UNDERSTANDING NEGATIVE SENTIMENT TOWARD FACEBOOK BY TECHIES: A QUALITATIVE & QUANTITATIVE ANALYSIS VIA USER GENERATED CONTENT

A Thesis
Presented to
The Academic Faculty

By

Abu Saleh Md Noman

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Informatics, Computing, and Engineering

Indiana University

December 2019

UNDERSTANDING NEGATIVE SENTIMENT TOWARD FACEBOOK BY TECHIES: A QUALITATIVE & QUANTITATIVE ANALYSIS VIA USER GENERATED CONTENT

Approved by:

Dr. Sameer Patil, Advisor
Luddy School of Informatics, Computing, and Engineering
Indiana University

Dr. Apu Kapadia
Luddy School of Informatics, Computing, and Engineering
Indiana University

Date Approved:

ACKNOWLEDGEMENTS

First and foremost I would like to thank my primary thesis supervisor Dr. Sameer Patil for his continuous help and support. I am also grateful to the members of my committee for their patience and support in overcoming numerous obstacles I have been facing through my research.

I would like to thank my fellow doctoral student Dr. Sanchari Das for her feedback, help in coding process and of course friendship. I am also grateful to Cavin D'souza and all undergraduate researcher who worked under my supervision for data collection related helps. I also acknowledge the long and tedious manual coding work of Emma Lashley with patience. Lastly, I would thank Emily Swiatek, Morgan Brockman, and Vivi Armacost for their patient proofreading of my thesis.

I am thankful to my wife Silvia Karim and my son Sabeer Arfaz Noman for their unconditional support throughout this journey. It would not have been possible without their love and struggle. Last but not the least, I would like to thank my parents and sister for supporting me spiritually throughout writing this thesis and my my life in general.

TABLE OF CONTENTS

Acknowledgments v

List of Tables x

List of Figures xi

Chapter 1: Introduction 1

1.1 Problem Statement and Motivation 1

1.2 Research Questions 2

1.3 Overview of Approach 3

1.4 Contributions 4

1.5 Thesis Overview 5

Chapter 2: Related Works 6

2.1 CMC, Blogs, Slashdot 6

2.2 HCI and Social Media Usage Practice 7

2.3 Technology Non-use 7

2.4 Negative Consequences of Facebook Use 9

2.5 Sentiment Analysis, Topic Modeling, and Text Classification on Social Media Data 10

Chapter 3: Overview of the Qualitative Analysis 13
3.1 Data Collection 13
3.1.1 Slashdot 13
3.1.2 Schneier on Security 15
3.2 Comment Selection 17
3.3 Coding 18
3.3.1 Sentiment Coding 18
3.3.2 Thematic Coding 19

Chapter 4: Overview of the Automated Analysis 22
4.1 Wordcloud and Topic Modeling 22
4.2 Sentiment Analysis 24
4.3 Text Classification 24
4.3.1 Data Pre-processing 25
4.3.2 Feature Engineering 26
4.3.3 Model Building 28

Chapter 5: Results of Thematic Analysis 33
5.1 Privacy and Security 33
5.2 User Experience 36
5.3 Advertisements 37
5.4 Psychosocial Well-being 38
5.5 Personal Disposition 39
5.6 Alternate Options 40

5.7 Uninteresting Content 41

5.8 Politics 43

5.9 Fake Accounts and Bots 43

5.10 Miscellaneous 45

5.11 Explicit Non-use 46

Chapter 6: Results of Automated Analysis 50

6.1 Wordcloud 50

6.2 Topic Modeling 51

6.3 Sentiment Analysis 52

6.4 Text Classification 55

6.4.1 Evaluation Measures 55

6.4.2 Results Analysis 57

Chapter 7: Discussion 65

7.1 Qualitative Coding 65

7.1.1 Comparison between Slashdot and Schneier's Blog 67

7.1.2 Explicit Non-use Only Comments Analysis 68

7.2 Results of the Text Classifier 71

7.3 Implications of the Findings 73

Chapter 8: Conclusion 75

8.1 Limitations 75

8.2 Future Work 76

8.3 Concluding Remarks 76

References 91

LIST OF TABLES

3.1 Categories assigned by two independent raters on 4,000 Slashdot and 1000 Schneier's blog comments. (NS = related to *negative sentiment*, NRNS = not Related to *negative sentiment*) 19

4.1 Dataset used for different automated analysis. 22

5.1 Results of thematic coding showing the comments with negative sentiment toward Facebook falling under each theme for each of the two data sources along with the fraction of those comments that explicitly indicated Facebook non-use. Percentages for each column do not add to 100% because the themes are not mutually exclusive. 34

6.1 LDA topic modeling on the dataset. 52

6.2 NMF topic modeling on the dataset. 53

6.3 Average performances of different classifier on different features (cleaned data). P = Precision, R = Recall, AUC = Area Under the Curve. 59

6.4 Average performances of different classifier on different features (uncleaned data). P = Precision, R = Recall, AUC = Area Under the Curve. 61

7.1 List of common themes with prior work. 69

LIST OF FIGURES

3.1 Flow diagram of qualitative analysis 14

3.2 Typical Slashdot comment thread 15

3.3 Typical Schneier of Security blog post 16

4.1 Flow diagram of Text Classification task 25

4.2 SVM classifier 31

5.1 Sample word tree of comments starting with “quit.” 48

5.2 Sample word tree of comments starting with “stop.” 49

6.1 Visualization of 2,364 *negative sentiment* comments. 50

6.2 Box plots of compound sentiment scores of each theme using VADER tool. (OA = Overall, UE = User Experience, PD = Personal Disposition, AD = Advertisements, PO = Politics, FB = Fake Accounts & Bots, WB = Psychosocial Well-being, AO = Alternate Options, UC = Uninteresting Content, MI = Miscellaneous, PS = Privacy & Security) 54

6.3 Confusion matrix. TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives. 55

6.4 Confusion Matrix for Logistic Regression based on character level TF-IDF (cleaned data). Note that, NS = related to negative sentiment and NRNS = not related to negative sentiment. 60

6.5 Precision-Recall curve for Logistic Regression based on character level TF-IDF (cleaned data) 62

6.6 Confusion Matrix for Logistic Regression based on character level TF-IDF (uncleaned data). Note that, NS = related to negative sentiment and NRNS = not related to negative sentiment. 63

6.7 Precision-Recall curve for Logistic Regression based on character level TF-IDF (uncleaned data) 64

7.1 Number of comments per theme by year. 70

SUMMARY

Owing to its exponential rise in popularity and adoption, social media usage has received extensive research attention in the past decade. Recently, however, researchers have started to recognize the need to pay attention to deliberate *non-use* and non-adoption in order to uncover important shortcomings and usage barriers in these systems. We contribute to these efforts by investigating user-generated comments on posts related to Facebook on Slashdot and Schneier on Security, two news blogs with a tech-savvy audience. We found that nearly half (47.28%) of 5,000 randomly selected reader comments indicated some form of negative sentiment. A deeper look at these comments via qualitative coding and automated natural language analyses revealed Facebook system architecture and privacy concerns as key factors. A machine learning based text classification approach for automatically identifying such negative sentiment and non-use related comments achieved 77.6% area under Precision-Recall curve. Based on the findings, we derive implications regarding how expert users can surface important usage aspects that could potentially be informative and transformative in shaping usage practices and preferences of non-experts.

CHAPTER 1
INTRODUCTION

1.1 Problem Statement and Motivation

Emergence of technology and continuous integration of internet based services in everyday life has introduced the era of “Big Data” [1]. While such advancement has flourished user experience in many aspects, it has engendered other user related issues as well, such as privacy concerns [2], data breaches, etc. Such issues are important to take into account while designing a secure and reliable system. For example, privacy concerns can lead users to reduce and even abandon further engagement with the system [3] which makes this a burning question for system developers, UX designers, and advertisers alike. The proliferation of social media sites and corresponding user interaction thus invoke issues that might lead to user dissatisfaction and lack of usage. Factors that contribute to such user negative sentiment and non-use can threaten the sustainability and success of these systems. Therefore, a comprehensive understanding of non-usage from technical and sentimental points of view is required for building a successful ecosystem of user base.

Since their inception in the early 2000s, social network sites have transformed everyday interaction practices. Today, the most popular of these sites, Facebook, claims more than 2 billion global active users.[1] Such explosive growth and popularity has resulted in a great deal of research attention towards the use and positive impacts of social media in general, and Facebook in particular. For instance, it has been shown that the use of Facebook can be associated with a number of benefits, such as enhancing social connectedness, increasing social capital, and boosting self esteem [4, 5]. However, more recent research indicates that increased use of Facebook may also lead to a number of negative effects, including addic-

[1] `https://dustinstout.com/social-media-statistics/#facebook-stats`

tion, feelings of jealousy, depression, decreased well-being, invasion of privacy, reduced work productivity, cyberbullying, etc. As a consequence, people have reported efforts to reduce their use of social media via tactics such as taking a "vacation" from Facebook or deleting their accounts. For instance, the recent Facebook scandal that exposed large-scale harvesting of its user data by the British firm Cambridge Analytica resulted in the trending hashtag #DELETEFACEBOOK.

Although prior research focused on adoption of technologies and various benefits of usage [6], little has been studied on those who reduce the usage, refrain from using, or totally abandon a system. Few studies have attempted to address this issue by explicit questioning of regular users [7, 8]. Although these studies have yielded important and useful insight, they suffer from two notable limitations. First, the average person in these samples may not possess adequate technical knowledge to understand and notice the underlying operational mechanisms of social networking services that are often a major contributing factor to the various features perceived negatively, such as behavioral tracking, personalized advertising, third party data sharing, etc. Second, participants in survey studies may refrain from expressing negative opinions to the fullest extent as they might when sharing their views in a more naturalistic and unprompted setting.

1.2 Research Questions

To address the shortcomings mentioned in previous section, we investigated negative sentiment toward Facebook by analyzing user generated comments produced by technically savvy individuals in two technology oriented blogs. From a wide range of choices we picked Slashdot[2] and Schneier on Security.[3] Specifically, this thesis aims to address the following research question:

RQ1: What factors lead technically savvy individuals to hold negative views toward social networking sites?

[2] https://slashdot.org/
[3] https://www.schneier.com/

The other goal of this thesis is to augment the findings of manual qualitative approach with automated analysis and to provide a generalized quantitative result in regards to negative sentiments. It is important to note here that, in the presence of a sufficiently large dataset, such manual analysis can be tedious and cumbersome. In order to extend the analysis to a larger scale (for example to do a temporal analysis of the factors) and to reduce human efforts to classify negative sentiment related responses we also wanted to address the following research question:

RQ2: Can we automatically identify negative sentiment related comments to posts in the blogs using supervised machine learning and natural language processing based techniques?

1.3 Overview of Approach

To answer these research questions we chose to focus on Facebook because of its popularity and size of the userbase. Comments of users were scraped and crawled from two technology focused blogs (Slashdot and Schneier on Security) on Facebook related posts. Slashdot ("News for Nerds. Stuff that Matters," as they say) is a technology-related news website, the summaries of stories and links to news articles are submitted by its own readers, and each story becomes the topic of a threaded discussion among users. Unlike the general technology coverage of Slashdot, the posts on Schneier on Security are limited to matters of security and privacy pertaining to technology. In contrast to the reader-submitted posts of Slashdot, each post on Schneier on Security is created and curated by the renowned security expert Bruce Schneier. We included Schneier on Security due to the blog's focus on security and privacy which often feature as prominent aspects in research and news regarding Facebook. The additional source also allowed us to confirm that our insight applies more broadly beyond a single site. It is safe to assume these users are technology geeks, and we get spontaneous unprompted content from their comments.

We used qualitative sentiment coding to determine the comments related to negative sentiment toward Facebook (4000 comments from Slashdot and 1000 from Schneier on Security). Two coders individually coded the randomly selected scraped comments. The method of randomization was used to avoid any form of biases which can occur while collecting the data during a particular time-period. Later, on the comments coded as *related to negative sentiment* by either one of the coders, we did extensive qualitative coding to probe more on the underlying factors. In this thesis, we investigated what reasons (privacy and security concerns, socio-economic and individual issues etc.) lead to negative sentiment and eventually to lurking, resisting, or dropping out of Facebook. We observed that aspects related to Facebook, privacy concerns, and personal preference were 3 of the 10 key factors responsible for non-use. The findings confirm several previous factors related to Facebook negative sentiment and unveil some new aspects that can be caused by evolution of Facebook as a company or other societal developments.

Later we adopted Natural Language Processing (NLP) based techniques for sentiment analysis, topic modeling, and wordcloud generation on the negative sentiment and non-use comments. For the text classification task we experimented with various classifiers after vectorizing the comments into different set of features. Our analysis suggested that Logistic Regression classifier with character level TF-IDF (Term Frequency-Inverse Document Frequency) achieved an *F-score* of 71% with performance of 77.6% area under the Precision-Recall curve. A shallow qualitative error analysis suggested that adequate training data and nuanced semantic representation of the input text can further improve the performances. Such analysis will enable us to investigate the temporal pattern of the evolution of the underlying themes of non-use practices on a larger scale. This gives an exciting new dimension to this work since no prior work to our knowledge has extended the study on non-use and negative sentiment toward social media system to this extent.

1.4 Contributions

The contribution of this thesis is multifaceted:

- We extend the Facebook non-use related discourse to include negative sentiment as higher level aspect connected to different non-use practices.

- We confirm the previously reported non-use practices although the nature and experiences of the platform has evolved. Moreover, there are additional drivers of negative sentiment that we discovered.

- We target technically savvy population and show that these individuals may exhibit larger amounts of negative sentiment toward Facebook compared to the average person.

- We show that analysis of high quality user generated content can serve as a useful method for surfacing user sentiment and studying user experiences and behaviors on a large scale.

- We confirm the fact that automatic identification of such negative sentiment related reviews is possible and can be further improved. Such implementation enables us to scale up the analysis in the presence of a larger dataset.

1.5 Thesis Overview

In the chapters that follow, we first situate our work based on prior work. Next, we describe how we collected, selected, and coded the user generated content on the selected sites. We proceed to present the findings of our analyses followed by a discussion of that unpacks the insight. We further highlight practical implications of the insight and point out important limitations. At the end, we suggest a few future directions for extending our work and offer concluding remarks.

CHAPTER 2
RELATED WORKS

Our study of relevant literature comprised of Computer-Mediated Communication (instant messaging, forums, etc.), blogs like Slashdot in particular, technology use, and non-use. In the sections to follow we will elaborate prior works on our chosen data sources, social media usage practice, technology non-use, negative sentiments related to Facebook use, and different natural language processing based approaches on social media data.

2.1 CMC, Blogs, Slashdot

Computer-mediated communication of different forms has been studied by researchers for various purposes [9]. The psychology and social aspects of using CMC was discussed by Kieslar et al. [10]. Some argued that nature of internet use and CMC varies by countries [11] while some studies [12] showed online discourse and flaming are by-products of CMC. As a major form of CMC different blogs and forums like Slashdot appeared and quickly became popular for their wide range of applications, such as education [13, 14, 15], politics [16, 17], participation study [18], technology [19], etc. So current research on Blogosphere is promising as researchers continue to study topics like content delivery, conversation analysis, ways to improve social media research through blogs [20, 21, 22]. One study [23] on whether blogs are echo chambers or not was particularly interesting because it adopted both empirical and algorithmic approach on comments from 33 of world's top blogs.

Slashdot as a technology news site has gained popularity and attracted researchers from different fields. Kunegis et al. [24] analyzed Slashdot zoo corpus by representing it as a graph containing positive and negative endorsements. Statistical methods show Slashdot discussion threads have strong heterogeneity and self-similarity and thus invoke less controversial topics [25]. Slashdot, as a public sphere was analyzed for understanding the mechanism of other similar environments [26, 27]. Quantitative analysis on Slashdot as a network and its user comments [25, 28] is not well explored yet. Lampe et al. [29] tried to analyze how users filter out comments leveraging the comment rating feature of Slashdot. Characteristics of sample posters, typical post distribution, and topic categories in Slashdot were studied by Halavais [30]. Unlike Slashdot, Schneier on Security has not received similar research attention but it shares several of Slashdot's characteristics, such as technically savvy target audience, moderated comments, and open discussion. Unlike the general coverage of technology related matters on Slashdot, Schneier on Security focuses exclusively on privacy and security matters, which can provide a useful topic focus given that privacy and security concerns are prominently mentioned in relation to the Facebook platform.

2.2 HCI and Social Media Usage Practice

Previous works in HCI addressed both usage practices [31] and non-use. Lampe et al. [32] described the anticipated audience of Facebook users and how their Facebook usage is shaped. Homophily in location surveillance and friendship network were studied by Guha and Wicker [33]. Lampe et al. [34] argued why concerns about privacy, context collapse, limited time, and channel effects are some of the reasons for not joining Facebook. The change of perception and interaction with Facebook over the course of time was studied by Lampe et al. [35] by analyzing interview data of three different years. They showed Facebook use remains constant in this time period, however, the privacy concern changes. Baumer et al. [36] introduced the Delphi method as measure of understanding the sentiment of social media use or non-use.

2.3 Technology Non-use

The study and approach of human relationship with technology is required to develop a stringent understanding of use and non-use [37]. A lot of previous work addressed lurking and non-usage [38, 39, 40] on online forums and mailing lists, however, they are now obsolete because of changes in technology and social aspects [8]. It is evident from these studies that non-use can't be strictly defined rather it is a complex phenomena and can take multiple forms [41]. According to Mason [42], lurkers do not feel competent enough to post on social media. This requires extensive study (both quantitative and psychological) on non-usage to improve community experience [43, 44]. Researchers have studied the necessity to understand how interventions can influence the known determinants of IT adoption and use. Baumer et al. [7] contribute to developing an understanding of the sociological processes of determining what technologies are (in)appropriate and in which contexts. Their qualitative study focuses on those users who left Facebook and investigates the sociological process of non-use. Comparison of users and non-users, inhibiting factors of social media use was analyzed by some work [45, 46, 47]. Another work [48] focused on refusing, limiting, and departing from technology and associated study such as theories, methods, foundational texts, and central research questions in the study of non-use. Madden [3] analyzed different privacy management on social media sites and concluded that users choose restricted privacy setting and unfriending people is increasing. Their analysis unpacked some of the aspects of user behavior related to non-use. Social media reversion, where users engage in deliberate non-use but later decide to continue using was studied by Baumer et al. [49]. They examined survey responses by people who left Facebook for 99 days but returned sooner. They showed some significant factors affecting the reversion process, such as prior use of Facebook, privacy and surveillance, use of other social media, etc.

For better understanding of non-use and negative sentiment, different strategies have been adopted so far, such as questionnaires [50, 51], interviews[52] and surveys [49, 53]. Manual analysis can be tedious for large dataset, and to tackle this problem researchers have been studying how statistical machine learning based approaches such as topic modeling can be incorporated to the existing grounded theory [54]. Another aspect of non-use study is non-users' perception of social media in developing countries. Wyche and Baumer [55] interviewed non-users living in rural Zambia and discovered what they think of Facebook, what they think are the benefits of joining and what are the inhibiting factors for their non-use. Satchell and Dourish [56] defined varieties of non-use, such as *lagging adoption*, *active resistance*, *disenchantment*, *disenfranchisement*, *displacement* and *disinterest*. They argued HCI mostly studies "use" rather than non-use because it is more visible and showed scenarios where study of non-use might be useful.

The study of non-use is not only limited to social media rather it has been more generic, such as (non)use of internet and information & communication technologies (ICT) in everyday life [57, 58, 59, 60]. By interviewing 1001 adults, Selwyn et al. [57] tried to answer the questions such as, who is using and not using internet in everyday life and how non-users can be encouraged to make use of internet. In addition to this study, the authors also examined [58] the established discourse of non-use of ICT, need for reformulation of factors for non-use and proposed a novel multi-layered model to address this issue. Different frameworks has been proposed so far for understanding technology use discontinuation [59, 60, 61, 62, 63, 64, 65]. These may be political [60, 61], personality factors [62, 63, 64], functionality-need mismatch [65], etc. Based on strategies of segmentation and differentiation, a new policy framework is suggested by Verdegem and Verhoest [59] to increase the ICT acceptance among those who are currently non-users.

2.4 Negative Consequences of Facebook Use

Prior research has extensively focused on the positive aspects of Facebook use. For example, Facebook use has been linked to a variety of benefits such as acquiring social capital [5], increasing well-being [66], finding employment [67], receiving social support during stressful events [68], and so on.

However, in recent years researchers have noticed growing set of concerns among the users with different aspect of their interaction with the system. Apart from privacy concerns that have been an issue from the early days of Facebook, it has been linked with several other issues, such as envy [69], depression [70], cyberstalking [71], cyberbullying [72], etc. In fact, Islam and Patil [73] suggest that there may be a threshold beyond which the amount of benefit derived from Facebook use begins to reverse, and Wisniewski et al. [74] show that not getting adequate privacy engenders a reduction in usage by users.

In the past couple of years, the Facebook platform has been implicated in the spread of targeted misinformation, especially pertaining to important political events, such as the 2016 Presidential election in the US [75] and the Brexit [76] referendum vote in the UK. Throughout 2018 Facebook was criticized because of third party data sharing and increased amount of surveillance [77], leading the US Congress to invite testimony on these matters from Facebook CEO Mark Zuckerberg [78]. As a result of these developments, the general public seems to be becoming increasingly aware of the negative aspects of using the Facebook platform [79]. A recent development of #DELETEFACEBOOK hashtag and the growing popularity of different anti-Facebook platforms also bear evidence to that.

Since most of the prior works were conducted before these events took place, the associated negative sentiment of user base was not addressed. In our work we attempt to fill up that void and try to validate if previous negative sentiments still exist. Moreover, the Facebook platform as well as the demographic characteristics and technical literacy of its user base are continually evolving. It should be interesting to examine if such evolution has contributed to solve the previously reported issues. To that end, we compare our findings those from previous studies of Facebook non-use.

2.5 Sentiment Analysis, Topic Modeling, and Text Classification on Social Media Data

Sentiment analysis and topic modeling on user generated content in naturalistic reviews are well researched topics. In particular, sentiment analysis in blogs [80] and social media [81, 82, 83] is very common to understand user satisfaction, and it has wide range of applications, such as product review [84, 85, 86, 87], opinion mining [88, 89], politics [90], financial marketing [91], etc. LDA Topic modeling is another well studied area for social science and NLP researchers as it is very good at finding inherent topics from given corpus. We only focused on the papers that covered topic modeling on social media dataset [92, 93], blogs in particular [94, 95].

Text classification on social media data like Reddit, Twitter has been previously studied to detect helpful/thoughtful comments, cyberbullying & hate speech, analyze sentiments, etc. Most of these machine learning approaches are either supervised (based on labelled/annotated corpus) or semi-supervised (relatively small set of labelled data compared to unlabelled instances). Van Hee et al. [96] performed a fine-grained annotation of cyberbullying corpus to train a linear support vector machine which yielded an F_1 score of 64$\%$. In recent years there has been increased popularity in content [97] and network based features [98] such as syntatic and sentiment information for more fine-grained classification. Our problem of detecting negative sentiment related comments share a common difficulty with problems such as cyberbullying detection: lack of annotated corpus and the inherent fuzziness of the hand labelled comments that are often difficult to classify.

Short text classification such as movie reviews on IMDB movie review dataset was conducted by numerous researchers, for example, Narayanan et al. [99] experimented simple Naive Bayes classifier with word n-grams and achieved 88.80% accuracy. Mac Kim and Calvo [100] developed a category and dimension based emotion prediction model on five universal emotion categories. They used latent semantic analysis and non-negative matrix factorization for dimentsionality reduction and achieved performance improvement beyond 50$\%$. Barobosa and Feng [101] developed a sentiment classification model of tweets based on Parts of Speech (POS) tags. In a similar work, Agarwal et al. [102] performed a comparative study on the twitter sentiment analysis with unigram baseline. By extending the data preprocessing to formal language word they showed the POS based models infact perform better. Prusa et al. [103] experimented with selective filter-based feature selection and showed this reduces the sparsity in the feature representation and improves performance significantly.

Research on text classification on hand annotated social media dataset has received particular attention for different motivations such as detecting helpful posts [104], useful comments [105], etc. Kavuluru et al. [104] used a dataset of manually labelled 3000 comments from suicide watch subreddit. They used n-grams, word psychometric scores, and discourse relation graphs as features to achieve *F-scores* beyond 80% for binary classification of helpful comments for better moderation. Poche [105] analyzed 6000 annotated Youtube comments and observed 77% accuracy for detecting useful comments using SVM classifier. Wanas et al. [106] proposed an automatic Slashdot post ranking using SVM classifier with RBF kernel. Similarly, Brennan et al. [107] attempted to classify Slashdot comments with binary class labels (good or bad) using 10-fold cross validation with SVM classifier and achieved an overall precision of 82%. FitzGerald [108] presented a method for classifying the quality of Slashdot comments using Linear-Chain Conditional Random Fields (CRFs) which yielded a precision of 81.8%.

The literature reviewed above gives us the impression that although few previous works addressed the issue of social media negative sentiment and non-use; they were based on questionnaire [50, 51], interview[52] and survey [49, 53] based qualitative study [7, 8]. In this work, we tried to address this issue by analyzing expert user generated content. Automatic text classification for detecting negative sentiment related to non-use is also a novel augmentation to the existing qualitative studies. To the best of our knowledge the study of tech-savvy users having social media negative sentiment is a novel approach hence it gives a new dimension to the future work.

CHAPTER 3

OVERVIEW OF THE QUALITATIVE ANALYSIS

As mentioned earlier, we collected all Facebook-related posts and associated reader comments from two sites: Slashdot and Schneier on Security. This thesis adopts a mixed method research; first, a random sample of these comments were selected for in-depth qualitative analysis. Then, using these hand annotated dataset as training data, we later trained different machine learning models for automated classification. The analysis of the methods are splitted in two chapters for better understanding.

We cleaned, scraped, and filtered the collected data and performed empirical qualitative coding on a random sample of comments. The coding was performed in two phases. First, to find out the comments which are negative sentiments towards social media, specifically Facebook. At the same time we also marked whether such negative sentiments evolved into explicit non-use or users still used Facebook irrespective of their negative sentiments. Second, what are the factors that primarily underlie for the negative connotations towards Facebook. Figure 3.1 summarizes the qualitative methodology. In the following sections, we describe our data collection, comment filtering & sampling, and coding procedures.

3.1 Data Collection

For both selected data sources, we scraped the respective sites for all posts related to Facebook since Facebook's inception in 2004 until the time of our data collection in May, 2019. We also collected all reader comments associated with the scraped posts.

3.1.1 Slashdot

We compiled a new corpus of articles and related comments from a technology news site: Slashdot. This site was chosen for several reasons: Slashdot was voted one of Newsweek's

Figure 3.1: Flow diagram of qualitative analysis

favorite technology web sites and rated one of Yahoo's top 100 web sites as "Best Geek Hangout."[1] Therefore, we can infer that posters and commenters of Slashdot are tech-savvy users and have better insight of the pros and cons of state of the art technology. Thus their reviews are critical in understanding the existing non-use of social media (Facebook in our case) and related sentiments. Also, we were looking for a discussion based blog or forum from where we can get unprompted naturalistic responses. Comments on Slashdot are moderated by users of the site, each comment has a score between -1 to +5 which ensures high quality contribution and discourages spams or vandalism. Each moderator in Slashdot is able to modify the score of a given comment by ± 1. Furthermore, each comment is classified in several categories: for good comments, the classes are: Interesting, Insightful, Informative, and Funny. For bad comments, the classes are: Flamebait, Troll, Off-Topic, and Redundant. Posts are associated with topic tags, such as political, facebook, social media, science, etc. This informative moderation environment makes Slashdot a useful source for user-generated content [30, 109, 110]. Figure 3.2 shows the threaded structure of the comments and the associated scores.

Slashdot comments are displayed in a threaded tree type discussion conversation. Commenters can reply in a given thread and subsequent replies for that comment appear in hierarchical order. We used Python *BeautifulSoup* and *Scrapy* for crawling and the data was saved into MongoDB in JSON format. For our purpose of analysis, we only considered the

[1] `http://digitalenterprise.org/cases/slashdot.html`

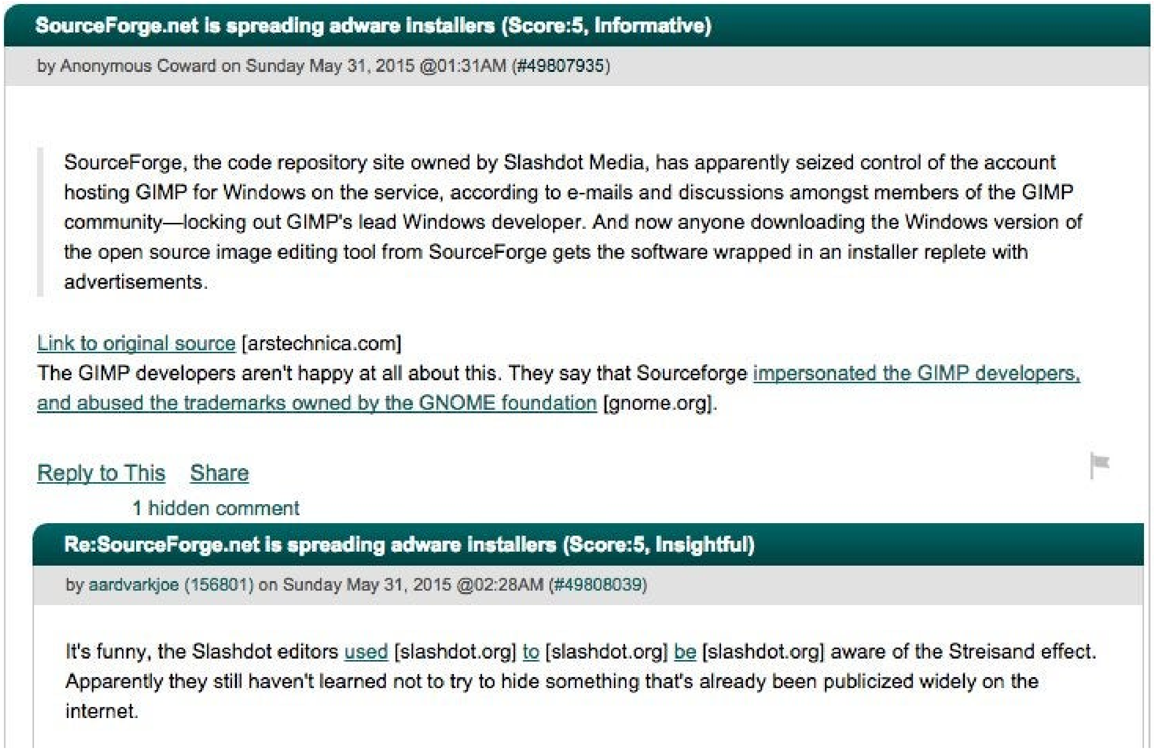

Figure 3.2: Typical Slashdot comment thread

posts which have a topic tag “Facebook.”

3.1.2 Schneier on Security

All posts on Schneier on Security are written by the renowned security professional Bruce Schneier. Like Slashdot, each post is tagged with one or more tags relevant to the content of the post, such as Algorithms, Identification, Privacy, Drones, etc. Readers of the blog can leave comments without needing to sign up for an account on the site. The comments are moderated, thus ensuring that off-topic or spam comments are rare. Comments are usually longer in length thus add additional insight and has general coverage of privacy/security compared to other similar sources, such as Krebs on Security with a heavy cybercrime bent. We decided against BoingBoing and Reddit as we couldn’t assume a reasonable tech savvy readership owing to diversity of topics and readers. Figure 3.3 shows a typical Schneier blog post.

We found that selecting posts with the tags ‘Facebook,’ and ‘Social Media’ yielded only

Figure 3.3: Typical Schneier of Security blog post

69 posts since Facebook's inception in 2004. On the other hand, selecting posts based on the presence of these two terms in the title or content of the posts returned a large number of posts irrelevant to our research. As a result, we manually examined every blog post from 2004 onward and selected 237 posts that were marked by three independent coders with 100% agreement as pertaining to our research question regarding Facebook. We then collected 11,365 reader comments associated with these 237 posts.

3.2 Comment Selection

From the collected combined dataset, we first filtered out comments that were less than 100 characters in length and those that did not contain either of the terms ‘Facebook,’ or ‘FB.’ For Slashdot comments, we further narrowed the set to include only those comments with a moderation score of +2 or higher. These criteria ensured that we filtered out comments without enough content to enable understanding the text as well as those that were marked as ‘low quality’ by the Slashdot moderation process. From this filtered set of comments, we randomly selected 5,000 comments for deeper examination and analysis, 4,000 from Slashdot and 1,000 from Schneier on Security, respectively. These 5,000 comments cover a period spanning 2006–2019 (2010–2019 for Slashdot and 2006–2017 for Schneier on Security, respectively). This sampling was done in separate times:

- Initial sample of 3,000 comments (2,000 from Slashdot and 1,000 from Schneier on Security) that we call `Dataset1`.
- For more training data for the text classifier (discussed in Chapter 4), we added 2,000 more comments from Slashdot later (in total 4,000 from Slashdot and 1,000 from Schneier on Security) that we call `Dataset2`. `Dataset1` is therefore a subset of `Dataset2`.

We used all 5,000 comments of `Dataset2` to do the initial sentiment coding to classify the comments related to negative sentiment toward Facebook. However, the thematic analysis was performed on the comments those were marked as negative sentiment related from `Dataset1`. The additional 2,000 comments were added later and used for the binary text classification purpose only.

3.3 Coding

We performed qualitative coding on `Dataset2` and `Dataset1`, respectively, in two stages:

3.3.1 Sentiment Coding

5,000 comments from `Dataset2` were used for the initial sentiment coding. The collected comments were manually coded by two independent coders to check if they belong to either one of the class *negative sentiment* or *not related to negative sentiment*. If a particular comment reflected harsh criticism of Facebook, active or passive usage, reduced usage or complete abandonment; the coders classified that comment as *negative sentiment*. Other comments which discussed all other different topics related to Facebook were marked as *not related to negative sentiment*. One typical such example of *negative sentiment* comment is:

> *"The future does not look so bright for facebook as they will probably suffer the same fate as usenet. Sounds like a positive outcome to me. The world would be better off without the cancer that is facebook."*

To make sure both the coders are on same wavelength, we checked the agreement after coding 500 comments from Slashdot. It turned out we achieved 78.3% agreement with the Cohen's κ=0.56. Later, we checked a subset of 20 of the disagreements and had a discussion to resolve the discrepancies and come to an agreement. Then we individually re-examined all other disagreement cases and checked if we wanted to keep or change our original classification. It was discovered that the raters encountered one major point where both of them disagreed mostly. Initially all the comments that explicitly stated or expressed some form of *negative sentiment*, such as "...deleted my account," "...not using for long," "...I hate Facebook," etc. were marked as negative sentiment only by one rater while the other rater also insisted to incorporate comments related to criticism, concern or sarcasm

Rater	**Slashdot**		**Schneier**	
	NS	**NRNS**	**NS**	**NRNS**
NS	1776	218	297	12
NRNS	50	1946	11	680

Table 3.1: Categories assigned by two independent raters on 4,000 Slashdot and 1000 Schneier's blog comments. (NS = related to *negative sentiment*, NRNS = not Related to *negative sentiment*)

of Facebook. One such example is:

> *"...How do the stream sorting algorithms work? If Facebook can't divulge that, I see no reason to trust them."*

This was reasonable since in such cases users might not explicitly mention about the *negative sentiment* but their aversion towards Facebook is visible which might lead them or encourage other users to non-use. After the discussion on the remaining disagreements, the agreement rate was 92.8% with the Cohen's κ=0.856. After coding all 5,000 comments, the overall agreement rate was 93.98% with the Cohen's κ=0.872. Table 3.1 summarizes the categories assigned by the raters after coding all 4,000 comments from Slashdot and 1000 comments from Schneier's blog respectively. Only 268 (6.7%) times for Slashdot and 23 (2.3%) times for Schneier's blog the raters disagreed. Since we got a high agreement score, we considered 2,044 (1776+218+50) and 320 (297+12+11) comments respectively (total of 2,364) which were marked by one or both the raters as related to *negative sentiment*.

3.3.2 Thematic Coding

After coding a total of 5,000 comments from `Dataset2` as discussed above, we wanted to expand more on finding the reasons of negative sentiment and non-use by the Facebook users. Following the grounded theory approach [111], an inductive and comprehensive qualitative thematic analysis [112] was done on the comments of `Dataset1` which were marked as *negative sentiment* related in the initial sentiment coding (1,342 such comments).

All the *negative sentiment* comments were read thoroughly for potential codes or themes by open coding (by the two coders). In general, similar comments were clustered together which were commonly mentioned by the users. Following grounded theory's constant comparison and axial coding scheme, the underlying factors were first identified, compared and later grouped together to constitute individual themes.

The key idea behind picking all the comments that were related to *negative sentiment* instead of just the ones expressing non-use was to capture the overall negative emotions of the users rather than a narrow focus on non-use only. Also that allowed us to get a bigger dataset for thematic coding. Simultaneously, individual comments were marked as *explicit non-use* if they directly expressed passive use, reduced use or non-use (e.g., "...I don't have an account," "...I have stopped using Facebook," etc.)

The two coders who coded the sentiment coding, coded a first set of 100 comments from all the *negative sentiment* comments. Initially we only allowed a single code per comment for the first set of coding. We noted the inter-coder reliability score was 70.4% with the Cohen's κ=0.637. The rating was fairly good, however we discussed about the coding scheme and noted that due to the nature of some of the comments we could not classify one comment only under a single theme. After discussing further, we decided on coding a single comment for multiple themes. For example, the following quote by one of the commenters not only shows that they are disliking ads on Facebook but also indicates their lack of trust to this media. Both the coders coded this comment under *Advertisements* as well as *Privacy and Security*.

> *"Maybe I'll see no ads after changing my Facebook age to something ridiculous? Why does anybody put their real age/birthday into Facebook?"*

After the thematic coding of a total of 100 comments, we found the inter-coder reliability score of 89.8% with the Cohen's κ=0.87. With this inter-coder reliability score, we coded the rest of the comments and found that the inter-coder reliability score after coding all the *negative sentiment* comments was 97.3% with the Cohen's κ=0.966. The iterative and independent assignment of the codes followed by the review ensured trustworthiness. Table 5.1 summarizes the 10 themes which were surfaced with their relative frequency.

CHAPTER 4
OVERVIEW OF THE AUTOMATED ANALYSIS

In order to answer *RQ2*, this thesis attempted to augment the findings of the qualitative study with different automated analysis. This is important since manual analysis is not feasible when there are substantial amounts of data. Also if we want to do some bigger scale analysis, for example how different negative sentiment related factors evolve over the period of time, it is imperative to analyze the overall dataset. To get an overview of the text data that we selected for analysis we did an initial unigram, bigram, and trigram wordcloud generation and topic modeling. These two techniques give us an top level idea of the underlying themes. We also did a basic sentiment analysis on the overall set of comments. Later we did text classification for automatic detection of negative sentiment related comments. To avoid confusion, Table 4.1 lists the dataset used for individual analysis. In the sections to follow we will discuss the steps in detail.

4.1 Wordcloud and Topic Modeling

Wordclouds have been particularly useful to get a top level idea and has been widely used for text mining [113]. We picked the 2,364 comments that were marked as negative sentiment related and built wordcloud using Python *wordcloud* package. The initial wordcloud had irrelevant terms which are not informative, such as "facebook," "would,' "say" etc.

Analysis	**Dataset**
Wordcloud	2,364 Negative sentiment comments of `Dataset2`
Topic Modeling	2,364 Negative sentiment comments of `Dataset2`
Sentiment Analysis	1,342 Negative sentiment comments of `Dataset1`
Text Classification	Overall `Dataset2`

Table 4.1: Dataset used for different automated analysis.

Therefore, we created a custom list of stopwords ourselves outside of the Python package *nltk* provided list of stopwords. To get better insight, we created bigram and trigram wordcloud on the same dataset using *tm* library of R. Data was preprocessed and cleaned using the process described in [114]. While we removed the stopwords for unigram and bigram wordclouds, it was not done for trigrams as we wanted phrases like “i hate facebook,” “i don’t care,” etc.

Besides wordclouds we performed Latent Dirichlet Allocation (LDA) topic modeling [115] to surface the hidden topics (sentiment words, non-use related terms) from the comments. LDA is a statistical method that assumes a corpus is a mixture of topics and a topic is mixture of words [116, 114]. We used the python library *Gensim* and trained the model on 2,364 *negative sentiment* comments after cleaning the data (removing punctuation, proper names, tokenizing, and lemmatizing). The number of topics were chosen as $K = 8$ because the topics started to repeat after this threshold hence becoming less interpretable. We made sure to eliminate words that only appeared a few times and generic words that appeared several times. This helped improve the information gain and find better topics by fast convergence. Each topic is associated with weighted words under that topic and the high weight represents most depictive word of that particular topic.

To confirm our results of LDA topics, we implemented another popular topic modeling approach called Non-negative Matrix Factorization (NMF) because it is better at compact semantic representation of documents in small data setting [117]. Using TF-IDF from *scikit-learn*, we converted each document in corpus to vectors. One important thing to note is, we had to set minimum and maximum document arguments to get rid of too generic or too unlikely words. The number of topics were chosen as 8 for the similar reasons mentioned above.

4.2 Sentiment Analysis

To understand the polarity of sentiments of users, we performed sentiment analysis on the comments. We used VADER (Valence Aware Dictionary and Sentiment Reasoner), a lexicon and rule-based sentiment analysis tool that is specifically engineered to extract sentiments expressed in social media. It incorporates qualitative and quantitative methods and performs as well as other standard benchmarks for sentiment analysis (e.g., LIWC, ANEW, the General Inquirer, and machine learning based techniques) [118]. We performed overall and theme wise sentiment analysis using *vaderSentiment* package in Python and reported the compound sentiment score and associated sentiment (positive or negative). One important adjustment that we made while doing this analysis was: we filtered out non sentiment related terms and picked the opinion terms instead. This technique was proposed by Pang and Lee [119] to remove objective sentences by extracting subjective ones. Often times in lexicon based sentiment analysis in a review or comment (especially long ones), it is observed that the non-sentiment related sentences affect the overall sentiment polarity of the review. To overcome this, we adopted a simple approach to filter the opinion sentences from the corpus. We took advantage of the list of positive and negative words compiled by Hu and Liu [120] (2,005 positive and 4,783 negative words). From each comment, the sentences that contain at least one sentiment word from above list (positive or negative) were retained. We call them the *opinion sentences*. The remaining analysis was routinely performed using VADER tool, and the categorization of scoring and interpretation of results are given in Chapter 6.

4.3 Text Classification

As explained previously, 5,000 comments from `Dataset2` were used as the training set for the text classification task. Figure 4.1 explains the steps involved in the machine learning based supervised text classification. We will examine the steps in details in the subsec-

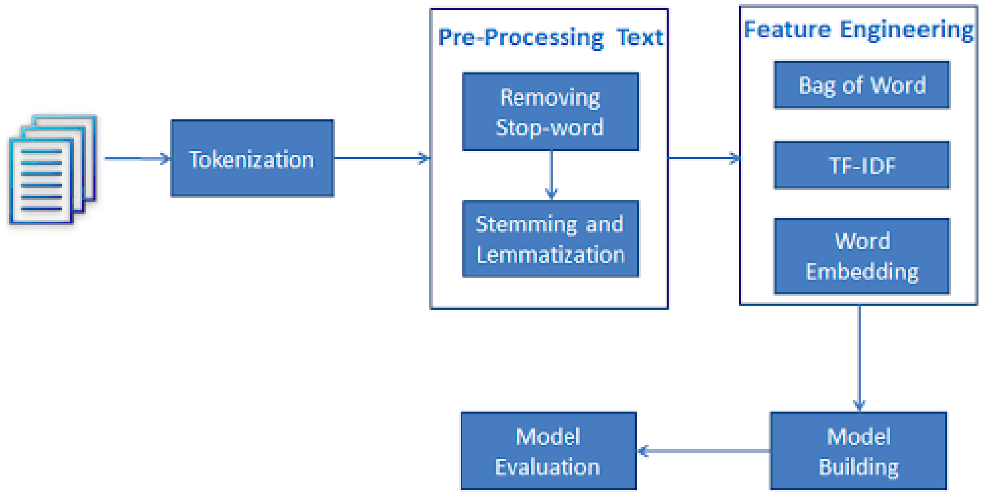

Figure 4.1: Flow diagram of Text Classification task

tions to follow:

4.3.1 Data Pre-processing

After the hand labeled dataset is ready, we need to do basic cleaning and pre-processing before we can apply the machine learning models. Following are some common techniques used for cleaning text data [121]:

1. **Converting to lowercase**: All words in the comments body are converted to lowercase so that multiple copies of same word does not exist. Also it doest not alter the semantic importance of a given word. For instance, 'Work' and 'work' should be treated as the same.

2. **Removing numerical values and punctuations**: Numerical values and special characters used in punctuation($,! etc.) introduces overhead and does not provide any extra information so they are removed.

3. **Removing extra white spaces**: To avoid problem during tokenization, extra white spaces that often occur in social media data as part of typos are removed.

4. **Tokenizing into words**: This refers to splitting a text string into a list of 'tokens,' where each token is a word. For example, the sentence 'I hate Facebook' will be converted to ['I,' 'hate,' 'Facebook'].

5. **Removing non-alphabetical words and 'Stop words'**: Commonly occurring words or 'stop words' such as, 'the,' 'is,' 'are,' etc. does not provide any extra information for the classifier and hence are removed.

6. **Stemming and Lemmatization**: Stemming refers to the removal of suffices, like 'ing,' 'ly,' 's,' etc. by a simple rule-based approach. Lemmatization is a pretty rad technique which converts similar words to their base meaning. For example, the words 'crying' and 'cried' will both be converted into their root word 'cry.'

We used basic python *NLTK* library to perform all of the aforementioned tasks.

4.3.2 Feature Engineering

The next step is the feature engineering step. In this step, raw text data will be transformed into feature vectors and new features will be created using the existing dataset. We will implement the following different ideas in order to obtain relevant features from our dataset.

1. Count Vectors as features

2. N-grams as features

3. TF-IDF Vectors as features

4. Topic Models as features

Count Vectors as features

Count Vector is a matrix notation of the dataset in which every row represents a document from the corpus, every column represents a term from the corpus, and every cell represents the frequency count of a particular term in a particular document.

N-grams as features

N-grams are basically a group of consecutive words clustered together. For example, N-grams with N=1 are called unigrams. Similarly, bigrams (N=2), trigrams (N=3) and so on can also be used. N-grams are particularly useful to capture the inherent language structure. For instance, bigram and trigram can be more informative compared to unigrams.

TF-IDF Vectors as features

Term Frequency (TF) is simply the ratio of the count of a word present in a sentence, to the length of the sentence. Therefore, we can generalize term frequency as:

$$TF(T) = \frac{\text{number of times term T appears in the particular row}}{\text{number of terms in that row}}$$

Inverse Document Frequency (IDF), computed as the logarithm of the number of the documents in the corpus divided by the number of documents where the specific term appears.

$$IDF(T) = \log \frac{\text{total number of documents}}{\text{number of documents with term T in it}}$$

TF-IDF(Term Frequency-Inverse Document Frequency) normalizes the document term matrix. It is the product of TF and IDF. Words with high TF-IDF in a document occur most of the times in that particular document and must be absent in the other documents. In this thesis, we experimented with word level, n-gram level, and character level TF-IDF features.

Topic Models as features

Topic Modelling is a technique to identify the groups of words (called a topic) from a collection of documents that contains best information in the collection. I have used Latent Dirichlet Allocation for generating Topic Modelling Features. LDA is an iterative model which starts from a fixed number of topics. Each topic is represented as a distribution over words, and each document is then represented as a distribution over topics. Although the tokens themselves are meaningless, the probability distributions over words provided by the topics provide a sense of the different ideas contained in the documents.

4.3.3 Model Building

The final step in the text classification framework is to train a classifier using the features created in the previous step. There are many different choices of machine learning models which can be used to train a final model. Kowsari et al. [122] analyzed the different text classifiers with their advantages and pitfalls. We will implement the classifiers described below for the purpose of this thesis.

Naive Bayes

As per [123], Naive Bayes classifiers are linear classifiers that are known for being simple yet very efficient. The probabilistic model of naive Bayes classifiers is based on Bayes' theorem, and the adjective naive comes from the assumption that the features in a dataset are mutually independent. A Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. In tasks like spam email detection with small sample size Naive Bayes has performed well despite being based on a somewhat unrealistic assumption.

Let's Consider the following problem of classification $x \in X^p \longmapsto y \in \{1, 2, \ldots, M\}$. Here, $x = (x_1, x_2, \ldots, x_p)$ is a vector of descriptors (or features) : $\forall i \in \{1, 2, \ldots, p\}, x_i \in X$, with $X = \{1, 2, \ldots, K\}$.

Goal : Learning $p(y|x)$ Bayes formula gets us :

$$p(y|x) = \frac{p(x|y)\, p(y)}{p(x)}$$

The Naive Bayes method consists in assuming that the features x_i are all conditionally independent from the class, hence :

$$p(x|y) = \prod_{i=1}^{p} p(x_i|y)$$

Then, the Bayes formula gives us:

$$p(y|x) = \frac{p(y)\prod_{i=1}^{p} p(x_i|y)}{p(x)} = \frac{p(y)\prod_{i=1}^{p} p(x_i|y)}{\sum_{y'} p(y')\prod_{i=1}^{p} p(x_i|y')}$$

Logistic Regression

Logistic Regression is a linear classifier which is an estimation of Logit function [124]. Logit function is simply a log of odds in favor of the event. This function creates a s-shaped curve with the probability estimate, which is very similar to the required step wise function. Given a set of inputs X, we want to assign them to one of two possible categories (0 or 1). Logistic regression models the probability that each input belongs to a particular category. The underlying theory [125] is presented here:

The target variable: $y \in \{0, 1\}$ where 0 = negative, 1 = positive (binary and multiclass classification problem exist). For **Logistic Regression**, $0 \leq h_\theta(x) \leq 1$. Linear regression is unreliable because the value of sigmoid funtion ($h_\theta(x)$) can stray out of scope.

Sigmoid function: $h_\theta(x) = g(\theta^T x)$, where $g(z) = \frac{1}{1+e^{-z}}$ and $z : R$

So,

$$h_\theta^{(x)} = \frac{1}{1 + e^{-\theta^T x}}$$

Predicts $y = 1$ when $h_\theta(x) \geq 0.5$, or when $\theta^T x \geq 0$.

Cost Function: Loss/cost functions are a measure of how well the algorithm performs using random weights:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\text{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

Where, $\text{Cost}(h_\theta(x), y) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$

If $y = 1$, the second half of the cost function is ignored.

If $y = 0$, the first half of the cost function is ignored.

Gradient Descent To minimize the loss function by adjusting the weights we perform the derivative of the loss function with respect to each weight:

$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)}\log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))]$

$\min_\theta J(\theta)$:

Repeat{

$\theta_j := \theta_j - \alpha\frac{\partial}{\partial\theta_j}J(\theta)$

} (simultaneously update all θ_j)

$\frac{\partial}{\partial\theta_j}J(\theta) = \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$

Support Vector Machine (SVM)

Like naive bayes, SVM doesn't need much training data to start providing accurate results and it is very robust hence widely used for small text classification tasks. As described in [126], the simplest SVM implementation is a non-probabilistic binary linear classifier which attempts to find a line that can accurately separate data points into two classes. A non-linear function is desired if the data is not linearly separable considering the data on higher dimension. Support vector machines for binary classification attempt to construct a hyperplane and to maximise the distance from the hyperplane to the nearest data points. A very simple 2-dimensional example is shown in Figure 4.2.

The underlying theory [125] is presented here: For a decision hyper-plane $\mathbf{x}^T\mathbf{w} + b = 0$

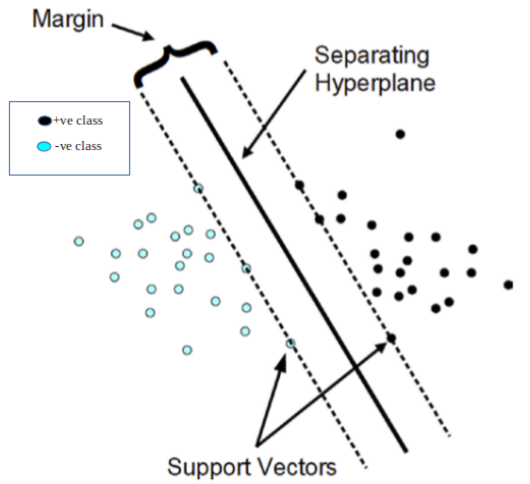

Figure 4.2: SVM classifier

to separate the two classes $P = \{(\mathbf{x}_i, 1)\}$ and $N = \{(\mathbf{x}_i, -1)\}$, it has to satisfy

$$y_i(\mathbf{x}_i^T\mathbf{w} + b) \geq 0$$

for both $\mathbf{x}_i \in P$ and $\mathbf{x}_i \in N$. Among all such planes satisfying this condition, we want to find the optimal one H_0 that separates the two classes with the maximal margin. The plane which is going to optimal is in between the two classes. The equality $(y_i(\mathbf{x}_i^T\mathbf{w} + b) \geq 1)$ holds for those points on the planes H_+ or H_- (H_+ and H_- are parallel to H_0). Such points are called *support vectors*, for which

$$\mathbf{x}_i^T\mathbf{w} + b = y_i$$

i.e., the following holds for all support vectors:

$$b = y_i - \mathbf{x}_i^T\mathbf{w} = y_i - \sum_{j=1}^{m} \alpha_j y_j (\mathbf{x}_i^T\mathbf{x}_j)$$

Therefore, our goal is to minimize the norm $||\mathbf{w}||$:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} = \frac{1}{2}||\mathbf{w}||^2 \qquad \text{(objective function)} \\ &\text{subject to} \quad y_i(\mathbf{x}_i^T\mathbf{w} + b) \geq 1, \;\; \text{or} \;\; 1 - y_i(\mathbf{x}_i^T\mathbf{w} + b) \leq 0, \qquad (i = 1, \cdots, m) \end{aligned}$$

Whenever the data points are not linearly separable, we need to use kernel trick. A kernel is a function that takes two vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ as arguments and returns the value of the inner product of their images $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$:

$$K(\mathbf{x}_1, \mathbf{x}_2) = \phi(\mathbf{x}_1)^T\phi(\mathbf{x}_2)$$

Some popular kernels involve: linear, polynomial, and RBF. In this thesis linear kernel was used.

Random Forrest

Random forest is an ensemble machine learning algorithm that is used for classification and regression problems. Random forest is like bootstrapping algorithm with Decision Tree model. Random forest tries to build multiple Decision Tree models on the random sample of the initial observations. It will repeat the process multiple times and then make a final prediction on each observation.

Note that, all of the aforementioned models were implemented using Python's *scikit-learn* library and the visualizations were done with the help of modules like *pandas*, *numpy*, *matplotlib*, and *seaborn*.

CHAPTER 5

RESULTS OF THEMATIC ANALYSIS

At the end of the sentiment coding described in the Chapter 3, the two coders agreed on classifying 1,185 out of the sample of 3,000 comments of `Dataset1` as containing negative sentiment toward Facebook (888 for Slashdot and 297 for Schneier on Security, respectively). The two coders also agreed that 1,658 comments did not express negative views regarding Facebook (978 for Slashdot and 680 for Schneier, respectively). In the case of the remaining 157 comments, one of the coders marked them as containing negative sentiment toward Facebook while the other coder did not. Given the high level of agreement between the two coders (2,843/3,000 = 94.8%), we decided to mark these 157 comments as containing negative sentiment because at least one of the coders classified it as such.

As described in Chapter 3, we engaged in thematic coding of these 1,342 comments (1,022 for Slashdot and 320 for Schneier on Secuity, respectively). The results of the coding are summarized in Table 5.1. The following sections describe in detail each of the 10 themes that emerged during the thematic coding process followed by a section on explicit mentions of non-adoption or non-use.

5.1 Privacy and Security

Unsurprisingly, privacy and security featured prominently among the expressed concerns. Facebook's data mining, selling of their personal information, third party data sharing, etc. were thoroughly criticized by unhappy users. One commenter mentioned:

> "*Facebook is actively developing new, innovative ways to invade your privacy, and this particular bit of data mining technology has become reliable enough that they felt it would be good PR [Public Relations] to create a feel-good,*

Theme	#Comments (Slashdot)		#Comments (Schneier's Blog)	
	Overall	*Explicit non-use*	Overall	*Explicit non-use*
Privacy and security	432 (42.27%)	138 (44.37%)	212 (66.25%)	111 (68.94%)
User experience	341 (33.36%)	118 (37.94%)	41 (12.81%)	18 (11.18%)
Advertisements	104 (10.17%)	21 (6.75%)	14 (4.37%)	5 (3.1%)
Psychosocial well-being	9 (0.88%)	1 (0.32%)	7 (2.18%)	3 (1.86%)
Personal disposition	147 (14.38%)	42 (13.5%)	94 (29.37%)	68 (21.86%)
Alternate options	18 (1.76%)	8 (2.57%)	18 (5.62%)	10 (6.21%)
Uninteresting content	39 (3.81%)	8 (2.57%)	28 (8.75%)	12 (7.45%)
Politics	30 (2.93%)	5 (0.01%)	1 (0.31%)	0 (0%)
Fake accounts and bots	5 (0.49%)	2 (0.64%)	22 (6.87%)	8 (4.96%)
Miscellaneous	35 (3.42%)	11 (3.53%)	34 (10.62%)	14 (8.69%)

Table 5.1: Results of thematic coding showing the comments with negative sentiment toward Facebook falling under each theme for each of the two data sources along with the fraction of those comments that explicitly indicated Facebook non-use. Percentages for each column do not add to 100% because the themes are not mutually exclusive.

> *help-the-disabled feature out of it.*"

Social media usage has been shaped and reformed over the course of time due to issues related to user privacy and security [127]. One major point of concern has been identity theft. The fact that Facebook uses user profile information made people skeptical. One such user wrote,

> "*By signing up for Facebook, you surrender your identity as it is – maybe this is to take it to the next level. They will stream everything these people do for our entertainment.*"

Another set of users did not like Facebook's Terms Of Service (TOS) requiring the use of their real name: "*Not to mention, Facebook's TOS is that you must use your real name when creating an account. If FB cared one whit about privacy it would let people uses aliases. It really is incredible how ghetto and scammy FB is with their tactics and policies.*" However, one user argued that although he does not have a personal account, he might consider having a corporate account since it is a good way of reaching people:

> "*. . . Never had an account on FB. . . never will. . . I don't want to give any per-*

sonal info to them, but I would consider a business only account if they offered one.”

Slashdotters were also apprehensive of the lack of security of Facebook source code, and some of them wanted it to be open source: *“If Facebook source code was open, security flaws which are disclosed every one day or another, probably ruining life of thousands due to privacy leaks, would have been closed faster…”* Privacy violations and security breaches were the major points of concern of many individuals who considered that Facebook does not take them seriously enough:

> “*…Of course, Facebook isn’t the only company guilty of this type of thing – and I suspect that until there is some serious consequence associated with this type of security hole, most companies won’t take it seriously enough.*”

Privacy and security were the primary concerns for negative sentiment and non-use of Schneier’s blog users as well: *“So if G and FB are OK with privacy being dead, why don’t they share the details of their advertising operations and source code? Or do they mean privacy is dead for everyone but themselves?”* The amount of surveillance (storing metadata, emails, phone call, etc.) and tracking that comes with a Facebook account was worrying for many:

> *“…If Facebook can do it, the NSA can do it, too. In a few years, everybody can do it. This is one of the reasons why I don’t have pictures of myself on Facebook.”*

Time and again it was argued that Facebook has monopolized the social media business and the major incentive is money instead of better service. People described how they attempt to avoid being tracked: *“I have a fake account of Facebook, it’s useful when researching. Considering I use it as an information resource and since I value my privacy, there is no way I would put my personal information on there! Muffin got it right. Don’t use it.”*

5.2 User Experience

This category encompasses all of the Facebook features that people did not like, e.g., bandwidth consumption, problems in surfing the site, video features, etc. One unhappy commenter mentioned, "*Not a fan of this idea. Facebook Messenger spams could exceed the data caps on mobile plans ($$$) and create a backlash from angry customers.*" The ability to control the profile (e.g., customizing the privacy and security settings, enabling/disabling features) is a factor important to users. When that ability is inhibited, user dissatisfaction escalates, such as:

> *"Just went into my profile to try to remove / disable this POS and you are not even given the option to do so... I am so close to closing my Facebook account it is not even funny anymore."*

Also, a few people complained about the plainness of the Facebook interface: "*I'm surprised no one mentioned the disgusting interface Facebook has. It really seems badly made and unsophisticated...*" Similar arguments were made by another commenter who said,

> *"If Facebook continues to make its site user-unfriendly, I'll simply stop using Facebook. I've already dropped back on my usage because I cannot view my timeline the way I want to view it... "* One user mentioned Facebook users are attention-seeking and that *"An hour on Facebook will make you hate humanity."*

Poor Facebook user experience was a theme also prevalent among Schneier's blog commenters. A lot of them mentioned that it is annoying that Facebook gathers information of people who do not even have an account:

> "*Scarier, is that you don't even need to have a Facebook account for them to have a dossier on you. All it takes is countless "friends" uploading their contact list to try to maximize friendship.*"

Another user complained about Facebook's friend suggestions: *"The whole "people you may know" thing is kind of creepy. I once had Facebook suggest a person who had died not long before the friend suggestion."* Similar negative sentiment was found due to false positives in friend matching:

> *"... I find it quite hard to believe that Facebook matches people who happened to be in the same location for some time. That would generate an awful high number of false positives... "*

5.3 Advertisements

Advertising was a specific aspect of user experience that turned many off. This category was often connected to privacy and security concerns. One disgruntled person sarcastically commented, "*Don't be silly – when Facebook taps into your pleasure center, it won't be to notify your friend that you're horny, it will be to give you a dopamine hit every time you view an advertisement. Within a few days you'll want to do nothing else.*" They felt that they should be compensated:

> *"if Facebook is making $1 off of my inconvenience, Facebook should pay me at least .50c of that money... "*

While a few individuals liked targeted ads, many expressed contempt toward the data mining that drives them: "*... Facebook probably mines the unencrypted messages to help form an "advertising profile" for you so they can better target ads at you when you're on Facebook.*" One Slashdotter thought stopping the use of Facebook is the only remedial measure:

> *"Maybe it's time for you to stop using Facebook altogether. Your continued use of the site IS their explicit permission that they can serve you their content – ads and all."*

However, some commenters, although dissatisfied, accepted this as the inevitable: *"After all, Facebook users have NO CHOICE but to use Facebook and allow them to force you to watch ads. Really, you have no choice, none at all. Suck it up and get used to it."* Somewhat similar arguments were made by Schneier's blog commenters who were more concerned with the associated security hazards caused by ads. They felt that Facebook collects all information if browser cookies are enabled, tracking visits to every site with a 'Like' button: *"...I'm sure Facebook has much, much more data on a given user (at least in 2012); they would have to in order to make the targeted ads... right?"* In fact, advertising did contribute to people leaving the platform:

> "*even though I'm spiteful of advertisements/marketers for a variety of reasons, it was an ad that really struck me and contributed to the many things that got me off of Twitter/Facebook and all the other wannabe sites.*"

5.4 Psychosocial Well-being

Depression caused some to stop or reduce their use of Facebook because they found either that their lives were too boring compared to their online friends or that the highlights shared by the friends were fake: "*The negative impact social media has on the human psyche is, in my opinion, quite significant. FOMO, self-esteem issues and F4ceb00k depression are real things and they exist with a measurable amount of people who live through mass social media...*" Facebook was accused of creating frustration and depression by encouraging social comparison:

> *"... Quitting Facebook was one of the best decisions that I ever made. Nothing good comes from it. It encourages us to compare ourselves negatively to others whom may have more money or more success. Facebook is psychologically damaging."*

It was argued that being on a social media like Facebook can affect the mental health of people who are already suffering: "*The problem with social media (not just FB) and depression is not that people do nothing but stare at FB and get depressed. The problem (or, one of them) is that if you're already depressed, viewing social media can make it worse...*" One commenter of Schneier on Security tried to explain the co-relation of mental health and social media by saying a majority of the US ("a pill-popping nation") population is suffering from anxiety disorder and depression and *"... the masses are unstable and the trend is worsening – Social media correlations with higher rates of reported mental illness (ironically this study used Twitter)... "* Some also mentioned that Facebook takes a toll on family life and creates loneliness:

> "*... If your kids did not grow up in the age of Facebook, then it is uncharted territory to you. You don't know the dangers, the addiction, the toll it takes on a family as they drop all activities and everything, for Facebook activity...*"

5.5 Personal Disposition

If a comment did not provide sufficient content or context for the expressed negative sentiment, we attributed the sentiment to personal disposition or preferences. Also, comments such as *"I do not care," "waste of time," "do not like using it,"* or merely containing hate/swear words, were inferred as personal opinions. One such person indicated they are not in any social media but doing well: *"... I did 4 years ago and I'm surviving just fine. No Twitter, no Facebook, no Instagram. Nothing."* As mentioned earlier, a lot of commenters considered Facebook a pure waste of time that eventually becomes an addiction:

> *"... Facebook is a waste of time simply because Facebook neuters your ability to reach people."*

Lack of contextual information for deliberate non-use was also observed in some Schneier's blog comments: *"People should delete their Facebook account and other social media accounts."* A common practice was mocking social media sites with derogatory names; e.g., “*you really don’t need to follow the herd to Faceache and Twatter.*” Many felt that the popularity of Facebook is a fad that will soon be replaced by the ‘next big thing.’ Sometimes personal disposition was accompanied by other concerns, such as privacy/security:

> “*Protect yourself by forcing the authentication to happen over TLS [Transport Layer Security]. Or stop logging in to Facebook from public networks. OR, do the sensible thing and don’t use Facebook at all.*”

We also identified comments where people mentioned using Facebook reluctantly due to social or professional reasons, despite their negative personal disposition toward the service: *“It keeps me vaguely in the loop of events and makes me available for my friends and family to contact me. But I keep my privacy options up-to-date and don’t use FB as authentication for anything else. Nor do I use many FB apps or games.”*

5.6 Alternate Options

A number of comments described alternate options for achieving the same purposes without being impacted by the negative aspects of Facebook: “*For people with a large audience, there’s Twitter, until Facebook buys it. . .* ” Many Slashdotters observed that Facebook is considered old-fashioned and teenagers do not use it anymore:

> *“The kids have already moved to Snapchat. Old people will no doubt stay with Facebook forever, but that’s the end of growth and growth is holy in Silly Valley. . . ”*

Those who found Facebook not worth their time preferred to communicate via other sources e.g., email, text, etc.: “*But I tell ya what – Facetwat is not on my phone. (yes, Facetwat is my derisive mashup of Facebook and Twitter.). . . You wanna reach me? Email me. Text me. iMessage me. Or fucking call me.*”

Similar tendencies were observed among Schneier’s blog commenters. Those who believed that Facebook has security issues and found it not trustworthy, preferred other modes of communication:

> *“...In any case, I wouldn’t trust anything from Facebook as far as security/privacy goes. I use Signal.”*

Another user discussed privacy-friendly alternative: *“...There is already an alternative project being created. Does nobody know about Diaspora? It’s open source and distributed, with no central company collecting all your private data...If you have privacy issues with Facebook, spread the word about Diaspora!...”*

5.7 Uninteresting Content

This theme appeared in multiple comments where the users expressed their detest to inappropriate, irrelevant, or uninteresting content shared by others. For example:

> “*I see very little content that’s actually from the people I follow...I hardly ever use Facebook anymore because most of that reshared content doesn’t interest me anyway.*”

Facebook was considered a source of spreading propaganda, hoax, spamming, etc. Commenters also considered it as a medium for assaulting people personally by trolling and abusing which caused both *explicit non-use* and implicit negative sentiment:

> “*I still don’t have a Facebook account, and am no worse the wear for it. I have noted that of my family and friends who do have accounts, the ones who typically talk about their Facebook activity the most are definitely the women, and a lot of that talk seems to swing between gossip and outright vicious assaults. I’ll just stay out of that mess, thanks.*”

Another user remarked, *“...Facebook these days seems to just be an outrage platform with different sides all screaming at each other and nobody listening...”* One user mentioned why he has negative sentiment due to the infinite trolling and fake news:

> "*I wouldn't be surprised if readership is falling. After I had to open a Facebook account last year, it looked interesting for about a week, then it became an annoyance, now it seems to be troll land. . . And as quickly as fake news is deleted, new ones will pop up, and will make note of being deleted, which will feed into conspiracies.*"

We read varied comments belonging to this category in the Schneier on Security dataset. It was commonly opinioned that Facebook users live in a world of delusion and they 'survive on their own filtered image.' One commenter wrote, *". . . If Facebook contact, which is hardly face to face and certainly not person to person, is your social life, then you may have no life at all. . . Facebook et al, remind me of something I have observed in the schools of America today, called the "liars club". . . "*

Content and audience are influential factors for technology adoption. People care about how the platform moderates its audience and controls its content. Commenters of Schneier on Security tended to be harsher with Facebook's disregard for fake profiles and pseudonyms. One such user also thought that Facebook is full of useless stuff:

> "*the irony I see in this is that if Facebook didn't try to imply that they care about using real names/make it so much of a "MEATSPACE PROFILE" (they don't, people use pseudonyms/create fake profiles lately), I wonder if Facebook would be as popular or as filled with uselessness as it tends to be.*"

Another user compared Facebook to a 'yearbook': *". . . I know next to nothing about MySpace or Facebook. I'm guessing you were not the type to purchase a yearbook and have all your friends scribble notes in it?. . . I can't help but think of these sites in those terms. . . "*

5.8 Politics

Politics was explicitly mentioned as a type of Facebook content that people found inappropriate and undesirable: "*I quickly learned to use the "mute" feature in Facebook because of those idiots who post zero impact political comments…*" Here the user is clearly frustrated and expressed how his interaction with Facebook has changed due to the audience activity. Some of the users questioned the credibility of the news sources in Facebook:

> "*Facebook does have credibility? Was that before or after Facebook admitted to promoting pro-Hillary and suppressing pro-Trump stories/outlets?*"

Occasionally commenters explicitly indicated political contents as the driver for non-use:*"And this is why you don't go on Facebook ever. Well that, and I literally couldn't give less of a damn about the kind of leftist drivel that populates most of any social media."* Meanwhile others blamed Facebook for keeping them in a 'political bubble,' such as:

> "*Yes, Facebook is keeping me in a political bubble, but not nearly to the extent that National Review did in the early 90s…*"

One commenter of Schneier's blog indicated his apprehension towards the idea that Facebook CEO Mark Zuckerberg is capitalizing on the data to create his own intelligence service, calling it some kind of 'oligarch plum.' Some criticized Facebook by claiming it is the worldwide censor of those who manipulate civic discourse:

> "*…Mr. Zuckerberg is reportedly preparing for a presidential run. The appearance is with his big money and data access he is trying to buy the presidency…Can Facebook and its executives be trusted?*"

5.9 Fake Accounts and Bots

Commenters were often worried about malicious activities via fake Facebook accounts. One user was worried that the flexibility of creating multiple accounts could stimulate crime. For instance, “*Since Facebook doesn’t verify your identity when signing up for an account…how long before the bad guys start setting up fake accounts or hacking FB accounts and ransoming people?*”

Fake accounts and bots were criticized [128] since they can be used to purchase fake likes or post reviews for businesses:

> “*Because I have a hard time believing that 1/7 of the world population is actively on Facebook and I also know full well that there are millions of bot accounts whose likes are purchased.*”

Commenters of Schneier on Security also addressed this issue. They chastised Facebook for the ‘like business’: *“1,000 Facebook likes for $29.99. The unknowable number of times Facebook has betrayed, bamboozled, lied to, and scammed their own user base makes it quite difficult for me to have even one byte of sympathy for Mr. Z and his shaky marketing company.”* One user thought negative reputation online could be shed easily by creating a new persona and thus positive reputation is vulnerable to ‘Sybil attacks’:

> “*…If these two problems are solved a lot of problems would immediately vanish off the net. i.e., like sellers on Facebook…fake reviews in online shops. Solving this key problem of pseudonymity in a cheap way without breaking anonymity seems hard.*”

From our data it was apparent that Schneier on Secuirty commenters were more critical compared to Slashdotters regarding misinformation and intrusion. One of them suggested,

> *“Technologists and inventors need to work on cutting Facebook off at the knees. Disposable identities, fake searches, encrypted fake traffic, photos of non-existent people, artificial-intelligence tools for generating fake blogs and comments, fake clicks, a tsunami of false information…”*

5.10 Miscellaneous

Comments that mentioned any issues apart from the above nine major themes were gathered under a single theme of miscellaneous concerns. For instance one person wrote,

> "*The overwhelming majority of people use Facebook. They don't give a shit about the ideals of freedom and decentralization. And they outnumber us hugely. We can't 'win' in an essentially democratic system wherein a tiny 'we' is up against a massive horde of 'them.'*"

Some people mentioned not liking the Facebook app because it drained the battery: "*Years ago I deleted the Facebook app due to excessive battery drain.*" Another common complaint for user negative sentiment was stalking. A number of commenters were annoyed by employers requiring potential employees to have a Facebook account. This led to dissatisfaction because not having a Facebook account made people ineligible for a job that they were otherwise qualified for:

> *"I don't use Facebook. . . If an employer wants my years of experience they will take me as I am. If they are going to reject me because I don't waste time on Facebook, then I probably wouldn't last long there. Their loss."*

On a similar note another Slashdotter said, "*The reason employers want everyone to be on social media: They can use it to gather information about you that would be illegal or inappropriate to ask in a job interview.*" On rare occasion, we encountered unconventional views, for example:

> "*Due to the fact that Facebook is nothing more than prostituting your social life in exchange for some IT service it makes sense that Facebook should be rated XXX and that Facebook should be prosecuted for boosting child prostitution if they do not take care that no child gets a Facebook account.*"

5.11 Explicit Non-use

As mentioned earlier, the negative sentiment comments which explicitly described some form of non-use of Facebook were marked as *explicit non-use*. It turned out that, 311 out of 1,022 negative sentiment comments from Slashdot and 161 out of 320 negative sentiment comments from Schneier's blog reflected *explicit non-use*. Table 5.1 summarizes the findings (percentages are calculated on number of *explicit non-use* only comments per dataset). Wordtree visualization [129] of *explicit non-use* for Slashdot and Schneier's blog comments (those include the words "quit" or "stop") are shown in Figure 5.1 and Figure 5.2 respectively (the branches explains the reasons).

Privacy and security was the majority theme in the *explicit non-use* subset (44.37% for Slashdot and 68.94% for Schneier's blog). The next two prominent themes for *explicit non-use* for Slashdot were *user experience* and *personal disposition*, and vice versa for Schneier's blog. Since *privacy and security* and *user experience* comprise substantial proportions of *explicit non-use* only comments, we examined them more deeply. We found that non-users were concerned about Facebook's default privacy and security settings, identity theft, selling of personal data for ads, tracking users even after logging out, automatically posting recent activity, scam applications & ads, fake profiles, retaining user data of deactivated accounts, and a lack of control for news feed content. One commenter explained:

> "*I used to use Facebook since the early days... But then I deleted it... Got sick of the privacy issues, having my personal information being sold for money (while I get no benefit from it)...*"

In terms of *user experience*, people complained about poor user interface, unsolicited friend suggestions, tagging non-users in photos, unwanted game requests, improper content blocking, irrelevant ads, inaccurate face recognition in photo tags, etc.:

"*. . . I know a LOT of people that have quit Facebook. . . Why? because FB has made it an annoying piece of crap. you cant block all stupid game requests by default. . . Facebook utterly sucks compared to a year ago. . . It's fricking turning into MySpace.*"

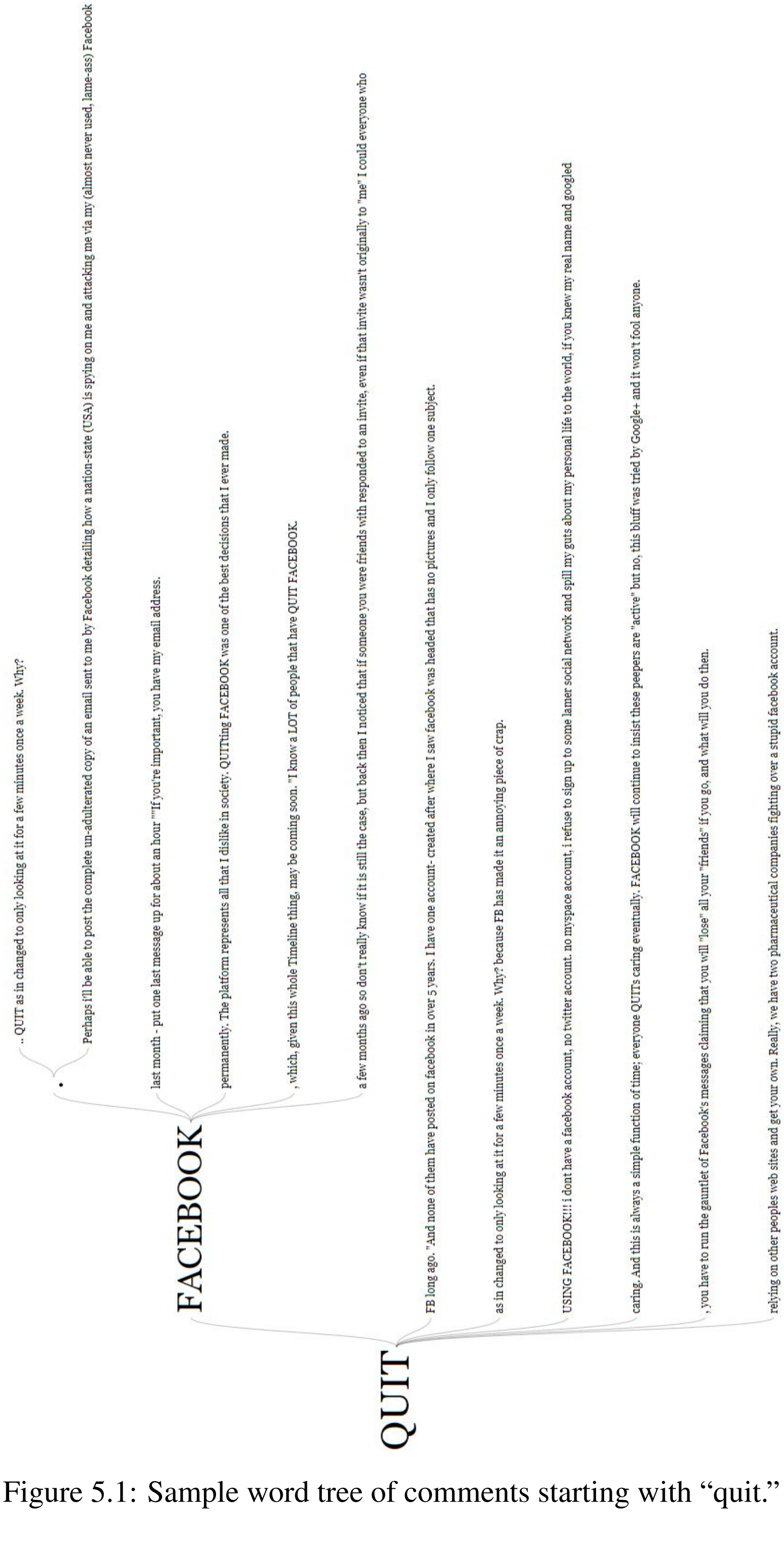

Figure 5.1: Sample word tree of comments starting with "quit."

Figure 5.2: Sample word tree of comments starting with "stop."

CHAPTER 6

RESULTS OF AUTOMATED ANALYSIS

In this chapter, we will present the findings from the automated analysis described in Chapter 4.

6.1 Wordcloud

We performed multiple wordcloud based analysis on our dataset for initial exploratory understanding. The unigram wordcloud that was made on overall *negative sentiment* related comments (regardless of theme) is shown in Figure 6.1(a). Drawing a bigram (Figure 6.1(b)) and trigram (Figure 6.1(c)) helped us surface further innate emotions and reasoning of Facebook haters.

A closer look at the unigram wordcloud reveals predominant words like "social," "privacy," "fake," "ad," "data," "news," "personal," "security," "stopped," "deleted" which are associated with comment phrases like "social media," "privacy and security concerns," "fake news," "selling personal data," "stopped using," "deleted account," etc. It also contains sentiment terms, such as "shit," "fuck," "hate," etc. that reflect user aversion. The

(a) Unigram wordcloud (b) Bigram wordcloud (c) Trigram wordcloud

Figure 6.1: Visualization of 2,364 *negative sentiment* comments.

key terms in the bigram wordcloud in Figure 6.1(b) appear to be "fake news," "stop using," "news feed," "personal information," "privacy setting," "delete account." Most visible key phrases of trigram from Figure 6.1(c) are: "i don't use," "i don't have," "don't use Facebook," "don't have a," "i don't care," "stop using Facebook." These visualizations are informative since they give us a top level idea of if people are rejecting Facebook and why.

6.2 Topic Modeling

Our trained LDA model produced 8 topics. Each topic is associated with a weight-distributed set of words. For each weight w_i, the weights are normalized such that, $\sum_{i=1}^{n} w_i = 1$. Table 6.1 lists each topic associated with the word distribution and few selected highly representative words from that distribution those the authors found a match while coding the comments.

To confirm the results obtained from LDA model, the topics found from Non-negative Matrix Factorization (NMF) were compared. Table 6.2 summarizes the top 5 topic words of each topic (Highlighted are the most intereseting and relevant topic words.) The interpretation of LDA and NMF topics reported in Table 6.1 and Table 6.2 also supports our findings so far. Top words associated with each topic assumes the corpus is a mixture of these topics. Table 6.1 shows the topic words useful for our analysis; for example "hate," "privacy" in (Topic 1); "news," "ads," "block" in (Topic 2); "data," "private" in (Topic 3); "privacy," "information" in (Topic 4); "stopped," "left," "dropped" in (Topic 5); "social," "personal" in (Topic 6); "deleted," "fake," "stop" in (Topic 7); "money," "problem" in (Topic 8), etc. The NMF based approach (Table 6.2 (highlighted)) extracts somewhat similar topics other than few exceptions like "networking," "friend," "private," "hell," "shit," "hate," "share," etc. The repetition of similar words indicates that the manual coding criteria adopted by the coders and automated analysis tend to converge.

Topic #	Word Distribution	Representative Words
1	0.008*“app” + 0.005*“company” + 0.004*“reason”+ 0.004*“feed” + 0.004*“speech” + 0.004*“apps” + 0.004*“data” + 0.004*“privacy” + 0.004*“hate” + 0.004*“always”	hate, privacy
2	0.012*“social” + 0.008*“stop” + 0.007*“news” + 0.006*“privacy” + 0.006*“ads” + 0.005*“device” + 0.005*“friend” + 0.004*“block” + 0.004*“seen” + 0.004*“political”	news, ads, block
3	0.011*“social” + 0.009*“information” + 0.007*“stuff” + 0.006*“data” + 0.004*“mean” + 0.004*“provide” + 0.004*“source” + 0.004*“private” + 0.004*“government” + 0.004*“looking”	data, private
4	0.010*“privacy” + 0.007*“email” + 0.006*“private” + 0.006*“service” + 0.006*“ads” + 0.005*“social” + 0.005*“profile” + 0.004*“information” + 0.004*“world” + 0.004*“name”	privacy, information
5	0.008*“privacy” + 0.007*“union” + 0.007*“social” + 0.005*“data” + 0.005*“stopped” + 0.005*“personal” + 0.004*“access” + 0.004*“code” + 0.004*“dropped” + 0.004*“left”	stopped, dropped
6	0.013*“information” + 0.010*“social” + 0.009*“personal” + 0.007*“data” + 0.004*“play” + 0.004*“upload” + 0.004*“online” + 0.004*“number” + 0.004*“take” + 0.004*“face”	social, personal
7	0.011*“news” + 0.006*“deleted” + 0.005*“tag” + 0.005*“information” + 0.005*“articles” + 0.004*“fake” + 0.004*“business” + 0.004*“stop” + 0.004*“privacy” + 0.004*“free”	deleted, fake
8	0.008*“social” + 0.007*“data” + 0.007*“news” + 0.005*“money” + 0.005*“might” + 0.004*“problem” + 0.004*“us” + 0.004*“making” + 0.004*“information” + 0.004*“last”	money, problem

Table 6.1: LDA topic modeling on the dataset.

6.3 Sentiment Analysis

Initially we performed sentiment analysis with VADER tool on 1,342 negative sentiment comments of `Dataset1`. Then we looked into each of the ten themes to see how the user sentiment compares according to comments belonging to each theme. The VADER tool returns four different measures of sentiment: *pos*, *neg*, *neu* and *compund score*. The compound score can be found by summing the valence scores of each word in the lexicon,

Topic #	Topic Words
1	**social** networking network face email
2	**news fake** feed articles trump
3	**friend info** word profile mean
4	**privacy** information **private personal** public
5	**data** business personal good fact
6	app installed **ads** guess android
7	accounts **stopped hell** play platform
8	**shit hate share** service public

Table 6.2: NMF topic modeling on the dataset.

adjusted according to the rules. Then it is normalized between -1 to 1, where -1 being extreme negative and +1 being extreme positive. Typically the threshold adopted by researchers [118] is: *compound score* $\geq$ 0.5 is positive; *compound score* $\leq$-0.5 is negative and anything between -0.5 and 0.5 is neutral. Values of *pos*, *neg* and *neu* can be used instead if multidimensional sentiment analysis of a sentence is desired, where they indicate the percentage of the text that fell into those sentiment categories. In Figure 6.2, we show the box plots of compound sentiment scores for ten themes as well as the overall set of comments side by side.

The sentiment box plots (Figure 6.2) reveal median compound scores overall and for each theme. The overall distribution of scores is fairly symmetric (mean of .007) with 624 out of 1,342 (46.51%) comments are marked as negative by VADER. This gives a rough idea of overall sentiments of all comments, however, that does not tell the distribution of positive/negative comments according to each thematic category. We can see that out of all thematic categories, *Privacy & Security* has the lowest median score (negative) which means comments of this theme are more associated with negative sentiment compared to any other theme (the lower the score, the more negative it is). It is expected because these comments are associated with slang and swearing terms. This category also holds the most negative score or most negative comment (score -0.97). Although the box plot has long tail on the right (indicating few positive scores), the median is closer to third quartile so

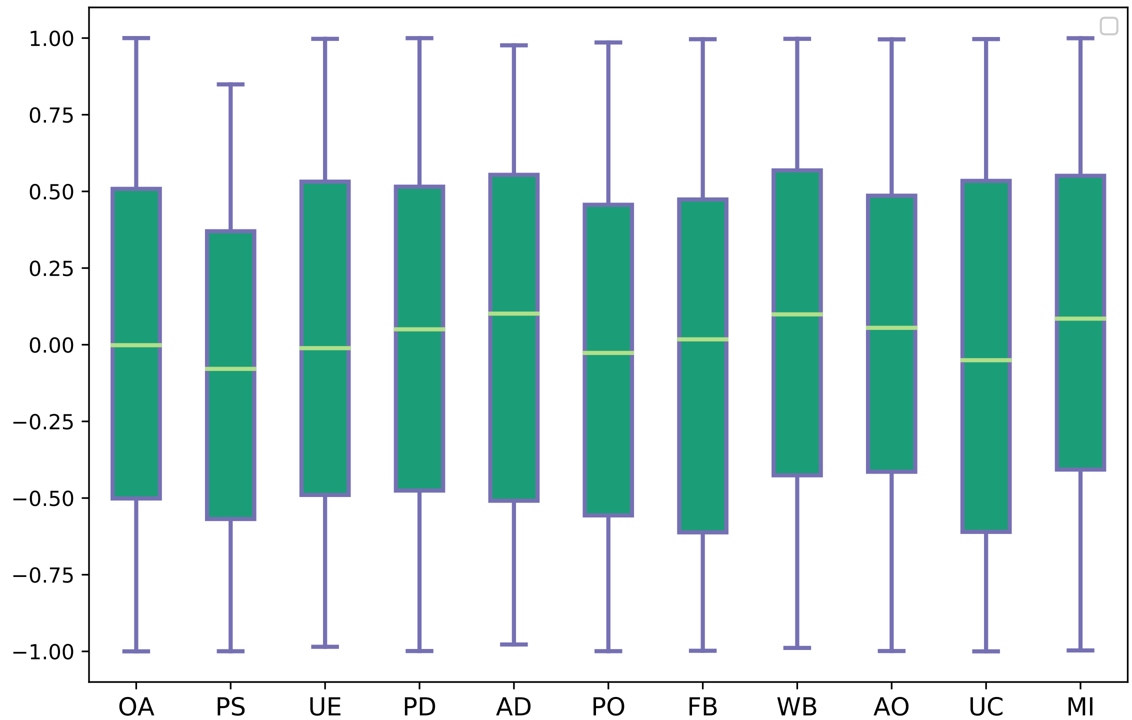

Figure 6.2: Box plots of compound sentiment scores of each theme using VADER tool. (OA = Overall, UE = User Experience, PD = Personal Disposition, AD = Advertisements, PO = Politics, FB = Fake Accounts & Bots, WB = Psychosocial Well-being, AO = Alternate Options, UC = Uninteresting Content, MI = Miscellaneous, PS = Privacy & Security)

there is no conclusive evidence on the skewness. In general, other themes that have median scores below 0 (meaning more negative) include *User Experiene*, *Politics*, and *Uninteresting Content*. The most positive theme with a mean score of 0.14 is *Advertisements* and that is explained by the fact that most of the comments of this theme do not involve harsh criticism or negative terms (just personal feelings). The box plot also indicate that the scores of this theme are skewed to left with $mean(0.14) < median(0.17)$. The same argument also holds for *Psychosocial Well-being* and *Miscellaneous* which have relatively higher compound score.

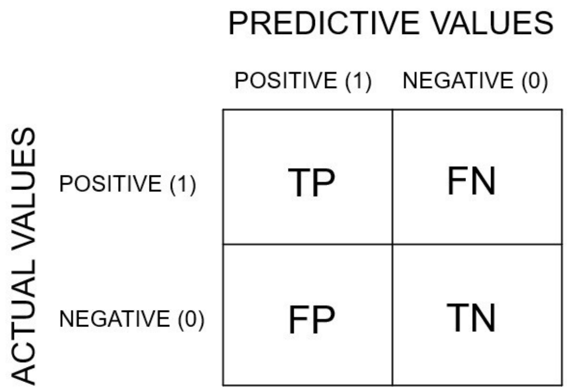

Figure 6.3: Confusion matrix. TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

6.4 Text Classification

Note that, our task is a binary classification on `Dateset2` comprising of 5,000 Facebook related comments. The target label is either *related to negative sentiment* or *not related to negative sentiment*. After the sentiment coding, it was observed that 2,364 comments were hand annotated as *related to negative sentiment* which implies this is not a perfectly balanced dataset. Formally, we assume that we have a set of N training examples $(c_i, y_i)_{i=1}^{n}$, where c_i is a comment and y_i is a binary label indicating whether c_i is a comment related to negative sentiment of Facebook. With a set of feature functions, we can represent c_i by a feature vector $\mathbf{x}(c_i)$ (which we refer to as x_i). We can then use standard classification algorithms to learn a classifier from $(\mathbf{x}_i, y_i)_{i=1}^{n}$. This classifier can be used to predict y for any unseen c. In this section we will explain the evaluation metrics and the outputs of different classification algorithms.

6.4.1 Evaluation Measures

There are multiple measures of the performance of machine learning classifiers for example, accuracy, precision, recall, F-1 score, etc. For a comprehensive understanding we need to understand the confusion matrix first (refer to Figure 6.3).

Here are some important terminologies [130]:

- **True Positives (TP)**: The cases in which we predicted YES and the actual output was also YES.
- **True Negatives (TN)**: The cases in which we predicted NO and the actual output was NO.
- **False Positives (FP)**: The cases in which we predicted YES and the actual output was NO.
- **False Negatives (FN)**: The cases in which we predicted NO and the actual output was YES.

Accuracy for the matrix can be calculated by taking average of the values lying across the "main diagonal" i.e.,

$$\text{Accuracy} = \frac{\text{True Positives + False Negatives}}{\text{Total number of samples}}$$

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. It means of all comments that labeled as related to negative sentiment, how many actually are? High precision relates to the low false positive rate.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$$

Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes. It means, of all the comments that truly are negative sentiment related, how many did we label?

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$$

F1 score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if there is an uneven class distribution.

$$\text{F1 Score} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall+Precision}}$$

Area Under Curve(AUC) is another widely used metrics for evaluation. It is used for binary classification problem. AUC of a classifier is equal to the probability that the classifier will rank a randomly chosen positive example higher than a randomly chosen negative example. AUC has a range of [0, 1]. The greater the value, the better is the performance of our model. Since our dataset is imbalanced and has lesser positive samples it makes more sense to use *F1 score* and *AUC* as performance measures.

6.4.2 Results Analysis

In this subsection we will report all the findings from the different text classifiers. We experimented with all possible combinations of features and different models. For the purpose of brevity, we will try to keep the analysis limited to the best performing scenarios. The performance measures are determined from distinct 80%-20% train-test split and 10-fold cross-validation of the `Dataset2`. We have shown average accuracy, precision, recall, AUC considering *related to negative sentiment* class as the positive class. Unless mentioned otherwise, the appropriate measure for our dataset would be *F1 score* and *AUC*. We used *DummyClassifier* which is a classifier that makes predictions using simple rules, as the **baseline** model. Our dummy classifier follows stratified strategy i.e., generates predictions by respecting the training set's class distribution. The hyperparameter settings for the different models are given as:

```
DummyClassifier(random_state=3),
RandomForestClassifier(n_estimators=200,max_depth=3,
random_state=0),
SGDClassifier(random_state=3, loss='log'),
MultinomialNB(),
```

```
LogisticRegression(random_state=0),
```

There are multiple variables involved as part of the result analysis. For example, application of the text pre-processing, different set of features, different settings of hyperparameters. We did not experiment much with the hyperparameters, however, the effect of text cleaning (stemming, stopwordremoval, etc.) and feature set combinations were investigated.

With pre-processing

Here we will discuss the performances of different models for different feature vectors while the actual dataset was cleaned as described in Chapter 4. Table 6.3 gives the average binary classification performances for 10-fold cross validation.

From Table 6.3 it is evident that, all the models outperform the baseline dummy model easily. For Count Vector feature, Random Forrest classifier performs poorly (F- score 50%). While SVM, Naive Bayes, and Logistic Regression performs almost similar (SVM has an slight edge here). For all variants of TF-IDF based features, Logistic Regression clearly outperforms all other models in terms of both F-score and AUC. It is interesting to note that, Naive Bayes performs better on average than Random Forrest. The N-gram based TF-IDF has low F-scores for every model compared to other features. For this analysis, Logistic Regression appears to be the clear winner with SVM as runner-up followed by mutinomial Naive Bayes classifier.

Since *Logistic Regression* model on character level TF-IDF performs best with an *F-1 score* of 72% and *AUC* of 73.5% for the cleaned data, we wanted to investigate further. Figure 6.4 and Figure 6.5 shows the confusion matrix and Precision-Recall curve, respectively for this classifier. While the confusion matrix shows the exact number of times misclassification happened, Precision-Recall curve gives an idea how well the classifier discriminates between the classes.

Feature	**Model**	**P**	**R**	**F1**	**AUC**
Count Vector	Dummy	0.49	0.49	0.49	0.495
	RF	0.67	0.58	0.50	0.648
	SVM	0.68	0.68	0.69	0.676
	NB	0.68	0.67	0.66	0.711
	LR	0.67	0.67	0.67	0.699
TF-IDF(1)	Dummy	0.49	0.49	0.49	0.495
	RF	0.65	0.60	0.55	0.642
	SVM	0.68	0.68	0.68	0.719
	NB	0.71	0.68	0.67	0.717
	LR	0.70	0.69	0.69	0.741
TF-IDF(N)	Dummy	0.49	0.49	0.49	0.495
	RF	0.64	0.53	0.39	0.625
	SVM	0.64	0.63	0.62	0.664
	NB	0.65	0.65	0.64	0.692
	LR	0.64	0.64	0.64	0.701
TF-IDF(char)	Dummy	0.49	0.49	0.49	0.495
	RF	0.67	0.66	0.65	0.695
	SVM	0.67	0.67	0.67	0.714
	NB	0.66	0.64	0.62	0.694
	LR	**0.72**	**0.72**	**0.72**	**0.735**

Table 6.3: Average performances of different classifier on different features (cleaned data). P = Precision, R = Recall, AUC = Area Under the Curve.

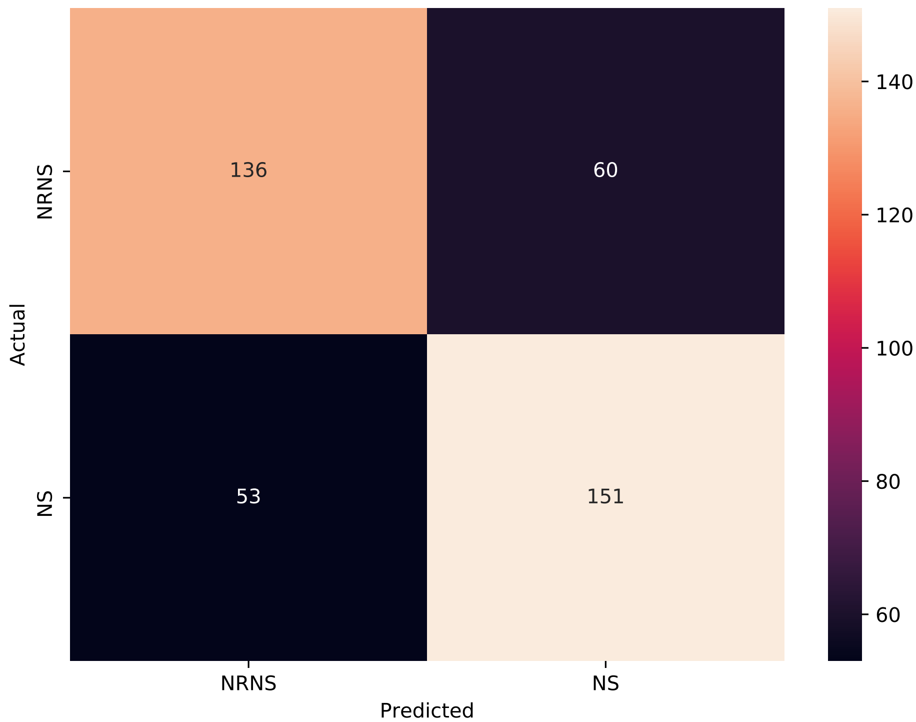

Figure 6.4: Confusion Matrix for Logistic Regression based on character level TF-IDF (cleaned data). Note that, NS = related to negative sentiment and NRNS = not related to negative sentiment.

Without pre-processing

Here we will discuss the performances of different models for different feature vectors while the actual dataset was not cleaned except stopwords removal. Table 6.4 gives the average binary classification performances for 10-fold cross validation.

From Table 6.4 it is evident that, all the models outperform the baseline dummy model easily. For Count Vector feature, Random Forrest classifier performs poorly again (F- score 56%). However, it has improved by 6% when we used the uncleaned data. For the other classifiers, the ranking is $LogisticRegression > NaiveBayes > SVM$. For all variants of TF-IDF based features, Logistic Regression clearly outperforms all other models in terms of both F-score and AUC in the uncleaned dataset as well. Once again, Naive Bayes

Feature	**Model**	**P**	**R**	**F1**	**AUC**
Count Vector	Dummy	0.49	0.49	0.49	0.495
	RF	0.67	0.61	0.56	0.657
	SVM	0.67	0.65	0.63	0.644
	NB	0.65	0.65	0.65	0.709
	LR	0.69	0.69	0.69	0.727
TF-IDF(1)	Dummy	0.49	0.49	0.49	0.495
	RF	0.66	0.65	0.64	0.712
	SVM	0.69	0.68	0.67	0.734
	NB	0.70	0.67	0.65	0.713
	LR	0.66	0.66	0.66	0.770
TF-IDF(N)	Dummy	0.49	0.49	0.49	0.495
	RF	0.62	0.59	0.57	0.65
	SVM	0.68	0.68	0.68	0.684
	NB	0.65	0.65	0.64	0.692
	LR	0.65	0.65	0.65	0.697
TF-IDF(char)	Dummy	0.49	0.49	0.49	0.495
	RF	0.69	0.68	0.68	0.749
	SVM	**0.71**	0.69	0.69	0.748
	NB	0.67	0.66	0.66	0.72
	LR	**0.71**	**0.71**	**0.71**	**0.776**

Table 6.4: Average performances of different classifier on different features (uncleaned data). P = Precision, R = Recall, AUC = Area Under the Curve.

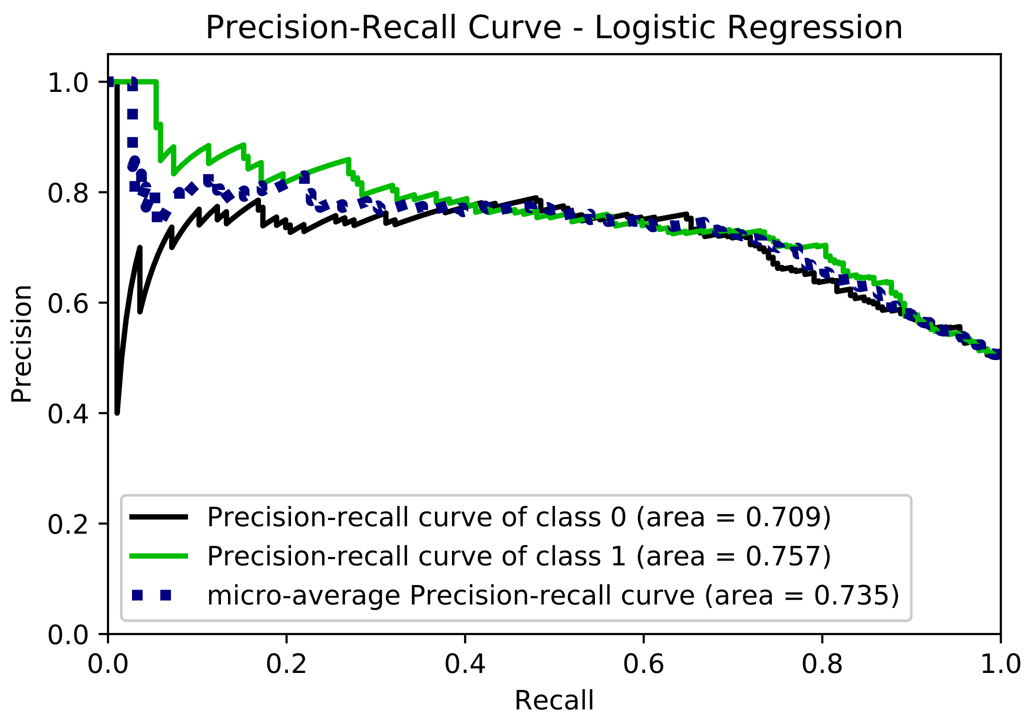

Figure 6.5: Precision-Recall curve for Logistic Regression based on character level TF-IDF (cleaned data)

performs better on average than Random Forrest. For this analysis too, Logistic Regression appears to be the clear winner with SVM as runner-up followed by mutinomial Naive Bayes classifier. Since *Logistic Regression* model on character level TF-IDF performs best with an *F-1 score* of 71% and *AUC* of 77.6% we wanted to investigate further. Figure 6.6 and Figure 6.7 shows the confusion matrix and Precision-Recall curve, respectively for this classifier.

Upon comparing Table 6.4 and Table 6.3, we did not notice a huge improvement with or without the data cleaning process. For example, the Logistic Regression classifier has seen improvement in terms of F-score with character level and unigram level TF-IDF after cleaning the data but it performed poorly for count vectors and n-gram level TF-IDF. Similar arguments can be made for SVM classifier too. In general, this is the ranking of the classifiers for our dataset over all the features (in terms of F-score and AUC):

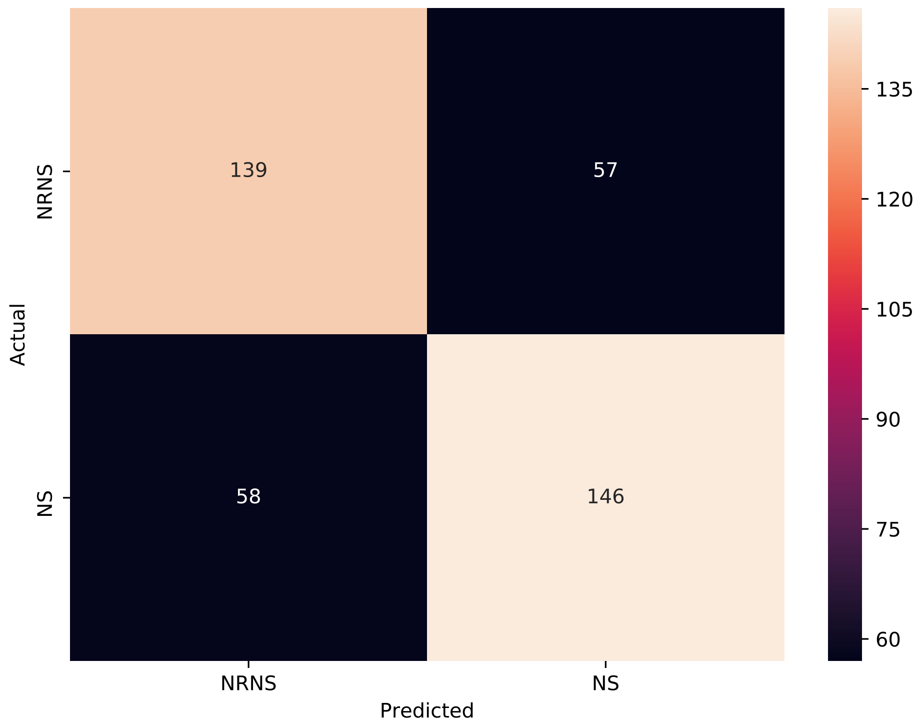

Figure 6.6: Confusion Matrix for Logistic Regression based on character level TF-IDF (uncleaned data). Note that, NS = related to negative sentiment and NRNS = not related to negative sentiment.

$$\text{Logistic Regression} > \text{SVM} > \text{Naive Bayes} > \text{Random Forrest}$$

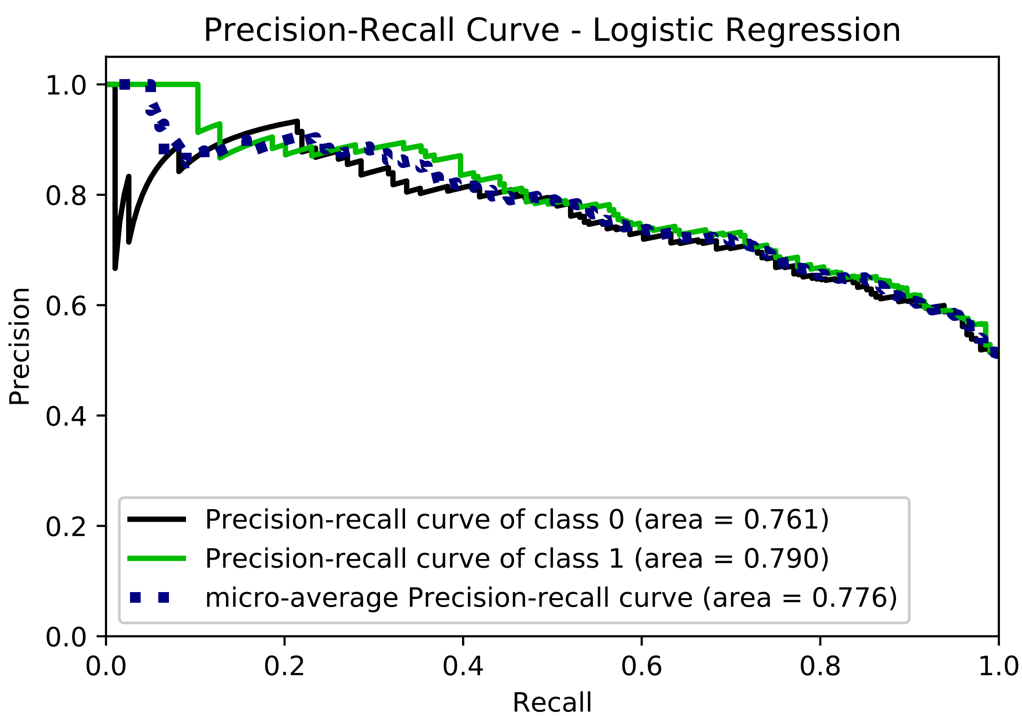

Figure 6.7: Precision-Recall curve for Logistic Regression based on character level TF-IDF (uncleaned data)

CHAPTER 7
DISCUSSION

7.1 Qualitative Coding

Our findings from the initial level of sentiment coding are very intriguing. Within our sample 4000 Facebook-related comments from Slashdot & 1000 from Schneier's blog, 51.1% & 32%, respectively were indicative of some form of Facebook negative sentiment and non-use. This shows the notably large propensity of experts to resist a system widely popular with the non-expert general population [131]. This discrepancy warrants an in-depth analysis of the factors that play pivotal role in such negative sentiment.

Considering all the Facebook-related comments of these two blogs belong to either one of the class *related to negative sentiment* or *not related to negative sentiment*, the probability of getting a Facebook non-user is nearly 47.28% in our data (which is very high considering the wide range of other topics belonging to the class *not related to negative sentiment*). This is striking given that tech savvy individuals are typically early and enthusiastic adopters of any technology. We believe that the high levels of negative sentiment and corresponding non-use can be attributed to our sample's deeper understanding of operational details that led them to recognize problematic aspects. Indeed, several of the illustrative comments we reported in the previous section refer to sophisticated technological matters, such as encryption, cookies, data mining, API, source code, etc. Moreover, in contrast to prior work, we did not find evidence of reversion [49], suggesting that the sentiments and convictions of our tech savvy sample may be stronger.

The second level thematic coding and further analysis attempt to answer our *RQ1*. The top 3 out of 10 common themes or primary influential factors for Facebook negative sentiment which emerged from our findings are: *privacy and security*, *user experience*, and *personal disposition*. This implies that users are dissatisfied with several aspects of the Facebook system. The myriad of *user experience* issues implies that Facebook is not wary of expert user's desired environment and they are often criticized for doing too much experimentation. A subset of these experiences were also observed in prior works [34, 7, 51], but this is in bigger scale and sometimes with more technical reasoning for our data. Apparently users do not find Facebook to be a reliable form of storage for their personal data due to variety of reasons. According to them, Facebook treats the user content as their property and they sell it for their sake. It should be noted that current users are particularly concerned about privacy violations by Facebook, such as data mining based on search history, compared to early days. Does this affect the way people interact with Facebook and how much does it hurt FB itself? Well, although there is no immediate Facebook collapse looming, incidents like the latest Cambridge Analytica debacle can haunt them if remedial measures are not taken. But why are tech-savvy users so prone to Facebook's privacy violation and security loopholes? A possible explanation is: not necessarily their background knowledge is associated with their secure behavior online [132]. *Personal disposition* accounts for almost 17.95% of the combined dataset. This category of comments often includes users castigating Facebook (e.g., by using slang or swearing terms, such as "fuck," "shit," "creepy," "sucks," etc.) but did not really point out why. An interesting discussion that came up in this category comments was: people don't actually think Facebook friends are real friends rather they underscore the importance of getting out and socializing in real life (Lampe et al. [34] described this as *channel effect*).

Careful inspection of 118 total comments of *advertisements* theme revealed that Facebook's business model of showing "targeted ads" was absolutely reprimanded by the users. They blamed Facebook for the selling of their personal information to third party or government funded surveillance operations. Consistent with the findings by Baumer et al. [49], some of the users found some non-Facebook media use is more user friendly or a better mode of communication compared to Facebook (e.g., Slack, Instagram, Snapchat, Reddit, etc.). Such comments indicated that it does not necessarily has to be another social media rather they mentioned that mere emails can be a good enough alternative. We observed *uninteresting content* (described as *banality* by Baumer et al. [7]) of Facebook; such as fake friends, distasteful confrontation among members of any Facebook group, presence of parents in friend list (also explained as *disenfranchisement* [56]), deliberate misinformation, cyberbullying, overuse of abusive language, religious bigotry, etc. caused significant Facebook criticism. We believe the over exposure to such non-relevant contents and ostentatious display of personal details of friends on Facebook cause damaging psychological impact. The most cited ones are depression, addiction to social media, etc., which is consistent with the findings of prior work [133, 134]. One possible explanation of this is to maintain a positive impression management on Facebook [135] or that Facebook users are primarily extroverted and narcissistic [46]. Another form of non-use, such as lurking, was observed in our data when users expressed why they are compelled to use Facebook against their will because they care about their Facebook friends and they simply find no other option but to use it since relatives/friends/employers take it for granted that everyone has to have an Facebook account. Baumer et al. [7] described such an event as *lagging resistance*. Among all these categories the common theme expressed by dissatisfied users is the fact that the cost of adopting the technology is simply outweighing the perceived benefits. These findings are congruous with Technology Acceptance Model [136, 137]. Apparently the users are risking their privacy, time, relationships and in return they are not getting enough return on investment. A more nuanced analysis on the risk factors and the threshold to which users agree to take the risk without compromising privacy and other factors could be a good future work.

7.1.1 Comparison between Slashdot and Schneier's Blog

Besides getting additional insight and data, the motivation to use two separate data sources was to compare between the findings of two different sets of users. We observed most

Slashdot comments are shorter in length (average 358 characters) compared to Schneier's comments (average 984 characters), they are typically direct in nature and sometimes lacks contextual information. The 3 most prominent themes are same for both datasets but in slightly different order (probably because of the difference in perspectives). It turns out the major source of concern is as expected: *privacy and security*. It is not surprising to see though the percentage of comments for Schneier's blog (66.25%) is higher than that of Slashdot (42.27%) since the former is predominantly a security blog and most discussed topics are state-of-the-art technology security issues. The interesting part is the second prominent theme. While it is *user experience* (33.36%) for Slashdot, Schneier's blog users expressed *personal disposition* (29.37%). It indicates that Schneier's blog users may be less concerned about *user experience* (12.81%) compared to their dissatisfaction with Facebook's handling of user privacy. Most of the privacy & security related concerns were common in both datasets. Interestingly, some of the commenters on Scheneir's blog believe that Facebook is covertly cooperating with the NSA and the CIA. It should be noted that for Slashdot, *personal disposition* was the third most cited theme (14.38%). A couple of other significant observation that can be made from Table 5.1 are: first, *psychosocial well-being* is not well represented compared to other themes and second, commenters of Schneier on Security are less troubled by *politics* (1 comment) compared to Slashdotters (30 comments). The explanation for this and nuances between the other themes for these 2 different sets of audiences require demographic information and further data.

7.1.2 Explicit Non-use Only Comments Analysis

311 out of 1022 negative sentiment comments (30.43%) of Slashdot and 161 out of 320 negative sentiment comments (50.31%) of Schneier's blog were marked as *explicit non-use*. The significant difference in percentages could be due to either differences in the sizes of samples or because commenters of Schneier's blog are comparatively more inclined to non-use in general. A closer look at the distribution of themes for *explicit non-use* only

comments (Table 5.1) reveal similar ordering (top 3) for both datasets as described above. As insinuated by Stieger et al. [138], *privacy and security* is the prime reason for committing "virtual identity suicide" via Facebook non-use. A majority of the themes echo the findings of prior studies [34, 7, 51, 129]. Table 7.1 describes the commonalities in terms of themes with previous studies on non-use. It should be noted that, such confirmation of prior results helps verify robustness (or datedness) in light of rapidly changing technology (especially social media) and corresponding societal changes as well as user learning/experience. Although these themes are common, some comments belonging to these themes contain subtle technical details. one such comment: *"...If they give this data to some advertiser (provided it's anonymized), it still doesn't concern me. I'd actually prefer to see ads that are relevant to my interests than not. If they provide the full, non-anonymized dataset to some third party though, that's where the concern comes in... "* This is a clear case where we see technical expertise and a concern from technical point of view.

Theme	**Lampe et al. [34]**	**Baumer et al. [7]**	**Schoenebeck [129]**	**Baker et al. [51]**
Privacy & security	Privacy	Privacy, Data use and misuse		Cybersafety concerns
Uninteresting content	Lack of interest	Banality		Lack of motivation
Alternate options	Channel Effect		Tradeoffs of not spending time elsewhere	Preference for other forms of communication
Psychosocial well-being		Addiction, withdrawal, and envy of the disconnected		
Personal disposition	Time constraint, Context collapse	Productivity	Spending too much time on social media	Poor use of time

Table 7.1: List of common themes with prior work.

Our findings thus indicate that the factors that lead to non-adoption and non-use have do not yet seem to have been properly addressed. In fact, we provide additional drivers of non-use connected to recent technological and societal developments, such as "fake

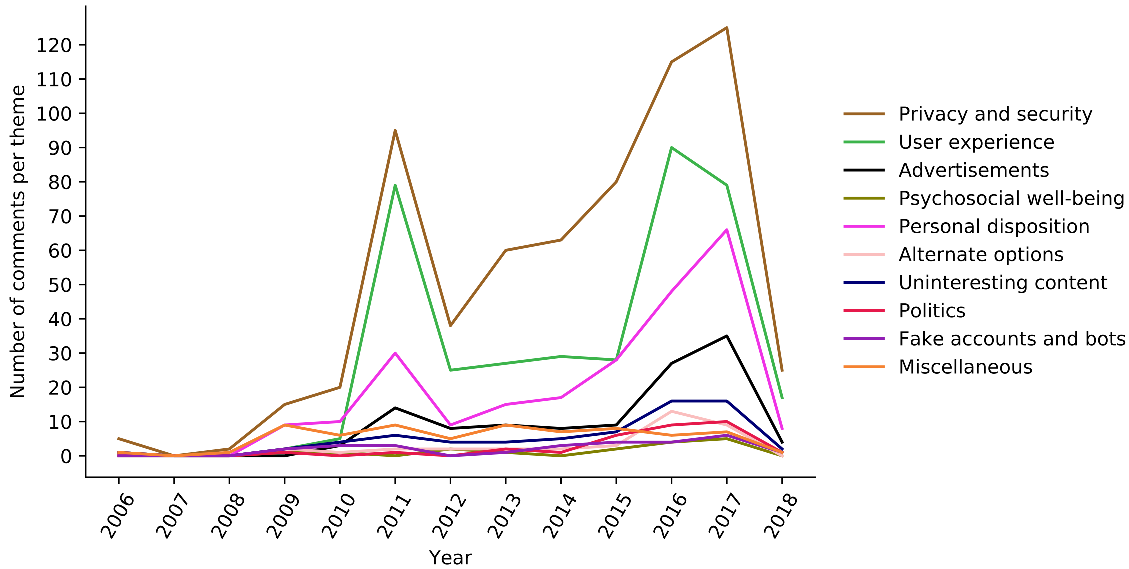

Figure 7.1: Number of comments per theme by year.

news," political arguments and polarization, personalized and targeted content, etc. While interpersonal privacy, such as privacy from the employer, was mentioned, we found a far greater mention of consumer privacy and government and corporate surveillance. Taken together these factors seemed to raise privacy and security concerns and degrade the user experience, thus increasing people's proclivity to view Facebook in a negative light and reducing or abandoning its use. These trends can be observed in Figure 7.1 that shows the number of comments from both data sources combined that fell under each theme per year. While Figure 7.1 provides a quick sense of temporal patterns, we should stress that the absolute numbers cannot be taken as statistically meaningful owing to our random sampling and qualitative coding techniques.

Our observations for both dataset are consistent with the concept of *lagging adoption* [56] and *diffusion theory* [58] where experts are on the left of technology adoption bell curve and non-users on the other side. Those who willingly did not join Facebook have strong reasons for doing so and cannot be explained by the notion of "laggards [56]". Similarly, other forms of non-use, such as *disenchantment* and *disinterest* [56] can be explained by *uninteresting content* and *personal disposition* respectively. Although withdrawal symptoms [49] were acknowledged by some, we did not find evidence of notable reversion. For example, one commenter wrote: *"I deleted my Facebook account years ago and I don't miss it at all. It's a great tool for companies to exploit consumers though, that's probably its main purpose these days."*

7.2 Results of the Text Classifier

Although compared to state-of-the art sentiment classifiers our classifier performance is not well matched but it has to be noted that our task was not mere sentiment classification and hence the results can not be compared as such. The low performance measures can be attributed to: first, inherent fuzziness of the data (often ironic and not directly harsh) and second, the lack of labeled data for training. There is no such existing work that investigated the classification of non-use comments. From that perspective, our findings shed light on *RQ2* and are exciting starting point for future work. One particular study on cyberbullying detection which resembles our work, dealt with similar fuzzy commentset and suffered from lack of sample [96]. After feature enginerring and hyperparameter optimisation of their models, a maximum F1 score of 64.32% was achieved. So the performance of our classifier is promising if not excellent.

As discussed in Chapter 6, character n-gram based TF-IDF feature achieved highest measures. It might be because they are likely to provide robustness to lexical variation in social media text, as compared to word n-grams [96]. In general, the Random Forrest classifier failed to perform well and this was possibly because in order to achieve good performance for a ensemble classifier like Random Forrest, extensive hyperparameter tuning is required. It was observed often that the data cleaning lowered performances for some classifier. A deeper look into this revealed that, sometimes stopwords and punctuation removal from the comments discarded important information while vectorizing and resulted into misclassification. We performed a qualitative error analysis on the misclassified comments for the Logistic Regression classifier with character level TF-IDF features. The observation was such misclassifications ocurred in scenarios when, in general, either one of the following was true:

- the comment is smaller in length.
- the comment is ironic or not explicit in nature . For example,

  > "*Makes one wonder what a Facebook account is really worth to a company (or pop group, or artist, whatever). On the one hand, the option of gaining & holding customers, and do lots of PR through the social network, on the other hand the possibility that at any time, if someone with same name (competitor?) creates a dispute about it, Facebook might close the account for no good reason.*"

  This comment was classified as *not related to negative sentiment* while it actually points to a criticism. Another example of misclassification was:

  > "*What are your Facebook privacy settings set to? All information public? If not, surely you don't having anything to hide? They may not be crimes, but still.*"

  This example was classified as *related to negative sentiment* maybe because the comment contains the term 'privacy,' and there are so many positive samples that contain the same term.

- during the sentiment coding phase one of the coder marked it otherwise (remember that for high agreement rate we counted them in).

To improve the accuracy, first and foremost we need more training data (either by hand labeling or rule based matching). A better hyperparameter tuning and aggregation of more nuanced features will guarantee better output. Also once we have more data we can try Deep Learning based approaches which have been proved to be better suited for representing fuzzy data with different word embeddings.

7.3 Implications of the Findings

The implications of the findings are manifold: from the perspective of researchers, users, service providers, and system designers [41]. Our findings matter because the common psychology in practice is that people go with the trend or they tend to follow the crowd [23]. Intention and sentiment are strong indicators of behavior [139]. As illustrated by our findings negative sentiment is thus likely to drive and indicate non-use. Moreover, expressed negative sentiment may strongly *imply* non-use even if it is not explicitly mentioned. Slashdot and Schneier's blog users are early adopters and their knowledge stems from deeper understanding of technology. Sentiments of these techies are important because: first, inclusive design should consider expert users. Second, people seek advice from expert family/friends. Third, with learning/experience, today's non-experts may be tomorrow's experts. Finally, experts do influence non-experts (e.g., op-eds/reviews) as well as policy makers (e.g., testimony, lobbying). Schneier writes op-eds and is often consulted by lawmakers.

Our *RQ1* results have implications on the way future systems could be designed and existing systems could be evaluated. This way product business related blog/forums could be benefited too because such blogs contain high volume unprompted reviews from users (e.g., Yelp, Amazon). Our approach will enable them to understand user behavior or sentiment and adjust their business model accordingly. We believe the most important implication for our *RQ1* results is practical. For instance, a Facebook system analyst can go through the finding of this work to get a top level idea of what is affecting user engagement and ways to fix them. Some of the expressed concerns, such as *User Experience*, could be addressed via design improvements. On the other hand, other aspects, such as targeted advertising, data sharing, etc., will most likely require improved public policy and regulations along with corresponding enforcement mechanisms. The European Union's (EU) General Data Protection Regulation (GDPR) takes a step in that direction but applies only to EU citizens.

It has to be understood that technology use is more a cultural phenomenon [56] so the study of non-use and use has to be considered from that perspective. Our findings do not necessarily point to a radical design change or complete disposal rather show what are design flaws and what changes can incorporate more users. Many users suggested how staying away from social media for a while helped them to better balance their social life. So Facebook designers can consider non-use positively and possibly facilitate it by allowing the users to take temporary breaks, strategies to balance user interaction with social media [129] and enabling more self-inhibiting options without 'undesigning' it [140]. As Baumer and Silverman [141] argued, designers should consider the implication when not to design [142] but this does not promote rejection of technology. One Slashdotter addressed this issue in his comment: "*You never really leave Facebook. I thought I had deleted my account but signed back in a month later and it was still there ...*"

It is evident that Facebook currently disregard the ongoing movement in terms of *non-use*. Our findings have critical implications for UI designers of Facebook in this regard. Incorporating ideas from non-users guarantee an engaged user-base and can bring positive sociotechnical changes [7]. Because today the non-users may not be an overwhelming majority but their perspective should be valued and dealt with [60, 56], especially if they are tech-savvy users. A more deeper analysis will reveal other system pitfalls and ways to overcome them to create an all inclusive environment that is devoid of the factors discovered in our findings.

CHAPTER 8
CONCLUSION

8.1 Limitations

We randomly sampled 4000 comments from Slashdot and 1000 comments from Schneier's blog and tried to answer our questions from those comments only. It is arguable that these comments are not representative of the whole mass population; therefore if the findings can be concluded as generic with same level of certainty. The whole point of random sampling was to get unrelated and unbiased data. We acknowledge the fact that there are more negative sentiment motives than what we have discovered due to the lack of sample. While analyzing the data, we noticed that the comments often involve multiple words including positive and negative words, however, the sarcastic tone altered the meaning completely. It is the inherent ambiguity in the human language that makes the analysis unclear at times which we tried to overcome to the best of our discretion. Moreover, while comments are indeed influenced by the original story, we coded for general sentiment/practices, not short-lived reactions. That said, even short-term negative sentiment and non-use can be informative.

One point of argument can be what is the explicit definition of "expertise" since we claimed Slashdotters and Schneier's blog users are tech-savvy expert population. There is no way of confirming that unless we have the users profile which is not feasible, but it is the type of blog and the nature of their discourse that is the core of our assumption here. That said, we do not have explicit measures of technical expertise and skills of individual commenters. Moreover, Slashdot and Schneier on Security receive comments from people all over the world. We do not know the extent to which our findings may be impacted by such cultural variation. Given that a majority of the comments on both sites appear to be based in the US, we estimate that the impact is likely to be minimal.

8.2 Future Work

Our findings engender interesting potential research questions: do we get similar user sentiments towards other non-Facebook privacy sensitive media? Does incorporating more data and application of unsupervised machine learning-based text classification approach guarantee interesting results? For our future work we want to explore other sources where users discuss their Facebook user experience to get a comprehensive understanding. In particular, this will be interesting to analyze if the underlying factors found in this study from expert users are consistent with other data sources consisting of general audience. This however, is not feasible to answer in a single thesis, since everyday more and more such platforms are emerging. Also, analysis of a particular site provides answers to problems of specific audience. Unsupervised machine learning-based approaches, such as clustering, topic modeling, and opinion mining can be applied to detect inherent motives automatically. Also, we would like to extend the knowledge gathered from this study to address more generic agenda, such as overall social media (not only Facebook) sentiment mining. We want to use more hand annotated data for future analysis as it will not only increase the accuracy but also help getting our domain specific sentiment.

8.3 Concluding Remarks

This thesis aims to address an under researched issue in the field of user interaction with technology, namely, Facebook negative sentiment and non-use. We believe this is imperative to understand the mindset of tech-savvy non-participants and the factors that play a pivotal role in this regard. We collected user-generated comments from a technology blog called Slashdot and a security blog called Schneier on Security and through qualitative coding, acquired significant responses related to negative sentiment. Further analysis unveiled several user dissatisfaction related to Facebook, for example, flaws in Facebook's architecture, privacy and security concerns, personal issues, etc. Our findings have manifold implications for non-expert users, system designers, and business analysts in designing and evaluating a reliable, fail-safe system. The text classification based automatic identification of non-use comments yielded promising findings (77.6% AUC) too. Besides, our work invokes several interesting questions, for example, does availability of more data enable us to perform temporal (time-series) sentiment analysis on non-participants? How do the same techniques apply to other sources of data comprised of diverse population (general and tech-savvy)?

REFERENCES

[1] S. Lohr, "The age of big data," *New York Times*, vol. 11, no. 2012, 2012.

[2] D.-H. Shin, "The effects of trust, security and privacy in social networking: A security-based approach to understand the pattern of adoption," *Interacting with computers*, vol. 22, no. 5, pp. 428–438, 2010.

[3] M. Madden, "Privacy management on social media sites," *Pew Internet Report*, pp. 1–20, 2012.

[4] K. Koroleva, H. Krasnova, N. Veltri, and O. Günther, "It's all about networking! empirical investigation of social capital formation on social network sites," in *In Proceedings of the International Conference on Information Systems (ICIS 2011)*, 2011.

[5] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of facebook "friends:" social capital and college students' use of online social network sites," *Journal of Computer-Mediated Communication*, vol. 12, no. 4, pp. 1143–1168, 2007. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1083-6101.2007.00367.x`.

[6] A. N. Joinson, "Looking at, looking up or keeping up with people?: Motives and use of facebook," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '08, Florence, Italy: ACM, 2008, pp. 1027–1036, ISBN: 978-1-60558-011-1.

[7] E. P. Baumer, P. Adams, V. D. Khovanskaya, T. C. Liao, M. E. Smith, V. Schwanda Sosik, and K. Williams, "Limiting, leaving, and (re)lapsing: An exploration of facebook non-use practices and experiences," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13, Paris, France: ACM, 2013, pp. 3257–3266, ISBN: 978-1-4503-1899-0.

[8] B. Nonnecke and J. Preece, "Why lurkers lurk," Jan. 2001.

[9] M. S. Ackerman and D. W. McDonald, "Answer garden 2: Merging organizational memory with collaborative help," in *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work*, ser. CSCW '96, Boston, Massachusetts, USA: ACM, 1996, pp. 97–105, ISBN: 0-89791-765-0.

[10] S. Kiesler, J. Siegel, and T. W. McGuire, "Social psychological aspects of computer-mediated communication.," *American psychologist*, vol. 39, no. 10, p. 1123, 1984.

[11] B. Wellman, A. Quan-Haase, J. Boase, W. Chen, K. Hampton, I. Díaz, and K. Miyata, “The social affordances of the internet for networked individualism,” *Journal of Computer-Mediated Communication*, vol. 8, no. 3, pp. 0–0, 2003.

[12] P. G. Lange, “What is your claim to flame?” *First monday*, vol. 11, no. 9, 2006.

[13] P. D. Duffy and A. Bruns, “The use of blogs, wikis and rss in education: A conversation of possibilities,” 2006.

[14] K. Pinkman, “Using blogs in the foreign language classroom: Encouraging learner independence,” *The Jalt CALL Journal*, vol. 1, no. 1, pp. 12–24, 2005.

[15] J. B. Williams and J. S. Jacobs, “Exploring the use of blogs as learning spaces in the higher education sector,” *Australasian journal of educational technology*, vol. 20, no. 2, pp. 232–247, 2004.

[16] R. Davis, *Typing politics: The role of blogs in american politics*. Oxford University Press, 2009.

[17] R. Koop and H. J. Jansen, “Political blogs and blogrolls in canada: Forums for democratic deliberation?” *Social Science Computer Review*, vol. 27, no. 2, pp. 155–173, 2009.

[18] S. Gaonkar, J. Li, R. R. Choudhury, L. Cox, and A. Schmidt, “Micro-blog: Sharing and querying content through mobile phones and social participation,” in *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys ’08, Breckenridge, CO, USA: ACM, 2008, pp. 174–186, ISBN: 978-1-60558-139-2.

[19] E. Khabiri, C. Hsu, and J. Caverlee, “Analyzing and predicting community preference of socially generated metadata: A case study on comments in the digg community,” in *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA, May 17-20, 2009*, 2009.

[20] R. E. Ferdig and K. D. Trammell, “Content delivery in the ‘blogosphere’,” *THE Journal (Technological Horizons In Education)*, vol. 31, no. 7, p. 12, 2004.

[21] S. C. Herring, I. Kouper, J. C. Paolillo, L. A. Scheidt, M. Tyworth, P. Welsch, E. Wright, and N. Yu, “Conversations in the blogosphere: An analysis “from the bottom up”,” in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 2005, 107b–107b.

[22] N. Hookway, “Entering the blogosphere’: Some strategies for using blogs in social research,” *Qualitative research*, vol. 8, no. 1, pp. 91–113, 2008.

[23] E. Gilbert, T. Bergstrom, and K. Karahalios, “Blogs are echo chambers: Blogs are echo chambers,” in *2009 42nd Hawaii International Conference on System Sciences*, 2009, pp. 1–10.

[24] J. Kunegis, A. Lommatzsch, and C. Bauckhage, “The slashdot zoo: Mining a social network with negative edges,” in *Proceedings of the 18th International Conference on World Wide Web*, ser. WWW ’09, Madrid, Spain: ACM, 2009, pp. 741–750, ISBN: 978-1-60558-487-4.

[25] V. Gómez, A. Kaltenbrunner, and V. López, “Statistical analysis of the social network and discussion threads in slashdot,” in *Proceedings of the 17th International Conference on World Wide Web*, ser. WWW ’08, Beijing, China: ACM, 2008, pp. 645–654, ISBN: 978-1-60558-085-2.

[26] A. O. Baoill, “Slashdot and the public sphere,” *First monday*, vol. 5, no. 9, 2000.

[27] N. Poor, “Mechanisms of an online public sphere: The website slashdot,” *Journal of Computer-Mediated Communication*, vol. 10, no. 2, pp. 00–00, 2005.

[28] A. Kaltenbrunner, V. Gómez, and V. López, “Description and prediction of slashdot activity,” in *2007 Latin American Web Conference (LA-WEB 2007)*, 2007, pp. 57–66.

[29] C. A. Lampe, E. Johnston, and P. Resnick, “Follow the reader: Filtering comments on slashdot,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ‘07, San Jose, California, USA: ACM, 2007, pp. 1253–1262, ISBN: 978-1-59593-593-9.

[30] A. M. C. Halavais, “The slashdot effect: Analysis of a large-scale public conversation on the world wide web,” PhD thesis, 2001.

[31] J. Grudin, “Why cscw applications fail: Problems in the design and evaluationof organizational interfaces,” in *Proceedings of the 1988 ACM Conference on Computer-supported Cooperative Work*, ser. CSCW ’88, Portland, Oregon, USA: ACM, 1988, pp. 85–93, ISBN: 0-89791-282-9.

[32] C. Lampe, N. Ellison, and C. Steinfield, “A face(book) in the crowd: Social searching vs. social browsing,” in *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, ser. CSCW ‘06, Banff, Alberta, Canada: ACM, 2006, pp. 167–170, ISBN: 1-59593-249-6.

[33] S. Guha and S. B. Wicker, “Do birds of a feather watch each other?: Homophily and social surveillance in location based social networks,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social*

Computing, ser. CSCW ’15, Vancouver, BC, Canada: ACM, 2015, pp. 1010–1020, ISBN: 978-1-4503-2922-4.

[34] C. Lampe, J. Vitak, and N. Ellison, “Users and nonusers: Interactions between levels of adoption and social capital,” in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, ser. CSCW ’13, San Antonio, Texas, USA: ACM, 2013, pp. 809–820, ISBN: 978-1-4503-1331-5.

[35] C. Lampe, N. B. Ellison, and C. Steinfield, “Changes in use and perception of facebook,” in *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, ser. CSCW ’08, San Diego, CA, USA: ACM, 2008, pp. 721–730, ISBN: 978-1-60558-007-4.

[36] E. P. Baumer, X. Xu, C. Chu, S. Guha, and G. K. Gay, “When subjects interpret the data: Social media non-use as a case for adapting the delphi method to cscw,” in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW ’17, Portland, Oregon, USA: ACM, 2017, pp. 1527–1543, ISBN: 978-1-4503-4335-0.

[37] E. P. Baumer, M. G. Ames, J. Burrell, J. R. Brubaker, and P. Dourish, “Why study technology non-use?” *First Monday*, vol. 20, no. 11, 2015.

[38] M. G. Ames, “Managing mobile multitasking: The culture of iphones on stanford campus,” in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, ser. CSCW ’13, San Antonio, Texas, USA: ACM, 2013, pp. 1487–1498, ISBN: 978-1-4503-1331-5.

[39] K. Crawford, “Following you: Disciplines of listening in social media,” *Continuum*, vol. 23, no. 4, pp. 525–535, 2009.

[40] L. Portwood-Stacer, “Media refusal and conspicuous non-consumption: The performative and political dimensions of facebook abstention,” *New Media & Society*, vol. 15, no. 7, pp. 1041–1057, 2013.

[41] E. P. Baumer, J. Burrell, M. G. Ames, J. R. Brubaker, and P. Dourish, “On the importance and implications of studying technology non-use,” *Interactions*, vol. 22, no. 2, pp. 52–56, 2015.

[42] B Mason, “Issues in virtual ethnography in ethnographic studies in real and virtual environments,” *Inhabited Information Spaces and Connected Communities. Edinburgh*, 1999.

[43] J. Preece, B. Nonnecke, and D. Andrews, “The top five reasons for lurking: Improving community experiences for everyone,” *Computers in human behavior*, vol. 20, no. 2, pp. 201–223, 2004.

[44] N. Schultz and B. Beach, “From lurkers to posters,” *Australian National Training Authority*, vol. 14, no. 1, pp. 4–23, 2004.

[45] E. Hargittai, “Whose space? differences among users and non-users of social network sites,” *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 276–297, 2007.

[46] T. Ryan and S. Xenos, “Who uses facebook? an investigation into the relationship between the big five, shyness, narcissism, loneliness, and facebook usage,” *Computers in Human Behavior*, vol. 27, no. 5, pp. 1658–1664, 2011.

[47] M. Warschauer, *Technology and social inclusion: Rethinking the digital divide*. MIT press, 2004.

[48] E. P. Baumer, M. G. Ames, J. R. Brubaker, J. Burrell, and P. Dourish, “Refusing, limiting, departing: Why we should study technology non-use,” in *CHI ’14 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA ’14, Toronto, Ontario, Canada: ACM, 2014, pp. 65–68, ISBN: 978-1-4503-2474-8.

[49] E. P. Baumer, S. Guha, E. Quan, D. Mimno, and G. K. Gay, “Missing photos, suffering withdrawal, or finding freedom? how experiences of social media non-use influence the likelihood of reversion,” *Social Media+ Society*, vol. 1, no. 2, p. 2 056 305 115 614 851, 2015.

[50] K. Hampton, L. Rainie, W. Lu, I. Shin, and K. Purcell, “Social media and the cost of caring,” *Washington, DC: Pew Research Center*, 2015.

[51] R. K. Baker and K. M. White, “In their own words: Why teenagers don’t use social networking sites,” *Cyberpsychology, Behavior, and Social Networking*, vol. 14, no. 6, pp. 395–398, 2011.

[52] Z. Turan, “The reasons for non-use of social networking websites by university students/razones por las que los alumnos universitarios no utilizan las redes sociales,” *Comunicar (English edition)*, vol. 21, no. 41, pp. 137–145, 2013.

[53] N. Brody, “Opting out of social media: Online communication attitudes mediate the relationship between personality factors and facebook non-use,” *Southern Communication Journal*, pp. 1–14, 2018.

[54] E. P. Baumer, D. Mimno, S. Guha, E. Quan, and G. K. Gay, “Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence?” *Journal of the Association for Information Science and Technology*, vol. 68, no. 6, pp. 1397–1410, 2017.

[55] S. Wyche and E. P. Baumer, “Imagined facebook: An exploratory study of non-users’ perceptions of social media in rural zambia,” *New media & society*, vol. 19, no. 7, pp. 1092–1108, 2017.

[56] C. Satchell and P. Dourish, “Beyond the user: Use and non-use in hci,” in *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, ser. OZCHI ’09, Melbourne, Australia: ACM, 2009, pp. 9–16, ISBN: 978-1-60558-854-4.

[57] N. Selwyn, S. Gorard, and J. Furlong, “Whose internet is it anyway? exploring adultsâ(non) use of the internet in everyday life,” *European Journal of Communication*, vol. 20, no. 1, pp. 5–26, 2005.

[58] N. Selwyn, “Apart from technology: Understanding peopleâs non-use of information and communication technologies in everyday life,” *Technology in society*, vol. 25, no. 1, pp. 99–116, 2003.

[59] P. Verdegem and P. Verhoest, “Profiling the non-user: Rethinking policy initiatives stimulating ict acceptance,” *Telecommunications Policy*, vol. 33, no. 10-11, pp. 642–652, 2009.

[60] S. M. Wyatt, N. Oudshoorn, and T. Pinch, “Non-users also matter: The construction of users and non-users of the internet,” *Now Users Matter: The Co-construction of Users and Technology*, pp. 67–79, 2003.

[61] N. E. Oudshoorn and T. Pinch, *How users matter: The co-construction of users and technologies*. MIT press, 2003.

[62] L. Lapointe and S. Rivard, “A multilevel model of resistance to information technology implementation,” *MIS Quarterly*, vol. 29, no. 3, pp. 461–491, 2005.

[63] H.-W. Kim and A. Kankanhalli, “Investigating user resistance to information systems implementation: A status quo bias perspective,” *MIS Quarterly*, vol. 33, no. 3, pp. 567–582, 2009.

[64] G. M. Marakas and S. Hornik, “Passive resistance misuse: Overt support and covert recalcitrance in is implementation,” *European Journal of Information Systems*, vol. 5, no. 3, pp. 208–219, 1996.

[65] J. Carroll, S. Howard, F. Vetere, J. Peck, and J. Murphy, “Just what do the youth of today want? technology appropriation by young people,” in *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, 2002, pp. 1777–1785.

[66] R. L. Nabi, A. Prestin, and J. So, “Facebook friends with (health) benefits? exploring social network site use and perceptions of social support, stress, and well-being,” *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 10, pp. 721–727, 2013, PMID: 23790356. eprint: `https://doi.org/10.1089/cyber.2012.0521`.

[67] M. Burke and R. Kraut, “Using facebook after losing a job: Differential benefits of strong and weak ties,” in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, ser. CSCW ’13, San Antonio, Texas, USA: ACM, 2013, pp. 1419–1430, ISBN: 978-1-4503-1331-5.

[68] Y. Al-Saggaf and P. Simmons, “Social media in saudi arabia: Exploring its use during two natural disasters,” *Technological Forecasting and Social Change*, vol. 95, pp. 3 –15, 2015.

[69] H. Krasnova, H. Wenninger, T. Widjaja, and P. Buxmann, “Envy on facebook: A hidden threat to users’ life satisfaction?” In *11. Internationale Tagung Wirtschaftsinformatik, Leipzig, Germany, February 27 – March 1, 2013*, 2013, p. 92.

[70] H. Appel, A. Gerlach, and J. Crusius, “The interplay between facebook use, social comparison, envy, and depression,” *Current Opinion in Psychology*, vol. 9, pp. 44–49, Jun. 2016.

[71] K. Chaulk and T. Jones, “Online obsessive relational intrusion: Further concerns about facebook,” *Journal of Family Violence*, vol. 26, no. 4, pp. 245–254, 2011.

[72] G. C. E. Kwan and M. M. Skoric, “Facebook bullying: An extension of battles in school,” *Computers in Human Behavior*, vol. 29, no. 1, pp. 16 –25, 2013, Including Special Section Youth, Internet, and Wellbeing.

[73] A. N. Islam and S. Patil, “Engagement and well-being on social network sites,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW ’15, Vancouver, BC, Canada: ACM, 2015, pp. 375–382, ISBN: 978-1-4503-2922-4.

[74] P. Wisniewski, A. N. Islam, B. P. Knijnenburg, and S. Patil, “Give social network users the privacy they want,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW ’15, Vancouver, BC, Canada: ACM, 2015, pp. 1427–1441, ISBN: 978-1-4503-2922-4.

[75] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211–36, 2017.

[76] C. Cadwalladr, “The great british brexit robbery: How our democracy was hijacked,” *The Guardian*, vol. 20, 2017.

[77] I. Lapowsky, *The 21 (and counting) biggest facebook scandals of 2018*, `https://www.wired.com/story/facebook-scandals-2018/`, 2018.

[78] ——, *Mark zuckerberg answers to congress for facebook's troubles*, `https://www.wired.com/story/mark-zuckerberg-congress-facebook-troubles/`, 2018.

[79] T. Hsu, *For many facebook users, a 'last straw' that led them to quit*, `https://www.nytimes.com/2018/03/21/technology/users-abandon-facebook.html`, 2018.

[80] N. Godbole, M. Srinivasaiah, and S. Skiena, "Large-scale sentiment analysis for news and blogs," in *Proceedings of the First International Conference on Weblogs and Social Media, ICWSM 2007, Boulder, Colorado, USA, March 26-28, 2007*, 2007.

[81] E. Kouloumpis, T. Wilson, and J. D. Moore, "Twitter sentiment analysis: The good the bad and the omg!" In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21*, 2011.

[82] S. Mohammad, S. Kiritchenko, and X. Zhu, "Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets," in *Proceedings of the 7th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2013, Atlanta, Georgia, USA, June 14-15*, 2013, pp. 321–327.

[83] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May, Valletta, Malta*, 2010.

[84] H. Cui, V. Mittal, and M. Datar, "Comparative experiments on sentiment classification for online product reviews," in *AAAI*, vol. 6, 2006, pp. 1265–1270.

[85] Y. Dang, Y. Zhang, and H. Chen, "A lexicon-enhanced method for sentiment classification: An experiment on online product reviews," *IEEE Intelligent Systems*, vol. 25, no. 4, pp. 46–53, 2010.

[86] H. Isah, P. Trundle, and D. Neagu, "Social media analysis for product safety using text mining and sentiment analysis," in *Computational Intelligence (UKCI), 2014 14th UK Workshop on*, IEEE, 2014, pp. 1–7.

[87] S. Mukherjee and P. Bhattacharyya, "Feature specific sentiment analysis for product reviews," *Computational Linguistics and Intelligent Text Processing*, pp. 475–487, 2012.

[88] B. Liu, “Sentiment analysis and opinion mining,” *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[89] B. Pang, L. Lee, *et al.*, “Opinion mining and sentiment analysis,” *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[90] T. Mullen and R. Malouf, “A preliminary investigation into sentiment analysis of informal political discourse.,” in *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, 2006, pp. 159–162.

[91] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, “New avenues in opinion mining and sentiment analysis,” *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, 2013.

[92] B. Hu and M. Ester, “Spatial topic modeling in online social media for location recommendation,” in *Proceedings of the 7th ACM conference on Recommender systems*, ACM, 2013, pp. 25–32.

[93] Y. Wang, E. Agichtein, and M. Benzi, “Tm-lda: Efficient online modeling of latent topic transitions in social media,” in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2012, pp. 123–131.

[94] M. Paul and R. Girju, “Cross-cultural analysis of blogs and forums with mixed-collection topic models,” in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, Association for Computational Linguistics, 2009, pp. 1408–1417.

[95] T. Yano, W. W. Cohen, and N. A. Smith, “Predicting response to political blog posts with topic models,” in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, 2009, pp. 477–485.

[96] C. Van Hee, G. Jacobs, C. Emmery, B. Desmet, E. Lefever, B. Verhoeven, G. De Pauw, W. Daelemans, and V. Hoste, “Automatic detection of cyberbullying in social media text,” *PLOS ONE*, vol. 13, no. 10, H. Suleman, Ed., e0203794, 2018.

[97] V. Nahar, S. AL Maskari, X. Li, and C. Pang, “Semi-supervised learning for cyberbullying detection in social networks,” Jul. 2014, pp. 160–171, ISBN: 978-3-319-08607-1.

[98] A. Squicciarini, S. Rajtmajer, Y. Liu, and C. Griffin, “Identification and characterization of cyberbullying dynamics in an online social network,” in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015*, J. Pei, J. Tang, and F. Silvestri, Eds., ser.

Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015, Association for Computing Machinery, Inc, Aug. 2015, pp. 280–285.

[99] V. Narayanan, I. Arora, and A. Bhatia, “Fast and accurate sentiment classification using an enhanced naive bayes model,” in *International Conference on Intelligent Data Engineering and Automated Learning*, Springer, 2013, pp. 194–201.

[100] S. M. Kim and R. A. Calvo, “Sentiment analysis in student experiences of learning.,” in *Educational Data Mining 2010, The 3rd International Conference on Educational Data Mining, Pittsburgh, PA, USA, June 11-13. Proceedings*, 2010, pp. 111–120.

[101] L. Barbosa and J. Feng, “Robust sentiment detection on twitter from biased and noisy data,” in *Proceedings of the 23rd international conference on computational linguistics: Posters*, Association for Computational Linguistics, 2010, pp. 36–44.

[102] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, “Sentiment analysis of twitter data,” in *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, 2011, pp. 30–38.

[103] J. D. Prusa, T. M. Khoshgoftaar, and D. J. Dittman, “Impact of feature selection techniques for tweet sentiment classification,” in *The Twenty-Eighth International Flairs Conference*, 2015.

[104] R. Kavuluru, M. Ramos-Morales, T. Holaday, A. G. Williams, L. Haye, and J. Cerel, “Classification of helpful comments on online suicide watch forums,” in *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, ACM, 2016, pp. 32–40.

[105] E. Poché, N. Jha, G. Williams, J. Staten, M. Vesper, and A. Mahmoud, “Analyzing user comments on youtube coding tutorial videos,” in *Proceedings of the 25th International Conference on Program Comprehension*, IEEE Press, 2017, pp. 196–206.

[106] N. Wanas, M. El-Saban, H. Ashour, and W. Ammar, “Automatic scoring of online discussion posts,” in *Proceedings of the 2nd ACM workshop on Information credibility on the web*, ACM, 2008, pp. 19–26.

[107] M. R. Brennan, S. Wrazien, and R. Greenstadt, “Learning to extract quality discourse in online communities,” in *Workshops at the Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[108] N. FitzGerald, G. Carenini, G. Murray, and S. Joty, "Exploiting conversational features to detect high-quality blog comments," in *Canadian Conference on Artificial Intelligence*, Springer, 2011, pp. 122–127.

[109] A. J. Chan, "Collaborative news networks: Distributed editing, collective action, and the construction of online news on slashdot. org," PhD thesis, Massachusetts Institute of Technology, 2002.

[110] C. Lampe and P. Resnick, "Slash(dot) and burn: Distributed moderation in a large online conversation space," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '04, Vienna, Austria: ACM, 2004, pp. 543–550, ISBN: 1-58113-702-8.

[111] C. B. Draucker, D. S. Martsolf, R. Ross, and T. B. Rusk, "Theoretical sampling and category development in grounded theory," *Qualitative health research*, vol. 17, no. 8, pp. 1137–1148, 2007.

[112] H. Joffe and L. Yardley, "Content and thematic analysis," *Research methods for clinical and health psychology*, vol. 56, p. 68, 2004.

[113] E. M. Younis, "Sentiment analysis and text mining for social media microblogs using open source tools: An empirical study," *International Journal of Computer Applications*, vol. 112, no. 5, 2015.

[114] Y.-C. Wang, M. Burke, and R. E. Kraut, "Gender, topic, and audience response: An analysis of user-generated content on facebook," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2013, pp. 31–34.

[115] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[116] H. Suresh and N. Locascio, "Autodetection and classification of hidden cultural city districts from yelp reviews," *ArXiv preprint arXiv:1501.02527*, 2015.

[117] J. Choo, C. Lee, C. K. Reddy, and H. Park, "Utopian: User-driven topic modeling based on interactive nonnegative matrix factorization," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 12, pp. 1992–2001, 2013.

[118] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Eighth international AAAI conference on weblogs and social media*, 2014.

[119] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd annual*

meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2004, p. 271.

[120] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2004, pp. 168–177.

[121] R. Agrawal, *Analyzing text classification techniques on youtube data*, https://towardsdatascience.com/analyzing-text-classification-techniques-on-youtube-data-7af578449f58, 2019.

[122] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, D. Brown, L. Id, and Barnes, "Text classification algorithms: A survey," *Information (Switzerland)*, vol. 10, Apr. 2019.

[123] S. Raschka, *Naive bayes and text classification*, https://sebastianraschka.com/Articles/2014_naive_bayes_1.html, 2014.

[124] M. Pellarolo, *Logistic regression from scratch in python*, https://medium.com/@martinpella/logistic-regression-from-scratch-in-python-124c5636b8ac, 2018.

[125] A. Staveley, *Adamstaveley/machine-learning-notes*, https://github.com/adamstaveley/machine-learning-notes/blob/master/notes.tex.

[126] D. J. Brand, "Automatic prediction of comment quality," PhD thesis, Stellenbosch: Stellenbosch University, 2016.

[127] J. Vitak, S. Blasiola, S. Patil, and E. Litt, "Balancing audience and privacy tensions on social network sites: Strategies of highly engaged users," *International Journal of Communication*, vol. 9, p. 20, 2015.

[128] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Commun. ACM*, vol. 59, no. 7, pp. 96–104, Jun. 2016.

[129] S. Y. Schoenebeck, "Giving up twitter for lent: How and why we take breaks from social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14, Toronto, Ontario, Canada: ACM, 2014, pp. 773–782, ISBN: 978-1-4503-2473-1.

[130] R. Joshi, *Accuracy, precision, recall f1 score: Interpretation of performance measures*, https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/, 2016.

[131] A. Lenhart, K. Purcell, A. Smith, and K. Zickuhr, "Social media & mobile internet use among teens and young adults. millennials.," *Pew internet & American life project*, 2010.

[132] R. Kang, L. Dabbish, N. Fruchter, and S. Kiesler, ""my data just goes everywhere:" user mental models of the internet and implications for privacy and security," in *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, Ottawa: USENIX Association, 2015, pp. 39–52, ISBN: 978-1-931971-249.

[133] M. A. Moreno, L. A. Jelenchick, K. G. Egan, E. Cox, H. Young, K. E. Gannon, and T. Becker, "Feeling bad on facebook: Depression disclosures by college students on a social networking site," *Depression and anxiety*, vol. 28, no. 6, pp. 447–455, 2011.

[134] H. B. Shakya and N. A. Christakis, "A new, more rigorous study confirms: The more you use facebook, the worse you feel," *Harvard Business Review*, vol. 10, 2017.

[135] J. Rosenberg and N. Egbert, "Online impression management: Personality traits and concerns for secondary goals as predictors of self-presentation tactics on facebook," *Journal of Computer-Mediated Communication*, vol. 17, no. 1, pp. 1–18, 2011.

[136] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS quarterly*, pp. 425–478, 2003.

[137] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989.

[138] S. Stieger, C. Burger, M. Bohn, and M. Voracek, "Who commits virtual identity suicide? differences in privacy concerns, internet addiction, and personality between facebook users and quitters," *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 9, pp. 629–634, 2013.

[139] I. Ajzen, "The theory of planned behavior," *Organizational behavior and human decision processes*, vol. 50, no. 2, pp. 179–211, 1991.

[140] J. Pierce, "Undesigning technology: Considering the negation of design by design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '12, Austin, Texas, USA: ACM, 2012, pp. 957–966, ISBN: 978-1-4503-1015-4.

[141] E. P. Baumer and M. S. Silberman, "When the implication is not to design (technology)," in *Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems, ser. CHI '11, Vancouver, BC, Canada: ACM, 2011, pp. 2271–2274, ISBN: 978-1-4503-0228-9.

[142] P. Dourish, "Implications for design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '06, Montréal, Québec, Canada: ACM, 2006, pp. 541–550, ISBN: 1-59593-372-7.