CoMoSAOVA: COMPUTATIONAL MODEL FOR SENTIMENT ANALYSIS AND ONLINE VISIBILITY ASSESSMENT

EJODAMEN, P. U. [1,3], EKONG, V. E[1] AND *ASUQUO, D. E. [1,2]

[1]*Department of Computer Science, Faculty of Science, University of Uyo, Nigeria*
[2]*TETFund Center of Excellence in Computational Intelligence Research, University of Uyo, Nigeria*
[3]*Admiralty University of Nigeria, Ibusa, Nigeria.*
danielasuquo@uniuyo.edu.ng

ABSTRACT

The online visibility of a brand, product or organization could influence customers' decision to patronize it. Online reviews are opinions or emotions of customers that reveal their perception about the product over a period of time. The manual identification of features and sentiments toward an entity is a difficult task. In this study, a computational model was proposed to measure online visibility by mining Twitter data on discourse about a corporate entity – University of The Gambia. The numbers of Twitter posts, followers, followings, likes, retweets, quotes, replies and mentions serves as metrics for online visibility assessment. A linear regression evaluation between the age of the entity's account and the number of its followers showed 96.81% correlation. This shows that the older an active Twitter account, the higher the chance of increasing its followers and its visibility. Another section of the model predicts the tweet sentiment of the entity's followers with an accuracy of 93.68% using support vector machine and multilayer perceptron neural network. The computed average sentiment score of the case study was 0.3531996 based on Valence Aware Dictionary of sEntiment Reasoning (VADER) model. This means that positive sentiments were expressed in discussions on various issues where the entity was mentioned. The model will enable decision makers understand the sentiments expressed towards an entity. It also estimates online visibility of the entity based on its number of followers and the accounts' lifespan. The perceived sentiments will aid better decisions that could advance loyalty to the entity. Future studies would examine other computational models to predict various Twitter features that increases an entity's online visibility.

Keywords: Computational Model; Social Media; Twitter; Sentiment Analysis; Online Visibility

INTRODUCTION

Twitter is a microblogging platform that enables users to communicate by posting messages called 'Tweets', which can be read by other users. Tweets may contain up to 280 text characters as well as pictures, videos, or website links (Madni *et al*., 2023), where the use of hashtags between words facilitates information search, retrieval, and processing (Radiuk *et al.,* 2022). The first tweet was posted on the 21st of March 2006 by the creator of Twitter – Jack Dorsey. Since then, over 6,000 tweets are generated every second, around 350,000 per minute, and an estimated 500 million posts daily (Twitter Usage Statistics, 2021).

There are over 7 billion visitors interacting on Twitter, generating volumes of various forms of data which can well be described as Big Data. Bakshy *et al*. (2011) noted that tweets from influential users provide cost-effective means to diffuse information. The more users post tweets and other users interact with them, the visibility of such account increases. A Twitter account can be created for an entity – that is an individual, organization, a product, or anything that exist independently for which information may be generated and stored. The awareness or knowledge about an entity on the internet is referred to as online visibility. It is difficult to determine the reach of a tweet by mere reading it. The size of tweets posted in real-time makes it more difficult for physical analysis of each tweet. Hence, this study proposes a model to estimate the online visibility of an entity using the age of the account and the number of followers it has.

Furthermore, the process of applying natural language processing (NLP) techniques and computer language to analytically identify, extract and manipulate relevant information about product reviews in order to classify the opinions into positive, negative or neutral polarity is known as sentiment analysis (Kumar and Zymbler, 2019). The outcome of such process can provide insights on the needs and expectations of customers, help decision makers respond quickly to changing customers' needs, improve customers' satisfaction and maintain brand loyalty. It is noteworthy that a negative sentiment about an entity expressed by an account with high visibility, can cause significant damage within a very short time. So, early detection of sentiments and analysis of their impact will help in making necessary decisions.

Features for Visibility Assessment on Twitter

In Twitter, a tweet can be seen by other users but the level of visibility depends on factors such as number of followers, likes, favourites and retweets. A Twitter user whose posts are well received will attract other users. When a user chooses to follow updates of another user, it is referred to as 'following'. Following a tweet is similar to subscribing to the tweets of that user. It is an indication that the "following" account is interested in posts from that user. A follower can also receive a direct message (DM) from the person. This type of relationship provides the opportunity to exercise some form of influence on other users. It means all the followers will have their timelines displaying posts from the

Open Access article published under the terms of a
Creative Commons license (CC BY).
http://wojast.org

*Ejodamen et al: CoMoSAOVA: Computational Model for Sentiment
Analysis and Online Visibility Assessment*
https://dx.doi.org/10.4314/WOJAST.v15i1.110

influencing user, thereby increasing the reach of their tweets in the network. While positive and beneficial tweets can be posted, there is also the tendency to post propaganda to large followers in the network. However, a follower can simply 'unfollow' any account deemed offensive or undesirable without the person being notified.

In most cases, influential Twitter users also follow others in the network. They like, retweet, and reply to other posts too. The user whose updates is followed up is known as a 'friend' or 'following'. The timeline of a user is updated by posts from accounts it is "following", while its post is displayed on timelines of its "followers". To measure users' social influence and associations, it is imperative to observe the number of their friends – both those they follow and those that follow them.

The 'Like' feature in Twitter, may increase the visibility of the entity since the liked tweet could be displayed on the "followers" home timeline. In Twitter, an entity can be included in a conversation using the "user mention" feature where the user's name is preceded by an '@' symbol; for example, @UniOfGambia. Mentions are necessary if such entity is an interested party or if their attention is deliberately being drawn to the post. A user is notified on the Twitter 'Notification tab' of any mentions but the post can only be visible to the public if the tweet of the sender is not protected. A retweet is a Twitter post that is publicly shared with followers of an account (Twitter Help Centre, 2022). One's account can be retweeted, as well as account of others, so that followers of the account can view the post. This greatly increases the online visibility of the account to all of its followers and followers of its followers.

Sentiment Analysis
Sentiment Analysis is known as opinion mining. It refers to the analysis of emotions and opinions of people from text (Swathi *et al.*, 2019) about an entity. Sentiment analysis of tweets is vital in gaining insights into customers' behaviours, citizen feelings towards their government, movie acceptance, response to acts of terrorism, as well as disease outbreaks and other opinions in social media (Çilgin *et al.*, 2022; Ekong *et al.*, 2021; Swathi *et al.*, 2019; Elbagirand Yang, 2019).

Valence Aware Dictionary of sEntiment Reasoning (VADER) model was developed, validated and evaluated by Hutto and Gilbert (2014). It is a parsimonious rule-based model optimized for sentiment score computation of text. Although, VADER depends on human-curated valence-based sentiment lexicon and it is generalizable to fit into other domains, it determines if a text is positive, negative, or neutral; the normalized compound of these is also computed. The model considers the brevity in social media communications including the sentiment properties of slangs, abbreviations, and emojis/emoticons. The authors reported its superior performance over human raters and compared remarkably to other major sentiment analysis models such as SentiWordNet, the General Inquirer, Linguistic Inquiry and Word Count (LIWC), Affective Norms for English Words (ANEW), and so on. It also

compared favourably to some machine learning (ML)-based techniques like the Maximum Entropy, Naïve Bayes, and support vector machines (SVM) algorithms.

VADER was designed to be used with any general task of sentiment detection. In addition, it is fine-tuned to handle the nature of text common with social media platforms such as Twitter and Facebook (Hutto and Gilbert, 2014). Although generalizable, VADER depends on human-curated valence-based sentiment lexicon which can be extended so as to fit into various domains. Its speed of execution makes it suitable for online streaming task executions. Many other sentiment detecting models such as LIWC and the General Inquirer have lexicons that have been widely validated. However, they do not consider some vital lexical items with sentiment properties like is found in emoticons, initialisms, acronyms, and slangs, among others. Also neglected is the variation of sentiment intensity in words. The words 'excellent' and 'good' are positive words but with varying intensity. VADER model has been used in previous studies with satisfactory results (Çilgin *et al.,* 2022; Elbagir and Yang, 2019). In addition, Laxmi *et al.* (2020) affirms that VADER maintains its accuracy while adopting a fast and efficient computational ability. This study leverages on the VADER model and proposes a computational model for sentiment analysis and online visibility assessment using University of The Gambia as case study.

A linear regression ML model is proposed for online visibility assessment on Twitter. The ML model determines Twitter user account and the potentials of increasing their visibility. The organization of the paper is as follows: Section 2 discusses relevant literature regarding online visibility and sentiment analysis approaches. Section 3 provides information about the proposed methodology. Section 4 presents the experimental results while Section 5 finally concludes the paper.

Online visibility has been studied by scholars due to its wide applicability, especially for online marketing. There have been studies that developed different methodologies for enhancing the visibility of an entity on social media platforms. Rathore and Tripathy (2021) proposed an exponential model aimed at measuring the visibility of tweets while following an account. They implemented a deep neural network (DNN) and polynomial regression to predict the visibility of an account on a social media network. In Song *et al.* (2021), a visibility estimation method based on deep label distribution learning (LDL) in a cloud environment was proposed. The model combined cloud computing and image processing to estimate visibility efficiently. Cloud computing made the front-end monitoring device thinner. The camera captured and compressed the images, while a high-performance network was utilized to transmit the image data from the camera to the cloud platform. This alleviated the lack of front-end computing power in the visibility estimation application. The method saved the required resources by utilizing high-performance

Open Access article published under the terms of a
Creative Commons license (CC BY).
http://wojast.org

*Ejodamen et al: CoMoSAOVA: Computational Model for Sentiment
Analysis and Online Visibility Assessment*
https://dx.doi.org/10.4314/WOJAST.v15i1.110

parallel computing of Graphics Processing Units (GPU) in cloud computing services.

In Huihui *et al*. (2021), a UserRBPM framework to predict the action status of a user given the action statuses of her near neighbours and her local structural information was proposed. Experiments on a large-scale real-world dataset have shown that the User RBPM significantly outperformed baselines with handcrafted features in user retweet behaviour prediction. Linhong and Kristine (2014) developed an intuitive model that used visibility of a name to predict new links. The model accounted for visibility of another user's name to a given user based on the number of messages containing new names a user receives in her social media stream and the frequency with which she visits the stream to view new messages. The proposed model can better predict new follow and co-mention links than alternative link prediction algorithms. It is observed that ML models can be used to enhance the visibility of an entity on a social media platform.

NLP models can be used to analyse and understand text-based contents, such as social media posts, comments, and reviews, in order to identify and extract relevant information, such as key topics, sentiments, and entities. Collaborative filtering models are another class that can be used to recommend content to users based on their interests and behaviours, such as the posts they have liked, shared, or commented on (Lei *et al.,* 2018).

Deep Learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been used to analyse images and videos, such as those shared on social media, in order to identify and extract visual features, such as objects, scenes, and emotions (Lei et al, 2018).Graph Embedding models are another class that can be used to analyse the structure and relationships between social media users and entities, such as universities, in order to identify and extract relevant communities, influencers, and opinion leaders. Sentiment analysis model on the other hand can be used to classify the feedback of the customer, this can help in understanding the customer's view on the university (Jain *et al.,* 2021).

Logistic regression is another model that can be used for binary classification tasks, such as determining whether a tweet about an entity (e.g. a university) is positive or negative. It can be a good model for predicting the visibility of an entity in a Twitter social media network, if the goal is to classify the sentiment of tweets about that entity. However, it is important to note that the performance of logistic regression will depend on the quality and representativeness of the data used to train the model. If the data is not representative of the target population or if it is imbalanced, the model may not perform well. Therefore, linear regression models are better suited for prediction of a continuous variable such as visibility. However, logistic regression can be used as a feature in a more complex model like a neural network (Lei *et al.,* 2018).

Furthermore, an ensemble method involving logistic regression has shown promise. In Roy et al. (2022), an ensemble learning system was developed using K-Nearest Neighbour, logistic regression, and decision trees with the predictions determined by a voting process. The system experimented on detecting anti-social behaviour such as cyber bullying. However, the impact of the visibility of such anti-social post on Twitter was not examined. In twitter, entities may bully others and can be bullied also. This may be accompanied by several negative implications which could be worsened by extensive visibility of such post.

It is also worth noting that Twitter is a dynamic platform and the visibility of an entity can be affected by various factors such as trending topics, advertisement, and so on. These factors can change rapidly and constantly, therefore, a more dynamic and adaptive model such as a linear regression or RNN might be a better fit for the task. Although depending on specific use cases, these models may be more or less appropriate (Jain *et al.,* 2021; Lei *et al.,* 2018).

Online visibility of an entity could be amplified by sharing information with the public. Ciccone *et al.* (2021) reported the use of Twitter engagement metrics in determining the extents of an entity's reach online. The study noted that the volume of posts and the follower base were important in estimating the online visibility of the entity.

In Park and Kaye (2019), it was reported that an author's and message's characteristics influences the expansion of its visibility on Twitter. While the message may determine if it will be retweeted, it was noted that the number of followers of the entity has great influence on how visible this will be.

MATERIALS AND METHODS

As at December 2022, there were over 36 million tweets posted every hour, 867 million tweets daily and 26 billion tweets posted monthly (Yaqub, 2022). Manually searching through every single tweet in the voluminous twitter data is unachievable in record time. Hence, a computational approach is adopted to sieve only data relevant to this study. Snowball sampling technique was adopted for selecting twitter features for analysis. The Snowball technique requires starting with a sample or respondent, then that respondent would help identify other respondents (Bhardwaj, 2019; Linhongand Kristine, 2014). In this case, the entity of interest serves as the seed data sample, and other samples needed are derived from it. Particularly, the twitter account identification for the entity is used to retrieve details about the entity and then pull the details of every follower of that entity.

The selection criteria for a tweet are that the tweet should contain a mention of the entity and should be an original tweet. For this experiment, "@UnivOfGambia" was used as the seed entity with which to start data mining; thereafter information of all the followers of the account was also mined. The architecture of the proposed computational model for sentiment analysis and online visibility assessment (CoMoSAOVA) is shown in Figure 1.

Open Access article published under the terms of a
Creative Commons license (CC BY).
http://wojast.org

Ejodamen et al: CoMoSAOVA: Computational Model for Sentiment
Analysis and Online Visibility Assessment
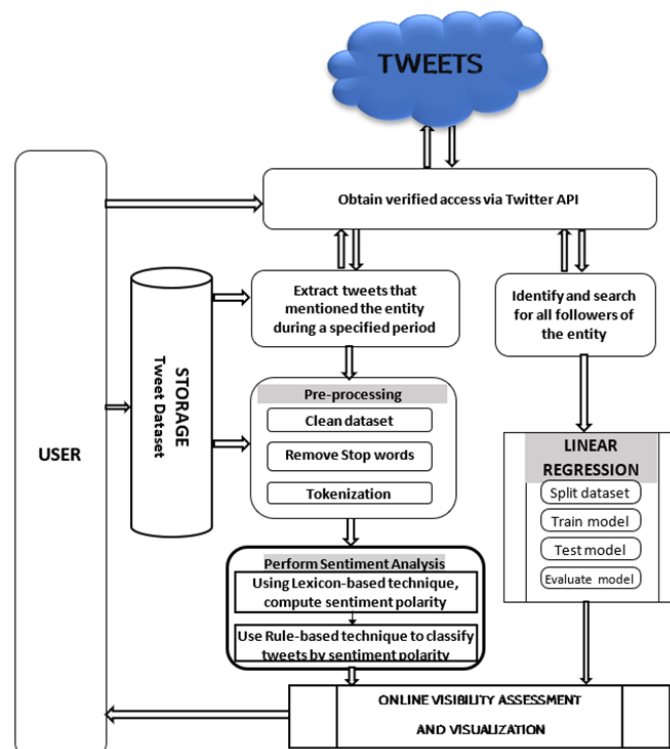https://dx.doi.org/10.4314/WOJAST.v15i1.110

Figure 1: Architecture of the CoMoSAOVA model

Twitter data provide useful resource to understand the opinion of people about entities. Python programming language was used to analyse the tweets after it was retrieved, cleaned, stored, pre-processed and classified according to the sentiment expressed. The mined tweets undergo pre-processing before sentiment analysis is performed on it using a hybrid of lexicon-based and rule-based techniques.

A linear regression model is also deployed to make predictions about the entity. Actionable insights derived from the analysis and visualization of the data is made available to the user. Panda library was used to create vectorised data frames, which enhanced the data wrangling processes. Also, the regular expression library provided data manipulation capabilities. Natural Language Toolkit (NLTK) was used to pre-process the tweet dataset.

Data Collection
Twitter Application Programming Interface (API) provides access to historic tweets. Roesslein's(2021)Tweepy library, a Twitter API wrapper, was used to download data from Twitter. For this study, an approved API access credentials were used for data mining of tweets from 25th September 2014 – when @UniOfGambia Twitter account was created – to 3pm (Greenwich Mean Time) on 31st December 2022. A total of 3552 tweets were sieved from the hundreds of millions of tweets posted during the period. Furthermore, the engagement metrics critical in the assessment of an entity's online visibility include the number of likes, retweets, replies, and quotes. Also, details of all the accounts following the entity were mined. It was noted that as at the

time of data collection, there were 6011 followers of the entity out of which 538 of them had mentioned the entity in their post.

Data Pre-processing
Downloaded dataset were stored as comma separated value (CSV) files and a visual assessment of the dataset to check for data quality and completeness. As is common with data from social media, the downloaded tweets had several noisy data such as user mentions, hashtags, stop-words, and so on. Stop-words are words (such as 'the', 'an', 'you', etc.) that do not add any lexicon value to a block of text. These were removed.

Similarly, other twitter users could be mentioned in a posted tweet in the form '@username'. For the sentiment analysis of the tweets, user mentions are irrelevant and excluded before lexicon polarity computation. Similarly, hashtags are words preceded with a '#' (e.g #WeCanRead), often used for emphasis about a topic. It is commonly used in social media to guide interested users to locate topics and interact with the community. While hashtags may provide some insights such as trending topics, this study does not consider it in its analysis. Also, web links were in several of the tweets and since they are irrelevant in this study were deleted from the tweets.

Sentiment Classification
Lexicon-based techniques assumes that the polarities of the individual phrases or words can be summed up to give the polarity of a sentence or documents. This study adopts the VADER model for polarity computation of words in a tweet

to determine its overall sentiment polarity. VADER has a corpus of words and sentiment scores assigned to each. These scores were derived from experts' opinion of each word. The sentiment score of a sentence/document/tweet is a summation of sentiments values of individual words present. A rule-based technique, developed by Hutto and Gilbert (2014), was then applied to classify tweets based on the sentiment polarity as shown in equation 1. The sentiment of a tweet, *s,* is *positive* if the polarity of the tweet is greater than or equal to plus 0.05 and *negative* if it is less than or equal to minus 0.05. Plain tweets users post that do not carry any sentiment or has negligible sentiment score between less than minus 0.05 and less than plus 0.05 are classified as *neutral.*

$$Sentiment\ (s) = \begin{cases} Positive & s \geq 0.05 \\ Neutral & -0.05 > s < 0.05 \\ Negative & s \leq -0.05 \end{cases} \quad (1)$$

Adopting VADER model for computing sentiment scores is preferred due to its generalisability to several domains. It has been validated by previous studies as an effective means to automatically compute word sentiment scores, especially in cases of rapid growth in volume of tweets (Abiola, *et al.,* 2023*;* Hutto and Gilbert, 2014; Laxmi, *et al.*, 2020).

Sentiment and online visibility assessment

The sentiment analysis module, as shown in Figure 1, classified tweets into positive, neutral and negative using the boundaries in Equation 1. The resultant dataset can be used fortraining with supervised classification algorithms using the derived sentiments as target/output variables. In this study, SVM and MLPNN were used to build the model. Previous studies that used these algorithms reported results that out-performed several other compared models (Naw, 2018). Similarly, linear regression can be used to predict continuous variables such as the counts of an entity's followers.

RESULTS AND DISCUSSION
Tweet Engagement Metrics

Twitter provides means for users to interact with the tweets of others. The main tools used for engaging a tweet are the number of likes, retweets, replies, and quotes. These engagement metrics can be used to determine the online visibility of the tweet. For instance, when a user retweets a post, all subscribers/friends of that user account will see the post; thereby extending its reach. A user with an extensive reach can, somehow, influence the decision of others.

It is worthy of note that most users chose to engage with tweets by liking it. Figure 2 shows an over-whelming number of likes than other engagements metrics. It can be seen that users seldom quoted tweets from the entity under study.
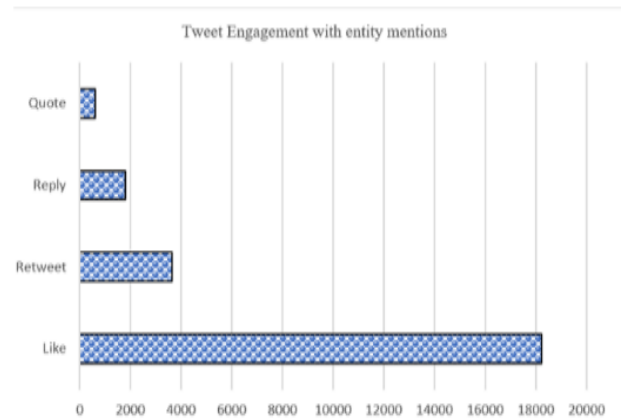

Figure 2: Visualization of tweet engagement

Furthermore, Table 1 shows the correlation between these tweet engagements. There was very strong relationship between "Like" and "Retweet" – 81.1%, "Like" and "Reply" – 80.6%as well as "Like" and "Quote" – 72.0%. However, the relationship between retweet and reply (59.8%) was the weakest in this dataset; those who retweet may not necessarily post a reply. Other correlations showed moderate relationships.

Table 1: Correlation of tweet engagement metrics

	Like	Retweet	Reply	Quote
Like	1.00000	0.81084	0.80610	0.71960
Retweet	0.81084	1.00000	0.59801	0.67203
Reply	0.80610	0.59801	1.00000	0.68244
Quote	0.71960	0.67203	0.68244	1.00000

Therefore, it can be rightfully stated that "Like" feature is the most important engagement metric for a tweet's visibility and in determining how well a tweet is accepted. Those who like a tweet could also retweet or reply it, giving it more guaranteed visibility.

Experimental Case Study

This model is designed, amongst other capabilities, to provide insights from the data. In this experiment, the University of The Gambia was used as the entity. The entity is mentioned 3,552 times in original tweets, exclusive of retweets. Visual assessment of relevant tweets posted, as shown in Figure 3, indicates the progression of entity mentions in tweets over the years. This information can help decision makers understand the rate of growth of the entity's visibility in Twitter. A visualization of the data reveals that all the tweets mentioning the entity were from followers of the account. It is an indication that account followers are major determinants of the diffusion of information from the entity, as well as its visibility.
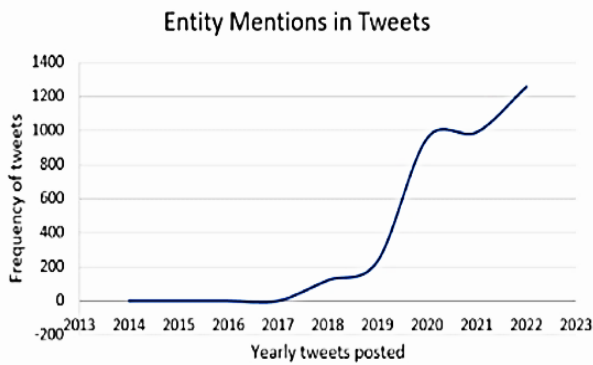
Figure 3: Number of tweets posted per year

Figure 3: Progression of entity mentions in tweets over the years

Out of 3,552 sample tweets analysed, 64% (2289) were positive sentiments while 9% (306) expressed negative sentiments as shown in Figure 4. Notably, 27% (957) were neutral; meaning they were neither positive nor negative in posts where the entity was mentioned. It should be noted that sentiments being analysed may not necessarily be about the entity solely. It could simply be that the entity is a stakeholder or should be involved in that discourse. Among the 9% of the entity's followers who mentioned it in their post, an average of 81% expressed positive sentiments, 8% were negative, while 11% were neutral.

(a)
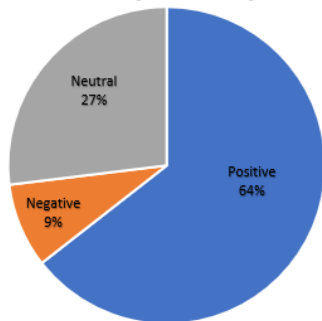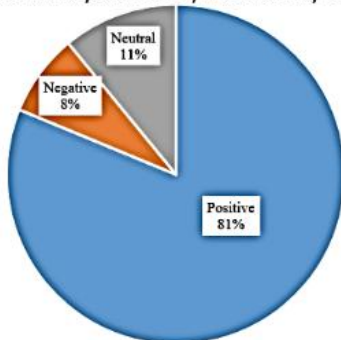
(b)

Figure. 4: Sentiment analysis of tweets for (a) entity mentions and (b) followers

Furthermore, the average sentiments expressed in the relevant tweets were computed. This ensures that all tweets posted and relevant to the discourse is vital in considering a comprehensive sentiment. The sentiment score, determined by the lexicon-based approach with the aid of the VADER model, was used to calculate the average sentiments. For example, the computed average sentiment score of the case study was 0.3531996. This means that, considering all tweets where the entity was mentioned, positive sentiments were expressed in discussions on various issues where the entity was mentioned. This assessment is necessary to aid decision making, as it shows the perceptions about the brand or issues related to it.

Machine Learning Models

The dataset is split into two, 70% for training the model and 30% for testing purpose. The importance of account followers cannot be overemphasized. In this study, it was observed that there is strong correlation between the number of followers of an entity and how old the account has been in existence. In other words, the number of followers of an entity (active account) is likely to increase with time. A simple linear regression model was developed with the age of the entity's account as the independent variable and the number of followers as the dependent variable as shown in Figure 5. The "Linear Regression" module in Sci-kit learn Python library was used.
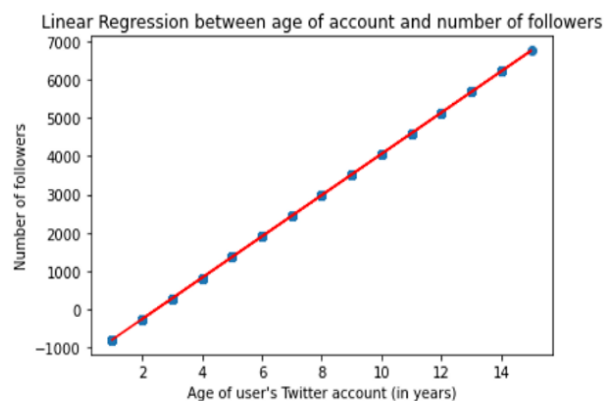

Figure 5: Linear Regression of account age and number of followers

The model when evaluated produced 96.81% accuracy, with a mean average percentage error (MAPE) of 3.19% and a root mean squared error (RMSE) of 26.90%. The result indicates that there is satisfactory correlation between the age of the Twitter account and the number of its followers. Hence, the number of followers of an active account could be predicted if the date of account creation is known. By extension, this is critical to estimating the level of an entity's visibility, since the number of followers of an account influences its information diffusion.

Furthermore, in designing this model, the entity's account features were used, namely; the age of the account, the total number of tweets the entity has posted during its lifetime, the number of followers and those it follows. These features

influence an entity's visibility in Twitter, as well as predict the sentiments that would be expressed in the next post by its followers.

An SVM linear kernel was also modelled, producing an accuracy of 93.68%. Similarly, a multilayer perceptron neural network (MLPNN) model with rectified linear unit activation function and a stochastic gradient-based weight optimizer also produced a 93.68% accuracy after 1000 maximum iterations. Consequently, this section of the model predicts the sentiment of its followers who sees the entity's post and could engage the tweet to give it more visibility. This is important because the sentiment of a tweet by a follower affects the general perception of an entity.

The model provides decision makers a means to obtain information about emotions towards an entity. There may be need to quickly detect negative sentiments and address the issues before it spreads beyond control. Linear regression model enables estimating the visibility of a tweet based on the number of the account followers. While a positive sentiment is desirable to spread extensively, a negative may need to be curtailed. This model provides the necessary information to aid swift decisions in this regard.

CONCLUSION
Twitter is a social media platform for online conversations and easy diffusion of information. While conversing, twitter users' express emotions - consciously or otherwise. A post may contain mentions of entities whose attention is being attracted to the conversation. Similarly, a post by an entity is available to its followers and the public who tend to 'like' or 'retweet' tweets they agree with. These increase the visibility of the entity who made the post or who was mentioned in it. In this study, a computational model is proposed to analyse the online visibility of an entity. It was observed that important features for determining entity visibility in Twitter includes the age of the account, the total tweets posted since then, and the number of followers. Also important are the number of entity's mentions in conversations, and the likes and retweets such post attracted.

A linear regression segment of the proposed model estimated the Twitter feature that influences online visibility, namely the number of followers of an account based on the date an active account was created and the number of its followers. The linear regression evaluation between the total number of tweets the user has posted and the number of followers of its account showed 96.81% accuracy, with a mean average percentage error (MAPE) of 3.19%. This showed that, the older an active Twitter account, the higher the chance of increasing its followers and its visibility. Subsequently, SVM and MLPNN models showed that the sentiment of such followers' tweets can be predicted with an accuracy of 93.68%.

This study used University of The Gambia as its case study in experimenting with the proposed model. The computed average sentiment score of the case study was 0.3531996 based on VADER model, which indicates that positive sentiments were expressed in discussions on various issues where the entity was mentioned. The dataset used was imbalanced, having most sentiments detected as positive. This limitation may be addressed in future by training the ML algorithms with a balanced dataset to improve its generalizationin prediction. Also, future studies would examine other computational models to predict various Twitter features that increases online visibility of an entity.

ACKNOWLEDGMENTS

REFERENCES
Abiola, O., Abayomi-Alli, A. Tale, O. A., Misra, S., and Abayomi-Alli, O. (2023). Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser. *Journal of Electrical Systems and Information Technology*, 10(5):1–20. https://doi.org/10.1186/s43067-023-00070-9

Bakshy, E., Hofman, J. M, Mason, W. A and Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on twitter. *In Proceedings of the 4th ACM International Conference on Web Search and Data Mining*(WSDM), pp. 65–74. https://doi.org/P10.1145/1935826.1935845

Bianchi, F., Cutrona, V and Hovy, D. (2011). Twitter Demographer: A flow-based tool to enrich twitter data. Available online at: https://doi.org/10.48550/arXiv.2201.10986. (Accessed 24-12-2022).

Bhardwaj, P. (2019). Types of Sampling in Research. *Journal of the Practice of Cardiovascular Sciences*, 5:157-163.

Ciccone, L. S., Van den Driessche G. A., and Gallardo-Williams, M. T. (2021). Using Twitter to Extend Your Department's Outreach and Visibility. *Journal of STEM Outreach*, 4(1): 1 – 5.

Çilgin, C., Baş, M., Bilgehan, H., and Ünal, C. (2022). Twitter Sentiment Analysis during Covid-19 Outbreak with VADER. *Academic Journal of Information Technology*. 13(49):72-89.

Ekong, V. E., Asuquo, D. E., and Ejodamen, P. U. (2021). Towards an Analytic Framework for Twitter influencers in Boko Haram Discourse. *In Proc. of the International Conference on Innovative Systems for Digital Economy* (ISDE). 32:204–212, Available online at: https://library.ncs.org.ng/conference-proceedings-vol-32-August-2021/ (Accessed 12-12-2021)

Elbagir, S. and Yang, J. (2019). Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment. *In Proc. of the International Multi-Conference of Engineers and Computer Scientists* (IMECS 2019), March 13-15, 2019, Hong Kong.

Huihui, G., Li, Y. and Zeyu, L. (2021). UserRBPM: User Retweet Behavior Prediction with Graph Representation Learning, *Hindawi Wireless Communications and Mobile Computing*, Vol. 2021. Available online at: https://doi.org/10.1155/2021/4431416 (Accessed 24-12-2022).

Hutto, C. J. and Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media* (ICWSM-14). Ann Arbor, MI, USA.

Jain, P. K, Rajendra, P., and Gautam, S. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews, *Computer Science Review*, 41: https://doi.org/10.1016/j.cosrev.2021.100413

Kumar, S. and Zymbler, M. (2019). A machine learning approach to analyze customer satisfaction from airline tweets. *Journal of Big Data,* 6(62). https://doi.org/10.1186/s40537-019-0224-1.

Laxmi, B. L., Yamini, B., Rakshitha, C. H., Keerthi, D., and Scholars, B. E. 2020. Twitter sentiment analysis using VADER on Python. *Journal of Emerging Technologies and Innovative Research*, 7(5): 1025 – 1031.

Lei, Z., Shuai, W and Bing, L. (2018). Deep Learning for Sentiment Analysis - A Survey, WIRES –*Wiley Interdisciplinary Reviews*, 8(4). Available online at: https://doi.org/10.48550/arXiv.1801.07883 (Accessed 28-12-2022)

Linhong, Z and Kristine, L. (2014). A visibility based model for link prediction in social media, *In Proc. of the ASE/IEEE conference on social computing.* Available online at: http://www.isi.edu/integration/people/lerman/papers/SocialCom14.pdf (Accessed 15-12-2022)

Madni, H.A., Umer, M., Abuzinadah, N., Hu, Y.C., Saidani, O., Alsubai, S., Hamdi, M., and Ashraf, I. (2023). Improving Sentiment Prediction of Textual Tweets Using Feature Fusion and DeepMachine Ensemble Model. *Electronics* 2023, 12, 1302.https://doi.org/10.3390/electronics12061302

Naw, N. (2018). Twitter Sentiment Analysis Using Support Vector Machine and K-NN Classifiers. *International Journal of Scientific and Research Publications*, 8(10), pp 407 – 411. http://dx.doi.org/10.29322/IJSRP.8.10.2018.p8252

Park, C. S. and Kaye, B. K. (2019). Expanding Visibility on Twitter: Author and Message Characteristics and Retweeting. Social media + society, Sage.

Radiuk, P., Pavlova, O. and Hrypynska, N. (2022). An Ensemble Machine Learning Approach for Twitter Sentiment Analysis, I*n Proc. of the 6th International Conference on Computational Linguistics and Intelligent Systems*, COLINS-2022, May 12–13, 2022, Gliwice, Poland.

Rathore N. C. and Tripathy S. (2021). Effective Visibility Prediction on Online Social Network. *IEEE Transactions on Computational Social Systems*, 8(2): 355-364.

Roesslein, J. (2021). Tweepy: Twitter for Python! Available online at: https://Github.com/Tweepy/Tweepy. (Accessed 11-12-2022)

Roy, P. K., Singh, A., Tripathy, A. K. and Das, T. K. (2022). Cyberbullying detection: an ensemble learning approach. *International Journal of Computational Science and Engineering*, 25(3): 315-324.

Song, M., Han, X., Liu, X.F. *et al.* (2021). Visibility estimation via deep label distribution learning in cloud environment. Journal of Cloud Computing, **10**(46). Available online at: https://doi.org/10.1186/s13677-021-00261-7 (Accessed 24-12-2022).

Swathi, H., Aravinth, S. S., Nivethitha, V., Saranya, T., and Nivethanandhini, R. (2019). Sentiment Analysis of Movie Review using data Analytics Techniques. *Iconic Research and Engineering Journals*, *2*(9): 48-52.

Twitter Help Center. How to Retweet. Available online at: https://help.twitter.com/en/using-twitter/how-to-tweet (Accessed 24-12-2022).

Yaqub, M. How Many Tweets per Day (2022). (New Data). Available online at: https://www.renolon.com/number-of-tweets-per-day/ (Accessed 31-12-2022).