

Report #2

GROUP 9

Contents

1	Student Performance Data Set:	1
2	Availability and Source	2
3	Motivation and Initial Goal	2

1 Student Performance Data Set:

The data is about student academic grades for math course of two Portuguese secondary public school during the 2005/2006 school year (Cortez and Silva, 2008). Student grades and their economic, social and demographic variables are included in the data. The data is based on the reports of two public schools (i.e., the grades of students and the number of absence days) and the questionnaire answers by the students. All of the demographic, economic and social variables were collected using the questionnaire. The dataset have 395 records in total with 5 numerical variables and 28 categorical variables. The numerical variables include age, number of school absences, 1st Period grade, 2nd period grade and final period grade. The categorical variables include sex, address(urban/rural), parents highest qualification, family size, commuting time, study time and so on. The table below shows the first 5 records with 10 variables.

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob
1	GP	F	18	U	GT3	A	4	4	at_home	teacher
2	GP	F	17	U	GT3	T	1	1	at_home	other
3	GP	F	15	U	LE3	T	1	1	at_home	other
4	GP	F	15	U	GT3	T	4	2	health	services
5	GP	F	16	U	GT3	T	3	3	other	other
6	GP	M	16	U	LE3	T	4	3	services	other

2 Availability and Source

The data is available at UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets/Student+Performance>). The relevant paper that shows more detail description of data attribute and data analysis was written by Cortez and Silva (2008). The paper can be accessed from <http://www3.dsi.uminho.pt/pcortez/student.pdf>.

3 Motivation and Initial Goal

We are motivated by this dataset since we are students of Math and Statistics school and interested in what features have associations with overall math grade. It would be interesting to explore the performance of students in math course by demographic variables (e.g., does math grade differ by gender?). It could also be interesting to find how economic status of parent of student relates to the grade. Our initial goal of this project is to find the important predictors of student grade out of many demographic, economic and social features.