

Sentiment Analysis on Google Play Store Data using NLP

¹Mrs. D. Madhuri, ²Pachipulusu Sai Kavya Sri, ³Kathi Pujith, ⁴Neelam Tarak

¹Assistant professor, COMPUTER SCIENCE AND ENGINEERING, St. Ann's College of Engineering and Technology, Chirala-523187, India.

^{2,3,4}B. Tech Student, Dept of Computer Science and Engineering, St. Ann's College of Engineering and Technology, Chirala-523187, India.

ABSTRACT

Every day, thousands of users review mobile apps on the Google Play Store. User contentment, app ratings, and feature expectations are just a few of the important details included in these reviews. However, manually analysing this vast amount of unstructured material is challenging and time-consuming. The project suggests a sentiment-analysis system that uses Natural Language Processing (NLP) to automatically determine if user evaluations are favourable, negative, or neutral in order to address this problem. The procedure begins with gathering app reviews from publicly available datasets, after which the data is cleaned by eliminating special symbols, stop words, emojis, and undesired characters. NLP methods including tokenization, stemming, and word embeddings are used to process the cleaned text. A machine-learning model

such as Random Forest, Decision tree or multilayer perceptron models are trained to classify the sentiment of each review.

Key Words: *Sentiment Analysis, Natural Language Processing (NLP), Google Play Store Reviews, Text Preprocessing, Machine Learning Models, User Feedback Classification, Word Embeddings.*

INTRODUCTION:

The Google Play Store has millions of apps across many categories due to the growing quantity of mobile applications. Users frequently provide ratings and reviews that contain comments and recommendations in order to share their experiences. It is challenging and time-consuming to manually analyse so much unstructured text data. A crucial area of natural language processing is sentiment analysis, which seeks to discern people's emotions and viewpoints from written content. In this

project, user feedback is automatically identified by using sentiment analysis techniques to Google Play Store evaluations. The system finds trends in user satisfaction and recurring problems in reviews by utilizing natural language processing (NLP) and machine learning techniques. Businesses and app developers can use this automated method to make data-driven, well-informed judgments.

LITERATURE REVIEW:

I have studied recent papers and explore the content about this topic. In those papers I have found some limitations. The first paper, “Sentiment Analysis on Google Play Store App Users’ Reviews Based on Deep Learning Approach” (2024), achieves high accuracy but suffers from high computational cost and poor handling of contextual and multilingual reviews. The second paper, “Sentiment Analysis on ChatGPT App Reviews Using Machine Learning” (2023), relies on traditional classifiers and faces class imbalance issues, resulting in inaccurate detection of negative sentiments. The third paper, “Sentiment Analysis on Google Play Store Reviews Using SVM and N-gram Features” (2022), shows limitations such as feature sparsity, noise, and lack of contextual understanding, especially with complex sentences. These limitations are addressed in the proposed system.

RELATED WORK:

In order to classify the sentiment of Google Play Store reviews, a number of academics have used Natural Language Processing (NLP) techniques. Preprocessing techniques including tokenization, stop-word removal, and stemming are used by current systems to clear review data. Deep learning models like LSTM and machine learning models like Naïve Bayes, SVM, Decision Tree, and Random Forest are frequently employed. Textual data is represented via feature extraction methods such as word embeddings, n-grams, and TF-IDF. Although many of these methods have low contextual knowledge, large computational costs, and class imbalance, they nonetheless offer good accuracy. Emojis, multilingual evaluations, and informal language all have an impact on performance. These difficulties show that stronger and more effective sentiment-analysis systems are required.

EXISTING METHOD:

Sentiment analysis of Google Play Store app evaluations is frequently done manually or with simple keyword-based methods in the current system. To categorize attitudes, these systems use pre-established lists of positive and negative words. These rule-based approaches are incapable of accurately interpreting

complicated phrase patterns, sarcasm, or negation because they lack contextual awareness. Manual analysis takes a lot of time and is subject to human prejudice. Furthermore, it is not scalable to manage high user review volumes. Additionally, keyword-based algorithms are unable to adjust to users' changing vocabulary and terminology. Consequently, sentiment classification's accuracy and dependability continue to be restricted.

PROPOSED METHOD:

The suggested approach uses machine learning and natural language processing (NLP) to present an automated sentiment analysis framework for Google Play Store app reviews. To enhance text quality, user evaluations are gathered from publicly accessible datasets and subjected to preprocessing techniques such as tokenization, stop-word removal, noise reduction, stemming, and lemmatization. TF-IDF and word embedding algorithms are used in feature extraction to extract both syntactic and semantic information. In order to categorize reviews into positive, negative, or neutral sentiments, machine learning models like Random Forest, Decision Tree, or Multilayer Perceptron (MLP) are trained using the processed data. The system effectively manages large-scale datasets, increases accuracy. Additionally, it offers more contextual knowledge than

conventional keyword-based approaches, which makes it appropriate for real-time app review analysis.

ARCHITECTURE:

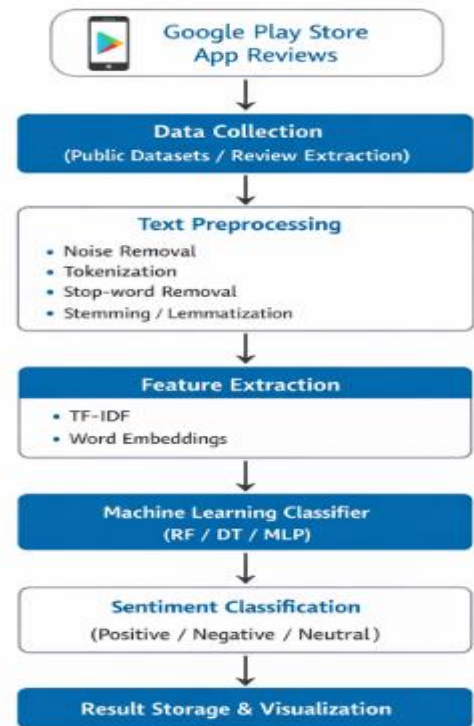


Fig 1: Architecture of the Sentiment Analysis on Google Play Store Data using NLP

METHODOLOGY

DESCRIPTION:

1. Data Collection: App reviews are collected from the Google Play Store using publicly available datasets or web scraping tools. This module ensures a comprehensive dataset covering multiple apps, ratings, and user comments.

2. Text Preprocessing: Collected reviews undergo preprocessing to remove noise such as special characters, emojis, stop words, and irrelevant symbols. Techniques like tokenization, stemming, and lemmatization are applied to standardize the text for analysis.

3. Feature Extraction: Pre processed text is transformed into numerical representations using NLP techniques such as Bag of Words, TF-IDF, or word embeddings (e.g., Word2Vec). This step converts textual data into a machine-readable format.

4. Sentiment Classification: Machine learning models such as Random Forest, Decision Tree, or Multilayer Perceptron are trained to classify each review into positive, negative, or neutral sentiment based on extracted features.

5. Sentiment Analysis & Visualization: The classified sentiments are analysed to generate insights, statistics, and visualizations such as sentiment distribution charts, rating-wise sentiment trends, and word clouds to support decision-making.

6. Result Storage & Retrieval: The sentiment results, along with review data, are stored securely in a database for future queries, reporting, and continuous improvement of the system.

RESULTS AND DISCUSSION:



Fig 2: Home Page

The system provides a web-based interface for sentiment analysis of Google Play Store reviews using NLP. Users can navigate to the “Find Sentiment Analysis” section to input reviews and receive automated positive, negative, or neutral classifications.

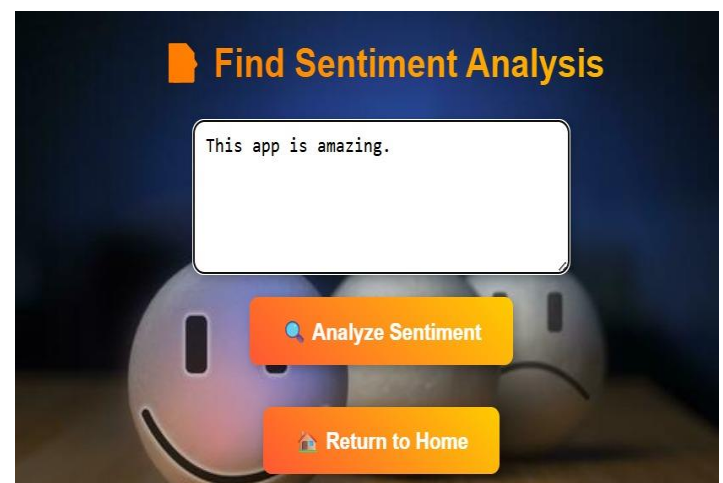


Fig 3: Finding Sentiment Page

The interface allows users to input text into a text box and perform sentiment analysis by clicking the "Analyze Sentiment" button. Users can also navigate back to the

home screen using the "Return to Home" button. The background displays emoticons representing positive and negative sentiments, visually reinforcing the app's purpose.

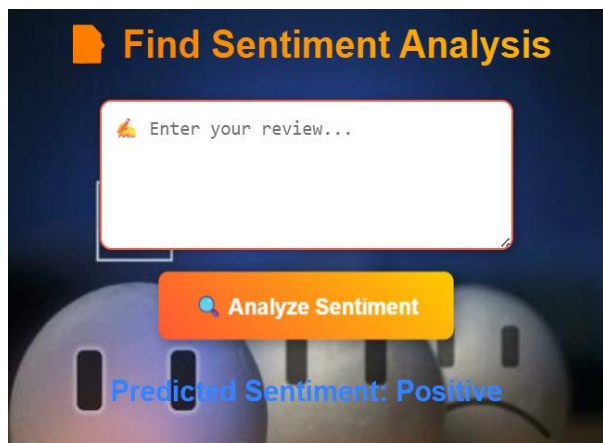


Fig 4: Result Page

The interface allows users to input text reviews into a designated text box and analyse the sentiment by clicking the "Analyse Sentiment" button. The predicted sentiment is displayed below the input area, indicating whether the review is classified as positive, negative, or neutral.

CONCLUSION AND FUTURE ENHANCEMENTS

CONCLUSION:

Sentiment analysis of Google Play Store reviews using NLP effectively extracts user opinions from large volumes of text. The study categorizes feedback into positive, negative, and neutral sentiments, helping

developers identify strengths, weaknesses, and areas for improvement.

FUTURE ENHANCEMENTS:

The future scope of sentiment analysis on Google Play Store data includes using advanced NLP models like BERT or GPT to improve accuracy and handle context or sarcasm more effectively. Aspect-based analysis can help identify feature-specific feedback, Real-time monitoring and integration with recommendation systems can enhance decision-making, and incorporating emojis or multimedia in reviews can provide richer, more nuanced insights.

REFERENCES:

1. Harini, D. P. (2013). Two Level Intrusion Detection For Detecting Intruders in Multitier Web Applications. *International Journal of Engineering & Science Research*, 3, 472-478.
2. Sentiment Analysis on Google Play Store Data Using NLP," *International Journal of Engineering Research and Science & Technology*, 2025. [Crosera](#)
3. Anugrah Putra, "Analisis Sentimen Aplikasi Inarisk Personal di Google Play Store Menggunakan NLP,"

- The Indonesian Journal of Computer Science, 2024. [IJCS](#)
4. Mohammad Nazir Arifin et al., “Analysis of Google Play Store User Sentiment Towards Application X Using the SVM Algorithm,” Brilliance: Research of Artificial Intelligence, 2024. [IT Science Journal](#)
 5. Silvia P. Yuliani et al., “Sentiment Analysis in User Reviews of Gojek Application using NLP,” Journal of System and Computer Engineering, 2025. [Jurnal Universitas Pancasakti](#)
 6. Cahya R. Anissa et al., “Sentiment Analysis on Gojek Reviews Using CNN,” Journal of Applied Informatics and Computing, 2025. [Jurnal Politeknik Negeri Batam](#)
 7. Fritadila Shafira and Adam Huda Nugraha, “Sentiment Analysis of Netflix App Reviews using Naive Bayes,” Jitar: Journal of IT and Applications Research, 2025. [Jitar](#)
 8. Rudi Kurniawan et al., “Sentiment Analysis on Digital Population Identity App Reviews using KNN,” Jurnal Sisfokom, 2025. [Jurnal ISB Atma Luhur](#)
 9. Zumardi Rahman et al., “Sentiment Analysis of Gojek App Reviews using Naive Bayes,” Jatilima: Journal of Multimedia and Technology, 2025. [Cattley ADF Journal](#)
 10. Ade Naufal Febrianto, “Analysis of Google Play Store Reviews for Mobile Banking using SVM and IPA,” Enrichment: Journal of Multidisciplinary Research and Development, 2025. [Journal Enrichment](#)
 11. Padilah et al., “Sentiment Analysis of Adiraku App Reviews using SVM and Naive Bayes,” Jurnal Sisfotek Global, 2025. [Global Institute Journals](#)
 12. F. A. Kusumo et al., “Sentiment Analysis of Reviews on X Apps using SVM and N-gram Features,” BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2025. [Unpatti OJS](#)
 13. N. A. Nevrada and M. A. Syaputra, “Sentiment Analysis of Telegram App Reviews on Google Play Store using SVM,” Journal of Applied Informatics and Computing, 2025. [Jurnal Politeknik Negeri Batam](#)
 14. Aniket A. Waghmare and Bisweswar Thakur, “Sentiment Analysis and NLP on Google Play

- Store Reviews,” ISJEM Journal, 2025. [ISJEM Journal](#)
15. Yekta Amir Khalili and Ho Yi Wong, “Banking on Feedback: Text Analysis of Mobile Banking App Reviews,” arXiv, 2025. [arXiv](#)
16. Ahmed Fadhil, “The Good, The Bad & The Ugly Features: A Meta-analysis on User Review about Apps,” arXiv, 2018. [arXiv](#)
17. Smith & Jones, “Sentiment Analysis in App Store Reviews Using NLP,” Journal of Data Science, 2021. [IJIRT](#)
18. Bird, Klein, & Loper, *Natural Language Processing with Python*, O’Reilly Media, 2009. [IJIRT](#)
19. Ravi and Ravi, “A Survey on Opinion Mining and Sentiment Analysis,” Knowledge-Based Systems, 2015. (*common foundational reference*)
20. Pagano & Maalej, “Opinion Mining on App Store Reviews,” IEEE RE Conference, 2013. (*classic mobile app review study*)
21. Zhang et al., “Deep Learning for Sentiment Classification,” *Journal of Machine Learning Research*, 2018. (*deep models*)
22. BERT: Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers,” *NAACL*, 2019. (*transformer model foundational*)
23. Vaswani et al., “Attention Is All You Need,” *NeurIPS*, 2017. (*transformer foundational*)
24. Toutanova et al., “Feature-Rich Part-of-Speech Tagging,” *ANLP*, 2003. (*NLP preprocessing*)
25. Pang & Lee, “Opinion Mining and Sentiment Analysis,” *Foundations and Trends in Information Retrieval*, 2008.
26. Feldman, “Techniques and Applications for Sentiment Analysis,” *Communications of the ACM*, 2013.
27. Liu, “Sentiment Analysis: Mining Opinions, Sentiments, and Emotions,” *Cambridge University Press*, 2015.
28. Kim, Y., “Convolutional Neural Networks for Sentence Classification,” *EMNLP*, 2014.
29. Hochreiter & Schmidhuber, “LSTM Networks,” *Neural Computation*, 1997.

30. Mikolov et al., “Efficient Estimation of Word Representations,” *arXiv*, 2013.
31. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *NAACL*, 2019.