IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, November 2024

Fake Social Media Detection Using Machine Learning Techniques

Dhanashree R. More[1], Dhanshree G. Gavali[2], Sahil S. Gaikawad[3], Payal R. Pawshe[4]
Students, Computer Science & Engineering[1,2,3,4]
MIT ADT (Arts Design & Technology), Loni Kalbhor, Pune, India

Abstract: *The proliferation of fake accounts on social media platforms, particularly Instagram, has raised concerns regarding misinformation, fraud, and compromised user trust. This study explores the efficacy of machine learning techniques for detecting fake accounts by leveraging two prominent datasets—InstaFake and IJECE. Using preprocessing techniques like SMOTE for class balancing and models such as Random Forest, Decision Trees, and Neural Networks, we aim to identify attributes that distinguish real from fake accounts. The Random Forest model emerged as the best performer with an F1-score of 0.95, demonstrating its robustness in identifying fraudulent accounts. This paper discusses the methodologies, results, and opportunities for enhancing social media platform integrity through advanced detection mechanisms.*

Keywords: Fraud Detection, Machine Learning, Random Forest, Class Imbalance, Social Media Security, Data Preprocessing

I. INTRODUCTION

Social media platforms like Instagram have revolutionized global communication, fostering creativity and collaboration. However, their growth has also seen a surge in fake accounts used for spamming, phishing, and creating fake engagement. These malicious accounts exploit platform features for personal or financial gain, often deceiving users and businesses alike. Traditional detection methods, including manual moderation and rule-based algorithms, struggle to keep up with the sophisticated tactics employed by fraudsters. This has led to the adoption of machine learning (ML) techniques capable of analyzing large datasets and identifying subtle patterns indicative of fake behavior. This paper presents a comprehensive approach to fake account detection, employing two datasets
InstaFake and IJECE. By comparing various ML algorithms and evaluating their performance on metrics like precision, recall, and F1-score, we aim to identify the most effective model for this task.

II. METHODOLOGY

The research methodology comprises the following steps:

Dataset Description

- **Sources:**
  Datasets such as *InstaFake* and *IJECE* were used, containing labeled examples of fake and authentic accounts. These datasets were chosen to ensure diversity in user behavior, profile metadata, and content characteristics.
- **Objective:**
  To gather a comprehensive dataset representing real-world scenarios of fake account activities.

Data Preprocessing

- Missing Value Handling: Imputation techniques (mean and mode imputation) were applied to handle missing data.
- Train-Test Split: The data was divided into training (80%) and testing (20%) sets to evaluate model performance.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, November 2024

Model Implementation

Three primary machine learning models were implemented and tested:

- Decision Tree: A simple baseline model for initial testing.
- Random Forest: An ensemble method that combines multiple decision trees to improve accuracy.
- Neural Networks: Constructed using Keras with two hidden layers and ReLU activation.

Evaluation Metrics

- Accuracy: The proportion of correctly classified transactions.
- Precision: Measures the accuracy of positive predictions.
- Recall: Measures how many actual fraudulent transactions are correctly identified.
- F1-Score: Balances precision and recall for an overall performance measure.

III. IMPLEMENTATION AND RESULTS

The models were evaluated on the hold-out test set to measure their effectiveness in detecting fraudulent transactions. Each model was trained using various hyperparameters, and extensive cross-validation was performed to ensure the results were not due to random chance. The key results for each model are summarized below:

Decision Tree

The Decision Tree model was used as a baseline to establish the fundamental structure of the data and to identify key features that differentiate fraudulent transactions from legitimate ones. Although the Decision Tree achieved an accuracy of 89%, it exhibited a tendency to overfit on the training data. This was evident from its significantly lower precision and recall scores on the test set, making it unsuitable for a highly imbalanced dataset like this. The model struggled with generalization due to its nature of creating overly specific rules for each class.

- Accuracy: 89%
- Precision: 0.81
- Recall: 0.73
- F1-Score: 0.77

Random Forest

The Random Forest model, an ensemble of multiple decision trees, performed significantly better. By aggregating the results of multiple trees, the model reduced overfitting and captured complex patterns in the data. It achieved the highest overall performance with an F1-score of 0.95 and an ROC-AUC score of 0.98. The high ROC-AUC indicates that the model was able to differentiate between fraudulent and non-fraudulent transactions effectively.
Additionally, implementing SMOTE (Synthetic Minority Over-sampling Technique) for balancing the dataset improved the recall score by 15%, ensuring more fraudulent transactions were correctly identified. This is crucial in fraud detection, where missing even a single fraudulent transaction could result in significant financial loss.
Accuracy: 97%
Precision: 0.94
Recall: 0.96
F1-Score: 0.95

Logistic Regression

Logistic Regression (LR) was employed in the project as a baseline model for detecting fake social media accounts. As a widely-used statistical and machine learning method, LR is effective for binary classification tasks and provides interpretable results, making it a reliable choice for initial model evaluations

- Accuracy: 80.94%
- Precision: 81.0%

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, November 2024

- Recall: 80.9%
- F1-Score: 80.9%

Comparison and Analysis

Model	Accuracy	Precision	Recall	F1-Score	Remarks
Random Forest (RF)	90.09%	90.7%	90.1%	90.1%	Achieved the highest overall performance due to its ensemble nature and robustness.
Logistic Regression (LR)	80.94%	81.0%	80.9%	80.9%	Reliable for linear relationships, but struggled with complex feature interactions.
Multilayer Perceptron (MP)	81.73%	81.8%	81.7%	81.7%	Effective in capturing complex patterns but required significant computational resources.
Naive Bayes (Gaussian)	73.12%	75.9%	73.1%	72.4%	Performed well on probabilistic relationships but less effective with diverse features.

Visual Analysis of Model Performance

To better understand the model's performance, confusion matrices and ROC curves were plotted for each model:

Confusion Matrix: Showed the distribution of true positives, false positives, true negatives, and false negatives. The Random Forest model showed a high number of true positives, indicating its effectiveness.

ROC Curve: Plotted for each model to visualize the trade-off between the true positive rate and false positive rate. The Random Forest model had the highest area under the curve, signifying its strong performance.

These visual insights helped pinpoint areas of improvement and guided future enhancements for handling edge cases

IV. CONCLUSION

In conclusion, The Fake Social Media Detection System developed in this project successfully demonstrates the potential of machine learning techniques to identify fraudulent accounts across social media platforms. By leveraging algorithms like Random Forest and Logistic Regression, along with carefully engineered features such as user activity, profile metadata, and content analysis, the system was able to achieve high accuracy in distinguishing between fake and real accounts. Overall, this project contributes to the development of an automated and efficient system capable of improving the security and integrity of social media platforms. By identifying fake accounts early, the system helps mitigate risks

REFERENCES

[1] **Ahmed, M., & Mahmood, A. N. (2018).** "A survey of fake account detection techniques in social media platforms." *International Journal of Computer Applications*, 180(4), 45-51.

[2] Boshmaf, Y., et al. (2011). "A survey of social network-based fake account detection techniques." *Proceedings of the 2011 International Conference on Computational Social Networks*, 123-134.

[3] Chia, K. M., & Wong, L. L. (2020). "Fake account detection in social media using machine learning algorithms." *Journal of Artificial Intelligence and Soft Computing Research*, 10(2), 121-131.

[4] Feng, Y., & Jiang, X. (2017). "Fake profile detection in social media: A machine learning perspective." *Social Network Analysis and Mining*, 7(1), 50.

[5] Liu, H., & Zhai, Y. (2019). "Fake account detection on social media using feature fusion and machine learning." *Journal of Computational Science*, 30, 56-65