Project: Data Science :: Persistency of a Drug

Week 7: Deliverables

Presented by: Sammy Suliman, Olivia Foster, and Tahsin Azad

Table of Contents:

Team Member Details:

Group Name: Call it Version 1.0

Member 1:

- **Name:** Sammy Suliman
- **Email:** sammysuliman@gmail.com
- **Country:** USA
- **College/Company:** UC Santa Barbara
- **Specialization:** Data Science

Member 2:

- **Name:** Olivia Foster
- **Email:** livia.n.foster@gmail.com
- **Country:** USA
- **College/Company:** Frisco Independent School District
- **Specialization:** Data Science

Member 3:

- **Name:** Tahsin Azad
- **Email:** tahsinazad31@gmail.com
- **Country:** USA
- **College/Company:** University of California, Santa Barbara
- **Specialization:** Data Science

Problem Statement:

One of the challenges faced by Pharmaceutical companies is the persistence of a drug (that is, the extent to which a patient will act in accordance with the prescribed time interval, and dose of a medication) as the physician prescribed it. In this problem, we will automate the process of classifying factors that determine the persistence of a drug through Machine Learning and Python.

Drug persistence is a task of classifying different disorders and a patient's medical history to determine the dose and length of dose. In order to train our model, we will need to classify risk factors, medical histories, and disorders. To do this, we will be using a dataset based on over 3000 patients' records.

Project Deadlines:

Week:	Due Date:	Plan:
Week 7	1/19/2023	Problem Statement, Data Collection, Data Report
Week 8	1/26/2023	Data Preprocessing
Week 9	2/2/2023	Feature Extraction
Week 10	2/9/2023	Building Model
Week 11	2/16/2023	Model Result Evaluation
Week 12	2/23/2023	Flask Development and Web Application
Week 13	2/28/2023	Final Submission (Report + Code + Final Submission)

Data Intake Report:

Name: Final Project -- Data Science:: Healthcare - Persistency of a drug:: Group Project
Report date: 1/17/2023
Internship Batch: LISUM16
Version:<1.0>
Data intake by: Olivia Foster
Data intake reviewer: Tahsin Azad
Data storage location:
https://github.com/LiviaNFoster/DataGlacierFinalProject/blob/main/Healthcare_dataset.xlsx

Tabular data details:

Total number of observations	3424
Total number of files	1
Total number of features	69
Base format of the file	xlsx
Size of the data	922 KB

Proposed Approach:

- Unique Row identified using with the "Ptid" column, a unique ID for each patient
- No duplicates of Unique ID confirmed
- Assumptions
    - No false positives or errors during any of the testing
    - No dishonest and unbiased data recorded

Github Repo link:

https://github.com/LiviaNFoster/DataGlacierFinalProject.git