



JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Detection of Fraud Applications using Sentiment Analysis

¹Prof.Swati Khokale, ²Rutuja Kakuste, ³Priyanka Thombare, ⁴Rutvik Patil, ⁵Shubham Pawar

¹Professor, ^{2,3,4,5}Students

^{1,2,3,4,5} Bachelor of Engineering (Information Technology),

^{1,2,3,4,5} Sandip Institute of Technology and Research Centre (SITRC), Nashik, MH, INDIA

Abstract: Analysis With the widespread use of mobile applications in everyday life, it is critical to keep track of which ones are safe and which are not. The reviews that are mentioned for each application cannot be used to determine how safe and true the application is. As a result, it is necessary to keep track of and develop a system to ensure that the apps available are genuine or not. The goal is to create a system that detects fraudulent apps before the user downloads them using sentimental analysis and data mining. Sentimental analysis is used to help determine the emotional undertones of words expressed online. This method is useful for monitoring social media and getting a quick sense of the public's opinion on specific issues. on the internet, the user may not always find accurate or true product reviews. We can search for sentimental comments from users across multiple applications. The reviews could be fake or real. We can determine whether the app is genuine or not by analysing the rating and reviews together, which include both user and admin comments. The machine can learn and analyse sentiments, emotions, and other texts using sentimental analysis and data mining. One of the most important aspects of App ranking fraud is the manipulation of reviews. Analyzing reviews and comments using sentimental analysis and data mining might help in finding the best app for both the Android and iOS platforms.

Index Terms - Sentiment analysis, fraud app detection, Machine Learning, data mining

I. INTRODUCTION

As we all know, the mobile application market is growing as mobile users increase in number. Sensible phone users employ those alternatives of mobile apps for entertainment, knowledge, and so on. Android phones have recently become popular, and consumers make extensive use of the play store. According to a recent survey, the number of applications in the Google Play store has increased from 1 million to 3.5 million. Because of its quick rise in daily usage, sales, and advances, it has become a serious problem in the area of business intelligence. This increases market competition. Companies and application developers are competing fiercely with one another to demonstrate the quality of their goods and to attract clients in order to support their future advancement.

In comparison, there are 2.2 million apps in Apple's App Store. The review area of every app on the Play Store could potentially be an honest way to research an app. Fraud applications may cause phone damage as well as data theft. However, once downloaded, all mobile consumers prefer high hierarchical apps. To transfer a program on a smartphone, the user must go to a play store such as Google Play or Apple's App Store. When a user visits the Play Store, he will be able to access the various application lists. Sometimes the user does not believe the appliance, regardless of whether the applications are useful or useless. When a user downloads an application, he or she may discover that it does not work or is ineffective. This indicates that there is fraud in the mobile application list. As a result, such programs must be labeled so that customers of the Google Play store can avoid them. Our application will provide the user reviews for the application that they wish to download. This could be a technique for developers to identify their vulnerabilities and improve the development of a new one while keeping the demands of the people in mind. Not only that, but criminal developers may deceive the recognition of their apps or utilize it as a platform to distribute malware everywhere. This is commonly accomplished by applying so-called "bot ranches" or "human water armed forces" to rapidly extend the Application downloads evaluations and audits.

II. LITERATURE REVIEW

In [5] focuses on analysing reviews in written English language belonging to various KSA telecommunication companies for opinion mining, and they used supervised machine learning algorithms for classification. Furthermore, they used TF-IDF (Term frequency-inverse document frequency) to determine the importance of a specific word in a review. [6] creates a sentiment analysis approach based on public Arabic reviews and Facebook post comments.

They used supervised machine learning algorithms such as Support Vector Machine (SVM) and Nave Bayes, as well as binary model (BM) and TF-IDF, to investigate the effect of various term weighting functions on sentiment analysis accuracy. In [7], they used natural language analysis for Arabic language text and applied sentiment analysis to a Twitter dataset of 4700 for Saudi dialect sentiment analysis with (k=0.807). Researchers in [8] presents a sentiment analysis for Egyptian dialect based on a corpus of reviews

and product reviews. They use natural language processing to comprehend Egyptian dialect. They also classified the data using a lexicon-based classification.

Fraud Detection and Prevention Using Machine Learning Algorithms: A Review, G. Jaculine Priya; S. Saradha, 2021 - In this paper AI and ML techniques are used to monitor key patterns that might help differentiate a real vs fraud transaction. Capturing Customer information like Geolocation, authentication, session, and device IP address can be maintained. Machine Learning and the application of Artificial Intelligence will play an important part in learning and detecting fraud patterns automatically.

Towards De-Anonymization of Google Play Search Rank Fraud, Mizanur Rahman; Nestor Hernandez; Bogdan Carbutar; Duen Horng Chau, 2021 - The author of this research introduces the fraud de-anonymization challenge, which goes beyond fraud detection to reveal the human masterminds behind search rank fraud on peer-review sites. The author gathered and analysed data from crowdsourced search rank fraud tasks, as well as surveyed the capabilities and behaviours of 58 search rank fraud workers recruited from six crowdsourcing platforms. The author gathered a gold standard dataset of Google Play user profiles assigned to 23 crowdsourced workers and investigated their fraudulent behaviour in the field. Dolos is a fraud de-anonymization system that uses attributes and behaviours extracted from studies to attribute identified fraud to crowdsourcing site workers, and hence to real identities and bank accounts.

III. PROPOSED METHODOLOGY

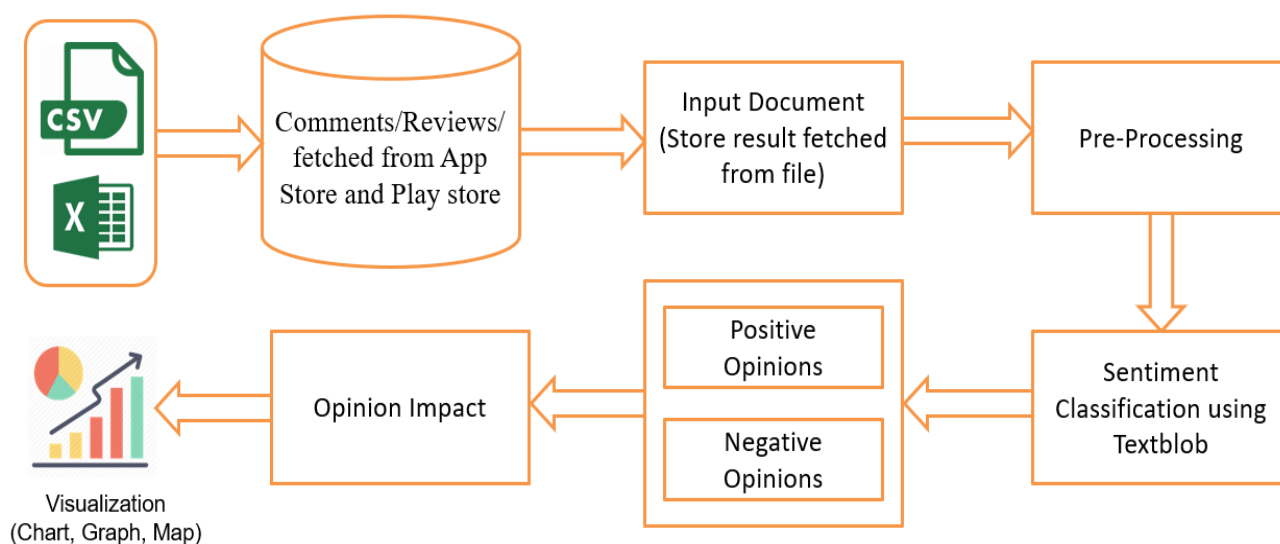


Figure 1 - System Architecture

Most of us now use Android and IOS phones, and we frequently access the play store or app store. Both marketplaces provide a huge number of applications, but unfortunately, a small number of those applications are fraudulent. Such programs can cause phone harm and data theft. As a result, such programs must be labeled so that store users can identify them. As a result, we propose a web application that will process the information, comments, and application evaluation. As a result, determining if an application is fraudulent will be easy.

The practice of determining a user's emotion toward a topic or a product is known as sentiment analysis. Opinion mining determines whether a user has a positive, negative, or neutral opinion of a product, topic, or event. Opinion mining and summarization are comprised of three primary steps: opinion retrieval, categorization, and summarization. Review websites are used to collect the review text. Opinion material, such as that found in blog posts, reviews, and comments, conveys subjective information about a topic. Positive and bad reviews are separated. By considering frequently recurring features of a topic, opinion summaries are constructed based on feature opinion sentences. The complete framework designed for sentiment analysis is as follows:

1. Read CSV file – In this step, we are accessing a CSV file that is exported from the app store/play store. The file contains the app name, review, username, date and time, etc.

2. Pre-process or Clean data

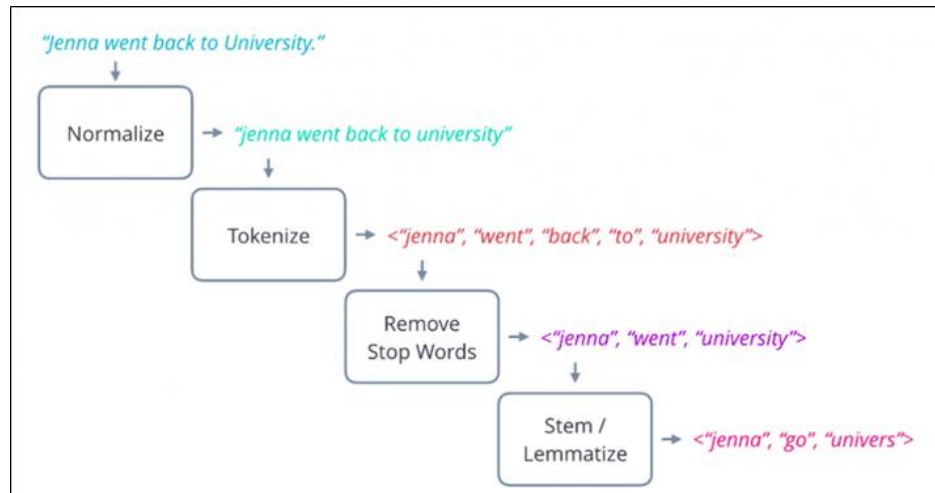


Figure 2 - data pre-processing

- Tokenization - Tokenizing sentences to break the text down into sentences, words, or other units by detecting whitespace in words.
 - Removing stop words - Removing stop words like “if,” “but,” “or,” and so on.
 - Lemmatization - Normalizing words by condensing all forms of a word into a single form
3. Perform Sentiment Analysis – In this step, we are checking the sentiment of a given review using the TEXT BLOB library.

Loves the German bakeries in Sydney. Together with my imported honey it feels like home	Positive
@VivaLaLauren Mine is broken too! I miss my sidekick	Negative
Finished fixing my twitter...I had to unfollow and follow everyone again	Negative
@DinahLady I too, liked the movie! I want to buy the DVD when it comes out	Positive
@frugaldougal So sad to hear about @OscarTheCat	Negative
@Mofette brilliant! May the fourth be with you #starwarsday #starwars	Positive
Good morning thespians a bright and sunny day in UK, Spring at last	Positive
@DowneyisDOWNEY Me neither! My laptop's new, has dvd burning/ripping software but I just can't copy the files somehow!	Negative

Figure 3 - Sentiment Analysis

4. Visualization – After successfully detecting sentiment for the given review we are showing the outcome in the form of a bar graph and word cloud.

IV. PROJECT STEPS

1. Start
2. Enter Review / Upload CSV file
3. Read App reviews from CSV file
4. Perform Tokenization
5. Remove stop words from reviews
6. Lemmatization
7. Calculate Sentiment for each review using Text Blob Library
8. Categorize review – Positive, Negative, and Neutral
9. Visualize the Result
10. End

V. CONCLUSION

The main objective of the proposed work was to review the fraud detection of apps and to use a sentiment analysis approach to differentiate the particular fraud apps. The experimental analysis is carried out on differing types of apps with the proposed method for the detection of fraud apps. Our system will detect ranking frauds using Natural Language Processing and machine learning techniques. Further, an optimization-based aggregation method combines all three pieces of evidence to detect fraud. Different class value and threshold value gives different results of accuracy of your time required for execution.

REFERENCES

- [1] Levy, M. (2016). Playing with Twitter Data. [Blog] R-bloggers. Available at: <https://www.r-bloggers.com/playing-with-twitter-data/> [Accessed 7 Feb. 2018].
- [2] Popularity Analysis for Saudi Telecom Companies Based on Twitter Data. (2013). Research Journal of Applied Sciences, Engineering and Technology. [online] Available at: <http://maxwellsci.com/print/rjaset/v6-4676-4680.pdf> [Accessed 1 Feb. 2018].
- [3] [3] Zhao, Y. (2016). Twitter Data Analysis with R – Text Mining and Social Network Analysis. [online] University of Canberra, p.40. Available at: <https://paulvanderlaken.files.wordpress.com/2017/08/rdataminingslides-twitter-analysis.pdf> [Accessed 7 Feb. 2018].
- [4] [4] Alrubaiee, H., Qiu, R., Alomar, K. and Li, D. (2016). Sentiment Analysis of Arabic Reviews in e-Learning. Journal of Computer Science. [online] Available at: <http://thesaipub.com/PDF/jcssp.2016.553.563.pdf> [Accessed 7 Feb. 2018].
- [5] [5] Qamar, A., Alsuhibany, S. and Ahmed, S. (2017). Sentiment Classification of Twitter Data Belonging to Saudi Arabian Telecommunication Companies. (IJACSA) International Journal of Advanced Computer Science and Applications, [online] 8. Available https://thesai.org/Downloads/Volume8No1/Paper_50Sentiment_Classification_of_Twitter_Data_Belonging.pdf [Accessed 1 Feb. 2018].
- [6] [6] R. M. Duwairi and I.Qarqaz, "A framework for Arabic sentiment analysis using supervised classification", Int. J. Data Mining, Modelling and Management, Vol. 8, No. 4, pp.369-381, 2016.
- [7] [7] Hossam S. Ibrahim, Sherif M. Abdou, Mervat Gheith, "Sentiment Analysis For Modern Standard Arabic And Colloquial", International Journal on Natural Language Computing (IJNLC), Vol. 4, No.2, pp. 95-109, April 2015.
- [8] [8] Assiri, A., Emam, A. and Al-Dossari, H. (2016). Saudi Twitter Corpus for Sentiment Analysis. International Journal of Computer and Information Engineering, [online] 10. Available at: <http://waset.org/publications/10003483/saudi-twitter-corpus-for-sentiment-analysis> [Accessed 1 Mar. 2018].
- [9] [9] L. Wasser and C. Farmer, "Sentiment Analysis of Colorado Flood Reviews in R", Earth Lab, 2018. [Online]. Available: <https://earthdatascience.org/courses/earth-analytics/get-data-usingapis/sentiment-analysis-of-twitter-data-r/>. [Accessed: 01- Mar- 2018].
- [10][10] D. Robinson, "Text analysis of Trump's reviews confirms he writes only the (angrier) Android half", Variance explained, 2016.
- [11][11] "A Common Database Interface (DBI)", cran.r-R, 2003. [Online]. Available: <https://cran.rproject.org/web/packages/DBI/vignettes/DBI-1.html>. [Accessed: 25- Mar- 2018].
- [12][12] V. Kharde and S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications, vol. 139, p. 11, 2016.
- [13][13] Product Quality Assessment using Opinion Mining in Persian Online Shopping Fatemeh HosseinzadehBendarkheili; Rezvan MohammadiBaghmolaei; Ali Ahmadi 2019 27th Iranian Conference on Electrical Engineering (ICEE) Year: 2019 | Conference Paper | Publisher: IEEE
- [14][14] News Comments Modeling for Opinion Mining: The Case of Senegalese Online Press Lamine Faty; Marie Ndiaye; Edouard Ngor Sarr; Ousmane Sall; Sény Ndiaye Mbaye; Tony Tona Landu; Babiga Birregah; Mamadou Bousso 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE) Year: 2020 | Conference Paper | Publisher: IEEE
- [15][15] Chinese Explanatory Opinion Relationship Recognition Based on Improved Target Attention Mechanism Xue Cao; Chenghao Zhu; Chengguo Lv 2020 3rd International Conference on Advanced Electronic Materials, Computers, and Software Engineering (AEMCSE) Year: 2020 | Conference Paper | Publisher: IEEE
- [16][16] Opinion Mining Using Live Twitter Data Andleeb Aslam; Usman Qamar; Reda Ayesha Khan; Pakizah Saqib; Aleena Ahmad; Aiman Qadeer 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC) Year: 2019 | Conference Paper | Publisher: IEEE