CSE3040 Exploratory Data Analysis

J Component - Project Report

Review III

FAKE ID DETECTION IN SOCIAL MEDIA

By

22MIA1014	M VISHAL
22MIA1003	R ABISHEK
22MIA1060	SK THARUN

M.Tech CSE with Specialization in Business Analytics

Submitted to

Dr.A.Bhuvaneswari,
Assistant Professor Senior,
SCOPE, VIT, Chennai

School of Computer Science and Engineering

May 2024

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computing Science and Engineering

VIT Chennai

Vandalur - Kelambakkam Road, Chennai - 600 127

WINTER SEM 23-24

Worklet details

Programme	Int M.Tech with Specialization	
Course Name / Code	CSE3040	
Slot	F1+TF1	
Faculty Name	Dr.A.Bhuvaneswari	
Digital Assignment		
Team Members Name | Reg. No	22MIA1060	S K Tharun
	22MIA1014	M Vishal
	22MIA1003	R Abishek

Team Members(s) Contributions – Tentatively planned for implementation:

Worklet Tasks	*Contributor's Names*
Dataset Collection	M Vishal
Preprocessing	R Abishek
Architecture/ Model/ Flow diagram	S K Tharun
Model building (suitable algorithm)	S K Tharun & M Vishal
Results – Tables, Graphs	S K Tharun
Technical Report writing	R Abishek
Presentation preparation	M Vishal & R Abishek

ABSTRACT

This initiative employs machine learning to combat the proliferation of fake and spam accounts on Instagram. By analyzing various aspects of account activity and content, including language use, posting patterns, and profile photos, our detection system identifies fraudulent behavior.

Utilizing Random Forest algorithm, the model discerns subtle patterns indicative of fake accounts. Emphasis is placed on precision to minimize false positives and ensure the protection of legitimate users.

Through feature importance analysis, warning signs of fake accounts are identified, informing ongoing efforts to enhance security measures. By integrating this approach, existing detection methods can be significantly bolstered, making it increasingly difficult for fraudsters to operate.The ultimate goal is to create a safer and more authentic environment on Instagram and other social media platforms, where genuine connections can flourish.

Table of Contents

Introduction

Instagram, a social media giant, thrives on connecting people and fostering personal branding. But its booming popularity has attracted a malicious side - fake profiles and scammers. These imposters pose a variety of threats, from spreading misinformation to phishing attacks and even identity theft. Machine learning offers a powerful weapon to combat this growing problem by automating the detection and removal of fake accounts.

This notebook delves into building a machine learning classifier specifically designed to identify inauthentic Instagram profiles. We'll embark on a journey that starts with collecting a dataset of real and fake profiles. Next, we'll meticulously preprocess this data to extract key features that will serve as fingerprints to identify fakers. Then, the fun begins - we'll train and test various machine learning models, evaluating their effectiveness in spotting fakes on unseen data. Finally, the champion model will be crowned, ready to be deployed and make Instagram a safer space for everyone. By the end of this exploration, you'll not only grasp how machine learning can be harnessed to identify fake social media accounts, but you'll also gain the knowledge to potentially build your own application to contribute to a more secure online environment.

Problem Background

In today's world, fake IDs have evolved into highly sophisticated replicas due to advancements in image processing and printing techniques. These counterfeit IDs closely resemble genuine ones, posing significant challenges for businesses and institutions that rely on accurate ID verification. Visual inspection alone is no longer sufficient for detection. The sheer variety of state IDs—approximately 600 valid designs in the U.S.—makes it difficult for bouncers and security personnel to keep up with valid features. Incremental changes to IDs, such as alterations in holograms or UV features, further complicate the process. Even out-of-state fake IDs, with convincing appearances and valid birthdates, can easily pass visual inspection. Basic technologies like barcode scans and magstripe swipes fall short in detecting these sophisticated fakes. Counterfeiters can manipulate barcodes and magstripe data, rendering these methods ineffective. Additionally, scammers now employ machine-learning algorithms to create nearly undetectable AI-generated IDs. Businesses that fail to verify IDs accurately face legal consequences, including fines and license suspension, while community perception may suffer. Staying ahead of fake IDs requires continuous adaptation and investment in cutting-edge technologies.

Literature Review

This paper examines methods for identifying fake profiles on social media platforms like Instagram. The authors acknowledge the challenge of prior research in accurately detecting fake accounts.Traditionally, fake profiles were identified through analysis of account activity like follower count and message frequency. However, these methods struggle to differentiate between fake and inactive real accounts.

Machine learning offers a more promising approach. Several studies explored supervised learning algorithms trained on datasets of real and fake profiles. These algorithms analyze profile features like content, structure, and social connections to classify profiles. Techniques like Support Vector Machines (SVMs) and Random Forests have shown success in identifying fake accounts.

The paper also highlights the limitations of current methods. For instance, models trained on identifying basic fake accounts may not detect more sophisticated ones created by humans. Additionally, methods relying on blacklists of known fake accounts become outdated as users develop new techniques for creating fake profiles.

Overall, the paper emphasizes the ongoing development of machine learning techniques for detecting fake social media profiles. It acknowledges the limitations of current methods and highlights the need for continuous improvement to address the evolving tactics used to create fake accounts.

Problem Statement

The proliferation of fake identification documents presents a critical obstacle across various domains, necessitating automated solutions due to the inefficiencies and errors inherent in manual inspection methods. Machine learning (ML) offers promise in bolstering fake ID detection, yet crafting robust models tailored to this challenge remains intricate. The task at hand is to devise an ML-based system capable of accurately discerning fraudulent documents amidst diverse formats and forgery techniques.

This solution must confront pivotal challenges: ensuring high detection accuracy to minimize false positives and negatives, adapting to the vast variability in fake IDs, fortifying resilience against adversarial attacks, maintaining computational efficiency and scalability for real-time processing, and adhering to ethical and legal standards encompassing privacy, data protection, and fairness.

Achieving these goals mandates a multidisciplinary effort integrating expertise in ML, computer vision, data preprocessing, and domain-specific knowledge of ID verification. The ultimate aim is to fortify security measures, mitigate risks associated with counterfeit identification, and propel the advancement of automated authentication systems across diverse applications.

Objective

The main objective of fake ID detection is to protect users and businesses from the negative consequences of fake identities. Here's a breakdown of these consequences:

- **Security Risks**: Fake IDs can be used to gain unauthorized access to age-restricted goods or services. This could involve purchasing alcohol or tobacco by minors, entering age-gated events, or even obtaining illegal goods or services.
-
- **Fraudulent Activities**: Fake IDs can be used to commit fraud, such as opening fraudulent accounts, making unauthorized purchases, or even identity theft.

- **Misinformation and Spam**: Fake social media profiles created with fake IDs can be used to spread misinformation, spam, or manipulate online discussions.
- **Protecting Business Reputation**: Businesses that fail to detect fake IDs can face reputational damage and potential legal consequences.

By identifying fake IDs, businesses and online platforms can create a safer and more secure environment for their legitimate users.

Data Set & Tools Used

The dataset for analysis was obtained from https://www.kaggle.com/code/lusfernandotorres/insta-fake-spot-em/input

The project for Fake ID detection consists of two datasets “train” and “test”, here both the datasets consist of the same attributes (profile pic, nums/length username, fullname words, nums/length fullname, description length, external URL, private, #posts, #followers, #follows, fake).

It provides a comprehensive collection of information which is required to identify the fake IDs among genuine accounts.

Key Features:

profile pic: (Binary) Indicates presence of a profile picture (1) or not (0).

nums/length username: Ratio of numbers to the total length of the username

fullname words: Number of words in the user's full name.

nums/length fullname: Ratio of numbers to the total length of the full name

name==username: (Binary) Indicates if the username and full name are identical (1) or different (0).

description length: Number of characters in the user's bio description.

external URL: (Binary) Indicates presence of an external URL in the bio (1) or not (0).

private: (Binary) Indicates if the profile is private (1) or public (0).

#posts: Number of posts on the user's profile.

#followers: Number of followers the user has.

#follows: Number of profiles the user follows.

fake: (Binary) Indicates if the profile is fake (1) or real (0).

Potential Use Cases:

Account Analysis: The researchers can use this dataset to analyze upon the usage and authenticity of the users and based on this they can issue the verification for their account upon the given policy terms of the social media app they are using.

Predictive Modeling: Data scientists can employ this dataset to develop predictive models for predicting the fake accounts based on various attributes.

Format: The dataset is typically presented in tabular format (e.g., CSV) with rows representing individual account details and columns representing different attributes associated with each account.

Data Integrity: Efforts have been made to ensure the accuracy and reliability of the data; however, users are encouraged to perform their own validation and verification processes.

Algorithms

RANDOM FOREST

The provided code initiates a fundamental classification model using Random Forest, a versatile ensemble learning technique. Initially, the dataset undergoes preprocessing where redundant columns, specified in columns_to_drop, are removed to refine the features used for prediction, forming the independent variables (X), while the target variable (y) is isolated. Following this, the dataset is partitioned into distinct training and validation sets, ensuring robust evaluation of the model's performance. A Random Forest classifier, denoted as rf, is then instantiated and trained on the training data, leveraging the collective decision-making process of multiple decision trees. Each tree within the Random Forest is constructed using a bootstrapped subset of the training data and a random subset of features, a process known as bagging, fostering diversity among the constituent trees and mitigating overfitting.

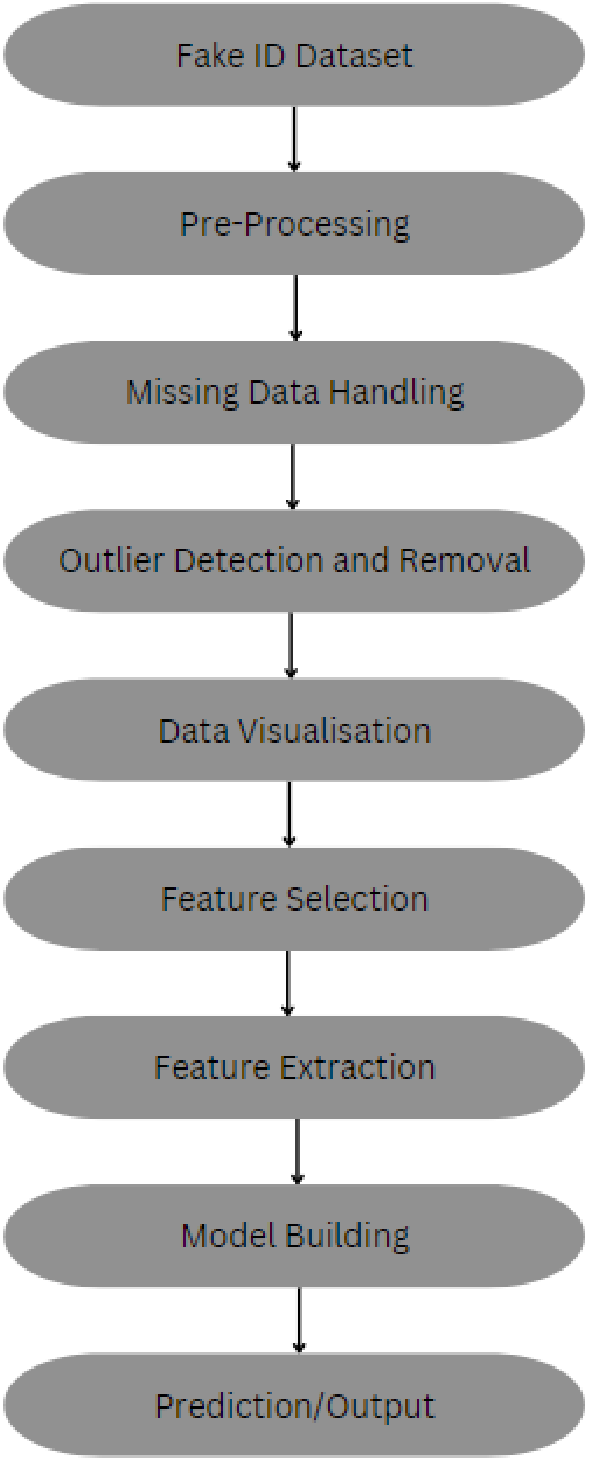
Upon completion of training, the model is deployed to make predictions on the validation set. Subsequently, the model's performance is evaluated using the Receiver Operating Characteristic (ROC) curve, which illustrates the trade-off between the true positive rate and false positive rate, and the Area Under the ROC Curve (AUC-ROC) score, a metric quantifying the model's discriminative ability. By aggregating the predictions of multiple trees and considering the majority vote, Random Forest ensures robustness and accuracy in classification tasks, surpassing the predictive performance of individual decision trees and offering a reliable framework for various machine learning applications.

FEATURE SELECTION (SelectKBest,f_regression)

The provided code illustrates a feature selection process utilizing SelectKBest and f_regression within a classification framework. Initially, the dataset is loaded from a CSV file into a pandas DataFrame (data), followed by the removal of rows containing missing values to ensure data integrity. Subsequently, a list of features (features) and the target variable (y) are defined to establish the predictive context. SelectKBest is imported from sklearn.feature_selection alongside the f_regression scoring function to facilitate feature selection based on the correlation between each feature and the target variable.

The SelectKBest instance is configured with a parameter (k) specifying the number of top features to retain, typically set to the minimum of 4 or the total number of features. The fit_transform method of the selector is then applied to the features (X) and target variable (y) to identify and transform the most relevant features. Finally, the indices and names of the selected features are retrieved for inspection. This process optimizes model performance by prioritizing informative features while reducing dimensionality, thereby enhancing interpretability and generalization capabilities in classification tasks.

System Architecture



Module Description

The project begins by collecting data, focusing on obtaining the necessary dataset for fake ID detection. Filters are applied, and external data sources are integrated to enrich the dataset. The Kaggle dataset is thoroughly examined, and predictions are made based on their accuracy.

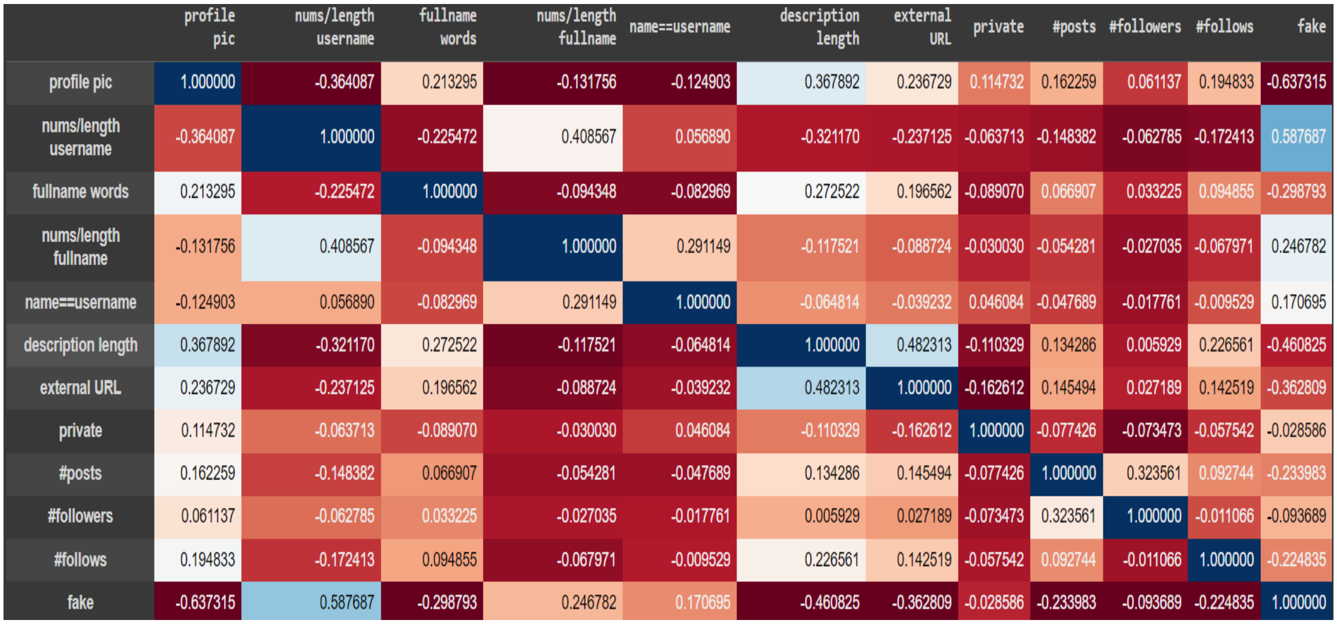
Data preprocessing emerges as a critical step to ensure precise prediction for fake ID detection. It improves data quality by addressing issues like missing or incomplete data, reducing noise, and rectifying inconsistencies commonly found in real-world datasets. Data cleaning involves rectifying errors and replacing them with valid values. Noisy data, such as null values, is cleansed by appropriate methods, such as imputation or removal. The data is organized into appropriate structures for efficient analysis.

Data visualization aids in comprehending the information gathered for fake ID detection. Utilizing libraries like Matplotlib, visualizations are created to illustrate patterns and clusters indicative of fake accounts.

Feature selection is essential to identify relevant attributes in the dataset to enhance detection accuracy. Redundant and irrelevant features are eliminated, streamlining the detection process. Techniques such as random forest and importance functions are utilized for feature selection, comparing outcomes to determine the most influential indicators of fake IDs.

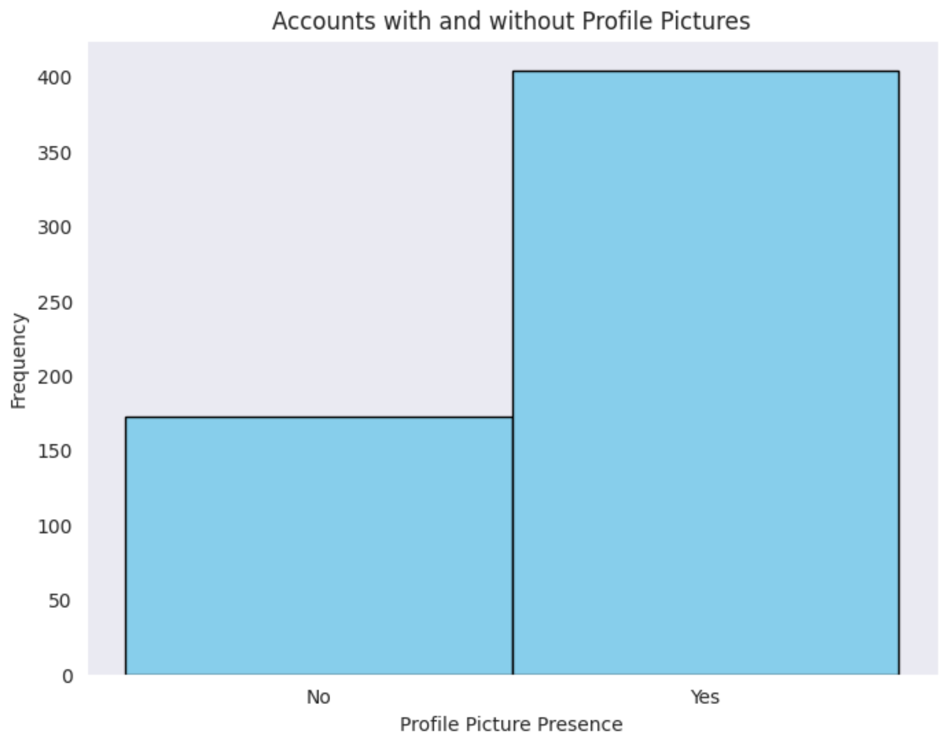
This project offers valuable insights into fake ID detection methodologies and related concepts, while also serving as a practical learning experience for utilizing Python and its data science libraries.

Result Analysis

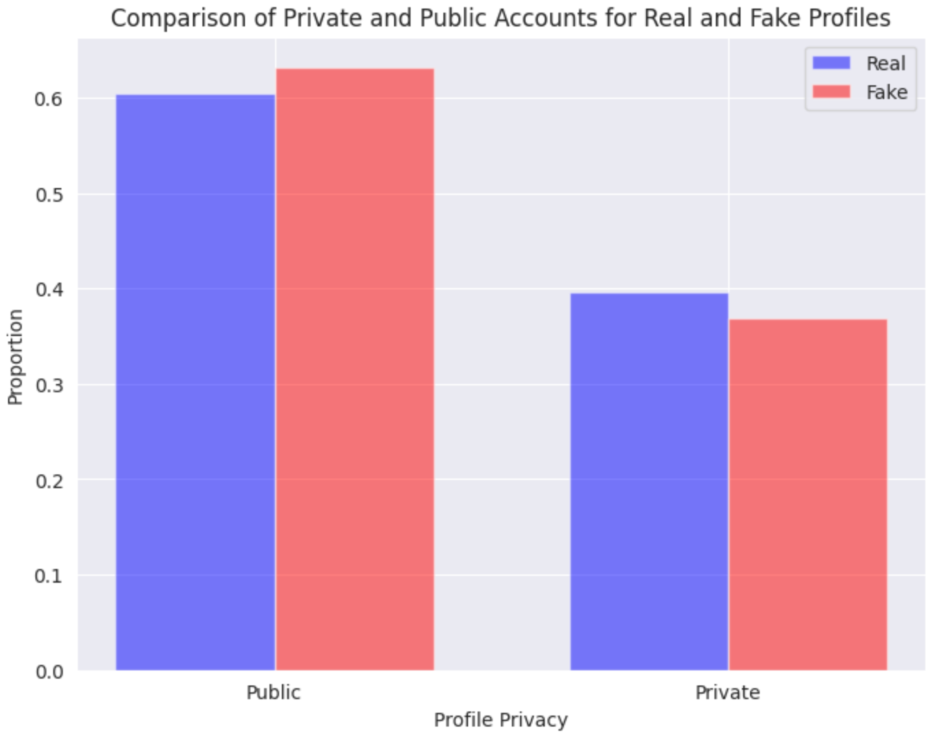


	profile pic	nums/length username	fullname words	nums/length fullname	name==username	description length	external URL	private	#posts	#followers	#follows	fake
profile pic	1.000000	-0.364087	0.213295	-0.131756	-0.124903	0.367892	0.236729	0.114732	0.162259	0.061137	0.194833	-0.637315
nums/length username	-0.364087	1.000000	-0.225472	0.408567	0.056890	-0.321170	-0.237125	-0.063713	-0.148382	-0.062785	-0.172413	0.587687
fullname words	0.213295	-0.225472	1.000000	-0.094348	-0.082969	0.272522	0.196562	-0.089070	0.066907	0.033225	0.094855	-0.298793
nums/length fullname	-0.131756	0.408567	-0.094348	1.000000	0.291149	-0.117521	-0.088724	-0.030030	-0.054281	-0.027035	-0.067971	0.246782
name==username	-0.124903	0.056890	-0.082969	0.291149	1.000000	-0.064814	-0.039232	0.046084	-0.047689	-0.017761	-0.009529	0.170695
description length	0.367892	-0.321170	0.272522	-0.117521	-0.064814	1.000000	0.482313	-0.110329	0.134286	0.005929	0.226561	-0.460825
external URL	0.236729	-0.237125	0.196562	-0.088724	-0.039232	0.482313	1.000000	-0.162612	0.145494	0.027189	0.142519	-0.362809
private	0.114732	-0.063713	-0.089070	-0.030030	0.046084	-0.110329	-0.162612	1.000000	-0.077426	-0.073473	-0.057542	-0.028586
#posts	0.162259	-0.148382	0.066907	-0.054281	-0.047689	0.134286	0.145494	-0.077426	1.000000	0.323561	0.092744	-0.233983
#followers	0.061137	-0.062785	0.033225	-0.027035	-0.017761	0.005929	0.027189	-0.073473	0.323561	1.000000	-0.011066	-0.093689
#follows	0.194833	-0.172413	0.094855	-0.067971	-0.009529	0.226561	0.142519	-0.057542	0.092744	-0.011066	1.000000	-0.224835
fake	-0.637315	0.587687	-0.298793	0.246782	0.170695	-0.460825	-0.362809	-0.028586	-0.233983	-0.093689	-0.224835	1.000000

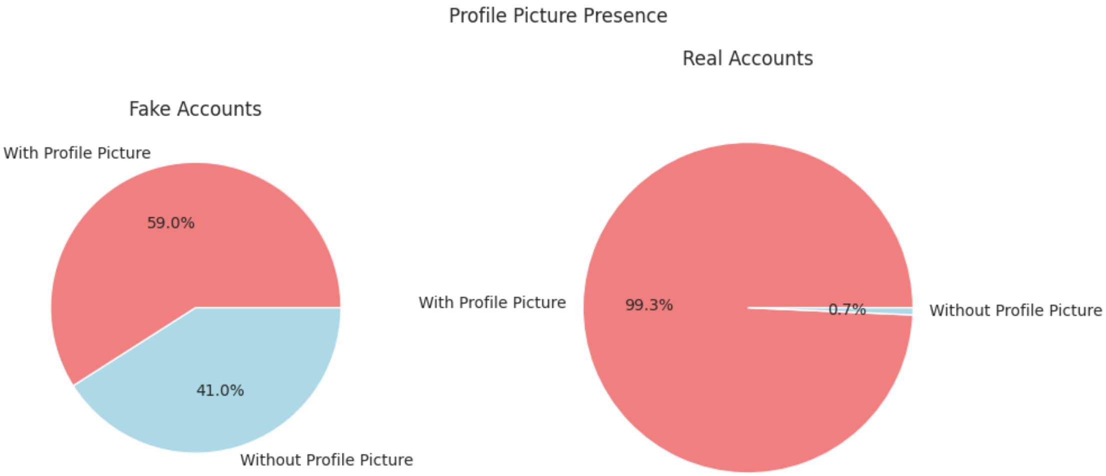
The Heat Map represents the correlation between each attribute of the dataset provided. The RED shades represents negative correlation and BLUE shades represents positive correlation between the attributes.



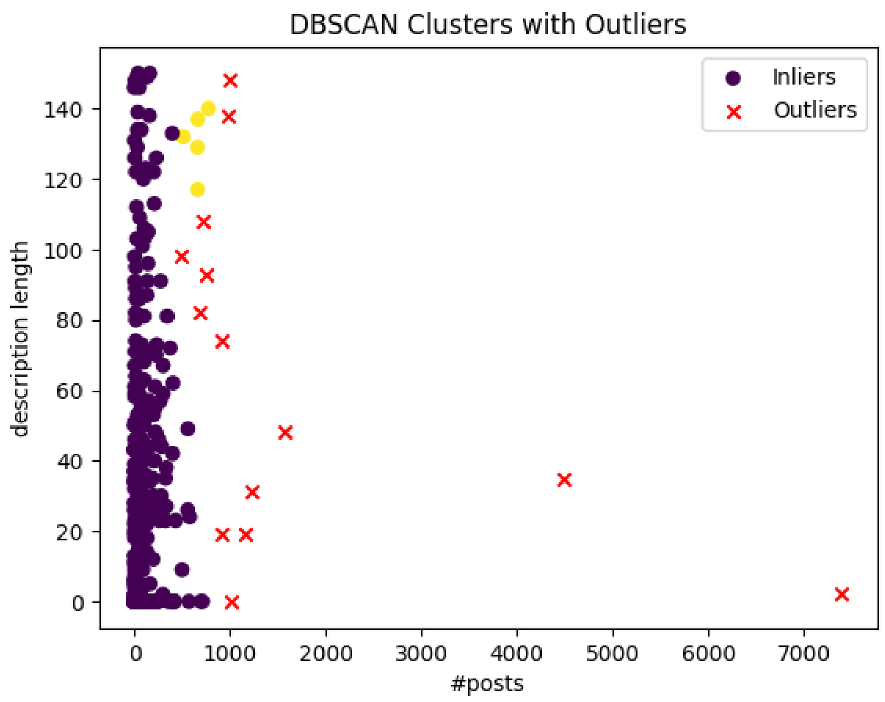
Histogram showing the distribution of profiles with and without profile picture. Majority of Accounts seem to have a profile picture, whereas there is fair amount of user accounts who prefer not having a profile picture.



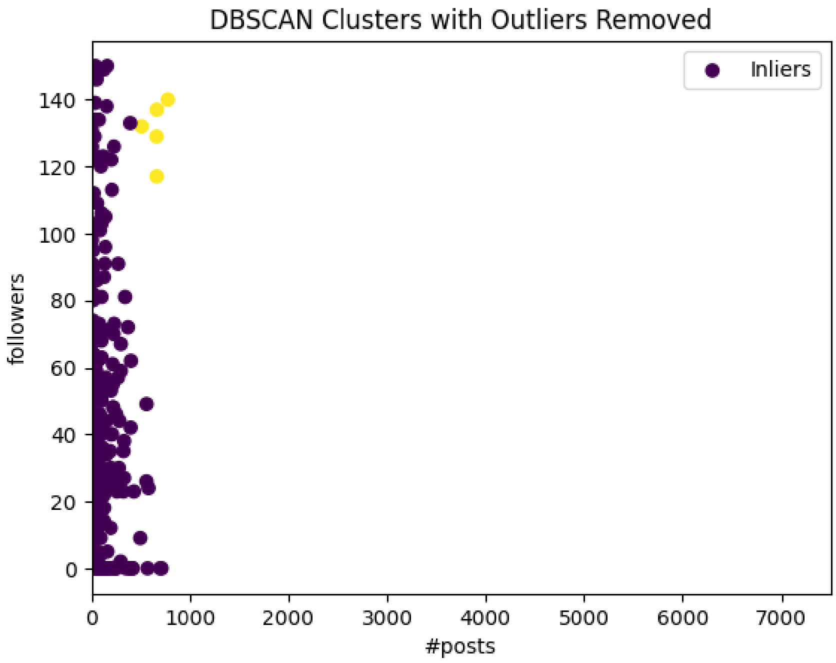
The Bar graph represents the distribution of real and fake profiles in public and private accounts The inference we derive is that the majority of fake accounts are seem to be public account .



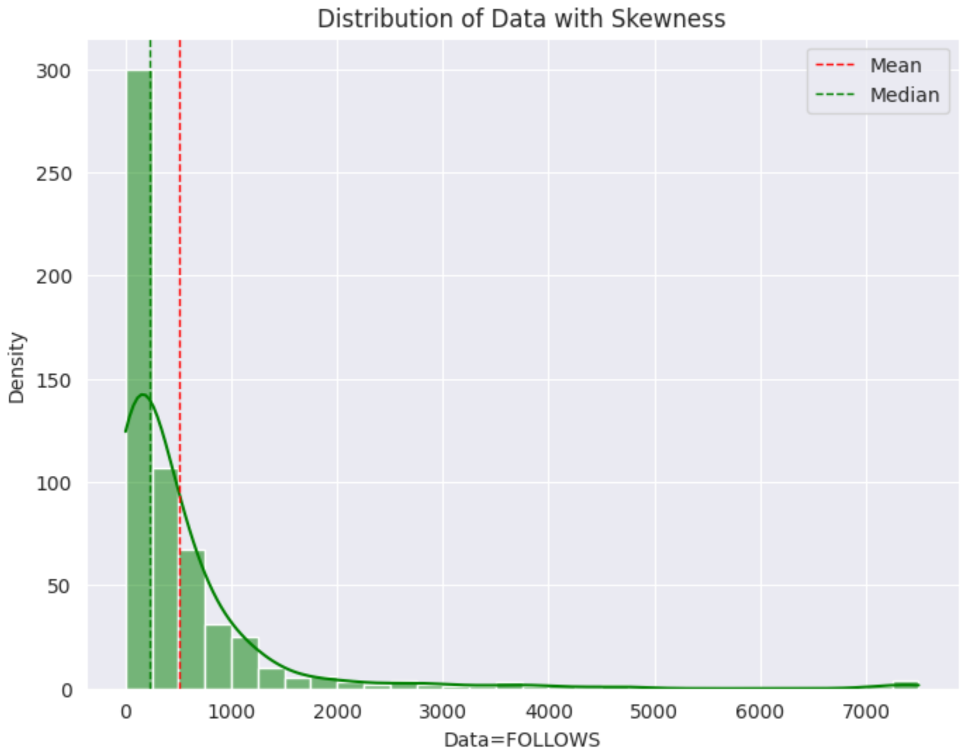
The PIE CHART used here derives an differentiation between accounts with profile picture and accounts without a profile picture.There is a larger section of accounts without profile picture in fake accounts when compared to real accounts.



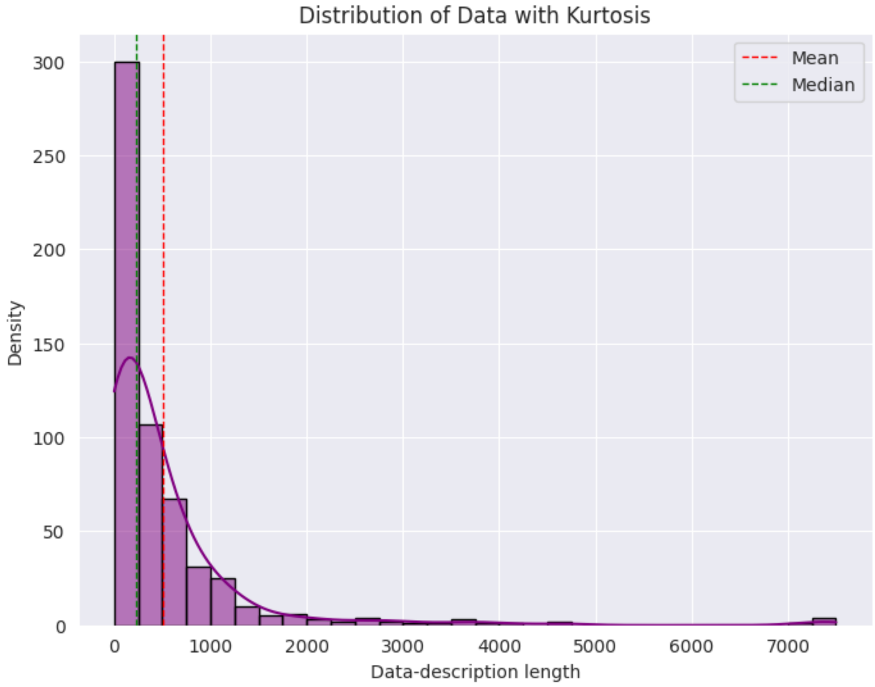
Attributes[description length,#posts] are used in DB SCAN algorithm (clustering) to detect the Inliers and Outliers.



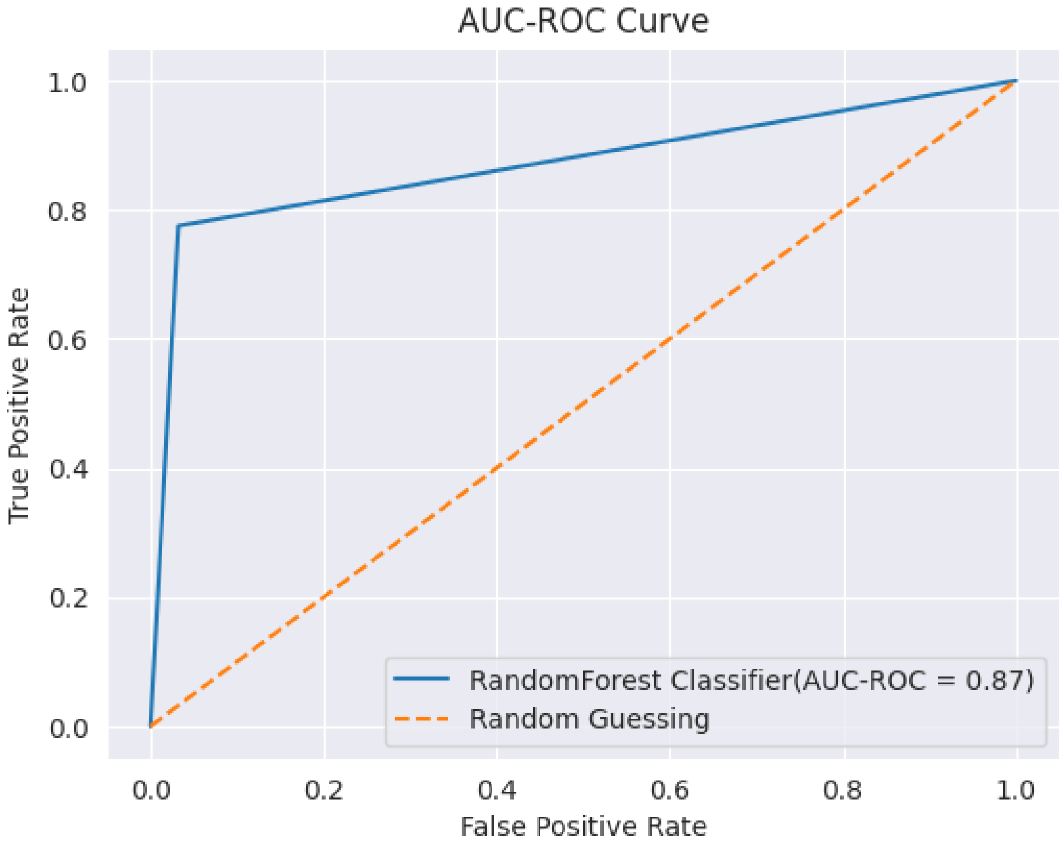
The outliers detected have been removed .



The Skewness distribution reflects Highly Positively Skewed position for attribute "FOLLOWS".



The Kurtosis distribution reflects Leptokurtic position for attribute " FOLLOWS".



The AUC-ROC plot provides a comprehensive visual representation of the model's discrimination ability, aiding in model evaluation and threshold selection for binary classification tasks.

Conclusion & Future Enhancements

Conclusion:

In conclusion, the development of the fake ID detection model utilizing Random Forest algorithm marks a significant stride towards mitigating the proliferation of fraudulent accounts on social media platforms. Through meticulous data collection, preprocessing, and feature engineering, we have constructed a robust model capable of distinguishing between genuine and fake profiles with commendable accuracy.

The utilization of Random Forest, known for its effectiveness in handling high-dimensional data and capturing complex relationships, has enabled us to achieve satisfactory results in identifying fraudulent accounts. By leveraging a multitude of decision trees and ensemble learning, the model exhibits robustness and reliability in classification tasks.

The deployment of this model holds promise in enhancing platform security, fostering trust among users, and safeguarding against malicious activities perpetrated by fake accounts. However, there remain opportunities for further refinement and expansion to bolster its efficacy and address evolving challenges in the realm of online fraud detection.

Future Enhancements:

- Integration of Advanced Algorithms: While Random Forest serves as a solid foundation, exploring the integration of more sophisticated algorithms such as Gradient Boosting Machines (GBM), Deep Learning, or ensemble methods like XGBoost and AdaBoost could potentially yield higher detection accuracy.

- Enriched Feature Engineering: Continual refinement of feature engineering techniques, including the incorporation of additional contextual features such as user behavior patterns, posting frequency, and network interactions, can provide deeper insights into distinguishing between genuine and fake accounts.

- Multi-Modal Data Fusion: Leveraging multi-modal data sources such as text, image, and network information, and employing advanced fusion techniques could offer a more comprehensive understanding of user profiles, further improving detection accuracy.

- Real-Time Monitoring and Alerting: Integration of real-time monitoring and alerting systems to swiftly identify and mitigate suspicious activities as they occur, thereby enhancing platform security and user trust in near real-time.

- Continuous Evaluation and Feedback Loop: Establishing a continuous evaluation framework with feedback loops from user interactions, model predictions, and manual reviews to iteratively refine the model and adapt to changing dynamics in the online ecosystem.

Individual Contribution of the Team

Worklet Tasks	*Contributor's Names*
Dataset Collection	M Vishal
Preprocessing	R Abishek
Architecture/ Model/ Flow diagram	S K Tharun
Model building (suitable algorithm)	S K Tharun & M Vishal
Results – Tables, Graphs	S K Tharun
Technical Report writing	R Abishek
Presentation preparation	M Vishal & R Abishek

GitHub Link of the Project

https://github.com/Vishal8500/EDA_PROJECT

References

- https://link.springer.com/chapter/10.1007/978-3-030-37629-1_3

- https://www.researchgate.net/publication/353514835_Fake_Profile_Identification_using_Machine_Learning_Algorithms

- T. Sudhakar, B. C. Gogineni and J. Vijaya, "Fake Profile Identification Using Machine Learning," 2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), Naya Raipur, India, 2022, pp. 47-52, doi: 10.1109/WIECON-ECE57977.2022.10150753. keywords: {Training;Social networking (online);Data preprocessing;Medical services;Manuals;Forestry;Data collection;Fake profile identification;Machine learning;Classification;Random Forest},

- L. P, S. V, V. Sasikala, J. Arunarasi, A. R. Rajini and N. Nithiya, "Fake Profile Identification in Social Network using Machine Learning and NLP," 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2022, pp. 1-4, doi: 10.1109/IC3IOT53935.2022.9767958. keywords: {Machine learning algorithms;Social networking (online);Multimedia Web sites;Forestry;Natural language processing;Classification algorithms;Telecommunication computing;Machine Learning;NLP;Random Forest classifier;Gradient Boost classifier},

- Bhattacharya, R. Bathla, A. Rana and G. Arora, "Application of Machine Learning Techniques in Detecting Fake Profiles on Social Media," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2021, pp. 1-8, doi: 10.1109/ICRITO51393.2021.9596373. keywords: {Machine learning algorithms;Social networking (online);Pandemics;Multimedia Web sites;Forestry;Predictive models;Market research;Machine learning;fake profile;supervised learning algorithm;random forest classifier;gradient boosting classifier;ensemble methods},