

Dataset name: Student Grades

Dataset file: student_grades.csv

Response variable: Pass

Scenario

Student performance in assessments is a complex problem to solve, with many factors able to contribute to the outcomes achieved by a student. However, effective prediction of student pass rates and the identification of the factors most closely associated with high grades being achieved may allow for a better understanding to be gained and best practice to be developed. The dataset presents data on a variety of heterogenous factors, and the assessment outcome for the student (given as a pass or a fail to an assignment). Your task is to build a classification model capable of predicting the value of the target variable, and to attempt to identify which are the most important features associated with a student passing or failing the assessment.

Data Dictionary

Feature ID	Feature name	Feature label	Data type	Response feature
school	School attended	Name of school (abbreviated)	Categorical	
sex	Sex	Student's sex	Categorical	
age	Age	Student's age	Numeric	
address	Address type	Student's home address type (u = urban, r = rural)	Categorical	
famsize	Family size	Student's family size (LE3 = less or equal to 3, GT3 = greater than 3)	Categorical	
Pstatus	Parental status	Parent's cohabitation status (t = together, a = apart)	Categorical	
Medu	Mother's education	Mother's education level (1 = no qualifications, 2 = school-level qualifications, 3 = further education, 4 = higher education)	Categorical	
Fedu	Father's education	father's education level (1 = no qualifications, 2 = school-level qualifications, 3 = further education, 4 = higher education)	Categorical	

Mjob	Mother's job	Occupation of the mother	Nominal
Fjob	Father's job	Occupation of the father	Nominal
reason	Reason	Reason for choosing the school	Nominal
guardian	Guardian	Student's guardian	Nominal
traveltime	Travel time	Time taken to travel to school (1 = <15 mins, 2 = 15-30 mins, 3 = 30mins-1 hour, 4 = > 1 hour)	Categorical
studytime	Study time	Weekly study time (1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)	Categorical
failures	Failures	Number of previous assessment failures	Numeric
schoolsup	School support	Extra educational support	Boolean
famsup	Family support	Family educational support	Boolean
paid	Paid support	Student has extra paid for classes	Boolean
activities	Activities	Student engages in extra-curricular activities	Boolean
nursery	Nursery	Attended nursery school	Boolean
higher	Higher	Wants to attend higher education	Boolean
internet	Internet	Has internet access at home	Boolean
romantic	Romantic	Is involved in a romantic relationship	Boolean
famrel	Family relationships	Quality of family relationships (1 = very bad, 5 = very high)	Categorical
freetime	Free time	Free time after school (1 = very low, 5 = very high)	Numeric
goout	Going out	Going out with friends (1 = very low, 5 = very high)	Numeric
Dalc	Weekday alcohol consumption	How much alcohol is consumed on an average weekday (1 = very low, 5 = very high)	Numeric
Walc	Weekend alcohol consumption	How much alcohol is consumed on an average weekend day (1 = very low, 5 = very high)	Numeric
health	Health	Current overall health status (1 = very bad, 5 = very good)	Numeric
absences	Absences	Number of school absences	Numeric
Pass	Pass	Whether the student passed the assessment (1 = yes, 0 = no)	Boolean