International Journal of Engineering Research and Science & Technology

IJERST

ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

SENTIMENT ANALUSIS ON GOOGLE PLAY STORE DATA USING NLP

Korivi Dhana Lakshmi, **21PD1A0543**, Department of CSE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

Nayudu Surya Sree, 21PD1A0558, Department of CSE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

Yadavalli Sena Abhai Kumar, 21PD1A0586, Department of CSE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

Amudala Hari Hara Krishna , 22PD5A0501, Department of CSE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

Dr. M.Aravind Kumar,Professor, Department of ECE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

B.Raju, Assistant Professor, Department of CSE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

ABSTRACT

Sentiment analysis on Google Play Store reviews plays a vital role in understanding user feedback and enhancing app performance. This study applies Natural Language Processing (NLP) and machine learning models—Random Forest, Decision Tree, and Support Vector Machine (SVM)—to classify sentiments as positive, negative, or neutral. User reviews are preprocessed using tokenization, stopword removal, and vectorization. The models are evaluated based on accuracy, precision, recall, and F1-score, with Random Forest offering strong predictive power, Decision Tree providing interpretability, and SVM ensuring high classification accuracy. A comparative analysis highlights the most effective model, emphasizing the value of machine learning in supporting data-driven app development decisions.

INTRODUCTION

The exponential rise in mobile applications has led to a surge in user reviews on platforms like the Google Play Store, offering critical feedback for improving app performance and user satisfaction. However, manually analyzing this vast, unstructured data is inefficient, prompting the need for automated sentiment analysis—a key NLP technique that classifies opinions as positive, negative, or neutral. Machine learning algorithms such as Random Forest, Decision Tree, and SVM are leveraged for their predictive accuracy and pattern recognition, especially when combined with preprocessing techniques like tokenization, stopword removal, and vectorization. This study compares these models to identify the most effective approach for sentiment classification, ultimately enabling developers to make informed, data-driven enhancements that boost user engagement and app success.

EXISTING SYSTEM

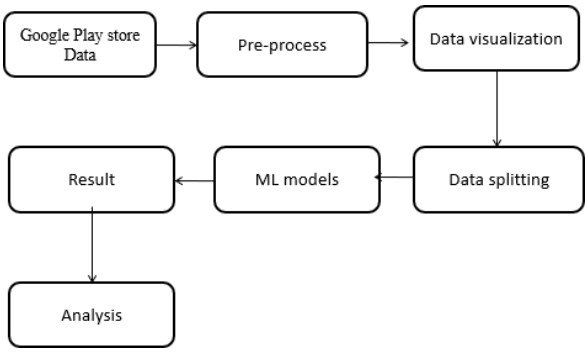
International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

This research employs the Valence Aware Dictionary and Sentiment Reasoner (VADER) to perform sentiment analysis on 7,267 user reviews from the Lazada app on the Google Play Store, collected between 2019 and 2022 using the Google-play-scraper Python script. VADER, a rule-based tool suited for short text and social media, uses a predefined lexicon to assign sentiment scores, classifying reviews as positive, negative, or neutral. The analysis revealed 4,229 positive, 2,857 negative, and 181 neutral reviews, reflecting an overall positive user experience. This lexicon-based approach helps extract actionable insights from user feedback, supporting businesses in assessing customer satisfaction and identifying improvement areas.

PROPOSED SYSTEM

The proposed sentiment analysis method for Google Play Store reviews utilizes Natural Language Processing (NLP) and machine learning techniques in a structured workflow. After collecting and preprocessing the data through text cleaning, tokenization, stopword removal, and TF-IDF vectorization, features are extracted to convert text into numerical form. Three classifiers—Random Forest, Decision Tree, and Support Vector Machine (SVM)—are employed to categorize sentiments as positive, negative, or neutral. Random Forest boosts accuracy through ensemble learning, Decision Tree offers interpretability, and SVM excels at defining optimal decision boundaries. Model performance is assessed using accuracy, precision, recall, and F1-score, enabling comparative analysis to identify the most effective model. This approach supports developers in making informed, data-driven decisions based on user feedback.

ARCHITECTURE



To perform sentiment analysis on Google Play Store data using NLP, you can start by collecting a dataset of app reviews from the Google Play Store, which can be done using web scraping tools or APIs like Google Play Scraper. After gathering the reviews, preprocess the text by removing stop words, special characters, and performing tokenization. Next, use Natural Language Processing techniques such as lemmatization or stemming to standardize the words. For sentiment analysis, you can apply pre-trained models like VADER (Valence Aware Dictionary and sEntiment Reasoner) or fine-tune a model using machine learning techniques such as Naive Bayes, Support Vector Machines, or deep learning models like LSTM. These models will classify the reviews into positive, negative, or neutral categories based on the sentiment conveyed. Finally, evaluate the model's performance using metrics like accuracy, precision, recall, and F1 score.

SOFTWARE REQUIREMENTS

1. Operating System : Windows

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

2. Version : Anaconda navigator3.9.2
3. Programming language : Python

HARDWARE REQUIREMENTS

1. Processor : i3core
2. Speed : 3.2GHZ
3. Ram : 4 GB(min)
4. Hard Disk : 512GB

CONCLUSION

This study underscores the importance of sentiment analysis for interpreting user feedback from Google Play Store reviews. Utilizing Natural Language Processing (NLP) and machine learning models like Random Forest, Decision Tree, and Support Vector Machine (SVM), reviews were classified into positive, negative, or neutral sentiments. The analysis revealed Random Forest as robust, Decision Tree as interpretable, and SVM as highly accurate, helping identify the most effective model. These insights enable app developers to make informed, data-driven improvements, highlighting the value of automated sentiment analysis in enhancing app functionality, user satisfaction, and performance.

FUTURE SCOPE

To enhance sentiment analysis, advanced deep learning models like LSTM and BERT can be implemented for improved prediction accuracy. Incorporating Aspect-Based Sentiment Analysis (ABSA) allows for evaluating sentiments tied to specific features such as app performance, user interface, and customer support. Expanding to multilingual sentiment analysis provides broader insights across diverse user groups. A hybrid model combining deep learning and machine learning techniques can further boost accuracy. Real-time sentiment analysis enables continuous monitoring of user feedback, offering instant insights to developers. Additionally, processing emojis, slang, and abbreviations enhances classification quality, while sentiment-based recommendation systems can suggest targeted improvements based on detected negative trends.

REFERENCES

1. **Pang & Lee (2008)** – Offered a comprehensive overview of opinion mining and sentiment analysis, laying the foundational theories, methods, and challenges in the field.
2. **Medhat, Hassan & Korashy (2014)** – Presented an extensive survey of sentiment analysis algorithms and applications, comparing various techniques across domains and datasets.
3. **Liu (2012)** – Delivered a detailed synthesis of sentiment analysis and opinion mining, including approaches, tools, and real-world use cases, with a focus on both subjective and objective information.
4. **Cambria et al. (2013)** – Explored emerging trends and future directions in sentiment analysis, advocating for a shift from traditional techniques to more concept-based approaches.
5. **Go, Bhayani & Huang (2009)** – Introduced a method for Twitter sentiment classification using distant supervision, paving the way for social media-based sentiment research.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

6. **Hirschberg & Manning (2015)** – Reviewed key advances in natural language processing (NLP), including developments that have significantly impacted sentiment analysis capabilities.