TripAdvisor Sentiment Analysis Report

Objective: -

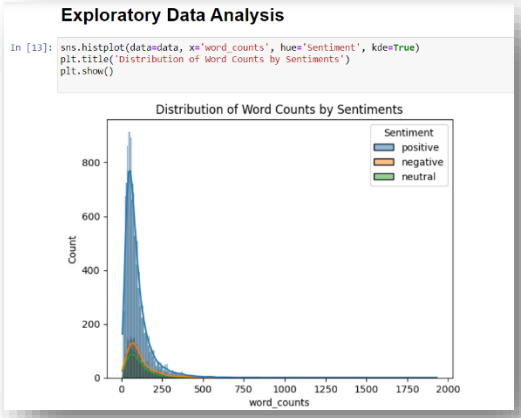
The sentiment analysis of TripAdvisor reviews. It analyses the text of reviews to categorize them as positive, negative, or neutral. By looking at how many words are used in reviews with different emotions (sentiments), it can provide insights for businesses listed on TripAdvisor. Imagine it as a way to understand if customers are raving (short, positive reviews) or ranting (longer, negative reviews) on TripAdvisor.

The libraries are used in the code:

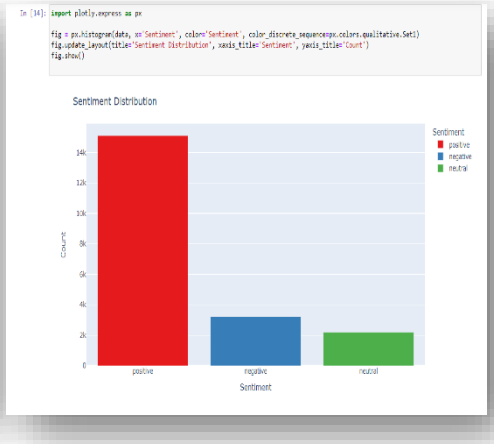
- Pandas, NumPy, NLTK, seaborn, matplotlib, Text blob, Plotly, TF-IDF,

Exploratory Data Analysis

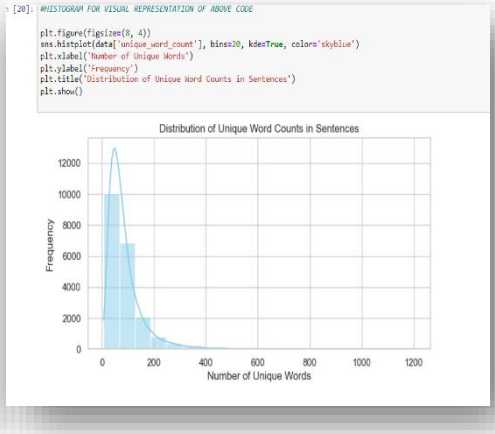
Plot - A



PLOT – B



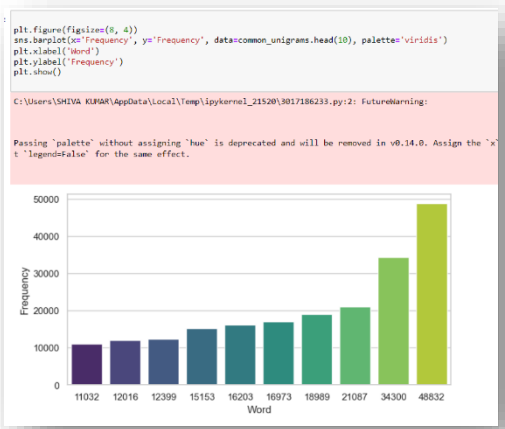
PLOT - C



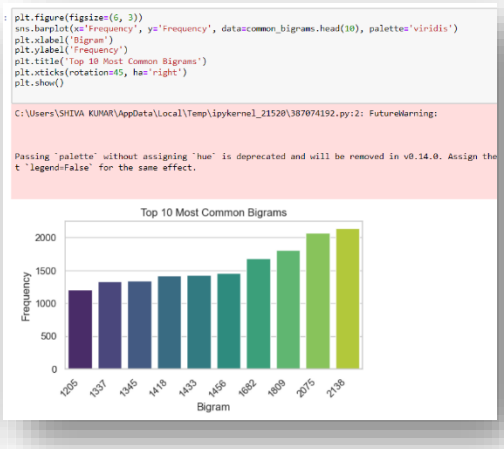
Plot-B

The code analyzes sentiment (positive, neutral, or negative) of TripAdvisor reviews by looking at word counts and how many words are used in reviews with different feelings. This can help businesses understand customer feedback on TripAdvisor.

PLOT – D



PLOT- E



PLOT - F

PLOT-C: - The histogram shows that the distribution of word counts varies across different sentiment categories. For instance, it might reveal that positive reviews tend to be shorter than negative reviews on TripAdvisor, which could be because positive experiences are easier to summarize in fewer words.

This kind of sentiment analysis can be useful for businesses listed on TripAdvisor to understand how customers feel about their experience. By analyzing the sentiment of reviews, businesses can gain insights into what aspects of their service customers find positive or negative, and use this information to improve their offerings.

PLOT- F: - The chart reveals how many reviews fall into each sentiment category and how many words they tend to use. For example, it might show that positive reviews on TripAdvisor are shorter on average compared to negative reviews. This could be because positive experiences are easier to describe in fewer words.

Sentiment analysis like this can be valuable for businesses listed on TripAdvisor. It helps them understand customer perception of their services. By analyzing the sentiment of reviews, businesses can gain insights into what aspects of their offerings resonate with customers (positive sentiment) and what areas need improvement (negative sentiment). This can help them improve their customer experience and attract more positive reviews

The Models applied on Trip Advisor Hotel Reviews: -

> Random Forest classifier: 0.96
> Naive Bayes: 0.94
> Logistic Regressor: 0.96
> Support Vector Machine (SVM): 0.96
> Decision Tree: 0.75
> AdaBoost: 0.82
> Gradient Boosting: 0.84

Analysis and Comparison: -

Top Performers:

Logistic Regression and SVM both achieved the highest accuracy of 0.96, making them the best performers for this task.
Random Forest also performed very well with an accuracy of 0.95, closely following the top models.
Moderate Performers:
Naive Bayes achieved an accuracy of 0.94, which is still high and quite competitive.
Gradient Boosting with an accuracy of 0.84 and AdaBoost with 0.82 are decent but significantly lower than the top models.

Low Performer:

Decision Tree had the lowest accuracy at 0.75, indicating that it may not be the best choice for this particular sentiment analysis task.

Insights:

Logistic Regression and SVM: The high performance of these models suggests that linear models are particularly effective for this sentiment analysis task. SVM's capability to find an optimal hyperplane for classification likely contributed to its success.

Random Forest: This ensemble method's strong performance indicates that combining multiple decision trees helps in capturing the nuances in the sentiment data better than a single decision tree.

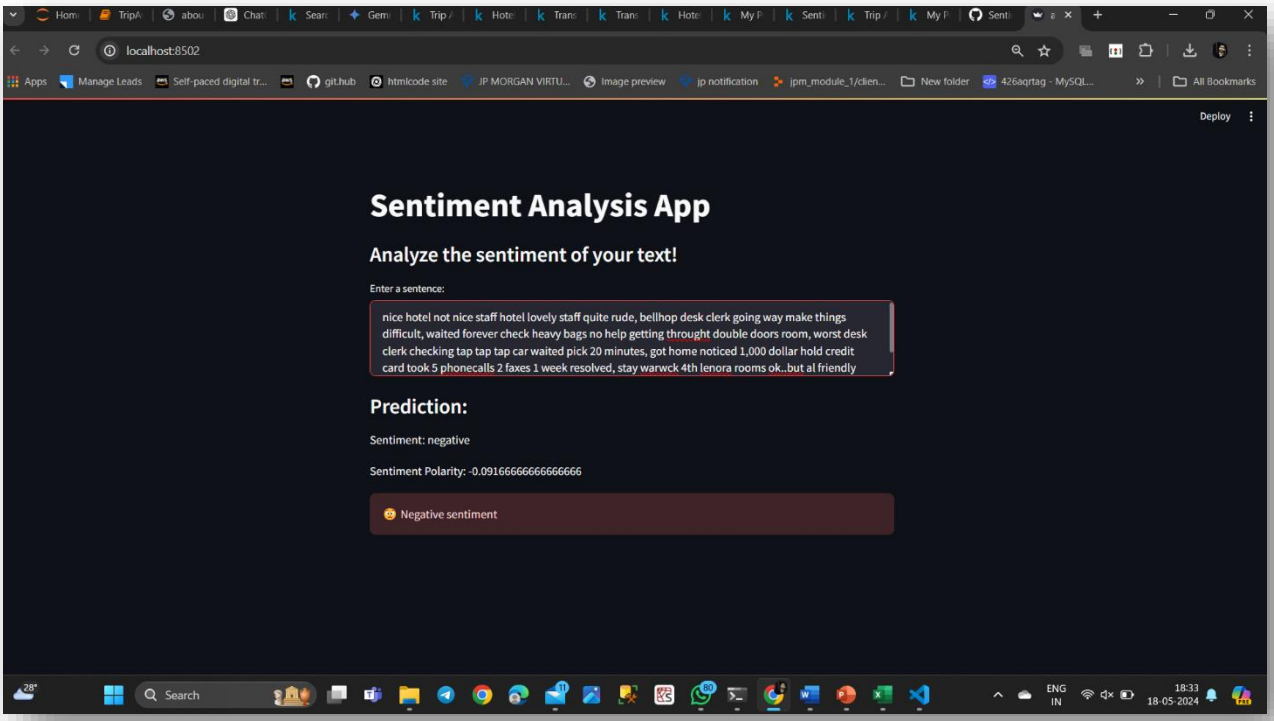
Naive Bayes*:* Despite its simplicity and the assumption of feature independence, Naive Bayes performed quite well, likely due to the effectiveness of the TF-IDF transformation and TextBlob features.

Boosting Methods: Both AdaBoost and Gradient Boosting performed moderately well, showing that boosting can enhance the performance of weaker learners but might require further tuning to match the top performers.

Decision Tree*:* The relatively poor performance suggests that a single decision tree might be overfitting or not capturing the complexity of the data well enough compared to ensemble methods or linear classifiers.

Model Deployment

- Used model deployment Stream-lit.
- Used for importing stream-lit tool, request, render template before creating the rf_model.pkl by importing pickle, which helps us to create a deployment folder and deployment app.



Conclusion:

For sentiment analysis of hotel reviews using the mentioned preprocessing techniques (lemma, TF-IDF, TextBlob), Logistic Regression and SVM are the most effective classifiers. While Random Forest is also a strong contender, simpler models like Decision Tree do not perform as well in this context. Boosting methods give a moderate performance.

The sentiment analysis model, based on RandomForestClassifier, achieved an impressive accuracy of *96%.* This indicates its strong ability to classify hotel reviews into positive, negative, or neutral sentiments. With its high accuracy, the model offers valuable insights for hotel management, enabling them to address customer feedback effectively and enhance guest experiences.