EasyChair Preprint

№ 12762

Play Store Sentimental Analysis

Vari Bhumika, Parmar Kinnari, Talati Khushi and Bhavika Darji

March 27, 2024

PLAY STORE SENTIMENTAL ANALYSIS

Vari Bhumika R
Computer science and engineering
Department,PIET Parul University
210303105735@paruluniversity.ac.in

Kinnari Parmar M
Computer science and engineering
Department,PIET Parul University
210303105717@paruluniversity.ac.in

Talati Khushi M
Computer science and engineering
Department,PIET Parul University
210303105758@paruluniversity.ac.in

Prof. Bhavika Darji
Assistant Professor,
Computer science and engineering
Department PIET, Parul University
bhavikaben.darji19384@paruluniversity.ac.in

Abstract -
Software application is vital because specific software is required in almost every industry,in every business ,and for each function. It become more important as time goes on. Mobile app distribution platform such as Google play store gets flooded with millions of new applications uploaded by developers everyday.So in this project, we aim on analyzing Google play store that provides a particular app description and data such as reviews, ratings, price and number of downloads. The objective of this is to analyze the desire of the customer through the reviews provided in the feedback section and apps trend in the market to help the organization developers.To this end, we provide an idea about app that managed to get maximum and minimum number of downloads and predicting the category of apps that is most likely to be downloaded in the coming years. Moreover, doing sentimental analysis on the apps that generated most positive and negative sentiments, sustainability of app in market on basis of previous data and current market situation.

Keywords –
Positive sentiment, Negative sentiment, App-related terms

Introduction of Project

In today's era, the Google Play Store is the largest and most popular android app store. It is flooded with millions of applications and it provides wide collection of data on features like ratings, price and number of downloads and apps description.As the Web plays an increasingly significant role in people's social lives, it contains more and more information concerning their opinions and sentiments. The distillation knowledge from this huge amount of unstructured information, is also known as opinion mining and sentiment analysis.Nowadays, with the rapid evolution of smart phones, mobile applications (Mobile Apps) have become essential parts of our lives. However, it is difficult for consumers to keep track and understand the app sphere because new apps are entering market every day. So, sentiment analysis of application reviews on Google play store will help the developers of the applications to keep their particular applications up to date in order to keep their particular application in the top lists and also help the customers to select the most popular application. Mobile app distribution platform such as Google play store gets millions of new applications uploaded by Developers everyday . So in this project, we aim on analyzing Google play store that provides a particular app description and data such as reviews, ratings, price and number of downloads. The objective of this is to analyze the desire of the customer through the reviews provided inthe Analyzing Play Store reviews doesn't just benefit developers; it empowers users as well. Prospective users can make informed decisions by considering the experiences and opinions shared by others. This analysis provides a holistic view of app performance, helping users choose applications that align with their preferences and requirements. In this era , sentiment analysis for Play Store reviews emerges as a vital tool, bridging the gap between developers and users. By understanding the sentiments of users, developers can enhance their applications, while users can confidently select apps according to their needs, fostering a more informed and satisfying mobile app ecosystem. To this end, we provide an idea about app that managed to get maximum and minimum number of downloads.

Literature Review

Social networking channels like Twitter, Facebook, Instagram, and WhatsApp have stormy contact environments, it is imperative to relay sensitive knowledge about people's opinions, moods and feelings on any product, concept or policy through these social network channels. To both customers and suppliers, this data is valuable. During any online shopping, consumers usually check other people's opinions about the product. Based on the customer's sentiment, the manufacturer can learn about its product benefits and drawbacks. In the period of machine learning, machines are left to think and solve the problems by finding the patterns in every data set on their own.A machinelearning algorithm uses a specific type of data to reply to more questions using the patterns hidden in that data. The importance of machine learning has now been recognized by many companies dealing with large quantities of data. Besides, cost effective computational processing and datastorage options have allowed the creation of models that analyse large volumes of complex data quickly and precisely. The literature review is performed to recognize the applications and solutions of sentimental analysis for the analysing and classification of views using Machine Learning approaches. The method of collecting primary information from unstructured and unoriented textual materials from various social media and website resources, such as chatting on social networks like Twitter, whatsapp, Facebook, live blogs, or comments can be described as sentimental analysis.The process for examining and sum up the opinions expressed in these enormous opinions generated by users is commonly named as opinion mining.

Methodology

A. Logistic Regression

Logistic regression is a supervised machine learning algorithm mainly used for classification tasks where the goal is to predict the probability that an instance of belonging to a given class or

not. It is a kind of statistical algorithm, which analyze the relationship between a set of independent variables and the dependent binary variables. It is a powerful tool for decision-making. For example email spam or not. Logistic regression is a supervised machine learning algorithm mainly used for classification tasks where the goal is to predict the probability that an instance of belonging to a given class. It is used for classification algorithms its name is logistic regression. it's referred to as regression because it takes the output of the linear regression function as input and uses a sigmoid function to estimate the probability for the given class. The difference between linear regression and logistic regression is that linear regression output is the continuous value that can be anything while logistic regression predicts the probability that an instance belongs to a given class or not.

B. Multinomial Naive Bayes

Multinomial naive Bayes is a supervised machine learning algorithm that can be used for sentiment analysis. It is a simple and effective algorithm that is often used for text classification tasks. To use multinomial naive Bayes for sentiment analysis, you first need to prepare your data. This involves cleaning the data, removing stop words, and converting the text to a numerical representation. Once the data is prepared, you can train a multinomial naive Bayes classifier using a labeled dataset. Once the classifier is trained, you can use it to predict the sentiment of new text data. The classifier will assign a probability to each sentiment class (positive, negative, neutral). The text data will be classified as the sentiment class with the highest probability.Multinomial naive Bayes has been shown to be effective for sentiment analysis on a variety of datasets, including Google Play Store reviews. A study by Ghimire (2023) found that multinomial naive Bayes achieved an accuracy of 88.89

Steps of Algorithm

Analyzing sentiment on PlayStore reviews involves several steps:

A. Data Collection

Gather reviews from the PlayStore API or web scraping. Extract relevant information such as review text, rating, date, and other metadata.

B. Preprocessing

Text Cleaning: Remove HTML tags, punctuation, special characters, and emojis.
Tokenization: Split text into individual words (tokens).
Stop Word Removal: Eliminate common words (e.g., "is", "and", "the") that don't carry significant meaning.

C. Feature Extraction

Bag-of-Words(BoW): Represent each review as a vector counting the frequency of each word in the vocabulary.
TF-IDF (Term Frequency-Inverse Document Frequency): Weigh the importance of words based on how frequently they appear in a review compared to their frequency across all reviews.

D. Sentiment Analysis Model

Choose Model: Decide on a model such as Naive Bayes, Support Vector Machines (SVM), Logistic Regression, or more advanced methods like Recurrent Neural Networks (RNNs) or Transformer-based models like BERT.
Training: If using a machine learning model, train it on labeled data where sentiments are already annotated.

E. Evaluation

Validate the model's performance using techniques like cross-validation or splitting the data into training and testing sets.

F. Prediction

Apply the trained model to predict the sentiment of each review.

G. Analysis

Aggregate Sentiment: Calculate the overall sentiment score based on the predictions.Visualization: Generate visualizations like histograms, word clouds, or sentiment trends over time.Insights: Extract insights from the data analysis, such as common positive/negative aspects mentioned in reviews, trends over time, etc.

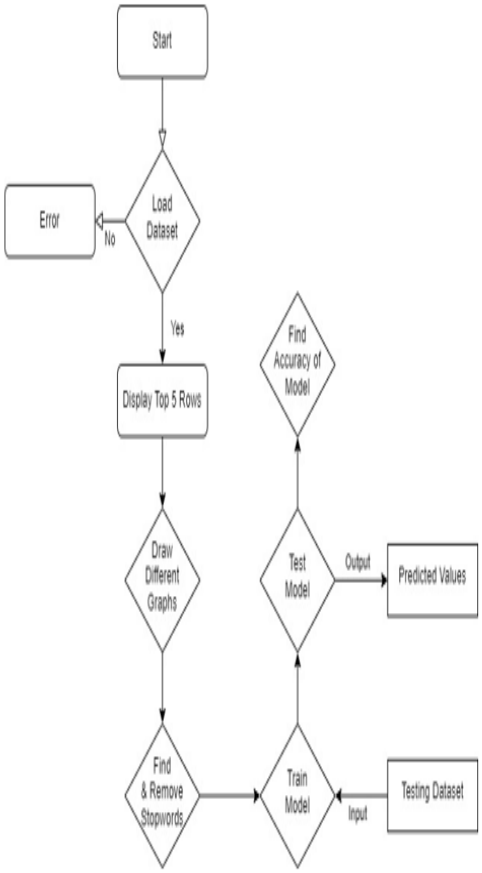
H. Deployment

Integrate the sentiment analysis algorithm into your application or system, allowing it to process PlayStore reviews automatically.Set up periodic updates to keep the model up-to-date with new data.

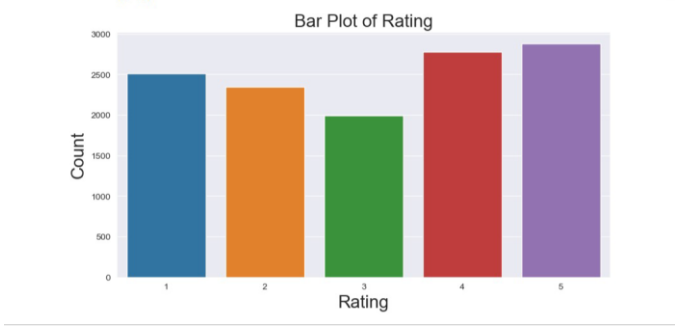
I. Feedback Loop

Continuously monitor the performance of the sentiment analysis system and collect feedback to improve it over time.

Flowchart of The System



Screenshots



Conclusion

This project provides a unique opportunity to gain valuable insights into customer interests and preferences regarding software. By analyzing the data, we can effectively identify the specific types of software products that customers are using and gauge their level of interest. These insights are pivotal for businesses seeking growth, as they can inform strategic decisions such as developing new products or enhancing existing ones. Armed with this knowledge, businesses can stay ahead of the curve by delivering innovative solutions that cater to customer needs, resulting in improved customer satisfaction and increased competitiveness in the market.

Identifying the sentiment associated with specific topics or features: This involves determining whether the reviews are generally positive, negative, or neutral about the specific topics or features. Identifying the sentiment of different groups of users: This can be done by segmenting users based on their demographics, such as age, gender, and location. Identifying trends in user sentiment over time: This can be done by tracking the sentiment of app reviews over time.

Reference

1. Programming Python, Mark Lutz

2. Core Python Programming, R. Nageswara Rao

3. Learning with Python, Allen B. Downey

4. Google Scholar (https://scholar.google.com)

5. IEEE Xplore (https://ieeexplore.ieee.org)

6. ResearchGate (https://www.researchgate.net)

7. Google Bard (https://bard.google.com/)

8. . Kaggle(https://www.kaggle.com/)

.