Predictive Model Development for Drug Eligibility

##Introduction:

The development of drugs is essential in providing effective therapeutic options for patients suffering from chronic and terminal illnesses. In this assignment, we aim to build a predictive model that can determine whether a patient will be eligible for a specific drug, referred to as the "Target Drug," in the next 30 days.

Steps for Model Development:

** A. Positive and Negative Sets: **

Positive Set: For the positive set, we need to identify patients who have taken the "Target Drug" as their first prescription. We will use historical data to determine when each patient took their first prescription of the drug, considering the date as the day when the patient became eligible for the drug.

Negative Set: To create the negative set, we need to select patients who did not take the "Target Drug" as their first prescription within the specified timeframe, i.e., patients who were not eligible for the drug.

** B. Feature Engineering: **

Frequency-based Features: We can extract features related to the frequency of drug prescriptions for each patient. These features may include the total number of drug prescriptions, the average time interval between prescriptions, or the number of prescriptions within a specific time window.

Time-based Features: Time is a critical aspect in this prediction task. We can create features that capture the time since the patient's first prescription, the time of the last prescription, or the time intervals between prescriptions.

Deep Learning Techniques: If necessary and applicable, we can leverage deep learning techniques to automatically extract relevant patterns and representations from the data. This may include using Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks to handle sequential data.

** C. Model Evaluation and Optimization: **

Validation Set: We will split the available data into training and validation sets. The model will be trained on the training set, and its performance will be evaluated on the validation set.

Evaluation Metrics: The primary evaluation metric for this task is the F1-Score, which takes into account both precision and recall. We will focus on optimizing the F1-Score to reduce false positives and false negatives.

Model Optimization: To improve the model's performance, we can experiment with

different algorithms, hyperparameters, and feature combinations. Additionally, we can consider techniques like oversampling or undersampling to handle imbalanced data, if necessary.

** D. Model Prediction and Final Submission: **

After developing the predictive model and optimizing its performance, we will use the model to generate predictions for patients in the test dataset (test.parquet).

Each patient in the test dataset will be labeled as 1 (eligible for the "Target Drug") or 0 (not eligible) based on the model's predictions.

The final generated predictions will be submitted in the "final_submission.csv" file.

** Conclusion: **

Developing a predictive model to determine a patient's eligibility for a specific drug is crucial for physicians in making informed decisions about treatment options. By leveraging feature engineering techniques and exploring deep learning methods, we can build an accurate and efficient predictive model. The F1-Score will be used as the evaluation metric to prioritize candidates with the highest performance in the final submission.