Team Member Details

Group Name: *SoloVision Analytics*

Name: Ankita Roy
Email: roy.anki39@gmail.com
Country: India
College/Company: Institute: Great Learning, Gurgaon/ Internship: Data Glacier
Specialization: Data Science

Problem Description

Drug persistency is a key metric in determining the long-term success of a treatment regimen. It reflects whether patients continue using a prescribed therapy as advised. Pharmaceutical companies like ABC Pharma often struggle to identify the factors that cause patients to discontinue treatment.

To address this, ABC Pharma has sought the help of data science to automate the detection of persistency patterns among patients using their drug. The aim is to develop a **classification model** that predicts whether a patient is likely to persist with the treatment based on demographic, clinical, and behavioural data.

The binary target variable for this task is **Persistency_Flag**, where the goal is to classify patients as either **persistent** or **non-persistent**.

Business Understanding

High drug persistency is crucial for ensuring effective treatment outcomes, patient health, and business growth. A drop in persistency can signal potential issues such as side effects, poor patient experience, or lack of physician engagement. Identifying patients at risk of discontinuing therapy enables timely interventions.

Business Goals:

- Understand key drivers influencing drug persistency.

- Predict non-persistent patients early using data insights.

- Improve overall patient management and commercial strategies using ML models.

Project Lifecycle and Timeline

Phase	Description	Deadline
1. Problem & Business Understanding	Define objective, align data with business impact.	Week 1
2. Data Understanding	Explore dataset shape, data types, value distributions.	Week 2

Phase	Description	Deadline
3. Data Cleaning, Feature Engineering & EDA	Handle nulls, encode categoricals, feature transformation.	Week 3–4
4. EDA Presentation & Recommended Models	Make EDA presentation and recommend models to be used further	Week 5
5. Model Selection & Model Building	Train baseline models (Logistic Regression, Random Forest, XGBoost) and save final model	Week 6
8. Reporting & Insights	Summarize results, insights, challenges, and recommendations.	Week 7

Data Intake Report

- **Source**: Provided by ABC Pharma (or via Kaggle/Shared Drive)

- **Format**: CSV

- **Size**: [Enter file size if known]

- **Features**: 20+ columns including demographics, clinical history, risk factors, and therapy adherence

- **Target Variable**: Persistency_Flag

- **Preliminary Observations**:

    o Mix of categorical and numeric features

    o Missing values in several clinical indicators (e.g., T-Score, Risk Segment)

    o Potential class imbalance to be checked

    o Derived features like "change in risk segment" may need encoding

- **Actions Planned**:

    o Exploratory Data Analysis (EDA)

    o Imputation and encoding

    o Feature scaling and selection

    o Model training and validation

GitHub Repository Link
https://github.com/Ankita-ar/PERSISTENCY-OF-A-DRUG