Prediction of User Behaviour through the Interaction in Social Media Using Deep Learning Approach

P. Mathiazhagan [1], C. Gopi [2], A. Arun [3]

[1] Assistant Professor, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore, Tamil Nadu, India

[2, 3] PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore, Tamil Nadu, India

Abstract: Predicting human behavior and personality from the social media applications like Facebook, Twitter and Instagram is achieving tremendous attention among researchers. Statistical information about the human thoughts expressed via status on social media is essential assets for research in predicting various human behaviour and personality. In addition to accurate prediction, having the capacity to understand the roles of human behavior determinants and to provide explanations for the predicted behaviors is also important. However, most prediction models do not provide explanations for the behaviors they predict. In this paper, user behaviour prediction is achieved using the Natural Language Processing (NLP) and Two-Level Classifier approach. User behaviour prediction on social media is usually defined as the problem of estimating the rating scores, view counts, or click through of a post. The proposed methodologies outperform compare with the existing approaches.

Keywords: Social Media, User Behaviour, Deep Learning, Prediction, User Interaction, Classification and NLP.

1. INTRODUCTION

With the development of information technologies, Internet access has become more available and in turn, online social networks have gained an increasing number of users and are currently used for a variety of purposes. They have become digital interaction and communication platforms. The transformation of communication and social networks has considerable effects on economics, politics, society and many other global aspects. Social networking platforms do not just lead the changes in cultural transformation; they are also the important information resources for commerce, research, business, and many other fields [1]. The digital world's takeover of communication has accelerated the development of social media platforms. According to the We Are Social's 4th quarterly report of 2019, 5,155 billion people, 67% of the world population, are using the Internet and 3,725 billion people, 48% of the world population, are using social media. The report states that the number of social media users has increased by 3% since the last quarter. There is a 10% increase in Internet and social media use with respect to last year. Almost all Internet users are associated with social media. According to the GlobalWebIndex report of 2019, Internet users spend a daily average of 2 hours and 15 minutes on social media, one third of the average of time users spend on the Internet. Many business segments, such as corporations, enterprises and entrepreneurs, can benefit from the interactions on social media. Misleading interactions between users on social networks may cause financial and intangible danger for businesses, producers, etc. Therefore, it is important to determine whether users on social network are malicious. User interactions on Facebook form the measurable datasets. Accordingly, social network analysis can reveal the relationship structures based on the analysis of user behavior [2].

Social media hosts different forms of abuse, one of which is fake interactions. These interactions can manipulate people in many ways. Fake interactions on e-commerce platforms with social media accounts sabotage the secure shopping environment for consumers. It is possible to create interactions such as news, comments and likes. Praises or criticize certain products and the perceptions created by these interactions can mislead consumers about certain products and services. Fake interactions coming from malicious accounts can manipulate statistical data and investment–support processes. Consequently, fake interactions can affect political processes along with the economy

[3]. Using fake followers, reputation and popularity can be achieved, and a strong profile image can be built in the eyes of society. This manipulative method can be used by some politicians to increase their number of interactions on social media with fake-bot accounts to increase their reputation. Cyber criminals use Facebook for spamming, spreading malicious messages, sending phishing links or filling the network with fake accounts, and for other malicious activities [4].

This creates a large challenge for social media platforms. Social media platforms need to provide the ability to connect companies with users. The field of advertising on social media is new. Platforms want to justify their effectiveness to potential advertisers. Measuring this effectiveness after advertising is possible. Yet, companies prefer a way to calculate return on investment (ROI) before investing. One way to calculate ROI is to measure user engagement. This study predicts user engagement for advertisement posts. The user engagement metrics include share count, comment count, and comment sentiment. This study explores predicting user engagement with different types of deep learning models. The models are unique in their incorporation of text and image data. The goal is to understand which model-types best predict user engagement. The research applies these models to a real-world use-case. The research uses models, which can predict user engagement, in order to vet for better performing ads. Platforms could use these models to inform advertisers which of their ads will perform best on the platform. This allows advertisers have their ads vetted. The vetting could prevent advertisers from spending a lot of money showing worse ads. Moreover, the vetting would allow advertisers to only show ads that will perform best.

2. RELATED WORK

Most current approaches for predicting the behaviour of Social Media users are based either on user-generated content or on the social ties. Initially, which has been investigated thoroughly, uses textual features from tweets to build location predictive models [11]. The latter arises from an observation that a user often interacts with people in nearby areas, and exploits the network connections of users.

Kamalesh, M. D., & Bharathi, B. (2022), focuses on guessing user personality based on big five personality traits. An intelligent Sentence analysis model is built to extract personality features [5]. Predicting human behavior and personality from the social media applications like Facebook, Twitter and Instagram is achieving tremendous attention among researchers. Statistical information about the human thoughts expressed via status on social media is essential assets for research in predicting various human behaviour and personality. A new Binary-Partitioning Transformer (BPT) with Term Frequency & Inverse Gravity Moment (TF-IGM) is proposed that identifies relationships among feature sets and traits from datasets. The proposed work outperforms the all feature extraction average baseline set on multiple social datasets.

Abousaleh, F. S., et al, (2020), proposed a deep learning model, called visual-social CNN (VSCNN), which predicts the popularity of a posted image by incorporating various types of visual and social features into a unified network model [6]. The proposed work finds the solution for raises the problem of predicting image popularity on social media. The popularity of an image can be affected by several factors, such as visual content, aesthetic quality, user, post metadata, and time. VSCNN first learns to extract high-level representations from the input visual and social features by utilizing two individual CNNs. The outputs of these two networks are then fused into a joint network to estimate the popularity score in the output layer. The simulation results demonstrate that the proposed VSCNN model significantly outperforms state-of-the-art models.

Mohbey, K. K. (2020), studied user behavior from the unstructured tweets shared on social media is an interesting yet challenging task [7]. A social platform such as Twitter yield access to the unprompted views of the wide-ranging users on particular events like election. These views cater government and corporate to remold strategies, assess the areas where better measures need to be put forward and monitor common opinion. The authors introduced Multi-class classification fabricated with novel deep learning approach to analyses the user opinion. Moreover, comparative analysis between tradition approaches such as Naïve Bayes, SVM, decision tree, logistic regression and employed approach with deep learning method is presented. Experimental results revel that the proposed model can reach up to highest accuracy.

SM, H. K., et al., (2021), described a system that was built to analyze data from social media platforms such as Twitter, Instagram, Facebook and Reddit [8]. The explosion of social media data which holds important, massive, rapidly evolving unstructured information has provided an opportunity to research public opinion and know people's

feelings. The data is gathered from various services and programs which involves using the official API's and scraping tools. Machine Learning & Image Processing algorithms are applied on the gathered data to gain insights about the individual users' interests, usage patterns, media reachability. These models are then relayed to a web application where various visualizations of the insights can be obtained.

Kour, H., & Gupta, M. K. (2022), proposed a model called a hybrid of two deep learning architectures, Convolutional Neural Network (CNN) and bi-directional Long Short-Term Memory (biLSTM) [9]. The key objective of our study is to explore the possibility of predicting a user's mental condition by classifying the depressive from non-depressive ones using Twitter data. Using textual content of the user's tweet, semantic context in the textual narratives is analyzed by utilizing deep learning models. CNN-biLSTM model is compared with Recurrent Neural Network (RNN) and CNN model and also with the baseline approaches. Experimental results based on various performance metrics indicate that our model helps to improve predictive performance. To examine the problem more deeply, statistical techniques and visualization approaches were used to show the profound difference between the linguistic representation of depressive and non-depressive content.

Xiao, Y., et al., (2020), proposed a user participant behavior prediction model of social hotspots based on a multimessage interaction-driving mechanism and the BP neural network [10]. Finally, the multimessage interaction has an iterative guiding effect on user behavior, which easily causes overfitting of the BP neural network. To avoid this problem, the traditional BP neural network is optimized by a simulated annealing algorithm to further improve the prediction accuracy. The authors mainly focus on a prediction model of user participation behavior during multiple messaging of hot social topics.

3. STATEMENT OF THE PROBLEM

Almost 70% of adults use Facebook. Running such a large site for many users is expensive. Yet, users pay no fee for using Facebook. Google is like Facebook in that it also provides free services to users. The services are free because platforms generate revenue from advertising. Google generates more than 70% of its revenue from advertising. Facebook generates more than 80% of its revenue from advertising. Advertising is not without difficulties [12]. Most people do not like advertisements. In fact, people generally feel annoyed by advertisements. Users will even avoid advertisements. This results in users using social media less. Fewer users mean that advertisements receive less engagement. Lower user engagement means the platform makes less money. Fewer users also mean that platforms are unable to reach as many people. In this way, advertisements can hurt the platform and advertisers. The challenge is to deliver relevant content to users. The content might agree with user interests, hobbies, and preferences. If advertisers can deliver relevant content, then advertisers, users, and the platform benefit. The users buy content. The advertiser's sell merchandise and the platform makes money. Yet, there are millions of users on social media and many advertisements. This research work aims to explore this problem by predicting user engagement. This helps companies better understand how their advertisement will perform on social media. It also gives advertisers feedback [13]. More confident advertisers will invest more on the platform.

The proposed work performs its analysis on the Facebook post with image and text data from a Facebook post as input. The research performs computations on the data and predicts user interaction. The aim of the proposed work is to predict user interaction. It will explore this prediction with different deep learning models. The analysis used the Facebook post since it contains image and text data. Its data is also available via Facebook's Graph API. The study collected and trained on exactly 350k posts. The Facebook post is generic. Other platforms have similar representations of the Facebook post. These include tweets and Pinterest posts. The posts in this study are from advertisers. This limitation might simplify the models. Previous studies have modeled user interaction on social media. The proposed work refers to post metrics as post-performance. Facebook refers to the metrics as user interactions, or engagements (Facebook API). The Facebook post may include text, an image, and meta-data. The meta-data includes shares, comments, and data from the post's Facebook page. The metrics denoting user interaction include the number of shares and the comments. The scope of the research will consist of predicting these metrics. The research creates many types of machine learning models. The research will uncover which can best predict user interaction. One goal is to model user interaction with both text and image data. The study will gauge how well models can predict user interaction.

4. SCOPE OF THE RESEARCH WORK

The main purpose of this study is to analyze the interaction of Facebook users, which is the most popular social networking platform with millions of users. This study presents a new approach to classify users on the Facebook social network as active, passive and malicious [14, 15]. An up-to-date and comprehensive data set was created, and the users were classified with the help of deep learning algorithms. In this work, the proposed method was summarized by a systematic methodology. To classify Facebook users as malicious, active or passive, 10 different metrics were used: density, centrality, reciprocity, diameter, modularity, account age, follow-up, and message routing, liking and tweeting averages. The profile of a total of 4,200 social media users was based on their current interaction data covering eight months. Data cleaning, integration, size reduction and optimization processes were performed on the data set. Metric data for all users were imaged by applying visualization methods to the data set after the filtering steps were completed. By applying deep learning algorithms to the data set, the classification performances of the active, passive and malicious categories were examined.

The main contributions of this work are summarized below: -

- The make-up of static and dynamic interaction of users in the network was extracted effectively by using fewer metrics. This is important for cost–performance considerations.
- For the first time, behavioral analysis was achieved in this study by simultaneously using 10 different dynamic interaction metrics, such as the number of posts and account age.
- Deep learning algorithms were applied to the interaction data of the users. - In this study, the quality of social network accounts was rated – in addition to malicious account detection and behavioral analysis – based on user interaction.

5. PROPOSED METHODOLOGY

The study explores three types of models. These are text-based, image-based, and those that combine text and image data. The scope of the study limits both its input and output data. Its input data includes the post's text and an image if present. The output is user interaction. The user interactions considered include share count, comment count, and comment sentiment [16]. These outputs were the available user interactions from the Facebook API. These inputs are ubiquitous on Facebook and common on other social media platforms. There was enough Facebook data to train all but one type of the machine learning models. Even more, each post always has some share count, comment count, and comment sentiment. The study has interest in the benefits and effects of combining image and text data. The study will also determine if the model is useful in practice. The study will compare the model's performance with a random model.

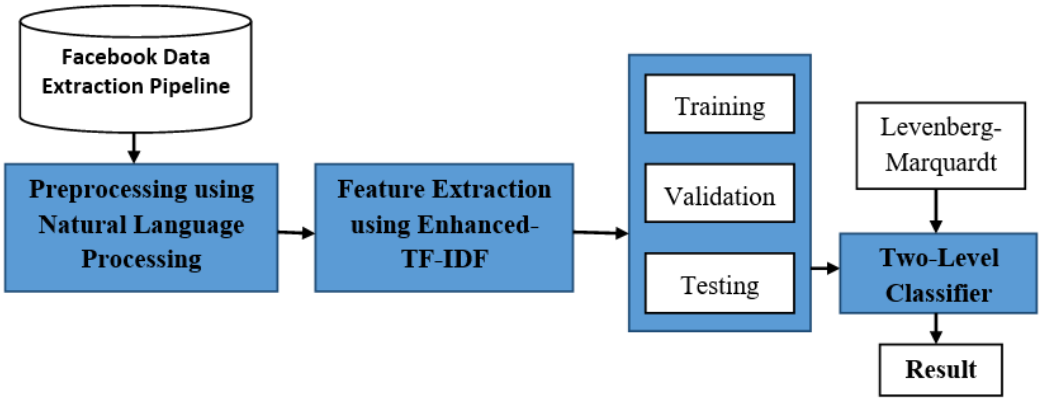

Fig. 4.1: Workflow of Proposed Methodology

Phase 1: - Preprocessing of Facebook Dataset

After getting all necessary parameters from the Facebook, it is done some preprocessing operations on the data to make them compatible for the sentiment analysis. In most of the research papers, they were also used to some extent same operations for the pre-processing. A little number of researchers used stemming, stop word removal and spell

checking. A growing number of them used stemming, stop word removal, text indexing, dimensionality reduction and term weighting. Only these methods cannot be used for the pre-processing for sentiment analysis. For example, if we do stop word removal, the accuracy of sentiment analysis will reduced. Few number of researchers used tokenizing. We cannot tokenize the sentences in order to do.

Phase 2: - Feature Extraction using Enhanced TF-IDF

TF-IDF means Term frequency-inverse document frequency. It is a weighting scheme frequently used in text mining. It is statistical measure which helps to assign a weighting scheme to a word. The tf-idf scheme works better than the basic BOW model, by finding how significant a word is to a document by giving weightage to words in the document. A new in-class parameter has been introduced to overcome the drawbacks of TF-IDF. Usually TF-IDF will calculate weightage of a word to the whole document only. To improve this, we have introduced a new parameter which calculates the term weightage for the document and also the term weightage inside the class. We renamed this as TF-IDF-NCF, where NCF means new class function.

Phase 3: - Classification using Two-level Classifier

The classification task in this study involves detecting event from news posts into one of their corresponding category (conflict, natural disaster, disease, terrorism, and airplane crash) using different classification techniques.

Convolutional Neural Networks (CNN) is commonly used in image classification problems. In CNN, layers are arranged in three dimensions: width, height and depth. Additionally, neurons in one layer are connected to only a small portion of neurons in the next layer, not to all neurons. Finally, the final output is reduced to a single vector of probability points arranged across the depth dimension. In Convolutional Neural Networks, attributes are extracted and reproduced in the convolution layer. There are layers such as convolution, pooling and full connectivity in this architecture.

6. PERFORMANCE EVALUATION METRICS

The executions of the three methods were assessed using the predictive classification table, known as Confusion Matrix by calculating the standard metrics of accuracy, precision, sensitivity, specificity and F-measure [17].

To distinguish the instances of different classes the confusion matrix acts as a valuable tool for evaluating the algorithms. It reveals the amount of correct and wrong predictions prepared by the model compared with the actual categorizations in the dataset. True Positive (TP) and True Negative (TN) are helpful to identify, when the algorithm is generating the actual data. The parameter for the evaluation measure such as False Positive (FP) and False Negative (FN) are used to know, when the classifier is producing the faulty information [18].

Accuracy:

It is the proportion between the quantity of right predictions and complete number of predications.

acc**=**$TP+TNTP+TN+FP+FN$**acc=TP+TNTP+TN+FP+FN**

Precision:

It is the proportion between the quantity of right positives and the quantity of true positives in addition to the quantity of false positives.

(p**=**$TPTP+FP$**Precision(p)=TPTP+FP**

Recall:

It is the proportion between the quantity of right positives and the quantity of true positives in addition to the quantity of false negatives.

$recall$**=**$TPTP+FN$**recall=TPTP+FN**

F-score:

It is known as the consonant mean of precision and review.

acc**=112(1**p**+1**r**)=2**$prp+r$**acc=112(1p+1r)=2prp+r**

7. CONCLUSION

Recently, deriving an effective computational model to characterize human behavior or predict decision making has become an emergent topic. In this study, a multimodal deep learning framework is developed for predicting the popularity of images on social media. First, it analyzed and extracted different types of image visual content features and social context information that significantly affect image popularity. Then, proposed a novel two level classifier visual-social computational model for image popularity prediction. This model uses individual networks to process input data with different modalities, and the outputs from these networks are then integrated into a fusion network to learn joint multimodal features and estimate the popularity score.

8. REFERENCES

[1] Do, T. H., Nguyen, D. M., Tsiligianni, E., Cornelis, B., & Deligiannis, N. (2017). Multiview deep learning for predicting twitter users' location. arXiv preprint arXiv:1712.08091.

[2] Phan, N., Dou, D., Piniewski, B., & Kil, D. (2016). A deep learning approach for human behavior prediction with explanations in health social networks: social restricted Boltzmann machine (SRBM+). Social network analysis and mining, 6, 1-14.

[3] Luceri, L., Braun, T., & Giordano, S. (2019). Analyzing and inferring human real-life behavior through online social networks with social influence deep learning. Applied network science, 4(1), 1-25.

[4] Al-Qurishi, M., Alrubaian, M., Rahman, S. M. M., Alamri, A., & Hassan, M. M. (2018). A prediction system of Sybil attack in social network using deep-regression model. Future Generation Computer Systems, 87, 743-753.

[5] Kamalesh, M. D., & Bharathi, B. (2022). Personality prediction model for social media using machine learning Technique. Computers and Electrical Engineering, 100, 107852.

[6] Abousaleh, F. S., Cheng, W. H., Yu, N. H., & Tsao, Y. (2020). Multimodal deep learning framework for image popularity prediction on social media. IEEE Transactions on Cognitive and Developmental Systems, 13(3), 679-692.

[7] Mohbey, K. K. (2020). Multi-class approach for user behavior prediction using deep learning framework on twitter election dataset. Journal of data, Information and management, 2(1), 1-14.

[8] SM, H. K., Hegde, S., Santosh, G., Shivakumar, M., & Srihari, R. (2021, April). User Interest Prediction based on Social Network Profile with Machine Learning. In 2021 6th International Conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE.

[9] Kour, H., & Gupta, M. K. (2022). An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM. Multimedia Tools and Applications, 81(17), 23649-23685.

[10] Xiao, Y., Li, J., Zhu, Y., & Li, Q. (2020). User behavior prediction of social hotspots based on multimessage interaction and neural network. IEEE Transactions on Computational Social Systems, 7(2), 536-545.

[11] Xu, K., Lin, Z., Zhao, J., Shi, P., Deng, W., & Wang, H. (2020, October). Multimodal deep learning for social media popularity prediction with attention mechanism. In Proceedings of the 28th ACM International Conference on Multimedia (pp. 4580-4584).

[12] Xue, D., Wu, L., Hong, Z., Guo, S., Gao, L., Wu, Z., ... & Sun, J. (2018). Deep learning-based personality recognition from text posts of online social networks. Applied Intelligence, 48(11), 4232-4246.

[13] Sun, J., Kunegis, J., & Staab, S. (2016, December). Predicting user roles in social networks using transfer learning with feature transformation. In 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW) (pp. 128-135). IEEE.

[14] Guimaraes, R. G., Rosa, R. L., De Gaetano, D., Rodriguez, D. Z., & Bressan, G. (2017). Age groups classification in social network using deep learning. IEEE Access, 5, 10805-10816.

[15] Lima, A. C. E., & De Castro, L. N. (2014). A multi-label, semi-supervised classification approach applied to personality prediction in social media. Neural Networks, 58, 122-130.

[16] Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. Journal of Ambient Intelligence and Humanized Computing, 1-24.

[17] Chaudhary, K., Alam, M., Al-Rakhami, M. S., & Gumaei, A. (2021). Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics. Journal of Big Data, 8(1), 73.

[18] Al-Garadi, M. A., Hussain, M. R., Khan, N., Murtaza, G., Nweke, H. F., Ali, I., ... & Gani, A. (2019). Predicting cyberbullying on social media in the big data era using machine learning algorithms: review of literature and open challenges. IEEE Access, 7, 70701-70718.

DOI: https://doi.org/10.15379/ijmst.v10i3.3719