

# Data Intake Report - Data Science Healthcare - Persistency of a Drug – Classification – Week 7

Group Name: <Igor Azevedo de Queiroz>

Email: <[igor\\_queiroz17@yahoo.com.br](mailto:igor_queiroz17@yahoo.com.br)>

Country: <Ireland>

College: <Dublin Business School>

Specialization: <Data Science - Classification>

Report date: <20/09/2021>

Internship Batch:<LISUM02>

Version:<1.0>

Data storage location: <https://github.com/IgorQueiroz32/Data-Science-Healthcare---Persistency-of-a-Drug-Classification/tree/main/week%207>

Problem Description: <Data Science Healthcare - Persistency of a Drug - Classification>

## Tabular data details:

<b>Total number of observations</b>	< 3424rows >
<b>Total number of files</b>	<1>
<b>Total number of features</b>	< 69 columns >
<b>Base format of the file</b>	<.ipynb, .csv, .txt, .png>
<b>Size of the data</b>	<891 in KB>

# Healthcare - Persistency of a Drug - Classification

## 1. Business Description/ Problem.

One of the challenges for all Pharmaceutical companies is to understand the persistency of drug as per the physician prescription. To solve this problem ABC pharma company approached an analytics company to automate this process of identification.

With an objective to gather insights on the factors that are impacting the persistency, it is necessary to build a classification for the given dataset, using the variable 'Persistency\_Flag' as target variable and other attributes as prediction variables.

## 2. Business Understanding.

ABC it is a private pharma company. Due to the problem to the persistency of drug as per the physician prescription, a data science project is applied to predict the classification of 'Persistency\_Flag' variable. In other words, based on the previously patients characteristics it is possible predict if futures patients will use the drugs during the role treatment or if they won't.

The object of this project is providing answer of the main questions made by the company's CEO, which are:

- What is the 'Persistency\_Flag' classification for future patients?

The answer for those questions is presented in two different methods:

- A webapp with all necessary prediction attributes in order to predict the classification of the 'Persistency\_Flag' for future patients.
- A dashboard with several hypotheses and insights to help the company CEO with future decisions.O decisions.

The tools used for this project are: Python 3.8, Pycharm, Jupyter Notebook, Streamlit and Heroku.

### 3. Data Understanding.

There is 1 dataset provided:

<https://www.kaggle.com/harbhajansingh21/persistent-vs-nonpersistent>

healthcare\_dataset.csv – this file includes characteristics of several patients.

Variables Description:

Here I'm describing the columns in detail:

Patient Details:

- **Patient ID:** Unique ID of each patient;
- **Persistency\_Flag:** Flag indicating if a patient was persistent or not;
- **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 IDN;

Provider Attributes:

- **NTM - Physician Specialty:** Specialty of the HCP that prescribed the NTM Rx;

Clinical Factors:

- **NTM - T-Score:** T Score of the patient at the time of the NTM Rx (within 2 years prior from rxdate);
- **Change in T Score:** Change in Tscore before starting with any therapy and after receiving therapy (Worsened, Remained Same, Improved, Unknown);
- **NTM - Risk Segment:** Risk Segment of the patient at the time of the NTM Rx (within 2 years days prior from rxdate);

- **Change in Risk Segment:** Change in Risk Segment before starting with any therapy and after receiving therapy (Worsened, Remained Same, Improved, Unknown);
- **NTM - Multiple Risk Factors:** Flag indicating if patient falls under multiple risk category (having more than 1 risk) at the time of the NTM Rx (within 365 days prior from rxdate);
- **NTM - DEXA Scan Frequency:** Number of DEXA scans taken prior to the first NTM Rx date (within 365 days prior from rxdate);
- **NTM - DEXA Scan Recency:** Flag indicating the presence of DEXA Scan before the NTM Rx (within 2 years prior from rxdate 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 rxdate);
- **Fragility Fracture During Therapy:** Flag indicating if the patient had fragility fracture during their first continuous therapy;
- **NTM - Glucocorticoid Recency:** Flag indicating usage of Glucocorticoids ( $\geq 7.5$ mg strength) in the one year look-back from the first NTM Rx;
- **Glucocorticoid During Therapy:** Flag indicating if the patient had a Glucocorticoid usage during the first continuous therapy;

#### Disease/Treatment Factors:

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

## **4. Project lifecycle along with deadline.**

- Problem understanding
- Data Understanding
- Data Cleaning and Feature engineering
- Model Development
- Model Selection
- Model Evaluation

All those steps are done.