**Project: Evaluation of Student Performance**

**Introduction**

**Dataset Description**

This data approach student achievement in secondary education of two Portuguese schools. The data attributes include student grades, demographic, social and school related features) and it was collected by using school reports and questionnaires. Two datasets are provided regarding the performance in two distinct subjects: Mathematics (mat) and Portuguese language (por). In [Cortez and Silva, 2008], the two datasets were modeled under binary/five-level classification and regression tasks. Important note: the target attribute G3 has a strong correlation with attributes G2 and G1. This occurs because G3 is the final year grade (issued at the 3rd period), while G1 and G2 correspond to the 1st and 2nd period grades. It is more difficult to predict G3 without G2 and G1, but such prediction is much more useful (see paper source for more details).

In this project, we abled to train two models or algorithms on student's data, which are a Support vector machine and a Decision tree to make a model predict the Final Grade for new students easily through taking some data about students there are two types of prediction.

The first is regression we used it to test training accuracy of models, and the second type is Classification we used it also to test training accuracy of models, and in the end will show you result, what I did to enhance my model to predict well.

before start at apply algorithms, I did:

* Marge two files to one
* Remove missing data.
* Remove duplicate data.
* Modify data type for all columns.
* Define correlation between G3 and all columns others.
* Convert columns to binary columns.
* Applied summary statistical description.
* Visualization some main columns

Then after we did all processes to make the data ready then we will do follow.

**Split data to Training-set and test-set**

We used this code to do that, as the following figure.

Graphical user interface, text, application, email

Description automatically generated

Figure1: Code-For-Split-data

after we splatted the data into two train-set and test-set became train-set size 835 rows and test-set 208, We will make a model and train it on train-set data and use test-set to test it accuracy.

**Make Model ML or Apply Algorithms To Regression the output (G3)**

First model is SVR:

This algorithm require parameters to build it like `kernel` , `degree` ,`gamma`, `C`

To choose their params easily using `GridSearch` algorithm which is define the best parameter with my data to choose it, and I used this code as following picture.

Graphical user interface, text, application

Description automatically generated

As shown at prev picture the best param for SVR algorithm is when C equal 9, then we used that and train a model with train-set data then the result is the training accuracy equal 86%

This was well but when I test model on another data (test-set data) the testing accuracy equal 38%





There are overfitting in my model because it train well but when testing it is bad so we can use another algorithm like Decision Tree

Second model is DecisionTreeRegressor:

This algorithm requires parameters to build it like `max-depth’, ‘min-samples-leaf`, `min-samples-split`

To choose their params easily using `GridSearch` algorithm which is define the best parameter with my data to choose it, and I used this code as following picture.

Graphical user interface, text, application

Description automatically generated

As shown at prev picture the best param for DecisionTreeRegressor() algorithm is when `max-depth` equal 4, then we used that and train a model with train-set data then the result is the training accuracy equal 87%

This was well but when I test model on another data (test-set data) the testing accuracy equal 44%





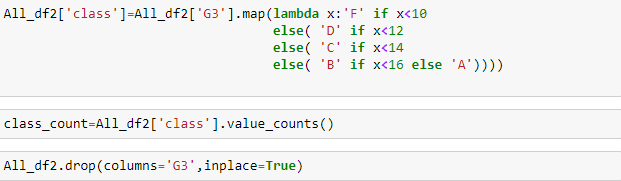
There are also overfitting in my model because it train well but when testing it is bad so may be the data that I have small too much to train a models or may be use another idea to raise the accuracy when I make G3 as Class which has 5 types to classify the output G3 like

{`A`,`B`,`C`,`D`,`F`}

This all code are used to make two models at stage the regressionText

Description automatically generated

If the column G3 < 10 =>F , G3 < 12 =>D , G3 < 14 =>C , G3 < 16 =>B ,G3 > 16 =>A let’s do it using code.



Then we do split data again to train-set and test-set and will repeat some step to train new models

**Make Model ML or Apply Algorithms To Classification the output (G3)**

First model is SVC:

This algorithm require parameters to build it like `kernel` , `degree` ,`gamma`, `C`

We will use GridSearch to choose params for this model using code as following

Text, letter

Description automatically generated

As shown at prev picture the best param for SVC algorithm is when C equal 9, then we used that and train a model with train-set data then the result is the training accuracy equal 82%

This was well and when I test model on another data (test-set data) the testing accuracy equal 71% also well

In this step the model became work well a little from the previous model SVR, and I will try Decision tree classifier also

Second model is DecisionTreeClassifier:

This algorithm requires parameters to build it like `max-depth’, ‘min-samples-leaf`, `min-samples-split`

To choose their params easily using `GridSearch` algorithm which is define the best parameter with my data to choose it, and I used this code as following picture.

Graphical user interface, text, application

Description automatically generated

As shown at prev picture the best param for DecisionTreeClassifier() algorithm is when `max-depth` equal 3, then we used that and train a model with train-set data then the result is the training accuracy equal 77%

This was well and when I test model on another data (test-set data) the testing accuracy equal 72%

And also in this step the model became work well a little from the previous model Decision Tree regressor, let’s go to predict to know which models is prefect SVC or DTC

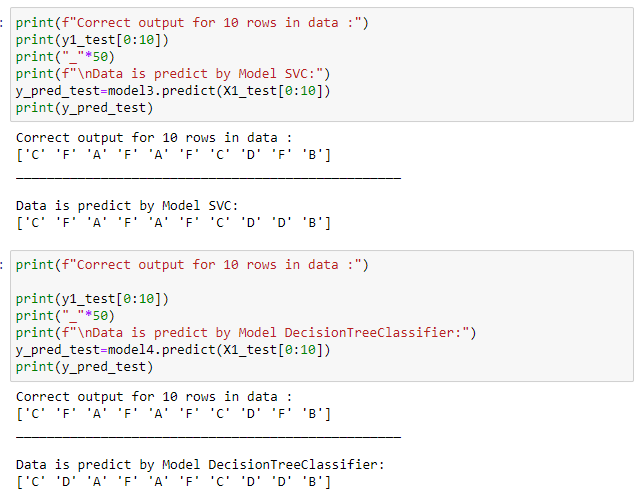
This all code are used to make two models at stage the Classification

A picture containing text

Description automatically generated

Test Predict Model

We abled to use this code as following to test that



As shown at prev picture the best model is **SVC** because it can predict General good.