**Short Report**

My Method and Outcomes :

For this project, I developed a model to determine if a movie review is positive or negative. I began by sanitizing the text—eliminating HTML tags, special symbols, numerical values, and frequent stopwords. Additionally, I transformed all content to lowercase and used stemming to consolidate similar words (such as playing and played) into a unified base form.

After cleaning the data, I converted the text into numerical form using a TF-IDF Vectorizer with 5000 features. This process enabled me to concentrate solely on the most significant words in the dataset. Subsequently, I trained a Multinomial Naive Bayes classifier, which is a straightforward yet efficient method for classifying text. Upon evaluating the model, it achieved approximately 88% accuracy. Both precision and recall, along with the F1-score, were balanced, indicating that the model performed equally well on positive and negative reviews. I also analyzed which words most significantly impacted the predictions, providing insights into what the model had "learned." In the end, I saved both the model and the vectorizer so that I could reuse them later without needing to retrain. Overall, I found that Naive Bayes combined with TF-IDF is quite effective for sentiment analysis, and I was pleased with the outcomes.