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 3, October 2024

Leveraging Artificial Intelligence and Machine Learning for Detecting Fake Accounts on Social Media

Prof. Snehal Mangale[1], Devika Naik[2], Pratham Matte[3], Shubham Kharade[4], Vaibhavi Mane[5]
Assistant Professor, Department of Computer Engineering[1]
Students, Department of Computer Engineering[2,3,4,5]
Dr. D. Y Patil College of Engineering and Innovation, Talegaon, Pune, India

Abstract: *Detecting fake accounts is a significant challenge for social media platforms, e-commerce sites, and online services. Malicious users create fake profiles to spread misinformation, commit fraud, or manipulate system behavior. Traditional rule based methods often fail to catch sophisticated fake accounts that resemble genuine user activities. This paper examines the application of Artificial Intelligence (AI) and Machine Learning (ML) techniques to improve the detection of these deceptive accounts. Using KNN model and Gradient Boosting algorithm offer a more robust solution to identifying and combating fake accounts in Social Media Platform.*

Keywords: Fake account, social Media, AI and ML algorithms, KNN model , Gradient Boosting Algorithm

I. INTRODUCTION

With the explosive growth of online social networks, the issue of fake accounts has become increasingly significant. These accounts are used to disseminate misinformation, engage in fraudulent activities, and manipulate public opinion. Addressing this challenge requires advanced machine learning (ML) and artificial intelligence (AI) techniques to identify and classify fake accounts in real-time. The surge in fake accounts negatively impacts platforms like Instagram, Facebook, and Twitter, necessitating thorough investigation. Given the sheer volume of these accounts, manual deactivation is impractical. However, technological advancements have enabled the implementation of ML in various applications to assist in such tasks. The use of AI and ML for fake account detection is gaining momentum, particularly on social media, e-commerce platforms, and online services where user authenticity is crucial. Malicious actors create fake accounts to spread false information, commit fraud, manipulate ratings or reviews, and engage in other harmful activities. This paper explores the implementation of the K-Nearest Neighbors (KNN) model for classifying fake accounts and the use of gradient boosting algorithms to enhance model performance. Additionally, we propose developing a Graphical User Interface (GUI) to facilitate user interaction and gather relevant features of problematic accounts. By analyzing behavioral patterns, user interactions, and other account features, these ML techniques can effectively identify fake accounts and help end-users take preventive actions, providing them with accuracy percentages for transparency and trust.

II. METHODOLOGY

1. Data Collection Module

- Functionality: Collect data from various social media platforms.

Details:

- Use APIs (e.g., Twitter API, Facebook Graph API) to gather user profiles, activity logs, and interaction patterns.
- Ensure data collection complies with privacy policies and regulations.
- Store collected data in a secure database for further processing.

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 3, October 2024

2. Data Preprocessing Module

- Functionality: Clean and preprocess the collected data.

Details:

- Handle missing values by imputing or removing incomplete records.
- Normalize data to ensure consistency (e.g., converting text to lowercase, removing special characters).
- Remove irrelevant information that does not contribute to the detection process.

3. Feature Extraction Module

- Functionality: Extract relevant features from the data.

Details:

- Posting Frequency: How often the user posts.
- Content Analysis: Sentiment analysis, keyword extraction, and topic modeling of posts.
- Network Behavior: Analysis of friends/followers, interaction patterns, and network centrality.
- User Interactions: Frequency and type of interactions (likes, comments, shares).

4. Model Development Module

- Functionality: Develop and train machine learning models.

Details:

- Implement various algorithms like Random Forest, Support Vector Machines, and Neural Networks.
- Split data into training and testing sets to evaluate model performance.
- Use cross-validation techniques to ensure model robustness.

5. Model Evaluation and Optimization Module

- Functionality: Evaluate and optimize the performance of the models.

Details:

- Use metrics such as accuracy, precision, recall, and F1- score to assess model performance.
- Perform hyperparameter tuning to optimize model parameters.
- Implement techniques to handle class imbalance, such as oversampling or undersampling.

6. Real-time Detection Module

- Functionality: Implement real-time detection of fake accounts. Details:
- Integrate the trained models into a real-time monitoring system.
- Continuously analyze incoming data and flag suspicious accounts.
- Implement alert mechanisms to notify administrators of potential fake accounts.

7. User Interface Module

- Functionality: Develop a user-friendly interface for the system. Details:
- Create a dashboard for users to input data and view detection results.
- Provide options to manage flagged accounts (e.g., review, approve, or delete).
- Ensure the interface is intuitive and accessible.

III. ALGORITHMS

1. KNN Classification Model:

a) Training the KNN Model

- Model Initialization: Initialize the KNN classifier with a chosen value of 'k', which represents the number of nearest neighbors to consider.

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 3, October 2024

- Model Training: Fit the KNN model to the training data, allowing it to learn the patterns and relationships in the data.

b) Making Predictions and Evaluation

- Predictions: Use the trained KNN model to make predictions on the test data.
- Evaluation: Assess the model's performance by calculating the accuracy, which measures the proportion of correctly classified instances.

c) Optimizing the Model

- Cross-Validation: Perform cross-validation to determine the optimal value of 'k'. This involves testing different values of 'k' and selecting the one that yields the highest accuracy.

d) Conclusion

- Effectiveness: KNN can effectively classify social media accounts as fake or real based on the selected features.
- Future Work: Consider exploring more advanced models like XGBoost or Gradient Boosting for potentially better performance and robustness.

2. Gradient Boosting Algorithm:

- Gradient Boosting is a popular boosting algorithm in machine learning used for classification and regression tasks.
- Boosting is one kind of ensemble learning method which trains the model sequentially and each new model tries to correct the previous model.
- It combines several weak learners into strong learners.

IV. FEATURE CONSIDERATIONS

For detecting fake social media accounts, we will consider features such as:

- **Profile Picture Presence:** Whether the account has a profile picture.
- **Follower Count:** Number of followers.
- **Following Count:** Number of accounts the user is following.
- **Post Count:** Number of posts or tweets.
- **Bio Description:** Length and content of the bio.
- **Engagement Metrics:** Likes, comments, retweets, etc.
- **Account Age:** How long the account has been active.

V. LITERATURE SURVEY

Yang et al. (2024): Analyzed Twitter accounts using AI generated faces for profile pictures, highlighting their use in spreading scams, spam, and amplifying coordinated messages. The authors presented a dataset of 1,353 such accounts and discussed the implications of these inauthentic activities.

Sruthi and Vasavi (2023): Explored machine learning and natural language processing (NLP) techniques for detecting fake accounts on various social networks. The authors emphasized the challenges of cyberbullying and misinformation, and proposed a framework for identifying fake profiles based on behavioral patterns and linguistic features.

Hammoodi and Obaid (2023): Investigated the use of machine learning and neural networks to detect fake accounts, focusing on the importance of social networking in daily life. The study evaluated the performance of various algorithms, including Random Forests and neural networks, in identifying fake profiles.

A Study of Different Methodologies (2022): This review paper discussed various machine learning methodologies for detecting fake accounts on social media. The authors compared the effectiveness and limitations of different algorithms, such as decision trees, support vector machines, and ensemble methods.

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 3, October 2024

Fake Profile Detection Using Machine Learning Techniques (2021): Utilized Twitter datasets to assess the authenticity of profiles using LSTM, XGBoost, Random Forest, and Neural Networks. The study highlighted the importance of feature selection and model tuning in improving detection accuracy.

Characteristics and Prevalence of Fake Social Media Profiles (2024): Systematically analyzed Twitter accounts with AI-generated faces. The authors presented a dataset of fake profiles used for inauthentic activities and discussed the prevalence of such accounts in spreading misinformation and scams.

Fake User Account Detection in Online Social Media Networks (2023): Explored machine learning and neural network techniques for detecting fake accounts. The study emphasized the role of Random Forests and other algorithms in identifying fake profiles based on user behavior and interaction patterns.

A Study of Different Methodologies to Detect Fake Account on Social Media (2022): Reviewed various machine learning techniques for detecting fake accounts, highlighting their strengths and weaknesses. The authors proposed a comprehensive approach combining multiple algorithms to improve detection accuracy.

VI. PRODUCT OVERVIEW

The social media fake account detection system is crafted to accurately classify and detect fake accounts. Initially, our system collects relevant features of the suspicious account via a user-friendly graphical user interface (GUI). Users facing issues with fake accounts report these features, which are then processed through our K-Nearest Neighbors (KNN) model.

This model classifies and predicts the authenticity of the account based on training and test datasets. The accuracy percentage of the prediction is then calculated. If the accuracy falls below 60 percent, a gradient boosting algorithm is employed to enhance the model's performance. The system re-calculates the accuracy percentage, and if it remains below 60 percent, it informs the user of the account's fake status, providing the relevant accuracy percentage.

This process empowers users to make informed decisions based on the predicted accuracy percentage, such as whether to block the account or continue communication. This not only helps users safeguard themselves but also maintains the integrity of online interactions.

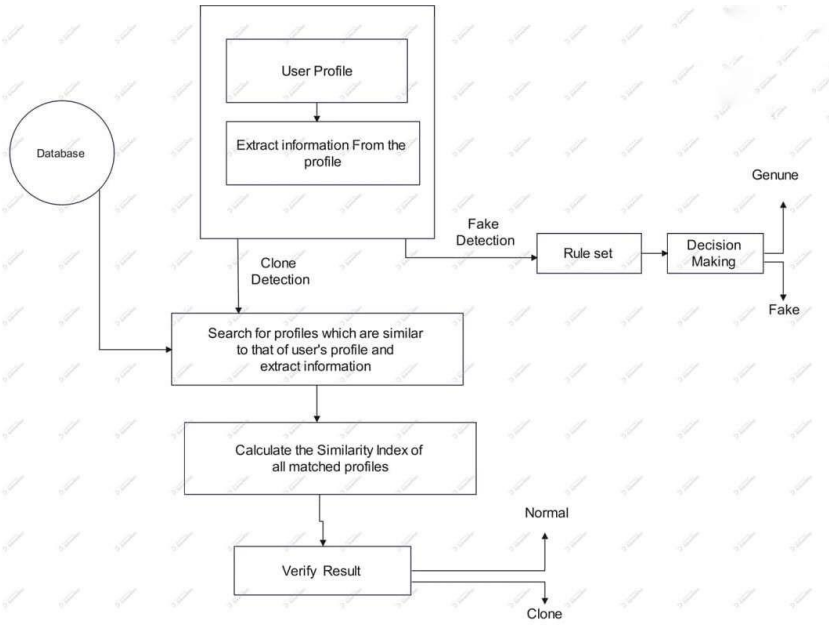

Fig. Algorithm Flowchart

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 3, October 2024

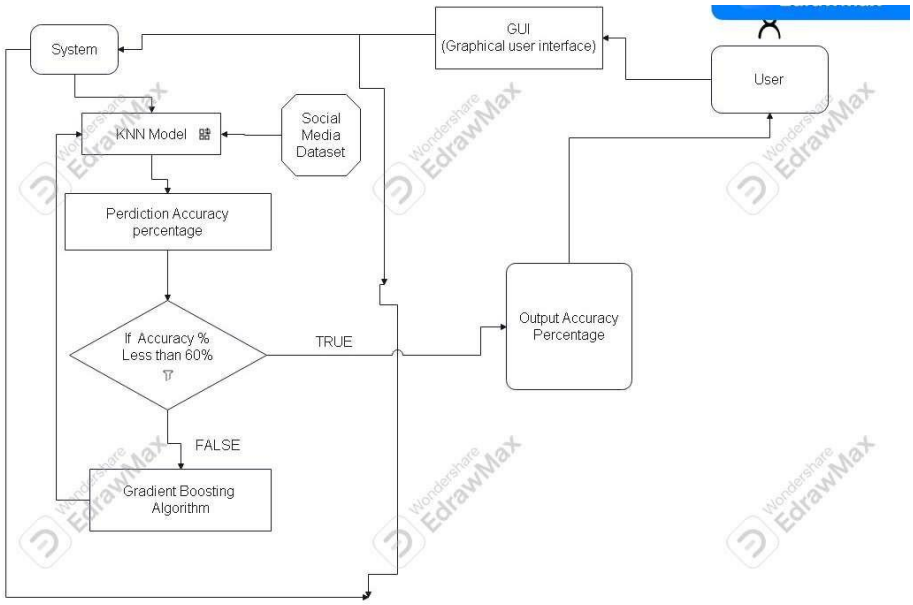

Fig. System Flowchart

VII. CONCLUSION

This social media fake account detection system effectively classifies and identifies fake accounts through an advanced K Nearest Neighbors model, enhanced by gradient boosting for improved accuracy. By leveraging user-reported features and predictive algorithms, it provides users with a reliable accuracy percentage, empowering them to make informed decisions on whether to block or continue interacting with accounts, thus safeguarding online interactions and maintaining platform integrity.

VII. ACKNOWLEGDEMENT

We sincerely thank Ms. Snehal Mangale for their invaluable guidance and mentorship throughout this research project. Their expertise in computer science has been crucial to completing this review paper. We also express our gratitude to the authors of referenced papers, whose work has significantly enhanced our understanding of fake account detection using AI and ML algorithms. Finally, we appreciate the institutions and organizations for providing the resources and opportunities necessary for this research.

REFERENCES

[1]. Yang, Y., Liu, B., Zhang, J., and Wang, X. (2024). Kai-Cheng Yang 1, 2, Danishjeet Singh1, and Filippo Menczer1

[2]. 1Observatory on Social Media, Indiana University, Bloomington, USA 2Network Science Institute, Northeastern University, Boston, USA “AI – Generated Faces in the Real World: A Large-Scale Case Study of Twitter Profile Images”.

[3]. Sruthi, R. and Vasavi, S. (2023).N.Divya Sruthi, 1 , CH.Vasavi 2 1Assistant Professor, 2Assistant Professor, Department of CSE, Geethanjali Institute of Science Technology, Nellore, A.P. , India "Detection of Fake Profiles on Online Social Network Platforms: Performance Evaluation of Artificial Intelligence Techniques".

[4]. Al-garadi, M.A., Varathan, K.D., Ravana, S.D: Cybercrime detection in online communications: “The experimental case of cyberbullying detection in the Twitter network. Comput.” Hum. Behav. Hum. Behav. 63, 433–443 (2016).

[5]. Hammoodi and Obaid (2023): "Fake User Account Detection in Online Social Media Networks Using Machine Learning and Neural Network Techniques".

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 3, October 2024

[6]. Sharma, B., Raju, A. (2023). "Fake Account Detection Using Machine Learning." International Journal of Creative Research Thoughts (IJCRT)5.

[7]. Ezarfelix, J., et al. (2022). "Fake Account Detection in Social Media Using Machine Learning Methods." Bulletin of Electrical Engineering and Informatics4.

[8]. Maniraj, S. P., et al. (2019). "Fake Account Detection using Machine Learning and Data Science." International Journal of Innovative Technology and Exploring Engineering (IJITEE)3.

[9]. Khaled, S., et al. (2022). A Systematic Literature Review: "Instagram Fake Account Detection Based on Machine Learning." Jurnal EMACS2."