

Data Explorer Team Details

Group Name: Data Explorer

Name: Mohammad Tohin Bapari

Email: tohin@gmx.de

Country: Germany

University: Bergische Universität Wuppertal

Specialization: Data Science

Problem Description

Pharmaceutical companies need to understand the persistency of drug usage as per physician prescriptions to improve patient adherence. ABC Pharma aims to automate the process of identifying factors impacting persistency.

Business Understanding

Persistency in medication adherence is critical for the effectiveness of treatments. By understanding and predicting persistency, pharmaceutical companies can design better patient support programs, improve health outcomes, and optimize resource allocation.

Project Lifecycle and Deadline

Problem Understanding: July 10, 2024

Data Understanding: July 15, 2024

Data Cleaning and Feature Engineering: July 20, 2024

Model Development: July 25, 2024

Model Selection: July 30, 2024

Model Evaluation: August 5, 2024

Reporting and Documentation: August 10, 2024

Model Deployment: July 15, 2024

Data Intake Report

Attribute	Description
Source	Provided by ABC Pharma
Format	Excel
Sheets	'Feature Description', 'Dataset'
Total Features	69 columns
Target Variable	Persistency_Flag
Feature Categories	Demographics, Clinical Factors, Provider Attributes, Adherence
Missing Values	No missing values detected
Categorical Variables	Gender, Race, Ethnicity, Region, Ntm_Speciality, Ntm_Specialist_Flag, Ntm_Speciality_Bucket, Risk_Family_History_Of_Osteoporosis, Risk_Low_Calcium_Intake, and others
Encoding	One-hot encoding applied to categorical variables
Train-Test Split	80% training set, 20% test set
Training Set Size	2739 samples, 116 features
Test Set Size	685 samples, 116 features
Initial Processing Steps	<ul style="list-style-type: none">- No missing values handled as there were none.- One-hot encoding applied to categorical variables.- Dropped unnecessary columns like 'PtId'.

GitHub Repo: <https://github.com/iamtohin/Data-Analytst-Internship-at-Data-Glacier/tree/main/Week%207>