
CAPSTONE PROJECT

SENTIMENTAL ANALYSIS

Presented By:

Vinnakota Nitish Raj – GITAM – Computer Science Engineering(Data Science)

OUTLINE

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach (Technology Used)**
- **Algorithm & Deployment**
- **Result**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

- The goal is to develop a sentiment analysis model to classify restaurant reviews as positive or negative. This model aims to assist restaurant owners and managers in understanding customer feedback and improving their services based on the sentiment of the reviews.

PROPOSED SOLUTION

- The proposed system will analyze restaurant reviews to determine the sentiment behind them. The solution involves the following components:
- **Data Collection:**
 - Gather a dataset of restaurant reviews from sources such as Yelp, Google Reviews, or other review platforms.
 - The provided dataset includes reviews and corresponding sentiment labels.
- **Data Preprocessing:**
 - Clean the data by removing stop words, punctuation, and other irrelevant elements.
 - Text normalization and tokenization are performed to prepare the data for analysis.
- **Machine Learning Algorithm:**
 - Implement a machine learning model such as Naive Bayes or a deep learning model like LSTM to classify the reviews.
 - NLP techniques are used to extract features from the text data.
- **Deployment:**
 - Develop an interface or application for users to input reviews and get sentiment analysis results.
 - Deploy the solution on a scalable platform to handle multiple requests efficiently.
- **Evaluation:**
 - Assess the model's performance using metrics such as accuracy, precision, recall, and F1 score.
 - Fine-tune the model based on feedback and performance monitoring.

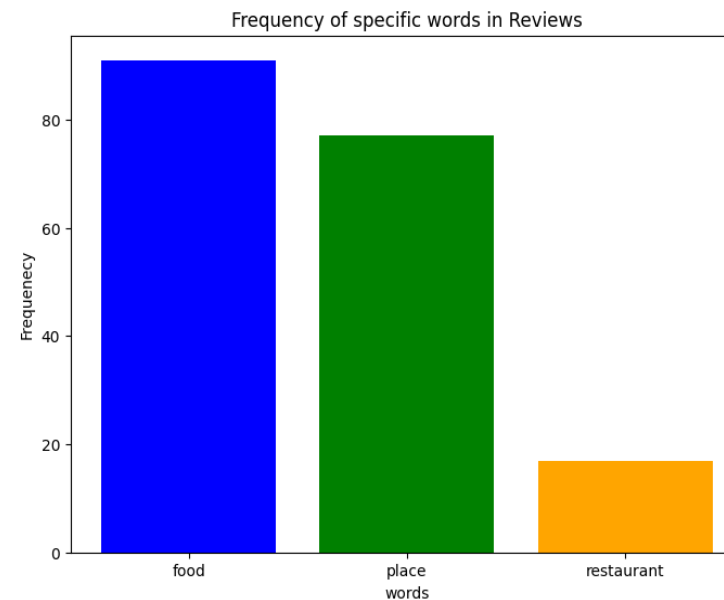
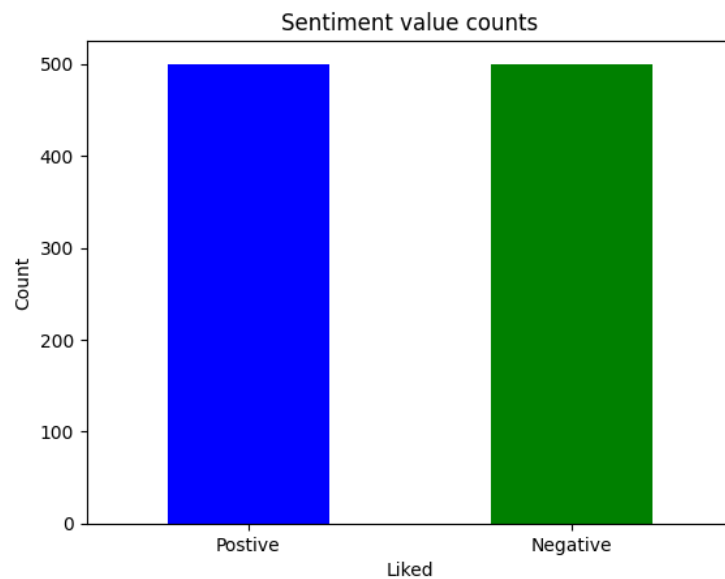
SYSTEM APPROACH

- **System Development Approach:**
- **System Requirements:**
 - Python programming language
 - I5 Processor, Windows Operating System, Google Colab/Jupyter Notebook
 - Libraries: Pandas, Scikit-learn, NLTK, TensorFlow/Keras, Matplotlib, Seaborn
- **Technology Used:**
 - Natural Language Processing (NLP)
 - Machine Learning
 - Data Cleaning and Preprocessing
- **Steps Involved:**
 - Load and explore the dataset
 - Visualize the data distribution
 - Generate word clouds and bar plots for data insights

ALGORITHM & DEPLOYMENT

- **Algorithm:**
- **Algorithm Selection:**
 - Choose Multinomial Naive Bayes for its effectiveness in text classification tasks. Additionally, TfidfVectorizer was used to convert text data into numerical format.
- **Data Input:**
 - Input features include the text of the reviews. Like and Review Columns.
- **Training Process:**
 - Split the data into training and testing sets (80/20 split). Train the model using the training set with techniques like cross-validation and hyperparameter tuning.
- **Prediction Process:**
 - Use the trained model to predict the sentiment of new reviews. Incorporate real-time data inputs during the prediction phase.
- **Deployment:**
 - Develop a user-friendly interface for predicting the sentiment of new reviews. Deploy the model using Flask or a similar framework for real-time predictions.

RESULT



CONCLUSION

- The sentiment analysis model accurately classified restaurant reviews with an accuracy of 0.8%, demonstrating strong performance in both positive and negative sentiment detection. Visualizations like the confusion matrix confirmed the model's reliability.
- Challenges included handling imbalanced data and preprocessing text by removing stop words, punctuation, and normalizing text. These issues were mitigated through resampling techniques and careful preprocessing.
- Sentiment analysis helps restaurant owners understand customer feedback, enabling them to improve services and address concerns. This leads to enhanced customer satisfaction, loyalty, and business growth.

FUTURE SCOPE

- Potential enhancements include:
 - Incorporating additional data sources (e.g., social media comments).
 - Optimizing the algorithm for better performance.
 - Expanding the system to cover multiple languages and regions.
 - Integrating advanced machine learning techniques or deep learning models for improved accuracy.

REFERENCES

- **Pandas Documentation:** McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 445, 51-56. Retrieved from <https://pandas.pydata.org/>
- **NLTK Documentation:** Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media Inc. Retrieved from <https://www.nltk.org/>
- **Scikit-learn Documentation:** Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. Retrieved from <https://scikit-learn.org/stable/>
- **TensorFlow/Keras Documentation:** Chollet, F., & others. (2015). Keras. Retrieved from <https://keras.io/> and <https://www.tensorflow.org/>
- **Data Source for Restaurant Reviews:** The dataset used for sentiment analysis was sourced from Kaggle, containing restaurant reviews labeled as positive or negative. Retrieved from <https://www.kaggle.com/>



THANK YOU