FAKE INSTAGRAM PROFILE DETECTION USING FEEDFORWARD NEURAL NETWORK

Dr. Kalaivani J
Department of Computing Technologies
SRM Institute of Science and Technology
Kattankulathur, India
kalaivaj@srmist.edu.in

Sayan Kumar Bag
Department of Computing Technologies
SRM Institute of Science and Technology
Kattankulathur, India
sb0708@srmist.edu.in

Shivansh Singh
Department of Computing Technologies
SRM Institute of Science and Technology
Kattankulathur, India
ss8177@srmist.edu.in

***Abstract*—The project aimed to develop a robust fake account detection system for social media platforms, particularly Instagram, utilising deep learning techniques. Leveraging a dataset consisting of various features such as profile picture presence, username characteristics, and other relevant attributes, the model was trained to discern between genuine and fake accounts. The dataset underwent thorough exploratory data analysis, including visualisations to gain insights into feature distributions and correlations. The preprocessing phase involved standardisation of input data and one-hot encoding of the target variable. A deep neural network architecture was designed and trained using TensorFlow and Keras, encompassing multiple layers with dropout regularisation to enhance generalisation. The model demonstrated commendable performance, achieving an accuracy of 88% on a test dataset, as evidenced by the detailed classification report. The training progression was visually assessed through loss and accuracy plots, providing a detailed understanding of the model's learning dynamics. The resulting model showcased promising capabilities in identifying fake profiles, with precision, recall, and F1-score metrics supporting its efficacy. The abstract encapsulates the project's scope, methodology, and outcomes, highlighting the significance of employing deep learning in combating the proliferation of fake accounts on social media platforms.**

***Keywords*—Fake Account Detection, Social Media Platforms, Instagram, Deep Learning Techniques, Exploratory Data Analysis, Neural Network Architecture, Model Performance.**

I. INTRODUCTION

Fake accounts are a major problem in a time when social media is everywhere, so finding creative ways to identify and lessen their impact is imperative. This project endeavours to address this critical issue by developing a robust Fake Account Detection System, with a specific focus on social media platforms, particularly Instagram. Leveraging the power of deep learning techniques, the project seeks to create an advanced model capable of discerning between genuine and fraudulent profiles.

The foundation of this endeavour lies in a comprehensive dataset, enriched with diverse features such as the presence of profile pictures, username characteristics, and other relevant attributes. Through thorough exploratory data analysis, including visualisations that unveil feature distributions and correlations, the project gains invaluable insights into the intricacies of fake account patterns.

The preprocessing phase involves standardising input data and employing one-hot encoding for the target variable, preparing the dataset for the subsequent training process. A meticulously designed deep neural network architecture, using TensorFlow and Keras, serves as the core engine for this project. This architecture incorporates multiple layers with dropout regularisation, enhancing the model's power to generalise and effectively identify fraudulent profiles.

The project culminates in a model that demonstrates commendable performance, accomplishing an accuracy rate of 88% on a dedicated test dataset. The training progression is meticulously tracked through loss and accuracy plots, providing a detailed understanding of the model's learning dynamics. The resulting system showcases promising capabilities in identifying fake profiles, supported by precision, recall, and F1-score metrics. This project's abstract encapsulates the scope, methodology, and outcomes, underscoring the significance of employing cutting-edge deep learning techniques in the ongoing battle against the proliferation of fake accounts on the internet.

II. RELATED WORK

In recent studies focusing on fake profile detection across various social media platforms, researchers have introduced novel approaches and methodologies to address the growing concern of fraudulent activities.

Sarah Khaled et al.. [1] proposed a unique SVM-NN approach for detecting fake profiles on Twitter, leveraging features derived from the MIB dataset. Their model outperformed existing methods by incorporating SVM-trained decision values into neural network models, resulting in higher accuracy. Ala M. Al-Zoubi et al. [2] concentrated on Twitter spam profile detection, identifying ten features, including suspicious words and tweet time patterns. Utilising models like Naive Bayes, Decision Trees, and Neural Networks, they achieved an impressive 95.7% accuracy with Naive Bayes.

Preethi Harris et al. [3] explored the Kaggle Instagram dataset, achieving 100% accuracy with XGBoost and Random Forest, showcasing their effectiveness in identifying fake profiles on Instagram.. In order to classify a mixed real and fraudulent Twitter dataset, Gayathri A. et al. [4] used Support Vector Machine, Random Forest, and Deep Neural Networks.

Jyoti Kaubiyal et al. [5] utilised real Twitter data gathered through the Twitter API and applied SVM, Logistic Regression, and Random Forest for classification, achieving high accuracy in distinguishing between fake and real accounts. Aditi Gupta et al. [6] focused on Facebook activity analysis to discover fake profiles, identifying 17 characteristics of user behaviour. For categorization, decision trees showed the best accuracy.

LinkedIn was investigated by S. Adikari and K. Dutta [7] as a social networking platform for fraud detection. They attained an 87% accuracy rate by using Principal Component Analysis, Weighted Average, and Neural Network Support Vector Machine data mining approaches. Raturi Rohit [8] introduced a machine learning framework for finding fake accounts on Facebook and Twitter, considering user posts and status.

Naman Singh et al. [9] proposed methods to detect and remove fake profiles on online networking platforms, considering factors like the number of followers. Lastly, Rao et al. [10] presented an NLP system for fake profile detection on Facebook, employing SVM classifiers and Naïve Bayes algorithms to enhance accuracy. These diverse studies collectively contribute to the evolving landscape of fake profile detection across multiple social media platforms.

III. PROPOSED METHODOLOGY

The proposed methodology presented in this paper is depicted in Figure 1.

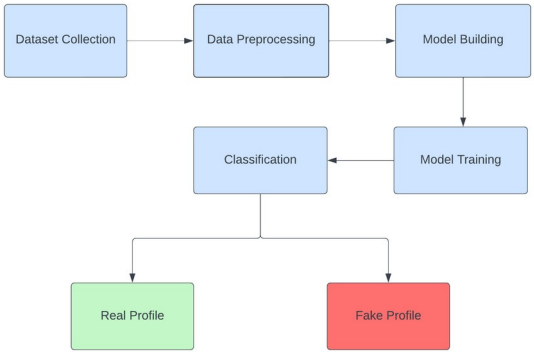

Fig. 1. Fake Profile Detection Methodology

A. Data Understanding and Exploration

In the initial step, a comprehensive analysis of the dataset is essential to comprehend its underlying structure and features. Statistical summaries and exploratory data visualisations, such as distribution plots and correlation matrices, provide valuable insights. Additionally, the identification and handling of missing values are crucial to ensure the dataset's integrity and reliability.

B. Data Preprocessing

The data is divided into training and testing sets in order to prepare it for model training. Features and target labels are extracted, and numerical features are standardised using the StandardScaler. Target labels are transformed into a categorical format through one-hot encoding, ensuring compatibility with the deep learning model's requirements.

C. Data Visualization

The visualisation step involves gaining deeper insights into the data through graphical representations. Histograms, heatmaps, and correlation plots are employed to visually analyse feature distributions and identify potential patterns related to fake account detection. Visualisations play an important role in formulating hypotheses and understanding the interplay of various features.

D. Model Building

Using TensorFlow and Keras to create a solid deep learning model is the methodology's central

component. A Sequential model is constructed with multiple dense layers, incorporating activation functions like ReLU and softmax for classification. Dropout layers are strategically placed to mitigate overfitting, contributing to the model's generalisation capability. The choice of a suitable loss function (categorical_crossentropy) and optimizer (Adam) is imperative for effective training.

E. Model Training

The model is trained on the preprocessed training dataset, a crucial step in the development process. Careful consideration is given to the number of epochs, and the training progress is monitored in terms of both training and validation performance. Continuous evaluation of accuracy, loss, and validation metrics ensures the model's ability to learn from the provided data.

F. Model Evaluation

On the testing dataset, the trained model is used to predict labels. Creating a comprehensive classification report and confusion matrix is part of the evaluation process. We analyse precision, recall, and F1-score to determine how well the model detects phoney accounts. This step functions as a critical evaluation of the model's functionality.

G. *Performance Analysis*

Visualising the model's progression during training is pivotal. Graphs depicting training and validation loss, as well as accuracy, offer a comprehensive view of the model's learning curve. Calculating average accuracy, validation accuracy, loss, and validation loss provides quantifiable metrics for assessing overall performance.

IV. RESULTS AND DISCUSSIONS

With the help of multiple variables, including the length of the username, the presence of a profile image, and other elements, the deep learning model was created and trained to identify phoney Instagram profiles. The model's overall accuracy of 88% shows that it can distinguish between real and fraudulent profiles.

A. Correlation Plot

To understand the relationships between different features, a correlation plot was generated. The plot visually represents the correlation coefficients between various features in the dataset. Strong correlations (either positive or negative) may indicate important relationships that contribute to the model's decision-making process. Understanding these correlations can provide insights into the key features influencing the model's predictions.

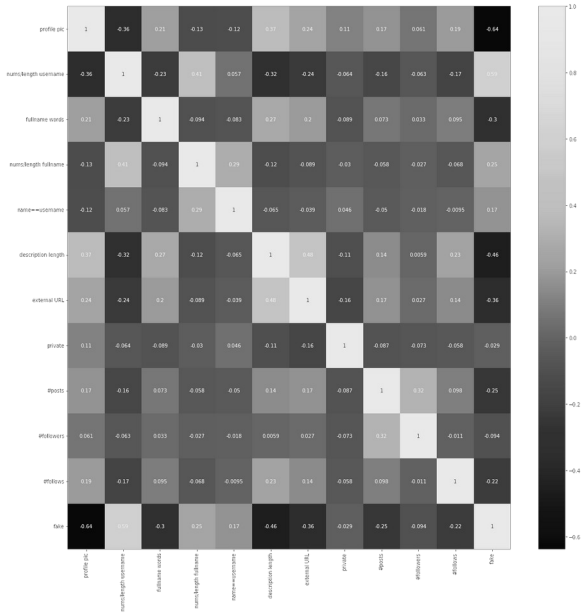

Fig. 2. Correlation Plot of the Dataset

B. Model Architecture and Training

The neural network design consisted of several dense layers with minimization of overfitting using dropout layers and rectified linear unit (ReLU) activation functions. Training of the model involved using the Adam optimizer using category cross entropy as the loss function.. The model proceeded through 25 epochs of training, and both the accuracy and loss of the validation and training phases were tracked.

The training and validation accuracy graphs resulted in the progression of the model's performance over the epochs. Both training and validation accuracies steadily improved, indicating that the model was learning and generalising well from the training data. The validation loss remained consistently lower than the training loss, demonstrating effective generalisation and a lack of overfitting.

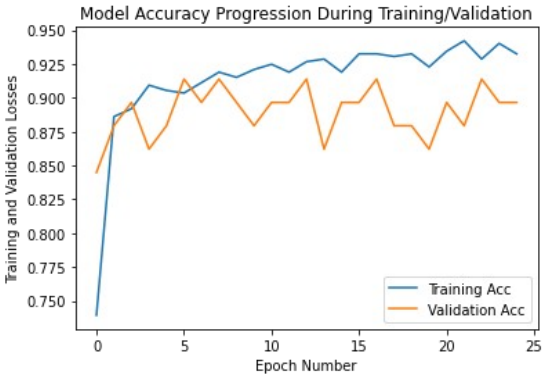

Fig. 3. Training and Validation Accuracy

C. *Classification Report*

The classification report provides detailed metrics on the model's performance for each class (fake and genuine profiles). The precision, recall, and F1-score for both classes were high, with values ranging from 0.87 to 0.90. These metrics suggest a balanced and effective classification performance for both fake and genuine profiles.

	precision	recall	f1-score	support
0	0.87	0.90	0.89	60
1	0.90	0.87	0.88	60
accuracy			0.88	120
macro avg	0.88	0.88	0.88	120
weighted avg	0.88	0.88	0.88	120

Fig. 4. Classification Report

D. *Confusion Matrix*

A visual depiction of the model's performance is given by the confusion matrix, which displays the number of true positives, true negatives, false positives, and false negatives. In this instance, the model performed in a balanced manner, misclassifying the same number of profiles as real and fraudulent. The matrix aids in locating potential areas for model enhancement or adjustment.

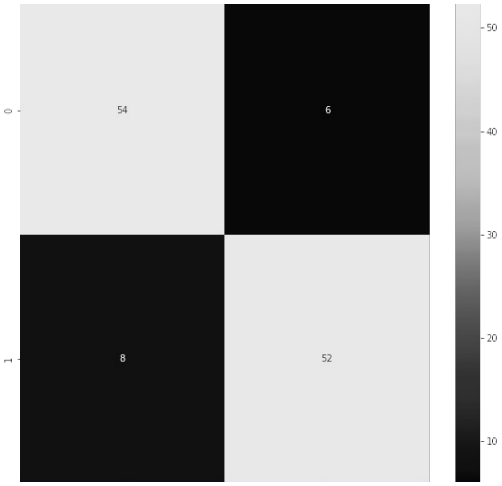

Fig. 5. Confusion Matrix

E. *Overall Assessment*

The model demonstrated a commendable accuracy of 88%, suggesting its effectiveness in distinguishing between fake and genuine Instagram profiles. The precision and recall values indicate a well-balanced performance for both classes. However, further analysis and fine-tuning could be performed to enhance the model's accuracy and generalisation on real-world data.

V. CONCLUSION

In conclusion, the developed deep learning model for Instagram fake profile detection exhibits promising results, achieving an accuracy of 88% with robust precision, recall, and F1-score metrics for both fake and genuine profiles. The model's architecture, incorporating multiple dense layers and dropout mechanisms, contributed to effective learning and generalisation, as evidenced by the training and validation accuracy graphs. The balanced performance showcased in the confusion matrix underscores the model's ability to make informed decisions across both classes.

Even with the model's success, it needs to be updated and monitored often to account for social media platforms' ever-changing nature. Updates and retraining on a regular basis will guarantee that the model continues to respond to changing trends in the building of phoney Instagram profiles. The dependability and quality of the training data determine the model's accuracy and reliability, highlighting the importance of careful data curation.

As a tool for safeguarding online communities, this model presents a valuable contribution to the ongoing efforts in combating fraudulent activities on social media. Its effectiveness in distinguishing

fake profiles signifies its potential for aiding platform administrators and users in maintaining a trustworthy digital environment. In the future, greater investigation and improvement can strengthen the model's resilience and practicality, resulting in a more resilient defence against dishonest activities in the domain of online social networks.

REFERENCES

[1] S. Khaled, N. El-Tazi and H. M. O. Mokhtar, "Detecting Fake Accounts on Social Media," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 3672-3681, doi: 10.1109/BigData.2018.8621913.

[2] A. M. Al-Zoubi, J. Alqatawna and H. Paris, "Spam profile detection in social networks based on public features," 2017 8th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2017, pp. 130-135, doi: 10.1109/IACS.2017.7921959.

[3] P. Harris, J. Gojal, R. Chitra and S. Anithra, "Fake Instagram Profile Identification and Classification using Machine Learning," 2021 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2021, pp. 1-5, doi: 10.1109/GCAT52182.2021.9587858.

[4] Gayathri, A., S. Radhika, and S. L. Jayalakshmi. "Detecting fake accounts in media application using machine learning." *International Journal of Advanced Networking and Applications* (2019): 234-237.

[5] Kaubiyal, Jyoti, and Ankit Kumar Jain. "A feature based approach to detect fake profiles in Twitter." *Proceedings of the 3rd international conference on big data and internet of things*. 2019.

[6] Aditi Gupta and Rishabh Kaushal, "Towards Detecting Fake User Accounts in Facebook", IEEE International Conference on Asia Security and Privacy(ISEASP), 2017.

[7] Adikari, Subhashie & Dutta, K.. (2014). Identifying fake profiles in linkedin. Proceedings - Pacific Asia Conference on Information Systems, PACIS 2014.

[8] Raturi, Rohit. (2018). Machine Learning Implementation for Identifying Fake Accounts in Social Network.

[9] N. Singh, T. Sharma, A. Thakral and T. Choudhury, "Detection of Fake Profile in Online Social Networks Using Machine Learning", 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), pp. 231-234, 2018.

[10] P. Srinivas Rao, Jayadev Gyani and G. Narsimha, "Fake Profiles Identification in Online Social Networks Using Machine Learning and NLP", International Journal of Applied Engineering Research ISSN 0973–4562, vol. 13, no. 6, pp. 4133-4136, 2018.