Name: Jiahao Yu

Email: jiahao.yu.2003@outlook.com

Country: United Kingdom

College/Company: University of Bristol

Specialization: Data Science

Business understanding:

This project aiming at understanding the persistency of drug based on the data provided. Use Machine Learning approaches to gather insights on the affecting features and build a classification model.

Project lifecycle along with deadline:

|  |  |  |
| --- | --- | --- |
| Week | Deliverables | Deadline |
| 7 | Format(s): pdf | 19-Nov-22 |
| Business understanding |
| Project lifecycle along with deadline |
| Data Intake report |
| 8 | Format(s): pdf | 26-Nov-22 |
| Data understanding |
| Type of data for analysis |
| Problems in the data |
| Approaches to overcome problems |
| 9 | Format(s): pdf, ipynb | 02-Dec-22 |
| Data cleansing and transformation |
| At least 2 techniques to clean the data |
| 10 | Format(s): pdf, ipynb | 09-Dec-22 |
| EDA performed on the data |
| Final Recommendation |
| 11 | Format(s): pptx | 16-Dec-22 |
| EDA presentation for business users |
| Proposed modeling technique |
| 12 | Format(s): ipynb | 23-Dec-22 |
| Model Selection |
| Model Building |
| 13 | Format(s): pptx, ipynb | 30-Dec-22 |
| Final Project Report |
| Code |

Data Intake Report

Name: Healthcare - Persistency of a drug

Report date: 14 November 2022

Internship Batch: LISUM14

Version: 1.0

Data intake by: Jiahao Yu

Data intake reviewer:

Data storage location:

**Tabular data details:**

Healthcare\_dataset.xlsx

|  |  |
| --- | --- |
| **Total number of observations** | 3424 |
| **Total number of files** | 1 |
| **Total number of features** | 68 |
| **Base format of the file** | .xlsx |
| **Size of the data** | 920 KB |