FAKE PROFILE DETECTION ON SOCIAL MEDIA USING MACHINE LEARNING ALGORITHMS

Project Reference No.: *47S_BE_1556*

College ***:*** *New Horizon College of Engineering, Marathalli*
Branch ***:*** *Department of Computer Engineering*
Guide(s) ***:*** Mrs. Roja R and Dr. T. Kavitha
Student(S) ***:*** Mr. Melbin Biju
Mr. Alistair Kissinger
Mr. Sanjay G.
Mr. Prince Enoch

Keywords:

Online Social Networks (OSNs), Support Vector Machine, Decision Tree, Neural Network, Accuracy, Random Forest.

Introduction:

Online social networks (OSNs) have become immensely popular ways for people to connect, share information, and communicate. However, the open and real-time nature of OSNs also makes them vulnerable to misuse by malicious actors. One major threat is from automated fake accounts known as social bots. Researchers estimate that 9-15% of Social Media accounts are social bots that mimic human behavior. While some bots can be benign or useful, a growing number are being used for harmful political and social manipulation. Recent incidents have highlighted how social bots were weaponized to spread misinformation and influence major political events like the 2016 US presidential elections and Brexit referendum.

Detecting and removing fake accounts is critical for social networks. Some machine learning techniques have been proposed for this purpose but more research is needed. Maintaining privacy and security of genuine users is also a major concern with the growth of social media. inauthentic accounts present an evolving challenge that can erode trust and enable the spread of misinformation on online social platforms. Ongoing efforts are needed to detect these accounts and mitigate their impacts through policy changes and technological solutions.

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. Decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

Objectives:

To collect a robust, label dataset of Social Media user profiles and news articles classified as reliable or unreliable using the fact-checking site PolitiFact. This served as ground truth data for training and testing models. To extract a range of features from the user profiles and news content that could potentially signal unreliability. This included text of articles, user metadata like number of followers, and context around how news spread between users. Feature engineering transformed raw data into informative inputs for models. To develop a deep neural network architecture that could accurately classify user profiles as reliable or unreliable based solely on the engineered features. Performance was benchmarked against SVM and KNN classifiers. The goal was over 90% accuracy based on ground truth labels. To create a variant of the predictive model that additionally incorporated signals of influence between users, based on propagation patterns of similar news articles. This explored whether indications of impact on others improves detection of misinformation spreaders.

Methodology:

1) Information Assortment : To Scrape Media Profile Information To Create A Dataset, But Ran Into Issues With Profile Terms Of Service Preventing Scraping. Consider Focusing Manual Data Collection On A Narrow Domain Where Quality Matters More Than Quantity. Or Investigate If Social Media Makes Certain Data Available Through Official Channels Like Analytics Partners.
2) Uploading Dataset And Information Preprocessing:

   Data Analyzing Is The First To Boost The Model Quality. It Can Be Done In Four Steps

   a) Information Data Assortment
   b) Handling Unavailable Information
   c) Handling Different Type Of Information
   d) We Have Used Data For Preprocessing
3) Ensemble Learning Strategies: Ensemble Learning Combines Multiple Machine Learning Models Together To Produce Better Predictions Than Any One Model Could Produce Alone.

   a) Bagging - Training Multiple Instances Of The Same Model On Different Random Subsets Of The Data, Then Averaging/Voting To Predict. Examples: Random Forests, Extra-Trees.

b) Boosting - Training Models Sequentially, With Each New Model Focusing More On The Errors Of The Previous Model. Examples: Adaboost, Gradient Boosting Machines.

c) Stacking - Training A Secondary Meta-Learner Model To Combine The Predictions From Multiple Primary Base Models.

4) Accuracy Comparison And Prediction: The More Signals Like These That Indicate Inauthentic Behavior, The More Likely An Account Is Fake Or Spreading Spam/Malicious Content. Analyzing Multiple Factors Together, Rather Than Just Any One Indicator, Can Help Make The Most Accurate Detection.

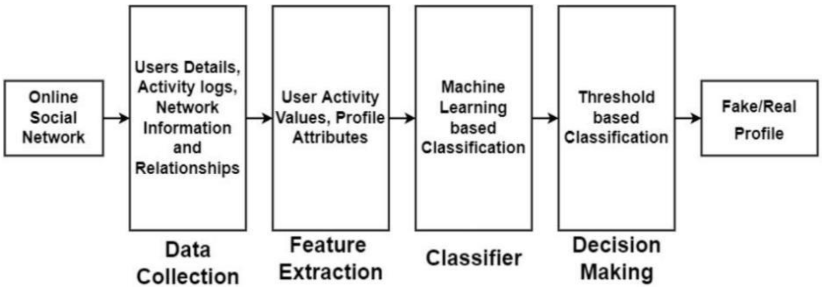
Online Social Network
Users Details, Activity logs, Network Information and Relationships
User Activity Values, Profile Attributes
Machine Learning based Classification
Threshold based Classification
Fake/Real Profile
Data Collection
Feature Extraction
Classifier
Decision Making

Fig 1. Block Diagram

Conclusion:

These inauthentic accounts are used to spread misinformation, inflammatory content and malicious links. Detecting and disabling them has become critical. Random Forest method have exploited machine learning techniques to distinguish between genuine and fake user accounts. The key idea is to train ML models on historical profile, post and network data labeled as belonging to real users or bots/fakes. Models learn inherent patterns that characterize each class. The trained system can then be applied to new, unseen account data to predict if they are likely real or fake

The A variety of features derived from user metadata, posts, network structure and temporal activity have proven useful. This includes profile properties, tweet content and topics, posting frequency, network connections, etc. Feature extraction and selection is crucial to good performance. Models like random forests, SVM, neural networks, graph-based methods have shown success. Ensemble approaches that combine multiple algorithms also effective. Social graph analysis techniques that model user interactions also hold promise for this graph-structured data.

Accuracy rates of 90-95% on benchmark datasets highlight viability, but ongoing maintenance required. Adversarial attacks that fool models also need addressing as part of model robustness. Deploying protections ahead of public availability also necessary to prevent infiltration.

Scope for future work:

One method searched profile attributes like name and location to find suspicious, similar profiles and compared them to a "normal" profile to flag potential fakes with low resemblance. A limitation was possibly mislabeling real accounts as fake. Another framework detected clones by separating accounts, identifying suspicious ones via similarity schemes, and removing matched accounts from friend lists. However, performance declined with large datasets. Some research has used real-time fake account behaviour data and graph topology analysis to improve detection. Machine learning techniques have also been leveraged, including training models on features extracted from profile content and communications.

A range of techniques have shown promise identifying fake and cloned social media profiles by spotting patterns in account characteristics and behaviour. Hybrid methods combine machine learning, graph analysis, and heuristics. Key challenges include improving computational efficiency, managing imbalanced classes, and avoiding misclassification of legitimate accounts when applied broadly.