DATA SCIENCE HEALTHCARE PROJECT

INDIVIDUAL PROJECT NAME: Health and Care

Name : Abina Azees

MAIL ID: abinaazees@gmail.com

Country : UAE

College: Amal Jyothi College of Engineering and Technology

Specialisation: Data Science

Problem Description

Pharmaceutical companies often struggle to assess how consistently patients adhere to their prescribed treatments, a concept known as **drug persistency**. This persistency plays a vital role in determining not only the **effectiveness of medical therapies** but also their **impact on patient outcomes** and the **commercial success** of the drugs.

To address this challenge, **ABC Pharma** has collaborated with an analytics firm to build a **predictive model** capable of identifying patients who are likely to be **persistent or non-persistent** with their therapies. The goal is to leverage a combination of **demographic, clinical, treatment-related, and adherence data** to uncover key factors influencing persistency. These insights will help enhance **patient engagement strategies**, optimize **treatment planning**, and ultimately improve **healthcare outcomes and business performance**.

Business Understanding

Drug persistency is a key indicator of how effectively patients adhere to their prescribed therapies. Gaining insights into persistency patterns provides valuable opportunities to enhance both **clinical outcomes** and **business performance**.

Why Drug Persistency Matters:

- **Improved Patient Targeting:** Enables more tailored and proactive interventions for patients at risk of non-adherence.

- **Optimized Sales and Marketing:** Facilitates data-driven strategies to promote therapies to the right patient segments.

- **Better Treatment Outcomes:** Helps reduce relapses and health complications by ensuring continuous medication intake.

- **Operational Efficiency:** Supports cost-effective resource allocation by focusing on patients who are less likely to persist.

Stakeholder Objectives:

- **ABC Pharma:**

  o Increase overall drug persistency rates

  o Enhance therapeutic effectiveness and patient satisfaction

  o Maximize market share and revenue growth

- **Analytics Partner:**

  o Build a robust, interpretable, and high-performing predictive model

  o Provide actionable insights into the drivers of persistency

  o Deliver a solution that can be seamlessly integrated into ABC Pharma's decision-making processes

Project Lifecycle

The project duration varies from June 19 to July 30.

Phase	Task Description
1 Problem Understanding	Define objectives, identify business and ML problems.
2 Data Understanding	Explore the dataset, identify features, understand variable types, distributions, and missing values.

Phase	Task Description
3 Data Cleaning & Feature Engineering	Handle missing values, encode categorical variables, normalize/scale numerical features, and create derived features if needed.
4 Model Development	Split data, train multiple classification models (e.g., Logistic Regression, Random Forest, XGBoost).
5 Model Evaluation & Selection	Evaluate using Accuracy, Precision, Recall, ROC-AUC. Choose best-performing model.
6 Model Deployment	Deploy the model using Flask or Streamlit (web app or API).
7 Reporting	Document process, create presentation, and write final report including: insights, performance metrics, feature importance, business impact, and challenges.
Final Submission	Submit GitHub repo, PDF report, and deployment link.

Data Intake Report

Name : Persistency of a Drug – Classification Model

Client: ABC Pharma

Date: June 19

Dataset Overview

Number of Observations: 3424

Total Features: 68

File Format: xlsx

Feature Group	Description	Example Variables
Patient Identification	Unique ID of patients	Patient_ID
Target Variable	Indicates persistency of therapy	Persistency_Flag
Demographics	Patient characteristics	Age, Race, Gender, Ethnicity, Region

Feature Group	Description	Example Variables
Provider Attributes	Prescriber-related features	NTM - Physician Specialty, IDN Indicator
Clinical Factors	Risk indicators and scan history	NTM - Risk Segment, Change in T Score, Dexa Scan Recency, Fragility Fracture During Therapy
Disease/Treatment Factors	Conditions and drug usage before/during therapy	NTM - Comorbidity, Glucocorticoid Usage, Injectable Experience, Concomitancy
Adherence	Overall adherence level	Adherence

Proposed Approach

- Perfom EDA on the Dataset
- Handle missing values
- Encode categorical variables
- Normalize/scale numerical features (if required)
- Feature engineering for derived insights (e.g. patient risk profiles)
- Train/test split and model development