Week 7: Project Overview and Plan

Team Members: Apurwa Bhausaheb Sontakke, Vedant Wagh

Group Name - HealthData Innovators

Name – Apurwa Bhausaheb Sontakke

Email – sontakke.ap@northeastern.edu

Country - USA

College– Northeastern University

Specialization – Data Science

Problem Description

Pharmaceutical companies face a significant challenge in understanding why patients continue or discontinue their prescribed medications. To address this, ABC Pharma has sought the help of an analytics company to automate the identification process of factors influencing drug persistency. The aim is to develop a classification model that predicts whether a patient will persist with a prescribed drug (Persistency_Flag).

Business Understanding

- Understanding drug persistency is crucial for pharmaceutical companies to:
- Enhance patient adherence to prescribed therapies.
- Identify key factors that influence drug persistency.
- Improve physician engagement by providing insights into prescription patterns.
- Optimize marketing strategies to target specific patient demographics.

Project Lifecycle

Problem Definition and Business Understanding (May 15, 2024 – May 19, 2024)

- Clarify the problem.
- Grasp the business context and goals.
- Create a data intake report.

Data Understanding and Preparation (May 20, 2024 – May 26, 2024)

- Access and explore the provided dataset.
- Understand the structure, attributes, and relationships within the data.
- Identify data issues (such as missing values, outliers, and skewness) and devise strategies to address them.

Data Cleaning and Transformation (May 26, 2024 - June 2, 2024)

- Handle missing values, outliers, and inconsistent data.
- Perform data transformations and normalization.

Exploratory Data Analysis (June 3, 2024- June 9, 2024)

- Perform EDA on the dataset.
- Provide final recommendations based on the analysis.

EDA Presentation and Proposed Modeling Technique (June 10, 2024- June 16, 2024)

- Prepare an EDA presentation tailored for business users.
- Include recommendations for models suitable for technical users on the last slide.

Model Selection and Model Building/Dashboard (June 17, 2024- June 23, 2024)

- Model Development
    - Split the data into training and testing sets.
    - Select appropriate algorithms for classification.
    - Train multiple models and tune hyperparameters.
- Model Selection
    - Evaluate models using cross-validation.
    - Compare performance metrics such as accuracy, precision, recall, and ROC-AUC.
    - Select the best-performing model.

Final Project Report and Code (June 23, 2024- July 1, 2024)

- Document the entire process, including challenges faced and decisions made.
- Prepare a comprehensive report detailing the model's performance and business impact.
- Present findings to stakeholders.