

A Prediction Model for Information Diffusion in Online Social Network

John K. Omoniyi

Computer Science Dept.
Federal University of Technology
Akure, Ondo State, Nigeria

Folasade Adedeji

Computer Science, Babcock
University, Ilesan-Remo, Nigeria

Joshua J. Tom

Department of Mathematics &
Computer Science, Elizade
University Ilara Mokin,
Nigeria

ABSTRACT

The emergence of online social networks has brought many new platforms, e.g., Facebook, Flickr, YouTube, Sina Microblog, LinkedIn, and Twitter. These platforms are imperative constituents within the diffusion of information at an expansive scale, and Twitter is among the foremost utilized microblogging and online social organizing administrations. In Twitter, a title, phrase, or point tweeted at a greater rate than others are called a "trending topic" or "trend," and it becomes imperative to make available ways to evaluate this phenomenon. Assessing information diffusion appears to be an unsolvable perplex as these "trending topics" constitute a flood of views, thoughts, recommendations, considerations, proposals, etc., produced by users of these social networks. This paper thoroughly examined Twitter's trending topics in September 2019. We accessed Twitter's trends API for the month's trending topics and concocted six criteria to assess the dataset. These six criteria are location, lexical analysis, trending time, tweet volume, promo/giveaway, and social media influencer. Based on the criteria earlier mentioned, a prediction model was developed based on these criteria. Their results were used to predict how a piece of information would diffuse on the Twitter platform.

General Terms

Simulation, modeling

Keywords

Social network, information diffusion, twitter, Predictive model, trend

1. INTRODUCTION

Nowadays, people share or spread information using modern information and communication technologies. Because of the development in technology and proliferation in communication devices and gadgets, individuals are now active publishers and communicators of information rather than being just passive receivers of information, as was the case in the past. This active participation of individuals in information dissemination had boosted the online social network enabled by the internet. The online social network plays a massive role in diffusing information across the internet using the now available social network platforms such as Twitter, Facebook, LinkedIn, etc. The information diffusion process occurs when information spreads in a network from one node to another. Information diffusion has been a concerning public issue, especially in marketing and rumors of emergencies [18]. As earlier opined, users are no longer simple receivers; therefore, individuals' behavior plays a vital role in their social network evolution and propagation process. Individuals exchange information with their neighbors, which constructs their social relations. Much research has been done to explore the diffusion of

information.

Most research attention is directed at which segments influence information dissemination, what information diffuses all the more rapidly, and how information is scattered or diffused. Be that as it may, the diffusion of information in online social networks may be a complex and dynamic process [16]. Monitoring information diffusion is challenging but achievable by analyzing different information diffusion factors, monitoring information diffusion, and forecasting information diffusion. A decent presentation model is imperative to achieve information diffusion. The model shows how information diffusion can be anticipated and influenced, including rumor control, behavioral analysis, public opinion assessment, psychological phenomena study, and resource allocation in public health care systems [18]. Because of this threat that we encounter, it has become imperative to know how information diffusion occurs, to be able to counter such or contain it at least and possibly find means to use it as an advantage.

Online social networks play a significant role in the dissemination of information on an unprecedented scale. The extremely huge amounts of information with various details have hastened much research in information diffusion in online social networks. This work will allow people to understand the information diffusion process better, enabling them to improve their business performance (e.g., improving marketing campaigns), track events (e.g., tracking disruptive waves), and solve problems.

1.1 Theoretical Development

1.1.1 Information Diffusion

Information diffusion plays an essential role in numerous human interactions such as the spread of innovations, knowledge and information security management, social influence in marketing, and more. It is worthy to note that information diffusion cannot take place without communication. Communication implies transmitting messages using commonly perceived signs, images, and semiotic guidelines, starting with one individual or group onto another. There is always one sender, a message (Information to be transmitted), a channel, and a beneficiary or received in any communication process.

The advent of social networks has further increased information diffusion, making online social platforms prevalent in recent years. In addition to disseminating information, online social media users can audit or gather information. Whether genuine or false, news has since the beginning of social media life spread quicker than any time in recent years. Today, news feeds via web-based networking media pages and pop-up messages are sent straight to user devices such as phones, thereby allowing people to stay up

with the latest on what is going on around them.

1.1.2 Twitter

In the last two decades, online social media sites (such as Facebook, Twitter, YouTube) have revolutionized how we communicate and transform everyday practices. Among these Online Social Networks (OSNs), Twitter is an American online microblogging and social networking service. It is a free long-range microblogging website that enables enlisted individuals to communicate short posts called tweets [2]. Twitter Inc., the provider of Tweeter, is situated in San Francisco, California. It has over 25 workplaces worldwide and was made in March 2006 by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams and propelled in July of that year. The administration quickly increased overall ubiquity. In 2012, above 100 million users posted 340 million tweets every day, and the administration dealt with more than 1.6 billion inquiry questions each day. In 2013, it was one of the ten most-visited sites and had been depicted as "the SMS of the Internet". Starting in 2018, Twitter had over 321 million monthly dynamic users. Since 2015 Twitter has been a hotbed of discussions and news covering legislative issues of nations. During the 2016 U.S. presidential political race, Twitter was the biggest wellspring of breaking news on the day, with 40 million political race-related tweets sent by 10:00 p.m. (Eastern Time) that day. Twitter users can impart tweets and take after other customer's tweets by using various exercises. Twitter is, as of now, the third most famous long-range informal communication administration, with over 335 million dynamic users.

A name, expression, or topic referenced at a more noteworthy rate than others are known as a "trending topic" or basically "trend". Trending topics become well known either through a deliberate exertion by users (as in advancing a political race candidate) or because of an occasion that prompts individuals to discuss a particular point, for example, a TV arrangement or seismic tremor [2]. A rundown of the top twenty trending topics is recorded on the site, which helps Twitter and their users comprehend what is going on the planet and what individuals' assessments are about. Users can subscribe to or "follow" fascinating individuals, by which a coordinated social network is framed. Short messages or "tweets" that are restricted to containing close to 280 characters show up on followers' screen. Individuals can forward or "retweet" selected posts on the screen to their followers. They are likewise permitted to "mention" the screen name of another user in tweets by utilizing the "@" sign (for example, @Obama).

Moreover, users frequently mark their posts with subject names or "hashtags" unequivocal topical labels, words, or expressions following a hash image, for example, #oscar2014 [26]. Twitter is a micro-blogging system that enables many individuals to communicate short messages through social associations. Twitter trends have indicated their amazing capacity on numerous open occasions. For example, in the 2018 World Cup played in Russia and decisions, for instance, in the 2019 Presidential election of Nigeria and market changes. Besides, governments and businesses analyze and understand the population's general mood dynamics to reach better results

2. RELATED WORKS

In [10], a predictive model named Time -Based Asynchronous Cascade (T-BaSIC) was used in resolving information diffusion in online social networks. The model

developed a graph-based diffusion model that fully integrates the temporal dimension, and its parameters are time-dependent. It suits what is seen in an online social network. The problem with their system is that it works with the "closed world" assumption. It only accounts for the network influence, thereby ignoring the possibility of external influence and hence does not take into cognizance that nodes can be affected by external sources.

[18] explained the disparities between the predictive and explanatory models with a survey carried out on predictive and explanatory models, the most widely used basic models of information diffusion. Different methods from each of these basic models were reviewed, analyzed, discussed, and compared. It was observed that the two models were discovered to be independent; they always complement each other.

In [1], their study reviewed the current researches and applications relating to information diffusion in social networks and therefore provide essential knowledge for building new models, algorithms, and systems in the broad area of the social network. The paper compared the different models available for information diffusion in terms of techniques such as optimization algorithms, data mining, data streams covering many problems such as influence spread maximization, misinformation limitation, and study of trends in online social networks.

[23] proposed using an explanatory model for information diffusion, which involves testing an app (VKontakte) with the dynamic network analysis that focuses on the network's stability. The study's objective is to introduce a measure to combat the spreading of unwanted information while disseminating information. One of the system's limitations is that the "Vkontakte" app used in the modeling had many closed profiles of users, which deteriorated the dataset for modeling.

In [14], a predictive model developed using a hydrodynamic information diffusion prediction model (hydro-IDP) with Sina-Weibo social platform as a case study. The system's objective is to analyze the spreading process of information from both temporal and spatial perspectives. It is also helpful in extracting information diffusion characteristics, e.g., the informational popularity, the user influence, and the diffusivity of a social platform. A notable limitation of the system is that it could not model the cross-platform diffusion of the information in online social networks and analyze influential users' infection rates.

[9] proposed a Social Media Simulation (SMSim) simulator using Barack Obama's political campaign in microblogging Twitter network during the 2012 United States presidential race as a case study. SMSim uses a stochastic multi-agent-based approach where each agent represents a user in the Online Social Network. It aims at using a real online social network to simulate the user's behavior based on what they post on online social networks to analyze how the information is spread across a network. It was observed that the simulator outcome was not accurate as it was unable to find influential users and top engaged users to evaluate their impact on the information flow.

[16] implemented a system dynamics modeling to specify and quantify the core mechanisms of petition diffusion online. It then estimated their resulting dynamic model based on empirical data, which was then compared with the Bass diffusion model. One of the model's limitations is that the case

study was based on a small sample of data. Hence the study assumed that the total population was fixed over time.

[20] provide a formal definition of the Information Diffusion problem in the multi-social-network scenario. The study implements a graph-based model for a multi-social-network scenario (MSNS) using Answer Set Programming (ASP). LiveJournal, Flickr, Twitter, and YouTube as a case study. The model considers the existence of more social networks and the presence of bridges and topics of interest for MSNS users.

[19] proposed two new models, the Scale-free network Linear Influence Model (SLIM) and Exponential Network Linear Influence Model (ELIM) for diffusion in Scale-Free (SF) and Exponential networks. The models follow power-law and exponential degree distributions, respectively. It shows that the mathematical structures of the two diffusion models match the well-known Bass. Also circumventing SLIM and ELIM's fitting errors due to evolving node degree distribution, a multi-stage network linear influence model (MLIM) which adapts the degree distribution parameter in different stages of diffusion model when assuming the node degree distribution is static during diffusion was used. The paper aims to develop macro-level diffusion models that incorporate specific network topology characteristics. It was observed that without the MLIM model, the proposed models would not have been successful.

[6] proposed to model the actions of influence in a multidimensional social network (MSN) using DEVS formalism, which has permitted explicit modeling and simulation of human behavior and the interaction between individuals as a network. DEVS is a timed, highly modular, hierarchical formalism for the description of reactive systems. It can be appropriated to implement networks, propagation, and human behavior. It aims to provide a simple but efficient and accurate framework to model an individual's behavior and simulate the propagation of information among a group of individuals and its influence on their behavior.

3. RESEARCH METHOD

The first step was to carry out a literature review and analyze different journals and materials that discussed other information, diffusion models. This phase is essential because it is necessary to gather enough knowledge required for this work. The limitations of the examined materials formed the basis for this paper's approach to the model's design and implementation. Online social network diffusion has two primary models for information diffusion, namely: explanatory and predictive models. This paper proposes a predictive model-based design using Twitter as a case study. The trending topics of a month will be accessed from the Twitter API, which will then be analyzed into six criteria, namely location, lexical analysis, Trending Time, Tweet Volume, Promo/Giveaway, and Social Media Influencer. These criteria will help to analyze and explain how these topics came to be trending. The data and result generated will be used to train a model with the ability to predict how information diffuse on the Twitter platform to these criteria (using some as input and others as the output)

There are 463 Where On Earth Identifier (WOEID) locations based on trends available on Twitter as of today. Nigeria is one of them, and some regions within Nigeria, such as Lagos, Ibadan, Benin, Port-Harcourt, Kaduna, and Kano. Twitter API gives trending topics concerning these regions. Lagos was selected as the case study in this project as it is one of the

supported locations for a trending topic. Arguably, according to the World Population Review, Kano may have been selected because it still has a considerable population, but it's not as much as that of Lagos. The population density in Lagos is five times higher than that of Kano. But the landmass of Kano state is at least six times that of Lagos State.

3.1 Proposed System Architecture

The proposed system reads a text or tweet as input and predicts the text's diffusion rate or tweet based on other information given.

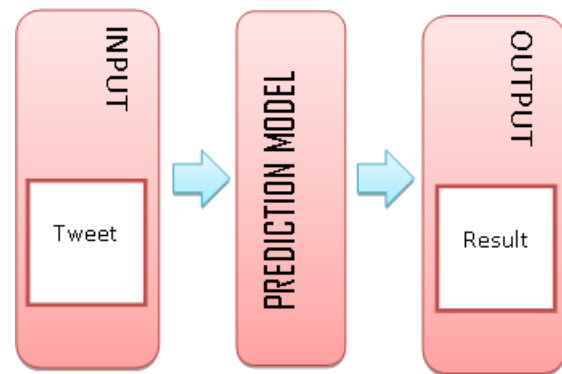


Figure 1: Prediction model overview

3.2 Dataset Analysis

The dataset was collected using Twitter's trends API (Application Programming Interface), which provides current trending topics for a given place in the format of WOEID. WOEID is a unique 32-bit reference identifier, initially identified by GeoPlanet and now assigned to Yahoo!, which identifies any attribute on earth. A WOEID is a locator that Yahoo! uses to identify unique locations worldwide. It is amazingly accurate and the easiest way to find your location. Contrary to the website trend list showing only 20 items, the result of the API consists of (mostly) 50 trending topics with their corresponding number of tweets (where available) using Lagos as a location to access the API for Lagos trending topics using a script for the whole month of September 2019. Also with 97.35% availability, the script accessed the API every hour. It should be noted that there could be some unknown states due to the script period, such as those trending topics that held the top rank for less than 10 minutes. The dataset comprises 12486 unique topics listed once or more in the top 50 trending topics.

The following algorithms were used to collect data from Twitter API (in Figure 2), merge the data (Figure 4), and analyze the collected data (Figure 5). We also present screenshots showing data before pre-processing and data after pre-processing in Figures 3 and 6, respectively.

```
Data = [ ]
```

```
Access twitter API # Using Client key and Client secret key
for each t ∈ T
```

```
if woeid == 2346342
```

```
trendholder = { }
```

```
trendholder.topic = t.topic # Get trending topics
```

```
trendholder.tweet_volume = t.tweet_volume # Get
```

```

        tweet volume
trendholder.hashtags = t.hashtags # Get trending
                                topics with hashtags
trendholder.influencer = t.influencer # Get topmost
                                influencer
trendholder.text_length = t.text_length # Get text
                                length of trending topics

data.push(t)
end if
end;

```

Figure 2: Algorithm for collecting data from twitter API

```

{
  "trends": [
    {
      "name": "#ChainedToTheRhythm",
      "url": "http://twitter.com/search?q=%23ChainedToTheRhythm",
      "promoted_content": null,
      "query": "%23ChainedToTheRhythm",
      "tweet_volume": 48857
    },
    {
      "name": "George Lopez",
      "url": "http://twitter.com/search?q=%22George+Lopez%22",
      "promoted_content": null,
      "query": "%22George+Lopez%22",
      "tweet_volume": 98590
    },
    {
      "name": "#FelizMiercoles",
      "url": "http://twitter.com/search?q=%23FelizMiercoles",
      "promoted_content": null,
      "query": "%23FelizMiercoles",
      "tweet_volume": 36103
    },
    {
      "name": "#wednesdaytsdom",
      "url": "http://twitter.com/search?q=%23wednesdaytsdom",
      "promoted_content": null,
      "query": "%23wednesdaytsdom",
      "tweet_volume": 36103
    },
    {
      "name": "#wednesdaytsdom",
      "url": "http://twitter.com/search?q=%23wednesdaytsdom",
      "promoted_content": null,
      "query": "%23wednesdaytsdom",
      "tweet_volume": 42916
    }
  ]
}

```

Figure 3: Snapshot of the data before pre-processing

```

import python libraries
files_in_directory = []
merge = []
for each file in files_in_directory
    if filename != none
        append file into merge
    remove file from files_in_directory

```

Figure 4: Algorithm for merging of data

```

import python libraries
import merge
remove duplicate file in merge
Sort merges by tweet volume
For each file ∈ merge

```

```

If file == '
    Remove file from merge
Check % of trends in merge
Check for giveaway in merge
Check for top influencers in merge
Return results

```

Figure 5: Algorithm for analyzing the data

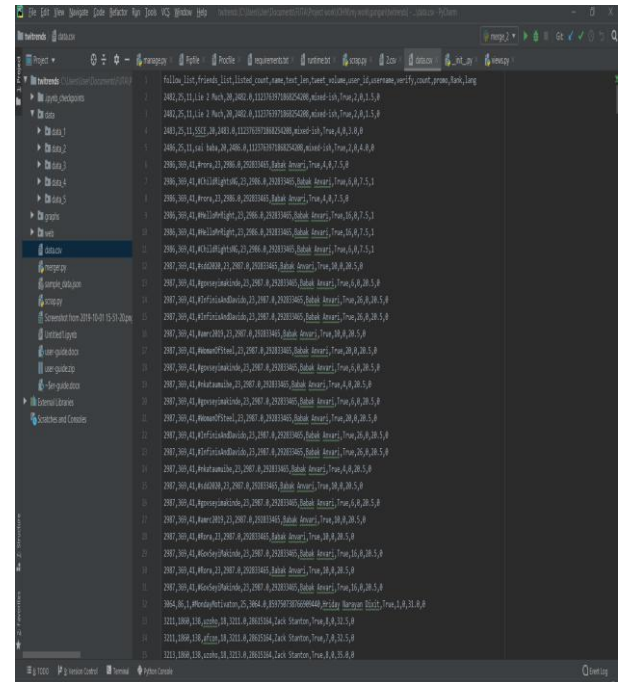


Figure 6: Snapshot of the data after pre-processing

3.3 Technological Tools

Various technological tools can be used for data preparation and analysis. They include Jupyter Notebook, Pandas, NumPy, Matplotlib, and Seaborn. In this paper, Jupyter Notebook was employed during the processing and analysis of the data. A snapshot of the Jupyter Notebook environment is given in Figure 7.

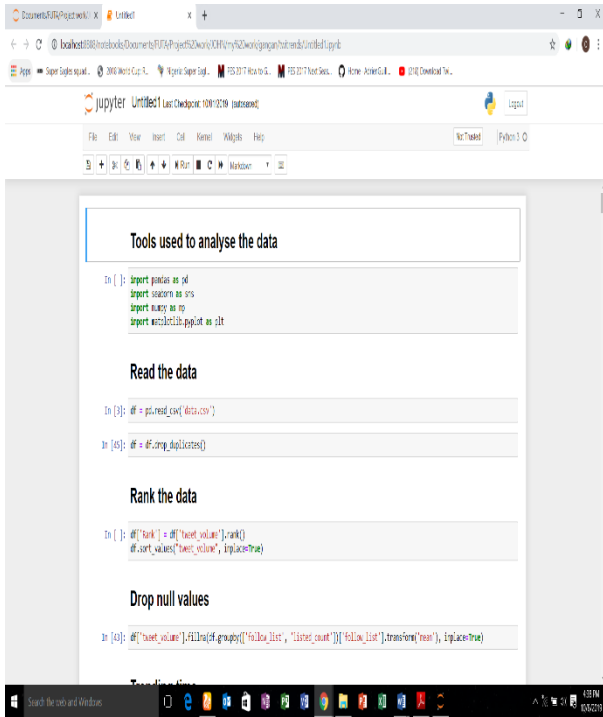


Figure 7: Snapshot of the Jupyter environment

3.4 Design of the Prediction Model

The model employs random forest, which is a predictor comprising of a group of randomized base regression trees $\{r_n(x, \theta_m, D_n), m \geq 1\}$, where $\theta_1, \theta_2, \dots$ are outputs of a randomizing variable θ . These random trees are consolidated to form the aggregated regression estimate

$$\bar{r}_n(X, D_n) = \mathbb{E}_\theta[r_n(X, \theta, D_n)]$$

where \mathbb{E}_θ denotes expectation with respect to the random parameter, conditionally on X and the data set D_n . The randomizing variable θ is used to determine how the successive cuts are performed when building the individual trees, such as selecting the coordinate to break up and position of the split. The variable θ is assumed to be independent of X and D_n the training sample. This excludes, in particular, any bootstrapping or resampling step in the training set. This also rules out any data-dependent strategy to build the trees, such as searching for optimal splits by optimizing some criterion on the actual observations. However, θ is to be based on a second sample, independent of but distributed as, D_n . Therefore, with these warnings in mind, it is assumed that each individual random tree is constructed in the following way. All nodes of the tree are associated with rectangular cells such that at each step of the construction of the tree, the collection of cells related to the leaves of the tree (i.e., external nodes) forms a partition of $[0,1]$. The root of the tree is $[0,1]$ itself. The following procedure is then repeated $\lceil \log_2 k_n \rceil$ times, where \log_2 is the base-2 logarithm, $\lceil \cdot \rceil$ the ceiling function and $k_n \geq 2$ a deterministic parameter, fixed beforehand by the user, and possibly depending on n .

- I. At each node, a coordinate of $X = (X^{(1)}, \dots, X^{(d)})$ is selected, with the j -th characteristic having a

likelihood $\text{pnj} \in (0,1)$ of being selected.

- II. At each node, as soon as the coordinate is chosen, the split is at the midpoint of the chosen side.

Each randomized tree $r_n(X, \theta)$ yields the normal overall Y_i for which the corresponding vectors X_i fall in the same cell of the random partition as X . In other words, letting $A_n(X, \theta)$ be the rectangular cell of the random partition containing X ,

$$r_n(X, \theta) = \frac{\sum_{i=1}^n Y_i 1_{[X_i \in A_n(X, \theta)]}}{\sum_{i=1}^n 1_{[X_i \in A_n(X, \theta)]}} 1_{\mathcal{E}_n(X, \theta)}$$

where the event $\mathcal{E}_n(X, \theta)$ is defined by

$$\mathcal{E}_n(X, \theta) = \left[\sum_{i=1}^n 1_{[X_i \in A_n(X, \theta)]} \neq 0 \right]$$

(Thus, by convention, the estimate is set to 0 on empty cells.) Taking finally expectation with respect to the parameter θ , the random forests regression estimate takes the form

$$\bar{r}_n(X) = \mathbb{E}_\theta[r_n(X, \theta)] = \mathbb{E}_\theta \left[\frac{\sum_{i=1}^n Y_i 1_{[X_i \in A_n(X, \theta)]}}{\sum_{i=1}^n 1_{[X_i \in A_n(X, \theta)]}} 1_{\mathcal{E}_n(X, \theta)} \right]$$

3.5 System Implementation

The prediction model design in this paper was run on a 2.0GHz Intel Core-I3 with 4giga-byte RAM and 465giga-byte HDD. The system also used the following Django rest framework, sklearn, langdetect, Django, and gunicorn.

4. RESULT AND DISCUSSION

The dataset which comprises 12,486 trending topics extracted from twitter API was analyzed using Pandas, NumPy, Matplotlib, and Seaborn to the six criteria. The result of the analysis is discussed below.

4.1 Location-based Analysis

Lagos is used as the location to extract the dataset from Twitter API and a total of 12,486 trending topics were generated for the dataset.

4.2 Lexical Analysis

The official language in Lagos is English, but in our dataset, which comprises 12,486 trending topics, 24.17% of it was in the English language. At the same time, the rest consists of pidgin English and other indigenous languages of Nigeria. Most of the trending topics did not have a hashtag (#) as only 39.6% of the whole trending topics had a hashtag. In terms of character length to tweet volume, "#showersOfBlessingByWhiteMan" is the highest with a 28-character length, "SSCE" is the lowest with a 4-character length and the average character length is 24. In terms of character length to trending time, "#breakthedevilshheartgiveaway" is the highest with a 28-character length, "Arla" is the lowest with a 4-character length and the average character length is 10.1. Figure 8 shows the distribution for a set of lowest character length of trending topics. The overall highest character length is 28. The lowest character length is 4, and 10.1 is the average overall character length.

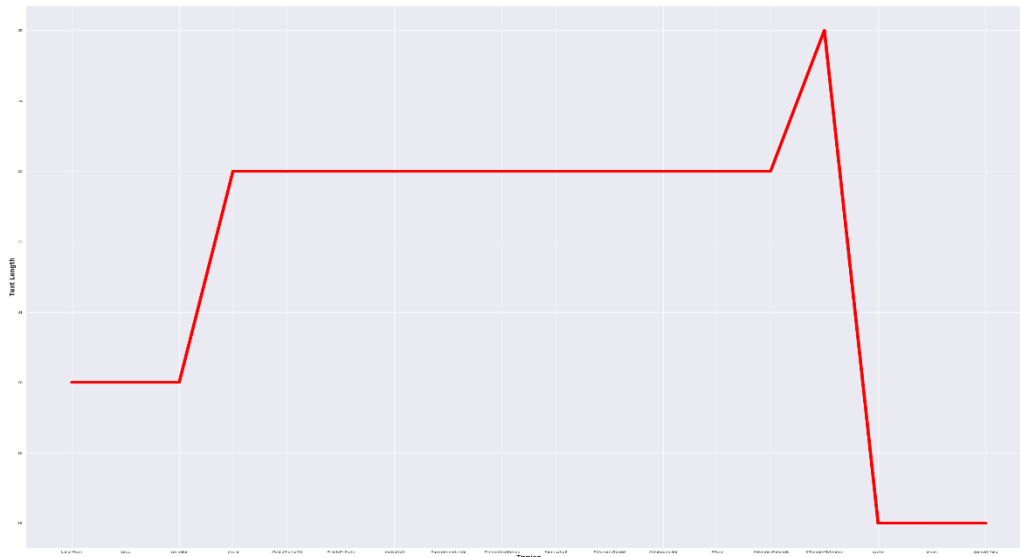


Figure 8: Overall smallest character length of trending topics



Figure 9: Lowest trending time of trending topics

4.3 Trending Time

Analysis based on trending time was carried out by counting the number of times a topic becomes trending as the Twitter API is accessed hourly. In terms of trending time to tweet volume, "Watford" is the highest with 105 counts for trending time, and "#MondayMotivation" is the lowest with 1 count for trending time. The average trending time has an 18.18 count. Figure 9 shows the distribution of the lowest trending time of trending topics.

4.4 Tweet Volume

The Tweet volume consists of the computation of tweet count, retweet, and likes generated by a trending topic. The dataset

"message title" has the highest tweet volume with 64,593,382 volumes, and the lowest tweet volume is "Lie 2 much," with 2,482 volumes. Typically, a trending topic will have an average tweet volume of 1871362.84. Figure 10 shows the lowest tweet volume of trending topics.

4.5 Promo/Giveaway

During September 2019, there were not that many giveaways; hence there are just a small fraction of topics having a relationship with promo/giveaway. The dataset consists of 12,486 trending topics, of which 138 trending topics are related to giveaways. The percentage of trending topics related to promo/giveaway in the dataset is 1.11%. Figure 11 shows the distribution for promo/giveaway of trending topics.

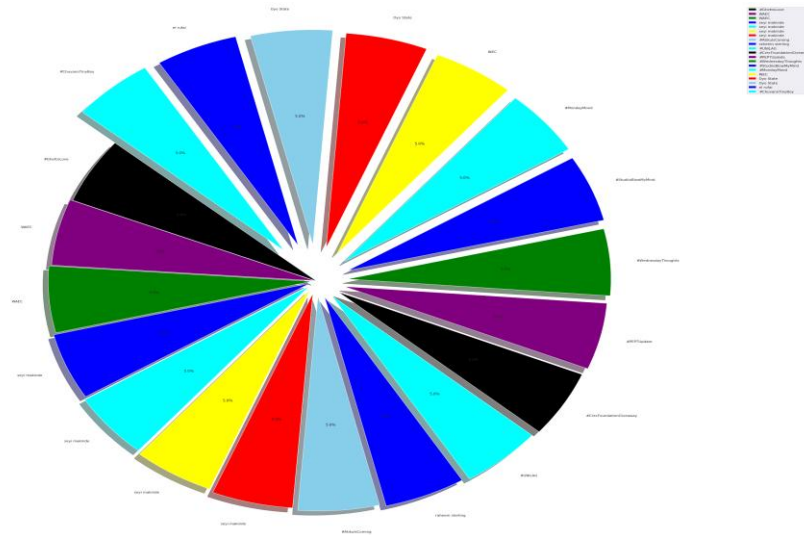


Figure 10: Lowest tweet volume of trending topics



Figure 11: Distribution for Promo/Giveaway of trending topics

4.6 Social Media Influencer

In the dataset, in terms of social media influencer to tweet volume, "Donald J. Trump" is the highest influencer in "message title" and "mixed-ish" is the lowest influencer in "Lie 2 Much". In terms of social media influencer to trending time, "AuronPlay" is the highest influencer in "Watford" and "Hriday Narayan Dixit" is the lowest influencer in "#MondayMotivation.". Figure 12 shows the distribution of social media influencers of trending topics

4.7 Testing of Prediction Model

The prediction model can be accessed online at <https://tweetprediction.herokuapp.com>. The model required inputs are the number of followers, following, and the tweet or statement that its diffusion process is to be predicted. The output consists of the inputted variables with their values alongside the tweet or statement's language, predicted tweet count, and predicted frequency (hours). Figure 13 shows the interface of the prediction model environment.

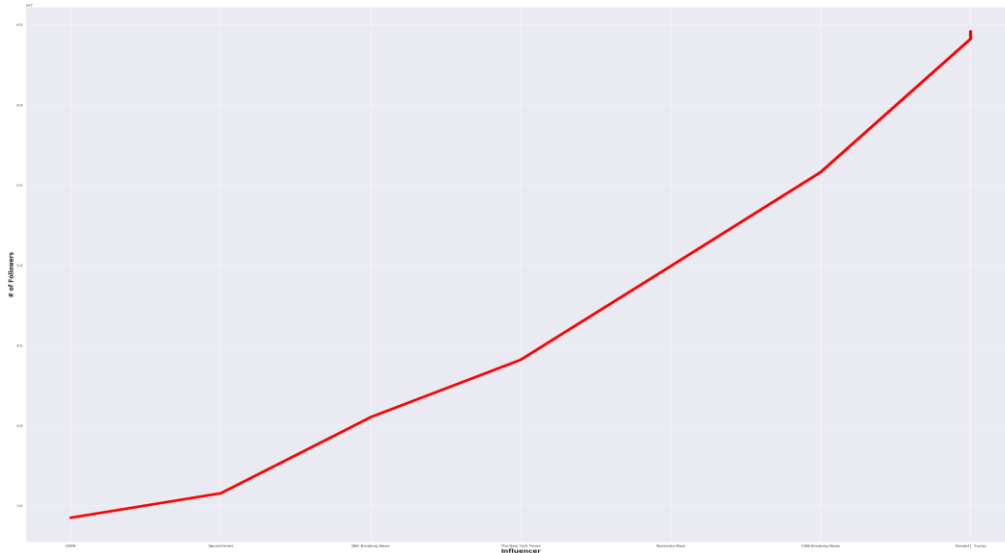


Figure 12: Snapshot of distribution for Social Media Influencer of trending topics

The interface shows a 'Tweet' field with a JSON output:

```
{ "output": "Welcome" }
```

. Below is an HTML form with input fields for 'followers(numbers)', 'following(numbers)', and 'Tweet', and a 'POST' button.

Figure 13: Snapshot of the Prediction model environment

The interface shows a 'Tweet' field with a detailed JSON prediction result:

```
{ "followers": "18600", "following": "56", "Tweet": "#SC19", "Language": "Local", "Predicted tweet count": "27832.726333333332", "Predicted frequency(hours)": "9.86" }
```

. Below is an HTML form with input fields for 'followers(numbers)' (18600), 'following(numbers)' (56), and 'Tweet' (#SC19), and a 'POST' button.

Figure 14: Snapshot of the Supercomputing (SC19) event prediction result

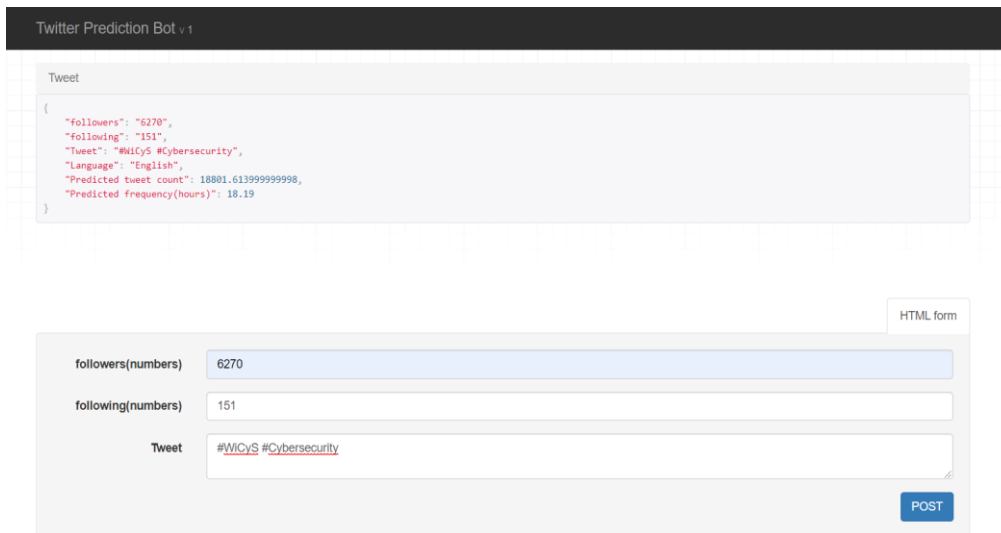


Figure 15: Snapshot of the Women in Cybersecurity(WiCyS) event prediction result

Figures 14 and 15 show a typical prediction by the prediction model on Supercomputing 2019 and Women in Cybersecurity events respectively.

5. CONCLUSION

A prediction model shows how information diffuse on the Twitter platform was developed using a dataset of trending topics of September 2019 to train it. The dataset consisted of 12,486 specific topics listed once or more in Twitter API's top 50 trending topics. The dataset was analyzed to six criteria that were later used in the prediction model to show how information diffuses on the Twitter platform as either an input or an output. The model implemented could be enhanced more effectively by feeding larger data to the model, and a more advanced learning curve, either deep learning or recommendation learning, could be used. Archived data belonging to parastatals could be made accessible, and a reduction in the cost of hosting the server will make the model more accessible.

6. REFERENCES

- [1] Agrawal, D., Budak, C., & Abbadi, A. El. (2011). Information diffusion in social networks: Observing and influencing societal interests. *Proceedings of the VLDB Endowment*, 4(12), 1512–1513.
- [2] Annamoradnejad, I., & Habibi, J. (2019). A Comprehensive Analysis of Twitter Trending Topics. 22–27.
- [3] Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. *WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web*, 519–528.
- [4] Behera, P. C. (2016). Data Mining Technique for Tracking of Information Diffusion in Online Social Network. *V(Iv)*, 102–105.
- [5] Biau, G. (2010). Analysis of a Random Forests Model Random forests. 1–40.
- [6] Bouanan, Y., Forestier, M., Ribault, J., Zacharewicz, G., Vallespir, B., & Moalla, N. (2015). Simulating information diffusion in a multidimensional social network using the DEVS formalism (WIP). *Simulation Series*, 47(8), 63–68.
- [7] Breiman, L. (2001). Random Forest Draft. 1–33.
- [8] Buskens, V., & Yamaguchi, K. (1999). A new model for information diffusion in heterogeneous social networks. *Sociological Methodology*, 29(1), 281–325.
- [9] De C Gatti, M. A., Appel, A. P., Dos Santos, C. N., Pinhanez, C. S., Cavalin, P. R., & Neto, S. B. (2013). A simulation-based approach to analyze the information diffusion in Microblogging Online Social Network. *Proceedings of the 2013 Winter Simulation Conference - Simulation: Making Decisions in a Complex World, WSC 2013, (Jennings 2001)*, 1685–1696.
- [10] Guille, A. (2013). *Information Diffusion in Online Social Networks*. New York: ACM.
- [11] Guille, A., Favre, C., Hacid, H., & Zighed, D. (2013). SONDY: An open source platform for social dynamics mining and analysis. *Proceedings of the ACM SIGMOD International Conference on Management of Data, (May 2014)*, 1005–1008.
- [12] Henry, D., Stattner, E., & Collard, M. (2017). Social media, diffusion under influence of parameters: Survey and perspectives. *Procedia Computer Science*, 109, 376–383.
- [13] Hu, Y., Aiello, M., & Hu, C. (2018). Information diffusion in online social networks: A compilation. *Journal of Computational Science*, 28, 204–205.
- [14] Hu, Y., Song, R. J., & Chen, M. (2017). Modeling for Information Diffusion in Online Social Networks via Hydrodynamics. *IEEE Access*, 5.
- [15] Iribarren, J. L. (2011). Information Diffusion Epidemics in Social Networks. *SSRN Electronic Journal*, 1–12.
- [16] Jalali, M. S., Ashouri, A., Herrera-Restrepo, O., & Zhang, H. (2016). Information diffusion through social networks: The case of an online petition. *Expert Systems with Applications*, 44, 187–197.
- [17] Kalogeratos, A., Scaman, K., Corinzia, L., & Vayatis, N. (2018). Information Diffusion and Rumor Spreading. *Cooperative and Graph Signal Processing*, 651–678.
- [18] Li, M., Wang, X., Gao, K., & Zhang, S. (2017). A survey on information diffusion in online social networks: Models and methods. *Information (Switzerland)*, 8(4).
- [19] Luu, M. D., Lim, E. P., Hoang, T. A., & Chua, F. C. T.

- (2012). Modeling diffusion in social networks using network properties. ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media, 218–225.
- [20] Marra, G., Nocera, A., Ricca, F., Terracina, G., & Ursino, D. (2014). Investigating information diffusion in a multi-social-network scenario via answer set programming. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8741, 191–196.
- [21] Social network analysis and Information Diffusion.
- [22] Myers, S. A., Zhu, C., Leskovec, J., & Bakshy, E. (2011). Information Diffusion and Social Influence in Online Networks. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 33–41.
- [23] Viksnin, I., Iurtaeva, L., Tursukov, N., & Muradov, A. (n.d.). The Model of Information Diffusion in Social Networking Service.
- [24] Wei, Z., Yanqing, Y., Hanlin, T., Qiwei, D., & Taowei, L. (2013). Improved SI Model for Information Dissemination. *Proceedings of the 2012 International Conference of MCSA*, 145–150.
- [25] Wellman, M. P. (n.d.). *Modeling Information Diffusion in Heterogeneous Information Networks*
- [26] Weng, L. (2014). Information diffusion on online social networks. *ProQuest Dissertations and Theses*, (April), 223.
- [27] Weskida, M., & Michalski, R. (2019). Finding influentials in social networks using evolutionary algorithm. *Journal of Computational Science*, 31, 77–85.