ML on colab cloud

Group members :(Zaid Mazen 0206186, Karam Alhanatleh 0201517)

Instructor name: Dr. Shernaz AL-HajBaddar

2022/2nd semester

## Executive Summary

In this project we had applied the supervised ML (machine learning) Logistic Regression model on the colab cloud environment with two different algorithms(newton-cg, lbfgs)that is supported by Google.

We have imported dataset from Kaggle that shows Patients and the drug that each one take based on his case.

then we used classification model to predict the drug that the Patient should take depending on some attributes about each of them shown in the database.

then we showed the model results and performance in graphs for each algorithm we used that showed that the model is more accurate at newton-cg.

**Introduction**

The problem that we are about to solve is that we have some attributes about some patients that they take medicine, and we should tell what is the right medicine that the patients should take.

ML (supervised) solution is very helpful in this case because we have some information about the patients that we can teach the machine to decide the correct drug for each patient depending on his information that is available. This will save a lot of time and effort on the doctors and the patients.

Google offer free access cloud environment with limited sources for public using python programming language that support ML on jupyter notebooks .

**Methodology**

**Model Description**

We have used Logistic Regression that is considered as a classification (supervised) model that is a process of modeling the probability of a discrete outcome given an input variable.

We tried with two different algorithms the first is the default one(lbfgs) and the second is(newton-cg, ).

Data is divided into 70% training (140 records) and 30% testing (60 records) randomly.

**Dataset Description**

We chose small and easy demand database for focusing more and ML understanding.

Drugs database imported from Kaggle contains 200 records of patients that have six attributes 1. Age (max 74, min 15), 2.sex (male, female) 3. blood pressure (high, normal, low)

4.cholostrol (high, normal, low) 5. sodium potassium pump [na\_to\_k] 5. Drug (A, B, C, X, Y)

First thing we uploaded the dataset as a data frame to make it easier to deal with

Then we used some function know more about the data:

1.Head: to show the data and its attributes

2. unique to show how many types of drugs we have

3.describe to make fast analysis showed that the min age is 15 and the max is 74

And na\_to\_k (max:38, min:6)

4.value\_count to showed that we have (drugY:91, drugX:54, drugA:23, drug C:16, drug B:16)

**Chart

Description automatically generatedTable

Description automatically generated**

Figure :report for first algo Figure : results for first algo

**Chart

Description automatically generatedTable

Description automatically generated**

Figure :report for second algo Figure :result for second algo

**Conclusions**

First experiment algorithm one **(**lbfgs)

Accuracy: 86.66666666666667% RAM: 1.42 DISK:38.34

Second experiment algorithm two(newton-cg)

Accuracy : 98.33333333333333% RAM: 1.34 DISK:38.34

AS the results shows that the second(newton-cg, ) algorithm way better than the first one(lbfgs) from all sides RAM usage and accuracy

Based on working on colab environment that made the work easier because its free and helpful for small project specially ML projects

This shows that google provide users with good resources to fit and run their projects

# References

Some of the packages used during the project:

matplotlib.pyplot

pandas

sklearn.linear\_model

google.colab

seaborn

The website we imported the dataset from: [www.kaggle.com](http://www.kaggle.com)

# Date and sign

Karam alhanatleh: 07/05/2022

Zaid aburasheid: 2022/05/07