Data Science project

Team members:

Name: Tomisin Abimbola Adeniyi E-mail: tomisin_adeniyi11@yahoo.com Country: Nigeria College/Company: Specialization: Data Science	Name: Fabio Pontecchiani E-mail: pontecchianifabio@gmail.com Country: Belgium College/Company: University of Sheffield Specialization: Data Science
Name: Bilikis Omolara Alayo E-mail: berlykis@gmail.com Country: United Kingdom College/Company: Specialization: Data Science	

Business Context: ABC Pharma Company faces the challenge of understanding drug persistence based on physician prescriptions. They want to automate the identification process to gain insights into the factors affecting persistence.

Problem Statement: The goal is to build a classification model using a given dataset. The target variable we're interested in is Persistency_Flag.

Key Objectives:

1. **Insights Gathering:** Understand the factors impacting drug persistence.
2. **Model Building:** Develop a classification model to predict the Persistency_Flag.

Dataset: The dataset contains relevant features related to drug usage, prescriptions, and other relevant information. The objective of this project is to analyze this data and create a model that can predict whether a patient will be a persistent drug user or not.

Persistency Flag:

- This binary target variable indicates whether a patient continues using the drug over time (True) or discontinues it (False).

Project lifecycle and Approach:

1. **Exploratory Data Analysis (EDA):** Explore the dataset to understand its structure, distributions, and relationships between features. (by week 8 - 26/04/24)
2. **Data Cleansing and Transformation:** try at least 2 techniques for NA values (by week 9 - 3/05/24)
3. **Feature Engineering:** Create relevant features or transform existing ones to improve model performance. (by week 10 - 9/05/24)
4. **Model Selection:** Choose appropriate classification algorithms (e.g., logistic regression, decision trees, random forests, etc.). (by week 11 - 16/05/24)
5. **Model Training and Evaluation:** Train the selected models, evaluate their performance using appropriate metrics (e.g., accuracy, precision, recall), and fine-tune hyperparameters. (by week 12 - 23/05/24)
6. **Insights Interpretation:** Analyze feature importance and interpret the model results to understand the impact of different factors on drug persistence. (by week 13 - 29/05/24)

Success Criteria: Build a well-performing classification model that accurately predicts the Persistency_Flag based on relevant features and provides actionable insights for ABC Pharma Company.

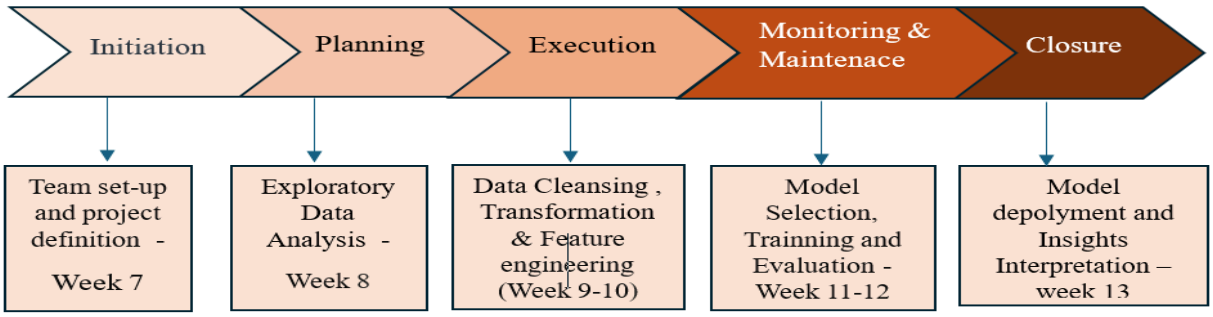

Fig 1: Project lifecyle

Data Intake Report

Name: <The Insights Team>

Report date: <19/04/2024>

Internship Batch:<LISUM31>

Version:<1.0>

Data intake by:<Bilikis O. Alayo>

Data intake reviewer:<Fabio Pontecchiani>

Data storage location:

<https://drive.google.com/file/d/1P_oMc6gOBlhw6dY5PxaqxV2swdHMUooK/view>

Tabular data details:

Healthcare_dataset.xlsx

Total number of observations	<3423>
Total number of files	<1>
Total number of features	<69>
Base format of the file	<.xlsx>
Size of the data	<1.8mb>