**Sentiment Analysis of IMDB Movie Reviews**

**Problem Statement:**

The problem revolves around categorizing movie reviews into sentiment categories such as positive, negative, or neutral based on the textual content. The objective is to develop models that can accurately predict sentiment categories for movie reviews, aiding in sentiment analysis of user opinions.

**Description of Dataset:**

* **Source for Scraping:** The dataset has been obtained from IMDb. IMDb might has been scraped using BeautifulSoup and APIs to retrieve movie reviews along with their corresponding sentiment labels.
* **Storing Scraped Data:** The scraped data is stored in a CSV file by ensuring uniformity in the format of scraped data. Through this csv file we can easily perform preprocessing techinques.

**Preprocessing Techniques Applied:**

Several preprocessing techniques might have been employed:

* **Text Cleaning:** Removing punctuation, special characters, and converting text to lowercase.
* **Tokenization:** Splitting text into tokens/words.
* **Stopword Removal:** Eliminating common stopwords.
* **Lemmatization:** Reducing words to their base or root form.
* **Handling Missing Values:** Removing any missing or null values in the text data.

**Models Selected and Why? (ML and DL):**

* **Machine Learning Models:** Chosen models Logistic Regression and Naive Bayes for their interpretability, simplicity, and efficiency with text classification tasks.
* **Deep Learning Models:** Utilized models Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) due to their ability to capture complex patterns in sequential data and achieve potentially higher accuracy for text analysis tasks.

**Implementation of Each Model:**

* **Parameters Used:** Hyperparameters such as Laplace smoothing in naive bayes, GlobalMaxPooling and ReLU activation in CNN, Embedding Layer and Softmax activation for multi-class classification in RNN
* **Number of Epochs (DL):** Trained for 5 epochs with a batch size of 64 in CNN and same for RNN.

The number of epochs chosen for training the deep learning models, determining the number of times the entire dataset was passed forward and backward through the model.

**Comparing ML with Deep Learning Models:**

* **Performance:** Comparison of the performance metrics like accuracy, precision, recall, F1-score, and computational efficiency between ML and DL models show that DL models are more efficient.
* **Complexity:** Utilizing various evaluation metrics such as accuracy, precision, recall, and F1-score in the classification report provides a comprehensive assessment of model performance. Interpretability of these metrics aids in understanding the model's strengths and weaknesses in classifying sentiments.

**Uniqueness of Your Project:**

* The uniqueness of this project lies in its comprehensive approach towards sentiment analysis of movie reviews using a diverse set of models, representation techniques, evaluation metrics, and practical implications in real-world scenarios. It showcases a broad spectrum of methodologies applied to a specific task, enabling comparisons and insights into their effectiveness.

**Real-World Application:**

* Sentiment analysis of movie reviews has practical applications in the entertainment industry, market research, and customer feedback analysis. Implementing these models can provide valuable insights for decision-making processes.