STUDENT PERFORMANCE FACTOR

A Machine Learning Analysis

INTRODUCTION

Education is, in this modern era, a primary need for every people, without which one may find it hard to survive and thrive in nowadays society. Education ideally helps one on achieving stability in life, by helping them acquire knowledge, improving their critical thinking skills, and provides them with a broad understanding of the world Formal education is typically obtained from school.

There are countless of factors that may affect one's performance in school, both internally and externally. This research analysis aims to find which factors affects a student's performance the most, from both internal and external factors. Through these findings,it is hoped that people can be more aware of these factors and may further help children around them in performing well in school. Tests area of measuring one's performance in school.

DATASET

The dataset used in this research was taken from kaggle.com. This dataset contains informations related to the student's backgrounds and their scores on 3 math tests. The mean of the three will then be analyzed.

Variables:

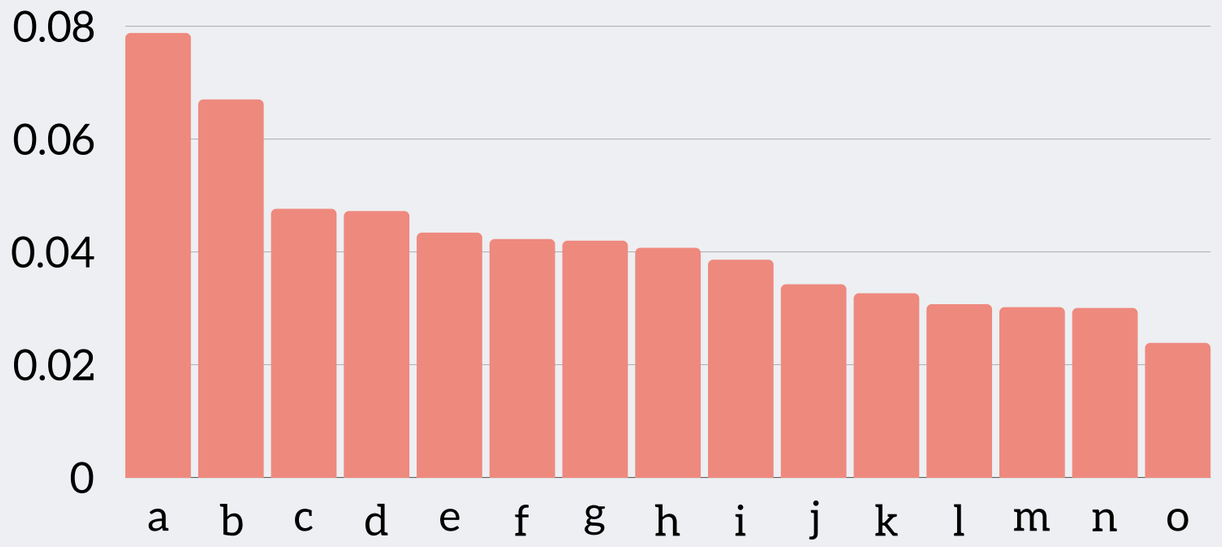
- school - school name
- sex
- age
- address - urban or rural
- famsize - family size
- pstatus - living together/apart
- medu & fedu - mother and father education
- mjob & fjob - mother and father job
- reason - why the school is chosen
- guardian
- traveltime - from home to school
- studytime
- famsup - family education support
- paid - extra paid class
- activities - extracurricular
- nursery - attended nursery school
- higher - wants to take higher education
- internet - access at home
- romantic - in a relationship
- famrel - quality of family relationship
- freetime
- goout - with friends
- dalc - weekday alcohol consumption
- walc - weekend alcohol consumpttion
- health - health status
- absence - number of absence
- G1 - 1st math score
- G2 - 2nd math score
- G3 - 3rd math score

The average of all G1, G2, and G3 is then calculated and classified into two categories, which are pass and fail. The student is considered pass when the average is above or equal to 10 or 50% of the total score and fail if otherwise.

METHODOLOGY

To find the top 15 most contributive factors of student performance, we use the random forest importance algorithm. The variables will then be used to perform a classification task, to predict whether or not a student will fail a class or not given their background, using the models K-Nearest Neighbor, Decision Tree, and Naive-Bayes.

RESULT



a. failures
b. absence
c. goout
d. Fedu
e. Medu
f. freetime
g. age
h. health
i. Walc
j. famrel
k. studytime
l. schoolsup
m. Dalc
n. sex
o. traveltime

The bar graph above shows the importance scale of the 15 most contributive factors to the outcome. From this graph, it is safe to conclude that a student's past failures and class absence has a high chance of contributing towards their performance in class.

From the three classification models, here are the performance comparison:

	accuracy	precision	recall	f1
KNN	0,72	0,72	0,84	0,78
DT	0,64	0,66	0,80	0,72
NB	0,58	0,58	1	0,74

In this dataset, it can be seen that KNN works best, as seen from its high accuracy, precision, recall, and f1 score compared to other models.