# Data Science Project

# Healthcare – Persistency of a Drug

# Week 12 works

Team member’s details:

Group Name: DS\_SS

Name: SONIYA SUNNY

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

Country: Canada

College/Company: Data Glacier

Specialization: Data Science

Problem Description

To identify the persistency of a drug, a pharmaceutical company approached to develop a model based on data analysis. Factors that affect the persistence of drugs should be identified, along with data insights with predictive analytics, to help the company for their smooth and efficient functioning, with the help of dataset provided by the company.

Github Repo Link

[Final\_Project\_DS\_SS/week\_12 at master · Soniyasunny1/Final\_Project\_DS\_SS (github.com)](https://github.com/Soniyasunny1/Final_Project_DS_SS/tree/master/week_12)

# Week 12: Model Selection and Model Building

Select your base model and then explore 1 model of each family if its classification problem then 1 model for Linear models, 1- Model for Ensemble, 1-Model for boosting and other models if you have time (like stacking)

Please make sure selected model fits in your business requirement. For example : If your business does not want black box model then select only those models which can be used to explain the prediction.

As this is group assignment hence upload the code of each team member and other deliverables in the single repo and share the URL of that repo.

**you are allowed to merge the code of each individual and work together to get good result.**

Data Intake Report

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

Report date: August 12, 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 |