ZOMATO CUSTOMER REVIEWS SENTIMENT ANALYSIS

Abstract:

The NLP project aims to performs sentiment analysis on restaurant reviews using a Naive Bayes classifier. Sentiment analysis is the process of identifying the sentiment of a text, such as whether it is positive or negative. It is a natural language processing (NLP) task that has many applications, such as analyzing customer reviews, social media posts, and news articles. The Naive Bayes classifier is a simple but effective machine learning algorithm for classification tasks. It works by assuming that the features of a text are independent of each other, given the sentiment of the text. This assumption is often not true, but it still works well in practice.

The Bag of Words (BOW) model is a simple but effective way to represent text data for machine learning tasks. It creates a feature vector for each text document, where each feature represents the number of times a particular word appears in the document. The project follows a supervised machine learning approach to sentiment analysis. This means that it uses a labeled training dataset to train a model to predict the sentiment of new reviews. The training dataset consists of restaurant reviews and their corresponding sentiment labels (positive or negative). The project uses the training data to train a Naive Bayes classifier. Once the model is trained, it can be used to predict the sentiment of new reviews.

Methodology:

1. Data preprocessing:
The code preprocesses the restaurant reviews by cleaning them, removing stop words, and stemming the words. This is done to improve the performance of the sentiment analysis model.

2. Feature extraction:
The code then creates a Bag of Words (BOW) model to represent the reviews as feature vectors. Each feature vector represents the number of times a particular word appears in the review.

3. Model training:
The code then trains a Naive Bayes classifier on the feature vectors and the corresponding sentiment labels. This involves calculating the probability of each feature occurring in each sentiment class.

4. Prediction:
Once the model is trained, it can be used to predict the sentiment of new reviews. This involves taking a new review as input and calculating the probability of it belonging to each sentiment class. The sentiment class with the highest probability is predicted to be the sentiment of the review.

Technical Terms:

1. Sentiment analysis:
Sentiment analysis is the process of identifying the sentiment of a text, such as whether it is positive, negative, or neutral.

2. Naive Bayes classifier:
Naive Bayes is a simple but effective machine learning algorithm for classification tasks. It works by assuming that the features of a text are independent of each other, given the sentiment of the text.

3. Bag of Words model:
The Bag of Words model is a simple but effective way to represent text data for machine learning tasks. It creates a feature vector for each text document, where each feature represents the number of times a particular word appears in the document.

4. Feature vector:
A feature vector is a representation of a text document in a numerical format. Each feature in the vector represents a different aspect of the document, such as the number of times a particular word appears in the document.

5. Training data:
Training data is a set of labeled examples that is used to train a machine learning model.