Project Documentation

Predicting Patient Eligibility for "TARGET DRUG"

Problem Statement:

The objective of this project is to develop a predictive model capable of determining whether a patient will be eligible for the "TARGET DRUG" within the next 30 days. This predictive capability aims to assist healthcare providers in making informed decisions about suitable **treatments, enhancing the overall quality of patient care and treatment management.**

Data Preprocessing:

1. Duplicate Handling: Removal of duplicate entries from the dataset to ensure data integrity and accuracy.
2. Feature Generation: Derivation of essential features such as 'Prescription_Count' and 'Time_diff' to analyze the patient's prescription history and time intervals effectively.
3. Positive Dataset Creation: Generation of a positive dataset comprising patients who have taken the "TARGET DRUG," facilitating the analysis of relevant patterns and trends.

Model Development:

1. Algorithm Selection: Utilization of the XGBoost classifier for its robust performance and efficiency in handling complex classification tasks.
2. Training and Testing: Division of the dataset into training and testing sets to train the model and evaluate its performance on unseen data.
3. Feature Utilization: Integration of critical features, including 'Prescription_Count' and 'Time_diff,' to enable accurate predictions of patient eligibility.

Model Evaluation:

1. Performance Metrics: Evaluation of the model using metrics such as the F1 score, accuracy score, and the ROC-AUC curve to assess its predictive capabilities and overall performance.
2. Insights and Analysis: Interpretation of the evaluation metrics to gain insights into the model's accuracy and efficacy in predicting patient eligibility for the "TARGET DRUG."

Overview of potential areas for further exploration and enhancement if more time were available for the project:

"If more time were available for this project, we could have delved deeper into several key aspects to further refine the predictive model and its applications. Some potential areas for exploration and improvement include:

1. Feature Engineering Refinement: The exploration of additional relevant features and the refinement of existing feature engineering techniques could enhance the model's predictive capabilities and improve its understanding of complex patient profiles and treatment histories.

2. Hyperparameter Tuning: Further experimentation with hyperparameter tuning for the XGBoost classifier could optimize the model's performance, potentially leading to improved accuracy and robustness in predicting patient eligibility for the 'TARGET DRUG'.
3. Ensemble Learning Techniques: Implementation of ensemble learning techniques, such as stacking or boosting, could potentially boost the model's performance by combining the strengths of multiple models and mitigating individual model weaknesses.
4. Data Augmentation Strategies: Integration of data augmentation strategies to expand the dataset and balance class distribution could address potential data imbalances, leading to a more comprehensive and representative analysis of patient eligibility for the 'TARGET DRUG'.

Exploring these avenues could further elevate the project's impact and relevance in the medical field, paving the way for more robust and effective healthcare solutions for patients and healthcare providers.