

Midocean University

College of Informatics

Department of Artificial Intelligence



Project Proposal

Sentiment Analysis for Mobile App Reviews:

A Comparative Study

Prepared by:

Sultan Khalaf Alotaibi (222000226)

Abdel Hafez Adeeb Alassar (251002398)

Supervised by:

Dr. Wessam Ahmed

Second semester 2025

Table of Contents

1. Introduction	3
2. Problem Statement	3
3. Research Objectives	4
4. Methodology	5
5. Proposed Framework Diagram	6
6. Evaluation Metrics	6
7. Tools and Technologies.....	7
8. References	8

1. Introduction

With the rapid expansion of mobile applications in the digital marketplace, user comments has emerged as a valuable source of data both for stakeholders and developers (Ballas et al., 2024). Out of all the different kinds of user-generated data, app reviews are crucial for influencing new downloads, forming user perception, and directing product enhancements for the business. Therefore, it is becoming increasingly important to be able to analyze these reviews in a systematic way (Afsheen Maroof et al., 2024).

The detection of sentiments from textual content is an area of natural language processing (NLP) that allows for the automated identification and classification of opinions expressed in text data (Adiningtyas & Auliani, 2024). Sentiment analysis helps extract useful information from massive amounts of user reviews by classifying the user reviews into positive, negative, and neutral sentiment (Sivakumar & Uyyala, 2021). This can help developers enhance overall user satisfaction, correct long-term issues, and understand how people perceive a product.

For unstructured text data analysis, processing, and interpretation, NLP methods provide robust methods. Unprocessed text input can be transformed by NLP models into meaningful structures suitable for machine learning as well as deep learning tasks using methods such as tokenization, lemmatization, and vectorization. These processes enable scalable and uniform analysis of feedback, well beyond human review capabilities (Qahar et al., 2024).

This project involves applying sentiment analysis to user reviews of the mobile game Mobile Legends: Bang Bang. By comparing different sentiment classification methods ranging from classical machine learning to deep learning architectures and transformer-based models. We want to determine the most efficient approaches to understanding user sentiment in reviews of mobile apps. This study's findings can provide useful insights to app developers, marketers, and UX designers wishing to improve app performance through feedback from users.

2. Problem Statement

Mobile app reviews are invaluable sources of feedback for developers, presenting information about user satisfaction and complaints as well as feature requests. Given the volume and unstructured nature of these reviews, manual analysis is futile and error-prone. Automatic sentiment analysis would enable the mining of trends and areas for improvement.

Existing methods have drawbacks despite their potential. The informal language, slang, and emojis used in the reviews are often ignored by traditional machine-learning models. Deep-learning and transformer-based methods promise extra accuracy, but they demand more computer resources and painstaking model tuning. What is more, very few comparative studies on mobile app review sentiment analysis exist.

This project will address these problems through evaluating and comparing several sentiment analysis models on a dataset of reviews for Mobile Legends to find the best techniques for extracting insight into user feedback.

3. Research Objectives

The primary goal of this study is to perform a comprehensive sentiment analysis on user reviews of the *Mobile Legends* app and compare the performance of various classification models.

Main Objectives:

- To develop a robust preprocessing pipeline for cleaning and preparing app review data.
- To apply and evaluate multiple sentiment analysis techniques, including traditional machine learning, deep learning, and transformer-based models.
- To compare the effectiveness of these models in classifying user sentiment accurately.
- To identify the most suitable approach for analyzing mobile app reviews based on performance and practicality.

Research Questions:

- Which sentiment analysis model performs best on mobile app reviews in terms of accuracy and reliability?
- How do traditional machine learning models compare to deep learning and transformer-based models in this context?
- What preprocessing steps contribute most to improving sentiment classification accuracy?
- Can insights from sentiment analysis help identify common user concerns and satisfaction drivers in mobile games?

4. Methodology

This research employs a systematic procedure in conducting sentiment analysis on mobile app reviews, integrating classical Natural Language Processing methods with both traditional and contemporary machine learning models. The proposed framework outline consists of the following key stages:

1. Data Collection

The dataset used in this study, titled *Mobile Legends Google Play Reviews*, comprises 52,000 user reviews collected between February and June 2025. It includes review text, ratings, timestamps, and other metadata, serving as the foundation for analysis.

2. Preprocessing

A comprehensive preprocessing pipeline will be implemented to enhance data quality. This includes removal of duplicates, irrelevant entries, punctuation, URLs, hashtags, emojis, and stop words. All text will be converted to lowercase, followed by tokenization and lemmatization. Reviews with insufficient or meaningless content will be filtered out to maintain dataset integrity.

3. Feature Extraction & Selection

For traditional machine learning models, we will use Bag-of-Words and TF-IDF with unigram representations. In deep learning models, one-hot encoded vectors will serve as input embeddings. This stage transforms raw text into structured numerical data suitable for model training.

4. Modeling

Multiple classifiers will be employed and compared. These include:

- **Machine Learning Models:** Random Forest, Logistic Regression, Linear SVC, Multinomial Naive Bayes, Bernoulli Naive Bayes
- **Deep Learning Models:** RNN, LSTM, GRU using Keras
- **Transformer-based Models:** BERT, RoBERTa, DistilBERT using PyTorch

Model performance will be evaluated using standard metrics such as accuracy, precision, recall, and F1-score. The goal is to identify the most effective approach for sentiment classification in mobile app reviews.

5. Proposed Framework Diagram

The proposed framework processes reviews by cleaning text, extracting features (like word embeddings or BERT tokens), and trains models using both traditional ML (like SVM) and deep learning (like LSTM or BERT). It evaluates performance to predict sentiment accurately. The approach combines simplicity (bag-of-words) with advanced techniques (transformers) for robust analysis. The diagram of the proposed framework is presented in Figure 1.

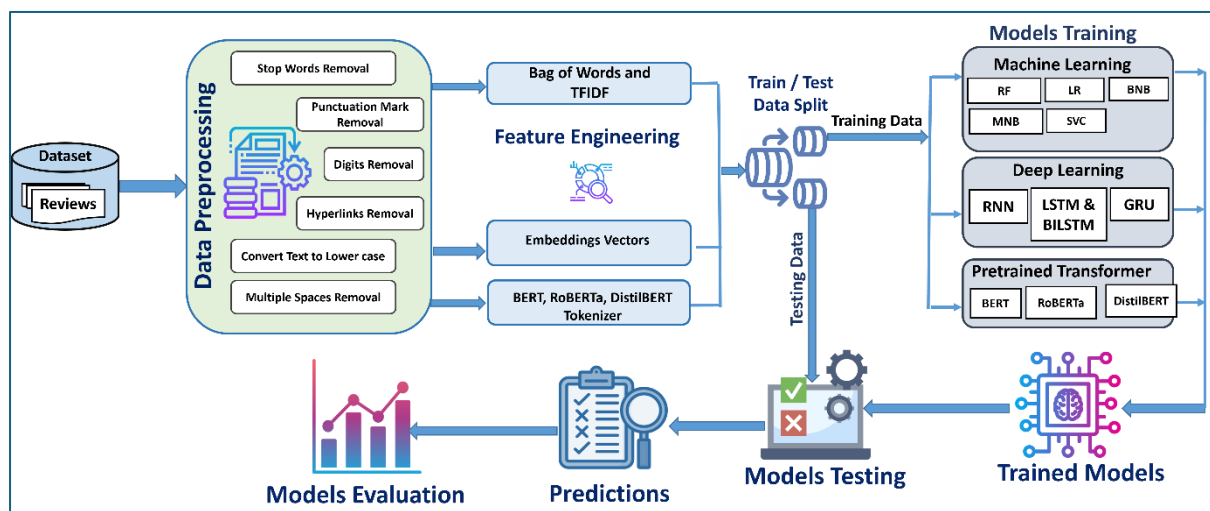


Figure 1: Proposed Framework

6. Evaluation Metrics

To evaluate sentiment classification models, this study will use common NLP metrics:

- **Accuracy:** How often the model is correct overall.
- **Precision:** How many predicted positives are actually correct.
- **Recall:** How many actual positives the model correctly finds.
- **F1-Score:** Balance between precision and recall.
- **Confusion Matrix:** Table showing correct vs. incorrect predictions for each class.

These metrics will be calculated for each model to compare their effectiveness in sentiment analysis of mobile app reviews.

7. Tools and Technologies

The project will utilize the following tools and technologies for implementation, experimentation, and analysis:

- **Programming Language:** Python, due to its strong ecosystem for data science and NLP.
- **Development Environment:** Jupyter Notebook for interactive development and Google Colab for leveraging GPU support in deep learning tasks.
- **Libraries and Frameworks:**
 - **NLTK:** For natural language preprocessing tasks such as tokenization and stop word removal.
 - **scikit-learn:** For implementing and evaluating traditional machine learning models.
 - **pandas & NumPy:** For efficient data manipulation and analysis.
 - **Keras (TensorFlow backend):** For building and training deep learning models like RNN, LSTM, and GRU.
 - **PyTorch & Transformers (HuggingFace):** For implementing and fine-tuning transformer-based models such as BERT, RoBERTa, and DistilBERT.
 - **matplotlib/seaborn:** For data visualization and performance plotting.

These tools will support the full lifecycle of the project—from data preprocessing and modeling to evaluation and visualization.

8. References

- Adiningtyas, H., & Auliani, A. S. (2024). Sentiment analysis for mobile banking service quality measurement. *Procedia Computer Science*, 234, 40–50. <https://doi.org/10.1016/j.procs.2024.02.150>
- Afsheen Maroof, Shaukat Wasi, Jami, S. I., & Siddiqui, M. S. (2024). Aspect-Based Sentiment Analysis for Service Industry. *IEEE Access*, 12, 109702–109713. <https://doi.org/10.1109/access.2024.3440357>
- Ballas, V., Konstantinos Michalakakis, Georgios Alexandridis, & Caridakis, G. (2024). Automating Mobile App Review User Feedback with Aspect-Based Sentiment Analysis. *Lecture Notes in Computer Science*, 179–193. https://doi.org/10.1007/978-3-031-60487-4_14
- Qahar, A., Yova Ruldeviyani, Ulfah Nur Mukharomah, Miftahul Agtamas Fidyawan, & Putra, R. (2024). Factor analysis influencing Mobile JKN user experience using sentiment analysis. *IAES International Journal of Artificial Intelligence*, 13(2), 1782–1782. <https://doi.org/10.11591/ijai.v13.i2.pp1782-1793>
- Sivakumar, M., & Uyyala, S. R. (2021). Aspect-based sentiment analysis of mobile phone reviews using LSTM and fuzzy logic. *International Journal of Data Science and Analytics*, 12(4). <https://doi.org/10.1007/s41060-021-00277-x>