Facebook Property Scam Analysis
Final Project Proposal

Student Name : Kamesh kumar

Enrollment ID : w19143688

Supervisior: Yongchao Huang

UNIVERSITY OF WESTMINSTER

Abstract

Scamming is dominating our era in world wide web than the pro's that it was built to serve. In this project I have my attention towards one of most renowned scams that are affecting thousands of users in daily routine. The Facebook property scamming in London area is taking over negative side. Using NLP (Sentimental analysis, Language detection, NER and Topic Modeling) and Machine learning algorithms like sigmoid, Decision Tree and SVM models I am building a prototype to spot these scams and analysis.

Background

Researchers have implemented that 62% of the users are being scammed in Facebook in various ways and nearly 80% of the users face scamming via social media every month and the rate is still increasing than ever before. Social media is one of the best way to connect our world and it makes our life easier in many ways.

Global statistics from 2017 reveal that over 16.7 million people fell victim to online scam [1]. In 2016 alone, the total amount of money stolen through scam exceeded 7 billion US dollars, with predictions that this figure would reach approximately 31 billion by 2020. A significant portion of these schemes transpire on Facebook each year. Many individuals prefer using Facebook as a platform for purchasing goods due to the ability to communicate and negotiate prices directly with sellers. Moreover, the seller's identity often influences a buyer's decision, as knowing the person behind the transaction provides a modicum of mental reassurance. However, despite these factors, online scam continues to persist, leaving us without a concrete solution.

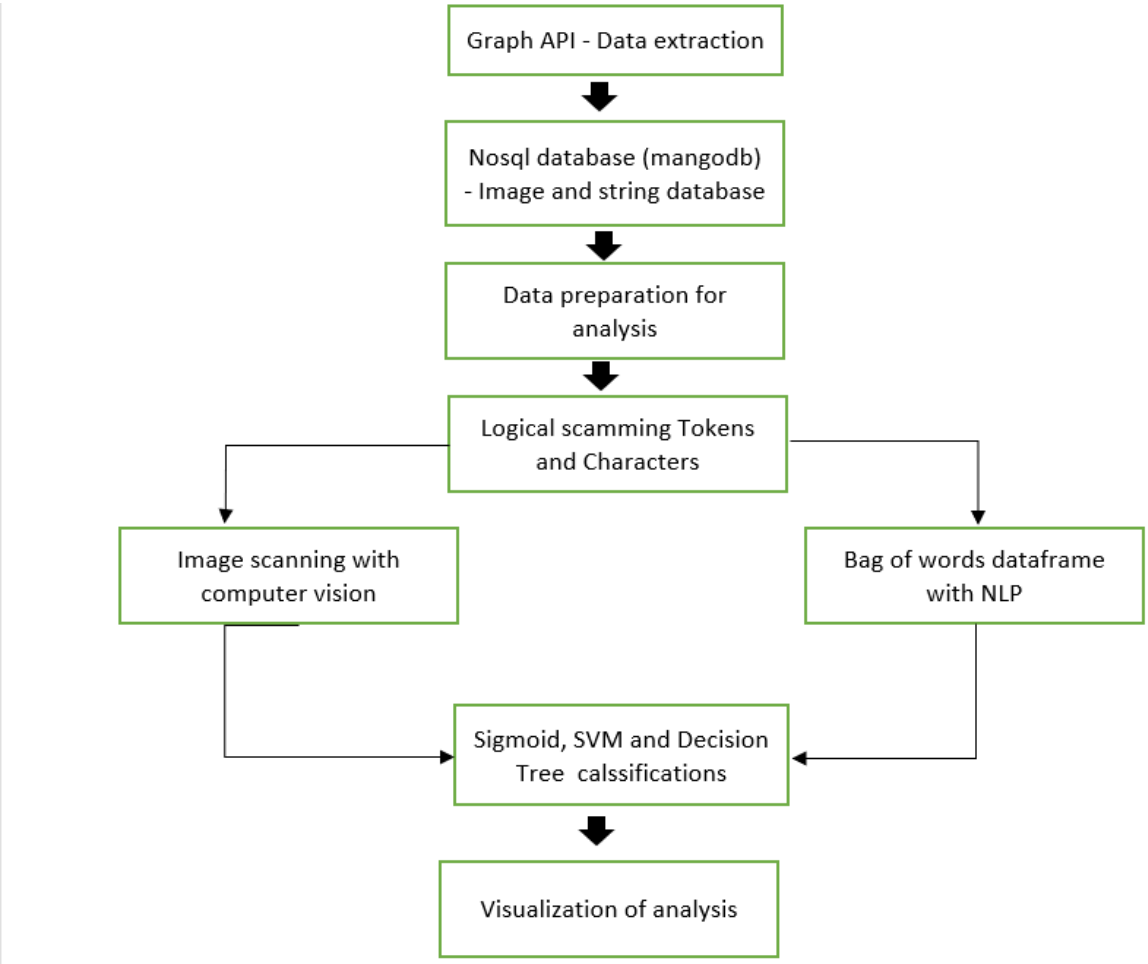
Model Flow:

The flowchart illustrates the process flow for r project work, which involves several steps for scam detection and analysis. Here's an explanation of each step:

1. Graph API to NoSQL Database for Image and String Storage: This step involves utilizing a Graph API to store both image and string data in a NoSQL database. The API interact with the database and store relevant information, such as images of potential scams and associated textual data.

2. Data Preparation for Analysis: In this step, the stored data is prepared for analysis. It may

involve data cleaning, preprocessing, and transformation to ensure the data is in a suitable format for further processing.



3. Logical Scamming Token and Character Classification: This step focuses on the classification of tokens and characters within the textual data to identify potential scam-related patterns or keywords. It involves applying NLP techniques to classify and extract relevant information that may indicate fraudulent activity.

4. Image Scanning with Computer Vision: Using computer vision techniques, this step involves scanning the stored images to identify visual cues or features that may be indicative of scams. Computer vision algorithms can extract relevant information, such as specific objects, text, or patterns present in the images.

5. Bag of Words Extraction into Dataframe for NLP: The textual data is further processed by applying the Bag of Words technique. This involves extracting individual words or tokens from

the text and converting them into a structured format, such as a dataframe, which can be used for NLP analysis.

6. Classification using Sigmoid/SVM/Decision Tree ML: The extracted data is used to train and apply machine learning models for classification. This step involves employing algorithms like Sigmoid, Support Vector Machines (SVM), or Decision Trees to classify and predict whether the given data represents a scam or not based on learned patterns and features.

7. Visualizing Reports in Webpage with Django: The final step involves visualizing the results and reports generated from the analysis in a web-based interface using Django. This allows for easy access and interpretation of the findings, presenting the classification results, statistical summaries, or any other relevant insights to users.

By following this flowchart, the project aims to detect and classify scams by leveraging both textual and visual data, employing NLP and computer vision techniques, and utilizing machine learning algorithms for accurate classification. The final results are then presented in a web-based interface for easy visualization and interpretation

Objective:

The objective of this project is to develop a comprehensive scam detection and analysis system using a combination of natural language processing (NLP), computer vision, and machine learning techniques. The system will utilize a Graph API to store image and string data in a NoSQL database. It will involve preprocessing and transforming the data for further analysis. The project aims to classify and identify scam-related patterns and keywords in textual data through logical token and character classification. Additionally, it will employ computer vision algorithms to scan stored images for visual cues indicative of scams. The extracted data will be used to train machine learning models, such as Sigmoid, SVM, or Decision Tree, for scam classification. The final results and reports will be visualized in a user-friendly web interface using Django, facilitating easy interpretation and access to the analysis outcomes. Overall, the project aims to provide an effective tool for scam detection and visualization, enhancing the ability to identify and mitigate fraudulent activities.

Resources

All the resources are already available freely and without any license or ethical issues.

Software Requirement

The software that we are going to use for the project is API's, Scraping tools(Beautifulsoup, Requests), Jupyter Notebook, NLP, Computer vision libraries, Mangodb, Django/Flask web frame and Machine Learning Models

Language: Python, Nosql

References

[1] E-commerce fraud attack rates hit new highs in 2017, Traci Krepper, 10 April 2018, (cited 20 April 2018) available: http://www.experian.com/blogs/insights/2018/04/e-commerce-fraud-attack-rates-hit-new-highs-in-2017.

[2] Prateek Dewan,Shrey Bagroy and Ponnurangam Kumaraguru, "Hiding in plain sight: Characterizing and detecting malicious Facebook pages," 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 193-196, San Francisco, CA, USA, 2016.

[3] https://www.socialmediatoday.com/news/Social-Media-Scam-Activity-Report-2022/638016/

[4] https://www.statista.com/statistics/1013474/facebook-fake-account-removal-quarter/#:~:text=Facebook%3A%20fake%20account%20removal%20as%20of%20Q4%202022&text=In%20the%20fourth%20quarter%20of,the%20first%20quarter%20of%2020