IMDB Sentiment Analysis Project Report

---

1. Final Model Accuracy & Evaluation Metrics

Model: Multinomial Naive Bayes

- 📊 **Accuracy**: 0.8571
- 🎯 **Precision**: 0.8533
- 🔁 **Recall**: 0.8638
- ⚖️ **F1-Score**: 0.8585

Confusion Matrix (summary)

- True Positives (Correct Positive Reviews): High
- True Negatives (Correct Negative Reviews): High
- False Positives & False Negatives: Moderate, but balanced

✅ The model demonstrates reliable performance with a good balance between **precision and recall**.

---

2. Short Write-Up (Approach + Results)

Approach

This project focused on **binary classification** of IMDB movie reviews into **positive** or **negative** sentiment using **Natural Language Processing (NLP)**.

Steps followed:

1. **Data Preprocessing**
   - Converted text to lowercase
   - Removed punctuation & stopwords
   - Applied stemming/lemmatization to normalize words
2. **Feature Extraction**
   - Used **TF-IDF Vectorizer (max 5000 features)** to transform text into numerical format
3. **Model Training**

    - Trained a **Multinomial Naive Bayes** classifier on the processed data
4. **Evaluation**
    - Measured accuracy, precision, recall, and F1-score
    - Analyzed confusion matrix and feature importance

Results

- The model achieved **~85.7% accuracy** on the test dataset.
- Precision (85.3%) and Recall (86.3%) show the model is good at both **catching positive reviews** and **avoiding false positives**.
- Top words influencing classification included:
    - **Positive reviews** → *excellent, amazing, fantastic, great*
    - **Negative reviews** → *boring, waste, bad, terrible*

Conclusion

The Naive Bayes sentiment classifier performed **reliably** with strong balance across all evaluation metrics. It can be used to automatically analyze and classify customer reviews effectively.

🔮 **Future Work**:

- Explore **Logistic Regression, SVM, or Deep Learning models (LSTMs/Transformers)** to improve accuracy further.
- Perform **hyperparameter tuning** and expand preprocessing with advanced text cleaning.