Fraud App Detection using Fuzzy Logic Model Based on Sentiment of Reviews

Monika Pandey[1], Prof. Tripti Sharma[2]

[1]Student, Department Of Computer Science & Engineering, RCET, Bhilai, Chhattisgarh, India
[2]Professore, Department Of Computer Science & Engineering, RCET, Bhilai, Chhattisgarh, India

---***---

Abstract:- *Fake behaviour is most popular in application stores like Google play store, Apple's application store, and so forth. The data in application stores, for example, graph rankings, user ratings, and user reviews, gives an exceptional opportunity to perceive user behaviour with mobile applications. There exists many of the fraud application detection tool which separates the fraud and genuine apps based on user comments, ratings etc. But this tool does not fits well because they do not consider sentiment of users which plays very important role while detection of fraud apps. In this paper we propose a noble mechanism using fuzzy logic system to detect fraud application based on user reviews. The user reviews are taken from application store and the sentiment score are calculated.*

***Key Words*: Sentiment, User Review, Fuzzy Logic System, Fake Review, Genuine Review, Fraud Application.**

1. INTRODUCTION

Presently a-days with crazy increment towards android mobiles the rage of mobile applications has additionally expanded. As indicated by the ongoing investigation, the quantity of utilization in Google play store, which is otherwise called Android Market developed from 1 million to 3.5 million [1] till December 2017. While, the quantity of utilization in Apple's play store are 2.2 million [2] from 2008 July to 2017 January. The application engineers attempt false component so that the application created by them ought to have high rank in the application leader board. Application leader board is the one which demonstrates the outline positioning of the well-known applications, and it is a route through which one can advance their mobile applications. High rank of the application in leader board leads not exclusively to the vast number of downloads of that application yet in addition million dollars in income. Thus, the designers of the application attempt different strategies to advance their applications like publicizing which encourages them to have higher rank in the application leader board. Nonetheless, rather than utilizing moral system to advance their applications, the application designers attempt exploitative intends to advance their applications which controls the graph positioning of the application in the leader board and consequently the application is positioned high in the leader board. This sort of deceptive component is by and large completed utilizing web water army‖. Web water armed force is a gathering of web professional writers who are paid to post online remarks with specific substance. In this way, this encourages the application designers to advance their applications utilizing counterfeit reviews and appraisals [3, 4].

Fig.-1

For the most part fraud recognizing frameworks characterizes reviews and evaluations of the applications into two gatherings i.e., good and bad. Be that as it may, a few reviews and rating are not ordered into their significant gatherings due to blended reviews. Along these lines, to legitimately arrange applications into their pertinent gathering we utilize fuzzy rationale component hitch chips away at all moderate potential outcomes between good The fuzzy rationale deals with the levels of conceivable outcomes of contribution to accomplish the unequivocal yield.

2. LITERATURE SURVEY

Liu M., Wu C, Zhao X. N., Lin Y.C. et al. [3] Proposed that due to convenience of giant quantity of apps, apps retrieval and recommendation of apps is incredibly necessary, after that there is a need of top quality of app retrieval and recommendation there ought to be an explicit app relationship calculation for recommendation. He finished that various strategies supported app's context, which may solely find whether or not 2 apps square measure downloaded, put in meantime or offer similar functions or not.

Mohammad Aamir and Mamta Bhusry et al.[4], explained a categorical assess of the sphere of recommender systems and Approaches for analysis of advice System to propose the advice methodology that may any facilitate to reinforce opinion mining through recommendations.

G.Angulakshmi and Dr.R.ManickaChezian et al.[5], discussed that the assorted tools and techniques of opinion mining, She

explained that Opinion mining may be a style of tongue process that may track the mood of the folks concerning any specific product by review. Opinion mining is additionally referred to as sentiment analysis attributable to giant volume of opinion that is made in internet resources obtainable on-line.

Vijay.B. Raut and D.D. Londhe et al[6]. discussed the ways for opinion extraction, classification, and summarisation. He additionally explained all approaches of summarization techniques, techniques utilized in method of opinion mining. He showed that approaches of machine learning and sentiment analysis of knowledge like motion-picture show, product, hotel etc., whereas lexicon based mostly approach is appropriate for brief text in micro-blogs, tweets, and comments knowledge on internet.

Kularbphettong. K, Somngam S., Tongsiri C , and Roonrakwit P et al.[7] proposed a diamond recommendation system by mistreatment K-Means and cooperative Filtering techniques. He recommended the model system by that user mechanically so as to maximize users' satisfaction. The total system designed and developed in the form of automaton (android operative system). He divided the result by the analysis functions into a pair of parts: developing the Mobile application for diamond recommender users their evaluation and testing. The results showed that the specialists and users are happy with the system at a decent level.

Beel, J., LangerS., Genzmehr M., Gipp B, Breitinger C., and Nürnberger A. et al[8], reviewed that of 176 publications and it showed that no agreement exists on a way to assess and compare analysis paper recommender approaches. He examined that ends up in the dissatisfactory state of affairs that despite the numerous evaluations, the individual strengths and weaknesses of the planned approaches stay for the most part unknown.

Nikolaos Polatidis and Christos Georgiadis et al.[9] explained the all mobile recommendation system technologies and gave the challenges long-faced the technologies. He additionally explained that mobile devices and for information retrieval various factors are crucial. He planned that development of recent Role-Based Access management (RBAC) model that improve privacy issues of permissions in apps and information accessed by user.

Varsha N. Patil and Nitin N. Patil et al.[10] have proposed a system for detect fraud app ranking using opinion analysis, The existing system consists of Mining leading session, Ranking based pieces of evidences, Rating based pieces of evidences, and Review based pieces of evidences. An efficient mechanism for review based evidence has been discussed in the paper. In this system applied opinion analysis algorithm for find the review opinion analysis. By using this algorithm they can get the proper result about the review of apps. Also, calculate ranking and rating score of apps given by user and combine all evidences and detect fraud apps, and here some

methods are studied like a Permutation-based model, score-based model etc.

Navdeep Singh, Prashant Kr. Pandey and Mr.Srinivasan et al.[11] have proposed a system for detect app fraud based on app rating, ranking and feedback or comments of app provided by users. In this paper session tracking is used to track user's every activity on browser side and manage the fake user download and rating of any application by user.

Harpreet Kaur, Veenu Mangat and Nidhi et al. [19] have presents a survey of sentiment analysis and classification algorithms.SVM and naïve bayes are most popular algorithms are for sentiment classification. Datasets from various sites like: Amazon, IMDB, flipkart are widely used for sentiment analysis. Also apply decision tree classifier and rule based classifier to classify the dataset.

3. METHODOLOGY

In this section we present proposed methodology in detail. Fig. 2. Shows the pictorial representation of complete process of sentiment based fraud application detection.

There are several modules in proposed system they are:

    A.   Different application user reviews dataset.

    B.   Three different sentiment dictionaries.

    C.   Data pre-processing.

    D.   Feature extraction.

    E.   Fuzzy classification and score calculation phase.
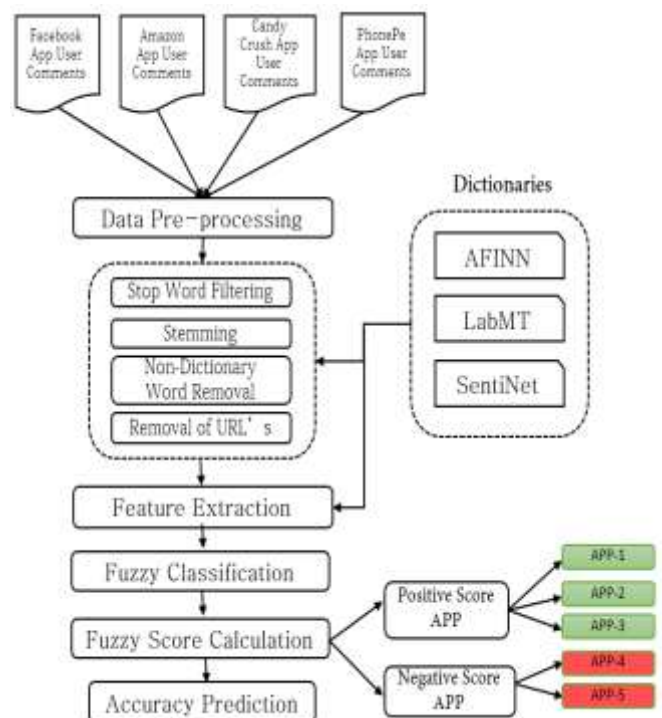

Fig.-1: Proposed System Architecture

A. User Review Dataset

We have collected datasets from Google play store. The reviews are manually copied. We have collected user reviews for 5 different application class:

1. Social

2. Shopping

3. Game

4. Finance

5. News

B. Dictionary

We have utilized sentiment dictionary for processing of user reviews. This sentiment dictionary contains the sentiment score of all the English words. The three dictionary used are:

1. AFINN

2. SentiWord Net

3. LabMT

1. AFINN

AFINN is a list of English words containing collection of words which are rate between -5 to +5 based on sentiment.

2. SentiWord Net

SentiWord Net is a lexical resource for opinion mining. SentiWord Net assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity.

3. LabMT

It is the collection of word along with its happiness rank. The dataset represents the sentiment of each word in the integer form.

C. Data Preprocessing

This steps involves the processing of user reviews for removal of noises and unwanted text. It includes,

1. Stop Word Filtering: This removes the stop words from the reviews. The stop words does not convey any sentiment. The stop words like the, is, what, when, where etc.

2. Stemming: Stemming removes the suffixes and prefixes from the word. It does not have any sentiment hence can be removed.

The suffixes like: -ies, -ing, -al, -ism, -ness.

The prefixes like: ante-, circum-, co-, dis-, ex-.

3. Non-Dictionary Word Removal: This phase is important because the processing of non-English word will lead to wrong sentiment.

4. Removal of URL's: The URL are also removed because it does not contain the sentiment.

D. Feature Extraction

In this phase the 3 dictionary are used for sentiment score calculation. The noun, verbs, adjective are collected and tokenize in order to get the sentiment score from dictionary.

We have also utilized the TF-DF for calculation of repeated term in document. The equation is:

$$tf(t,d) = 0.5 + \frac{0.5 * f_{t,d}}{max\{f_{t',d} : t' \in d\}}$$

E. Fuzzy Classification and Score Calculation Phase

In this phase we now proceed to calculate the sentiment score of each review. In earlier stage we have utilized sentiment calculation of only single words.

For fuzzy score calculation we have utilized three dictionaries which are AFINN, SentiWord Net and LabMT. The score are calculated using below equation:

$$f(\mu(s)) = 1 - (1 - \mu(s))^{\delta}$$

Where, $\mu(s)$ is SentiWord, AFINN and LabMT features and $\delta = 1/2$ as constant.

4. RESULT

In this section experimental results are presented. The dataset used are Google play store user review dataset. The snapshot of the dataset is shown in Fig. 3-6.
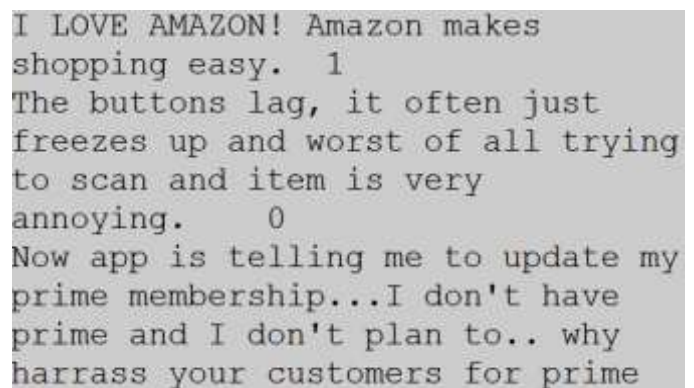

Fig.-3: Shows the Amazon dataset snapshot

Fig.-4: Shows the Candy Crush dataset snapshot

Fig.-5: Shows the Facebook dataset snapshot

Fig.-6: Shows the PhonePe dataset snapshot

Framework Output
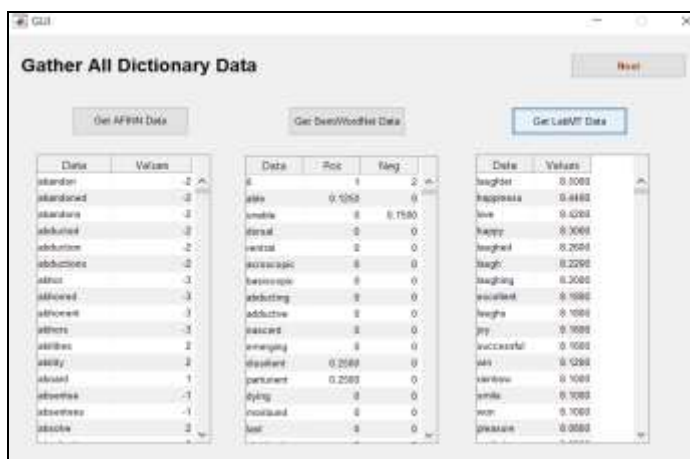
STEP 01: Getting all the dictionary data

Fig.-7: Dictionary Data

In Fig. 7, AFINN, SentiWord Net and LabMT dictioanry dataset are collected. It contains the sentiment score of all English words.

STEP 02: Gather all Review Dataset
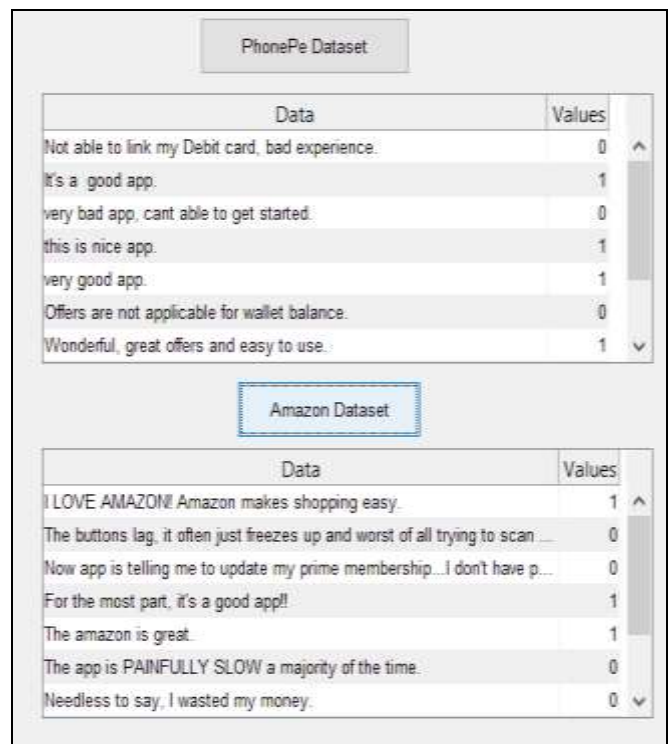

Fig.-8: Review Dataset

Fig.-9: Review Dataset

In Fig. 8 and 9 shows different categories of dataset are collected.

STEP 03: Pre-processing and Training

This phase pre-process and trained the dataset. The fuzzy system model is used for trianing of comments.
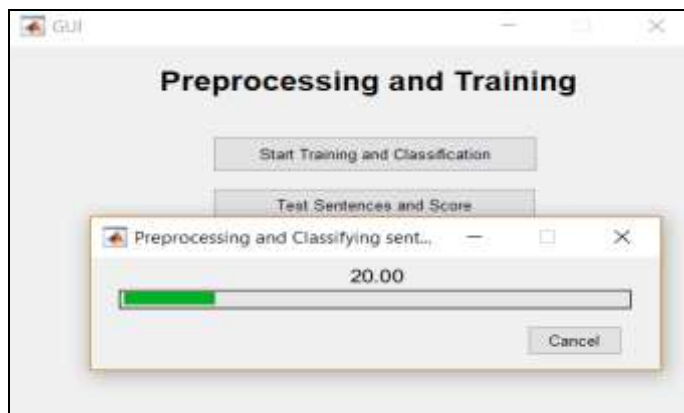

Fig.-10: Pre-processing and Training

Fig. 10. Show Pre-processing and trainning the datasets.

STEP 04: Accuracy of proposed system

Table 1 shows the accuracy obtained by proposed model.

Table-1: Accuracy of proposed system

SNO	Dataset Name	Number of Instances	Accuracy
1	Facebook	10	70 %
2	Amazon	10	90 %
3	PhonePe	10	100 %
4	Candy Crush	10	70 %
5	Aaj Tak	10	100 %
6	Hike	10	90 %
7	Ludo King	10	100 %
8	PayTm	10	70 %
9	Snapdeal	10	80 %
10	TOI	10	60 %
11	OVERALL ACCURACY	100	83 %

Table-1, show that Facebook and candy crush datset achives 70% of accuracy while Amazon and PhonePe achives 90% and 100% of accuracy respectively. There are more apps with more than 90% accuracy.
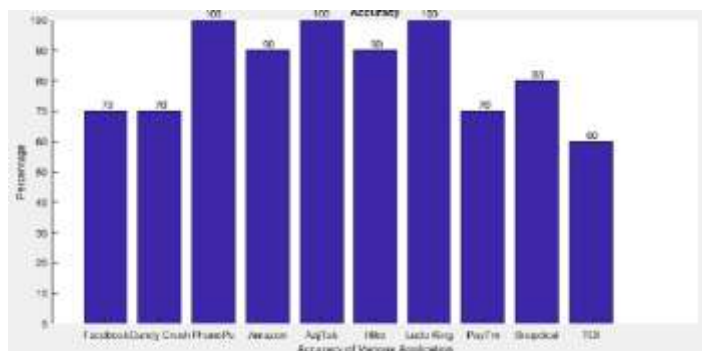

Chart-1: Accuracy of Applications.

STEP 05: Fraud application prediction

In this the fraud and genuine apps are classified.

Table-2: Classification of Geuine and Fraud Apps

SNO	Application Name	Positive Review	Negative Review	Fraud / Genuine
1	Facebook	8	2	GENUINE
2	Amazon	5	5	FRAUD
3	Candy Crush	8	2	GENUINE
4	PhonePe	6	4	GENUINE
5	Aaj Tak	4	6	FRAUD
6	Hike	5	5	GENUINE
7	Ludo King	9	1	GENUINE
8	PayTm	6	4	GENUINE
9	Snapdeal	2	8	FRAUD
10	TOI	2	8	FRAUD

Table 2 show classification of fraud and genuine apps.

5. CONCLUSION

The main purpose of this paper is to present the working model of fuzzy sentiment based fraud application detection. We have utilized several pre-processing and feature extraction process which removes noises and extract relevant feature based on the dictionary. It also helps in extraction the user review score through which fraud and genuine applications are classified.

The proposed system gains average accuracy of 83% with 5 different types of dataset. We can successfully able to classify the applications from different datasets.

REFERENCES

1. The mentioned statistics of google play store. Online.Available:https://www.statista.com/statistics/266210/number-ofavailable-applications-in-the-google-play-store/

2. The mentioned statistics of apple's app store. Online. Available:https://www.statista.com/statistics/276623/number-ofapps-available-in-leading-app-stores/

3. Liu M., Wu C, Zhao X. N., Lin Y.C., 2015. "APP Relationship Calculation: An Iterative Process". Published by the IEEE Computer Society. Issue No.08 - vol.27,pp: 2049-2063.

4. Mahmudur Rahman, Mizanur Rahman, Bogdan Carbunar, and Duen Horng Chau, — "Search Rank Fraud and Malware Detection in Google Play", IEEE Transactions on Knowledge and Data Engineering, Vol. 29, No. 6, June 2017, pp. 1329-1342.

5. Angulakshmi G and ManickaChezian R. "An Analysis on Opinion Mining: Techniques and Tools". International Journal of Advanced Research in Computer and Communication EngineeringVol. 3, Issue 7, 2014.

6. Vijay B. Raut and Prof. D.D. Londhe — "Surve On Opinion Mining And Summarization Of User Review On Web", IJCSIT International Journal Of Computer Science And Information Technologies, vol. 5(2), 2014,pp. 1026-1030.

7. Kularbphettong. K, Somngam S., Tongsiri C , 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.

8. Beel, J., LangerS., Genzmehr M., Gipp B, Breitinger C., and Nürnberger A.2013 "Research Paper Recommender System Evaluation: A Quantitative Literature Survey" . In ACM Recommender System Conference (RecSys). ACM InternationalConference Proceedings Series (ICPS).

9. Nikolaos Polatidis, "Mobile Recommender Systems: An Overview of Technologies and Challenges" ,published in Informatics and Applications (ICIA),2013 Second International Conferencepp. 282 -287, 2013.

10. Varsha A. Patil and Nitin N. Patil, — "Mobile Apps Opinion Analysis using Emoticon", IEEE International Conference on Global Trends in Signal Processing, Information Computing and Communication, 2016, pp. 203-207

11. Navdeep Singh, Prashant Kr. Pandey and Mr. Srinivasan, — "Improved Discovery of Rating Fake for Cellular Apps", IEEE International Conference on Science Technology Engineering and Management (ICONSTEM), 2016, pp. 135-140.

12. Hengshu Zhu, Hui Xiong, Yong Ge, and Enhong Chen, — "Discovery of Ranking Fraud for Mobile Apps", IEEE Transactions on Knowledge and Data Engineering, Vol. 27, No. 1, January 2015, pp. 74-87.

13. Jing Wan, Mufan Liu, Junkai Yi and Xuechao Zhang, — "Detecting Spam Webpages through Topic and Semantics Analysis", IEEE Global Summit on Computer and Information Technology (GSCIT), 2015, pp. 83-92.

14. Hengshu Zhu, Chuanren Liu, Yong Ge, Hui Xiong and Enhong Chen, — "Popularity Modeling for Mobile Apps: A Sequential Approach", IEEE Transactions on Cybernetics, Vol. 45, No. 7, July 2015, pp. 1303-1314.

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

16. Liu M., Wu C, Zhao X. N., Lin Y.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 Linguistics", vol.11, nos.1 and 2

17. Polatidis, N., & Georgiadis, C. K. 2013. "Mobile recommender systems: An overview of technologies and challenges". In Informatics and Applications (ICIA), 2013 Second International Conference on (pp. 282-287). IEEE

18. Jovian Lin, Kazunari Sugiyama, Min-Yen Kan, and Tat-Seng Chua, "Addressing Cold-Start in App Recommendation: Latent User Models Constructed from Twitter Followers" , Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 283-292, 2013.

19. Harpreet Kaur, Veenu Mangat and Nidhi, — "A Survey of Sentiment Analysis techniques", International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2017, pp. 921-923.