***Sentiment Analysis Report***

**1. Introduction**

The goal of this analysis was to build a model that can automatically classify movie reviews as either positive or negative. This was achieved using Natural Language Processing (NLP) techniques and machine learning algorithms.

**2. Dataset**

The analysis utilized the "aclImdb" dataset, a collection of movie reviews commonly used for sentiment analysis. The dataset was loaded and split into training and testing sets.

**3. Methodology**

The following steps were undertaken:

* **Text Preprocessing:** This stage leverages various NLP techniques to prepare the text data for analysis:
  + **Tokenization:** Splitting reviews into individual words using nltk.word\_tokenize.
  + **Stop Word Removal:** Removing common words (e.g., "the," "a," "is") that don't carry much meaning using nltk.corpus.stopwords.
  + **Lemmatization:** Reducing words to their base form (e.g., "running" to "run") using WordNetLemmatizer.
* **Feature Engineering:**
  + **TF-IDF Vectorization:** Converting text into numerical representations using TF-IDF (Term Frequency-Inverse Document Frequency) to capture word importance. This is done using TfidfVectorizer from sklearn. This process enables machine learning models to understand and analyze text data.
* **Model Training:**
  + **Logistic Regression:** A machine learning model was trained to classify reviews based on their TF-IDF vectors.
* **Model Evaluation:**
  + **Classification Report:** The model's performance was evaluated using metrics like precision, recall, and F1-score to measure its accuracy in classifying positive and negative reviews.

**4. Results**

To see the results of the model evaluation, run the code in the 'Model Evaluation' section.

**5. Custom Prediction**

The code provides a function predict\_sentiment(text) that allows users to input their own text and get a sentiment prediction.

**6. Conclusion**

The sentiment analysis pipeline successfully built a model for classifying movie reviews by utilizing NLP techniques and machine learning. Further improvements could include exploring different models, hyperparameter tuning, and incorporating more advanced NLP techniques.