

MODELING REAL AND FAKE NEWS SHARING IN SOCIAL
NETWORKS

by
Abishai Joy



A thesis
submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Computer Science
Boise State University

August 2021

© 2021

Abishai Joy

ALL RIGHTS RESERVED

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Abishai Joy

Thesis Title: MODELING REAL AND FAKE NEWS SHARING IN SOCIAL NETWORKS

Date of Final Oral Examination: 09 June 2021

The following individuals read and discussed the thesis submitted by student Abishai Joy, and they evaluated the student's presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Francesca Spezzano Ph.D.	Chair, Supervisory Committee
Jerry Alan Fails Ph.D.	Member, Supervisory Committee
Edoardo Serra Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Francesca Spezzano Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.

ACKNOWLEDGMENT

I am extremely grateful to my supervisor, Dr. Francesca Spezzano, this research work would not have been possible without her invaluable guidance and advice. My sincere thanks to Dr. Edoardo Serra and Dr. Jerry Alan Fails for their technical support and feedback on my study. I would also like to thank my teammates (Anu, Nikesh, and Steven) for their kind help. Finally, I would like to express my gratitude to my husband, my parents, my sister and my in-laws. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study. Thanks to the Almighty, who gave me good health and mind throughout this venture.

ABSTRACT

Online media is changing the traditional news industry and diminishing the role of journalists, newspapers, and even news channels. This in turn is enhancing the ability of fake news to influence public opinion on important topics. The threat of fake news is quite imminent, as it allows malicious users to share their agenda with a larger audience. Major social media platforms like Twitter, Facebook, etc., are making it easy to spread fake news due to the minimal moderation/ fact-checking on these platforms.

This work aims at predicting fake and real news sharing in social media. Specifically, we employ a multi-level influence, drawn from the Diffusion of Innovation (DOI) theory on a real-world dataset and predict whether and when a given user will share information in social media. We hypothesize that fake and real news sharing is better predicted by considering user, news, and network-level feature attributes together.

We are also predicting the time elapsed between the influencer and follower shares via survival analysis. Binary classifiers such as Support Vector Machine (SVM), Random Forest, etc. are used for the prediction of fake and real news sharing. This approach is demonstrated using a dataset comprising 1,572 users that are sampled from the FakeNewsNet repository. Our results show a 30% increase in the Area Under Receiver Operation Characteristics (AUROC) in comparison to the best baseline. Real and fake news sharing shows high dependency on user similarity, tie strength,

and explicit features.

Furthermore, the analysis shows that users with characteristic features like love, self-transcendence, ideals, conservation, and openness to change tend to share real news, whereas users with dominant features like self-enhancement, curiosity, closeness, structure, and harmony are more likely to share fake news.

Finally, survival analysis is employed to predict the time elapsed between influencer and follower shares. The Concordance Index (C-Index) for real news sharing is slightly lower compared to the baseline, and the C-Index of Random Survival Forest (RSF) is comparable to the baseline for fake news sharing. Furthermore, in comparison to the regression baseline models, the Mean Absolute Error (MAE) is significantly less in RSF for both real and fake news sharing.

TABLE OF CONTENTS

ACKNOWLEDGMENT	iv
ABSTRACT	v
LIST OF FIGURES	ix
LIST OF TABLES	xi
LIST OF ABBREVIATIONS	xii
1 INTRODUCTION	1
2 RELATED WORK	6
3 DATASET	12
3.0.1 Dataset Description	12
3.0.2 Dataset Creation	13
4 METHODS AND FEATURES	17
4.1 Methods	31
4.1.1 Survival Analysis	31
4.1.2 Censoring	32

5	EXPERIMENTS AND RESULTS	36
5.1	Experimental Setting	36
5.1.1	Comparison with Baselines	37
5.1.2	Analysis of Real and Fake News Sharing	38
5.1.3	Experimental Setting for Survival Analysis	43
6	CONCLUSIONS AND FUTURE WORK	51
6.0.1	Future Work	52
6.0.2	Limitations	53

LIST OF FIGURES

1.1	Schema for the Proposed Work	4
3.1	Glimpse of Fake (red) and Real News (blue) in Dataset	14
3.2	Distribution of Real and Fake news shares	15
4.1	Distribution of real and fake news sharing by the political alignment of users.	20
4.2	Distribution of fake and real news sharing by gender and age of the users.	21
4.3	Words and Phrases associated with low-stress (blue) and high-stress (red) users	23
4.4	Right Censoring	33
4.5	Random Survival Tree	34
5.1	Contribution from each Feature Groups	40
5.2	Feature Ablation Results (News)	41
5.3	Feature Ablation Results (Network)	42
5.4	Feature Ablation Results (User)	46
5.5	Feature Ablation Results (User)	47
5.6	Feature importance of personality features for fake (red) and real news sharing (blue).	48

5.7	Feature importance of key features for fake (red) and real news sharing (blue).	49
5.8	Simplified Key Feature Ablation Results	50

LIST OF TABLES

3.1	FakeNewsNet : Details of the dataset from Politifact media	13
3.2	Computed Dataset - Fake News Sharing	16
3.3	Computed Dataset - Real News Sharing	16
5.1	Classifier results: Results of classification for Fake News Sharing . . .	39
5.2	Classifier results: Results of classification for Real News Sharing . . .	39
5.3	Key Features: Results of classification for Fake News Sharing	45
5.4	Key Features: Results of classification for Real News Sharing	45
5.5	Dataset - Fake News Sharing	45
5.6	Dataset - Real News Sharing	45
5.7	Regression results: Results of Regression for Real News Sharing . . .	45
5.8	Regression results: Results of Regression for Fake News Sharing . . .	45

LIST OF ABBREVIATIONS

AUROC Area Under the Receiver Operating Characteristics

C-Index Concordance Index

DOI Diffusion of Innovation theory

ICM Independent Cascade Model

LIWC Linguistic Inquiry and Word Count

LTM Linear Threshold Model

MAE Mean Absolute Error

RSF Random Survival Forest

SMOTE Synthetic Minority Oversampling Technique

SVM Support Vector Machine

TFF Twitter Follower-Following ratio

CHAPTER 1:

INTRODUCTION

Middle schoolers in Philadelphia believed that the earth is flat because of the idea they picked up from basketball star Kyrie Irving, who said that on a podcast [1]. This seemingly funny anecdote highlights the alarming impact of misinformation in today's digital age. The Cambridge dictionary defines fake news as false stories that appear to be news, spread on the internet or using other media, usually created to influence political views or as a joke [2]. Fake news is typically characterized by its use of unverifiable content, source, or origin, often designed to appeal to emotions. Web 2.0 technology has further accelerated the consumption of fake news and its impact on society, leading to an urgent need for understanding the spread of misinformation in social media. One of the most prevalent social media platforms today is Twitter. As per Twitter statistics, around 145 million daily active users (out of the 330 million) spend an average of 3.39 minutes per session on online social media networks and are estimated to share 500 million tweets each day [3]. While a small fraction of the users actively propagate fake information, large platforms like Twitter amplify the impact based on the dynamics of sharing. Understanding the dynamics of news sharing (particularly fake news) will provide a means to prevent the dissemination of fake news preemptively or at least rapidly upon detection.

Traditional print media involves rigorous research, investigation, peer review and

cannot be edited easily after publication. However, news on social media is often unmoderated, can be altered by anyone, and the rate of dissemination can change dynamically at any time. Understanding the correlation between the news content, the user's behavior (or traits), and the rate of information spreading will help prevent spreading fake news. News features such as style, complexity, and psychology; user features like individual's personality, political alignment, demographics, and time between news sharing may provide useful insights to model online news (and fake news) propagation. The time of a tweet can be particularly useful in the study of fake news sharing and detecting of fake news. Like we pointed out earlier, fake news attempts to appeal to emotion, and this often leads to faster news dissemination. This work explores these correlations to model real and fake news sharing.

Classical models for information diffusion such as Independent Cascade and Linear Threshold models assume that a user will share the news with some probability only according to the fact that some of their friends have previously shared the same news [4]. However, recent works on fake news sharing in the social science domain have shown that a user's decision of sharing or not sharing a piece of given news does not only depend on the influence of their friends but also specific characteristics of the users (e.g., demographics, Twitter profile properties, Twitter behavior, and activity, etc.), the news received (e.g., title and content, etc.), and the social context (e.g., number of followers and following, tie strength, etc.) [5]. All these aspects align with what is theorized by the diffusion of innovation theory to explain how an innovation (which in our case is news) diffuses in a social network [6]. Moreover, in real life, users do not share news as soon as it is received, but there are some users who are exposed to the news for a while before sharing, while others become skeptical and do

not share at all.

Thus, our main objective through this thesis is to predict whether and when the user will share a piece of news from his influencer. For this, we are applying the diffusion of innovation theory to model how real and fake news is shared by Twitter users. Given evidence by social science studies, we hypothesize that real and fake news sharing is better predicted when user, news, and social network characteristics are all taken into account [5]. All these factors have never been combined into a unique predictive model or tested on a large scale before.

Specifically, we will address the following problems:

- given that a user ‘u1’ is influenced on some given news ‘n’ by at least one of their influencers ‘i1’ (i.e., ‘u1’ is following ‘i1’ and ‘i1’ has shared some news ‘n’ among their followers), predict whether the user ‘u1’ will also share news ‘n’ among their followers (u11 & u12); and
- given that a user ‘u1’ is influenced on some given news ‘n’ by at least one of their influencers ‘i1’, predict the time elapsed between when ‘i1’ shared news ‘n’ and the user ‘u1’ will also share ‘n’ among their followers (u11 & u12).

Like we mentioned earlier, Twitter is one of the most prevalent social media platforms. Hence we use a Twitter data set from FakeNewsNet¹ for our analysis. The data set includes news content, social context, and spatial-temporal information. It contains data like Twitter profile data, timeline tweets with content, news titles as well as content information, and follower relationship.

The overall goal is to predict whether and when the user will share the content in his/her social circle. Figure 1.1 gives an outline for the methodology we implement

¹<https://github.com/KaiDMML/FakeNewsNet/tree/master/code>

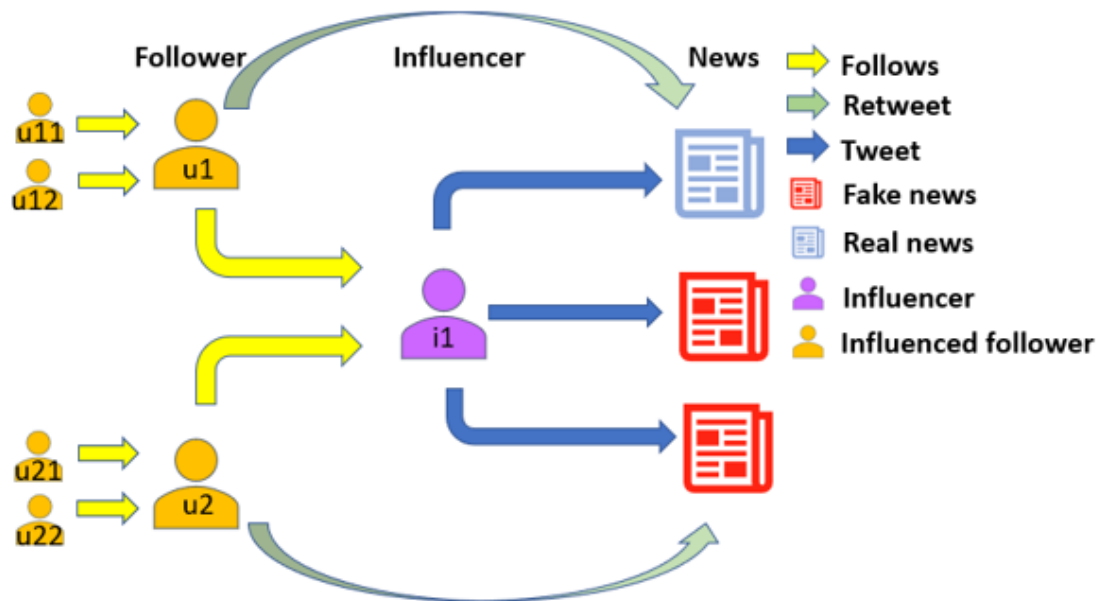


Figure 1.1: Schema for the Proposed Work

to predict user's news sharing behavior. The raw data includes information from the Twitter profile, tweets, news content, and follower details. A multi-level influence approach, drawn from the Diffusion of Innovation (DOI) theory, is used to explain how, why, and at what rate new ideas and technology spread through a specific population or social system [6]. The data is split into three categories (user, news, and network) to implement the multi-level influence approach. The data is further transformed and used to compute features like user behavioral, linguistic, and tie strength. These features are further fed to the classifiers and survival model. For the first part of our research, the classifiers such as logistic regression, extra trees, random forest, and support vector machines are used and evaluated using Area Under the Receiver Operating Characteristics (AUROC) and Average Precision. The second part focuses on predicting how long the user takes to share the news (time-to-share).

As per Vosoughi et al., fake news travels faster, farther, and more broadly than the real news [7]. Thus time to share is an important characteristic of fake news and knowing speed of propagation is important to understand the response time needed to prevent fake news diffusion. Therefore, time is crucial information that can help with both detection and prevention of fake news sharing. This analysis requires the computation of time elapsed between the tweets and retweets. Further, the random survival forest is utilized to predict the time-to-event.

Our results show that news sharing is better predicted when multiple features are considered. Among the features, news-based features outperformed, followed by user and network attributes. The approach of combining the news, network, and user features has boosted the overall AUROC by 30% in comparison to the best baseline. We used Linear Threshold and Independent Cascade diffusion models as a baseline. These models use propagation probabilities to infect an inactive user or follower and does not depend on news or user features. The survival model is evaluated using the concordance index and mean absolute error. The findings show that the mean absolute error obtained by the survival model is significantly less and more reliable for fake news sharing.

The thesis is organized as follows, Chapter 2 reports related work. Chapter 3 describes the dataset used for the experiments and analysis. Chapter 4 explains the features computed and the methods. The results of our analysis are tabulated in Chapter 5, Experiments, and Results. Finally, Chapter 6 gives the conclusion along with future work.

CHAPTER 2: RELATED WORK

Previous studies show that the typical user who shares news can be considered an opinion leader or an influencer. An opinion leader adopts new ideas and is often approached by their network for advice or information on the content shared by an influencer. While there is no global theory for studying news sharing in social networks, studies show that many previous works draw inferences from the DOI theory [5]. Aside from DOI, theories of social influence and the concepts of interactivity, political participation, and the uses and gratifications approach were somewhat relevant [5]. In all of these theories, seeking status, gaining reputation, and drawing people's attention to one's own views and ideas are the main motivations of news sharing [8]. Shu et al. [9, 10] focus on considering a tri-relationship among user, publisher, and news content. User engagements represent the news proliferation process over time, which provides useful auxiliary information to infer the veracity of news articles. Shu et al. [11] have proposed a detection method using explicit and implicit features from data, which has the potential to differentiate fake news. Some of the examples include register time and demographics details for explicit and implicit features, respectively. The register is the time of registration at the social media platform by the user, this information is included in the user profile data from Twitter. The demographics were computed using the id, name, screen name, description, language, image path, and

resized profile images of the users.

Recent studies also infer that people tend to share misinformation that is formerly tagged as inaccurate. The survey conducted by Pennycook et al. [12] shows that true headlines receive higher accuracy ratings by humans than false headlines. However, users in that research seemed to barely take this information into account when considering what to share on social media. The main reasons behind this behavior was to (1) attract more followers, (2) signal one's group membership, and (3) engage with emotionally evocative content that distracts the audience from the veracity of the news.

As mentioned in Chapter 1, considering multiple levels of influence is an effective way of studying news sharing behavior in social media [13]. The multiple levels of influence include diffusion networks, individual influence, and innovation attributes (news attributes). Ma et al. [13] also explain the importance of analyzing strong and weak ties along with the number of followers in a social circle. Not all demographic attributes are predictors for information sharing in social media. It has been proven that a lack of digital media literacy is one of the reasons for increased fake news sharing [14]. This suggests that the user's age is a good predictor for understanding fake news sharing in social media. Further, none of the other demographics variables - sex, race, education, and income - have a strong predictive effect on sharing fake news [14]. The study by Yaqub et al. shows that an addition of credibility indicator to the social media content can potentially increase the information literacy [15]. However, the author's findings show that their effectiveness varies based on the type of indicator and personal characteristics of the user [15]. The paper by Yaqub et al. also discusses the various rationales behind sharing intent [15]. Under control

conditions (where the user is not given a credibility indicator), the main motivations for a user to share or not share true or false news are: (1) user wanted / not wanted to include the interesting/uninteresting news on their social media page, (2) news will / will not trigger a discussion among friends, (3) user wanted/did not want to share a true/false news, and (4) user wanted/did not want to share this news because it is/is not relevant to his or her life [15]. Where the user was given a credibility indicator, participants in the survey chose not to share the fake news 30% of the time. These findings are biased because the users know the credibility of news before the survey. Moreover, the statistics were collected manually, and no scalable, automated analysis was developed [15].

Vosoughi et al. [7] explored the diffusion dynamics of true and false news. As per this study, fake news diffused significantly faster, farther, deeper, and more broadly than true news. The analysis showed that the news about politics and urban legends was the most viral. This article also emphasized that the removal of bots using a bot detection algorithm never changed their results on diffusion dynamics.

Existing research efforts exploit various features of the data, including network features. The paper by Shu et al. [16] also explains that the provenance of fake news indicates the originators. Provenance can help answer questions such as whether the piece of news has been modified during its propagation and how the creator of the piece of information is connected to the transmission of the statement. However, it does not help predict if user will share or speed of sharing of fake news.

Social network analysis is increasing its popularity, and one of the important research areas within this field is Information Diffusion. There are two widely used information diffusion models, namely (i) Threshold Model of Diffusion and (ii) Cas-

cade Model of Diffusion.

Independent Cascade Model (ICM) is a stochastic information diffusion model where the information flows over the network through a cascade. Nodes can have two states, (i) Active: It means the node is already influenced by the information in diffusion. (ii) Inactive: The node is unaware of the information or not influenced by the information in diffusion.

The process runs in discrete steps. At the beginning of the ICM process, a few nodes, known as seed nodes, have already shared the piece of news. Upon receiving the information, these nodes become active. In each discrete step, an active node tries to influence one of its inactive neighbors. The same node will never get another chance to activate the same inactive neighbor. The success depends on the propagation probability of their tie. The propagation probability of a tie is the probability by which one can influence the other node. In reality, propagation probability is relation-dependent, i.e., each edge will have a different value. The process terminates when no further nodes became activated from the inactive state [4].

In **Linear Threshold Model (LTM)**, a node v is influenced by each neighbor w according to a weight $b_{v,w}$. The dynamics of the process then proceeds as follows. Each node v chooses a threshold θ_v uniformly at random from the interval $[0, 1]$; this represents the weighted fraction of v 's neighbors that must become active in order for v to become active. Given a random choice of thresholds and an initial set of active nodes A (with all other nodes inactive), the diffusion process unfolds deterministically in discrete steps: in step t , all nodes that were active in step $t - 1$ remain active, and we activate any node v for which the total weight of its active neighbors is at least θ_v :

$$\sum_{\text{sum of } w' \text{ neighbor of } v} b_{v,w} \leq 1$$

Thus, the thresholds θ_v intuitively represent the different latent tendencies of nodes to adopt the innovation when their neighbors do [4].

In an epidemiological model such as SEIR [17] (Susceptible-Exposed-Infected-Recovered), a latent period ϵ is introduced to allow a random waiting time before the beginning of infectiousness. Biologically, this corresponds to a period during which the infection is establishing itself in its host but unable to jump to another host. The contact interval is assumed to be identically distributed for all pairs of ij where i infects j , which is unrealistic and the contact interval can vary for different users [17]. The effects of covariates (features) on the transmission of infection is a central concern that most of the previous studies have not addressed. Other models like the independent cascade model and the linear threshold model were proposed to understand the direction of influence [18]. However, the details of the user or news content have not yet been analyzed using these models. We have used Independent Cascade and Linear Threshold models as the baseline for this research and compared the performance with our proposed method.

To overcome the above-mentioned limitations, our work investigates the diffusion of innovation theory to explore individual-level, network-level, and news-attribute-level impact on the user’s decision to share. Furthermore, we conducted a survival analysis using random survival forest to predict the time to share for a given influenced user. A random survival forest is a meta estimator that fits a number of survival trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. This method is used for right-censored survival

data as described in Section 4.1.1.

CHAPTER 3:

DATASET

3.0.1 Dataset Description

FakeNewsNet is a multi-dimensional data repository that currently contains two datasets with news content, social context, and spatiotemporal information [19]. The dataset is constructed using an end-to-end system, FakeNewsTracker [19]. The constructed FakeNewsNet repository has the potential to boost the study of various open research problems related to fake news study. Because of the Twitter data sharing policy, it only shares the news articles and tweet ids as part of this dataset and provides codes to download complete tweet details, social engagements, and social networks.

The code repository can be used to download news articles from published websites and relevant social media data from Twitter. The scripts make use of keys from the tweets keys file, which are activated from a Twitter developer account. A summary of the dataset can be seen in Table 3.1.

The downloaded dataset from FakeNewsNet contains news ground truth gathered from Politifact. PolitiFact is a website operated by the *Tampa Bay Times*, where reporters and editors from the media fact-check the news articles. The website publishes the original statement of news articles and their fact-check results.

Table 3.1: FakeNewsNet : Details of the dataset from Politifact media

	Politifact
Count of Users	281,596
Real News + Fake News	560 + 432
Count of News	992
Count of Tweets	438,504
Count of Retweets	619,239

There are 560 real news and 432 fake news in the FakeNewsNet. The word clouds for the news content compositions of the datasets are presented in Figure 3.1. The word clouds represent the frequent words in a text, where the size of the word is proportional to the number of times words were used. We observe that fake news from the Politifact dataset has more political content compared to the real news. The social context information of 281K users consists of their posts, user behavior such as replies, re-posts, and likes as well as the metadata information for user profiles, user posts, and social network information. The dynamic context includes information such as timestamps of user engagements.

Tweets and retweets are the popular means of content sharing on Twitter. From FakeNewsNet, we had 438K tweets and 619K retweets with details of the user and the time of share. Figure 3.2 shows the distribution of shares for the real and fake news in the dataset. It is evident from the figure that the real news is shared more than and retweeted more frequently than fake news. This is not surprising as the study by Guess et al. shows that the sharing of articles from fake news domains is a rare phenomenon [14], but it can have a big impact.

3.0.2 Dataset Creation

After the exploratory analysis, we merged the data sets from Politifact media to identify the influencer and influenced user pairs. The user pair is a pair of users who



Figure 3.1: Glimpse of Fake (red) and Real News (blue) in Dataset

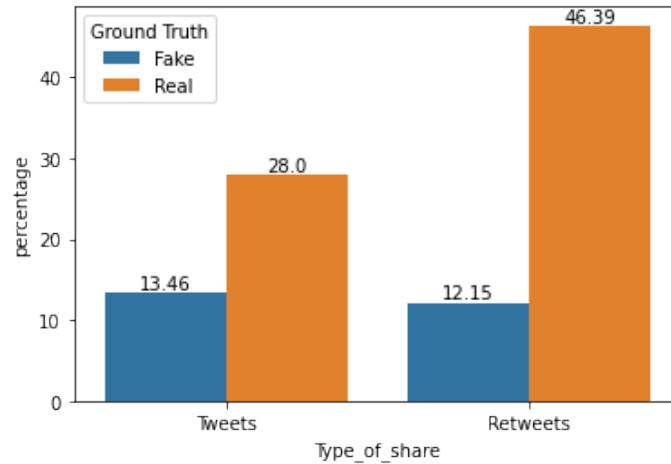


Figure 3.2: Distribution of Real and Fake news shares

tweeted and retweeted a given piece of news. The influenced user is a follower of the influencer user. For each influenced user, at least 5 instances of news sharing were considered in chronological order. The two most recently shared news were used in the creation of the dataset and the remaining shares were concatenated to compute features for the classification task, which are detailed in Chapter 4. In the final dataset, there were 2,403 influenced users, and each row defines a user (influencer) exposed to a piece of news and having at least one follower. If the follower (influenced user) shares the content, we labeled the instance as 1, 0 otherwise. The calculation of user-based features utilized timeline tweets of the influenced-user, which were separately crawled from Twitter. For this, we filtered timeline tweets of the influenced-user, posted within the time interval of the publish date of the news utilized in the main dataset per influenced-user. We only had timeline data on 1,572 influenced users which reduced the overall count of users. Our dataset was imbalanced with 3,144 and 14,936 instances labeled as 1 (influenced user) and labeled as 0 (not influenced user). The details of the computed dataset are given in Table 3.2 and Table

Table 3.2: Computed Dataset - Fake News Sharing

	Politifact
Count of Users	1,557
Count of Fake News	169
Influenced Users	527
Not Influenced Users	7,134

Table 3.3: Computed Dataset - Real News Sharing

	Politifact
Count of Users	1,572
Count of Real News	127
Influenced Users	2,617
Not Influenced Users	7,802

3.3 for fake and real news sharing, respectively. Finally, the time elapsed between the tweets and retweets of the same news for different pairs of users were computed to provide the ground truth for the problem of predicting when the user will share information as discussed in Section 4.1.1.

CHAPTER 4:

METHODS AND FEATURES

As discussed in Chapter 1, we have employed multi-level influence drawn from the DOI theory. This research focuses on three categories of features, namely user-based, news-based and social network-based features. While the explicit and implicit attributes became a feature for each user in the dataset, stylistic and complexity characteristics constituted the news-based features. This chapter describes the set of features we used in the research to analyze real and fake news sharing.

User Related features:

- **Personality Features**

The IBM Watson Personality Insights service uses linguistic analytics to infer individuals' intrinsic personality characteristics, including Big Five personality traits, Needs, and Values, from digital communications such as social media posts. In this research, all the timeline tweets were concatenated for a given influenced user to compute their personality characteristics. The features computed by this service are detailed in the following:

Big Five

The Big Five personality traits, also known as the five-factor model (FFM) and the OCEAN model is a widely used taxonomy to describe people's personality

traits [20]. The five basic personality dimensions described by this taxonomy are openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. For each personality dimension, IBM Watson Personality Insights also provides a set of additional six facet features. For instance, agreeableness's facets include altruism, cooperation, modesty, morality, sympathy, and trust.

Needs

These features describe the needs of a user as inferred by the text they wrote and include excitement, harmony, curiosity, ideal, closeness, self-expression, liberty, love, practicality, stability, challenge, and structure.

Values

These features describe the motivating factors that influence a person's decision-making. They include self-transcendence, conservation, hedonism, self-enhancement, and openness to change.

- **Explicit Features**

Protected, Verified and Register time - Protected, when true, indicates that this user has chosen to protect their tweets. Verified indicates whether it is a verified user. Register time is the number of days passed since the creation date. These are explicit features explained by Shu et al. that can better depict the user characteristics [11].

Status count and Favor count - Status count indicates the number of tweets (including retweets) issued by the user and favor count indicates the number of tweets this user has liked in the account's lifetime [11]. These features represent how active the user is in the social network.

- **Political Ideology**

This indicates the political alignment of the given user and involves computation of polar scores from the hashtags of the users using feature selection algorithms [21]. For the computation of polar scores, we considered the political dataset from the paper by Chamberlain et al., since its time range aligns well with our main dataset [22]. The dataset has tweets and the related hashtags of different politicians. As part of the process, TfidfVectorizer was used to create the feature vector, which was the hashtags. The chi-square algorithm was used to compute importance scores for each feature or hashtag. Following this, all the scores corresponding to the hashtags used by an influenced user were summed up. If the sum is negative, we label the user as left-leaning, otherwise right-leaning. Guess et al. also infers political affiliation as a statistically significant factor in the spread of fake news [14]. Out of the 1572 influenced users, we had political affiliation details on 712 users. Figure 4.1 shows the distribution of real and fake news sharing among different users by their political alignment.

- **Age and Gender**

As per Guess et al., one of the predictors of whether someone will share fake news can be their age [14]. Moreover, Shu et al. has proved that female users are more likely to spread fake news than male users [11]. We used the PyTorch implementation of the M3 (Multimodal, Multilingual, and Multi-attribute) system to determine the age and gender of the influenced user [23]. M3 is a deep learning system for the demographic inference that was trained on a massive Twitter dataset [23]. It features three major attributes: (1) Multimodal - the input can be an image and text, (1) Multilingual - it operates in 32 different

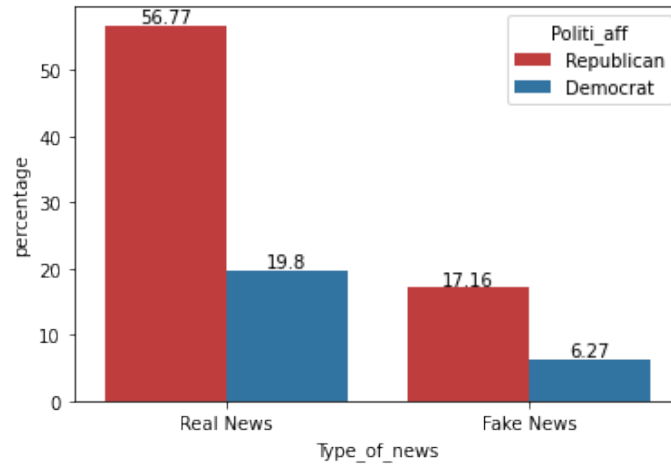


Figure 4.1: Distribution of real and fake news sharing by the political alignment of users.

languages, and (3) Multi-attribute - it can predict three demographic attributes (gender, age, and human-vs-organization status) [23]. For our analysis, we fed the id, name, screen name, description, language, image path, and resized profile images as input to m3inference, and the resulting user’s age and gender distributions are shown in Figure 4.2. Our output consisted of user demographics details on age and gender. It is distinct that older public and females are more vulnerable to misinformation.

- **Stress Analysis**

A series of literary studies have demonstrated that user’s mental health conditions can be predicted from their social media messages [24]. We prepared the input file comprising of concatenated and cleaned timeline tweets of each user from the dataset and used Pennebaker’s Linguistic Inquiry and Word Count (LIWC) approach for stress analysis. As part of the cleaning process, we replaced all emojis and emoticons in the stop words and punctuations - free texts.

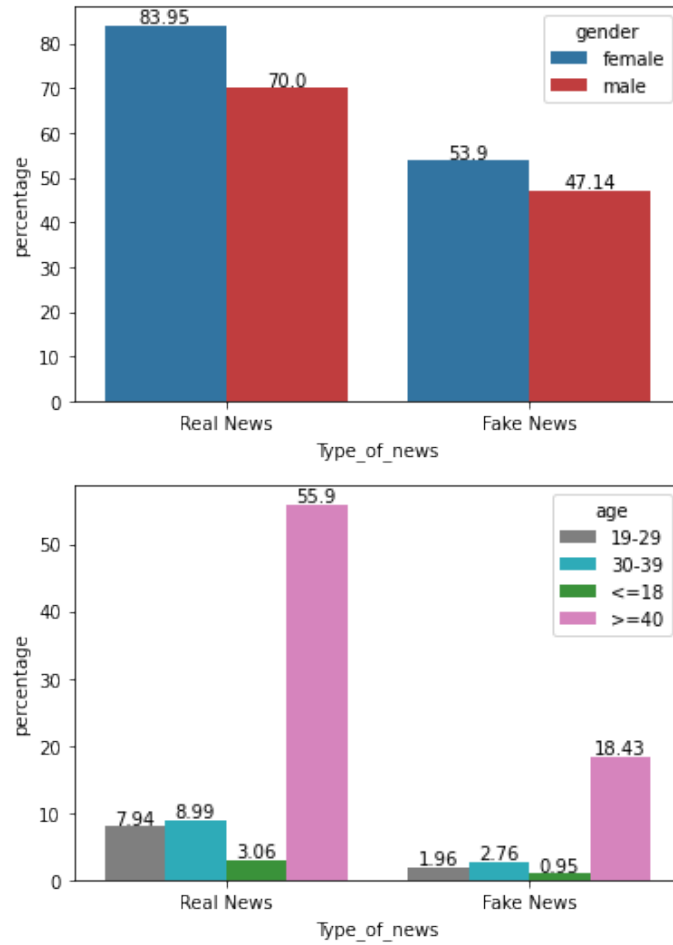


Figure 4.2: Distribution of fake and real news sharing by gender and age of the users.

The patterns, for example, HTTPS, RT, #via, were also removed.

LIWC is a transparent text analysis program that counts words in psychologically meaningful categories [25]. The LIWC program has two features—the processing component and the dictionaries. The processing feature is the program, which opens a series of text files—which can be essays, articles, blogs, and so on—and then each word in a given text file is compared with the dictionary file [26]. Empirical results using LIWC demonstrate its ability to detect meaning in a wide variety of experimental settings. It helps to show attentional focus, emotionality, social relationships, thinking styles, and individual differences. It allows users to look under the hood of works of literature. LIWC's design has made it a favorite for psychologists, but it also finds use in marketing, Twitter analysis, mental health diagnostics, and much more [25]. Initially, the LIWC software did not include a stress dictionary. Wang et al. created the stress dictionary following the procedures and steps established by Pennebaker to ensure desired psychometric properties [27]. Also, it is fair to assume that the stress dictionary is a sub-dictionary of the negative emotion dictionary in the LIWC software. Figure 4.3 depicts the words and phrases associated with high-stress and low-stress users. It shows the words frequently used among high and low-stress users.

- **Sentiment Analysis**

Sentiment analysis is useful to a wide range of problems, that are relevant to human-computer interaction researchers. It has also found applications in the fields such as sociology, marketing, and advertising, psychology, economics, and political science. The sentiment analysis of Twitter data has gained much



Figure 4.3: Words and Phrases associated with low-stress (blue) and high-stress (red) users

attention as a topic of research.

The positive, neutral, and negative sentiments within the timeline tweets of the influenced users were computed using the ‘VADER’ (Valence Aware Dictionary for Sentiment Reasoning) sentiment analysis tool. VADER sentimental analysis relies on a dictionary that maps lexical features to emotion intensities known as sentiment scores. The sentiment score of a text can be obtained by summing up the intensity of each word in the text. We concatenated all the timeline tweets for each user and the cleaned tweets were analyzed using VADER. As part of the cleaning process, we replaced all emojis and emoticons in the stop words and punctuations - free texts. Some of the patterns, for example, HTTPS, were also removed. The output polarity of 1 implies positive sentiments, 0 stands for neutral and -1 indicates negative sentiments.

- **User’s Interest and Similarity**

The main dataset for this research had influencer-user pairs along with the details on news that were shared. To compute the cosine similarity between the user’s interest and shared the news, we determined the user’s interest using the following two approaches : (1) User interest is the concatenation of all remaining shares of the influenced user, as discussed in Section 3.0.2. (2) User Interest is the concatenation of all timeline tweets of the influenced user.

For this feature computation, we created an LDA model on Wikipedia using Gensim and extracted 100 topics. Gensim is designed to process raw text using unsupervised machine learning algorithms [28]. The algorithms in Gensim, such as Word2Vec, FastText, Latent Semantic Indexing, Latent Dirichlet Allocation

(LDA), etc., discover the semantic structure of documents by inspecting statistical co-occurrence patterns within a corpus of training documents [28]. Once these patterns are found, they are used for retrieving topical similarity against other documents.

- **Emotional features**

We computed additional emotional features such as anger, joy, sadness, fear, disgust, anticipation, surprise, and trust by using the Emotion Intensity Lexicon (NRC-EIL) [29] and the approach proposed in [30]. The NRC Emotion Intensity Lexicon is a list of English words with real-valued scores of intensity for eight basic emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust). For a given word w and emotion e , the scores range from 0 to 1.

- A score of 1 indicates that the w conveys the highest amount of emotion e .
- A score of 0 indicates that the w conveys the lowest amount of emotion e .

The lexicon has close to 10,000 entries for eight emotions. It includes common English terms as well as terms that are more prominent in social media platforms, such as Twitter. It includes terms that are associated with emotions to various degrees. The concatenated timeline tweets were utilized to extract emotional intensity scores for each user.

- **Behavioral features**

We analyzed the temporal and topical signature of users' sharing behavior, showing how they exhibit distinct behavioral patterns. The behavioral features

of the users were calculated using the temporal information of retweets and timeline tweets. There were two features computed under this category. In the first feature, we calculated the time difference between the night posts and day posts by a user upon his/her total posts. A positive value indicates that the user was active mostly during nighttime. For our second feature, we calculated the average of the time taken by the user to tweet for a given set of timeline tweets.

News Related features:

In our implementation, we considered features that confirmed the findings of both Shrestha and Spezzano, and Horne and Adali [31, 32]. The following stylistic, psychological, and complexity features are computed for both the title and body text of the news.

- **Stylistic Features**

We used the subset of LIWC features that represent the the functionality of text, including word count (WC), words per sentence (WPS), number of personal (I, we, you, she/he – one feature each) and impersonal pronouns, number of exclamation marks (exlam), number of punctuation symbols (allPunc), number of quotes (quote).

Regarding the part of speech features, we used the Python Natural Language Toolkit part of speech (POS) tagger to compute the number of nouns (NN), proper nouns (NNP), personal pronouns (PRP), possessive pronouns (PRP), Wh-pronoun (WP), determinants (DT), Wh-determinants (WDT), cardinal numbers (CD), adverbs (RB), verbs (VB), past tense verbs (VBD), gerund

or present participle verbs (VBG), past participle verbs (VBN), non-3rd person singular present verbs (VBP), and third person singular present verbs (VBZ).

- **Psychology Features**

Social psychology is the study of the dynamic interaction between individuals and the people around them. Psychology plays an important role in the field of social media marketing. One needs to tap into the emotions for developing long-term customer relationships. The science of social psychology came into existence when scientists first started to formally measure the thoughts, feelings, and behaviors of human beings.

We computed the positive (pos) and negative (neg) sentiment metrics using the LIWC tool. As Shrestha and Spezzano, [32] and Ghanem et al., [33] recently showed that emotions play a key role in deceiving the reader and can successfully be used to detect false information. In addition to the sentiment metrics, we calculated emotional features, such as anger, joy, sadness, fear, disgust, anticipation, surprise, and trust by using the Emotion Intensity Lexicon (NRC-EIL) [29] and the approach proposed in [30]. The NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). We computed these scores for both text and title of different news.

- **Complexity Features**

SMOG - The complexity of text in natural language processing depends on how easily the reader can read and understand a text. We used the Simple Measure

of Gobbledygook Index (SMOG) readability measure as a complexity feature in our analysis [34]. Readability Formula estimates the years of education a person needs to understand a piece of writing. McLaughlin created this formula as an improvement over other readability formulas [34]. Higher scores of these readability measures indicate that the text is easier to read.

Lexical Diversity - It is a measurement of how many different lexical words there are in a text. Lexical words are words such as nouns, adjectives, verbs, and adverbs that convey meaning in a text.

Type-Token Ratio (TTR) - A type-token ratio (TTR) is the total number of UNIQUE words (types) divided by the total number of words (tokens) in a given segment of language. The closer the TTR ratio is to 1, the greater the lexical richness of the content.

Average Word Length (avg wlen) - Average length of the words (count of characters).

Network Related features:

- **Weak and Strong ties**

According to Ma et al., perceived tie strength in online social networks is positively associated with news sharing intention in social media [13]. M. Granovetter showed that strong ties are friends and weak ties represent acquaintances [35]. The paper talks about interpersonal relationships between disparate sets of people and how they hold their networks together. As per Granovetter, the information diffuses faster among people with strong ties. While strong ties

indicate a large number of shares between two people in a network, weak ties depict fewer shares between influencer and follower.

Although there can be different strategies for determining tie strength between the influencer and influenced user pairs, the research focuses on the following two definitions.

Receiver's perspective - For a given influencer-follower pair in the dataset, the percentage of retweets by the influenced user on the tweets by an influencer user was calculated.

Time-based analysis - This is computed as the average of the time taken by an influenced user to share from an influencer.

- **Twitter Follower-Following ratio (TFF) ratio**

This is computed as the ratio of follower to following counts. The data from FakeNewsNet has the details of users and their respective followers. The formula for the calculation of TFF is shown below. An addition of 1 keeps the ratio from falling to infinity or zero.

$$TFF = \frac{\#Follower + 1}{\#Following + 1}$$

While a ratio of 2.0 or above shows that you are a popular person and people want to hear what you have to say, a ratio of around 1.0 means that you are respected among your peers. TFF is proposed as a significant network feature by Shu et al. [11].

- **Pagerank and Degree Centrality**

PageRank accounts for link direction. It can help uncover influential or important nodes whose reach extends beyond just their direct connections. Each node in a network is assigned a score based on its number of incoming links (in-degree). The recursive equation for PageRank is given by,

$$PR(p_i) = \frac{1 - d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

where $M(p_i)$ is the set of nodes pointing to p_i , and $L(p_j)$ is the number of nodes p_j points to.

Degree centrality indicates the relative importance of a node within the network. In general, for a given node x it is calculated as a ratio between the number of nodes connected with node x and the total number of all nodes in the network (decreased by one). We computed in-degree and out-degree centrality using the library, `networkx`. The dataset used from FakeNewsNet contained details on users and their respective followers.

Since our research employs DOI theory, it requires features from each category of news, user, and network. This has increased the number of features in our study. The machine learning algorithms are capable of investigating many features compared to the manual analysis. Therefore, it suits the purpose.

4.1 Methods

For the first part of the research question mentioned in Section 1, the classifiers such as Logistic Regression, Support Vector Machine (SVM), Random Forest, XG-Boost, and Extra Trees Classifier will be used.

4.1.1 Survival Analysis

As mentioned in Chapter 1, we will also be predicting the time to share content by a user, using Random Survival Forest. The Random Survival Forest package provides a python implementation of the survival prediction method originally published by Ishwaran et al. [36].

Overview

Survival analysis is a sub-field of statistics where the goal is to analyze and model data where the outcome is the time until an event of interest occurs [37]. Broadly speaking, survival analysis methods can be classified into two main categories: statistical methods and machine learning-based methods. Statistical methods share a common goal with machine learning methods in that both are expected to make predictions of the survival time and estimate the survival probability at the estimated survival time. Machine learning methods are usually applied to high-dimensional problems, while statistical methods are generally developed to handle low-dimensional data.

Depending on the assumptions made and the way parameters are used in the model, the traditional statistical methods can be subdivided into three categories: (i) non-parametric models, (ii) semi-parametric models, and (iii) parametric models.

Machine learning algorithms such as survival trees, bayesian methods, neural networks, and support vector machines are included under a separate branch. Several advanced machine learning methods, including ensemble learning, active learning, transfer learning, and multitask learning methods, are also included in this category. The object of primary interest is the survival function, conventionally denoted S , which is defined as

$$S(t) = \Pr(T > t)$$

where t is time, T is a random variable denoting the time of an event, and "Pr" stands for probability. That is, the survival function is the probability that the time of an event is later than some specified time t .

4.1.2 Censoring

Censoring is common in survival analysis. It is a form of missing data problem in which time to event is not observed because,

- Of the termination of study before all recruited subjects have shown the event of interest or
- The subject has left the study before experiencing an event.

Types

Left Censoring - If the event of interest has already happened before the subject is included in the study but it is not known when it occurred, the data is said to be left-censored.

Right Censoring - If only the lower limit l for the true event time T is known such

that T , this is called right censoring. Right censoring will occur, for example, for those subjects whose birth date is known but who are still alive when they are lost to follow-up or when the study ends. We generally encounter right-censored data and the Figure 4.4 shows a pictorial representation with study time vs. subjects.

Interval Censoring - When it can only be said that the event happened between two observations or examinations, this is interval censoring.

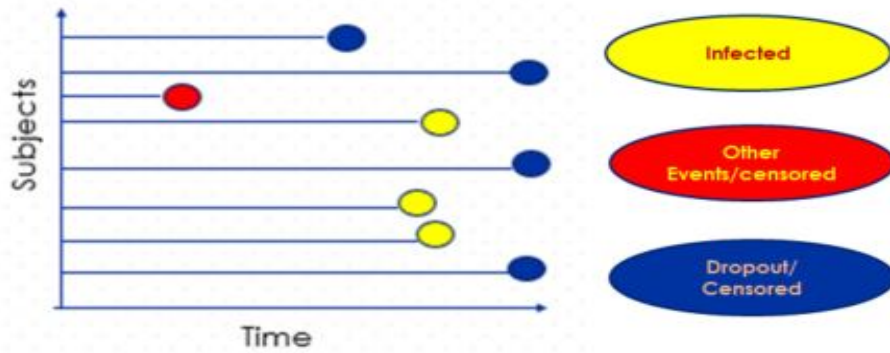


Figure 4.4: Right Censoring

Random Survival Forest

This method is used with right-censored survival data [36]. Right censoring occurs when a subject leaves the study before an event occurs, or the study ends before the event has occurred. A random survival forest consists of random survival trees. Using independent bootstrap samples, each tree is grown by randomly selecting a subset of variables at each node and then splitting the node using a survival criterion involving survival time and censoring status information. The tree is considered fully grown when each terminal node has no fewer than $d_0 > 0$ unique deaths. The estimated cumulative hazard function (CHF) for a case is the Nelson–Aalen estimator for the

case's terminal node. The ensemble is the average of these CHF's. Because trees are grown from in-bag data, an out-of-bag (OOB) ensemble can be calculated by dropping OOB cases down their in-bag survival trees and averaging. The predicted value for a case using the OOB ensemble does not use survival information for that case, and, therefore, it can be used for a nearly unbiased estimation of prediction error. From this, other useful measures can be derived, such as variable importance values for filtering and selecting variables.



Figure 4.5: Random Survival Tree

Survival Function and Time Estimation

Given a new instance i described by the feature vector X_i , survival analysis estimates a survival function S_i that gives the probability that the event for the instance i will occur after time t , i.e., $S_i(t) = Pr(T_i \geq t)$.

Let $i = (u, v)$ be an influencer - follower pair. Consider the case of user u being followed by user v and user v is predicted to share the content from user u in the time interval $[t_a, t_b]$. The time when the event of interest happens for the instance $i = (u, v)$ is denoted by $T_i (T_i \in [t_a, t_b])$. We also assume the time interval $[t_a, t_b]$ to be divided into k time periods, e.g., days, weeks, months, etc. From the survival analysis model, we will have the following probabilities: $S_i(t_a) = Pr(T_i > t_a), S_i(t_a + k) = Pr(T_i > t_a + k) \dots, S_i(t_b) = Pr(T_i > t_b)$. The probability that the sharing will occur

in the time interval $[t_a + (h-1)k, t_a + hk)$ is given by,

$$Pr(t_a + (h-1)k \leq T_i < t_a + hk) = Pr(T_i \geq t_a + (h-1)k) - Pr(T_i \geq t_a + hk)$$

From the survival function, the probability density function is that the event will occur in the given interval $[t_a + (h-1)k, t_a + hk)$. Let x denote all the intervals in the probability density function and time is the expected value, $E(x)$, which is estimated as,

$$E(x) = \int_{t_a}^{t_b} x \quad (Pr(t_a + (h-1)k \leq T_i < t_a + hk))$$

$$E(x) = \sum_{x=t_a}^{t_b} x \quad (Pr(t_a + (h-1)k \leq T_i < t_a + hk))$$

CHAPTER 5:

EXPERIMENTS AND RESULTS

5.1 Experimental Setting

We tested our features using the binary classification of whether the user will share or not a given piece of news from his/her influencer on various machine learning algorithms namely Logistic Regression, Support Vector Machine (SVM), Random Forest, XGBoost, and Extra Trees Classifier. Since the dataset was highly imbalanced (label 1: 3,144, label 0: 14,936), we used class weighting to deal with it. We also used the simple imputation method using the mean to obtain the missing observations. To evaluate the performances, we considered the Area Under the Receiver Operating Characteristics (AUROC) and Average Precision, which are well-suited for unbalanced data, and performed 10 - fold cross-validation. The XGBoost classifier outperformed among them for fake and real news sharing. The classifier given an average precision of 95.23 with an AUROC of 97.39 for real news sharing. For fake news sharing, it reports the AUROC and average precision as 97.34 and 88.43, respectively, as shown in Tables 5.1 and 5.2. XGBoost is a decision-tree-based ensemble machine learning algorithm that uses a gradient boosting framework.

5.1.1 Comparison with Baselines

We compared the results of the classifier with the baselines in Chapter 2: 1) Independent Cascade Model, and 2) Linear Threshold Model. Classification algorithms in machine learning use input training data to predict the likelihood that subsequent data will fall into one of the predetermined categories. The likelihood is the probability value. We used Bernoulli Distribution and Jaccard's Index to determine this probability for the Independent Cascade model [38]. Unlike the ML-based classifiers, the propagation probabilities of baselines do not depend on news or user features.

Bernoulli Distribution

In this model, the activated node 'v' influences inactive node 'u' using a fixed probability or threshold. A successful attempt is when the 'u' is activated. Each attempt, that is linked with some action, can be viewed as a Bernoulli trial. The success probability is the ratio of the number of successful attempts over the total number of trials [38]. Hence, the influence probability of 'v' on 'u' using Maximum Likelihood Estimator (MLE) is estimated as:

$$\rho_{v,u} = \frac{A_{v2u}}{A_v}$$

Jaccard Index

It is defined as the size of the intersection divided by size of the union of the sample under study. Goyal et al., adapted the Jaccard Index as follows [38],

$$\rho_{v,u} = \frac{A_{v2u}}{A_{u|v}}$$

In the Linear Threshold Model, each edge $e = (u, v)$ is associated with a weight $b_{u,v}$, as explained in Section 2. Also, there is a propagation probability required for information diffusion. In this research, we have used two approaches to calculate the probability. In the first approach, the inverse of the in-degree of the influenced user is multiplied with the count of influencers as activation probability [39]. The second approach takes the values from a set of [0.1,0.01,0.001] and it is multiplied with the count of influencers of each user in the main dataset [39]. Once the probabilities were computed for both Independent Cascade and Linear Threshold models, we used 10 - fold cross-validation and evaluated the model using AUROC and average precision. We also used Synthetic Minority Oversampling Technique (SMOTE) for handling imbalanced data.

The results of the baselines are also reported on Tables 5.1 and 5.2. Out of the two baseline models, Independent Cascade Model (ICM) outperformed with an AUROC of 67.67 and an average precision of 56.73 for fake news sharing. For real news sharing, ICM gave an AUROC of 64.42 and an average precision of 48.54. Therefore, we can conclude that our model performs better than the baseline models. Unlike the ML-based classifiers, baselines lack multi-features from the user, news, network and hence explains the low AUROC and average precision.

5.1.2 Analysis of Real and Fake News Sharing

The metrics that help in the interpretation of probabilistic forecasts for binary classification problems are ROC curves and Precision-Recall curves. ROC curves are appropriate when the observations are balanced between each class, whereas precision-recall curves are appropriate for imbalanced datasets. Shrestha and Spezzano, used

Table 5.1: Classifier results: Results of classification for Fake News Sharing

Model Name	AUROC	Average Precision
Logistic Regression	93.78	65.22
SVM	89.08	36.58
Random Forest	97.86	85.11
Extra Trees	96.82	76.75
XGBoost	97.34	88.43
Baselines		
Independent Cascade Model	67.70	53.03
Linear Threshold Model	63.78	87.45

Table 5.2: Classifier results: Results of classification for Real News Sharing

Model Name	AUROC	Average Precision
Logistic Regression	92.39	84.15
SVM	87.05	66.91
Random Forest	97.38	94.61
Extra Trees	95.26	85.94
XGBoost	97.39	95.23
Baselines		
Independent Cascade Model	64.48	46.05
Linear Threshold Model	59.35	63.77

the same dataset from Politifact for the detection of fake news. We used the news features from Shrestha and Spezzano, and received an AUROC close to 93% (using just the news features) for predicting fake and real news sharing [32].

The Figures in 5.1 explain the performance of classifiers for each group of features, such as user, network, and news. While the news features contribute to 93% of decision-making, this can be boosted using our network and user features. We had an overall of 4% increase in the AUROC as reported in Tables 5.1 and 5.2. The visualizations of the feature ablation study are in Figures 5.2, 5.3, 5.4 and 5.5. The news title and text features give consistently good AUROC scores for each classifier. Figure 5.3 shows that the tie strength is an important feature for the network category. Also, under the user category, the cosine similarity between user’s interests and shared news is a key feature along with the explicit features. Average precision indicates that every time a classifier predicts one class, what is the percentage that the classifier is correct. The low average precision for fake news sharing is because of its small sample size.

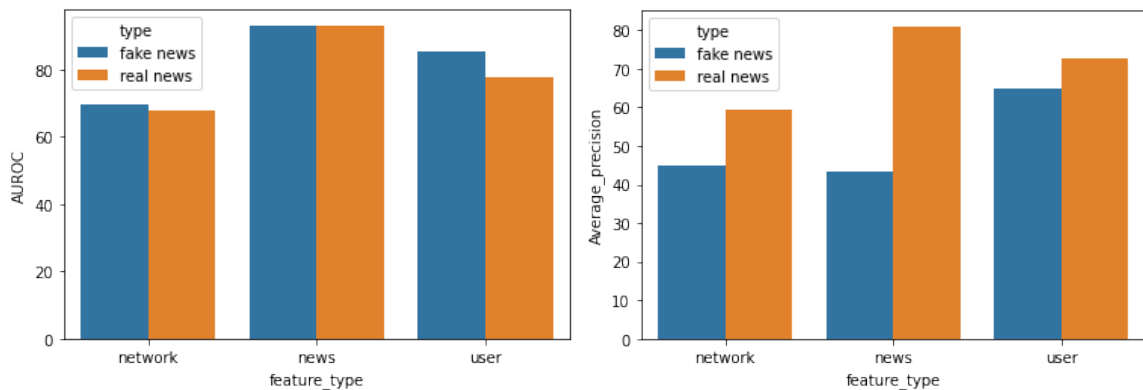


Figure 5.1: Contribution from each Feature Groups

The personality insights from IBM Watson were further studied using random

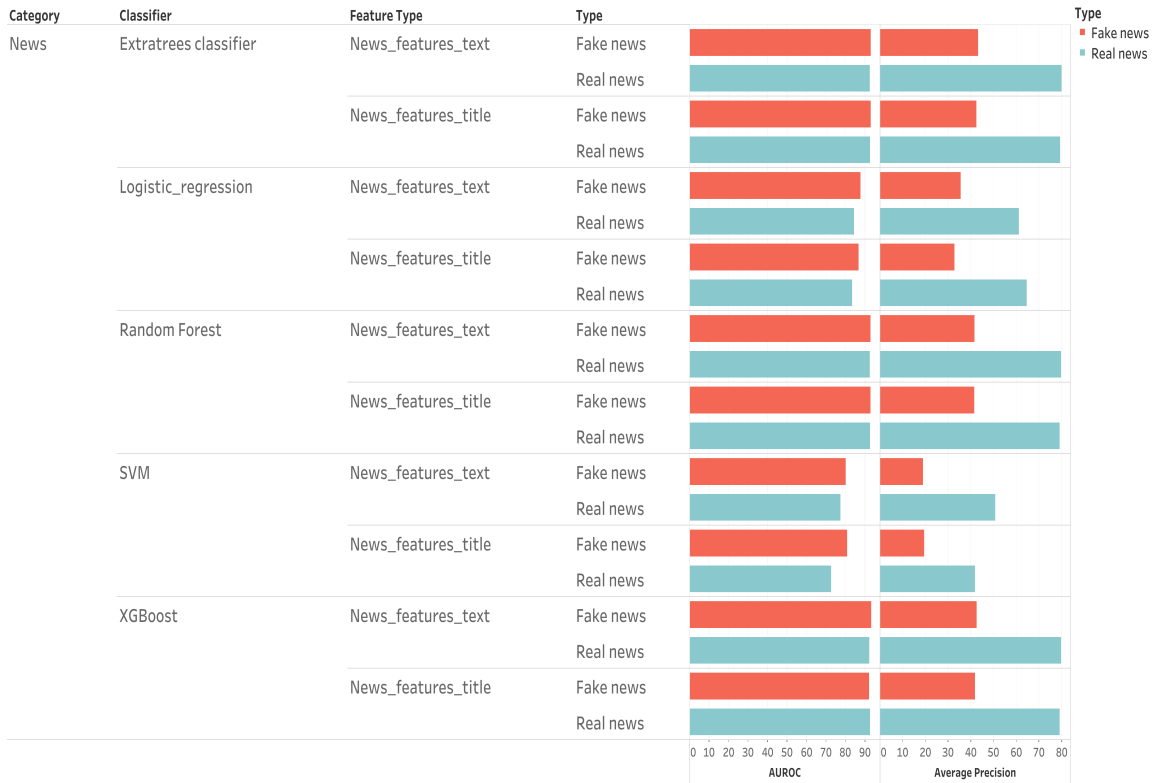


Figure 5.2: Feature Ablation Results (News)

forest based feature importance and the scores are displayed in Figure 5.6. Feature importance is a technique of assigning importance scores to the features. The score indicates the relative importance of each feature when making a prediction. It can be used to improve a predictive model. The top five features for both fake and real news sharing are selected as significant for this discussion. ‘Need’ describes which aspects of the Tweet are likely to resonate with a user and ‘Value’ indicates the motivating factors that influenced the user to share that piece of news. The top five ‘Need’ related features, show that real news sharing users, display characteristics like ideal and love. On the other hand, important ‘need’ features observed for users that share fake news include curiosity, closeness, structure and harmony. Among the



Figure 5.3: Feature Ablation Results (Network)

‘values’, self-enhancement and openness to change have high importance in fake news sharing. For real news sharing, self-transcendence, conservation, openness to change is important. This analysis provides insights into user characteristics and establishes some correlation between personality and news sharing habits.

This research incorporated multi-level influences of the user, news, and network attributes. After screening for the best features, we analyzed those features on the

same classifiers. Figure 5.8 shows the simplified view of all the key features that have the maximum AUROC and average precision. It is apparent that the tie strength, news features, and explicit and user similarity attributes, contribute more to the degree of separability. We further performed classification using the key features and the results are outlined in Tables 5.4 and 5.3. The XGBoost classifier outperformed in both real and fake news sharing. Using the key features, the XGBoost classifier given an AUROC of 97.24 for fake news sharing and attained an AUROC of 97.53 for real news sharing. In order to observe the distinguishing factors between fake and real news sharing, we conducted a features importance study using random forest on the key features. The results are shown on Figure 5.7. Among the top features, fake news sharing gives high importance to the number of proper nouns, adverbs, possessive nouns, words per sentence and negative emotions. And real news sharing depends more on the tie strength, similarity, punctuations, words per sentence, and negative emotions.

5.1.3 Experimental Setting for Survival Analysis

As mentioned in Section 4.0.2, we used Random Survival Forest (RSF) to predict the time to event for our study. The implementations of the model are available in Python [40]. The training time for this package was significantly high (more than 48 hours), hence we used approximately 800 random instances from the main data set for both real and fake news sharing. As machine learning algorithms tend to increase accuracy by reducing error, they do not deal with class distribution or imbalanced datasets. To deal with unbalanced data, oversampling using Synthetic Minority Oversampling Technique (SMOTE) was utilized and the algorithm was further run on 5-fold cross-

validation. The dataset had all the features computed for the user, news, and network categories. The details of the dataset after oversampling is shown in Tables 5.5 and 5.6. The time to event is the time taken by a user to respond to a given piece of news. The dataset was right-censored and assigned a value of -1 for the censored instances. For regression, we considered Ridge and Lasso regression. Since regression cannot deal with censored instances, we approximated the occurrence time of those instances with a time value $t_c \gg t_b$.

To compare the performance of RSF with baselines, we considered two metrics, namely concordance-index and Mean Absolute Error (MAE). The Tables 5.7 and 5.8 shows the results of our analysis. The difference between the predicted and actual time to share the content on Twitter constituted the MAE in the regression model. To compare the time series of the actual and predicted number of units that experienced the event at each time t in the random survival forest. We calculated the actual density function of the data using the Kaplan-Meier estimator and compared it to the average of all predicted density functions. From the analysis, while the baselines outperform the proposed method in real news sharing, RSF equals the c-index score with baselines for fake news sharing. The concordance index or c-index depends on the ordering of the instances in the dataset. The concordance index for real news sharing detection is slightly lower compared to baseline, and the concordance index of Random Survival Forest (RSF) is comparable to baseline for the fake news sharing. Furthermore, in comparison to the regression baseline models, the Mean Absolute Error (MAE) is significantly less in RSF for both real and fake news sharing.

Table 5.3: Key Features: Results of classification for Fake News Sharing

Model Name	AUROC	Average Precision
Random Forest	96.90	85.28
Extra Trees	96.59	81.90
XGBoost	97.24	87.43

Table 5.4: Key Features: Results of classification for Real News Sharing

Model Name	AUROC	Average Precision
Random Forest	97.08	94.24
Extra Trees	95.92	90.14
XGBoost	97.53	95.24

Table 5.5: Dataset - Fake News Sharing

	Politifact
Count of Users	80
Influenced Users	352
Not Influenced Users	352

Table 5.6: Dataset - Real News Sharing

	Politifact
Count of Users	80
Influenced Users	399
Not Influenced Users	399

Table 5.7: Regression results: Results of Regression for Real News Sharing

Model Name	C-Index	Mean Absolute Error
Random Survival Forest	0.68	1.29
Baselines		
Ridge Regression	0.768	53605900
Lasso Regression	0.77	53547000

Table 5.8: Regression results: Results of Regression for Fake News Sharing

Model Name	C-Index	Mean Absolute Error
Random Survival Forest	0.81	1.25
Baselines		
Ridge Regression	0.81787	40180300
Lasso Regression	0.80855	41328300



Figure 5.4: Feature Ablation Results (User)

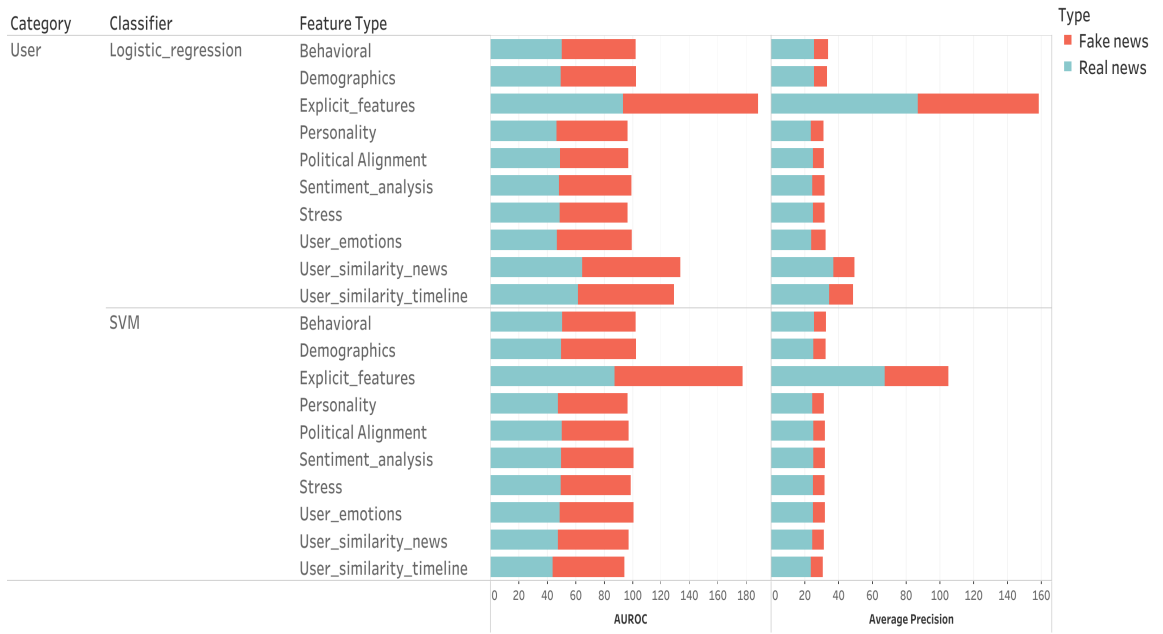


Figure 5.5: Feature Ablation Results (User)

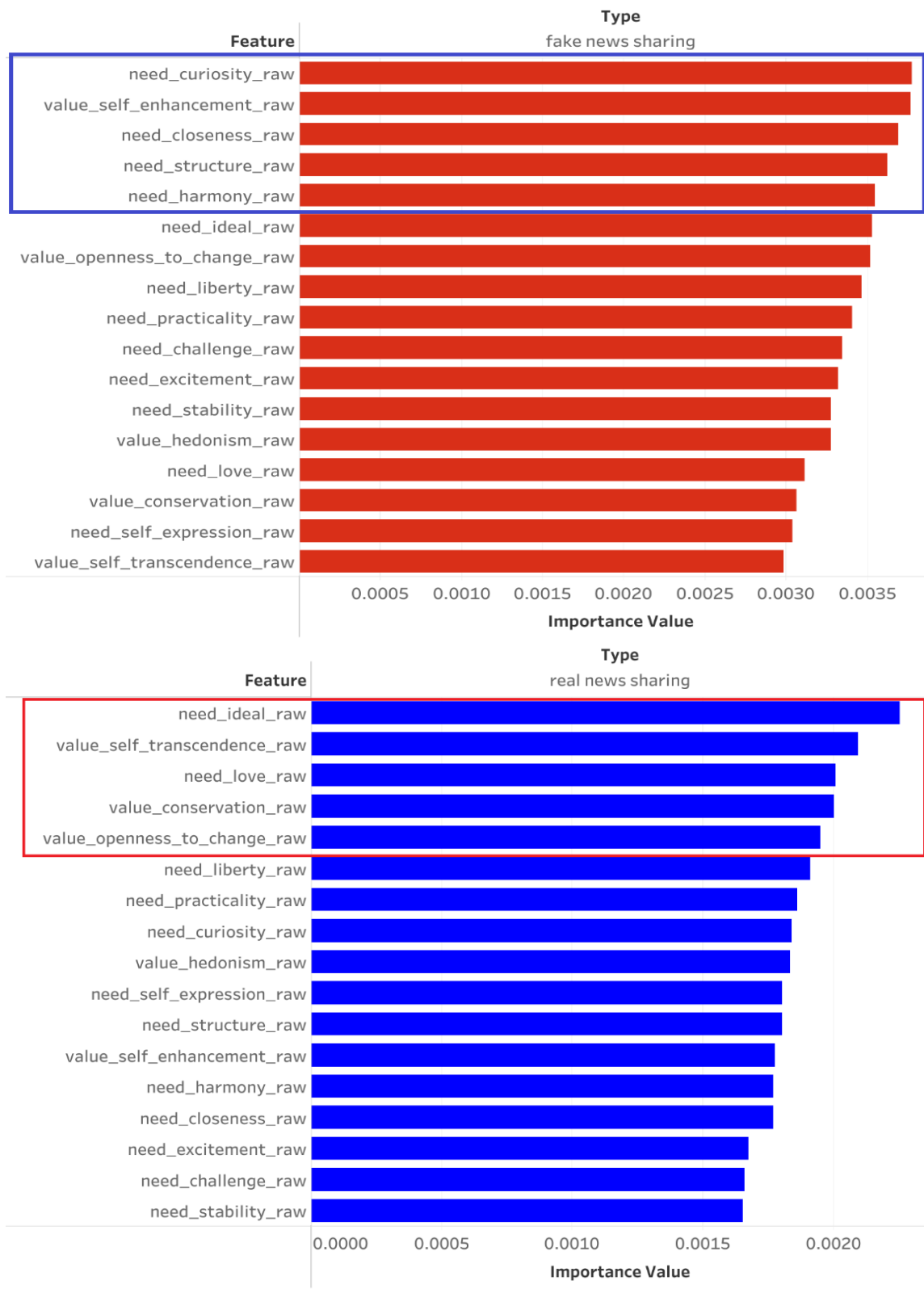


Figure 5.6: Feature importance of personality features for fake (red) and real news sharing (blue).

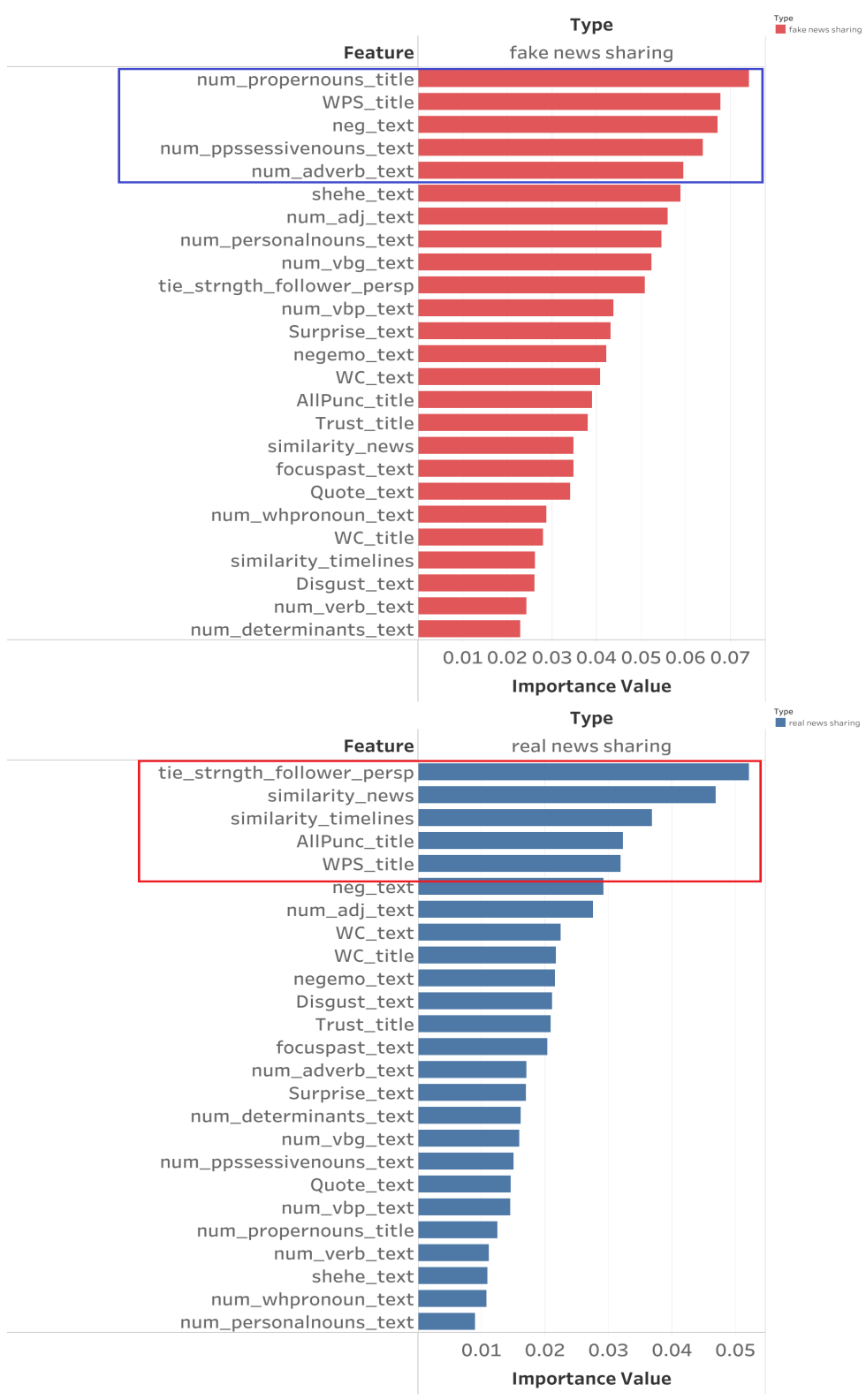


Figure 5.7: Feature importance of key features for fake (red) and real news sharing (blue).



Figure 5.8: Simplified Key Feature Ablation Results

CHAPTER 6:

CONCLUSIONS AND FUTURE WORK

Fake news detection has become critical in today's internet age. This study discussed the implementation of a DOI theory-based method to predict, if and when the user will share real or fake news. User behavior and factors that correlate with a tendency to share fake or real news were determined. Real and fake news sharing was better predicted with the user, news, and social network characteristics.

In this study, we presented the design and real-world evaluation of the prediction of social sharing. Building on DOI theory, we demonstrated the factors that predict fake and real news sharing among users. The analysis supported the hypothesis that real and fake news sharing is better predicted with the user, news, and social network characteristics.

The proposed approach of combining news, user, and network features boosted the AUROC by 30% in comparison to network-based baseline models. Among the key features, we found that real and fake news sharing shows a high dependency on user similarity, news text, news title, tie strength, and explicit attributes. For the second part of our research, the survival model was used to predict the time to share a given piece of news by a user. The mean absolute error attained by the random survival forest was low and reliable. Although the Concordance Index (C-Index) for real news sharing was slightly lower compared to baseline, the C-Index of Random

Survival Forest (RSF) is comparable to baseline for the fake news sharing.

6.0.1 Future Work

This work demonstrated how user and news data can be used to predict some aspects of fake news sharing. This approach can be further expanded and made more robust in several ways. Today, a lot of information is shared via graphics like images, gifs, etc. Analyzing in-tweet gifs and images can boost the prediction of fake news sharing and has the potential to enhance personality insights. The training of the model with a bigger data set was one of the challenges in this work. The non-parametric survival model used here took significant processing time. In the next stage of this work, a parametric survival model can be explored to see if this helps reduce the training time. Another interesting aspect that can be explored is the rate of news propagation. Fake news tends to spread faster and further compared to real news. So, estimating the speed of diffusion of news can potentially improve the detection of misinformation in social media. Yet another opportunity is expanding this study to other social media platforms and harnessing additional user and news features. Thus, further optimization of the methodology to improve the processing time, exploring more types of data, extracting dynamic propagation parameters of the news, and expanding these models to other social media platforms can yield more insights into fake news prediction and prevention.

6.0.2 Limitations

The survival models are employed to analyze the expected duration of time until one event occurs. In this research, the event is the user's decision to share or not-share a piece of news. One of the drawbacks for this study was that the package took significantly high training time for the main dataset. The issue was solved by using a smaller data set, 800 instances of sampled data. It will be insightful to check the performance on other survival models such as parametric models. The parametric model has the benefit of easily incorporating features into the model and inference procedures. Another limitation was that of the imputation method utilized. We computed features for the classification task and utilized simple imputation with mean to deal with missing observations. This approach is not recommended as it introduces bias.

The research also implemented Diffusion of Innovation theory (DOI) theory to integrate multiple levels of influence by user, innovation, and network. Our dataset was compact, the theory needs to be tested with a bigger dataset with conflicting data. This approach can contribute to either adapt the DOI theory or understand its effect on larger datasets.

REFERENCES

- [1] A Wolfman-Arent. The ongoing battle between science teachers and fake news. *National Public Radio (NPR): Morning Edition*, 2017.
- [2] Lance E Mason, Dan Krutka, and Jeremy Stoddard. Media literacy, democracy, and the challenge of fake news. *Journal of Media Literacy Education*, 10(2):1–10, 2018.
- [3] Ying Lin. 10 twitter statistics every marketer should know in 2020 [infographic].
- [4] Paulo Shakarian, Abhivav Bhatnagar, Ashkan Aleali, Elham Shaabani, Ruocheng Guo, et al. *Diffusion in social networks*. Springer, 2015.
- [5] Anna Sophie Kümpel, Veronika Karnowski, and Till Keyling. News sharing in social media: A review of current research on news sharing users, content, and networks. *Social media+ society*, 1(2):2056305115610141, 2015.
- [6] Everett M Rogers. *Diffusion of innovations*. Simon and Schuster, 2010.
- [7] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.
- [8] Chei Sian Lee and Long Ma. News sharing in social media: The effect of gratifications and prior experience. *Computers in human behavior*, 28(2):331–339, 2012.

- [9] Kai Shu, Suhang Wang, and Huan Liu. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 312–320, 2019.
- [10] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1):22–36, 2017.
- [11] Kai Shu, Suhang Wang, and Huan Liu. Understanding user profiles on social media for fake news detection. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 430–435. IEEE, 2018.
- [12] Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G Rand. Understanding and reducing the spread of misinformation online, 2019.
- [13] Long Ma, Chei Sian Lee, and Dion H Goh. Understanding news sharing in social media from the diffusion of innovations perspective. In *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, pages 1013–1020. IEEE, 2013.
- [14] Andrew Guess, Jonathan Nagler, and Joshua Tucker. Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science advances*, 5(1):eaau4586, 2019.
- [15] Waheeb Yaqub, Otari Kakhidze, Morgan L Brockman, Nasir Memon, and Sameer Patil. Effects of credibility indicators on social media news sharing intent.

- In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.
- [16] Kai Shu, H Russell Bernard, and Huan Liu. Studying fake news via network analysis: detection and mitigation. In *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, pages 43–65. Springer, 2019.
- [17] Eben Kenah. Non-parametric survival analysis of infectious disease data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(2):277–303, 2013.
- [18] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.
- [19] Kai Shu, Deepak Mahudeswaran, and Huan Liu. Fakenewstracker: a tool for fake news collection, detection, and visualization. *Computational and Mathematical Organization Theory*, 25(1):60–71, 2019.
- [20] Yair Neuman. *Computational personality analysis: Introduction, practical applications and novel directions*. Springer, 2016.
- [21] Libby Hemphill, Aron Culotta, and Matthew Heston. Polar scores: Measuring partisanship using social media content. *Journal of Information Technology & Politics*, 13(4):365–377, 2016.

- [22] Joshua M Chamberlain, Francesca Spezzano, Jaclyn J Kettler, and Bogdan Dit. A network analysis of twitter interactions by members of the us congress. *ACM Transactions on Social Computing*, 4(1):1–22, 2021.
- [23] Zijian Wang, Scott Hale, David Ifeoluwa Adelani, Przemyslaw Grabowicz, Timo Hartman, Fabian Flöck, and David Jurgens. Demographic inference and representative population estimates from multilingual social media data. In *The World Wide Web Conference*, pages 2056–2067, 2019.
- [24] Sharath Chandra Guntuku, Anneke Buffone, Kokil Jaidka, Johannes C Eichstaedt, and Lyle H Ungar. Understanding and measuring psychological stress using social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 214–225, 2019.
- [25] James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001, 2001.
- [26] Yla R Tausczik and James W Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54, 2010.
- [27] Wei Wang, Ivan Hernandez, Daniel A Newman, Jibo He, and Jiang Bian. Twitter analysis: Studying us weekly trends in work stress and emotion. *Applied Psychology*, 65(2):355–378, 2016.
- [28] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges*

- for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. <http://is.muni.cz/publication/884893/en>.
- [29] Saif M Mohammad. Word affect intensities. *arXiv preprint arXiv:1704.08798*, 2017.
- [30] Ashlee Milton, Leveson Batista, Garrett Allen, Siqi Gao, Yiu-Kai D Ng, and Maria Soledad Pera. “don’t judge a book by its cover”: Exploring book traits children favor. In *Fourteenth ACM Conference on Recommender Systems*, pages 669–674, 2020.
- [31] Benjamin Horne and Sibel Adali. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, 2017.
- [32] Anu Shrestha and Francesca Spezzano. Textual characteristics of news title and body to detect fake news: A reproducibility study. In Djoerd Hiemstra, Marie-Francine Moens, Josiane Mothe, Raffaele Perego, Martin Potthast, and Fabrizio Sebastiani, editors, *Advances in Information Retrieval - 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 - April 1, 2021, Proceedings, Part II*, volume 12657 of *Lecture Notes in Computer Science*, pages 120–133. Springer, 2021.
- [33] Bilal Ghanem, Paolo Rosso, and Francisco Rangel. An emotional analysis of false information in social media and news articles. *ACM Transactions on Internet Technology (TOIT)*, 20(2):1–18, 2020.

- [34] G Harry Mc Laughlin. Smog grading-a new readability formula. *Journal of reading*, 12(8):639–646, 1969.
- [35] Olaf Zorzi. Granovetter (1983): The strength of weak ties: A network theory revisited. In *Schlüsselwerke der Netzwerkforschung*, pages 243–246. Springer, 2019.
- [36] Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, Michael S Lauer, et al. Random survival forests. *The annals of applied statistics*, 2(3):841–860, 2008.
- [37] Ping Wang, Yan Li, and Chandan K Reddy. Machine learning for survival analysis: A survey. *ACM Computing Surveys (CSUR)*, 51(6):1–36, 2019.
- [38] Amit Goyal, Francesco Bonchi, and Laks VS Lakshmanan. Learning influence probabilities in social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 241–250, 2010.
- [39] Yuchen Li, Ju Fan, Yanhao Wang, and Kian-Lee Tan. Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1852–1872, 2018.
- [40] Stephane Fotso et al. PySurvival: Open source package for survival analysis modeling, 2019–.