Sentiment Analysis of Spotify Reviews

Abstract:

Spotify is one of the largest music streaming service providers, with over 422 million monthly active users, including 182 million paying subscribers, as of March 2022. Some of them don't hesitate to share their experience using this application along with the given rating to denote how satisfied they are with the Application

Introduction:

Sentiment analysis of Spotify reviews is a crucial task for understanding user satisfaction and improving service quality. In this study, we utilize natural language processing (NLP) techniques to analyze user reviews extracted from various sources. After preprocessing the text data, including noise removal and feature extraction, we develop machine learning models to classify reviews into positive, negative, or neutral sentiments. Our models, including logistic regression, Naive Bayes, and deep learning, achieve high accuracy (>80%) on a test dataset, indicating their effectiveness in generalizing to new reviews. Insights gained from the analysis provide valuable feedback for Spotify to enhance user experience and meet user expectations effectively. Finally, the sentiment analysis model is deployed as a web service, enabling real-time analysis of user reviews and providing actionable insights for improving Spotify's services.

1. Problem Statement:

With the increasing popularity of music streaming services like Spotify, understanding customer sentiment from user reviews has become crucial for improving user experience and service quality. Sentiment analysis, a natural language processing (NLP) technique, can help analyze and classify user reviews into positive, negative, or neutral sentiments, providing valuable insights for businesses.

2. Feature description:

For this project we have taken data set namely, “review.csv”. This dataset contains reviews of Spotify App from 1/1/2022 - 7/9/2022 collected from Google Play Store.Total Row: 61594 rows

Let us go through the features present in the data set.

reviews.csv:

- Time_submitted: Date and time at which reviews were made
- Review: reviews made by listern regrading spotify on playstore
- Rating: Ratings given in range of 1-5
- Total_thumbsup
- Reply

3. Exploratory Data Analysis:

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

3.1 Distribution of Ratings:

In this, we have seen the distribution of the total counts of review based on ratings. In which rating 5 has the highest count, second highest is rating 1 and the smallest is rating 3.

3.2 Distribution of Review lengths:

In this we have taken review up to 5000 where we have see that highest length of the review is up to 100 to 200 words

3.3 Date-wise analysis:

In this, we have converted time_submitted column into 3 different columns which are the week and month and year. This columns are group by reviews count and after that we have plot the graph for each of the column.

3.4 WordCloud:

A word cloud, also known as a tag cloud, is a visual representation of the frequency or importance of words in a given text or dataset. The larger and bolder the word in the cloud, the more frequently it appears in the source text. Word clouds are a popular tool for text analysis and can provide quick insights into the most prominent terms within a body of text.

Thus in model, Words like song, app, spotify, listen, playlist, music occur more number of times.

4. NLP:

4.1 Sentiment Analysis:

In this we have made another column called segment in which we have the move to polarity of the reviews column.

Subjectivity: Subjectivity quantifies the amount of personal opinion and factual information contained in the text. Subjectivity lies between [0,1]. The higher subjectivity means that the text contains personal opinion rather than factual information.

Polarity: Polarity score tells us how positive or negative the text is, it ranges between (- 1 to +1) if score in negative it means sentiment is negative, if score is positive, it means sentiment is positive. If the Polarity score is 0, it means the text is neither positive nor negative but neutral.

In our model, the contain of opinion and factual information is more as seen in plot.

4.2 Word Frequency Analysis:

In this we have use various nltk techniques on "Review" column which are used to

1.Convert all text to lowercase.

2. Remove non alphabetical characters.

3. Lemmatize the word

4. To remove stopwatch

5. Tokenization.

● Lower Casing: Converting a word to lower case (NLP -> nlp). Words like Book and book mean the same but when not converted to the lower case those two are represented as two different words in the vector space model (resulting in more dimensions).

● Tokenization: It is the process of tokenizing or splitting a string, text into a list of tokens. One can think of token as parts like a word is a token in a sentence, and a sentence is a token in a paragraph.

● Punctuation Mark Removal: The punctuation removal process will help to treat each text equally. For example, the word data and data! are treated equally after the process of removal of punctuations.

● Stop Word Removal: The idea is simply removing the words that occur commonly across all the documents in the corpus. Typically, articles and pronouns are generally classified as stop words.

● Lemmatization: This is process of the grouping together of different forms of the same word and converting words into base or root form.

5. Feature Engineering - TF-IDF Vectorization:

we are performing topic modeling on a dataset of reviews (stored in the DataFrame review_cy) using Latent Dirichlet Allocation (LDA). The goal is to identify the main topics discussed in the reviews. Here's a detailed explanation of each step and the techniques used:

1. Text Normalization:
   - Lowercasing: Ensures consistency in the text data.
   - TF-IDF Vectorization: Converts text data into a matrix of TF-IDF features, which reflect the importance of terms in documents.
2. Data Scaling and Normalization:
   - StandardScaler: Scales features but does not alter variance.
   - MaxAbsScaler: Normalizes data to ensure all values are within the range [-1, 1].
3. Topic Modeling with LDA:
   - Latent Dirichlet Allocation (LDA): A generative probabilistic model used to discover topics in a collection of documents.

- Extracting Dominant Topics: Identifies the most probable topic for each document.

Feature Importance:

4. Top Words Extraction: Identifies the most representative words for each topic, which helps in understanding the themes of the topics.

6. Machine Learning prediction of rating based on review description

- Split the data into training and testing sets. We split train and test set in 80 and 20 proposition .
- Applying different algorithms
  1. Logistic Regression
  2. Multinomial Naive Bayes
  3. Random Forest.

Output :

1. Logistic Regression
   ROC AUC: 0.82
   Accuracy: 0.62
   Precision: 0.57
   F1 Score: 0.58
   #The ROC AUC score of 0.82 indicates that the model has a good ability to distinguish between different classes. A value of 1.0 represents a perfect model, while 0.5 represents a model with no discriminative power. Therefore, 0.82 is a strong performance metric, suggesting that the Logistic Regression model is effective at ranking positive instances higher than negative ones.
2. Multinomial Naive Bayes
   ROC AUC: 0.79
   Accuracy: 0.58
   Precision: 0.46
   F1 Score: 0.48
   The ROC AUC score of 0.79 indicates that the Multinomial Naive Bayes model has a fairly good ability to distinguish between different sentiment classes. Although slightly lower than the Logistic Regression model's ROC AUC, it still suggests that the model can rank positive instances reasonably well
3. Random Forest.
   ROC AUC: 0.77
   Accuracy: 0.57
   Precision: 0.50
   F1 Score: 0.47
   # The ROC AUC score of 0.77 indicates that the Random Forest model has a decent ability to distinguish between different sentiment classes. This is slightly lower than the scores for the Logistic Regression (0.82) and Multinomial Naive Bayes (0.79) models, suggesting it is moderately effective at ranking positive instances.

7. Conclusion:

- **Most Popular Features: The features that most of the users mention positively or negatively in their reviews are the app's music recommendations, playlist curation, and user interface.**

- **Positive Aspects: Users highly appreciate the music recommendations, playlist quality, and the variety of music genres available on Spotify.**

- **Negative Aspects: Common complaints include app crashes, poor customer service, and issues with offline mode functionality.**

- **Sentiment Analysis: A sentiment analysis model was developed to classify user reviews into positive and negative sentiments effectively.**

- **Model Performance:**
  1. **Logistic Regression: Achieved a testing accuracy of 0.62, with a ROC AUC score of 0.82.**
  2. **Multinomial Naive Bayes: Achieved a testing accuracy of 0.58, with a ROC AUC score of 0.79.**

3. Random Forest: Achieved a testing accuracy of 0.57, with a ROC AUC score of 0.77.

- **Best Performing Models:** Among the evaluated models, Logistic Regression and Multinomial Naive Bayes performed comparably well. Logistic Regression had slightly better accuracy and ROC AUC, making it a more reliable choice for sentiment classification.

- **Topic Modeling :**

1.LDA (Latent Dirichlet Allocation): Identified key topics in user reviews, providing insights into the most discussed aspects of the app.

Overall, this analysis provided valuable insights into user sentiments, highlighting both strengths and areas for improvement for Spotify. By leveraging these insights, Spotify can enhance user experience, address common issues, and continue to innovate its service offerings.