Structured Data Assignment

Problem Statement

The dataset in question contains a comprehensive collection of electronic health records belonging to patients who have been diagnosed with a specific disease. These health records comprise a detailed log of every aspect of the patients' medical history, including all diagnoses, symptoms, prescribed drug treatments, and medical tests that they have undergone. Each row represents a healthcare record/medical event for a patient and it includes a timestamp for each entry/event, thereby allowing for a chronological view of the patient's medical history.

The Data has mainly three columns

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

Project Overview

Importance of Drug Development:

The project recognizes the critical role of drug development in providing therapeutic options for patients with chronic and terminal illnesses. Specifically, it highlights the development of "Target Drug" designed to improve patients' health without causing dependency or severe side effects.

Focused Treatment Approach:

"Target Drug" is designed to offer a more focused and effective approach to treatment by tailoring it to a particular disease or condition. This approach aims to minimize the risk of harmful reactions and enhance the patient's well-being.

Objective of the Project:

The primary objective of this assignment is to develop a predictive model. This model's purpose is to predict whether a patient will be eligible for "Target Drug" within the next 30 days.

Clinical Decision Support:

The project's predictive model serves as a valuable tool to assist physicians in making informed decisions about patient treatment. By predicting eligibility for "Target Drug," physicians can optimize treatment choices.

Data Description

Data Structure:

1) Patient-Uid - Unique Alphanumeric Identifier for a patient 2) Date - Date when patient encountered the event. 3) Incident - This columns describes which event occurred on the day.

Data Cleaning:

- Missing Data: In this dataset there is no misssing values present in it.
- Duplicates: I checked whether there is any duplicates are in the dataset and dropped the duplicates from the dataset

Model Selection

Algorithms Used:

- RandomForestClassifier
- XGBoostClassifier

Evaluation Metrics Used:

- F1 score
- AUROC score

Model Performance Summary:

- Among these two models xgboost's f1 score gave around 96.1% where RandomForest gave 96%
- So i chose the model XGBoostClassifier

Conclusion:

This project showcases a compelling blend of cutting-edge data science and healthcare expertise, delivering a predictive model with the potential to revolutionize patient care. By accurately forecasting a patient's eligibility for "Target Drug" 30 days in advance, this project not only demonstrates technical prowess but also an unwavering commitment to improving healthcare outcomes. The significance of this work extends beyond its technical achievements. It has the potential to make a profound impact on patient lives and clinical decision-making, aligning perfectly with our shared goal of advancing healthcare solutions. I invite you to explore the depths of this project and join us in harnessing the power of data science for positive healthcare transformation.