JOURNAL OF MULTIDIMENSIONAL RESEARCH & REVIEW

http://www.jmrr.org Volume: 6, Issue: 1, April 2025 | ISSN: 2708-9452

Enhancing Web App Protection With Machine Learning

Sublakshmi Priya C

Assistant Professor,PG & Research Department of Computer Sciences and Applications Vivekanandha College of Arts and Sciences for Women [Autonomous],Tiruchengode, Namakkal, Tamilnadu, India.

Fathima Roshini B

MCA Student,PG & Research Department of Computer Science and Applications Vivekanandha College of Arts and Sciences for Women [Autonomous],Tiruchengode Namakkal, Tamilnadu, India.

Abstract

Programmes that mimic the capabilities of actual, authentic, and trustworthy programmes are known as fake Web applications. These programmes engage in dangerous behaviours after they are installed, like aggressively showing advertisements to make money, intercepting sensitive data from your system, infecting devices, and so on. People commonly check the app's user evaluations before installing it because the majority of users can't tell the +framework for analysing apps based on user input. Utilising sentiment analysis, these comments are examined. Sentiment is an emotion or attitude brought on by the customer's sentiments. Opinion mining is another name for sentiment analysis since it uses user reviews to determine how well-liked an app is. Sentiment analysis is a function of machine learning. Sentiment analysis gathers and analyses the opinion or sentiment of the statement using natural language processing (NLP). It is well-liked since many individuals like seeking assistance from other users. The primary goal of the project is to identify genuine fraud apps using machine learning and the NLTK tool. The suggested approach classifies applications and determines whether they belong to a favourable or bad group.

Keywords: Web application security, Machine learning(ML)¸Anomaly detection, threat detection , AI drive security, Intrusion detection system(IDS).

1 Introduction

The rise in online fraud serves as a driving force behind the use of sentiment analysis and machine learning approaches to identify fraud in framework. As ecommerce, social networking, and other web-based platforms have grown, fraud has as well, including bogus reviews, phoney accounts, and exploitation of personal data. These actions can hurt consumers and businesses alike and erode confidence in the online community. Because it enables the study of text data, including customer reviews and social media posts, sentiment analysis in particular is a potent method for spotting fraud in web services. It is feasible to spot patterns and abnormalities that can point to fraud by examining the emotional state of this text data. For instance, a deep learning can be trained to identify specific terms or words that are frequently connected to false reviews, flagging any reviews that use them. Because machine learning enables the study of massive volumes of data in real-time, it is also a successful method for identifying fraud in web services. Machine learning makes it possible to examine network traffic and user behaviour to look for fraudulent activity. For instance, a machine learning model can be trained to spot user behaviour patterns, including click patterns or mouse movements that might point to fraudulent activities. Overall, by providing a more robust way to examine text data and other types of data, employing sentiment analysis and machine learning techniques to detect fraud in online applications can increase the effectiveness of fraud detection systems. This can enhance confidence in the online community while defending consumers and companies against fraud.

2 Literature Review

The main advantages of the sentiment analysis system are scalability as it can summarize large quantities of text, realtime analysis as it can generate results at run time, and consistent criteria as it is automated and free from bias compared to humans. A sentiment analysis system is also an essential tool for private and government organizations. Sentiment analysis can be used to improve traditional recommendation systems. It is helpful for manufacturers as it gives the sentiment orientation of customers about their products. It can also be used for market research and competitive analysis. Other domains of application include politics, government policy-making, investigation of legal matters [2]. Sentiment analysis may be performed at different levels of detail, aspect level sentiment analysis being the most informative one. Detection of explicit aspects is explored widely by researchers, and a variety of approaches are suggested. On the other hand, due to its complexity, less attention is given to detecting implicit aspects. A significant proportion of the text contains implicit aspects making detection of implicit aspects important for sentiment analysis. Many survey papers discuss and analyze different sentiment analysis approaches, but a handful of them have discussed aspect-level sentiment analysis. The following table lists these papers and their specifics. The importance of implicit aspect detection for sentiment analysis and the discussed limitations of the existing survey papers have motivated us to write a paper focusing on only implicit aspect detection[**1**].
In this study, the researchers carry out a comprehensive review of existing studies on spam review detection using the Systematic Literature Review (SLR) approach. Overall, 76 existing studies are reviewed and analyzed. The researchers evaluated the studies based on how features are extracted from review datasets and different methods and techniques that are employed to solve the review spam detection problem. Moreover, this study

analyzes different metrics that are used for the evaluation of the review spam detection methods. This literature review identified two major feature extraction techniques and two different approaches to review spam detection. In addition, this study has identified different performance metrics that are commonly used to evaluate the accuracy of the review spam detection models. Lastly, this work presents an overall discussion about different feature extraction approaches from review datasets, the proposed taxonomy of spam review detection approaches, evaluation measures, and publicly available review datasets. Research gaps and future directions in the domain of spam review detection are also presented. This research identified that success factors of any review spam detection method have interdependencies. The feature's extraction depends upon the review dataset, and the accuracy of review spam detection methods is dependent upon the selection of the feature engineering approach. Therefore, for the successful implementation of the spam review detection model and to achieve better accuracy, these factors are required to be considered in accordance with each other. To the best of the researchers' knowledge, this is the first comprehensive review of existing studies in the domain of spam review detection using SLR process[**2**].

3 Methodology

3.1 Project Overview

The app is said to fraud on the basis of 3 parameters: Ranking, Rating & Review of the app. In ranking based we check the historical ranking of the app, there are 3 different ranking phases, rising phase, maintaining phases & recession phase. The apps ranking rising to peak position on leader board (i.e. rising phase), to keep at the peak position on the leader board (i.e. maintaining phase), & finally decreasing till the end of event (i.e. recession phase). Their views are taken from the dataset and are converted into tokens on which sentiment analysis is performed. The rise in online fraud serves as a driving force behind the use of sentiment analysis and machine learning approaches to identify fraud in framework. As e-commerce, social networking, and other web-based platforms have grown, fraud has as well, including bogus reviews, phoney accounts, and exploitation of personal data. These actions can hurt consumers and businesses alike and erode confidence in the online community. Because it enables the study of text data, including customer reviews and social media posts, sentiment analysis in particular is a potent method for spotting fraud in web services.

3.2 System Overview

Recommender Systems are indispensable to provide personalized services on the Web. Recommending items which match a user's preference has been researched for a long time, and there exist a lot of useful approaches. First, discuss existing Collaborative Filtering methods with explicit feedbacks. Collaborative Filtering with explicit feedbacks that both positive and negative feedbacks are observed in the dataset. The Collaborative Filtering methods can be divided into the memory-based method, the model based method and the combination of the two. The memory-based method includes the Neighborhood method, which calculates the similarity of the users or items. The model-based method includes the Matrix Factorization model, the Probabilistic model and Cluster based model. The biggest problem in Collaborative Filtering is the sparseness of observed values. It means

feedbacks are observed in very small portion of all possible user-item pairs. However the Matrix Factorization model is known to work better than other models even if the data is sparse.

4 Existing System

Traditional web application security relies on rule-based systems and signaturebased detection, which have limitations in detecting new and evolving threats. Below are the key components of existing systems:
Web Application protection (WAP)
WAFs filter and monitor HTTP requests to block common attacks like SQL Injection (SQLi) and Cross-Site Scripting (XSS).They rely on predefined rules but struggle with zero-day attacks and adaptive threats.
Intrusion Detection and Prevention Systems (IDPS)
IDPS monitors network traffic for known attack patterns.These systems use signature-based detection, which is ineffective against new or obfuscated threats. **Disadvantages**

- Only analyzed ratings from user reviews.
- Fake reviews can't be analyzed by existing work.
- User can't be identify genuine reviews.
- Handle only limited number of product reviews.

5 Proposed System

A recommendation system has been implemented based on hybrid approach of stochastic learning and context based engine. We have tried to combine the existing application for recommendation to come up with a hybrid one. It improves the performance by overcoming the drawbacks of traditional recommendation systems. Recommender systems being a part of information filtering system are used to forecast the bias or ratings the user tends to give for an item. Among different kinds of recommendation approaches, collaborative filtering technique has a very high popularity because of their effectiveness. These traditional collaborative filtering systems can even work very effectively and can produce standard recommendations, even for wide ranging problems. For item based on their neighbor's preferences entropy based technique creates better suggestions than others. Whereas other techniques like content based suffers from poor accuracy, scalability, data sparsity and big-error prediction. To find these possibilities we have used user-based collaborative filtering approach. In this Item based filtering technique we first examine the User item rating matrix and we identify the relationships among various items, and then we use these relationships in order to compute the recommendations for the user. Then using cosine similarity which is a similarity weight is going to play an important role in the item based filtering approach and hence in order to maintain or select the trustable users from the given set of user. Hence they give us a method to increase or decrease the significance of a particular user or item. In the present methodology we are using adjusted similarity for computation of similar weights of items. **Advantages**

- System helps the user to find out correct review of the product.
- Handle large number of contextual information.
- User easily buy genuine products.
- Recommend the positive products based on user reviews.
- Automatic decision making system in product recommendation.

6 Result and Discussion

The proposed system demonstrates significant improvements in detecting fraudulent web applications and providing accurate recommendations through the integration of sentiment analysis and hybrid recommendation techniques. By combining machine learning algorithms with natural language processing using the NLTK toolkit, the system effectively analyzes user reviews to identify patterns indicative of fake or malicious behavior. The use of a hybrid recommendation approach, which integrates stochastic learning and context-aware collaborative filtering, overcomes common limitations found in traditional systems such as data sparsity, cold start problems, and lack of scalability. In terms of performance, the system was evaluated using several metrics including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), precision, recall, and F1-score. The results showed a notable reduction in MAE and RMSE, indicating higher accuracy in predictions. Additionally, higher precision and recall values confirm the system's ability to correctly classify genuine versus fraudulent applications, while an improved F1-score demonstrates a balanced and effective classification performance. The use of cosine similarity and adjusted similarity metrics enhanced the accuracy of item-based filtering by refining the weight calculations and improving the selection of trustworthy users and reviews. The incorporation of entropy-based ranking techniques further contributed to more reliable and prioritized recommendations, leading to better user engagement. The system's ability to process and adapt to real-time data changes ensures that it remains effective even as user preferences and threat patterns evolve. As a result, users receive more relevant and personalized suggestions, while being protected from deceptive or harmful applications. Overall, the discussion highlights how leveraging multiple filtering techniques and optimization strategies can significantly enhance the reliability and performance of recommendation systems. While the current implementation yields promising results, there is potential for further improvement by incorporating advanced deep learning models such as LSTM or BERT to deepen contextual understanding in sentiment analysis. Additionally, expanding the dataset and integrating real-time behavioral tracking could further strengthen the system's fraud detection capability. In conclusion, the hybrid model presents a robust and scalable solution that effectively balances accuracy, efficiency, and user safety in web application environments.

7 Conclusion

In this project, proposed an application recommendation system based on deep learning algorithm. The main advantages of method are a visual organization of the data based on the underlying structure, and a significant reduction in the size of the search space per

result output. And user can easily search the products anywhere and anytime. Ratings, reviews and emoticons are analyzed and categorized as positive and negative sentiments. Search the products based on price based filtering and reviews based filtering. The current results are notably better than proposed approach. However, feel that with a better dataset and a number of improvements to method, may achieve better results. Hybrid Recommendations is one of the main modules of the system which helps overcome the drawbacks of the traditional Collaborative and Content Based Recommendations. And have obtained promising results using current model.

References

[1] P. K. Soni, "A Survey on Implicit Aspect Detection for Sentiment Analysis: Terminology, Issues, and Scope."

[2] N. Hussain, "Spam Review Detection Techniques: A Systematic Literature Review."

[3] B. Heinold, "A practical introduction to Python programming," 2021.

[4] R. T. Kneusel, "Practical deep learning: A Python-based introduction," No Starch Press, 2021.

[5] A. J. Dhruv, R. Patel, and N. Doshi, "Python: the most advanced programming language for computer science applications," Science and Technology Publications, Lda, pp. 292–299, 2021.

[6] J. Sundnes, "Introduction to scientific programming with Python," Springer Nature, 2020.

[7] C. Hill, "Learning scientific programming with Python," Cambridge University Press, 2020.

Cite this article:

Sublakshmi Priya C & Fathima Roshini B, "Enhancing Web App Protection With Machine Learning", Journal of Multidimensional Research and Review (JMRR), Vol.6, Iss.2, pp.146-151, 2025