Data Intake Report

Name: Data Science :: Healthcare - Persistency of a drug:: Group Project

Report date: 24th April 2024

Internship Batch: LISUM30:30 Jan24 - 30 Apr 24

Version: 1.0

Data intake by: David Paul Nalumenya

Group name: None

Specialization: Data science

Email: nalumenyad@gmail.com

Company: Makerere University

Data intake reviewer: None

Data storage location:

Tabular data details:

Parameter	Value
Total number of observations	3424
Total number of files	1
Total number of features	69
Base format of the file	.xlsx
Size of the data	919,559 bytes

1.1   Problem Statement

One of the challenges for all pharmaceutical companies is to understand the persistency of drug as per the physician prescription. To solve this problem ABC pharma company approached an analytics company to automate this process of identification.

1.2   ML Problem

With an objective to gather insights on the factors that are impacting the persistency, build a classification for the given dataset.

2.0   Business understanding

Pharmaceutical businesses make significant investments in drug research and development (R&D) with the goal of increasing patient uptake and market domination. However, when individuals discontinue their drug regimens for a variety of reasons, pharmaceutical corporations face huge losses.

2.1   Impact of Low Persistence

- Reduced revenue when patients cease their medications, the corporation loses income.
- Ineffective R&D evaluation in projecting drug success due to low persistency can result in incorrect appraisals of a medicine's genuine potential, stifling future R&D efforts.
- Financial burden related with nonadherence, such as handling issues caused by missed drugs.
- Brand image damage reducing future sales and brand perception.

2.2   Need for Persistence Prediction

- Effective and focused prescriptions by identifying patients who are more likely to stick to a pharmaceutical regimen enables tailored prescribing, which improves treatment success and patient outcomes.
- Proactive monitoring by identifying patients at risk of dropping out enables interventions to address concerns and increase adherence.
- Improved R&D evaluation by accounting for persistency allows organizations to acquire a more accurate picture of a drug's genuine effectiveness during clinical trials and R&D review.

3.0   Project lifecycle along with deadline

Predicting Patient Drug Persistency with Machine Learning

Project Goal

Develop a machine learning model to predict patient persistency for a specific drug, enabling targeted prescriptions and improved patient outcomes.

Week 1

Data Acquisition and Exploration

Deadline: 1 Week

Activities:

- Acquire and pre-process all the health (patient) data, including specialty, demographics, medical history, scans and medication history.
- Explore the data to understand patient characteristics, medication usage patterns, and potential factors influencing persistency.
- Clean and address missing values and any outliers in the data.
- Exploratory data analysis to determine the distribution and relationships between the different variables in the data

Feature Engineering and Model Selection

Deadline: 1 Week

Activities:

- Identify and engineer relevant features from the data that might influence persistency (e.g., age, comorbidities, risk factors and demographic factors).
- Perform feature selection techniques to identify the most informative features for model building.
- Research and select appropriate machine learning algorithms for predicting binary classification (persistent vs. non-persistent).
- Consider algorithms like Logistic Regression, Random Forest, or XGBoost.

Week 2

Model Training and Evaluation

Deadline: 1 Week

Activities:

- Split the data into training, validation, and testing sets.

- Train the chosen machine learning models on the training set, fine-tuning hyperparameters using techniques like GridSearchCV or RandomizedSearchCV.

- Evaluate model performance on the validation set using metrics such as accuracy, precision, recall, and AUC-ROC curve.

- Compare the performance of different models and select the best performing model for prediction.

Week 3 (optional)

Model Deployment and Monitoring

Deadline: 1 Week

Activities:

- Develop a deployment strategy for integrating the model into existing healthcare workflows or a dedicated application.

- Consider deployment options like web service API or integration with electronic health record systems.

- Implement a monitoring plan to track model performance over time and identify potential degradation in accuracy.

Deliverables

- A documented machine learning model that predicts patient drug persistency.

- Performance metrics and evaluation reports.

Timeline

- This project is estimated to be completed within a timeframe of 3 weeks, with each phase having its own deadline.