

# University of Palestine



## Department of Software Engineering

### **Sentiment Analysis of Restaurant Reviews**

**By:**

**Salah El-Din Abu Saif**

**120190964**

**Supervisor:**

**Dr. Mahmoud Abu Alsebah**

**Dr. Mohammed Timraz**

**May, 2023**

# Content

<b>Abstract:</b> .....	1
<b>1. Introduction:</b> .....	2
<b>2. Methodology:</b> .....	2
2.1. Planning: .....	2
2.2. Data Collection: .....	2
2.3. Data Preprocessing: .....	2
2.4. Model Training: .....	3
2.5. Evaluation: .....	3
<b>3. Results:</b> .....	3
<b>4. Discussion:</b> .....	3
<b>5. Conclusion and Recommendations:</b> .....	4
<b>6. References:</b> .....	4

## **Abstract:**

This report presents a sentiment analysis of restaurant reviews using a Multinomial Naive Bayes algorithm. The objective was to classify the sentiment expressed in the reviews as either positive or negative. The dataset consisted of restaurant reviews collected from an online source. Preprocessing steps, including text cleaning, tokenization, stop word removal, and stemming, were applied to the dataset. The CountVectorizer from scikit-learn was used to convert the preprocessed text into numerical features. The dataset was split into training and testing sets, and the Multinomial Naive Bayes model was trained on the training set. The accuracy score and classification report were used to evaluate the model's performance on the testing set. The results showed that the model achieved a satisfactory accuracy score, indicating its effectiveness in classifying restaurant reviews based on sentiment. This report highlights the methodology employed, discusses the results, and provides recommendations for further analysis and improvement. The findings of this sentiment analysis can be valuable for businesses in the restaurant industry to gain insights from customer feedback and enhance their services accordingly.

## **1. Introduction:**

The purpose of this report is to present the results of sentiment analysis conducted on a dataset of restaurant reviews. The objective of the analysis is to determine the sentiment (positive or negative) expressed in the reviews. This report provides an overview of the data preprocessing steps, the machine learning model used, and the evaluation metrics employed to assess the performance of the model.

## **2. Methodology:**

### **2.1. Planning:**

The initial step involved acquiring the dataset of restaurant reviews from the provided source. The preferred conventions of structure and presentation were taken into consideration during the planning phase.

### **2.2. Data Collection:**

The dataset was obtained from the provided source, which consisted of a tab-separated values (TSV) file containing reviews and their corresponding sentiment labels.

### **2.3. Data Preprocessing:**

The data preprocessing steps included:

- Removing special characters and retaining only alphabetic characters.
- Converting all text to lowercase.
- Tokenizing the text by splitting it into individual words.
- Removing stop words to eliminate commonly occurring words with little semantic value.
- Stemming the words to reduce them to their root form.

## **2.4. Model Training:**

The CountVectorizer from scikit-learn was used to convert the preprocessed text into numerical features. The dataset was split into training and testing sets using the train\_test\_split function. The Multinomial Naive Bayes (MNB) algorithm was chosen as the classification model. The MNB model was trained on the training set using the fit function.

## **2.5. Evaluation:**

The trained model was used to make predictions on the testing set, and the accuracy\_score and classification\_report functions were employed to evaluate the performance of the model. The accuracy score provides a measure of the overall accuracy of the model, while the classification report presents metrics such as precision, recall, F1-score, and support for each class.

## **3. Results:**

The accuracy score achieved by the Multinomial Naive Bayes model on the testing set was 76.3.

The classification report provides a detailed breakdown of the performance metrics for each sentiment class (positive and negative). It includes precision, recall, F1-score, and support values for both classes.

## **4. Discussion:**

The results of the sentiment analysis demonstrate the effectiveness of the Multinomial Naive Bayes model in classifying restaurant reviews into positive and negative sentiments. The achieved accuracy score indicates a satisfactory level of performance.

## 5. Conclusion and Recommendations:

Based on the analysis conducted, it can be concluded that the sentiment expressed in restaurant reviews can be accurately classified using the Multinomial Naive Bayes algorithm. This approach can be useful for businesses in the restaurant industry to gain insights into customer feedback and make data-driven decisions to improve their services. Further exploration and experimentation with other machine learning algorithms and advanced natural language processing techniques may enhance the accuracy of sentiment analysis.

## 6. References:

1. Kaggle. (n.d.). Restaurant Reviews Dataset. Retrieved from <https://www.kaggle.com/datasets/maher3id/restaurant-reviewstsv>
2. Brownlee, J. (2019). How to Prepare Text Data for Machine Learning with scikit-learn. Machine Learning Mastery. Retrieved from <https://machinelearningmastery.com/prepare-text-data-machine-learning-scikit-learn/>
3. NLTK Project. (n.d.). NLTK 3.6.2 Documentation. Retrieved from <https://www.nltk.org/>
4. Seaborn. (n.d.). Seaborn: Statistical Data Visualization. Retrieved from <https://seaborn.pydata.org/>
5. Matplotlib. (n.d.). Matplotlib: Visualization with Python. Retrieved from <https://matplotlib.org/>