Epoch Core Selection - Round 1

Hackathon

17th June, 2022

1 Problem Statement

You are given a red-wine dataset. You can find the dataset here:
Epoch Hackathon Dataset - GitHub Link

This is a classification problem. Your goal is to find a model that can classify the wine as of:

1. high quality (1) or
2. low quality (0)

1.1 More details on the dataset:

The dataset includes 15 measured parameters such as amount of chlorides, sulphates, acidity, etc.,

The target variable for the dataset is quality.

1.2 Guidelines:

Any model(s) used for training and classification **should be written entirely from scratch and no in-built libraries apart from numpy, pandas and matplotlib should be used**. This is to check your understanding of mathematical concepts behind the algorithms.

2 Criteria for Evaluation

Your submission will be judged on the basis of the following parameters:

2.1 Originality and Quality of your Model

The submissions will be subjected to plagiarism checks in advanced softwares. If an applicant is found to have plagiarised, he/she will be disqualified from the selection process. So, please refrain from copying from the sources in the internet or from your friends.

2.2 Accuracy of your model

You are expected to keep a separate test dataset to report the final accuracy of your model. Have a train:test ratio of 80:20 for the test dataset that you are using in your report.

2.3 Data Pre-processing

Data Pre-processing, Exploratory Data Analysis and Feature Engineering are to be done before the training of your model. The relevant plots and results of exploratory data analysis should be included.

2.4 Report

You are expected to attach a final report explaining your code and the algorithms used. The documentation should contain an explanation of your approach and explanation of code snippets wherever required.

You have to include the accuracy of your model along with this.

3 Submission Details

The submission must be a single zip file containing the .ipynb file along with the code documentation and report as a .pdf file.
Please make sure you are naming the file with your roll number, so that we can evaluate easily.
Even if you aren't able to complete the entire code, we encourage you to submit whatever work you have done.

You can submit the files in this google form:
Hackathon Submission - Google Forms Link

Deadline: 11:59 PM, 23rd June 2022

4 Useful References

You may find these resources useful:

1. Andrew NG's Stanford CS229 - Machine Learning Course
2. Data Pre-Processing Concepts
3. Logistic Regression
4. Classification Algorithms

All the best for the Hackathon!