

Week 7 Project Deliverables

Group Name: Decent Healthcare Analytics

Name: Tolga Yaz

E-Mail: tolgayaz1991@gmail.com

Country: Turkey

Specialization: Data Science

Problem Description and Business Understanding:

- One of the challenges for pharmaceutical companies is to understand the persistency of drugs as per the physician prescription.
- To solve this problem a pharma company wants to automate this process of identification via using analytics.
- **Main Goal:** Predicting drug persistency of patients to provide better healthcare.
- **Main Objective:** Building a classification model to predict drug persistency of patients via using provided data and data science methods.

Project Lifecycle:

Objective	Deadline
Problem Description, Business Understanding and Data Intake Report Preparation	8 Nov 2021
Data Understanding Report	8 Nov 2021
Data Cleansing and Transformation	15 Nov 2021
Exploratory Data Analysis (EDA)	22 Nov 2021
EDA Presentation and Proposal of Modeling Technique(s)	29 Nov 2021
Model Selection and Model Building	6 Dec 2021
Evaluation of Results and Final Project Report	18 Dec 2021

Data Intake Report

Name: Healthcare - Persistency of a drug

Report date: 04.11.2021

Internship Batch: LISUM04

Version: 1.0

Data intake by: Tolga Yaz

Tabular data details:

Name of the File	Healthcare_dataset.xlsx
Total number of observations	3424
Total number of features	69
Base format of the file	.xlsx
Size of the data	893 kB

Proposed Approach:

- The data was transformed into Python Pandas DataFrame to be processed.
- The dataframe were observed carefully to identify the nulls.
- .duplicated() method of Python Pandas Library was used to identify duplicates.