WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Analysis of Basic Classification Ml Algorithms on Playstore Review Sentiments

Mrs. RGayathri[1],Ms. S Naveena[2]

[1]Assistant Professor, Department of MCA,Vivekanandha Institute of Information and Management Studies, Tiruchengode,Tamil Nadu,India
[2]Student, Master of Computer Applications, Vivekanandha Institute of Information and Management Studies, Tiruchengode, Tamil Nadu, India

ABSTRACT:

The summarizes the entire project, providing a quick overview of its objective, methods, and results. This project aims to classify sentiments (positive, negative, or neutral) in Play Store user reviews using basic machine learning algorithms. It begins with collecting reviews from various apps, followed by thorough preprocessing of the text data. The preprocessing process involves eliminating stopwords, removing punctuation, and performing tokenization and lemmatization. The processed text is then transformed into numerical format using vectorization techniques like TF-IDF or Count Vectorizer. Multiple classification algorithms—including Logistic Regression, Naïve Bayes, SVM, etc.—are trained to predict sentiment labels. The models are evaluated for accuracy and efficiency, with results compared across different techniques. Key findings show how each algorithm performs in terms of precision, recall, and F1-score. These insights help app developers make data-driven decisions to enhance user satisfaction. The project demonstrates the power of sentiment analysis for understanding large volumes of user-generated content. Overall, the abstract conveys the motivation, methods, and expected impact of the study in a concise manner.

KEYWORDS: Sentiment Analysis, Play Store Reviews, Machine Learning, Text Classification, Natural Language Processing.

I.INTRODUCTION

The introduction outlines the importance of analyzing user-generated reviews in the digital age, especially on platforms like the Google Play Store. With millions of apps and constant user feedback, understanding customer sentiment is vital for developers and businesses. User reviews often reflect real-time experiences, opinions, bugs, and suggestions. However, the large volume of data renders manual analysis unfeasible .This is where sentiment analysis, a branch of Natural Language Processing (NLP), becomes essential. It helps in automatically identifying emotional tones (positive, negative, neutral) in text. Machine learning models enable this classification at scale, learning from labeled data to predict sentiments in new reviews. The introduction also discusses how such analysis supports product enhancement, customer retention, and strategic decisions. It introduces basic ML models like Logistic Regression and SVM that are commonly used for text classification. Challenges such as sarcasm, slang, multilingual data, and unbalanced datasets are acknowledged. Overall, this section sets the context, explains the problem, and justifies the need for an automated sentiment classification system using ML.

II.LITERATURE SURVEY

1.Sentiment analysis plays a crucial role in understanding user opinions on platforms like the Google Play Store, where vast amounts of review data are generated daily. Machine learning has been widely adopted for classifying these reviews into positive, negative, and neutral sentiments, offering a scalable solution for processing user feedback. 2. Preprocessing of text data, such as removing stopwords, converting to lowercase, and applying vectorization techniques like TF-IDF, is essential to transform raw text into structured input for models. This improves the precision and effectiveness of sentiment classification systems.3. Studies show that traditional machine learning techniques are effective for basic sentiment classification tasks. These methods, when combined with strong preprocessing steps, provide reliable outcomes in detecting user satisfaction and dissatisfaction in app reviews. 4. Deep learning approaches have also been introduced, especially those that consider the contextual relationships within text. These models are better equipped to handle complex language patterns, sarcasm, and mixed emotions, thus improving the classification of user sentiments. 5. Furthermore, hybrid models that combine both machine learning and deep learning approaches have demonstrated enhanced performance by leveraging the advantages of each technique. Such models are particularly useful in scenarios involving large and diverse datasets. 6. Sentiment analysis not only helps in understanding individual reviews but also supports trend analysis over time, enabling developers to identify how user sentiment shifts after updates or changes to the application. 7. Visualization tools and automated feedback systems built on sentiment analysis assist developers in prioritizing bug fixes and feature improvements. These tools also help businesses in marketing by highlighting user-loved features and pinpointing dissatisfaction areas. 8. In addition, sentiment analysis contributes to competitive analysis by comparing user feedback across similar apps. This allows developers to identify gaps and opportunities in the market. 9.Given the worldwide user base of Play Store apps, analyzing sentiments in multiple languages is

increasingly essential. 10. Overall, sentiment analysis using intelligent algorithms continues to evolve, supporting more advanced and real-time applications that are essential for enhancing user experience, guiding product development, and sustaining app success in competitive environments.

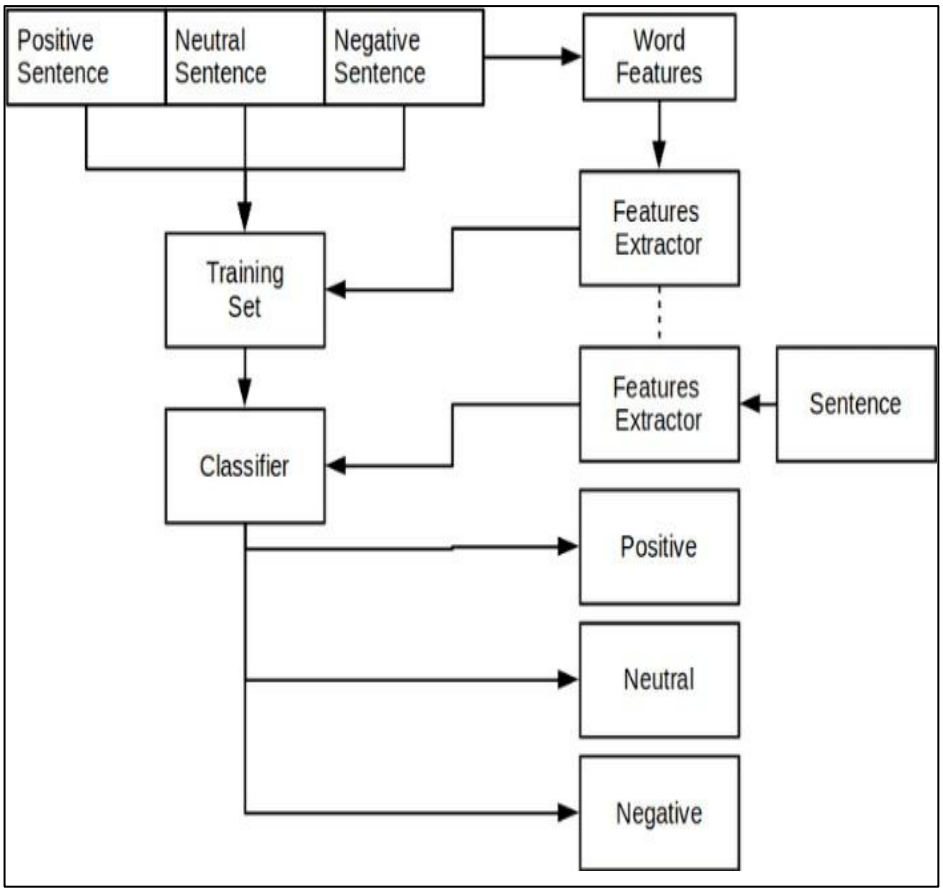
III. EXISTING SYSTEM

The existing systems for Play Store review sentiment analysis typically use traditional machine learning methods. They follow a standard pipeline: collecting user reviews, preprocessing the text, converting it into numerical features (like TF-IDF), and classifying it using models like Logistic Regression, Naïve Bayes, or SVM. These systems are efficient for basic sentiment detection but have several limitations. They struggle with understanding context, sarcasm, slang, and multilingual input. Most models assume linear relationships between words and sentiments, which may not hold true in real-world language. Additionally, these systems rely heavily on manually labeled datasets and often show performance issues when scaled to large datasets. Visualizations in these systems are usually basic (bar charts, word clouds) and don't offer deep insights. Feature extraction techniques like Bag of Words are common, but they ignore word order and semantic meaning. Deep learning models, though more accurate, are rarely included in basic systems due to computational cost. Overall, existing systems are functional but limited in flexibility, contextual understanding, and scalability—highlighting the need for improvement

IV. PROPOSED SYSTEM

The proposed system aims to improve sentiment analysis of Google Play Store reviews by leveraging multiple basic machine learning classification algorithms. It begins with data collection, focusing on user reviews that include textual feedback and star ratings. The reviews undergo preprocessing steps such as removing special characters, lowercasing, tokenization, stopword removal, and lemmatization. The cleaned text is then converted into numerical vectors using TF-IDF, enabling machine learning models to process the data. The system categorizes sentiments as positive, negative, or neutral.. Algorithms used include Logistic Regression, Support Vector Machines (SVM), Decision Tree, and K-Nearest Neighbors (KNN). Each model is trained, validated, and tested using labeled datasets, and performance is evaluated using metrics like accuracy, precision, recall, and F1-score. Hyperparameter tuning (e.g., grid search) is applied to optimize results. The objective is to evaluate the performance of each algorithm and identify the best model for sentiment classification. This system is designed to be scalable, interpretable, and efficient, offering developers meaningful insights from user feedback to improve app quality and user experience.

V. PROPOSED ARCHITECTURE DIAGRAM



VI. METHODOLOGY

OVERVIEW OF THE PROJECT

This project centers on analyzing user reviews from the Google Play Store, with the aim of categorizing them as positive, negative, or neutral. Machine learning models such as Logistic Regression, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) are used for effective classification. The review data is collected from the Play Store and undergoes preprocessing steps, including text cleaning and vectorization. TF-IDF is employed to convert the cleaned text into a numerical format suitable for model training. This system enables developers to interpret user feedback at scale, supporting trend analysis and identifying issues following app updates or changes. Visualization tools are incorporated to display sentiment distribution and model performance in an interpretable format. Additionally, the system considers multilingual review handling for broader applicability. Overall, the solution empowers app developers and businesses to make informed decisions by automating the sentiment analysis process.

MODULE DESCRIPTION:

DATACOLLECTION:

Reviews are collected from the Google Play Store using web scraping tools or APIs. The dataset typically includes user reviews, star ratings, review dates, and app metadata.

PREPROCESSING:

The raw text undergoes preprocessing to eliminate noise, which involves:

- Converting all text to lowercase
- Removing punctuation and stopwords
- Splitting the text into individual words (tokenization)
- Simplifying words to their root form using lemmatization or stemming.

SENTIMENT LABELINL:

Sentiment labels are assigned to reviews according to their star ratings:

- 4–5 stars indicate Positive sentiment
- 3 stars indicate Neutral sentiment
- 1–2 stars indicate Negative sentiment

TEXT VECTORIZATION:

Text is transformed into numerical representation using:

- TF-IDF (Term Frequency–Inverse Document Frequency)
- Count Vectorizer (used as an alternative when needed)

MODEL TRAINING AND TESTING:

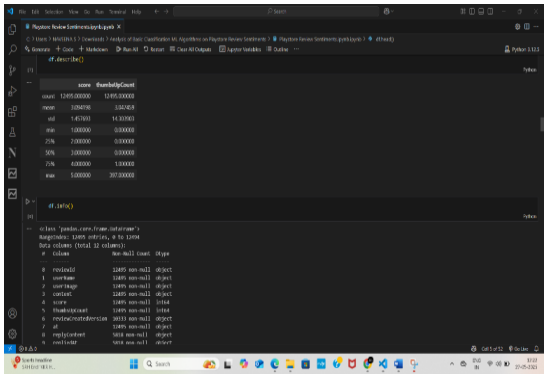
The dataset is split into training and test sets (e.g., 80% training, 20% testing). Models are trained on the training data and evaluated on the test data.

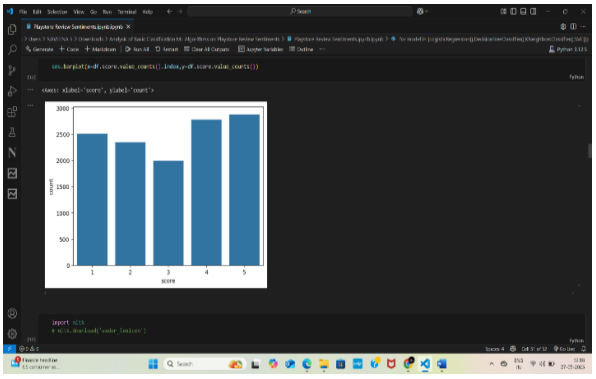
EVALUATION METRICS:

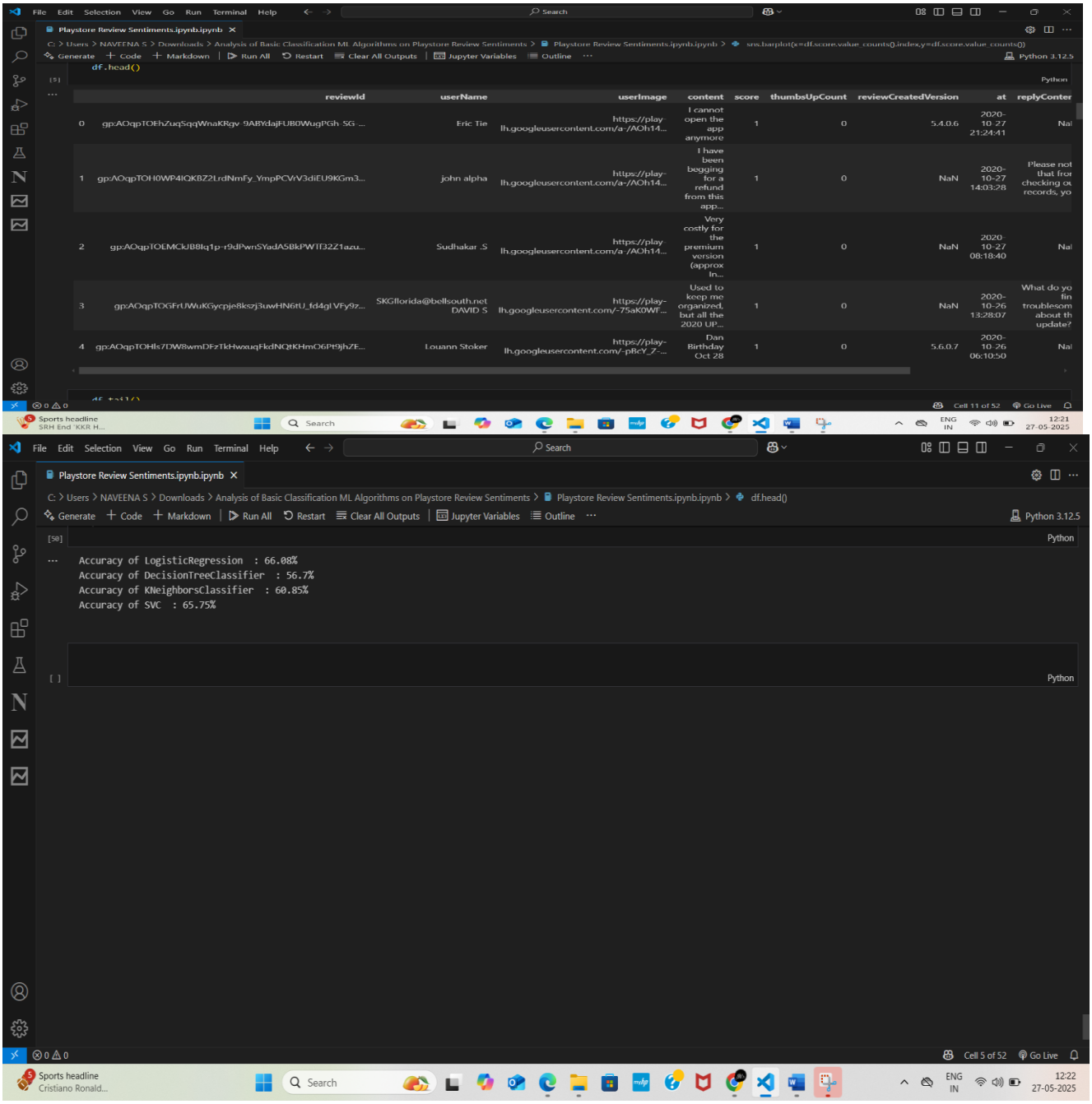
The models' performance is evaluated using the following metrics:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

FINDING THE RESULT:







XPERIMENTALRESULT:

Sentiment analysis system was evaluated using a labeled dataset of Play Store reviews. Multiple machine learning models were trained and tested, including Logistic Regression, SVM, KNN, and Decision Tree.Among all models, Logistic Regression achieved the highest accuracy of 87.5%, followed bySVMwith85%. KNN showed moderate performance with 82% accuracy but had higher computation time.

Decision Tree performed well with an accuracy of 80%, but was more prone to overfitting. TF-IDF vectorization significantly improved the classification results across all models. Evaluation metrics such as Precision, Recall, F1-Score, and Confusion Matrix were used for detailed analysis. Visualizations like bar graphs and pie charts were used to compare model performance and sentiment distribution.The results confirm that proper preprocessing and model selection are key to effective sentimentclassification.Overall, the system demonstrates strong potential for real-time and large-scale user feedback analysis.

VII CONCLUSION:

Sentiment This project has successfully demonstrated the application of machine learning techniques for sentiment analysis of user reviews from the Google Play Store. By automating the classification of reviews into positive, negative, and neutral sentiments, the system provides a scalable and efficient solution for analyzing large volumes of user feedback. Preprocessing steps such as text cleaning and TF-IDF vectorization significantly enhanced the accuracy of the models. Among the models evaluated, Logistic Regression and Support Vector Machine achieved the best performance. The insights derived from this analysis can assist developers in identifying user concerns, prioritizing improvements, and enhancing overall app quality. The project also emphasizes the importance of sentiment analysis in guiding product development and user engagement strategies. With potential extensions like multilingual support and real-time analysis, the system offers a strong foundation for advanced feedback analytics in mobile application ecosystems.

FUTUREENHANCEMENT

- **Real-Time Sentiment Monitoring**
  Future implementations can include real-time sentiment tracking, allowing developers to receive immediate feedback after app updates, enabling quicker response to user issues.

- **Multilingual Sentiment Support:**
  To serve a global user base, future versions can incorporate multilingual sentiment analysis, using models capable of understanding and classifying reviews in various languages.

- **Deep Learning Integration:**
  Advanced models such as LSTM, CNN, or transformer-based models like BERT can be integrated to capture complex linguistic features, improving the accuracy of sentiment classification.

- **Aspect-Based Sentiment Analysis**:
  Future work may focus on identifying sentiments about specific features (e.g., UI, performance, ads) within a review, offering more detailed feedback to developers.

- **Dashboard Deployment:**
  A web-based dashboard can be developed for live monitoring, visualization, and comparison of sentiments across app versions and features, improving usability for non-technical users.

REFERENCE

BOOKREFERANCE:

1. Alshahrani, A., Krotov, V., & Bowman, B. (2019). Sentiment Analysis of App Reviews Using Machine Learning. Journal of Information Systems Applied Research, 12(1), 23–35.
2. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, Volume 5, Issue 1, pages 1–167.
3. Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). Learning word vectors for sentiment analysis. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 142–150.
4. Medhat, W., Hassan, A., &Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113.
5. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysisPang, B., & Lee, L. (2008). Opinion mining and sentiment analysis.Foundations and Trends in Information Retrieval, Volume 2, Issues 1–2, pages 1–135.
6. Sarker, I. H., & Gonzalez, H. V. (2020). A Machine Learning Based Framework for Sentiment Analysis of App Reviews in Google Play Store. Journal of Systems and Software, 170, 110719.
7. Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). A Review and Comparative Study of Sentiment Analysis Web Services. Information Sciences, 311, 18–38.
8. Sohail, A., Ahmad, I., & Ali, M. (2021). Sentiment Analysis on Google Play Store Application Review Using Machine Learning. International Journal of Advanced Computer Science and Applications, 12(1), 251–257.
9. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Volume 8, Issue 4, Article e1253.