

Structured Data Assignment

Predicting Eligibility for "Target Drug"

Objective:

- Develop a predictive model to determine whether a patient will be eligible for "Target Drug" within the next 30 days.
- Achieve high accuracy in predicting eligibility for "Target Drug" to assist healthcare professionals in making informed treatment decisions.
- Generate a final submission of predictions for eligible patients based on the developed model.

Dataset Description:

- Patient-Uid - Unique Alphanumeric Identifier for a patient
- Date - Date when patient encountered the event.
- Incident - This column describes which event occurred on the day.

Positive and Negative Set Creation:

- Patients who have taken the "Target Drug" should be labeled as "1." For the negative set, patients who have not taken the "Target Drug." Ensure the balance of positive and negative sets to avoid class imbalance issues.
- Time since the last "Target Drug" prescription, time since the last symptom occurrence.

Model Development:

- The dataset is split into a training set and a validation set.
- Trained the model on the training set and evaluated its performance on the validation set.
- Focused on the F1-score as the primary evaluation metric to balance precision and recall. And the output on XGBoost Classifier (Accuracy: 0.71, Precision: 0.65, Recall: 0.95, F1 Score: 0.77).

Evaluation and Error Analysis:

- Analyzed false positives and false negatives to understand model errors.

Reduce False Positives and False Negatives:

- Adjusted classification threshold to balance precision and recall.
- Considered different machine learning algorithms that prioritize recall.