# Data Science Project

# Healthcare – Persistency of a Drug

# Week 7 works

Team member’s details:

Group Name: DS\_SS

Name: SONIYA SUNNY

Email: [soniyasunny1210@gmail.com](mailto:soniyasunny1210@gmail.com)

Country: Canada

College/Company: N/A

Specialization: Data Science

Problem Description

Data Science:: Healthcare - Persistency of a drug:: Group Project

**Problem Statement:**

One of the challenge 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.

**ML Problem:**

With an objective to gather insights on the factors that are impacting the persistency, build a classification for the given dataset.

**Target Variable:** Persistency\_Flag

**Task:**

* Problem understanding
* Data Understanding
* Data Cleaning and Feature engineering
* Model Development
* Model Selection
* Model Evaluation
* Report the accuracy, precision and recall of both the class of target variable
* Report ROC-AUC as well
* Deploy the model
* Explain the challenges and model selection

Business Understanding

Project Lifecycle along with deadline

Data Intake Report

Name: Data Science Final Project – ‘Healthcare – Persistency of a Drug’

Report date: July 16, 2022

Internship Batch: LISUM10: 30

Version:<1.0>

Data intake by: Soniya Sunny

Data intake reviewer:<intern who reviewed the report>

Data storage location: [Healthcare\_dataset.xlsx - Google Drive](https://drive.google.com/file/d/1P_oMc6gOBlhw6dY5PxaqxV2swdHMUooK/view)

**Tabular data details:**

|  |  |
| --- | --- |
| **Total number of observations** | 3425 |
| **Total number of files** | 1 |
| **Total number of features** | 69 |
| **Base format of the file** | .xlsx |
| **Size of the data** | 899 KB |

Github Repo Link