**IMDB Sentiment Analysis: Final Report**

**1. Introduction**

This project aimed to develop a sentiment analysis model that can accurately classify movie reviews as positive or negative. This task is crucial for understanding public opinion about movies and can be valuable for various applications, including movie recommendations and market research.

**2. Dataset**

The project utilized the "IMDB Dataset.csv" file, which contains a large collection of movie reviews and their corresponding sentiment labels (positive or negative). This dataset provided the necessary data for training and evaluating the sentiment analysis model.

**3. Methodology**

The project followed these key steps:

* **Data Preprocessing:** Cleaning and preparing the text data by removing stop words and unnecessary characters.
* **Feature Extraction:** Converting text reviews into numerical features using the TF-IDF (Term Frequency-Inverse Document Frequency) technique.
* **Model Selection:** Choosing Logistic Regression and Gaussian Naive Bayes as the machine learning models for sentiment classification.
* **Model Training:** Training the selected models on a portion of the dataset to learn patterns and relationships between text features and sentiment labels.
* **Model Evaluation:** Evaluating the trained models on a separate test set using metrics like accuracy, precision, recall, and F1-score.
* **Model Deployment:** Creating a Streamlit web application for users to input movie reviews and get real-time sentiment predictions.

**4. Results**

The Logistic Regression model achieved high accuracy in classifying movie reviews as positive or negative, with an accuracy score of approximately 88%. Other evaluation metrics such as precision, recall and F1-score were also satisfactory. The Gaussian Naive Bayes model showed slightly lower accuracy, implying Logistic Regression being the preferred choice for this task.

**5. Conclusion**

The IMDB Sentiment Analysis project successfully developed a robust sentiment analysis model using machine learning techniques. The model can be used to automatically classify movie reviews, providing valuable insights into audience sentiment.

**6. Future Work**

Possible improvements and extensions for the project include:

* **Exploring more advanced models:** Experimenting with deep learning models or ensemble methods to potentially enhance classification accuracy.
* **Incorporating additional features:** Including features like user ratings, movie metadata, or emotional lexicons to capture richer information.
* **Improving the user interface:** Enhancing the Streamlit app with better design and more features like sentiment visualization.
* **Deploying to a production environment:** Making the model accessible to a wider audience by deploying it as a web service or API.