1. Introduction :

In this assignment, our objective is to develop a sentiment analysis model solely using natural language processing (NLP) machine learning techniques.

2. Data Preprocessing:

Detail the steps taken to preprocess the text data for sentiment analysis, including:

Text cleaning (converting the text to lowercase, removing special characters, and punctuation)

Tokenization

Stopword removal

Lemmatization or stemming

3. Model Architecture and Parameters:

Describe the architecture of the sentiment analysis models, including:

Input representation (e.g., Bag-of-Words, TF-IDF)

Model type (e.g., Logistic Regression, Random Forest Classifier, Naive Bayes)

4. Training Process:

Explain the training process for the sentiment analysis models, including:

Evaluation Results and Analysis

Present the evaluation results of the sentiment analysis models, including:

Performance metrics (e.g., accuracy, precision, recall, F1-score)

Confusion matrix

Analysis of model predictions (e.g., examples of correctly classified and misclassified instances)

Sentiment distribution in the dataset

5. Conclusion

"Overall, I chose the Multinomial Naive Bayes classifier as it is a popular and effective choice for sentiment analysis tasks, especially when dealing with text data. Its simplicity, efficiency, and good performance make it well-suited for analyzing sentiment in text."

Successes and limitations of the model(s)

Insights gained from the analysis

Future directions for improvement or further research in sentiment analysis