Summary

Sentiment analysis is the process of determining the sentiment or emotion expressed in a piece of text, such as a review or comment. It plays a crucial role in understanding the overall opinion or attitude of users towards a product, service, or any other subject. In this project, sentiment analysis is performed specifically on ChatGPT iOS reviews.

The data for this analysis is collected from the App Store, focusing on reviews related to ChatGPT. Basic exploratory data analysis (EDA) is conducted to gain insights into the dataset. Preprocessing steps are applied to clean the text data, including converting words to lowercase, removing punctuation and emojis, removing stop words, and performing lemmatization to reduce words to their base form.

After the initial preprocessing, word clouds are created to visualize the most frequent words associated with positive, negative, and neutral sentiments. These word clouds provide a quick overview of the commonly expressed sentiments in the reviews.

Further preprocessing steps involve transforming the text data into a numerical representation using bag-of-words vectorization. To address the issue of imbalanced data, the SMOTE (Synthetic Minority Over-sampling Technique) algorithm is applied to generate synthetic samples of the minority class.

Next, various classification models are trained and evaluated to find the best-performing model for sentiment analysis. Evaluation metrics such as precision, recall, and F1 score are utilized to assess the model's performance in predicting sentiment accurately.

In conclusion, this project presents an end-to-end approach for sentiment analysis of ChatGPT iOS reviews. By applying preprocessing techniques, exploratory data analysis, and employing classification models, the project aims to gain insights into the sentiments expressed in the reviews and predict sentiment with reasonable accuracy.