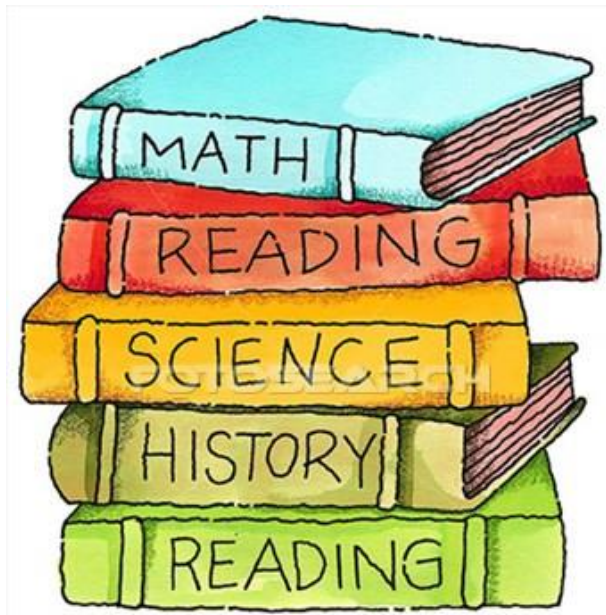


Final Grades Prediction



Classification

Model- 3

Prepared by:

Sara AlAbdulsalam

Fatimah AlShammari



Introduction:

Students, parents, and instructors seek to have an early alert of the performance of students to take actions based on it. This problem can be carried through machine learning classification model such as (KNN, Logistic Regression, and Decision Tree) to predict if the student will pass or fail which also makes the student knows whether he/she is in a position to reach his/her expectations or not. If the model shows that the student is going to fail and needs to improve then that student, with the help of the parents and instructors can prepare more for that semester to pass it. Lastly, we can get insights of what student features such as demographic, social, and school related attributes can affect student grades.



Data Description:

The dataset used to solve the problem is publicly available on Kaggle [1]:

The predictors:

School: student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)

Gender: student's sex (binary: 'F' - female or 'M' - male)

Age: student's age (numeric: from 15 to 22)

Address: student's home address type (binary: 'U' - urban or 'R' - rural)

Famsize: family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)

Pstatus: parent's cohabitation status (binary: 'T' - living together or 'A' - apart)

Medu: mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)

Fedu: father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)

Mjob: mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')

Fjob: father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')

Reason: reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')

Guardian: student's guardian (nominal: 'mother', 'father' or 'other')

Traveltime: home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)

Studytime: weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)

Failures: number of past class failures (numeric: n if $1 \leq n < 3$, else 4)

Schoolsup: extra educational support (binary: yes or no)

Famsup: family educational support (binary: yes or no)

Paid: extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)

Activities: extra-curricular activities (binary: yes or no)

Nursery: attended nursery school (binary: yes or no)

Higher: wants to take higher education (binary: yes or no)

Internet: Internet access at home (binary: yes or no)

Romantic: with a romantic relationship (binary: yes or no)

Famrel: quality of family relationships (numeric: from 1 - very bad to 5 - excellent)

Freetime: free time after school (numeric: from 1 - very low to 5 - very high)

Goout: going out with friends (numeric: from 1 - very low to 5 - very high)

Dalc: workday alcohol consumption (numeric: from 1 - very low to 5 - very high)

Walc: weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)

Health: current health status (numeric: from 1 - very bad to 5 - very good)

Absences: number of school absences (numeric: from 0 to 93)

the response variable will be to predict whether the student will pass or fail based on his/her grades for the three semesters periods which are the following:

G1 - first period grade (numeric: from 0 to 20)

G2 - second period grade (numeric: from 0 to 20)

G3 - final grade (numeric: from 0 to 20)

All the above grades will be scaled to 100 .

If the summation of the grades are at least 60% of the total grade the student will be classified as pass, otherwise fail.



Tools Description:

To achieve our goal, we will analyze and explore the data in Python by using Jupyter, and we will use different packages such as: SKLearn for modeling, Pandas, Matplotlib, Seaborn, and numpy.



Conclusion:

We aim to know the students' grades based on the students' social status and external influences by using machine learning classification models to predict if the student will pass or fail. In this document, we reviewed the problem that we want to solve, a description of the data we will work on, and finally the tools that we will use.



References:

- [1] <https://www.kaggle.com/dipam7/student-grade-prediction>