 15), (('forklare', 'mye'), 15), (('fungere', 'forklare'), 15), (('forklare', 'karakter'), 15), (('forklare', 'dei'), 15), (('forklare', 'leste'), 15), (('pge', 'forklare'), 15), (('gjelde', 'forklare'), 15), (('forklare', 'bo'), 15), (('forklare', 'komme'), 15), (('forklare', 'fag'), 15), (('irritere', 'forklare'), 14), (('forklare', 'be'), 14), (('tegn', 'forklare'), 14), (('forklare', 'nyte'), 14), (('forklare', 'mening'), 14), (('hverandre', 'forklare'), 14), (('forklare', 'snill'), 14), (('måned', 'forklare'), 14), (('la', 'forklare'), 14), (('forklare', 'riktig'), 14), (('kontakte', 'forklare'), 14), (('forklare', 'dum'), 14), (('venninne', 'forklare'), 14), (('forklare', 'leve'), 14), (('forklare', 'helsesøster'), 14), (('forklare', 'sint'), 14), (('ute', 'forklare'), 14), (('forklare', 'plutselig'), 14), (('jobb', 'forklare'), 14), (('forklare', 'studere'), 14), (('forklare', 'ønske'), 14), (('nekte', 'forklare'), 14), (('fravær', 'forklare'), 14), (('forklare', 'reise'), 14), (('lett', 'forklare'), 14), (('forklare', 'synes'), 14), (('d', 'forklare'), 14), (('fare', 'forklare'), 14), (('hate', 'forklare'), 14), (('fast', 'forklare'), 14), (('forklare', 'kjempe'), 14), (('forklare', 'forklaring'), 14), (('sann', 'forklare'), 14), (('vær', 'forklare'), 14), (('enkelt', 'forklare'), 14), (('små', 'forklare'), 14), (('penger', 'forklare'), 14), (('råd', 'forklare'), 14), (('synes', 'forklare'), 14), (('normal', 'forklare'), 14), (('forklare', 'spesiell'), 14), (('lengre', 'forklare'), 14), (('åsse', 'forklare'), 14), (('høy', 'forklare'), 14), (('forklare', 'irritere'), 13), (('klar', 'forklare'), 13), (('forklare', 'ho'), 13), (('bry', 'forklare'), 13), (('barn', 'forklare'), 13), (('forklare', 'nei'), 13), (('grunne', 'forklare'), 13), (('forklare', 'bilde'), 13), (('æ', 'forklare'), 13), (('betale', 'forklare'), 13), (('forklare', 'start'), 13), (('forklare', 'slutt'), 13), (('kalle', 'forklare'), 13), (('forklare', 'gi'), 13), (('forklare', 'elske'), 13), (('eksempel', 'forklare'), 13), (('mye', 'forklare'), 13), (('forklare', 'ute'), 13), (('sosial', 'forklare'), 13), (('sterk', 'forklare'), 13), (('forklare', 'stor'), 13), (('type', 'forklare'), 13), (('forklare', 'lage'), 13), (('forklare', 'psykisk'), 13), (('forklare', 'att'), 13), (('stemme', 'forklare'), 13), (('forklare', 'ringe'), 13), (('forklare', 'bort'), 13), (('forklare', 'betale'), 13), (('forklare', 'deprimert'), 13), (('forklare', 'glemme'), 13), (('forklare', 'syk'), 13), (('forklare', 'gange'), 13), (('forklare', 'søster'), 13), (('forklare', 'funke'), 13), (('eventuell', 'forklare'), 13), (('forklare', 'synse'), 13), (('videregående', 'forklare'), 13), (('forklare', 'hate'), 13), (('butikk', 'forklare'), 13), (('forklare', 'flau'), 13), (('etterpå', 'forklare'), 13), (('enkel', 'forklare'), 13), (('forklare', 'rette'), 13), (('forklare', 'vekk'), 12), (('lage', 'forklare'), 12), (('gråte', 'forklare'), 12), (('hvem', 'forklare'), 12), (('tro', 'forklare'), 12), (('forklare', 'depresjon'), 12), (('forklare', 'drikke'), 12), (('forklare', 'påvirke'), 12), (('forklare', 'handle'), 12), (('forklare', 'ødelegge'), 12), (('riktig', 'forklare'), 12), (('muskel', 'forklare'), 12), (('forklare', 'detalj'), 12), (('navn', 'forklare'), 12), (('følge', 'forklare'), 12), (('miste', 'forklare'), 12), (('kun', 'forklare'), 12), (('forklare', 'lite'), 12), (('slett', 'forklare'), 12), (('svar', 'forklare'), 12), (('såre', 'forklare'), 12), (('forklare', 'sann'), 12), (('forklare', 'va'), 12), (('forklare', 'såre'), 12), (('alder', 'forklare'), 12), (('forklare', 'forelsket'), 12), (('forklare', 'mobbe'), 12), (('forklare', 'interessere'), 12), (('forklare', 'nøyaktig'), 12), (('stor', 'forklare'), 12), (('forklare', 'plage'), 12), (('farlig', 'forklare'), 12), (('sider', 'forklare'), 11), (('nyte', 'forklare'), 11), (('forklare', 'matte'), 11), (('forelske', 'forklare'), 11), (('forklare', 'brev'), 11), (('drikke', 'forklare'), 11), (('forklare', 'høy'), 11), (('forklare', 'klar'), 11), (('forklare', 'hete'), 11), (('skade', 'forklare'), 11), (('forklare', 'frem'), 11), (('alene', 'forklare'), 11), (('ende', 'forklare'), 11), (('forklare', 'gjelde'), 11), (('slutt', 'forklare'), 11), (('elev', 'forklare'), 11), (('side', 'forklare'), 11), (('forklare', 'pleie'), 11),

	(('forklare', 'drive'), 11), (('forklare', 'enig'), 11), (('ca', 'forklare'), 11), (('bo', 'forklare'), 11), (('an', 'forklare'), 11), (('forklare', 'beskjed'), 11), (('tanke', 'forklare'), 11), (('forklare', 'begrep'), 11), (('kjefte', 'forklare'), 11), (('gi', 'forklare'), 11), (('sånne', 'forklare'), 11), (('stole', 'forklare'), 11), (('forklare', 'kompis'), 11), (('forklare', 'par'), 11), (('forklare', 'type'), 11), (('forklare', 'historie'), 11), (('slippe', 'forklare'), 11), (('engang', 'forklare'), 11), (('forklare', 'ærlig'), 11), (('forklare', 'penis'), 11), (('forklare', 'bytte'), 11), (('forklare', 'stoppe'), 10), (('pga', 'forklare'), 10), (('komme', 'forklare'), 10), (('forklare', 'hu'), 10), (('forklare', 'politi'), 10), (('forklare', 'bestemme'), 10), (('støtte', 'forklare'), 10), (('forklare', 'miste'), 10), (('forklare', 'tegn'), 10), (('forklare', 'følge'), 10), (('forklare', 'alkohol'), 10), (('forklare', 'vis'), 10), (('elske', 'forklare'), 10), (('forklare', 'selvom'), 10), (('forklare', 'fortsette'), 10), (('forklare', 'åre'), 10), (('forklare', 'masse'), 10), (('lønn', 'forklare'), 10), (('linje', 'forklare'), 10), (('forklare', 'arbeidsgiver'), 10), (('forklare', 'fast'), 10), (('forklare', 'nøye'), 10), (('dei', 'forklare'), 10), (('bestemme', 'forklare'), 10), (('barnevern', 'forklare'), 10), (('forklare', 'alvorlig'), 10), (('par', 'forklare'), 10), (('begge', 'forklare'), 10), (('forklare', 'samtale'), 10), (('forklare', 'selvmord'), 10), (('veit', 'forklare'), 10), (('forklare', 'ill'), 10), (('huske', 'forklare'), 10), (('hendelse', 'forklare'), 10), (('forklare', 'stem'), 10), (('forklare', 'vurdere'), 10), (('forklare', 'mi'), 10), (('forklare', 'hendelse'), 10), (('forklare', 'pga'), 10), (('hus', 'forklare'), 10), (('vgs', 'forklare'), 10), (('forklare', 'sove'), 10), (('gå', 'forklare'), 10), (('forklare', 'vgs'), 10), (('glemme', 'forklare'), 10), (('forklare', 'finnes'), 10), (('utdanning', 'forklare'), 10), (('mail', 'forklare'), 10), (('sliten', 'forklare'), 10))

Reference

Ackard, D. M., & Neumark-Sztainer, D. (2001). Health care information sources for adolescents: age and gender differences on use, concerns, and needs. *Journal of adolescent health, 29*(3), 170-176.

Aguirre, A., Silva, I., Billings, J., Jimenez, M., & Rowe, S. (2020). What are the barriers, facilitators and interventions targeting help-seeking behaviours for common mental health problems in adolescents? A systematic review.

Aichner, T., & Jacob, F. (2015). Measuring the degree of corporate social media use. *International Journal of market research, 57*(2), 257-276.

Alghamdi, R., & Alfalqi, K. (2015). A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl.(IJACSA), 6*(1).

Ampofo, L., Collister, S., O'Loughlin, B., Chadwick, A., Halfpenny, P., & Procter, P. (2015). Text mining and social media: When quantitative meets qualitative and software meets people. *Innovations in digital research methods. Thousand Oaks, CA: SAGE Publications Inc*, 161-192.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine, 16*(7), 16-07.

Anderson, J. E., & Lowen, C. A. (2010). Connecting youth with health services: Systematic review. *Canadian Family Physician, 56*(8), 778-784.

Andrews, G., Issakidis, C., & Carter, G. (2001). Shortfall in mental health service utilisation. *The British Journal of Psychiatry, 179*(5), 417-425.

Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. *R&D Management, 50*(3), 329-351.

Arbreton, A. J. A. (1994). When getting help is helpful: Developmental, cognitive, and motivational influences on students' academic help-seeking.

Archer, D. (2009). *What's in a word-list?: investigating word frequency and keyword extraction*: Ashgate Publishing, Ltd.

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American psychologist, 55*(5), 469.

Babbie, E. R. (1989). *The practice of social research*: Wadsworth Publishing Company.

Bae, S. D., & Park, D.-H. (2018). The Effect of Mobile Advertising Platform through Big Data Analytics: Focusing on Advertising, and Media Characteristics. *Journal of Intelligence and Information Systems, 24*(2), 37-57.

Baker, P., Gabrielatos, C., Khosravinik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society, 19*(3), 273-306.

Bargh, J. A., McKenna, K. Y., & Fitzsimons, G. M. (2002). Can you see the real me? Activation and expression of the “true self” on the Internet. *Journal of social issues, 58*(1), 33-48.

Barker, G. (2007). Adolescents, social support and help-seeking behaviour: an international literature review and programme consultation with recommendations for action.

Barker, V. (2009). Older adolescents' motivations for social network site use: The influence of gender, group identity, and collective self-esteem. *Cyberpsychology & behavior, 12*(2), 209-213.

Bell, D. R., Chiang, J., & Padmanabhan, V. (1999). The decomposition of promotional response: An empirical generalization. *Marketing Science, 18*(4), 504-526.

Bell, J., Mok, K., Gardiner, E., & Pirkis, J. (2018). Suicide-related internet use among suicidal young people in the UK: Characteristics of users, effects of use, and barriers to offline help-seeking. *Archives of Suicide Research, 22*(2), 263-277.

Beller, E. K. (1955). Dependency and independence in young children. *The Journal of genetic psychology, 87*(1), 25-35.

Bennett, S., Maton, K., & Kervin, L. (2008). The 'digital natives' debate: A critical review of the evidence. *British journal of educational technology, 39*(5), 775-786.

Best, P., Gil-Rodriguez, E., Manktelow, R., & Taylor, B. J. (2016). Seeking help from everyone and no-one: Conceptualizing the online help-seeking process among adolescent males. *Qualitative health research, 26*(8), 1067-1077.

Best, P., Manktelow, R., & Taylor, B. J. (2014). Social work and social media: Online help-seeking and the mental well-being of adolescent males. *The British Journal of Social Work, 46*(1), 257-276.

Beyer, M. A., & Laney, D. (2012). The importance of 'big data': a definition. *Stamford, CT: Gartner*, 2014-2018.

Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications, 105*(9).

Biddle, L., Donovan, J. L., Gunnell, D., & Sharp, D. (2006). Young adults' perceptions of GPs as a help source for mental distress: a qualitative study. *British Journal of General Practice, 56*(533), 924-931.

Birnbaum, M. L., Candan, K., Libby, I., Pascucci, O., & Kane, J. (2016). Impact of online resources and social media on help-seeking behaviour in youth with psychotic symptoms. *Early intervention in psychiatry, 10*(5), 397-403.

Birnbaum, M. L., Rizvi, A. F., Correll, C. U., Kane, J. M., & Confino, J. (2017). Role of social media and the I nternet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. *Early intervention in psychiatry, 11*(4), 290-295.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84.

Boldero, J., & Fallon, B. (1995). Adolescent help-seeking: what do they get help for and from whom? *Journal of adolescence, 18*(2), 193-209.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature, 489*(7415), 295-298.

Borzekowski, D. L., & Rickert, V. I. (2001). Adolescent cybersurfing for health information: a new resource that crosses barriers. *Archives of pediatrics & adolescent medicine, 155*(7), 813-817.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society, 15*(5), 662-679.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated communication, 13*(1), 210-230.

Bradford, S., & Rickwood, D. (2014). Adolescent's preferred modes of delivery for mental health services. *Child and Adolescent Mental Health, 19*(1), 39-45.

Bradley, K. L., Robinson, L. M., & Brannen, C. L. (2012). Adolescent help-seeking for psychological distress, depression, and anxiety using an internet program. *International Journal of Mental Health Promotion, 14*(1), 23-34.

Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of Retailing, 93*(1), 79-95.

Brennen, J. S., & Kreiss, D. (2016). Digitalization. *The international encyclopedia of communication theory and philosophy*, 1-11.

Brindis, C. D., Hair, E. C., Cochran, S., Cleveland, K., Valderrama, L. T., & Park, M. J. (2007). Increasing access to program information: a strategy for improving adolescent health. *Maternal and child health journal, 11*(1), 27-35.

Brito, P. Q., Soares, C., Almeida, S., Monte, A., & Byvoet, M. (2015). Customer segmentation in a large database of an online customized fashion business. *Robotics and Computer-Integrated Manufacturing, 36*, 93-100.

Brookes, G., & Harvey, K. (2016). Examining the discourse of mental illness in a corpus of online advice-seeking messages. In *Talking at Work* (pp. 209-234): Springer.

Brunetti, G., Forsyth, P., Feltracco, A., Papaiz, D., Hodgson, K., Moore, P., . . . Sedgwick-Walsh, S. (2001). Partnering for Social Change: Public Health Positively Affecting Physician Practices. *Social Marketing Quarterly, 7*(3), 57-62.

Burns, J. M., Davenport, T. A., Durkin, L. A., Luscombe, G. M., & Hickie, I. B. (2010). The internet as a setting for mental health service utilisation by young people. *Medical journal of Australia, 192*, S22-S26.

Callahan, A., & Inckle, K. (2012). Cybertherapy or psychobabble? A mixed methods study of online emotional support. *British Journal of Guidance & Counselling, 40*(3), 261-278.

Camara, M., Bacigalupe, G., & Padilla, P. (2017). The role of social support in adolescents: are you helping me or stressing me out? *International Journal of Adolescence and Youth, 22*(2), 123-136.

Carr, G., & Bednarek, M. (2019). Beyond risk and safety? Identifying shifts in sex education advice targeted at young women. *Discourse & Society, 30*(3), 225-247.

Chambers, C. T., Reid, G. J., McGrath, P. J., & Finley, G. A. (1997). Self-administration of over-the-counter medication for pain among adolescents. *Archives of pediatrics & adolescent medicine, 151*(5), 449-455.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems, 22*, 288-296.

Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems, 63*, 67-80.

Chen, N.-C., Drouhard, M., Kocielnik, R., Suh, J., & Aragon, C. R. (2018). Using machine learning to support qualitative coding in social science: Shifting the focus to ambiguity. *ACM Transactions on Interactive Intelligent Systems (TiiS), 8*(2), 1-20.

Chen, Y., Zhang, J., Guo, M., & Cao, J. (2017). *Understanding customer behaviour in urban shopping mall from WiFi logs.* Paper presented at the 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops).

Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics, 16*(1), 22-29.

Cioffi-Revilla, C. (2017). Computation and Social Science. In *Introduction to Computational Social Science: Principles and Applications* (pp. 35-102). Cham: Springer International Publishing.

Collin, P. J., Metcalf, A. T., Stephens-Reicher, J. C., Blanchard, M. E., Herrman, H. E., Rahilly, K., & Burns, J. M. (2011). ReachOut. com: The role of an online service for promoting help-seeking in young people. *Advances in Mental Health, 10*(1), 39-51.

Costin, D. L., Mackinnon, A. J., Griffiths, K. M., Batterham, P. J., Bennett, A. J., Bennett, K., & Christensen, H. (2009). Health e-cards as a means of encouraging help seeking for depression among young adults: randomized controlled trial. *Journal of medical Internet research, 11*(4), e42.

Cukier, K. (2010). *Data, data everywhere: A special report on managing information*: Economist Newspaper.

Cunningham, C. E., Walker, J. R., Eastwood, J. D., Westra, H., Rimas, H., Chen, Y., . . . Group, M. M. R. (2014). Modeling mental health information preferences during the early adult years: a discrete choice conjoint experiment. *Journal of Health Communication, 19*(4), 413-440.

Davies, S. C., Lemer, C., Strelitz, J., & Weil, L. (2013). Our children deserve better: prevention pays. *Lancet (London, England), 382*(9902), 1383-1384. doi:10.1016/s0140-6736(13)62004-8

Davis-McCabe, C., & Winthrop, A. (2010). Computerised CBT: University students experiences of using an online self-help programme. *Counselling Psychology Review*.

Davis, K. (2012). Tensions of identity in a networked era: Young people's perspectives on the risks and rewards of online self-expression. *new media & society, 14*(4), 634-651.

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian psychology/Psychologie canadienne, 49*(3), 182.

Dhar, S., & Varshney, U. (2011). Challenges and business models for mobile location-based services and advertising. *Communications of the ACM, 54*(5), 121-128.

Dickinson, G. E. (1978). Adolescent sex information sources: 1964-1974.

Divin, N., Harper, P., Curran, E., Corry, D., & Leavey, G. (2018). Help-seeking measures and their use in adolescents: a systematic review. *Adolescent Research Review, 3*(1), 113-122.

Druin, A., Foss, E., Hatley, L., Golub, E., Guha, M. L., Fails, J., & Hutchinson, H. (2009). *How children search the internet with keyword interfaces.* Paper presented at the Proceedings of the 8th International conference on interaction design and children.

Duggan, M., & Brenner, J. (2013). *The demographics of social media users, 2012* (Vol. 14): Pew Research Center's Internet & American Life Project Washington, DC.

Dyche, J. (2012). Big data 'Eurekas!'don't just happen. *Harvard Business Review Blog, 20*.

Early, E. A. (1982). The logic of well being: Therapeutic narratives in Cairo, Egypt. *Social science & medicine, 16*(16), 1491-1497.

Eichenberg, C. (2008). Internet message boards for suicidal people: A typology of users. *Cyberpsychology & behavior, 11*(1), 107-113.

Elbow Method. (2020). Retrieved from https://www.scikit-yb.org/en/latest/api/cluster/elbow.html

Ellis, L. A., Collin, P., Hurley, P. J., Davenport, T. A., Burns, J. M., & Hickie, I. B. (2013). Young men's attitudes and behaviour in relation to mental health and technology: implications for the development of online mental health services. *BMC psychiatry, 13*(1), 119.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research, 69*(2), 897-904.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National science review, 1*(2), 293-314.

Fan, S., Lau, R. Y., & Zhao, J. L. (2015). Demystifying big data analytics for business intelligence through the lens of marketing mix. *Big Data Research, 2*(1), 28-32.

Feng, X. L., & Campbell, A. (2011). Understanding e-mental health resources: Personality, awareness, utilization, and effectiveness of e-mental health resources Amongst youth. *Journal of Technology in Human Services, 29*(2), 101-119.

Frand, J. L. (2000). The information-age mindset changes in students and implications for higher education. *Educause review, 35*, 14-25.

Frost, M., Casey, L., & Rando, N. (2016). Self-injury, help-seeking, and the Internet: Informing online service provision for young people. *Crisis: The Journal of Crisis Intervention and Suicide Prevention, 37*(1), 68.

Gaikwad, S. V., Chaugule, A., & Patil, P. (2014). Text mining methods and techniques. *International Journal of Computer Applications, 85*(17).

Gall, S. N.-L. (1985). Chapter 2: Help-Seeking Behavior in Learning. *Review of research in education, 12*(1), 55-90.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management, 35*(2), 137-144.

Ghani, N. A., Hamid, S., Hashem, I. A. T., & Ahmed, E. (2019). Social media big data analytics: A survey. *Computers in Human Behavior, 101*, 417-428.

González-Bailón, S. (2013). Social science in the era of big data. *Policy & internet, 5*(2), 147-160.

Gopalakrishnan, R., & Venkateswarlu, A. (2018). *Machine Learning for Mobile: Practical guide to building intelligent mobile applications powered by machine learning*: Packt Publishing.

Gould, A. W., & Mazzeo, J. (1982). Age and sex differences in early adolescent's information sources. *The Journal of Early Adolescence, 2*(3), 283-292.

Gould, M. S., Munfakh, J. L. H., Lubell, K., Kleinman, M., & Parker, S. (2002). Seeking help from the internet during adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*(10), 1182-1189.

Gourash, N. (1978). Help-seeking: A review of the literature. *American journal of community psychology, 6*(5), 413.

Gray, N. J., Harvey, K., Macfarlane, A., & McPherson, A. (2008). 6: Help! Adolescent Health Language in Email Messages. *Journal of adolescent health, 42*(2), 5-6.

Gray, N. J., Klein, J. D., Noyce, P. R., Sesselberg, T. S., & Cantrill, J. A. (2005). Health information-seeking behaviour in adolescence: the place of the internet. *Social science & medicine, 60*(7), 1467-1478.

Greco, F., & Polli, A. (2020). Emotional Text Mining: Customer profiling in brand management. *International journal of information management, 51*, 101934.

Greidanus, E., & Everall, R. D. (2010). Helper therapy in an online suicide prevention community. *British Journal of Guidance & Counselling, 38*(2), 191-204.

Grunig, J. E. (1989). Publics, audiences and market segments: Segmentation principles for campaigns. *Information campaigns: Balancing social values and social change*, 199-228.

Gulliver, A., Griffiths, K. M., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC psychiatry, 10*(1), 113.

Gulliver, A., Griffiths, K. M., Christensen, H., & Brewer, J. L. (2012). A systematic review of help-seeking interventions for depression, anxiety and general psychological distress. *BMC psychiatry, 12*(1), 81.

Gwizdka, J., & Bilal, D. (2017). *Analysis of children's queries and click behavior on ranked results and their thought processes in google search.* Paper presented at the Proceedings of the 2017 conference on conference human information interaction and retrieval.

Hampton, K. N., Goulet, L. S., Marlow, C., & Rainie, L. (2012). Why most Facebook users get more than they give. *Pew Internet & American Life Project, 3*(2012), 1-40.

Haner, D., & Pepler, D. (2016). "Live Chat" clients at kids help phone: Individual characteristics and problem topics. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 25*(3), 138.

Hansen, D. L., Derry, H. A., Resnick, P. J., & Richardson, C. R. (2003). Adolescents searching for health information on the Internet: an observational study. *Journal of medical Internet research, 5*(4), e25.

Harvey, K. (2012). Disclosures of depression: Using corpus linguistics methods to examine young people's online health concerns. *International Journal of Corpus Linguistics, 17*(3), 349-379.

Harvey, K., Churchill, D., Crawford, P., Brown, B., Mullany, L., Macfarlane, A., & McPherson, A. (2008). Health communication and adolescents: what do their emails tell us? *Family Practice, 25*(4), 304-311.

Harvey, K., Locher, M. A., & Mullany, L. (2013). " Can I Be at Risk of Getting AIDS?" A Linguistic Analysis of Two Internet Columns on Sexual Health. *Linguistik online, 59*, 111-132.

Havas, J., de Nooijer, J., Crutzen, R., & Feron, F. (2011). Adolescents' views about an internet platform for adolescents with mental health problems. *Health Education*.

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014). *Word cloud explorer: Text analytics based on word clouds.* Paper presented at the 2014 47th Hawaii International Conference on System Sciences.

Hindman, M. (2015). Building better models: Prediction, replication, and machine learning in the social sciences. *The ANNALS of the American Academy of Political and Social Science, 659*(1), 48-62.

Holloway, S. L., & Valentine, G. (2003). *Cyberkids: Children in the information age*: Psychology Press.
Hong, L., & Davison, B. D. (2010). *Empirical study of topic modeling in twitter.* Paper presented at the Proceedings of the first workshop on social media analytics.
Horgan, A., & Sweeney, J. (2010). Young students' use of the Internet for mental health information and support. *Journal of psychiatric and mental health nursing, 17*(2), 117-123.
Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems, 12*(12), 2.
IBM Corporation. (2012). Big Data: The New Natural Resource. Retrieved from https://www.ibmbigdatahub.com/infographic/big-data-new-natural-resource
Ignatow, G., & Mihalcea, R. (2016). *Text mining: A guidebook for the social sciences*: Sage Publications.
Jeong, B., Yoon, J., & Lee, J.-M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International journal of information management, 48*, 280-290.
Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *It Professional, 15*(6), 43-48.
Joyce, D., & Weibelzahl, S. (2011). Student counseling services: Using text messaging to lower barriers to help seeking. *Innovations in education and teaching international, 48*(3), 287-299.
Jurafsky, D., & Martin, J. H. (2018). Speech and language processing (draft). *Chapter A: Hidden Markov Models (Draft of September 11, 2018). Retrieved March, 19*, 2019.
Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons, 53*(1), 59-68.
Karabenick, S. A. (1987). Computer Conferencing: Its Impact on Academic Help-Seeking.
Karabenick, S. A. (1998). *Strategic help seeking: Implications for learning and teaching*: Routledge.
Kauer, S., Buhagiar, K., & Sanci, L. (2017). Facilitating mental health help seeking in young adults: the underlying theory and development of an online navigation tool. *Advances in Mental Health, 15*(1), 71-87.
Kauer, S. D., Mangan, C., & Sanci, L. (2014). Do online mental health services improve help-seeking for young people? A systematic review. *Journal of medical Internet research, 16*(3), e66.
Kelle, U. (1997). Theory building in qualitative research and computer programs for the management of textual data. *Sociological research online, 2*(2), 10-22.
Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., & Hooker, G. (2009). Data-intensive science: a new paradigm for biodiversity studies. *BioScience, 59*(7), 613-620.
Kerner, B., Carlson, M., Eskin, C. K., Tseng, C. H., Ho, J. M. G. Y., Zima, B., & Leader, E. (2020). Trends in the utilization of a peer-supported youth hotline. *Child and Adolescent Mental Health*.
Kim, E.-H., Coumar, A., Lober, W. B., & Kim, Y. (2011). Addressing mental health epidemic among university students via web-based, self-screening, and referral system: a preliminary study. *IEEE Transactions on Information Technology in Biomedicine, 15*(2), 301-307.
Kim, W., Jeong, O.-R., & Lee, S.-W. (2010). On social Web sites. *Information systems, 35*(2), 215-236.
King, R., Bickman, L., Shochet, I., McDermott, B., & Bor, B. (2010). Use of the internet for provision of better counselling and psychotherapy services to young people, their families and carers. *Psychotherapy in Australia, 17*(1), 66.
Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big data & society, 1*(1), 2053951714528481.
Kowalenko, N. M., & Culjak, G. (2018). Workforce planning for children and young people's mental health care. *The Lancet Public Health, 3*(6), e266-e267.
Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity, 47*(4), 2025-2047.

Kummervold, P. E., Gammon, D., Bergvik, S., Johnsen, J.-A. K., Hasvold, T., & Rosenvinge, J. H. (2002). Social support in a wired world: use of online mental health forums in Norway. *Nordic journal of psychiatry, 56*(1), 59-65.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note, 6*(70), 1.

Lassemo, E., Sand, K., & Tøndel, G. (2020). Kartlegging spørsmål fra lhbtiq-ungdom, ung. no. *SINTEF AS (ISBN starter med 978-82-14-)*.

Lau, R. Y., Li, C., & Liao, S. S. (2014). Social analytics: Learning fuzzy product ontologies for aspect-oriented sentiment analysis. *Decision Support Systems, 65*, 80-94.

Lawrence, N. W. (2003). Social Research methods: Quantitative and Qualitative Approaches. *Boston, New York, san Fransisco: Pearson Education, Inc*.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., . . . Gutmann, M. (2009). Life in the network: the coming age of computational social science. *Science (New York, NY), 323*(5915), 721.

Lee, F. (1997). When the going gets tough, do the tough ask for help? Help seeking and power motivation in organizations. *Organizational behavior and human decision processes, 72*(3), 336-363.

Lehdonvirta, V., & Räsänen, P. (2011). How do young people identify with online and offline peer groups? A comparison between UK, Spain and Japan. *Journal of Youth Studies, 14*(1), 91-108.

Liew, J. S. Y., McCracken, N., Zhou, S., & Crowston, K. (2014). *Optimizing features in active machine learning for complex qualitative content analysis.* Paper presented at the Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science.

Livingstone, S. (2009). Changing childhood, changing media.

Lu, X., Ba, S., Huang, L., & Feng, Y. (2013). Promotional marketing or word-of-mouth? Evidence from online restaurant reviews. *Information Systems Research, 24*(3), 596-612.

Luhn, H. P. (1957). A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of research and development, 1*(4), 309-317.

Madigan, S. (2011). *Narrative therapy*: American Psychological Association.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., . . . Häussler, T. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures, 12*(2-3), 93-118.

Maitz, E., Maitz, K., Sendlhofer, G., Wolfsberger, C., Mautner, S., Kamolz, L.-P., & Gasteiger-Klicpera, B. (2020). Internet-Based Health Information–Seeking Behavior of Students Aged 12 to 14 Years: Mixed Methods Study. *Journal of medical Internet research, 22*(5), e16281.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*: McKinsey Global Institute.

Mar, M. Y., Neilson, E. K., Torchalla, I., Werker, G. R., Laing, A., & Krausz, M. (2014). Exploring e-Mental health preferences of generation y. *Journal of Technology in Human Services, 32*(4), 312-327.

Mars, B., Heron, J., Biddle, L., Donovan, J. L., Holley, R., Piper, M., . . . Gunnell, D. (2015). Exposure to, and searching for, information about suicide and self-harm on the Internet: Prevalence and predictors in a population based cohort of young adults. *Journal of Affective Disorders, 185*, 239-245.

Menchen-Trevino, E. (2013). Collecting vertical trace data: Big possibilities and big challenges for multi-method research. *Policy & internet, 5*(3), 328-339.

Michaud, P.-A., & Fombonne, E. (2005). Common mental health problems. *Bmj, 330*(7495), 835-838.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*: sage.

Miller, H. J. (2010). The data avalanche is here. Shouldn't we be digging? *Journal of Regional Science, 50*(1), 181-201.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). *Optimizing semantic coherence in topic models.* Paper presented at the Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing.

Miner, G., Elder IV, J., Fast, A., Hill, T., Nisbet, R., & Delen, D. (2012). *Practical text mining and statistical analysis for non-structured text data applications*: Academic Press.

Mitchell, T. M. (1997). *Machine Learning*: McGraw-Hill.

Moessner, M., Feldhege, J., Wolf, M., & Bauer, S. (2018). Analyzing big data in social media: Text and network analyses of an eating disorder forum. *International Journal of Eating Disorders, 51*(7), 656-667.

Morinaga, S., Yamanishi, K., Tateishi, K., & Fukushima, T. (2002). *Mining product reputations on the web.* Paper presented at the Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining.

Murphy, L. B. E. (1962). The widening world of childhood: Paths toward maturity.

Neal, D. M., Campbell, A. J., Williams, L. Y., Liu, Y., & Nussbaumer, D. (2011). "I did not realize so many options are available": Cognitive authority, emerging adults, and e-mental health. *Library & Information Science Research, 33*(1), 25-33.

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). *Automatic evaluation of topic coherence.* Paper presented at the Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics.

Newman, R. S. (1991). Goals and self-regulated learning: What motivates children to seek academic help. *Advances in motivation and achievement, 7*, 151-183.

Newman, R. S. (2000). Social influences on the development of children's adaptive help seeking: The role of parents, teachers, and peers. *Developmental review, 20*(3), 350-404.

O'Dea, B., & Campbell, A. (2011). Healthy connections: Online social networks and their potential for peer support. *Studies in health technology and informatics, 168*, 133-140.

Oh, S., & Park, M. S. (2013). Text mining as a method of analyzing health questions in social Q&A. *Proceedings of the American Society for Information Science and Technology, 50*(1), 1-4.

Palfrey, J. G., & Gasser, U. (2011). *Born digital: Understanding the first generation of digital natives*: ReadHowYouWant. com.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *arXiv preprint cs/0409058*.

Park, E., & Kwon, M. (2018). Health-related internet use by children and adolescents: systematic review. *Journal of medical Internet research, 20*(4), e120.

Parser. (2020). Retrieved from https://nlp.stanford.edu/software/lex-parser.shtml#:~:text=A%20natural%20language%20parser%20is,or%20object%20of%20a%20verb.&text=Their%20development%20was%20one%20of,language%20processing%20in%20the%201990s.

Part Of Speech. (2020) (Vols. 2020).

Pauwels, K., Hanssens, D. M., & Siddarth, S. (2002). The long-term effects of price promotions on category incidence, brand choice, and purchase quantity. *Journal of marketing research, 39*(4), 421-439.

Pokorný, J. (2015). *How to Store and Process Big Data: Are Today's Databases Sufficient?* Paper presented at the IFIP International Conference on Computer Information Systems and Industrial Management.

Pollach, I. (2012). Taming textual data: The contribution of corpus linguistics to computer-aided text analysis. *Organizational Research Methods, 15*(2), 263-287.

Prensky, M. (2001). Nativos digitais, imigrantes digitais. *On the horizon, 9*(5), 1-6.

Prensky, M. (2005). Listen to the natives.

Prensky, M. (2009). H. sapiens digital: From digital immigrants and digital natives to digital wisdom. *Innovate: journal of online education, 5*(3).
Prensky, M., & Berry, B. D. (2001). Do they really think differently. *On the horizon, 9*(6), 1-9.
Pretorius, C., Chambers, D., Cowan, B., & Coyle, D. (2019). Young People Seeking Help Online for Mental Health: Cross-Sectional Survey Study. *JMIR mental health, 6*(8), e13524.
Pretorius, C., Chambers, D., & Coyle, D. (2019). Young people's online help-seeking and mental health difficulties: Systematic narrative review. *Journal of medical Internet research, 21*(11), e13873.
Puustinen, M., & Rouet, J.-F. (2009). Learning with new technologies: Help seeking and information searching revisited. *Computers & Education, 53*(4), 1014-1019.
Qiu, L., Chan, S. H. M., & Chan, D. (2018). Big data in social and psychological science: theoretical and methodological issues. *Journal of Computational Social Science, 1*(1), 59-66.
Raviv, A., Raviv, A., Vago-Gefen, I., & Fink, A. S. (2009). The personal service gap: Factors affecting adolescents' willingness to seek help. *Journal of adolescence, 32*(3), 483-499.
Raviv, A., Sills, R., Raviv, A., & Wilansky, P. (2000). Adolescents' help-seeking behaviour: The difference between self-and other-referral. *Journal of adolescence, 23*(6), 721-740.
Reed, D. A., & Dongarra, J. (2015). Exascale computing and big data. *Communications of the ACM, 58*(7), 56-68.
Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook* (pp. 1-35): Springer.
Rickwood, D., Deane, F. P., Wilson, C. J., & Ciarrochi, J. (2005). Young people's help-seeking for mental health problems. *Australian e-journal for the Advancement of Mental health, 4*(3), 218-251.
Rickwood, D., & Thomas, K. (2012). Conceptual measurement framework for help-seeking for mental health problems. *Psychology research and behavior management, 5*, 173.
Rickwood, D. J., Deane, F. P., & Wilson, C. J. (2007). When and how do young people seek professional help for mental health problems? *Medical journal of Australia, 187*(S7), S35-S39.
Riffe, D., Lacy, S., Fico, F., & Watson, B. (2019). *Analyzing media messages: Using quantitative content analysis in research*: Routledge.
Rogers, E., & Storey, J. (1987). Communication Campaign. Dalam CR Berger & SH Chaffe (Eds.), Handbook of Communication Science. New Burry Park. In: CA: Sage.
Roth, S., Dahms, H. F., Welz, F., & Cattacin, S. (2019). Print theories of computer societies. Introduction to the digital transformation of social theory. *Technological Forecasting and Social Change, 149*, 119778.
Rothì, D. M., & Leavey, G. (2006). Mental health help-seeking and young people: A review. *Pastoral Care in Education, 24*(3), 4-13.
Rowe, S. L., French, R. S., Henderson, C., Ougrin, D., Slade, M., & Moran, P. (2014). Help-seeking behaviour and adolescent self-harm: a systematic review. *Australian & New Zealand Journal of Psychiatry, 48*(12), 1083-1095.
Rudin, C. (2015). Can machine learning be useful for social science. *The Cities: An essay collection from the Decent City initiative, 9*(1), 86-90.
Ruppel, E. K., & McKinley, C. J. (2015). Social support and social anxiety in use and perceptions of online mental health resources: Exploring social compensation and enhancement. *Cyberpsychology, Behavior, and Social Networking, 18*(8), 462-467.
Ryan, A. M., & Pintrich, P. R. (1997). " Should I ask for help?" The role of motivation and attitudes in adolescents' help seeking in math class. *Journal of educational psychology, 89*(2), 329.
Schoen, C., Davis, K., Collins, K. S., Greenberg, L., Des Roches, C., & Abrams, M. (1997). The Commonwealth Fund survey of the health of adolescent girls. *New York: The Commonwealth Fund, 252*.

Schoenherr, S. E. (2004). The digital revolution. *Available online: ermarketer. com (accessed on 7 October 2008)*.

Schonert-Reichl, K. A., & Muller, J. R. (1996). Correlates of help-seeking in adolescence. *Journal of Youth and Adolescence, 25*(6), 705-731.

Schroeder, R. (2016). Big data and communication research. In *Oxford Research Encyclopedia of Communication*.

Schwartz, H. A., & Ungar, L. H. (2015). Data-driven content analysis of social media: a systematic overview of automated methods. *The ANNALS of the American Academy of Political and Social Science, 659*(1), 78-94.

Sears, R. R., Maccoby, E. E., & Levin, H. (1957). Patterns of child rearing.

Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big data, digital media, and computational social science: Possibilities and perils. *The ANNALS of the American Academy of Political and Social Science, 659*(1), 6-13.

Siegel, E. (2013). *Predictive analytics: The power to predict who will click, buy, lie, or die*: John Wiley & Sons.

Singh, V. K., Tiwari, N., & Garg, S. (2011). *Document clustering using k-means, heuristic k-means and fuzzy c-means.* Paper presented at the 2011 International Conference on Computational Intelligence and Communication Networks.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research, 70*, 263-286.

Slade, T., Johnston, A., Oakley Browne, M. A., Andrews, G., & Whiteford, H. (2009). 2007 National Survey of Mental Health and Wellbeing: methods and key findings. *Australian & New Zealand Journal of Psychiatry, 43*(7), 594-605.

Sorensen, H., Bogomolova, S., Anderson, K., Trinh, G., Sharp, A., Kennedy, R., . . . Wright, M. (2017). Fundamental patterns of in-store shopper behavior. *Journal of Retailing and Consumer Services, 37*, 182-194.

Steadman, I. (2013). Big data and the death of the theorist. Wired, 25 January 2013. *URL: http://www. wired. co. uk/news/archive/2013-01/25/big-dataend-of-theory (date of access: 30.01. 2013)*.

Steinbach, P. (2012). Dynamic pricing pinpoints market value. *Athletic Business (Retrieved February 1, 2014 from www. athleticbusiness. com/articles/article. aspx*.

Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). *Exploring topic coherence over many models and many topics.* Paper presented at the Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning.

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: key issues in developing an emerging field. In: SAGE Publications Sage CA: Los Angeles, CA.

Subramaniam, M., Jean, B. S., Taylor, N. G., Kodama, C., Follman, R., & Casciotti, D. (2015). Bit by bit: using design-based research to improve the health literacy of adolescents. *JMIR research protocols, 4*(2), e62.

Suzuki, L. K., & Calzo, J. P. (2004). The search for peer advice in cyberspace: An examination of online teen bulletin boards about health and sexuality. *Journal of applied developmental psychology, 25*(6), 685-698.

Szlyk, H. S., Roth, K. B., & García-Perdomo, V. (2020). Engagement with crisis text line among subgroups of users who reported suicidality. *Psychiatric services, 71*(4), 319-327.

Tapscott, D. (1998). *Growing up digital* (Vol. 302): McGraw-Hill Companies San Francisco.

Thompson, L. K., Sugg, M. M., & Runkle, J. R. (2018). Adolescents in crisis: A geographic exploration of help-seeking behavior using data from Crisis Text Line. *Social science & medicine, 215*, 69-79.

Tierney, P. J. (2012). A qualitative analysis framework using natural language processing and graph theory. *International Review of Research in Open and Distributed Learning, 13*(5), 173-189.

van den Berg, B. J., & Parry, M. A. (1983). Adolescents' preference of source to obtain contraceptive information. *American Journal of Obstetrics & Gynecology, 147*(6), 719-721.

van der Rijt, J., Van den Bossche, P., van de Wiel, M. W., De Maeyer, S., Gijselaers, W. H., & Segers, M. S. (2013). Asking for help: A relational perspective on help seeking in the workplace. *Vocations and learning, 6*(2), 259-279.

Vasarhelyi, M. A., Kogan, A., & Tuttle, B. M. (2015). Big Data in accounting: An overview. *Accounting Horizons, 29*(2), 381-396.

Vesset, D., Woo, B., Morris, H. D., Villars, R. L., Little, G., Bozman, J. S., . . . Conway, S. (2012). Worldwide big data technology and services 2012–2015 forecast. *IDC report, 233485*.

Wajcman, J. (2008). Life in the fast lane? Towards a sociology of technology and time. *The British journal of sociology, 59*(1), 59-77.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics, 34*(2), 77-84.

Wang, Y., Chan, S. C.-F., & Ngai, G. (2012). *Applicability of demographic recommender system to tourist attractions: a case study on trip advisor.* Paper presented at the 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology.

Wetterlin, F. M., Mar, M. Y., Neilson, E. K., Werker, G. R., & Krausz, M. (2014). eMental health experiences and expectations: a survey of youths' Web-based resource preferences in Canada. *Journal of medical Internet research, 16*(12), e293.

Wiedemann, G., & Wiedemann. (2016). *Text mining for qualitative data analysis in the social sciences* (Vol. 1): Springer.

World Health Organization. (2012). Adolescent mental health: mapping actions of nongovernmental organizations and other international development organizations.

Yao, L., Mimno, D., & McCallum, A. (2009). *Efficient methods for topic model inference on streaming document collections.* Paper presented at the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122.

Zachrisson, H. D., Rödje, K., & Mykletun, A. (2006). Utilization of health services in relation to mental health problems in adolescents: a population based survey. *BMC public health, 6*(1), 1-7.

Zhang, J., Wang, W., Xia, F., Lin, Y.-R., & Tong, H. (2020). Data-driven Computational Social Science: A Survey. *Big Data Research*, 100145.

Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y., & Zou, W. (2015). *A heuristic approach to determine an appropriate number of topics in topic modeling.* Paper presented at the BMC bioinformatics.

Zheng, A., & Casari, A. (2018). *Feature engineering for machine learning: principles and techniques for data scientists*: " O'Reilly Media, Inc.".

Zillner, S., Becker, T., Munné, R., Hussain, K., Rusitschka, S., Lippell, H., . . . Ojo, A. (2016). Big data-driven innovation in industrial sectors. In *New Horizons for a Data-Driven Economy* (pp. 169-178): Springer, Cham.

Zimmerman, B. J., & Pons, M. M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American educational research journal, 23*(4), 614-628.