Full Paper

AN INVESTIGATIVE PROCESS MODEL FOR PREDICTING INFORMATION DIFFUSION ON SOCIAL MEDIA: INFORMATION SYSTEM PERSPECTIVE

I. P. Gambo
Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria
ipgambo@oauife.edu.ng

A. P. Adjicheboutou
Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria

R. N. Ikono
Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria

O. G. Iroju
Department of Computer Science, Adeyemi College of Education, Ondo, Ondo State, Nigeria

S. T. Yange
Department of Mathematics/Statistics/Computer Science, Federal University of Agriculture, Makurdi, Nigeria

ABSTRACT

The challenge of the information diffusion (ID) process across social media is the inability to know the provenance or source of information and the details of the person sending the information. This situation creates confusion within the organization of people, which hinders smooth communication. In this paper, a classification and a mathematical model for predicting ID on social media (SM) were formulated. This was with a view to know the source of tweeted information for clarity and decision-making. The formulated model was simulated and evaluated to ascertain its adequacy. Data were collected from Twitter by using the Twitter Search API. In particular, 3,200 tweets were extracted on related political matters in Nigeria. The results showed that the classification is a multiclass case. The macro-average ROC curve from it showed 0.93. The ROC from curve one to seven of each class is 0.90, 0.95, 0.98, 0.97, 0.90, 0.97 and 0.85, respectively. The results of the simulation also indicated that there were seven different types of source devices for the information, namely Android, BlackBerry, iPad, iPhone, rack, studio, and web, according to the dataset. Most of the information came from iPhones, with about one thousand, nine hundred and eighty-four (1,984) tweets out of the 3,200 (that is 62%); the studio was the source least frequently with seven tweets (that is, 0.22%). In terms of the rate of diffusion, information gets most diffused through iPhones (76.20% of the retweets), and least diffused with Android phones (0.56% of the retweet). The model evaluation shows an accuracy of 95.87%, a specificity of 97.64% and a sensitivity of 83.46%, making it suitable for prediction. The research not only unveiled the identity of devices used for sending information on SM but also revealed the rate at which information gets diffused with the devices to aid decision-making.

Keywords: Information diffusion, social media, Twitter, support vector machines, social networks.

1. INTRODUCTION

With the growing popularity of sophisticated and advanced technologies like Web 2.0, the world has become a global community (Berthon *et al.*, 2012; Kaplan and Haenlein, 2010) where communication is enhanced and made easy. Besides, the advent of information and communication technology (ICT) engendered the adoption and use of online social networks and media to facilitate smooth communication in various domains like politics, marketing, education, and research (Hoang and Mothe, 2018). Communication, in this context, also involves interaction where information can be shared across the board. However, lousy communication or information can lead to a wrong decision, thereby affecting individual and organizational Information Systems (IS) (Courtney, 2001; Walsham and Sahay, 2006).

Several channels of communication like face-to-face meetings, telephone calls, writing of letters and memos, and email messages exist in an organization. These channels of communication present some limitations that can be outlined as forgetting a message, missing notes taken during a meeting, and the frustration of searching for information in an enormous list of email messages. The limitation mostly encountered during communication is time wastage. In the course of this study, it was observed that social media could alleviate these limitations by creating a platform for information sharing and discussion (Kietzmann *et al.*, 2011; Lee and Ma., 2012; Agnihotri *et al.*, 2016), thereby making information diffusion (ID) a high possibility with ease.

The information gets diffused quickly through social media use in facilitating communication (Chang, 2010; Valenzuela, 2013; Hoang and Mothe, 2018). ID, in this

context, entails using possible channels for propagating information without physical contact. The source of information in the context of this study is the device where the information originates. On the other hand, information is anything that is communicated and meaningful to the receiver (Bates, 2006). The disseminated contents may be viewpoints, rumors, knowledge, product marketing information etc. (Zhang *et al*., 2016). The channel is the online social network and medium (Guille and Hacid, 2012; Guille *et al*., 2013; Anduiza *et al*., 2014), which organizations and individuals often use to spread information to countless users (Davoudi and Chatterjee, 2015).

While the social media (SM) signify the channel, medium or platform that permit the creation and conversation of user-generated contents (Icha and Agwu, 2015; Edosomwan *et al*., 2011), the social networks (SN) represent the platforms or sites that are utilized to ease the construction and the maintenance of social connections between individuals of same and diverse races and countries. Both SM and SN give individuals and organizations opportunities to share their interests, activities, or positive relations (Eke *et al*., 2014). Noticeably, the information can reach the ends of the earth within a short time, thanks to the SN's frequent user-engagement. Besides, the utilization of SM has expanded channels of communication and their helpfulness to the organizations.

It is noteworthy that information – such as text messages, images, videos diffused on SM – are increasing and need to be predicted for future purposes, decision-making support, and knowing their spread of propagation as well as their provenance or source (Hoang and Mothe, 2018; Taxidou and Fischer, 2013; Oh *et al*., 2013). For instance, Twitter produces over 400 million messages per day (Taxidou *et al*., 2015). Remarkably, Allcott and Gentzkow (2017) extracted 115 pro-Trump and 41 pro-Clinton fake stories on Facebook (as an example of SM) with a total of 30 million and 7.6 million views for Trump and Clinton, respectively. It is evident then that information shared across the SM and SN (whether fake or genuine) definitely gets to everyone connected.

The diffusion process on the SM generally involves the sender(s), receiver(s), and medium (Behera, 2016), while the diffusion through SN can be affected by the network structures, behavioural characteristics of network members, and the characteristics of the information being diffused (Bampo *et al*., 2008). Also, some information spreads faster compared to others, and that is dependent on individuals' interests. This paper opines that since much information passes across SM, prediction becomes crucial to uncover the sources (either the device used, location, or identity of the person) (Najar *et al*., 2012), understand the information in circulation, and find a way to manage this information better (Hoang and Mothe, 2018). From the IS perspective, doing those as mentioned earlier can facilitate good decision-making in service delivery and product development. Besides, it can unveil the identities of those spreading either fake or genuine information and can even increase competitiveness in business.

This paper, therefore, develops a model that can predict information diffused over SM and know the device source from which information is diffused to aid organizational decision-making. This is with the view to know the source of information tweeted with devices for clarity and decision-making. The paper focused mainly on the diffusion of text (image, video, and audio were not taken into account) by knowing its source on related political matters (that is, the means used to send the message). Twitter is the SM chosen for this study. The source in this context is limited to the devices used to tweet or to retweet. Also, the political domain in the Nigeria case was used for the investigative process and predictions.

The remaining part of this paper is organized as follows: In Section 2, existing works of literature were reviewed, and gaps were identified. Section 3 discusses the research methodology by explaining the approaches used, including data extraction and analysis, the mathematical formulation of the model, and the simulation and evaluation procedures. Section 4 contains the results and discussion; and finally, in section 5, the conclusions and future work are discussed.

2. RELATED WORKS

Diffusion of information has been a very active research domain, especially after the advent of online social networks. It is an active field of research that supports the complexity of social interactions over social media (Chen *et al*., 2013). There are many methods applied to predict ID on social media (Kafeza *et al*., 2014). Some researchers, for example, Conover *et al*. (2011), made use of linear support vector machines (SVM) with latent semantic analyzes (LSA) to predict the political alignment of Twitter users. Wang *et al*. (2012) applied the Diffusive Logistic (DL) equation to predict information diffusion in time-based and spatial dimensions. Still, the aspect of knowing the source of information is vital (Hoang and Mothe, 2018) for clarity and decision-making.

Many of the authors in the literature (for example, Yang and Leskovec, 2010; Romero *et al*., 2011; Guille and Hacid, 2012; Taxidou and Fischer, 2013; Guille *et al*., 2013; Stieglitz and Dang-Xuan, 2013; Kafeza *et al*., 2014; Barberá *et al*., 2015; Allcott and Gentzkow, 2017; Hoang and Mothe, 2018; Alkhodair et al., 2020) made use of Twitter as the social media (SM) for extracting their data, and to make prediction, investigative analysis, and validation. The widespread use of Twitter is connected to the microblogging services it offers (Stieglitz and Dang-Xuan, 2013), which has made it to get much consideration (Sakaki *et al*., 2010). Milstein et al. (2008) showed 330 million monthly active users, 134 million daily active users around the globe use Twitter to remain socially associated with companions, relatives, and collaborators using their PCs, mobile phones, or other means of connection, with 140 million daily tweets while 460,000 open new accounts daily. Again, the general use of Twitter is due to its reputation, which has ultimately made it a standard social laboratory for understanding information dissemination and impact (Cha *et al*., 2010; Kwak *et al*., 2010), most notably as a communication channel for political institutions (Tumasjan et al., 2011). Most importantly, Kursuncu *et al*. (2019) observed that Twitter gives multimodal information containing content, pictures, and videos, alongside relevant and social metadata, such as temporal and spatial data, and data about client connectivity connections.

In Yang and Counts (2010), descriptions on reasons why some features of tweets predict more magnificent

information propagated than others were observed. However, in contrast with other studies, it was found that the practice of mentioning to another user in a tweet via the *@username* convention could also be analyzed. The method of mentioning @username has a different influence on retweet prediction than just following the tweet as it indicates active user interaction. Yang and Counts (2010) research aimed to investigate how topics spread through the network structures, not a prediction on the source of information. Besides, Yang and Counts (2010) focused on building predictive models to provide support for: (i) speed, regardless of whether and when the primary diffusion instance will occur; (ii) scale, which is the quantity of influenced cases at the main degree; and (iii) range, to determine how far the diffusion chain can proceed inside and in-depth.

Kwak *et al*. (2010), investigated the relationship between the quantity in terms of adherents to authors' tweets and retweets were investigated. Accordingly, Kwak *et al*. (2010) considered more specifically, the number of retweets made by clients who were not immediate devotees of the initial information diffused or tweeted. Strikingly, they contended that clients with under 1000 supporters would, in general, have a similar usual number of retweets. Following the outcome of Kwak *et al*. (2010), Cha *et al.* (2010) specifically found that clients with a high volume of tweets and retweets are typically persuasive when it comes to new information being tweeted. However, we realised that some of the clients were not powerful enough to retweet. Cha et al.'s (2010) work borders mainly on the relationship between influential users and the different topics they tweet while neglecting the source of the tweeted and retweeted information for clarity.

Sakaki *et al.* (2010) proposed a method for predicting an earthquake's location using machine-learning algorithms. Sakaki et al. (2010) accounted that each Twitter user was considered a social sensor, thereby providing help in identifying a natural disaster location. Their algorithm was based on processing Twitter real-time data, which gave a high probability of 96% for detecting earthquakes. However, based on identifying fundamental characteristics of approaching disaster, it becomes inevitable to predict the catastrophe and take new measures.

The work in Conover *et al*. (2011) predicted the political alignment of Twitter users by applying linear SVMs. The emphasis of the authors was on content-based classifications and latent semantic examination. They considered the substance of clients' tweets to distinguish hidden structure in the information emphatically connected with the political alliance. Likewise, network clustering algorithms were used to extract data about the people with whom every client expresses and demonstrates that these topological properties can be utilized to improve classification accuracy even further.

Again, Najar *et al*. (2012) predicted ID over SNs with partial knowledge. In particular, the authors used the classical independent cascade models (ICM) and the linear threshold model (LTM) methods. Consequently, Najar *et al*. (2012) generated the artificial cascades over real SNs. They predicted the last spread values without necessarily having to model the whole process of diffusion at every step, as is the case with other models. Besides, the authors' prediction approach figured out how to anticipate the final spread of states from information produced by various models, which had an advantage over other models in situations where the data about the structure of the network ends up temperamental. However, Guille and Hacid (2012) introduced the Bayesian logistic regression approach to obtain a good prediction of ID in a dynamic situation. Remarkably, Wang *et al*. (2012) approached predicting ID in online SM by formulating a diffusive logistic (DL) equation, which could characterise the processes of information spread. When experimented on the Digg dataset, the prediction results gave an accuracy of over 92%. As observed by Wang *et al.* (2013), the linear diffusive model also achieved high accuracy and precision with the Digg data set.

Furthermore, Zhu *et al*. (2013) made predictions on the activity level of users on SNs. The authors had models personalized for each user by using an extension of the consistent part of the models. Subsequently, Zhu *et al*. (2013) introduced the time-delay strategy for addressing out-of-date data. Noticeably, Zhu et al.'s (2013) strategy was unified into the logistic regression classifier to achieve user activity level prediction. Most importantly, Taxidou and Fischer (2013) work on real-time analysis of ID by first subscribing to relevant hashtags as the data source. The goal was to provide an infrastructure for establishing real-time analysis in massive datasets containing structural information. Even though the research is promising, it is still at its early stage and requires an in-depth investigation.

Yang and Wang (2016) used a deep belief network method to classify images that were spread over the SM. The deep learning approach showed the potential to achieve excellent performance when large datasets are available to use. In Hu *et al.* (2017), the procedure for the spread of information over SM was introduced and used in a hydrodynamic model. The model portrays the procedure from temporal and spatial points of view, which helped collect quality information being diffused. For instance, the popularity of the information, its impact, and how the information diffused on the SM platform is part of the qualities assessed. Also, Hu *et al.* (2018) used the recurrent diffusion-LSTM network architecture to model how image contents get propagated through the SN. However, the source of information being diffused was not considered.

Liu et al. (2018) used the Random Recursive Tree (RRT) approach to build topologies for cascade tree to enable the capturing of essential structures as information is diffused. Besides, the authors' approach portrayed the behaviour of users and clarified the size of the cascade in WeChat. Still, the source of information being diffused was not addressed.

3. METHODS

This study was addressed through the experiment-based quantitative research approach (Chang and Ren, 2000; Arghode, 2012) involving simulation. A Twitter application programming interface (API) was used to harvest the dataset, and a Support Vector Machine (SVM) to classify the dataset. The Stochastic Differential Equation (SDE) was used to formulate a mathematical model for predicting ID. The experiment was simulated using the

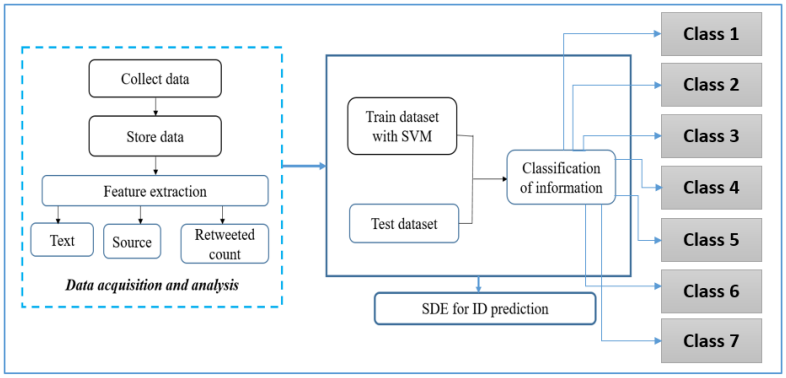

Figure 1: Conceptual View of the Developed Model

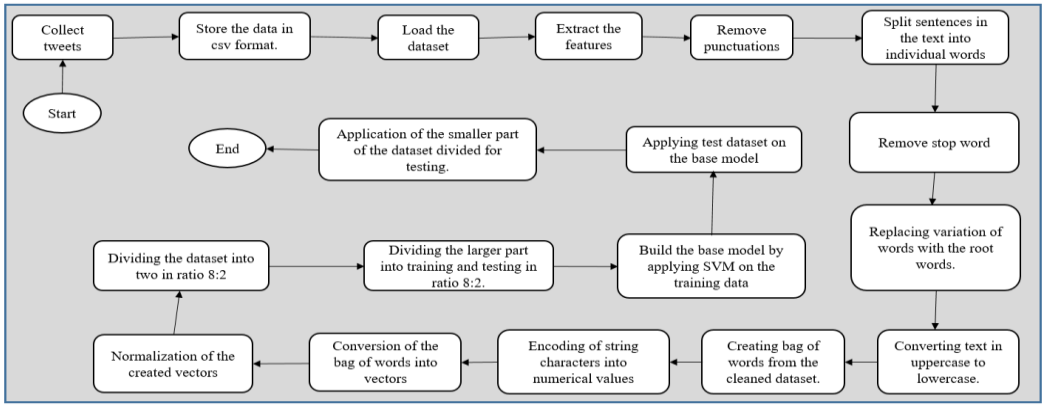

Figure 2: The Flow of Processes Involved in the Model Formulation

python programming language. The adequacy of the model performance was evaluated using accuracy, specificity, and efficiency as parameters. Figure 1 describes the conceptual flow of our approach. Moreover, Figure 2 shows the detailed flow of processes involved in the model formulation.

In Figure 1, the first activity was on data acquisition, and analysis where data (in this case, the tweets) are: (i) extracted from Twitter API; (ii) stored in a suitable format for pre-processing; and (iii) made to extract relevant features in the dataset (in this case, the text, source, and retweeted counts). Following the data acquisition and analysis process is the prediction of ID. Specifically, the extracted features of the tweets are subjected to a machine-learning algorithm. In this paper, the SVM was used for classification purposes (i.e., train the dataset). SVM is mostly useful for classification and regression problems (Flake and Lawrence, 2002; Frohlich and Zell, 2005). With the kernel trick technique used by SVM (Suykens et al., 2003; Lodha et al., 2006), it can detect an optimal boundary between the output and can perform complex data transformation (Amari and Wu, 1999). SVM is known for the classification of extant data and predictions of unknown data classes (Byun and Lee, 2002; Coussement and Van den Poel, 2008). It is conveniently suitable for both small and high dimensional data (Jonsson *et al.*, 2002), and efficient in handling both linearly and non-linearly separable datasets (Amari and Wu, 1999). Labelling took place after "dividing the larger part into training and testing in ratio 8:2" as indicated in Figure 2. The SVM classifier experimental setup, as shown in Figure 3 (Amarappa and Sathyanarayana, 2014), was used to classify the tweets.

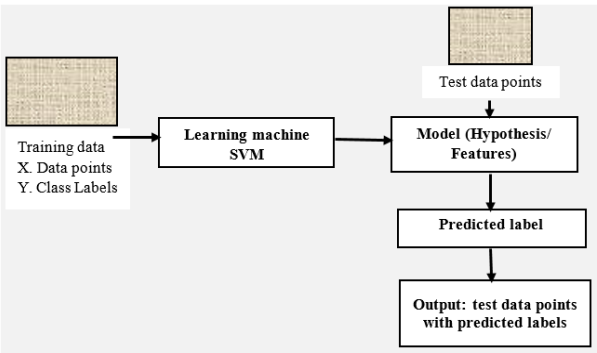

Figure 3: A Block Diagram of the SVM Classifier (adopted from Amarappa and Sathyanarayana, 2014)

3.1. Data Extraction and Analysis

Data were extracted from Twitter using the Twitter Search API by writing python code[1]. 3,200 tweets were extracted on related political matters in Nigeria. The initial plan was to combine data from Facebook and Twitter; however, access to Facebook was not granted based on privacy concerns. Figure 4 shows aspects of the data

[1] shorturl.at/jsE12

extracted from the Twitter API. The complete list of data extracted is found here[2]. The extracted data has eight columns: id, created_at, source, favorite_count, retweet_count, language, place, and text. The tweeted text sent by users, the source, and the retweeted count were used for the prediction.

After extracting the tweets, we pre-processed them by converting the tweets into a suitable scale format with all the significant features extracted. Pre-processing data, in this context, involves the transformation of raw data into a comprehensible format (Marquardt et al., 2004). In this case, the CSV format was used. Next, the data (tweets) were cleansed, and punctuations were removed before tokenization using the Natural Language Toolkit (NLTK). The goal of data cleaning is to get data in a clean, standard format for further analysis. The punctuations were removed from each word, the data was tokenized and remaining tokens that were not alphabetic were removed, and stop words filtered out. Figure 5 shows the removed punctuations, and Figure 6 shows the tweets tokenized. The complete files are as shown here (removed punctuation[3], tokenized tweets[4]).

3.2. Developed Mathematical Model

As stated earlier, the Stochastic Differential Equation (SDE) was used to formulate a mathematical model for predicting ID. SDE is suitable because it reflects reality in the case of the problem domain and class classification (Juhl *et al.*, 2016). The mathematical formulation has an exact and concise mathematical interpretation of the complex random propagation process. The SDE provides a statistical notion for presenting predictive results in appropriate formats (Cobb, 1981). Therefore, this research proposed a novel point process-based ID model, as shown in equation 1, and reformulated as an SDE in equation 2. The formulation was based on the theorem that considered and modeled a network consisting of *N* users and the information content $x_i(t)$ for node *i*, as observed in Wang *et al.* (2016) and Wang *et al.* (2018). Thus:

$$x_i(t) = b_i + \sum_{j=1}^{U} \alpha_{ij} K_\omega(t) * h(x_j(t)) dN_j(t)) = b_i + \sum_j \alpha_{ij} \sum_{t_j \in H_j(t)} K_\omega(t - t_j) h(x_j(t)) \quad (1)$$

As described in Wang *et al.* (2016), the base content is denoted represented as b_i; the influence weight from the given node *j* to *i* is represented as α_{ij}. Also, the specific function for the problem defined is represented as *h*; the point process used to model the events as they occur at random intervals relative to either time or space axis (Ripley and Kelly, 1977; Van Lieshout, 2000) is represented as $N_j(t)$. Here, the point process was categorized as temporal point processes consisting of discrete events localized in time. The history of the event is given as $H_j(t)$ up to time *t* for the user *j*. Moreover, the application dependent function is given as $h(x_j(t))$, while the exponential that triggers the kernel is defined as $K_\omega(t) = exp(-\omega t)$. Finally, a convolution operator is represented as *, as shown in Equation 1. To this end, Equation 2 shows the SDE form.

$$dx_i(t) = \omega(b_i - x_i(t))dt + \sum_j \alpha_{ij} \; h(x_j(t))dN_j(t)) \quad (2)$$

Proof: The convolution operator * of any two functions $f(t)$ and $g(t)$ is defined as:

$$f(t) * g(t) = \int_0^t f(t-s)g(s)ds$$

Next, the differential operator d was applied to $x_i(t)$, using the following properties:

i. $dk_\omega(t) = -\omega k_\omega(t)dt$ *for* $t \geq 0$ *and* $k_\omega(0) = 1$.
ii. The differential of the convolution of two functions is: $d(f * g) = f(0)g + g * df$.

By setting $f = k_\omega(t)$, and $g = \sum_j a_{ij} h(x_j) dN_j(t)$, the differential of $x_i(t)$ was taken. Moreover, the two properties mentioned above were used to get:

$$dx_i(t) = d(f * g) = \sum_{j=1}^{U} a_{ij} h(x_j) dN_j(t) - \omega \left(\sum_{j=1}^{U} a_{ij} k_\omega(t) * (h(x_j).\, dN_j(t)) \right) dt$$

$$= \sum_{j=1}^{U} a_{ij} h(x_j) dN_j(t) - \omega(x_i(t) - b_i)dt$$

$$= \omega\,(b_i - x_i(t))dt + \sum_{j=1}^{U} a_{ij} h(x_j(t) dN_j(t))$$

3.3. Simulation and Model Evaluation

The model was simulated in Jupyter Notebook IDE from Anaconda with the Python programming language. Also, the metrics used for the model evaluation include: (i) accuracy, which determines the over-all correctness of the classifier after prediction as shown in equation 3; (ii) the sensitivity, which defines the proportion of actual positive cases correctly classified; and (iii) the specificity, which defines the proportion of actual class negatively predicted. Equations 3, 4 and 5 describe the accuracy, sensitivity and specificity used for evaluating the performance of the model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (4)$$

$$Specificity = \frac{TN}{TN+FP} \quad (5)$$

From equations 3, equations 4, and equations 5, TP is the number of true positives; FN is the number of false negatives; FP is the number of false positives; TN is the number of true negatives.

[2] shorturl.at/bdxBH

[3] shorturl.at/CKYZ8

[4] shorturl.at/hlxJS

```
id,created_at,source,favorite_count,retweet_count,lang,place,text

1140690430394294273,2019-06-17 18:39:26,b'Twitter for iPhone',638,162,en,,b'In recent times we have established new Forward Operating
Bases in vulnerable communities. We\xe2\x80\x99re now set to launch\xe2\x80\xa6 https://t.co/bJdGBAKejV'

1140689671896403969,2019-06-17 18:36:25,b'Twitter for iPhone',691,181,en,,b'Condolences also to the families of the victims of the
killings and destruction of property in Sokoto and Zamfara S\xe2\x80\xa6 https://t.co/zvNm9rneWX'

1140688039104176129,2019-06-17 18:29:56,b'Twitter for iPhone',650,162,en,,"b'We will ensure that the perpetrators are found and brought to
justice, and we will also continue to give our law en\xe2\x80\xa6 https://t.co/s3wG57PtyB'"

1140687686036086787,2019-06-17 18:28:31,b'Twitter for iPhone',5242,1082,en,,"b'I commiserate with the families of the victims of
yesterday\xe2\x80\x99s Viewing Center bomb blasts in Mandarari, Konduga LGA\xe2\x80\xa6 https://t.co/yqQoZtIvEt'"

1139924383080833026,2019-06-15 15:55:26,b'Twitter for iPhone',719,175,en,,"b'We\xe2\x80\x98re not unmindful of business environment
challenges: It still takes too long for goods to clear at our seaports,\xe2\x80\xa6 https://t.co/ptGcvl8plh'"

1139921705445580800,2019-06-15 15:44:47,b'Twitter for iPhone',1060,244,en,,b'The Next Level is for small businesses and rural economies.
We will ensure greater access to microcredit to rural f\xe2\x80\xa6 https://t.co/vEpTXuXNkq'

1139918780316966912,2019-06-15 15:33:10,b'Twitter for iPhone',4824,1040,en,,b'We can be proud of our history since Independence: our
contributions to UN peace-keeping globally; our stabilizatio\xe2\x80\xa6 https://t.co/M6ynywFLBt'

1138833215161065473,2019-06-12 15:39:31,b'Twitter for iPhone',19341,3319,en,,"b'I\xe2\x80\x99m pleased I was able to watch the Super
Falcons match against South Korea, after the Democracy Day Lunch. Congrat\xe2\x80\xa6 https://t.co/seY65q7uCK'"

1138564792078221312,2019-06-11 21:52:54,b'Twitter for iPhone',1405,355,en,,"b'I urge all contestants who lost out to be gallant in defeat,
and to join hands with the victors. And to the victors\xe2\x80\xa6 https://t.co/wnoSQYdqdU'"

1138564368826818560,2019-06-11 21:51:13,b'Twitter for iPhone',1315,338,en,,"b'Stepping into the Next Level, the legislature has a big role
to play for the goals of our administration to be achi\xe2\x80\xa6 https://t.co/2nQNUiT1PZ'"

1138564047614500865,2019-06-11 21:49:57,b'Twitter for iPhone',750,289,en,,b'Let me make it clear that the Executive does not desire a
```

Figure 4: Extracted Data from Twitter API

File Edit Format View Help

```
['idcreatedatsourcefavoritecountretweetcountlangplacetext', '1140690430394294273207190617', '183926bTwitter', 'for', 'iPhone638162enbIn', 'recent', 'times', 'we', 'have', 'established', 'ne
'the', 'victims', 'of', 'yesterdayxe2x80x99s', 'Viewing', 'Center', 'bomb', 'blasts', 'in', 'Mandarari', 'Konduga', 'LGAxe2x80xa6', 'httpstcoyqQoZtIvEt', '113992438308083302620190615', '15
ter', 'for', 'iPhone193413319enbIxe2x80x99m', 'pleased', 'I', 'was', 'able', 'to', 'watch', 'the', 'Super', 'Falcons', 'match', 'against', 'South', 'Korea', 'after', 'the', 'Democracy', 'D
le', 'separation', 'of', 'powersxe2x80xa6', 'httpstcoK2a78DDS6r', '113856374788951654420190611', '214845bTwitter', 'for', 'iPhone624208enbWe', 'all', 'including', 'todayxe2x80x99s', 'winne
stcoojDQxftWcu', '113553287900867379220190603', '130510bTwitter', 'for', 'iPhone1322429enbI', 'thank', 'the', 'members', 'of', 'the', 'SARS', 'Reform', 'Panel', 'for', 'their', 'work', 'an
'our', 'Police', 'Officers', 'as', 'they', 'daily', 'put', 'thexe2x80xa6', 'httpstcomD3b4WqYKJ', '113553039075047424020190603', '125516bTwitter', 'for', 'iPhone802366enbWhere', 'the', 'rig
', 'of', 'the', 'Presidential', 'Panel', 'on', 'the', 'Reform', 'of', 'the', 'Special', 'AntiRobbery', 'Squad', 'SARS', 'We', 'coxe2x80xa6', 'httpstcoqIv6g9SRxK', '113198335966257971320190
412808790016201905O7', '172328bTwitter', 'for', 'iPhone0650enbRT', 'ProfOsinbajo', 'Each', 'one', 'of', 'the', 'piers', 'is', 'the', 'equivalent', 'of', 'an', '18', 'storey', 'building', '
ike', 'Qatarxe2x80x99s', 'support', 'as', 'it', 'is', 'not', 'a', 'project', 'thaxe2x80xa6', 'httpstcoiOyoEmZJj1', '112074104444908749020190423', '172742bTwitter', 'for', 'iPhone94681977en
on', 'reason', 'and', 'good', 'judgement', 'wexe2x80x98re', 'giving', 'violence', 'the', 'chance', 'to', 'take', 'over', 'and', 'make', 'a', 'bad', 'situation', 'mxe2x80xa6', 'httpstco86Ei
stcoVD5minopz5', '111735782102312960120190414', '092358bTwitter', 'for', 'iPhone1584355enbIt', 'gladdens', 'our', 'hearts', 'to', 'see', 'the', 'progress', 'being', 'made', 'by', 'the', 'y
l', 'bring', 'an', 'end', 'to', 'the', 'banditry', 'and', 'insecurity', 'We', 'are', 'axe2x80xa6', 'httpstcoIJqxLNyCes', '111590388566999449820190410', '090633bTwitter', 'for', 'iPhone1347
2551202enbLet', 'me', 'again', 'offer', 'our', 'deepest', 'condolences', 'to', 'all', 'the', 'victims', 'and', 'their', 'families', 'and', 'loved', 'ones', 'We', 'feel', 'your', 'painsxe2x
ttpstcoaHWVePDUxE', '111481769221567283220190407', '091024bTwitter', 'for', 'iPhone1703814enbWe', 'are', 'fully', 'determined', 'to', 'tackle', 'this', 'challenge', 'ferociously', 'until',
, 'of', 'the', 'World', 'Economic', 'Forum', 'on', 'the', 'Middle', 'East', 'and', 'North', 'Africxe2x80xa6', '111341307824954982420190403', '120858bTwitter', 'for', 'iPhone2359779enbWhile
ident', 'Deepest', 'Condolences', 'to', 'the', 'family', 'and', 'friends', 'of', 'Kolade', 'Johnson', 'victim', 'of', 'the', 'March', '31', 'fatal', 'shooting', 'in', 'Lagos', 'PoliceNxe2x
RT', 'ProfOsinbajo', 'Happy', 'Birthday', 'to', 'an', 'iconnnYou', 'are', 'a', 'visionary', 'who', 'redefined', 'governance', 'by', 'establishing', 'a', 'cabinet', 'on', 'merit', 'and', 'c
72487621459968320190325', '133201bTwitter', 'for', 'iPhone497180enbI', 'would', 'also', 'like', 'to', 'thank', 'the', 'Minister', 'of', 'Budget', 'amp', 'Nat', 'Planning', 'for', 'coordinat
tation', 'of', 'the', 'National', 'Mxe2x80xa6', 'httpstcoTw09uS2dDt', '111013122660052582520190325', '104804bTwitter', 'for', 'iPhone2589478enbWe', 'celebrate', 'the', 'purposefulness', 'a
', 'Committee', 'will', 'be', 'to', 'produce', 'a', 'comprehensive', 'document', 'that', 'will', 'guide', 'the', 'incoming', 'Cxe2x80xa6', 'httpstcoNWpY6ioCHQ', '11080387936538296322019031
74848bTwitter', 'for', 'iPhone2725633enbI', 'charge', 'the', 'Lagos', 'State', 'Government', 'to', 'do', 'all', 'that', 'is', 'needful', 'so', 'that', 'such', 'tragic', 'developments', 'do
0xa6', 'httpstcooXXU1C19XQ', '110476344211056640020190310', '151824bTwitter', 'for', 'iPhone137663720enbI', 'am', 'profoundly', 'saddened', 'by', 'the', 'news', 'of', 'the', 'crash', 'invo
, 'Women', 'in', 'the', 'MBuhari', 'Administration', '', 'httpstco60OOJprSOM', 'IWD2019', 'Baxe2x80xa6', '110332785111226777620190306', '161353bTwitter', 'Media', 'Studio59212184enbMy', 'M
'political', 'differences', 'I', 'still', 'hold', 'him', 'in', 'the', 'highest', 'esteem', 'because', 'his', 'contributions', 'to', 'the', 'dexe2x80xa6', 'httpstco7y3fCEahf2', '11030400528
'194141bTwitter', 'for', 'iPhone115322263enbI', 'congratulate', 'President', 'Macky', 'Sall', 'on', 'his', 'reelection', 'for', 'a', 'second', 'term', 'There', 'is', 'no', 'doubt', 'at', '
04624862003420190227', '164120bTwitter', 'for', 'iPhone47581118enbIxe2x80x99d', 'like', 'to', 'pay', 'tribute', 'to', 'those', 'who', 'lost', 'their', 'lives', 'either', 'through', 'accide
8020387840420190227', '071312bTwitter', 'for', 'iPhone30501196enbThe', 'new', 'Administration', 'will', 'intensify', 'its', 'efforts', 'in', 'Security', 'Restructuring', 'the', 'Economy', '
handling', 'the', 'difficult', 'taxe2x80xa6', 'httpstcobSwM7Q1BvY', '110065469151029657620190227', '071142bTwitter', 'for', 'iPhone2112824enbI', 'am', 'very', 'sad', 'at', 'the', 'grievous
ed', 'to', 'the', 'logistical', 'expenses', 'of', 'the', 'campaign', 'in', 'such', 'a', 'vast', 'country', 'like', 'Nigeria', '110065442989481984020190227', '071039bTwitter', 'for', 'iPhon
and', 'foremost', 'I', 'give', 'thanks', 'to', 'Almighty', 'God', 'Who', 'has', 'spared', 'our', 'lives', 'to', 'witness', 'another', 'victory', 'for', 'Nigerianxe2x80xa6', 'httpstcoscQqul
lic', 'of', 'Nigeria', '109883445061869158420190222', '063843bTwitter', 'for', 'iPhone4049980enbAs', 'citizens', 'there', 'is', 'no', 'greater', 'duty', 'than', 'this', 'and', 'no', 'great
'entrench', 'and', 'strengthen', 'our', 'democracy', '109883398741657190420190222', '063652bTwitter', 'for', 'iPhone1121417enbAs', 'we', 'thank', 'domestic', 'monitors', 'and', 'observers'
ve', 'worked', 'diligently', 'to', 'ensure', 'that', 'adexe2x80xa6', 'httpstco4VkJOk3P5Y', '109883370047586304020190222', '063544bTwitter', 'for', 'iPhone1074401enbAll', 'who', 'are', 'abl
he', 'exercise', 'of', 'your', 'rights', 'as', 'citizens', 'and', 'voters', 'tomorrow', '109883336201648537620190222', '063423bTwitter', 'for', 'iPhone1009376enbIt', 'is', 'only', 'upon',
', 'profound', 'and', 'weighty', 'duty', 'that', 'rests', 'upon', 'it', 'We', 'must', 'cast', 'aside', 'doubt', 'and', 'have', 'faith', 'txe2x80xa6', 'httpstcoZhzRdr9tAi', '109883276080079
827220190222', '063056bTwitter', 'for', 'iPhone1308467enbIt', 'requires', 'a', 'combination', 'of', 'patience', 'tolerance', 'compassion', 'diligence', 'wisdom', 'and', 'hope', 'These', 't
general', 'elections', '109856216297967616020190221', '123644bTwitter', 'for', 'iPhone03634enbRT', 'ProfOsinbajo', 'MisleadingFAKE', 'news', 'is', 'on', 'the', 'rise', 'especially', 'in',
'disorder', 'and', 'remain', 'peaceful', 'patriotic', 'and', 'unitedxe2x80xa6', 'httpstcot1rbnFolR0', '109671634514685952120190216', '102207bTwitter', 'for', 'iPhone1952874enbWhile', 'I',
C', 'themselves', 'have', 'given', 'assurances', 'day', 'after', 'day', 'and', 'almost', 'hour', 'after', 'hour', 'that', 'they', 'are', 'in', 'complete', 'readinessxe2x80xa6', 'httpstcog5
', 'rather', 'ruthless', 'I', 'did', 'things', 'a', 'certain', 'way', 'Now', 'as', 'a', 'democratically', 'electxe2x80xa6', 'httpstcoWilMF4OalY', '109603244159944294420190214', '130432bTwi
er', 'Media', 'Studio85363769enbFellow', 'Nigerians', 'this', 'is', 'our', 'story', 'Join', 'us', 'as', 'we', 'take', 'Nigeria', 'to', 'the', 'Next', 'Level', 'httpstco07ASFmUlGW', '109472
Y', '109463016150767206420190210', '161222bTwitter', 'for', 'iPhone44911765enbHow', 'Corruption', 'threatens', 'Nigeria', '', 'We', 'have', 'secured', 'high', 'profile', 'convictions', 'bu
n', 'they', 'had', 'the', 'chance', 'to', 'solve', 'thxe2x80xa6', 'httpstcoBaC0re9K9b', '109036481246544281720190129', '214323bTwitter', 'for', 'iPhone592205enbI', 'was', 'told', 'that', '
e', 'can', 'only', 'stand', 'a', 'chance', 'to', 'win', 'the', 'fight', 'against', 'Corruption', 'and', 'position', 'our', 'dear', 'nation', 'forxe2x80xa6', 'httpstco78p4XBogzG', '10888568
2554316820190125', '174951bTwitter', 'for', 'iPhone729433enbIt', 'also', 'explains', 'why', 'I', 'am', 'not', 'only', 'complying', 'immediately', 'but', 'with', 'some', 'degree', 'of', 're
'has', 'spoken', 'on', 'any', 'matter', 'severalxe2x80xa6', 'httpstcodr1dZghwHu', '108885608532121190420190125', '174815bTwitter', 'for', 'iPhone580386enb', 'and', 'after', 'quite', 'a', '
```

Figure 5: Removed punctuations from tweets

Figure 6: Tweets tokenized/ Word tokenization

4. RESULTS AND DISCUSSIONS

The classification results shown in Figure 7 described how the dataset was classified. Figure 8 illustrates the devices used for tweeting and re-tweeting. We discovered that seven classes of devices were used as sources of connection to Twitter from the dataset. As shown in Figure 8, the iPhone was the device that people on Twitter mostly used. Thus, the rate of diffusion of tweets coming from the iPhone was higher than others. Figure 9 contains the source code used for determining the proportion of the source of connection on Twitter.

Table 1 shows the count of the different tweets emanating from the same medium source on Twitter, and Table 2 shows the rate of diffusion of tweets. We have shown the rate of diffusion in Figure 10, as contained in Table 2.

The confusion matrix shown in Figure 11 contains information about the actual and predicted classifications used as a measure of the model performance. The confusion matrix visualizes the tasks performed in the classification. As Figure 11 reflects, the predicted classes are represented in a row, while the actual class predicted are described in the column. In the case of classification with more than two classes, the confusion matrix contains values of results for each class as it is in this research. Thus, all correctly predicted values are described diagonally in the confusion matrix, so it is easy to visually interpret errors in the prediction, which were obtained from values outside the diagonal. The classification was a multiclass case, as shown in Figure 11.

Figure 12 shows some of the calculations for determining the accuracy, sensitivity, and specificity based on equations 3 to 5. Table 3 shows the performance evaluation results for each class, and Table 4 shows the overall results of the evaluation for all the classes.

Going further, we have shown the ROC curves to aid the visualization of how the classifier performed. On the ROC curve, the TP rate is positioned on the *y*-axis, while the FP rate is positioned on the *x*-axis for every possible classification threshold. The goal of the ROC graph is to benchmark the result obtained from the classifiers. Figures 13, 14, 15, 16, 17, 18, and 19 show the ROC curve for each class. The ROC for the overall classes is shown in Figure 20.

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	0
2	0.00	0.00	0.00	0
3	0.00	0.00	0.00	0
4	1.00	0.86	0.93	481
5	0.21	1.00	0.34	5
6	0.79	0.74	0.76	155
7	0.00	0.00	0.00	0
micro avg	0.83	0.83	0.83	641
macro avg	0.25	0.33	0.25	641
weighted avg	0.94	0.83	0.88	641

Figure 7: Classification of Information

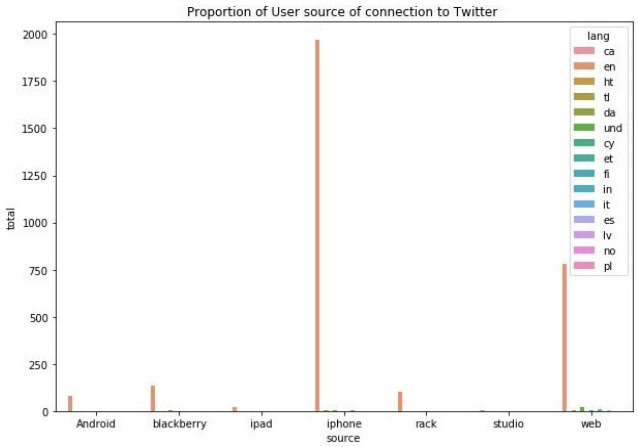

Figure 8: The Source of Connection on Twitter Proportion

Table 1: Retweeted Count of Tweet from the Different Source

Media Source of Tweet	Retweeted Count of the Tweet	Rank
iPhone	1016085	(1)
Web	203370	(2)
Blackberry	14204	(4)
Rack	67125	(3)
Android	7418	(7)
IPad	12517	(6)
Studio	12688	(5)
Total	1333407	

Table 2: The Rate of Diffusion of the Tweet

Media Source of Tweet	Rate of Diffusion of tweets
iPhone	76.20%
Web	15.25%
Rack	5.03%
Blackberry	1.07%
Studio	0.95%
IPad	0.94%
Android	0.56%

```
import itertools
import numpy as np
import pandas as pd
import seaborn as sn
import matplotlib.pyplot as plt
from time import time
from collections import OrderedDict
from sklearn.preprocessing import StandardScaler
import networkx as nx
from matplotlib.lines import Line2D
from datetime import datetime

df=pd.read_csv('twitter_data.csv')
df['source'] = df['source'].map({
   "b'Twitter for iPhone'":'iphone',
   "b'Twitter Web Client'": 'web',
   "b'Twitter Media Studio'": 'studio',
  "b'WriteRack'": 'rack',
   "b'Twitter for iPad'": 'ipad',
   "b'Twitter for BlackBerry'": 'blackberry',
   "b'Twitter for Android Tablets'": 'Android',
   })
dfa=df['source'].groupby(df['source'])
df['source'].value_counts().reset_index()
twts_politics = df[["source", "id", "lang"]].groupby(["source", "lang"]).agg(len) \
   .rename(columns={"id": "total"}).reset_index()
plt.figure(figsize=(10,7))
plt.title("Proportion of User source of connection to Twitter ")
ax = sn.barplot(x="source", y="total", hue="lang", data=twts_politics)
```

Figure 9: Python Source Code for the Proportion of User Source of Connection to Twitter

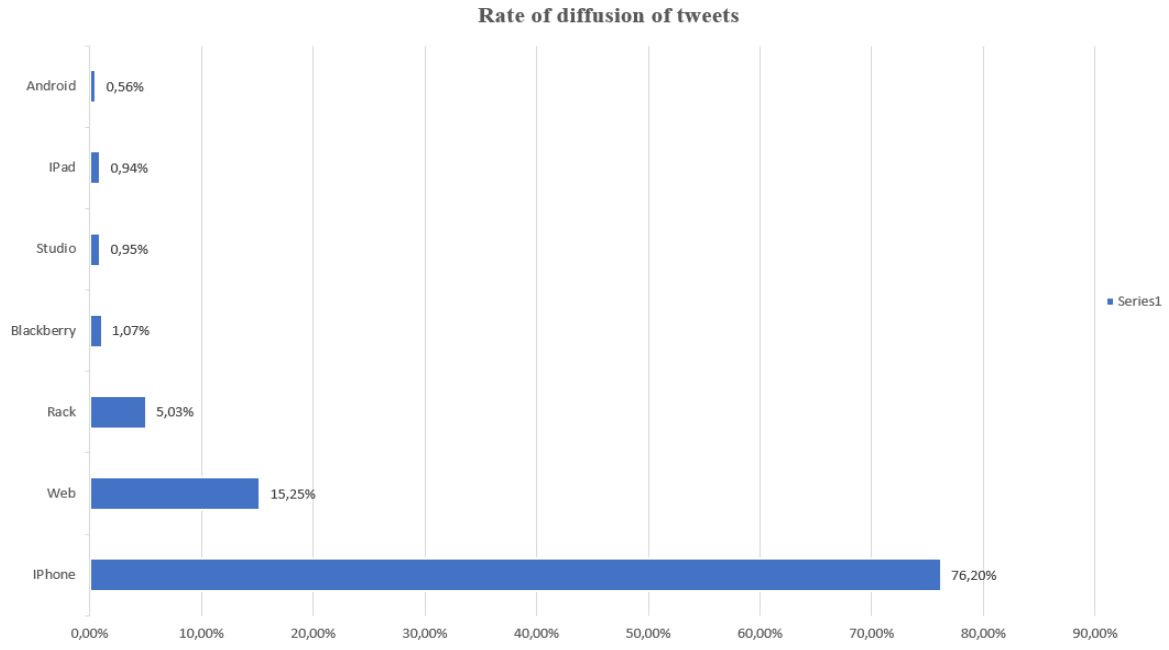

Figure 10: The Rate of Diffusion of Tweets

Table 3: Results of Performance Evaluation for Each Class: Accuracy and Specificity

Class	TP	TN	FP	FN	TP+TN	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$	$Spe = \frac{TN}{FP+TN}$
A	0	633	0	8	633	$ACC_A = \frac{633}{641} = 98.75\%$	$Spe_A = \frac{633}{0+633} = 100\%$
B	117	460	45	27	577	$ACC_B = \frac{577}{649} = 88.91\%$	$Spe_B = \frac{460}{460+45} = 91.09\%$
C	0	625	0	16	625	$ACC_C = \frac{0+625}{0+625+0+16} = 97.50\%$	$Spe_C = \frac{625}{0+625} = 100\%$
D	416	163	61	1	579	$ACC_D = \frac{416+163}{416+163+61+1} = 90.33\%$	$Spe_D = \frac{163}{61+163} = 72.77\%$
E	2	617	0	22	619	$ACC_E = \frac{619}{641} = 96.57\%$	$Spe_E = \frac{617}{617+0} = 100\%$
F	0	615	0	26	615	$ACC_F = \frac{615}{641} = 95.94\%$	$Spe_F = \frac{615}{0+615} = 100\%$
G	0	637	0	4	637	$ACC_G = \frac{0+637}{641} = 99.38\%$	$Spe_G = \frac{615}{0+615} = 100\%$
H	0	639	0	2	639	$ACC_H = \frac{0+639}{641} = 99.69\%$	$Spe_H = \frac{639}{0+639} = 100\%$

$TP_{\_all}$=535; $FP_{\_all}$=106; $FN_{\_all}$=106; $TN_{\_all}$=4389

$Total_{\_all} = TP_{\_all}+FP_{\_all}+FN_{\_all}+TN_{\_all} = 5136$

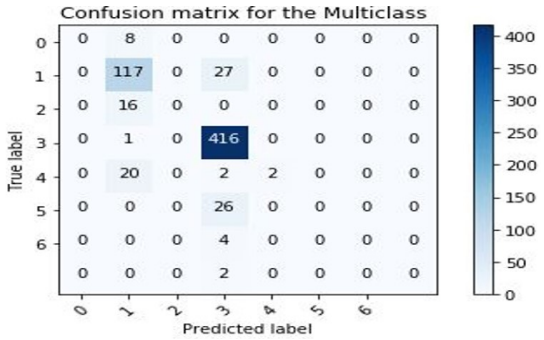

Figure 11: Confusion Matrix of the Multiclass Classification

	A	B	C	D	E	F	G	H	FN	TN
A	**0**	8	0	0	0	0	0	0	8	633
B	0	**117**	0	27	0	0	0	0	27	460
C	0	16	**0**	0	0	0	0	0	16	625
D	0	1	0	**416**	0	0	0	0	1	163
E	0	20	0	2	**2**	0	0	0	22	617
F	0	0	0	26	0	**0**	0	0	26	615
G	0	0	0	4	0	0	**0**	0	4	637
H	0	0	0	2	0	0	0	**0**	2	639
FP	0	45	0	61	0	0	0	0	106	
TP	0	117	0	416	2	0	0	0		

Figure 12: Data for Calculating the Accuracy, Sensitivity, and Specificity Based on Equations 3 to 5

Table 4: Overall results of performance Evaluation: Accuracy Sensitivity Specificity

SN	Metrics	Evaluation Results
1	Accuracy	$Acc_{\_all} = \frac{TP+TN}{Total\_all} = \frac{535+4389}{5136} = 95.87\%$
2	Sensitivity	$Sen\_all = \frac{TP}{FN+TP} = \frac{535}{641} = 83.46\%$
3	Specificity	$Spe_{\_all} = \frac{TN}{FP+TN} = \frac{4389}{4495} = 97.64\%$

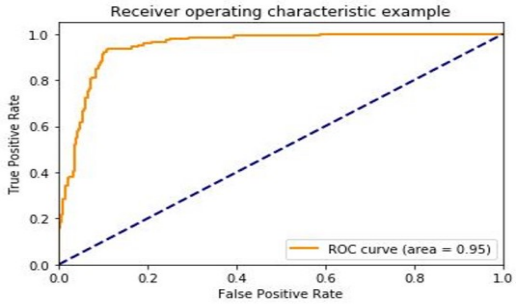

Figure 13: The ROC Curve of the First Class

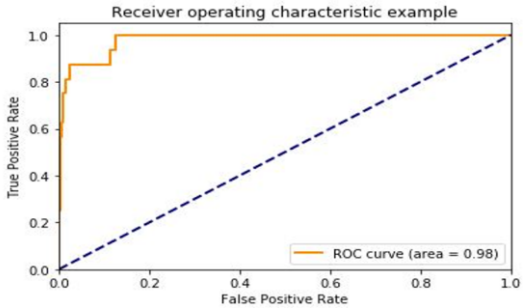

Figure 14: The ROC Curve of the Second Class

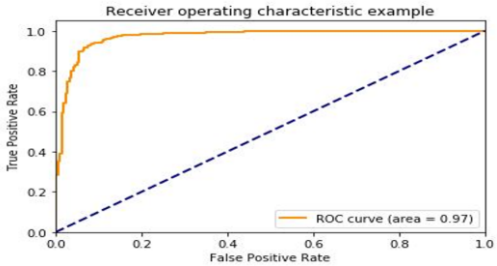

Figure 15: The ROC Curve of the Third Class

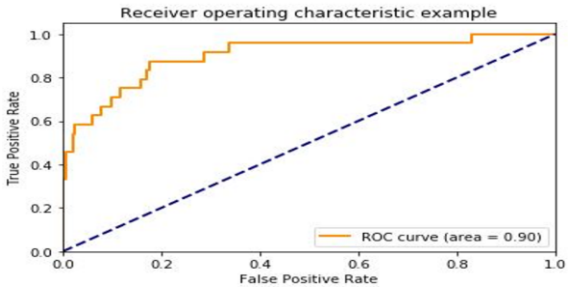

Figure 16: The ROC Curve of the Fourth Class

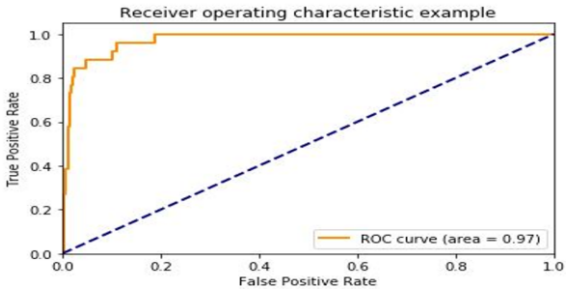

Figure 17: The ROC Curve of the Fifth Class

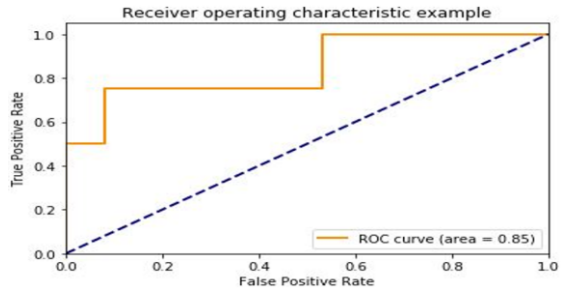

Figure 18: The ROC Curve of the Sixth Class

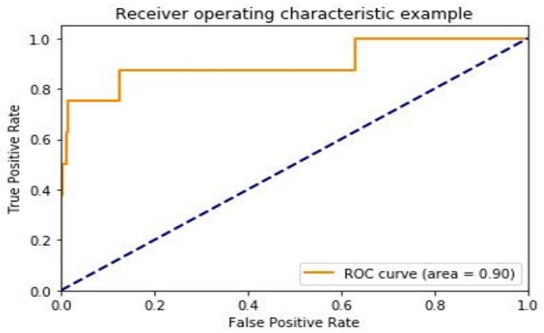

Figure 19: The ROC Curve of the Seventh Class

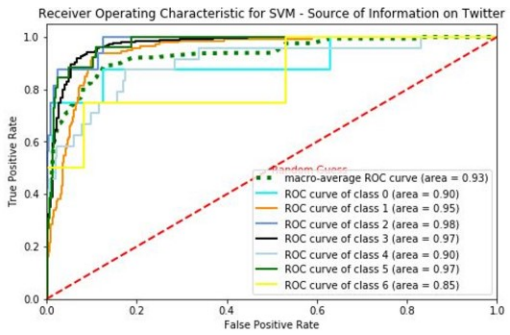

Figure 20: The ROC-SVM of the Source of Information on Twitter for Overall Classes

The macro-average ROC curve from the overall multiclass is 0.93. The ROC curve of each class from one to seven is 0.90, 0.95, 0.98, 0.97, 0.90, 0.97 and 0.85, respectively. As reflected in the dataset, the simulation results indicate that there were seven different sources of information; namely: Android, BlackBerry, iPad, iPhone, rack, studio, and web. However, the iPhone was seen to have the highest number of retweeted count of tweets over the social medium with 1,016,085 tweets. In comparison, Android was the lowest count of 7,418 tweets, as shown in Table 1, Table 2, and Figure 10. In terms of the rate of diffusion, information gets most diffused through iPhone (76.20% of the retweets) and least diffused with Android (0.56% of the retweet). The model has an accuracy of 95.87%, a specificity of 97.64%, and a sensitivity of 83.46%, indicating that the model's performance is good and suitable for prediction.

Therefore, the research has unveiled the possibility of determining the sources of information tweeted with devices, thereby providing clarity and aiding decision-making. The inference of this is that there is a very high purchase and use of iPhones compare with other devices. Besides, this is a good way of saying that manufacturers of these devices can have a rethink based on our findings on how to boost competitiveness, judging from the ways and manner their products are demanded and used. Moreover, our results indicate that in the world of politics, most politicians and those with a keen interest in political matters use iPhone, regardless of the price compare to other devices.

5. CONCLUSION

The research reported in this paper focuses on a predictive model for ID and the devices use to send information. An SDE was used to formulate the mathematical model for ID. The emphasis was on the diffusion of text (image, video, and audio are not taken into account) by knowing the type of device (as the source) used in sending information on related political matters (that is, the means used to send the message). Knowing the source of information in ID provides support for addressing the complexity of social interactions over SM (Chen *et al.*, 2013; Peters *et al.*, 2013; Jiang *et al.*, 2014). Besides, it is essential for clarity and decision-making. We extracted the dataset from Twitter, pre-processed the dataset, did feature extraction on the dataset before applying it to a machine-learning algorithm for classification and prediction. The prediction was made based on the present information diffused to determine the rate of diffusion.

Thus, this paper's contribution borders on the provision of a model that can efficiently unveil the identity of devices used for sending information on SM. The model provided can predict the ID process accurately and support decision-making for organizations producing these devices. As reflected in this research, the scope was limited to knowing the devices used to send tweets on Twitter. Further investigation is required to reveal the address mac of the device used for tweeting and retweeting to track the owner of the device. However, doing such an investigation will require access to an available and robust dataset that includes all the features needed for prediction. Additionally, it is essential to compare several machine-

learning algorithms to determine the best and most suitable.

REFERENCES

Agnihotri, R., Dingus, R., Hu, M. Y., & Krush, M. T. (2016). Social media: Influencing customer satisfaction in B2B sales. *Industrial Marketing Management*, *53*, 172-180. https://doi.org/10.1016/j.indmarman.2015.09.003

Alkhodair, S. A., Ding, S. H., Fung, B. C., & Liu, J. (2020). Detecting breaking news rumours of emerging topics in social media. *Information Processing & Management*, *57*(2), 102018. https://doi.org/10.1016/j.ipm.2019.02.016

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, *31*(2), 211-36. https://doi.org/10.1257/jep.31.2.211

Amarappa, S., & Sathyanarayana, S. V. (2014). Data classification using Support Vector Machine (SVM), a simplified approach. *Int. J. Electron. Comput. Sci. Eng*, *3*, 435-445.

Amari, S. I., & Wu, S. (1999). Improving support vector machine classifiers by modifying kernel functions. *Neural Networks*, *12*(6), 783-789. https://doi.org/10.1016/S0893-6080(99)00032-5

Anduiza, E., Cristancho, C., & Sabucedo, J. M. (2014). Mobilization through online social networks: the political protest of the indignados in Spain. *Information, Communication & Society*, *17*(6), 750-764. https://doi.org/10.1080/1369118X.2013.808360

Arghode, V. (2012). Qualitative and Quantitative Research: Paradigmatic Differences. *Global Education Journal*, *2012*(4), 155-163.

Bampo, M., Ewing, M. T., Mather, D. R., Stewart, D., & Wallace, M. (2008). The effects of the social structure of digital networks on viral marketing performance. *Information systems research*, *19*(3), 273-290.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber?. *Psychological* science, 26(10), 1531-1542. https://doi.org/10.1177/0956797615594620

Bates, M. J. (2006). Fundamental forms of information. Journal of the American Society for information science and technology, 57(8), 1033-1045. https://doi.org/10.1002/asi.20369

Behera, P. C. (2016). Data Mining Technique for Tracking of Information Diffusion in Online Social Network. International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS), 5(4), 102-105.

Berthon, P. R., Pitt, L. F., Plangger, K., & Shapiro, D. (2012). Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy. Business Horizons, 55(3), 261-271. https://doi.org/10.1016/j.bushor.2012.01.007

Byun, H., & Lee, S. W. (2002). Applications of support vector machines for pattern recognition: A survey. In International Workshop on Support Vector Machines, 213-236. Springer, Berlin, Heidelberg.

Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. P. (2010). Measuring user influence in twitter: The million follower fallacy. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, George Washington University, May 23, 2010 – May 26, 2010, 10-17

Chang, C. I., & Ren, H. (2000). An experiment-based quantitative and comparative analysis of target detection and image classification algorithms for hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing, 38(2), 1044-1063. https://doi.org/10.1109/36.841984

Chang, H. C. (2010). A new perspective on Twitter hashtag use: Diffusion of innovation theory. Proceedings of the American Society for Information Science and Technology, 47(1), 1-4. https://doi.org/10.1002/meet.14504701295

Chen, W., Lakshmanan, L. V., & Castillo, C. (2013). Information and influence propagation in social networks. Synthesis Lectures on Data Management, 5(4), 1-177. https://doi.org/10.2200/S00527ED1V01Y201308DTM037

Cobb, L. (1981). Stochastic differential equations for the social sciences. Mathematical frontiers of the social and policy sciences, 37-68.

Conover, M. D., Gonçalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011). Predicting the political alignment of twitter users. In Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on, Boston, MA, USA — October 9-11, 192-199. https://doi.org/10.1109/PASSAT/SocialCom.2011.34

Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. Expert systems with applications, 34(1), 313-327. https://doi.org/10.1016/j.eswa.2006.09.038

Davoudi, A., & Chatterjee, M. (2015) "Probabilistic Spreading of Recommendations in Social Networks", In MILCOM 2015 Proceedings of the IEEE International Conference on Big Data, December 5-8, 2016, Washington DC, USA, 1373-1378. https://doi.org/10.1109/MILCOM.2015.7357636

Edosomwan, S., Prakasan, S. K., Kouame, D., Watson, J., and Seymour, T. (2011). The history of social media and its impact on business. Journal of Applied Management and entrepreneurship, 16(3), 79-91.

Eke, H. N., & Odoh, N. J. (2014). The use of social networking sites among the undergraduate students of University of Nigeria, Nsukka. Library Philosophy and Practice, 0_1.

Flake, G. W., & Lawrence, S. (2002). Efficient SVM regression training with SMO. Machine Learning, 46(1-3), 271-290. https://doi.org/10.1023/A:1012474916001

Frohlich, H., & Zell, A. (2005, July). Efficient parameter selection for support vector machines in classification and regression via model-based global optimization. In Proceedings of 2005 IEEE International Joint Conference on Neural Networks, 2005, 3, 1431-1436. https://doi.org/10.1109/IJCNN.2005.1556085

Guille, A., & Hacid, H. (2012). A predictive model for the temporal dynamics of information diffusion in online social networks. In *Proceedings of the 21st international conference on World Wide Web, Lyon, France — April 16 – 20*, 1145-1152. ACM. https://doi.org/10.1145/2187980.2188254

Guille, A., Hacid, H., Favre, C., & Zighed, D. A. (2013). Information diffusion in online social networks: A survey. *ACM Sigmod Record*, *42*(2), 17-28. https://doi.org/10.1145/2503792.2503797

Hoang, T. B. N., & Mothe, J. (2018). Predicting information diffusion on Twitter–Analysis of predictive features. Journal of computational science, 28, 257-264. https://doi.org/10.1016/j.jocs.2017.10.010

Hu, W., Singh, K. K., Xiao, F., Han, J., Chuah, C. N., & Lee, Y. J. (2018). Who Will Share My Image?: Predicting the Content Diffusion Path in Online Social Networks. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Marina Del Rey, CA, USA — February 05 – 09, 252-260. https://doi.org/10.1145/3159652.3159705

Hu, Y., Song, R. J., and Chen, M. (2017). Modelling for Information Diffusion in Online Social Networks via Hydrodynamics. IEEE Access, Wuhan, China — February 6, 5, 128-135. https://doi.org/10.1109/ACCESS.2016.2605009

Icha, O., & Agwu, E. (2015). Effectiveness of social media networks as a strategic tool for organizational marketing management. Internet Bank Commer 2015, S2. http://dx.doi.org/10.4172/1204-5357.S2-006

Jiang, C., Chen, Y., & Liu, K. R. (2014). Evolutionary dynamics of information diffusion over social networks. IEEE Transactions on Signal Processing, 62(17), 4573-4586. http://dx.doi.org/10.1109/TSP.2014.2339799

Jonsson, K., Kittler, J., Li, Y. P. & Matas, J. (2002). Support vector machines for face authentication. Image and Vision Computing, 20(5-6), 369-375. https://doi.org/10.1016/S0262-8856(02)00009-4

Juhl, R., Møller, J. K., Jørgensen, J. B., & Madsen, H. (2016). Modeling and prediction using stochastic differential equations. In Prediction Methods for Blood Glucose Concentration (pp. 183-209). Springer, Cham. https://doi.org/10.1007/978-3-319-25913-0_10

Kafeza, E., Kanavos, A., Makris, C., & Vikatos, P. (2014). Predicting information diffusion patterns in twitter. In IFIP International Conference on Artificial Intelligence Applications and Innovations, 79-89. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44654-6_8

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons, 53(1), 59-68. https://doi.org/10.1016/j.bushor.2009.09.003

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. Business Horizons, 54(3), 241-251. https://doi.org/10.1016/j.bushor.2011.01.005

Kursuncu, U., Gaur, M., Lokala, U., Thirunarayan, K., Sheth, A., & Arpinar, I. B. (2019). Predictive Analysis on Twitter: Techniques and Applications. In Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining, 67-104. Springer, Cham. https://doi.org/10.1007/978-3-319-94105-9_4

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In Proceedings of the 19th international conference on World wide web, 591-600. ACM. https://doi.org/10.1145/1772690.1772751

Lee, C. S., & Ma, L. (2012). News sharing in social media: The effect of gratifications and prior experience. Computers in human behaviour, 28(2), 331-339. https://doi.org/10.1016/j.chb.2011.10.002

Liu, L., Qu, B., Chen, B., Hanjalic, A., & Wang, H. (2018). Modelling of information diffusion on social networks with applications to WeChat. Physica A: Statistical Mechanics and its Applications, April 15, 496, 318-329. https://doi.org/10.1016/j.physa.2017.12.026

Lodha, S. K., Kreps, E. J., Helmbold, D. P., & Fitzpatrick, D. (2006). Aerial LiDAR data classification using support vector machines (SVM). In IEEE Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), 567-574. https://doi.org/10.1109/3DPVT.2006.23

Marquardt, C. G., Becker, K., & Ruiz, D. D. (2004). A pre-processing tool for web usage mining in the distance education domain. In Proceedings of IEEE International Database Engineering and Applications Symposium, 2004. IDEAS'04, 78-87. https://doi.org/10.1109/IDEAS.2004.1319780

Milstein, S., Chowdhury, A., Hochmuth, G., Lorica, B., and Magoulas, R. (2008). Twitter and the Micro-Messaging Revolution: Communication. Connections, and Immediacy-140 Characters at a Time.

Najar, A., Denoyer, L., & Gallinari, P. (2012). Predicting information diffusion on social networks with partial knowledge. In Proceedings of the 21st International Conference on World Wide Web, Lyon, France — April 16 - 20, 1197-1204. ACM. https://doi.org/10.1145/2187980.2188261

Oh, O., Agrawal, M., & Rao, H. R. (2013). Community intelligence and social media services: A rumour theoretic analysis of tweets during social crises. Mis Quarterly, 407-426.

Peters, K., Chen, Y., Kaplan, A. M., Ognibeni, B., & Pauwels, K. (2013). Social media metrics—A framework and guidelines for managing social media. Journal of interactive marketing, 27(4), 281-298. https://doi.org/10.1016/j.intmar.2013.09.007

Ripley, B. D., & Kelly, F. P. (1977). Markov point processes. Journal of the London Mathematical Society, 2(1), 188-192.

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In Proceedings of the 20th international conference on World wide web, 695-704. https://doi.org/10.1145/1963405.1963503

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th international conference on World wide web, Raleigh, North Carolina, USA, 851-860. ACM. https://doi.org/10.1145/1772690.1772777

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behaviour. Journal of management information systems, 29(4), 217-248. https://doi.org/10.2753/MIS0742-1222290408

Suykens, J. A., Van Gestel, T., Vandewalle, J., & De Moor, B. (2003). A support vector machine formulation to PCA analysis and its kernel version. IEEE Transactions on neural networks, 14(2), 447-450. https://doi.org/10.1109/TNN.2003.809414

Taxidou, I., & Fischer, P. (2013). Real-time analysis of information diffusion in social media. Proceedings of the VLDB Endowment, Riva del Garda, Trento, Italy — August 26 – 30, 6(12), 1416-1421. https://doi.org/10.14778/2536274.2536328

Taxidou, I., De Nies, T., Verborgh, R., Fischer, P. M., Mannens, E., & Van de Walle, R. (2015). Modelling information diffusion in social media as provenance with W3C PROV. In Proceedings of the 24th International Conference on World Wide Web, Florence, Italy — May 18 – 22, 819-824. ACM. https://doi.org/10.1145/2740908.2742475

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Election forecasts with Twitter: How 140 characters reflect the political landscape. Social science computer review, 29(4), 402-418. https://doi.org/10.1177/0894439310386557

Valenzuela, S. (2013). Unpacking the use of social media for protest behaviour: The roles of information, opinion expression, and activism. American Behavioral Scientist, 57(7), 920-942. https://doi.org/10.1177/0002764213479375

Van Lieshout, M. N. M. (2000). Markov point processes and their applications. World Scientific.

Walsham, G., & Sahay, S. (2006). Research on information systems in developing countries: Current landscape and future prospects. Information technology for development, 12(1), 7-24.

Wang, F., Wang, H., & Xu, K. (2012). Diffusive logistic model towards predicting information diffusion in online social networks. In Distributed Computing Systems Workshops (ICDCSW), 2012 32nd International Conference on, Macau, China — June 18-21, 133-139. IEEE. https://doi.org/10.1109/ICDCSW.2012.16

Wang, F., Wang, H., Xu, K., Wu, J., & Jia, X. (2013). Characterizing information diffusion in online social networks with the linear diffusive model. In Distributed Computing Systems (ICDCS), 2013 IEEE 33rd International Conference On, Philadelphia, PA, USA — July 8-11, 307-316. https://doi.org/10.1109/ICDCS.2013.14

Wang, Y., Theodorou, E., Verma, A., & Song, L. (2016). A stochastic differential equation framework for guiding information diffusion. arXiv preprint arXiv:1603.09021.

Wang, Y., Theodorou, E., Verma, A., & Song, L. (2018). A stochastic differential equation framework for guiding online user activities in closed loop. In International Conference on Artificial Intelligence and Statistics, 1077-1086.

Yang, C., & Wang, Y. (2016). Online social network image classification and application based on deep learning. DEStech Transactions on Engineering and Technology Research (iceta). https://doi.org/10.12783/dtetr/iceta2016/6970

Yang, J., & Leskovec, J. (2010). Modelling information diffusion in implicit networks. In 2010 IEEE International Conference on Data Mining, December 13 -17, Sydney, Australia, 599-608. https://doi.org/10.1109/ICDM.2010.22

Yang, J., & Counts, S. (2010). Predicting the speed, scale, and range of information diffusion in twitter. In Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC, USA, 355-358.

Zhang, Z. K., Liu, C., Zhan, X. X., Lu, X., Zhang, C. X., & Zhang, Y. C. (2016). Dynamics of information diffusion and its applications on complex networks. Physics Reports, 651, 1-34. https://doi.org/10.1016/j.physrep.2016.07.002

Zhu, Y., Zhong, E., Pan, S. J., Wang, X., Zhou, M., & Yang, Q. (2013). Predicting user activity level in social networks. In Proceedings of the 22nd ACM international conference on Information & Knowledge Management, San Francisco, CA, USA — October 27 - November 01, 159-168. https://doi.org/10.1145/2505515.2505518