
FRAUD APP DETECTION USING SENTIMENTAL ANALYSIS

Abhijit Vasant Kote*¹, Shreenath Amol Tambe*², Mahavir Devendra Kasar*³,
Atharva Vijay Otari*⁴, Ms. A.A. Chougule*⁵

*^{1,2,3,4}Department Of Diploma In Computer Engineering, Third Year, Sharad Institute Of Technology,
Polytechnic Yadrav, Ichalkaranji, Kolhapur, Maharashtra, India.

*⁵Lecturer, Department Of Diploma In Computer Engineering, Sharad Institute Of Technology,
Polytechnic Yadrav, Ichalkaranji, Kolhapur, Maharashtra, India.

ABSTRACT

The proliferation of fraudulent mobile applications has raised serious concerns, necessitating innovative strategies for detection and mitigation. This review paper delves into the promising domain of fraud app detection, with a primary emphasis on leveraging sentiment analysis techniques. Sentiment analysis, a branch of natural language processing, plays a pivotal role in identifying deceptive mobile applications by analysing user reviews, app descriptions, and other textual data for emotional cues and opinions. This review paper comprehensively surveys the existing literature on fraud app detection, offering a detailed account of methodologies, tools, and advancements in the field. It categorizes sentiment analysis-based approaches into supervised, unsupervised, and hybrid models, each with its strengths and limitations. Moreover, it discusses the pivotal role of user reviews, ratings, app descriptions, and developer information as key features in detecting malicious applications. The paper also addresses challenges like fake reviews and biased sentiment expressions in sentiment analysis for fraud app detection. Emphasizing the role of datasets and ethical considerations, this review paper underscores the critical importance of sentiment analysis in enhancing the security of mobile app ecosystems. By providing insights into current techniques, challenges, and future prospects, this paper aims to catalyse the development of more robust fraud app detection systems.

Keywords: Fraudulent App, Sentimental Analysis, Fake Review.

I. INTRODUCTION

The number of mobile apps has full-grown at wide-ranging rate over past few years. As an example, as of top of April 2013, there is a unit over 1.6 million apps at Apple's app store and Google play. To stimulate the event of mobile apps, several app stores launched daily App leader boards, that demonstrate the chart ranking of preferred Apps. Indeed, the App leader board is one in all the foremost viral ways in which for promoting mobile Apps. A better rank on the leader board typically result in an enormous range of downloads and millions of greenbacks in revenue. Therefore, App developers tend to explore varied method like advertising companies to market their Apps to own their Apps hierarchal as high as potential in such App leader boards. However, as recent tend, rather than hoping on ancient promoting solutions, shady App developers resort to some deceitful suggests that to deliberately boost their apps associated eventually manipulate the chart ranking on an App store. This can be typically enforced by victimization questionable "bot farms" or "human water armies" to inflate the app downloads, rating and reviews during {a very short time. As an example, writing from Venture Beat [4] reportable that, once associate App was promoted with the assistance of ranking manipulation, it may well be propelled from number one, 800 to prime the highest} twenty-five in Apple's top free leader board and over fifty,000-100,000 new users may well be noninheritable at intervals a handful of days. Such ranking down on App developers WHO commit ranking fraud [3] within the Apple's App Store. Within the literature, whereas there area unit some connected work, like internet ranking spam detection [22], [25], [30], on-line review spam detection [19], [27], [28], and mobile apps recommendation [24], [29], [31], [32], the matter of police work ranking fraud mobile apps continues to be under-explored. To fill this important void, during this paper we propose to develop a ranking fraud detection system for mobile apps, on this line, we determine many vital challenges.

II. LITERATURE SURVEY

1. Liu m., wu c, Zhao x. N., liny.c., 2015. "app Relationship calculation: an iterative process". Published by the IEEE computer society. Issue No.08 - vol.27, pp: 2049-2063 Lovins b.j.,1968. "development of a stemming algorithm". "mechanical translation and computational

Liu M., et al present the convenience of big amount of apps, apps retrieval and recommendation of apps is unbelievably necessary, subsequently, there's a requirement of top of the range of app retrieval and recommendation there need to be a certain app relationship calculation for the recommendation. He finished that varied methods supported app's context, which can entirely realize whether or not or not two apps unit downloaded, place in the meantime or supply similar functions or not.

2. Aamir M. and Bhusry .M., "Recommendation System: State of the Art Approach". International Journal of Computer Applications, Volume 120 -No.12. 2015

Aamir M., et.al [2] explained that the sentimental analysis involves a theme to gather and classify opinions a few products to trace the humour of the individuals for an exact product. However positive document of any product doesn't signify that the reviewer holds a positive opinion on all the aspects of the merchandise.

3. Beel, J., Langer S., Genzmehr M., Gipp B, Breitinger C., and Nürnberg A. 2013 "Research Paper Recommender System Evaluation: A Quantitative Literature Survey". In ACM Recommender System Conference (RecSys). ACM International Conference Proceedings Series (ICPS).

Beel, J., Langers., Genzmehr M., Gipp B, Breitinger C., And Nürnberger, et.al [4] presents a showed that no agreement exists on some way to assess and compare analysis paper recommender approaches. He examined that winds up within the unsatisfying state of affairs that despite the various evaluations, the individual strengths and weaknesses of the planned approaches keep for the foremost thalf unknown.

4. Kularbphettong. K, Somngam S., Tongsiric, and Roonrakwit P. 2014. "A recommender system using Collaborative Filtering and K-Mean Based on Android Application". Journal of theoretical & Applied Information Technology, Vol. 70 Issue 1, p90-94. 5p.

Kularbphettong. K, Somngam S., Tongsiric, And Roonrakwit P, et.al [5] proposed a diamond recommendation system by mistreatment K-Means and cooperative Filtering techniques. He counselled the model system by that user mechanically thus on maximize users' satisfaction. The total system designed and developed within the type of automaton (android operative system). He divided the result by the analysis functions into a combination of parts: developing the Mobile application for diamond recommender users their evaluation and testing. The results showed that the specialists and users area unit proud of the system at an honest level.

III. PROBLEM STATEMENT

The admin is allowed to add and create new applications along with the links to the actual app in the play or app store. A set of data is collected for that specific application from both the stores and saved in the database from a specific period of time. Several data pre-processing methods are used in order to clean the data which has been given by the user. As in the architecture, it can be logically visualized with the tokenization, stop word removal and stemming algorithms being used. Here the user's comments and reviews are stored in the database act as the input to the algorithm. Now the number of positive and negative words that appear in reviews are counted. If the number of positive word appearances is greater than the number of negative word appearances, the system returns a positive sentiment, and vice versa. If the numbers are even, the system will return a neutral sentiment. Now the training set will be fitted to the SVM classifier. To create the SVM classifier, we will import SVC class from Sklearn. svm library. we have used kernel='linear', as here we are creating SVM for linearly separable data. However, we can change it for non-linear data. And then we fitted the classifier to the training dataset (x_train, y_train). The model performance can be altered by changing the value of C(Regularization factor), gamma, and kernel. Predicting the test set result Now, we will predict the output for test set. For this, we will create a new vector y_pred. After getting the y_pred vector, we can compare the result of y_pred and y_test to check the difference between the actual value and predicted value. Now we will see the performance of the SVM classifier that how many incorrect predictions are there as compared to the Logistic regression classifier. To create the confusion matrix, we need to import the confusion matrix function of the sklearn library. After importing the function, we will call it using a new variable cm. The function takes two parameters, mainly y_true(the actual values) and y_pred (the targeted value return by the classifier).

FRONT-END: In front-end coding, we use React is for creating the web page. Material Ui is used to give proper design to page with CSS, it is used to control the layout of multiple web pages all at once and JavaScript is used for providing functioning to elements.

BACK-END: Python is used to create sentiments from the comments provided by user. As our project is based on Machine learning in which Sentiment Analysis is one of its subtract, Python is very powerful and manageable language as world is moving towards Machine Learning and other aspects of it.

IV. PURPOSED SYSTEM

- ❖ In this project, we propose the system by developing web application which help to detect fraud apps using sentiment comments and data mining.
- ❖ In today's era, because of speedy development at intervals the mobile technology and mobile devices, the applications i.e. mobile apps area unit being really fascinating and stylish conception. As there is sizable quantity of mobile Apps, ranking fraud is the troublesome consider front of the mobile App market.
- ❖ Ranking fraud is the term used for relating dishonest or suspicious activities having the intention of boosting up the Apps at intervals the standard list. In fact, App developer's area unit exploitation troublesome means oft for increasing their Apps sales.
- ❖ The main aim is to develop such system that understand ranking, rating and review behaviours for investigation review based totally evidences, rating based totally evidences and ranking based totally evidences then aggregation supported improvement to combine all the evidences for detection of fraud.
- ❖ For every users reviews will be fetched separately and analysed for positive negative rating. The overall sentiment analysis of each app is calculated and then the final predicted result is displayed showing the app may be fraud or not.

V. CONCLUSION

In this paper, a review of the existing techniques for both emotion and sentiment detection is presented. As per the paper's review, it has been analysed that the lexicon-based technique performs well in both sentiment and emotion analysis. However, the dictionary-based approach is quite adaptable and straightforward to apply, whereas the corpus-based method is built on rules that function effectively in a certain domain. As a result, corpus-based approaches are more accurate but lack generalization. The performance of machine learning algorithms and deep learning algorithms depends on the pre-processing and size of the dataset. Nonetheless, in some cases, machine learning models fail to extract some implicit features or aspects of the text. In situations where the dataset is vast, the deep learning approach performs better than machine learning. Recurrent neural networks, especially the LSTM model, are prevalent in sentiment and emotion analysis, as they can cover long-term dependencies and extract features very well. But RNN with attention networks performs very well. At the same time, it is important to keep in mind that the lexicon-based approach and machine learning approach (traditional approaches) are also evolving and have obtained better outcomes. Also, pre-processing and feature extraction techniques have a significant impact on the performance of various approaches of sentiment and emotion analysis.

VI. REFERENCES

- [1] Liu m., wu c, Zhao x. N., liny.c., 2015. "app Relationship calculation: an iterative process". Published by the IEEE computer society. Issue No.08 - vol.27, pp: 2049-2063 Lovins b.j.,1968. "development of a stemming algorithm". "mechanical translation and computational.
- [2] Aamir M. and Bhusry .M., "Recommendation System: State of the Art Approach". International Journal of Computer Applications, Volume 120 -No.12. 2015.
- [3] Beel, J., LangerS., Genzmehr M., Gipp B, Breitinger C., and Nü rnberger A.2013 "Research Paper Recommender System Evaluation: A Quantitative Literature Survey". In ACM Recommender System Conference (RecSys). ACM International Conference Proceedings Series(ICPS).
- [4] Kularbphettong. K, Somngam S., Tongsiric, and Roonrakwit P. 2014. "A recommender system using Collaborative Filtering and K-Mean Based on Android Application". Journal of theoretical &Applied Information Technology, Vol. 70 Issue 1,p90-94. 5p.