The extent of sentiment in sexual health information between moderated and non-moderated websites

Jaimy Lai
STUDENT NUMBER: 6211291

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE APPLIED DATA SCIENCE
DEPARTMENT OF SCIENCE
SCHOOL SCIENCES
UTRECHT UNIVERSITY

Thesis committee:

Mirjam Visscher, Annemarie van Oosten
Remco Veltkamp

Utrecht University
School of Sciences
Department of Applied Data Science
Utrecht, The Netherlands
July 2022

Preface

As part of the MSc Applied Data Science program, students are required to write a final thesis to graduate from the program at Utrecht University. In this thesis, research was conducted to assess the extent of sentiment on sexual health topics between moderated and non-moderated websites.

This academic year has seen a lot of up's and down's, and the program was very insightful and educational in terms of the field that interests me the most: Data Science. I am happy to have chosen this Masters and to end my journey with this topic.

I would like to thank Mirjam Visscher and Annemarie van Oosten for their great supervision. I am very grateful to have had not only one but two supervisors who paid attention to their students, made us feel understood and supported at any time during the thesis progress. Writing the thesis would have definitely not been as enjoyable without supervisors who are this fun. So thank you for all the mornings and evenings you had to proof-read my work, and thank you for making this experience so easy and enjoyable. A shout-out goes out to Remco Veltkamp, who will be assessing this thesis, and who has helped the team be more critical in terms of our methods and argumentations.

A big thank you goes out to Esther Sernández and Jo Scheurs for being incredibly fun teammates. I hope the both of you will be happy in whatever you choose to pursue after this program.

Aside from my supervisors and teammates, I would like to mention de Kindertelefoon. For inviting us over to their office and talking to us about their platform. It was incredibly insightful to learn more about their admirable volunteer: helping young children who need someone to talk or who need advice. Thank you Mirjam, for organising this.

Furthermore, I'd like to thank my family for always being so supportive and for taking good care of me. I hope you can be proud. Last but not least, a thank you goes out to Jakob Lechner, for proof-reading this lengthy thesis.

The extent of sentiment in sexual health information between moderated and non-moderated websites

Jaimy Lai

Sexual health information shared online is not always credible. Due to the nature of the internet, and how it allows anyone to create or spread content, often misinformation occurs, and being misinformed about (sexual) health can be dangerous. This sparks a research interest to create a model that can predict credibility, in this case, sexual health information. To create a model, we must identify which factors mediate and modulate credibility. In this study, the aim is to evaluate 'sentiment' as a marker of credibility prediction. Using a moderated and a non-moderated source, we can compare if there is a significant statistical difference between the sentiment on sexual health information between the two sources. A statistical difference would indicate that sentiment is a promising candidate for credibility predictions, and the sentiment can tell sources apart in terms of credibility. Using a rule-based method (Pattern.nl) to compute sentiment and statistical analysis methods, it was concluded that there was no statistically significant difference between the credible and non-credible sources in terms of sentiment. However, some intruiging patterns surfaced such as the non-credible source scoring higher on high levels of sentiment, or a subtopic within sexual health information that did return statistically significant with a small effect size. Therefore, more research should be conducted to further analyze this marker.

1. Introduction

The internet is a place for youth to find information, connect with peers, for entertainment, etc. According to a rapport by the Central Bureau of Statistics (CBS) on the behavior of youth online, 96% of the people between the ages of 12 to 25 years old in the Netherlands are online daily. In this same rapport, it was indicated that 79% use the internet for finding information about goods and services. Almost 63% use the internet for reading the news, 65% reported finding information on health on the internet and roughly 64% use it for uploading pictures or music. Furthermore, the internet is the first and most preferred source for youth to learn about "embarrassing" health topics (de Graaf et al. 2017). These are topics that young people feel too embarrassed about to mention to educators, health care providers, or parents (Gray et al. 2002). Topics include sexuality, body changes, pregnancy, sexual fantasies, contraception, and sexually transmitted infections (STIs).

A report by (de Graaf et al. 2017), where interviews were conducted with people ages 12 to 25 found that most people find information about sex online, followed by asking friends for advice, and a small part of the group mentioned asking their mother for advice. In more recent years, schools have started to incorporate a more diverse set of sexual health topics to educate the youth on than before. The focus has shifted

from mostly only educating on STIs and pregnancy prevention to education on healthy (sexual) relationships, positive experiences of sex, present-day topics (e.g. the influence of online behaviors), homosexuality, and different genders (Meijer 2019). Young people seem to care increasingly more and are more open to learning about sexual health, and it is, therefore, one of the most searched topics by youth on the internet (Borzekowski and Rickert 2001). However, despite the recognition of the importance of correct and open sexual health information, the knowledge of STIs seems to have decreased since 2012 (Marra, de Graaf, and Meijer 2020), with 4 out of 10 youngsters thinking you won't get an STI if you wash thoroughly after sex.

This problem is concerning and therefore sparks an interest that led to the creation of this study. Nowadays, most people are using the internet and a large group of the population is using the internet very extensively to the point where researchers have been studying whether it is an addiction or a possible emerging lifestyle (Bergmark, Bergmark, and Findahl 2011). This means online users are constantly confronted with online content created by anyone, and thus makes them susceptible to misinformation when it is presented to them considering the plethora of content available. This content is often not checked on credibility. Especially social media can be dangerous in this regard, which is popular for news consumption due to easy access, fast dissemination, and low efforts. However, this also enables the propagation of misinformation online. It is, therefore becoming increasingly important to find ways to regulate and prevent misinformation or recognize a falsehood.

A good demonstration of why research of this nature (analyzing the credibility of online news) is so important is a study by Greene and Murphy. In this study, they showed participants a fabricated story about privacy concerns with a national contact tracing app, which consequently led to participants being less willing to download the app. Furthermore, research by Pennycook and Rand concluded that people often fail to discern truth from fiction because they do not stop and reflect on the accuracy of what they see online. However, it was mentioned that digital literacy tips and prompts to shift people's attention to be critical of what they read could increase the quality of news people share online. To tackle this emerging problem of fake news influencing people's behavior, we must find ways to prevent people from believing fake news or to help identify them. Previous studies have aimed at creating models which can predict credibility (McGlynn, Baryshevtsev, and Dayton 2020; Kakol, Nielek, and Wierzbicki 2017). The ability to be able to predict whether a text is credible is valuable research for this current social problem. In these studies, the aim is to make progress in the research of the credibility of online content to avoid misinformation generated by online users. It is important, especially in very sensitive topics such as sexual health to be correctly informed to avoid possible dangers. However, the same principle transcends sexual health information shared online and can be applied in many ways. For instance, research by Zhou et al. and Gundapu and Mamidi aimed to create a model to predict whether corona news is credible in order to stop the spread of false news. Another example is a study by Singh and Sharma, where fake images spread online were analyzed. Advancing in such research can have big advantages. Platforms where most of the fake news is hosted (e.g. on fora but also Facebook, and YouTube) could start implementing these algorithms to analyze content and remove the content which contains only falsehoods. In order to create a sound model, we must identify which factors mediate and modulate credibility as variables we can assign to the model. The definition of credibility as considered in this paper is the confidence that can be placed in the truth of the information, news or findings shared. One of the possible markers is called sentiment. Sentiment is the negative, positive or neutral attitude towards an

entity e.g. products, organizations, individuals, issues, and topics (Zhang, Wang, and Liu 2018). Most of the literature on sentiment in relation to credibility demonstrates that there is a correlation between negative sentiment and reduced credibility (Newman et al. 2003; Ott, Cardie, and Hancock 2013). Additionally, within the field of psychology, research has been done on linguistic traits of lies and their psychological effect and Kwon et al. showed rumors to be less likely to have a positive sentiment when analyzing two-thousand tweets using a dictionary-based sentiment analysis tool.

However, there is also research that indicates different results. Castillo, Mendoza, and Poblete demonstrated how non-credible information tends to exhibit both positive and negative sentiment, but especially positive sentiment. Additionally, Hu et al. analyzed the difference between spammers and other social media uses when it comes to sentiment and concluded that spammers yielded a more positive sentiment. In this research, 62K Twitter users' tweets were used for analysis using linear regression for sentiment analysis. According to the researchers, this could be related to spammers mimicking social bots. Additionally, in an interview with Carolien Gravemaker and Roelie Heijmans (interview, May 30, 2022) from de Kindertelefoon (a website for young people to ask questions and advice under moderation from experts) described the sentiment experienced by them on the forum as, "Children are generally very positive. Especially when someone needs encouragement to ask someone out or what to do when it is their first time or first kiss. The cases where children are very negative is when someone posts a comment or story in which they tell a wild fantasy or fetish that is unbelievable. Children would then often react in disbelief and suspicion when someone says that they, for instance, are sexually abused but enjoyed it and thus would want to try it again." Indicating, that children, when met with something that is crazy, are suspicious and more negative than when they are met with a believable story.

Sentiment in the context of Natural Language Processing (NLP) is a sub-area that aims at automatically detecting the polarity of a text based on textual information. Sentiment polarity of an element defines the orientation of the expressed sentiment (positive, neutral, or negative sentiment). There are two main approaches to performing sentiment analysis and determining the polarity of a text: the rule-based (lexicon-based) approach and the machine learning approach.

Rule-based approaches entail making predictions using a dictionary of opinion words (e.g. 'nice' or 'awful') which are rated at a certain value to determine the polarity of a text document. The machine learning approach is based on annotated data in which data is collected, and an annotator will label the data to then feed it to a model as training data. Predictions are made using the model it created. Both approaches come with different considerations. Rule-based methods need to be maintained more. It needs an implementation that can distinguish words based on their context of use, as well as it needs to take words individually into consideration and give them all a polarity value. On the other hand, machine learning approaches are domain specific and need a large amount of labeled data for them to perform well e.g. a model that has been trained on restaurant reviews from Google, will not perform similarly on Twitter data on the US elections (Aue and Gamon 2005). It also needs annotators, meaning there has to be a set of rules anyone can follow and there needs to be insurance that the annotations between different annotators have similar accuracy for the model to work well.

Sentiment analysis is considered a classification problem and comes with its own challenges. When conducting sentiment analysis, the possible challenges that need to be taken into account are term presence and frequency, Parts-Of-Speech (POS), opinion words (e.g. 'good or bad'), and expressions (e.g. 'it cost me an arm and a leg'), as well as negations (e.g. 'not so bad' is positive, despite the two negative words). These

problems arise mostly from computers, which do not understand semantics the same way humans do. It is therefore important to take these issues into consideration when working with NLP.

The goal of this research is to identify whether sentiment is a possible factor that can help mediate and modulate credibility predictions, in particular for the topic of sexual health information. In order to answer this question, we analyze whether there is a difference between a credible source (a moderated forum) and a less credible source (a non-moderated forum). Furthermore, there is an interest in the orientation of the sentiment (positive, negative, or neutral sentiment) and if there are nuances between subtopics within sexual health information. Based on the literature, we can make the following hypothesis: there is a difference in sentiment between sources that are credible and less credible, with the less credible (non-moderated) source holding more sentiment.

2. Data

2.1 Datasets

The experiments were performed using datasets that are obtained by web scraping a moderated and non-moderated forum, namely de Kindertelefoon and FOK! forum respectively.

2.1.1 Webscraping: Beautiful Soup. Web scraping was performed using the Python library Beautiful Soup. It is a library that is used to pull out the data from HTML or XML files (Rietvelt 2019). It is considered a simple way to extract data from HTML, given that the pages of the website are well structured. It transforms the textual files into an object iterating, searching, and modifying Python parse tree. Furthermore, it can help you to scrape websites but also to clean the data obtained.

The data was web scraped by making a request using page content and transforming it into a BS4 (Beautiful Soup) object, which can be used for web scraping the specific data that is sought after (Richardson 2007). Using HTML tags, the desired variables e.g. URLs, comments, or time a post is created, can be collected.

2.1.2 De Kindertelefoon. To find a distinction between credible and not credible information that is shared online by users, a source that is moderated by experts was used. De Kindertelefoon is a Dutch helpline, for children of age 8 to 18 who need advice, information, or a conversation about anything. The organization aims to create a safe space where children can freely and confidentially talk about subjects that they do not dare, cannot, or do not wish to discuss in their environment. Additionally, children can also post a topic on the forum of de Kindertelefoon, where other young people (under the watchful eye of a moderation team with experts) can provide answers, advice, or support to the thread. De Kindertelefoon has around 700 volunteers who have followed appropriate and extensive internal training to moderate, help and advise children. Within the forum, comments are reviewed on falsehoods. When a child gives false information to another child, the moderators will step in with a 'mod break' to correct the incorrect information so anyone reading the thread will be provided with the right information. Therefore, this forum has been selected as the source which is expected to contain more credible information.

2.1.3 FOK! forum. To contrast a moderated source, a non-moderated website has to be considered to find whether there is a difference in sentiment. FOK! Is a Dutch website and is one of the largest online communities in the Netherlands. The forum's users are from diverse backgrounds and ages but the general content is geared toward a younger audience. The content found on FOK! contains e.g. a forum, polls, giveaways, columns, and news. Its forum contains a wide variety of topics of which 'sexuality' is the topic that will be considered within this research. This topic contains, currently, roughly 25,000 threads. A moderator is in charge of the topic of sexuality but does not correct falsehoods. The moderator of the topic makes sure rules are followed by removing spam or degrading threads and comments.

2.1.4 Dataset setup. The specific data that has been used from de Kindertelefoon and FOK! forum is the id of the topic (TopicID), the content of the comment posted (Content or Comment), the date on which the comment has been posted (CreateTime), and whether it is the first post of the user (FirstPost). Essentially, only the comment will be used for analysis.

As for the threads dataset, the following was web scraped: the id of the thread (ID), the title of the thread (Title), when the thread was created (CreationTime), the last response within the thread (LastReplyTime or LastResponse depending on the forum) and the number of replies, views and likes the thread has (NReplies, NViews, and NLikes). In this case, mostly the ID column will be utilized within this research for identifying topics.

The differences in naming are due to the HTML tags that the designers of the websites have decided on. The reason to keep separate datasets for the comments and the threads is for better data management (see Appendix A). Using the column ID in the threads dataset and the column TopicID in the comments dataset, a thread belonging to a certain comment can be found more efficiently.

2.2 Data exploration

In this section, the results of the data exploration will be discussed.

Both datasets contain only Dutch comments from online anonymous users but English words can be found in the dataset (due to the popularity of English slang). The data is only a sample of the sexual health category provided on the platforms and ranges between the years 2014 and 2021. De Kindertelefoon contains 87706 comments, and 116962 comments which have been collected are from the FOK! forum. Interestingly, there are more topics web scraped from de Kindertelefoon, namely 10947 topics and 2053 from FOK! forum (Table 1).

	Threads	comments	Average comments per thread	Highest count of replies in a thread	Lowest count of replies in a thread	Empty comments
de Kindertelefoon	10947	87706	∼37	463	0	3656
FOK!	2053	116962	∼56	301	0	3437

Table 1
Dataset exploration details of de Kindertelefoon and FOK! forum

After web scraping and removing the empty cells, the number of comments per forum was found to be imbalanced. There are more comments of FOK! Forum available

than de Kindertelefoon, which is due to FOK! forum being the more active website (see Figure 1).

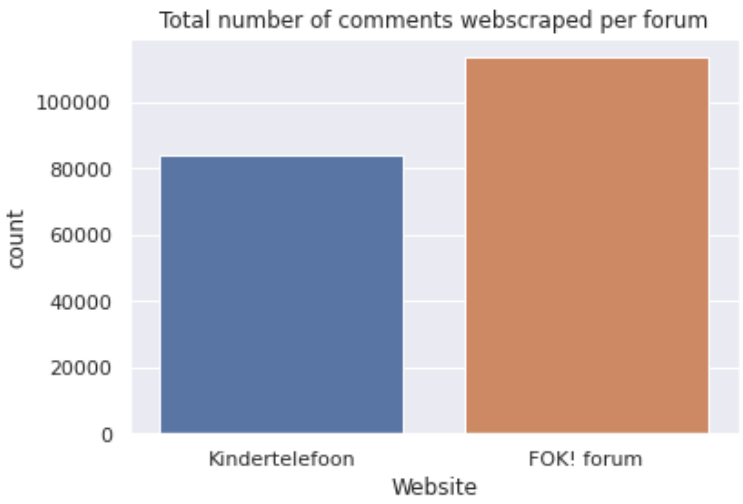

Figure 1
Count plot of the number of data points (comments) collected for each forum.

To tackle this problem, a sample was taken from the FOK! forum using random sampling to match the number of comments web scraped from de Kindertelefoon. Random sampling ensures that the sampling is performed without bias.

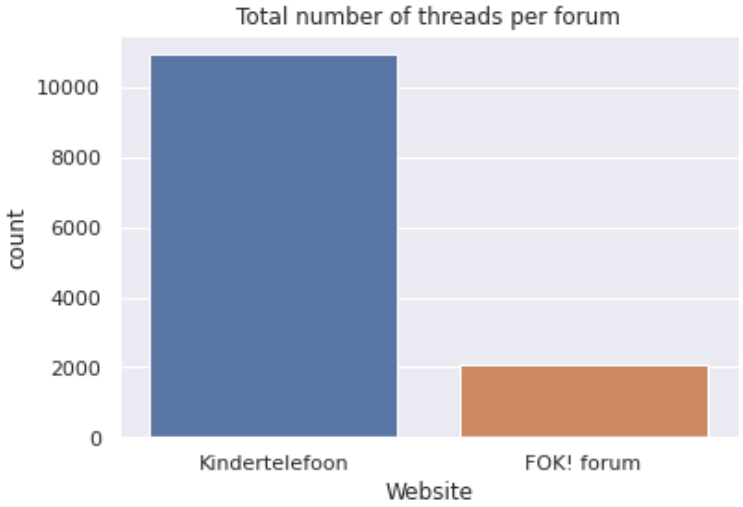

Figure 2
Count plot of the number of threads collected for each forum.

2.2.1 Thread analysis. As can be seen in Figure 2, the number of threads for the Kindertelefoon is 5 times as high as the threads found on FOK! forum. Meaning that

given the number of comments retrieved from both fora, de Kindertelefoon has fewer comments per thread than FOK! forum.

This pattern can also be found when plotting the number of comments per thread. On the x-axis, the number of comments is depicted and on the y-axis the number of threads that have this amount of comments. As shown in Figure 3 and Figure 4, there seems to be a different distribution of comments per thread.

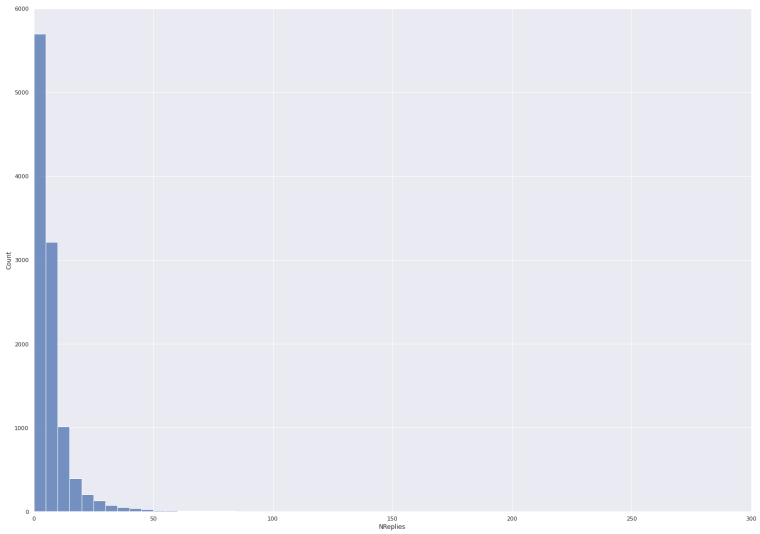

Figure 3
De Kindertelefoon distribution of comments per thread (with an x-axis cutoff of 300 for the readability of the graph).

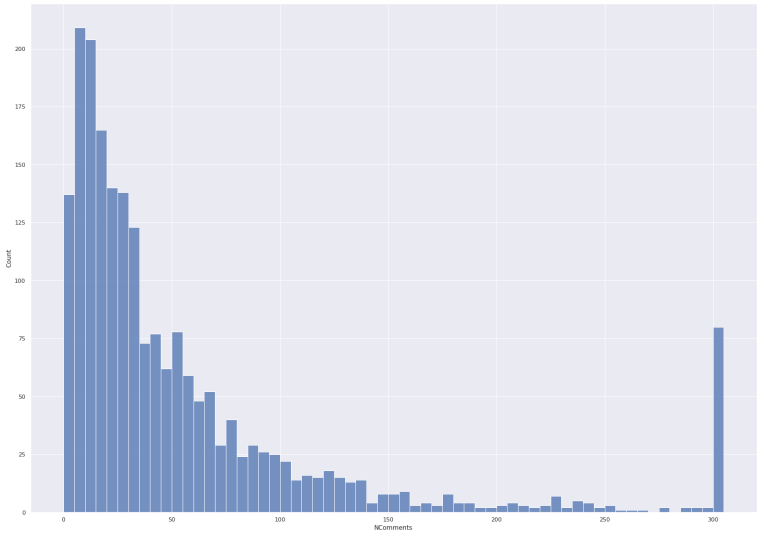

Figure 4
FOK! forum distribution of comments per thread.

2.2.2 Keyword analysis. A keyword analysis was conducted to find the most commonly used words per forum. This was performed by taking the entire list of comments per forum, disregarding the topic, and creating a word cloud to find the most used keywords (see Figure 5 and 6).

Interestingly, many words appear in both fora. Noticeably, the word 'seks' is the most apparent word in both word clouds. Furthermore, words that indicate sexual acts such as 'klaarkomen', 'pijpen', 'vingeren' are also noticable in both word clouds. Lastly, words that implicate a gender e.g. 'meisjes', 'jongens', 'vriendin', 'man', and 'vrouw' is evident in both de Kindertelefoon and FOK! forum.

Figure 5
Kindertelefoon word cloud

Figure 6
FOK! Forum word cloud

2.2.3 Preprocessing. For preprocessing, there are two main steps. The preprocessing for the sentiment analysis, and the preprocessing for the topic modelling. Using sentiment analysis, we can find the sentiment per comment, and using topic modelling we can find subtopics within the data for each forum.

Sentiment analysis
The datasets of both de Kindertelefoon and FOK! forum contained many empty cells.

These empty cells were removed. Resulting in 113526 data points for FOK! forum and 84050 data points for de Kindertelefoon.

Due to the imbalance in the data, the datasets were balanced out using random undersampling. Random undersampling ensures that balancing the data was performed without bias. In this study, the Dutch sentiment analysis tools require no further preprocessing. After an evaluation of different tools (see method section), the sentiment analysis tool 'Pattern' will be used and it contains built-in preprocessing units, whole sentences were used in order to retrieve results. As indicated by the documentation on Github1, the library can handle tasks such as tokenization, and lemmatization on its own.

The only preprocessing step taken was the removal of rows with empty comments. De Kindertelefoon's dataset contained 3656 empty comments, and FOK! forum's dataset contained 3437 empty comments.

Topic modelling
In order to get the different topics per fora, topic modeling was used. Topic modeling is an unsupervised machine learning method that can analyze data and determine, based on words and phrases, which documents cluster together using natural language processing techniques (NLP). Topic models discover hidden themes throughout a set of documents and annotate them according to those themes, then a document coverage distribution is generated which provides new ways to explore the data in the structure of topics (Tong and Zhang 2016).

To perform the topic modelling, the following preprocessing steps were taken:

- Tokenization: breaking raw texts into small chunks (of words)
- Removing 'gewijzigd' messages: the data would also record all messages that included automated notification that a comment has been
- 'changed' and therefore is not representative of the users' comment.
- Remove improperly quoted content: content that was quoted very often would reappear in the data and therefore is redundant
- Removal of Dutch and English stopwords: by removing low-level information words, more focus is given to important information
- Tag removal: removal of comments which contain another user's comment
- Lemmatization: in order to group words together that have the same meaning but may be spelled differently e.g. 'plays' and 'playing' will turn into 'play'
- Removing numbers
- Removal of empty comments: could perhaps be comments with images or wrongly scraped comments due to unknown reasons

2.3 Ethical and legal considerations

Ethical and legal considerations should be evaluated before web scraping. According to previous legal cases within the Netherlands, the Court of Justice ruled that it is not allowed to use data from other websites when this is forbidden in their Terms and

Conditions. It is thus not allowed to web scrape any website. In this study, permission was asked from both de Kindertelefoon and FOK! forum. De Kindertelefoon granted permission to web scrape the data from their forum and is aware of the nature of this study. FOK! forum did not reply to the request but their website nor Terms and Conditions mentioned it to be forbidden to use their data that is publicly available on their forum.

Additionally, we must mention the ethical aspect of web scraping data available on online platforms where online users share thoughts and opinions. The users whose comments have been web scraped are not aware of their data being used for research purposes, specifically sentiment analysis. There is a valid reason to believe this to be an unethical practice, as the participants are not aware that their data is used. On the contrary, one can argue that putting anything out on a publicly available website has the possibility of their data being used for any sort of purpose, and participants are aware that anyone can see their comments. Furthermore, to protect the users whose comments have been used for this study, all comments have been anonymized and no usernames are included in the data. However, users on both de Kindertelefoon and FOK! forum are already anonymous, to begin with, and none of the users can be traced to their identity. It is strongly unethical to use the comments of real-life people for research purposes of which they are not aware and their identity is revealed without their consent.

3. Method

The goal in this research is to answer whether sentiment is a suitable factor to take into consideration as a variable for credibility prediction in the domain of sexual health information. In order to answer the question, we must find out whether there is a distinction between credible and non-credible sources in regard to sentiment. In other words, is there (statistically) a significant difference in sentiment for credible and non-credible sexual health information sources?

To evaluate whether there is a significant difference, certain steps must be followed. First of all, to find a significant difference in sentiment between credible and non-credible sources, we must calculate the sentiment of the comments for each source in question. This can be done using sentiment analysis.

Pattern

Sentiment expressed in comments on both fora was decoded by performing sentiment analysis using Pattern.nl (De Smedt and Daelemans 2012), a submodule from Pattern. Pattern.nl is an open-source Python library for NLP that is developed and maintained by the Computational Linguistics group at Universiteit Antwerpen (CLiPS) and contains a submodule for the Dutch language (Gatti and van Stegeren 2020). The submodule contains a rule-based sentiment analyzer based on a built-in lexicon of about 4000 Dutch lemmata with each a polarity and subjectivity score for each word. The key aspect of sentiment analysis is to analyze a body of text for the sentiment and to comprehend the opinion expressed by it. This can be quantified using positive and negative values and is often called 'polarity'. Polarity is essentially a measure of the overall combination of the positive and negative emotions in a body of text. The subjectivity, although not the focus of this study, is not a measure to be overlooked. It is the overall subjectiveness of a body of text and is an indication of how subjective a body of text can be taken. Sentiment and emotion are subjective and perhaps random matters.

Sentiment and polarity scores are calculated using the dictionary within Pattern.nl. As mentioned before, the dictionary contains 4000 Dutch lemmata, which have a subjectivity and polarity score assigned to them. By identifying positive and negative words in any body of text, Pattern.nl can use the lemmata to then calculate the sentiment and subjectivity of the text overall. Scores for polarity range between -1 and 1 for negative and positive sentiment respectively, and range between 0 (not subjective) and 1 (highly subjective) for subjectivity.

In order to evaluate Pattern.nl's fitness for this study, a small comparison was conducted between Pattern.nl and other Dutch sentiment analysis methods. In the evaluation, we compare BERTje (de Vries et al. 2019), RobBERT (Delobelle, Winters, and Berendt 2020), and Pattern.nl.

BERTje and RobBERT are two state-of-the-art transformer-based architectures for the Dutch language and based on the BERT architecture originally released for English (Devlin et al. 2018). The BERT model stands for 'Bidirectional Encoder Representation from Transformers' and is based on the Transformers model (Vaswani et al. 2017), which is a deep learning model for NLP problems that adopts the mechanism of self-attention by tracking relationships in sequential data (e.g. sentences). Unlike the models preceding it, BERT is designed to pre-train deep bidirectional representations from an unlabeled text by jointly conditioning on both left and right context in layers (Devlin et al. 2018), meaning it can read a text in both directions at once to understand its context. While both BERTje and RobBERT are based on the BERT model, their training data differs. BERTje is pretrained on a 12GB Dutch corpus composed of different topics: Wikipedia, news, books, and social media. Whereas RobBERT is trained on the OSCAR corpus's Dutch section (Suárez, Sagot, and Romary 2019), which is the largest web crawl corpus available to this day. The motivation to test BERTje and RobBERT is due to a limit of Dutch sentiment analysis models available, and compared to their predecessors were found to be state-of-the-art models.

All three methods were evaluated based on the correctness of classification for each sentiment (positive, negative, and neutral). 20 sentences per sentiment were given to the methods to classify. Pattern classified the best, followed by BERTje and then RobBERT. Although BERTje and RobBERT classified well for positive sentiment, negative sentiments were classified often incorrectly. As for the neutral sentiment, both BERTje and RobBERT did not classify neutral sentiment - only positive and negative sentiments.

	Positive (N=20)	**Negative (N=20)**	**Neutral (N=20)**
BERTje	18	13	0
RobBERT	17	4	0
Pattern.nl	18	17	19

Table 2
Evaluation of correctly classified positive, negative and neutral comments per method

After testing, it can be established that BERTje and RobBERT, although performed well in papers, would not be suitable for this research (see Table 2). First of all, this is due to the models not fitting the domain. As mentioned before, rule-based methods are widely employed for general-purpose sentiment analysis (Crocamo et al. 2021) and can therefore be used on this domain. In the case of machine learning models such as BERTje and RobBERT, the models' sentiment analysis modules have not been trained in the domain of sexual health information and thus are not a well fit due to a lack of

understanding of the context within this domain. The data they have been trained on are book reviews. Secondly, the models are trained to label comments as 'positive' or 'negative', leaving the 'neutral' polarity much to be desired. This was established when testing neutral sentences e.g. 'De boom is droog' (translation: 'The tree is dry.') would yield a highly positive sentiment when it is given a neutral sentence.

As reflected in Table 2, RobBERT did not classify negative sentiments well. Notably, highly negative sentences such as "Ik haat je" (translation: "I hate you") were classified as positive. After multiple runs, it was finally able to classify it as a negative sentiment, meaning that the model is not entirely reliable due to the inconsistency. An important note is that given more domain-specific training, both BERTje and RobBERT would vastly improve but the BERTje model is indicated to be more reliable.

Therefore, Pattern.nl was evaluated as the most fitting and significantly accurate model. Pattern.nl makes a distinction between positive, negative, and neutral sentiment, and gives a sentiment score representative and most often agreeable of the comment it is given. Furthermore, after multiple runs, it is the most consistent and fairly computationally inexpensive. Lastly, it is believed to be monitored by a specialized computational linguistics team at the University of Antwerp, and is considered a powerful NLP tool that has been trained on extensive data retrieved from Twitter, Wikipedia, Bing, and Google.

Latent Dirichlet Allocation In order to perform topic modelling, the Latent Dirichlet Allocation (LDA) model was chosen. LDA is a heavily cited dimensionality reduction machine learning method. Essentially, it is a model that generates topics based on word-frequencies. It looks at a set of documents, and can connect which documents belong in which category based on the word frequency and by using labelling (Ramage et al. 2009).

Its ability to perform topic modelling is widely used in order to find the most popular topics within a large set of data (Hagen 2018). It is likely due to the variety of its potential applications and its performance regarding feature reduction (Wang et al. 2020). LDA excels at receiving large amount of information, and classifying it without much loss of information.

4. Results

In this section, we will explore the results of the analysis.
To interpret the results, an analysis was conducted to evaluate the behavior of the data on sentiment retrieved after processing all comments using Pattern.nl. We compare the mean of all comments per forum. This is not a sound metric, due to the possibility of a forum being both equally significantly negative and significantly positive, the scores would even out into an overall neutral score due to extremes balancing out. Therefore, it is also interesting to evaluate the sentiment for each forum based on how positive or how negative it is. Thus, another comparison is made based on only positive comments, and only negative comments. Finally, a statistical analysis was conducted using z-tests on the dataset to test for statistical significance in the difference between sentiments.

Finally, we also look at the differences per topic that arose from the topic modeling analysis and how the sentiment differs per forum and evaluate the difference using statistical analysis with z-tests.

4.1 General analysis

A general analysis was conducted using the raw outcome of Pattern.nl's sentiment analysis.

	Sentiment score		**Subjectivity score**	
	M	sd	M	sd
De Kindertelefoon	0.088	0.25	0.52	0.31
FOK! forum	0.065	0.26	0.43	0.34

Table 3
The average scores of sentiment analysis

As shown in Table 3, de Kindertelefoon has an average sentiment score of M=0.088, sd=0.25, across the forum's sexuality category and the FOK! forum has a slightly lower score of M=0.065, sd=0.26. The subjectivity score showed a score of M=0.52, sd=0.31 for de Kindertelefoon and M=0.43, sd=0.24 for FOK! forum, meaning that the interpretation of negative or positive is highly subjective, indicating that the polarity score is not entirely generalizable. High polarity indicates that the sentiment could be vastly higher or lower depending on the reader.

Furthermore, a calculation has been made on the scores of negative and positive sentiment only (see table 4).

De Kindertelefoon has a positive sentiment score of M=0.25, sd=20, and a negative sentiment score of M=0.20, sd=0.19 and FOK! Forum has a positive sentiment of M=0.27, sd=0.21, and negative sentiment of M=-0.25, sd=0.22. FOK! forum, therefore, showed higher absolute sentiment scores for negative and positive values than de Kindertelefoon. As can be seen in Figures 7 and 8, the data seems to be distributed fairly normally. Additionally, a large number of 0 values can be found in the results. Upon further inspection, it was found to contain a high number of empty cell comments that were not understood by Pattern.

	Sentiment score (positive)		**Sentiment score (negative)**	
	M	sd	M	sd
De Kindertelefoon	0.25	0.20	-0.20	0.19
FOK! forum	0.27	0.21	-0.25	0.22

Table 4
Positive and negative sentiment scores (only) of de Kindertelefoon and FOK! forum

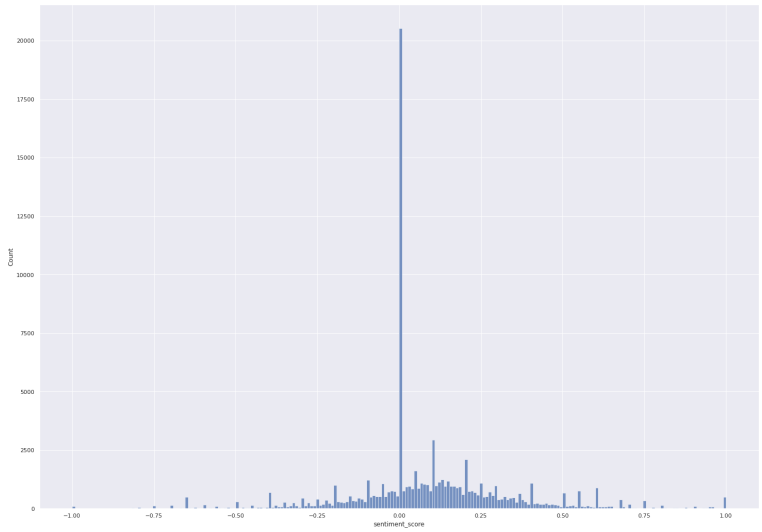

Figure 7
Histogram of all sentiment scores in de Kindertelefoon dataset.

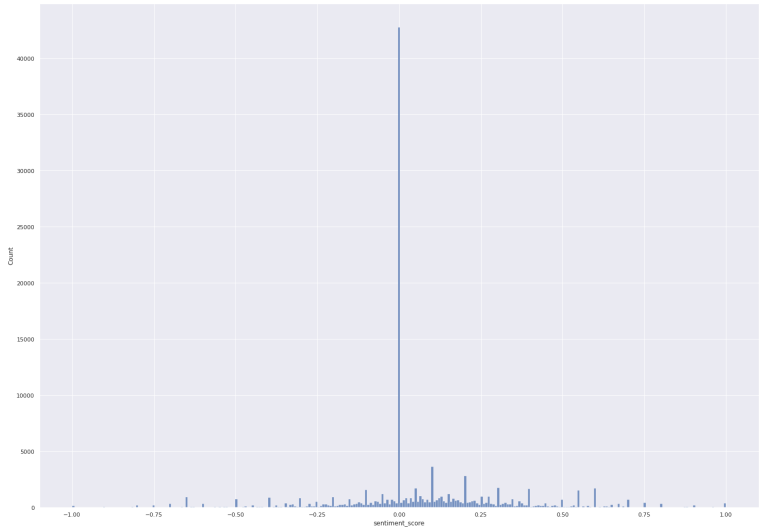

Figure 8
Histogram of all sentiment scores in the FOK! forum dataset

After careful analysis, it became apparent that the comments which were not understood by Pattern.nl are mostly abstract comments, which are in most cases too short for Pattern.nl to understand the sentiment, contain a high amount of Dutch slang, or perhaps words that are not in Pattern.nl's dictionary leading it to not understand the user-generated comments (e.g. niche words).

	Word count mean	Sentence count mean
Understood	77.25	9.77
Not Understood	10.99	2.041

Table 5
De Kindertelefoon word count and sentence count analysis

	Word count mean	Sentence count mean
Understood	45.19	3.95
Not Understood	7.72	1.4

Table 6
FOK! forum word count and sentence count analysis

According to the word count and sentence analysis in Table 5 and Table 6, the word count and sentence count of the comments that the sentiment analysis model understood are significantly higher than the comments that the model did not understand. The word count of understood comments from de Kindertelefoon and FOK! forum was roughly 8 times higher than the word count of not understood comments, and the sentence counts were roughly 3 to 4 times higher than the not understood comments. This implies that Pattern.nl does not handle short sentences well, or does not understand short sentences, due to them possibly not providing enough context for Pattern.nl to understand the sentiment.

In order to give a more representational view of the sentiment per forum, the rows within the dataset that indicated that the model does not understand certain comments and thus gave a score of 0.0 sentiment and 0.0 subjectivity, have been omitted. This reduced the amount of 0 scores and thus the bias of the 0 score drastically (see Figures 9 and 10).

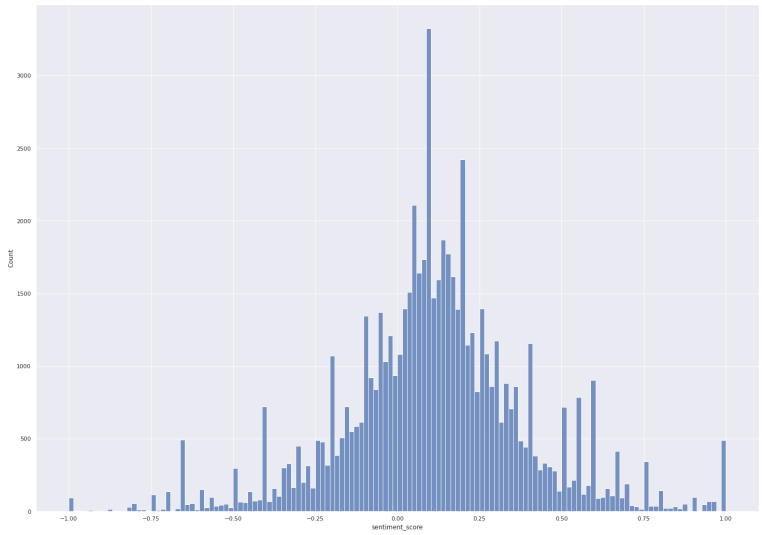

Figure 9
De Kindertelefoon sentiment scores distribution after removal of 'not understood' comments by the model

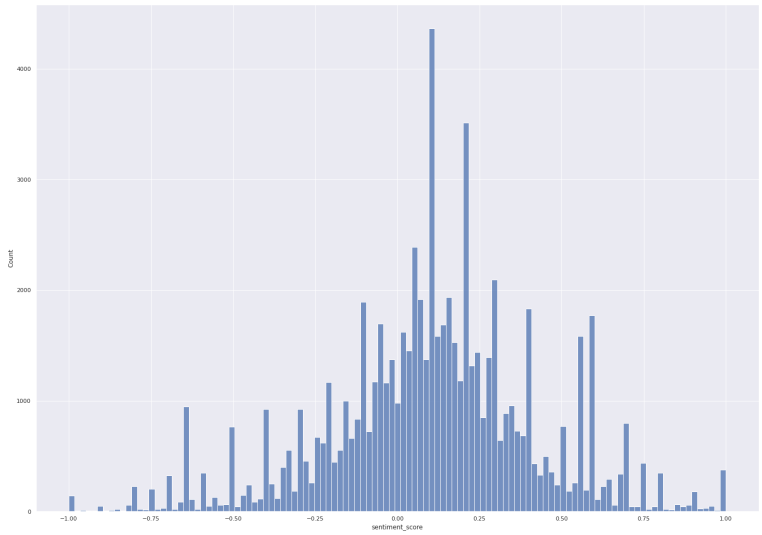

Figure 10
FOK! Forum sentiment scores distribution after removal of 'not understood' comments by the model

Re-evaluating the overall sentiment per forum using the mean we get the following results (see tables 7 and 8).

There seems to be an increase in (positive) sentiment as shown in table 7 and a score of M=0.65, sd=0.19 in subjectivity for both fora. As previously mentioned, it should be considered that this is an unclear metric as using the average can even out two extremes on the positive and negative side.

	Sentiment score		Subjectivity score	
	M	sd	M	sd
De Kindertelefoon	0.11	0.28	0.65	0.19
FOK! forum	0.10	0.32	0.65	0.19

Table 7
Sentiment and subjectivity scores after removal of not understood comments by the sentiment analysis model

	Sentiment score (positive)		Sentiment score (negative)	
	M	sd	M	sd
De Kindertelefoon	0.24	0.20	-0.20	0.19
FOK! forum	0.27	0.21	-0.24	0.22

Table 8
Positive and negative sentiment scores of de Kindertelefoon and FOK! Forum

Conducting the negative and only positives analysis again yields the identical scores as in table 3 as the 0 neutral scores were left out entirely in the analysis of only positive and negative results (see table 8).

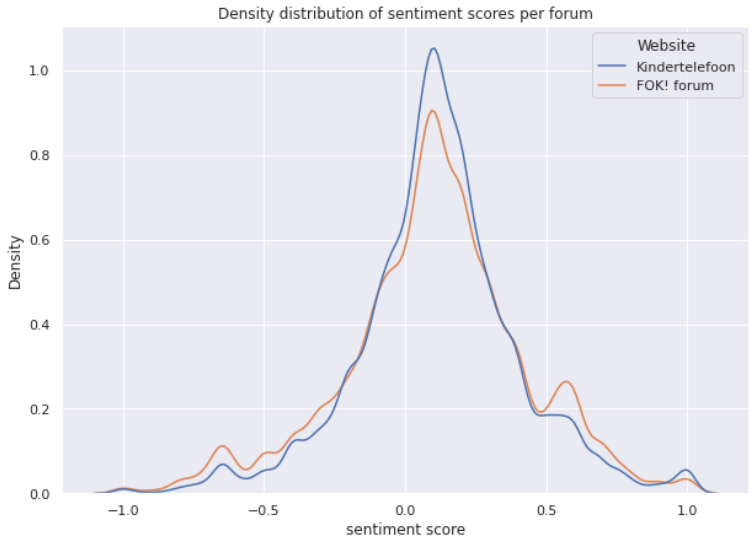

Figure 11
Distribution of sentiment scores per forum after removal of not understood comments.

In figure 11, the distribution of the sentiment scores per forum is shown, and it reflects what can be observed in Tables 7 and 8. In the tables, you can observe that de Kindertelefoon has a slightly higher average than FOK! forum, closely around the neutral 0 sentiment score. Moving away from the 0 (neutral sentiment score), we can observe that there is a higher number of FOK! forum comments that have sentiment

scores that score high (in negative and positive sentiment). Interestingly, we can observe that both de Kindertelefoon and FOK! forum follow a similar pattern. On the negative sentiment orientation (sentiment <0), there seems to be a small increase in the number of comments with a negative orientation around roughly -0.7 sentiment, and on the positive sentiment orientation (sentiment>0) is an increase around the 0.6 sentiment score as well the sentiment score of +/-1 for both fora.

As mentioned previously, de Kindertelefoon has a higher volume in sentiment scores that are located close to the 0 or the range that is classified as neutral. On the contrary, FOK! forum seems to have more scores that are highly negative or highly positive than de Kindertelefoon. Only at maximum positive sentiment (sentiment score of 1) does de Kindertelefoon exceed FOK! forum, aside from around the neutral scores.

A comparison between the moderated (M=0.11, sd=0.29) and non-moderated forum (M=0.10, sd=0.32) was performed using a two-sample independent z-test to understand if there is a significant difference between the scores of each respective group. The statistical analysis performed fits the assumptions of the z-test, as the data is continuous, normally distributed for both groups, large population (N>30), and the standard deviations are known. The result (z=6.36, p<0.001, two-tailed), d=0.033 was statistically significant, meaning the null hypothesis (H0: the populations are the same) could be rejected. However, conducting Cohen's d indicates that the effect size is trivial and therefore the difference is not relevant.

Topic modelling

In order to analyze whether in different sexual health subtopics there is a statistical difference in sentiment, different topics must be identified and compared. LDA was performed on both the dataset of de Kindertelefoon as well as FOK! forum. This was performed by looking at the most prominent topics within a thread, and clustering threads that fit together.

The results of the LDA model resulted in the following 16 topics per fora. Using the words shown per topic in Figures 12 and 13, one can get an instinct of the different topics identified. The higher the blue colored bar, the more the topic modelling model is certain about the topic.

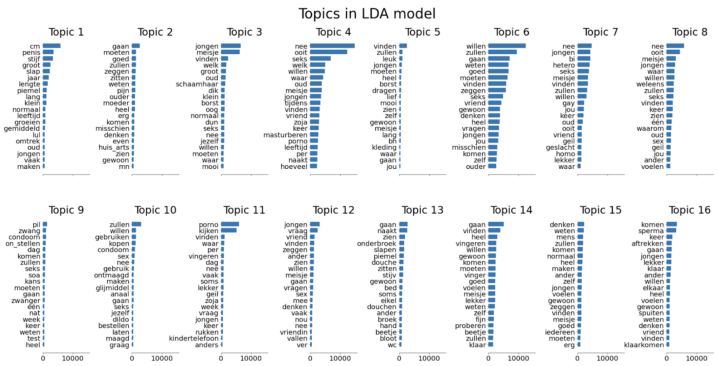

Figure 12
LDA topic models for de Kindertelefoon

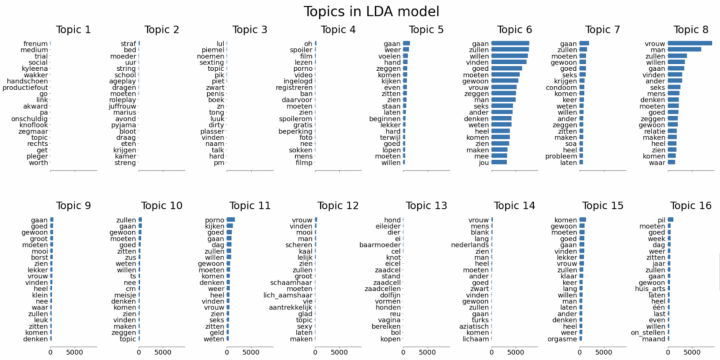

Figure 13
LDA topic models for de FOK! forum

Based on which topics were most prominent and found in both results of de Kindertelefoon and FOK! Forum, the following topics have been chosen for analysis:

- Shape of genitals - Kindertelefoon: topic 1, FOK!: topic 2
- Birth control, pill, and pregnancy - Kindertelefoon: topic 9, FOK!: topic 7
- STD testing and condom use - Kindertelefoon: topic 2, FOK! Topic 10
- Porn, sexual performance, libido, and fantasy - Kindertelefoon: topic 11, FOK!: topic 11

The validity of the topic pairings was first chosen based on the results of the topic modeling and then manually evaluated by analyzing the comments within each cluster. Therefore, the pairings made might not seem representative of what is shown in the figure. In figure 13, topic 3 would be more suitable for the topic of 'Shape of genitals'. However, when validating the data, it was apparent that topic 2 is more suitable for the topic of 'Shape of genitals' despite the LDA result as shown in the figure which did not reflect the expected outcome when analyzing the data. A possible explanation is due to the low term frequency, as shown in the graph, the blue colored bars represent the term frequency of each word. The bars are very shallow for both topics 2 and 3, and could possibly lead to less representative results shown in the graph. Thus, a manual validation concluded that FOK! forum's topic 2 would represent the topic of 'Shape of genitals' more accurately.

The results of the sentiment analysis per topic returned the following results (see table 9). Results were obtained by taking the mean of the data. Interestingly, de Kindertelefoon's comments were more negative than FOK! forum on the topics 'Shape of genitals' and 'STD testing and condom use', higher for 'Porn, sexual performance, libido, and fantasy' and equally for 'Birth control, pill and pregnancy'.

Comparing the positive and negative scores only (in tables 10 and 11), a similar pattern occurs as seen in the sentiment analysis of both fora in general (see table 8). For all topics, FOK! forum scores higher in positive sentiment than de Kindertelefoon, the same principle stands for negative sentiment.

	De Kindertelefoon				FOK! forum			
	Sentiment		Subjectivity		Sentiment		Subjectivity	
	M	sd	M	sd	M	sd	M	sd
Shape of genitals	-0.014	0.25	0.5	0.24	0.066	0.27	0.39	0.36
Birth control, pill, and pregnancy	0.047	0.22	0.52	0.26	0.047	0.25	0.42	0.33
STD testing and condom use	-0.015	0.24	0.53	0.32	0.052	0.26	0.43	0.34
Porn, sexual performance, libido, and fantasy	0.15	0.20	0.48	0.34	0.064	0.26	0.44	0.34

Table 9
Sentiment scores mean per topic per forum including all comments the topic modelling method classified.

Positives	**de Kindertelefoon**		**FOK! forum**	
	M	sd	M	sd
Shape of genitals	0.26	0.24	0.31	0.22
Birth control, pill, and pregnancy	0.19	0.16	0.25	0.20
STD testing and condom use	0.21	0.20	0.27	0.21
Porn, sexual performance, libido, and fantasy	0.24	0.16	0.27	0.18

Table 10
Sentiment score means of positive orientation per forum

Negatives	**de Kindertelefoon**		**FOK! forum**	
	M	sd	M	sd
Shape of genitals	-0.18	0.074	-0.27	0.23
Birth control, pill, and pregnancy	-0.17	0.16	-0.24	0.22
STD testing and condom use	-0.21	0.18	-0.25	0.22
Porn, sexual performance, libido, and fantasy	-0.23	0.23	-0.25	0.22

Table 11
Sentiment score means of negatives orientation per forum

Furthermore, a statistical analysis (z-test) was conducted for each topic. Due to an imbalance in the data, random undersampling was performed on FOK! forum in order to get a more representative statistical result with the caveat of information loss.

Results indicate that all but 'STD testing and condom use' was statistically insignificant (see table 12). 'Shape of genitals' ($z=-4.48$, $p>0.05$) was insignificant but received a high Cohen's d ($d=0.37$), indicating a low to medium-high effect size but no difference between the samples. 'Birth control, pill, and pregnancy' was statistically insignificant ($z=0.21$, $p>0.05$) with a low effect size ($d=0.03$), and 'Porn, sexual performance, libido, and fantasy' was also found to be statistically insignificant ($z=4.061$, $p>0.05$) but had a low to medium-high effect size ($d=0.3$).

As mentioned before, 'STD testing and condom use' was the sole topic that has statistically significant results (z=-2.64, p<0.01) with a small effect size of d=0.27. Indicating that statistically there is a small difference between the two fora on the topic of 'STD testing and condom use'.

	de Kindertelefoon		**FOK! forum**		**Statistics**		
	M	sd	M	sd	Z-test (z=)	Z-test (p=)	Cohen's d (d=)
Shape of genitals	-0.015	0.24	0.053	0.26	-4.48	7.29	0.37
Birth control, pill, and pregnancy	0.047	0.22	0.041	0.28	0.21	0.83	0.03
STD testing and condom use	-0.016	0.24	0.053	0.26	-2.64	0.0082	0.27
Porn, sexual performance, libido, and fantasy	0.15	0.20	0.074	0.28	4.061	4.89	0.30

Table 12
Statistical results of conducted z-tests per topic

5. Discussion

The aim of this study was to assess a possible factor (or marker), namely sentiment, that can indicate the credibility of sexual health information shared online. In order to evaluate the marker, we must establish that there is a statistical difference between the sample of a credible and non-credible source. We can do this by using a moderated (on falsehoods) website and a nonmoderated website. Results indicate a slightly higher sentiment for both negative and positive sentiment in the non-moderated and thus less credible forum, whereas the credible forum had a higher number of neutral comments. The difference between the credible and noncredible regarding sentiment was found to be statistically significant (using a z-test), however, the effect size indicates that the difference is trivial in effect size and therefore, the statistically significant difference has no effect. This finding can therefore be regarded as an observed effect that cannot be distinguished from a difference that would appear by chance. A possible explanation for the statistical difference could be due to the large populations in both samples which can cause a statistical test to return significant quickly.

However, the result of the z-test is in line with the majority of literature which demonstrates that a negative correlation can be found between negative sentiment and a source with reduced credibility (Newman et al. 2003; Ott, Cardie, and Hancock 2013), and this study reinforces this by validating the significance of a difference between a credible and non-credible source. Sentiment can be a promising candidate as a variable to consider when creating models that can predict credibility, but the low effect size in this study indicates that it is not a good variable to depend a model on. There is not enough effect size for the difference to be distinctive enough for prediction.

Possible explanations for this result could be the domain we are analyzing. There could be a chance that people online generally talk very neutral about sexual health information and try to give other people objective advice and information. In a study by Dahal, Kumar, and Li, it was found that people on Twitter are very negative regarding politics or extreme weather events. Or perhaps the nature of the community of both fora that have been analyzed are generally similar in behavior. Another possible explanation could be the usage of Pattern.nl for sentiment analysis. Due to the fact that Pattern.nl is a more general-purpose method, a method that is more specific to this domain would have classified the comments more precisely, revealing more underlying hidden sentiment. While conducting the analysis, as mentioned in the result section, many comments were removed due to Pattern not grasping the context. For future

research, it would be interesting to analyze sexual health information using a machine learning model that is specifically tailored to sexual health contexts, as in this domain, people tend to converse using specific jargon about sexual health. Although no such machine learning model exists yet, we have gone over the models that do exist for sentiment analysis in Dutch. BERTje and RobBERT's sentiment analysis modules are trained on training data that is about book reviews. Models for sentiment analysis cannot be applied to every domain, and so far observed, sentiment models in Dutch primarily focus on reviews (e.g. books or restaurants). This is a limitation in the Dutch NLP field caused by most studies within sentiment analysis focusing on models that use the English language, and thus most advancements are made in English. It is also the language that has the most variety of models available, whereas Dutch models lack a good variety of options.

Additionally, the fora that have been analyzed in this study are predominantly aimed at the younger audience, who are prone to use a lot of modern-day slang. Within the sexual health domain, there is also domain-specific slang available. Rule-based approaches are only tailored to general positive or negative words, and not to any specific domain. This may have caused a loss in sentiment detected and influenced the final sentiment scores. There is consistency and reliability in the model knowing general words that indicate sentiment, yet the lack of knowledge in sexual health jargon causes speculation in how well Pattern classified the comments by sentiment. In future research, it is recommended to use a rule-based or machine learning approach that is familiar with Dutch slang and the jargon used within the topic of sexual health.

As for the topic modelling, four topics have been analyzed. Of the four topics, only one topic indicated a statistically significant difference between the credible and non-credible sources and had a small effect size. This topic was 'STD testing and condom use'. When analyzing the overall sentiment per topic, people were found to be mostly neutral for both fora on each topic. Only de Kindertelefoon was more negative on the topic of 'Shape of genitals' and 'STD testing and condom use'. The comments on these topics are likely more negative due to most conversations mentioning worry about certain genital unusualities. Conducting the topic modelling analysis it was found to have a data imbalance. Due to the nature of the LDA topic modelling algorithm, it groups together the most similar threads and returns the most prominent N amount of threads, disregarding threads that do not fit in any frequently found topic. Due to de Kindertelefoon having 5 times the amount of threads that FOK! forum has, the number of comments per thread was thus significantly less for de Kindertelefoon and led to an imbalance of the data after topic modelling. Balancing the data would lead to too much data loss on FOK! forum's part. Although, in this case, it can be argued that it led to a representative view as it is the nature of both fora if one would base their full study on the differences per subtopic it would lead to a challenging limitation when making comparisons. It is therefore advised to find a forum that is similar in activity per thread to conduct a reliable and generalizable analysis of the sentiment per topic. Additionally, as can be observed in the results of the LDA model (figure 9 and 10), the topics do not completely overlap. By looking at the results in the figures and manual observation, topics were chosen based on what is available in both LDA results and could be a good match. Some topics were available in de Kindertelefoon but not FOK! forum and vice versa. These topics were disregarded.

The implications of the results of this study come with a warrant. The results are not supposed to indicate one forum to be more or less positive, and a simple generalization of a forum's community's attitude cannot be made with this study. The main objective is to find whether there is a distinction between the sentiment of a credible and a

more noncredible source. It should be mentioned that despite using FOK! forum as the noncredible variable of this study does not implicate that every information shared on FOK! forum is not credible, but it is considered less credible due to the fact that there is no moderator available that heeds others of misinformation.

Furthermore, research of this nature is important in the interest of those who seek advice and help through online platforms. Identifying key markers for credibility prediction can lead to reduced misinformation available online. As mentioned in the literature, certain knowledge seems to have decreased from years before, e.g. knowledge on STIs (Marra, de Graaf, and Meijer 2020). Being able to automatically tell people certain information online is false, could help people practice more safe sex and take responsible care of their health.

On the other hand, we must also be careful with the possible consequences of creating a model that has all the markers to identify credibility. As mentioned many times before, there is a difference between domains, meaning the sentimental attitude to sexual health might differ vastly from the sentimental attitude towards politics or maths. The sentimental difference in one domain should not hold the standard, and cannot be generalized to all other domains. In other words, the problem of identifying credibility and its markers is not a simple generalizable task. There are many layers of complexity other than domain, there are also differences in platforms, and even different languages.

Another point that must be raised is that creating a model which is able to predict credibility, should not be a watchdog for the entire internet. Many online platforms contain misinformation, sometimes unintentionally. If we were to adopt a rigid control over what can or cannot be posted based on its credibility, the attitude of online users may also change drastically. In a scenario where a good model is being created and deployed on e.g. a forum with the consequence of uncredible comments getting deleted, it might create unethical censorship and this strips away one's freedom of speech. This also raises the question, what if someone is asking for validation on something that is not credible and this gets flagged as not credible? This would be an unfair consequence. Not to mention, if a credibility method or model is used rigidly, people may get comfortable with the idea of "everything they read online must be correct" and could have the repercussion that people become more gullible (e.g. fall for scams) considering they do not need to be critical anymore. Another point is that there is also (fan)fiction that can be found on the fora that have been analyzed in this study, where people are aware that it is not real. What happens to them? Instead, the purpose of a model or method that can predict credibility should be a tool and an aid for users online to be critical of the content they get presented in front of them.

6. Conclusion

This study aims to assess the extent of the difference in sentiment on sexual health topics between Dutch moderated and non-moderated fora, in order to evaluate its potential as a marker for credibility prediction. This research was conducted using the moderated forum, de Kindertelefoon, and the non-moderated forum: Fok! forum. Based on a qualitative analysis of the sentiment between the two fora, it can be concluded that there is a significant difference between the moderated and non-moderated sources. Namely, the non-moderated forum has both a higher negative and positive sentiment than the moderated forum whereas the moderated forum exceeds the non-moderated forum in neutral sentiment. The outcome, therefore, is in line with the majority of research done on credibility and sentiment and therefore does support the findings in the majority of

previous literature in which results eluded there to be a correlation between sentiment and credibility. However, this finding is contradicted by the statistical analysis. The z-test indicated a statistically significant difference between the sample of the moderated and non-moderated sources, but the effect size indicates this significance to be trivial and thus the statistical difference is found to be merely random. However, research using more tailored models should be conducted to confirm the validity of this research as well as an in-depth analysis of differences in other platforms, domains, and languages.

References

Aue, Anthony and Michael Gamon. 2005. Customizing sentiment classifiers to new domains: A case study. In *Proceedings of recent advances in natural language processing (RANLP)*, volume 1, pages 2–1, Citeseer.

Bergmark, Karin Helmersson, Anders Bergmark, and Olle Findahl. 2011. Extensive internet involvement—addiction or emerging lifestyle? *International journal of environmental research and public health*, 8(12):4488–4501.

Borzekowski, Dina LG and Vaughn I Rickert. 2001. Adolescent cybersurfing for health information: a new resource that crosses barriers. *Archives of pediatrics & adolescent medicine*, 155(7):813–817.

Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684.

CBS. Wat voeren jongeren uit online?

Crocamo, Cristina, Marco Viviani, Lorenzo Famiglini, Francesco Bartoli, Gabriella Pasi, and Giuseppe Carrà. 2021. Surveilling covid-19 emotional contagion on twitter by sentiment analysis. *European Psychiatry*, 64(1).

Dahal, Biraj, Sathish AP Kumar, and Zhenlong Li. 2019. Topic modeling and sentiment analysis of global climate change tweets. *Social network analysis and mining*, 9(1):1–20.

De Smedt, Tom and Walter Daelemans. 2012. Pattern for python. *The Journal of Machine Learning Research*, 13(1):2063–2067.

Delobelle, Pieter, Thomas Winters, and Bettina Berendt. 2020. Robbert: a dutch roberta-based language model. *arXiv preprint arXiv:2001.06286*.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Gatti, Lorenzo and Judith van Stegeren. 2020. Improving dutch sentiment analysis in pattern. *Computational linguistics in the Netherlands journal*, 10:73–89.

de Graaf, Hanneke, Marieke van den Borne, Sanne Nikkelen, Denise Twisk, and Suzanne Meijer. 2017. Seksuele gezondheid van jongeren in nederland anno 2017. *Delft, The Netherlands: Rutgers and Soa Aids Nederland*.

Gray, Nicola J, Jonathan D Klein, Judith A Cantrill, and Peter R Noyce. 2002. Adolescent girls' use of the internet for health information: issues beyond access. *Journal of medical systems*, 26(6):545–553.

Greene, Ciara M and Gillian Murphy. 2021. Quantifying the effects of fake news on behavior: Evidence from a study of covid-19 misinformation. *Journal of Experimental Psychology: Applied*, 27(4):773.

Gundapu, Sunil and Radhika Mamidi. 2021. Transformer based automatic covid-19 fake news detection system. *arXiv preprint arXiv:2101.00180*.

Hagen, Loni. 2018. Content analysis of e-petitions with topic modeling: How to train and evaluate lda models? *Information Processing & Management*, 54(6):1292–1307.

Hu, Xia, Jiliang Tang, Huiji Gao, and Huan Liu. 2014. Social spammer detection with sentiment information. In *2014 IEEE international conference on data mining*, pages 180–189, IEEE.

Kakol, Michal, Radoslaw Nielek, and Adam Wierzbicki. 2017. Understanding and predicting web content credibility using the content credibility corpus. *Information Processing & Management*, 53(5):1043–1061.

Kwon, Sejeong, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. In *2013 IEEE 13th international conference on data mining*, pages 1103–1108, IEEE.

Marra, Elske, Hanneke de Graaf, and Suzanne Meijer. 2020. Seks onder je 25e in de residentiële jeugdzorg.

McGlynn, Joseph, Maxim Baryshevtsev, and Zane A Dayton. 2020. Misinformation more likely to use non-specific authority references: Twitter analysis of two covid-19 myths. *Harvard Kennedy School Misinformation Review*, 1(3).

Meijer, Suzanne. 2019. Interviewvragen seksualiteit bij jongeren. *Bijblijven*, 35(5):33–38.

Newman, Matthew L, James W Pennebaker, Diane S Berry, and Jane M Richards. 2003. Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin*, 29(5):665–675.

Ott, Myle, Claire Cardie, and Jeffrey T Hancock. 2013. Negative deceptive opinion spam. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies*, pages 497–501.

Pennycook, Gordon and David G Rand. 2021. The psychology of fake news. *Trends in cognitive sciences*, 25(5):388–402.

Ramage, Daniel, David Hall, Ramesh Nallapati, and Christopher D Manning. 2009. Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 conference on empirical methods in natural language processing*, pages 248–256.

Richardson, Leonard. 2007. Beautiful soup documentation. *Dosegljivo: https://www. crummy. com/software/BeautifulSoup/bs4/doc/.[Dostopano: 7. 7. 2018]*.

Rietvelt, DCJC. 2019. Influence of neutral word removal on sentiment analysis. B.S. thesis, University of Twente.

Singh, Bhuvanesh and Dilip Kumar Sharma. 2021. Predicting image credibility in fake news over social media using multi-modal approach. *Neural Computing and Applications*, pages 1–15.

Suárez, Pedro Javier Ortiz, Benoît Sagot, and Laurent Romary. 2019. Asynchronous pipeline for processing huge corpora on medium to low resource infrastructures. In *7th Workshop on the Challenges in the Management of Large Corpora (CMLC-7)*, Leibniz-Institut für Deutsche Sprache.

Tong, Zhou and Haiyi Zhang. 2016. A text mining research based on lda topic modelling. In *International conference on computer science, engineering and information technology*, pages 201–210.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

de Vries, Wietse, Andreas van Cranenburgh, Arianna Bisazza, Tommaso Caselli, Gertjan van Noord, and Malvina Nissim. 2019. Bertje: A dutch bert model. *arXiv preprint arXiv:1912.09582*.

Wang, Zheng, Feiping Nie, Lai Tian, Rong Wang, and Xuelong Li. 2020. Discriminative feature selection via a structured sparse subspace learning module. In *IJCAI*, pages 3009–3015.

Zhang, Lei, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1253.

Zhou, Xinyi, Apurva Mulay, Emilio Ferrara, and Reza Zafarani. 2020. Recovery: A multimodal repository for covid-19 news credibility research. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 3205–3212.

Appendix A: Datasets overview

	TopicID	CommentID	Comment	Time	sentiment_score	subjectivity_score
0	2510437	188320433	Deel VI (6)!!! Het is ons gelukt! Ruim zeven j...	dinsdag 6 augustus 2019 @ 20:50:11	-0.004167	0.816667
1	2510437	188320540	Tvp	dinsdag 6 augustus 2019 @ 20:53:24	0.000000	0.000000
2	2510437	188320584	Thuisfeestjes ligt ons niet, zelfde geldt dus ...	dinsdag 6 augustus 2019 @ 20:54:45	0.291667	0.709524
3	2510437	188320621	Vers Topic!	dinsdag 6 augustus 2019 @ 20:55:47	0.000000	0.100000
4	2510437	188320700	en dit dus , gewoon grote club veel mensen , r...	dinsdag 6 augustus 2019 @ 20:58:56	0.244444	0.633333
...	...	...	...	...	...	...
116957	2294598	161234941	Neem hem maar niet serieus, volgens mij is hij...	woensdag 6 april 2016 @ 21:23:35	-0.150000	0.600000
116958	2294598	161234954	admiraal_anaal.	woensdag 6 april 2016 @ 21:23:40	0.000000	0.000000
116959	2294598	161234975	Ik was serieus Wat zou je doen als je dochter ...	woensdag 6 april 2016 @ 21:23:59	0.108333	0.825000
116960	2294598	161234999	wat ken je me goed!	woensdag 6 april 2016 @ 21:24:13	0.687500	0.900000
116961	2294598	161235023	Goed, dit heeft geen zin. Als TS een beetje on...	woensdag 6 april 2016 @ 21:24:29	0.037500	0.825000

113526 rows × 6 columns

Figure 1
FOK! Forum comments dataset overview

	TopicID	CommentID	Content	CreateTime	FirstPost
0	424171	1000000	Hoi Ik ben een jongen van 14 jaar oud en wil...	2021-05-30	True
1	424171	1000001	Als jij daar gelukkig van wordt moet je het ...	2021-05-30	False
2	424171	1000002	Ik heb precies hetzelfde, het is dus totaal ...	2021-05-30	False
3	424171	1000003	Bij de decatlon kun je zelf afrekenen. Bij e...	2021-05-30	False
4	424171	1000004	Heyy Ik vind er niks raars aan hoor. Er zi...	2021-05-30	False
...	...	...	...	...	...
87701	103152	1087701	Nee, hier heb ik wat informatie voor je opgezo...	2014-01-12	False
87702	103152	1087702	Nounou helptaltijdx, wat een research, knap ho...	2014-01-14	False
87703	103152	1087703	Het is niet pedo, maar ik zou deze dame laten ...	2014-01-15	False
87704	103152	1087704	ik vind van niet. mijn buurman en buurvrouw he...	2014-01-16	False
87705	103152	1087705	als je 40 bent boeit dat toch ook niet meer ik...	2014-12-12	False

87706 rows × 5 columns

Figure 2
de Kindertelefoon comments dataset overview

	ID	Title	TopicCat	LastResponse	NComments	NViews	URL
0	2510437	[TVT] Parenclub: Done or not done? VI	Centrale	2021-05-19 17:05:25	127	41798	https://forum.fok.nl/topic/2510437/parenclub-d...
1	1834491	[TVT] Erotische verhalen en ervaringen #4	Centrale	2021-04-29 18:14:22	251	395816	https://forum.fok.nl/topic/1834491/erotische-v...
2	1466693	[TVT] Intieme Piercing IV	Centrale	2021-04-12 10:01:14	66	52247	https://forum.fok.nl/topic/1466693/intieme-pie...
3	918389	[TVT] Hoe was jouw eerste keer?	Centrale	2021-03-29 19:21:17	277	279549	https://forum.fok.nl/topic/918389/hoe-was-jouw...
4	1391746	[TVT] Te kort toompje - Te strakke voorhuid I ...	Centrale	2020-12-22 22:44:50	135	120426	https://forum.fok.nl/topic/1391746/te-kort-too...
...	...	...	...	...	...	...	...
2048	2298364	anale kriebels	apr 2016 - jun 2016	2016-04-25 16:18:58	24	828	https://forum.fok.nl/topic/2298364/anale-krieb...
2049	2296452	Masturberen	apr 2016 - jun 2016	2016-04-16 15:29:07	9	784	https://forum.fok.nl/topic/2296452/masturberen...
2050	2296421	Website om direct een neukpartner te vinden	apr 2016 - jun 2016	2016-04-16 11:37:43	7	685	https://forum.fok.nl/topic/2296421/website-om-...
2051	2295550	Meid opgewonden van stiefmoeder.....	apr 2016 - jun 2016	2016-04-11 21:19:47	16	848	https://forum.fok.nl/topic/2295550/meid-opgewo...
2052	2294598	Leeftijdsverschil	apr 2016 - jun 2016	2016-04-06 21:24:29	49	1444	https://forum.fok.nl/topic/2294598/leeftijdsve...

2053 rows × 7 columns

Figure 3
FOK! Forum threads dataset overview

	ID	Title	CreationTime	LastReplyTime	NReplies	NViews	NLikes	URL
0	424171	badpak als jongen	2021-05-29T23:15:28+0000	2021-05-30T08:13:40+0000	3	34	0	https://forum.kindertelefoon.nl/seksualiteit-3...
1	315613	vragenlijst voor meisjes en jongens	2019-07-03T10:52:09+0000	2021-05-30T07:34:30+0000	392	33640	4	https://forum.kindertelefoon.nl/seksualiteit-3...
2	424164	ik trek heel veel af	2021-05-29T21:32:03+0000	2021-05-30T06:38:24+0000	4	61	0	https://forum.kindertelefoon.nl/seksualiteit-3...
3	424172	Sex	2021-05-29T23:28:24+0000	2021-05-29T23:41:18+0000	1	33	0	https://forum.kindertelefoon.nl/seksualiteit-3...
4	423971	gedragen slipjes van mijn zus.	2021-05-28T13:18:28+0000	2021-05-29T22:51:01+0000	16	348	1	https://forum.kindertelefoon.nl/seksualiteit-3...
...	...	...	...	...	...	...	...	...
10942	161874	Te grote schaamlippen ?	2015-05-29T10:10:57+0000	2015-06-01T14:06:13+0000	7	550	0	https://forum.kindertelefoon.nl/seksualiteit-3...
10943	161856	Hoe ver zou jij gaan? Voor de meiden.	2015-05-28T22:06:26+0000	2015-06-01T08:46:01+0000	5	973	0	https://forum.kindertelefoon.nl/seksualiteit-3...
10944	112955	Zwanger op je 14e?	2014-03-02T09:00:39+0000	2015-04-09T09:05:56+0000	16	3434	0	https://forum.kindertelefoon.nl/seksualiteit-3...
10945	144565	Seks op de wc	2014-12-16T20:44:00+0000	2015-01-01T23:14:44+0000	6	3077	0	https://forum.kindertelefoon.nl/seksualiteit-3...
10946	103152	Is dit pedo?	2013-12-07T10:12:22+0000	2014-12-12T02:36:12+0000	9	1982	0	https://forum.kindertelefoon.nl/seksualiteit-3...

10947 rows × 8 columns

Figure 4
de Kindertelefoon threads dataset overview