INTERNSHIP TASK WEEK 7

NAME: HIMANI ARYAN
EMAIL: himani1689@gmail.com
COUNTRY: UNITED KINGDOM
UNIVERSITY: UNIVERSITY OF LIVERPOOL

SPECIALIZATION: DATA SCIENCE

PROJECT TITLE:

Data Science: Healthcare - Persistency of a drug

1. **Problem Description:**

ABC Pharma's difficulty is to comprehend the persistence of drug use based on doctor prescriptions. By gaining understanding of the variables influencing persistency, they aim to automate the identification process. The provided dataset must be used to create a classification model in order to accomplish this.

2. **Business Understanding:**

   Pharmaceutical firms like ABC Pharma must comprehend medicine persistency. A higher rate of persistency results in better patient outcomes, more money coming in, and lower healthcare expenses. ABC Pharma can efficiently customise marketing campaigns, enhance patient adherence, and optimise treatment programmes by automating the identification process and examining factors affecting persistency.

   Understanding Drug Persistency for ABC Pharma:

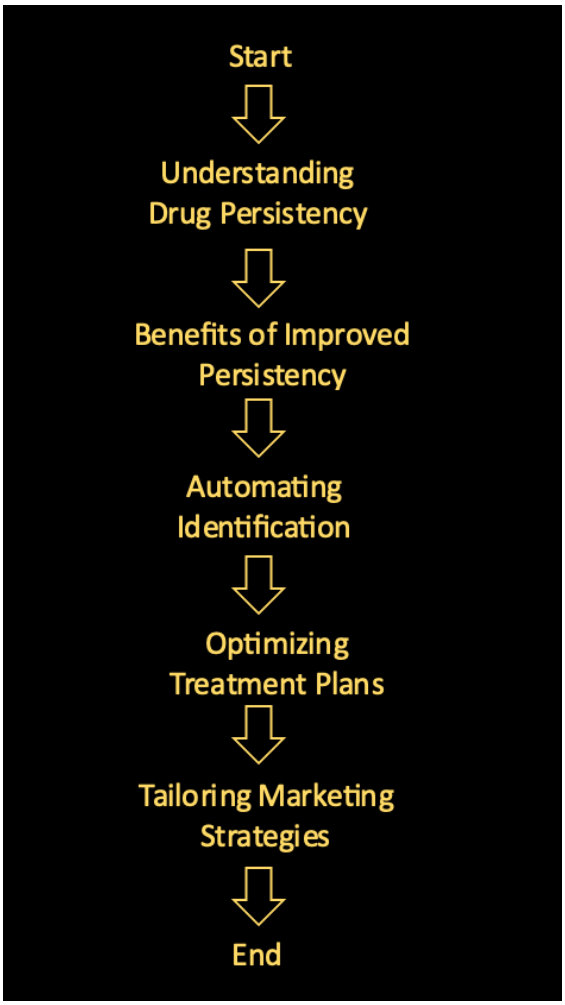

Figure 1: Flowchart describing drug persistence in pharmacology

↔ This flowchart illustrates the steps involved in comprehending drug persistency, its significance, and the advantages of enhancing persistency.
↔ It demonstrates how automating the identification process, factor analysis, and marketing strategy customization can improve patient outcomes, boost revenue, and lower healthcare expenses.

2.1 Detailed description of each step:

Importance of Drug Persistency
This step represents the core concept of why analysing drug persistency is significant in the context of pharmaceutical companies like ABC Pharma.

- **Crucial for Pharmaceutical Companies**: Understanding drug persistency is vital for pharmaceutical companies to gauge treatment effectiveness and patient satisfaction.
- **Improved Patient Outcomes:** Higher drug persistency leads to better treatment results and overall patient well-being.
- **Increased Revenue**: Improved drug persistency results in a more consistent revenue stream as patients continue taking medications for longer periods.
- **Reduced Healthcare Cost**: Effective drug persistency lowers healthcare expenses by reducing disease progression and the need for costly medical interventions.

Benefits of Improved Persistency
It covers the following points:

- **Better Patient Outcomes**: Improved drug persistency leads to more successful treatment results and overall well-being for patients.
- **Increased Revenue**: Higher drug persistency results in a more stable and consistent revenue stream for pharmaceutical companies.
- **Reduced Healthcare Costs**: Effective drug persistency lowers healthcare expenses by minimizing disease progression and the need for costly medical interventions.

Automating Identification

- The process involves analysing factors influencing drug persistency to streamline and automate the identification of patients with higher persistency levels.

Optimizing Treatment Plans

- This step focuses on tailoring treatment plans based on insights from drug persistency analysis, with the goal of enhancing patient adherence to prescribed medications and treatment regimens.

Tailoring Marketing Strategies

- In this step, pharmaceutical companies use insights from drug persistency analysis to customize their marketing approaches, targeting specific patient groups and implementing effective strategies to promote medication adherence and engagement.

3. Dataset:

↔ The dataset contains various variables categorized into different buckets, providing information related to patients, demographics, provider attributes, clinical factors, and disease/treatment factors.

↔ "Persistency_Flag," the target variable, shows whether a patient was persistent in taking the recommended medication or not. The dataset is probably going to have useful data that can be used to create a classification model that forecasts patient persistence. Prior to using machine learning techniques, it could be necessary to perform data cleansing, feature engineering, and cautious management of missing values and outliers.

Here is a brief description of each variable:

Demographics:

- Age: Age of the patient during their therapy.
- Race: Race of the patient from the patient table.
- Region: Region of the patient from the patient table.
- Ethnicity: Ethnicity of the patient from the patient table.
- Gender: Gender of the patient from the patient table.
- IDN Indicator: Flag indicating patients mapped to an Integrated Delivery Network (IDN).

Provider Attributes:

- NTM - Physician Specialty: Specialty of the Healthcare Professional (HCP) who prescribed the NTM (New Therapy Medication) Rx.

Clinical Factors:

- NTM - T-Score: T Score of the patient at the time of the NTM Rx (within 2 years prior from the Rx date).
- Change in T Score: Change in T score before starting with any therapy and after receiving therapy (categorized as Worsened, Remained Same, Improved, or Unknown).
- NTM - Risk Segment: Risk Segment of the patient at the time of the NTM Rx (within 2 years days prior from the Rx date).
- Change in Risk Segment: Change in Risk Segment before starting with any therapy and after receiving therapy (categorized as Worsened, Remained Same, Improved, or Unknown).

- NTM - Multiple Risk Factors: Flag indicating if the patient falls under multiple risk categories (having more than 1 risk) at the time of the NTM Rx (within 365 days prior from the Rx date).
- NTM - Dexa Scan Frequency: Number of DEXA scans taken prior to the first NTM Rx date (within 365 days prior from the Rx date).
- NTM - Dexa Scan Recency: Flag indicating the presence of Dexa Scan before the NTM Rx (within 2 years prior from the Rx date or between their first Rx and Switched Rx; whichever is smaller and applicable).
- Dexa During Therapy: Flag indicating if the patient had a Dexa Scan during their first continuous therapy.
- NTM - Fragility Fracture Recency: Flag indicating if the patient had a recent fragility fracture (within 365 days prior from the Rx date).
- Fragility Fracture During Therapy: Flag indicating if the patient had a fragility fracture during their first continuous therapy.
- NTM - Glucocorticoid Recency: Flag indicating the usage of Glucocorticoids (>=7.5mg strength) in the one-year look-back from the first NTM Rx.
- Glucocorticoid Usage During Therapy: Flag indicating if the patient had a Glucocorticoid usage during the first continuous therapy.

Disease/Treatment Factor:

- NTM - Injectable Experience: Flag indicating any injectable drug usage in the recent 12 months before the NTM OP (Outpatient) Rx.
- NTM - Risk Factors: Risk Factors that the patient is falling into. For chronic Risk Factors, a complete lookback is applied, and for non-chronic Risk Factors, a one-year lookback from the date of first OP Rx is used.
- NTM - Comorbidity: Comorbidities are divided into two main categories - Acute and chronic, based on the ICD codes. For chronic disease, a complete lookback is taken from the first Rx date of NTM therapy, and for acute diseases, the time period before the NTM OP Rx with a one-year lookback is applied.
- NTM - Concomitancy: Concomitant drugs recorded prior to starting with a therapy (within 365 days prior from the first Rx date).
- Adherence: Adherence for the therapies.

4. **Project Lifecycle:**

Problem Understanding and Data Collection:
- Deadline: 2 days from project initiation (by 21st July)
- Understanding the problem statement and gather the dataset from ABC Pharma or the analytics company.

Data Understanding and Exploration:

- Deadline: 5 days from project initiation (by 26th July)
- Performing exploratory data analysis (EDA) to understand the data's nature, distribution, and relationships between variables.
- Identifying the type of data available for analysis, such as numerical, categorical, or text data.
- Analysing the dataset for problems like missing values (NA values), outliers, skewed distributions, and other data quality issues.

Data Cleaning and Transformation:

- Deadline: 2nd August
- Addressing the problems identified in the data during the EDA phase.
- For handling NA values, will try at least 2 techniques such as mean, median, mode imputation, model-based imputation, or Weight of Evidence (WOE) approach depending on the nature of the missing data and the context of the problem.
- For handling outliers, will apply at least 2 techniques such as z-score, IQR method, or model-based approaches like robust regression to identify and handle the outliers appropriately.

Model Development and Selection:

- Deadline: 9th August
- strategy would be to split the dataset into training and testing sets.
- Selecting appropriate machine learning algorithms for classification.
- Training multiple models and fine-tune their hyperparameters using techniques like cross-validation and grid search.
- Evaluating the models based on appropriate metrics and select the best-performing one.

EDA and Final Recommendation:

- Deadline: 9th August
- will conduct a comprehensive exploratory data analysis (EDA) to gain insights into the dataset, relationships between variables, and potential patterns.
- will summarize key findings from the EDA and provide actionable recommendations based on the insights to address the persistency problem for ABC Pharma.

Model Evaluation:

- Deadline: 16th August
- will evaluate the selected model on the testing dataset to assess its performance.
- Calculate accuracy, precision, recall, and ROC-AUC for both classes of the target variable (persistent and non-persistent patients).

EDA Presentation for Business Users:

- Deadline: 16th August
- will create an EDA presentation for business users, showcasing the important insights and findings from the data analysis.

Following is a Gantt Chart for the progress of the project:

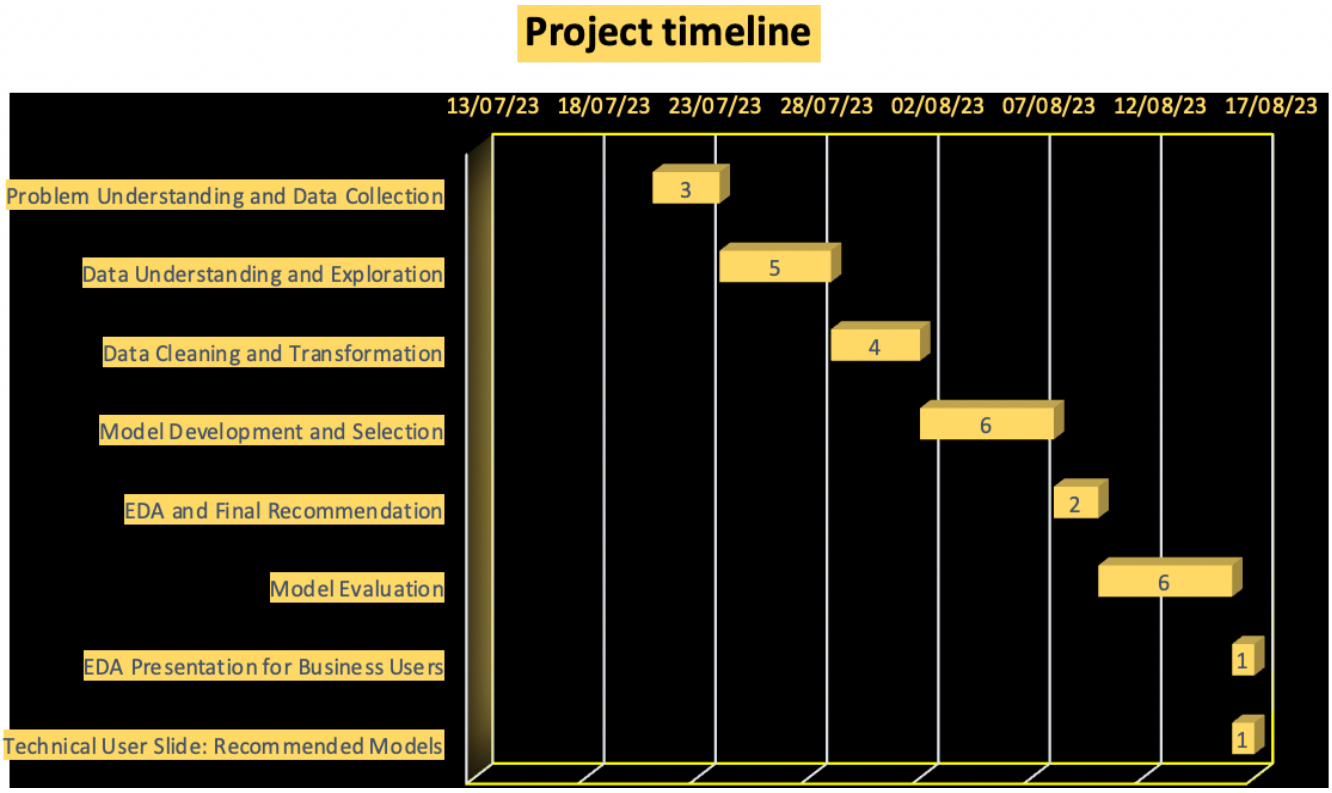

Figure 2: Project timeline