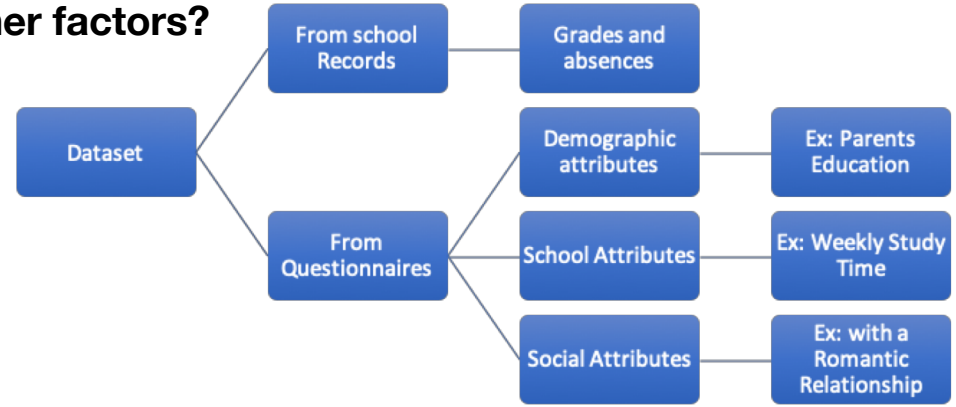
Hong Lyu
Xiang Yao
Yuan Liu

The Blacklist

lyuhong1997@gmail.com
yaoxiang@ucla.edu
liulinda921@ucla.edu

What Influences Academic Success(Final Grades)?

- Source:
  - From 2 secondary schools in Portugal, collected by Cortez and Silva, School records combine with questionnaires
- **395** students in math course, **649** in Portuguese language course, **382** overlap, and **33** variables
- “Students achievement are highly influenced by their past performances. “
- **What about other factors?**

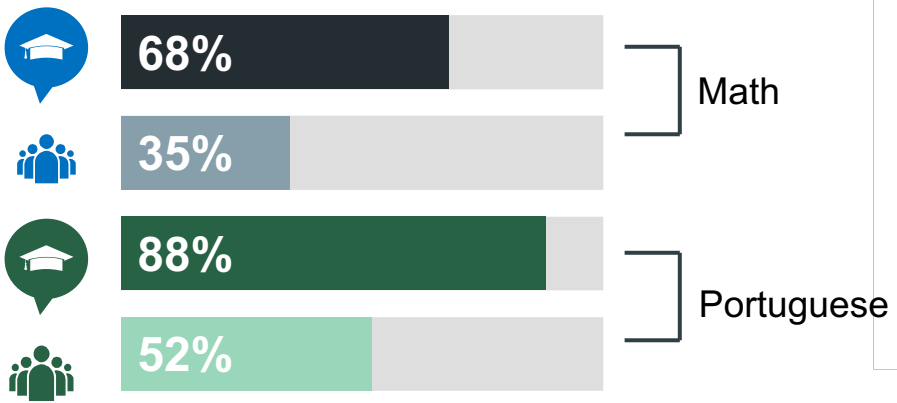


Findings in Exploratory Data Analysis: Part I

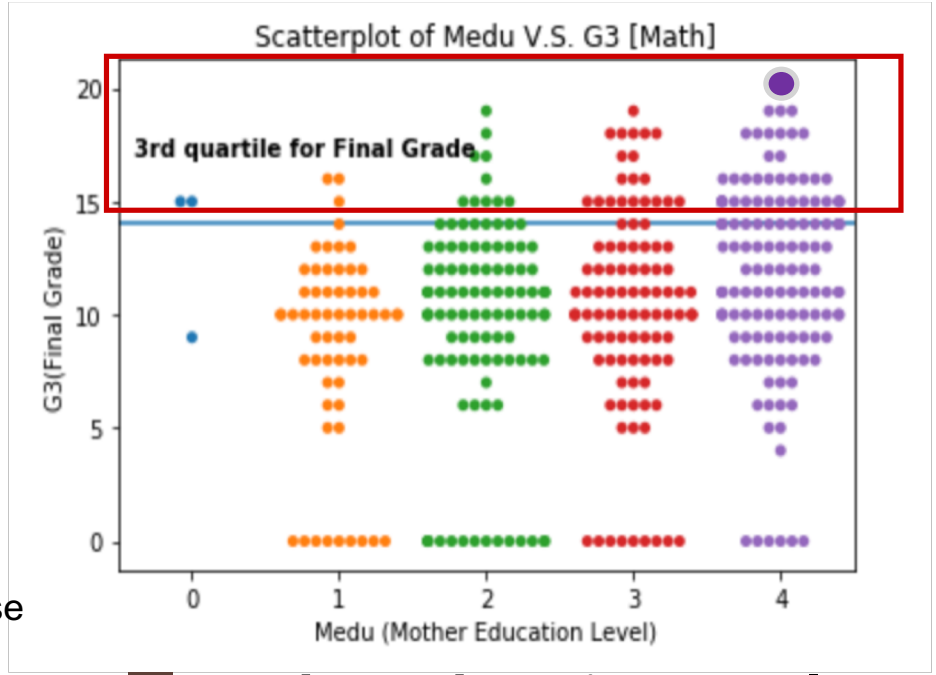
Plan for Higher Education(Y/N)

20% Students who obtained the TOP 20% final grades all plan for higher education.

Proportion of Passes



➔ Mother education level goes up,
- ◆ More students in top 25%.



Mother Education Level(1-5)

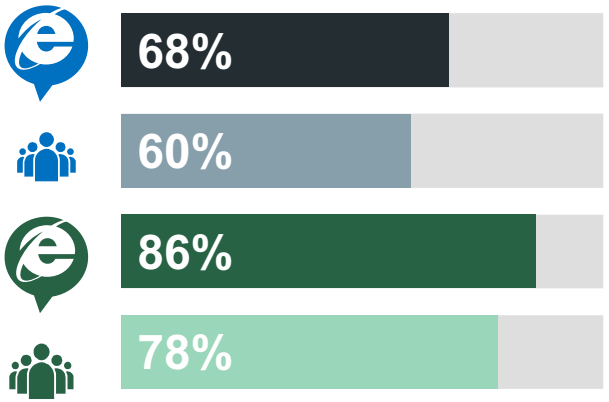
Findings in Exploratory Data Analysis: Part II

Internet Access at Home (Y/N)

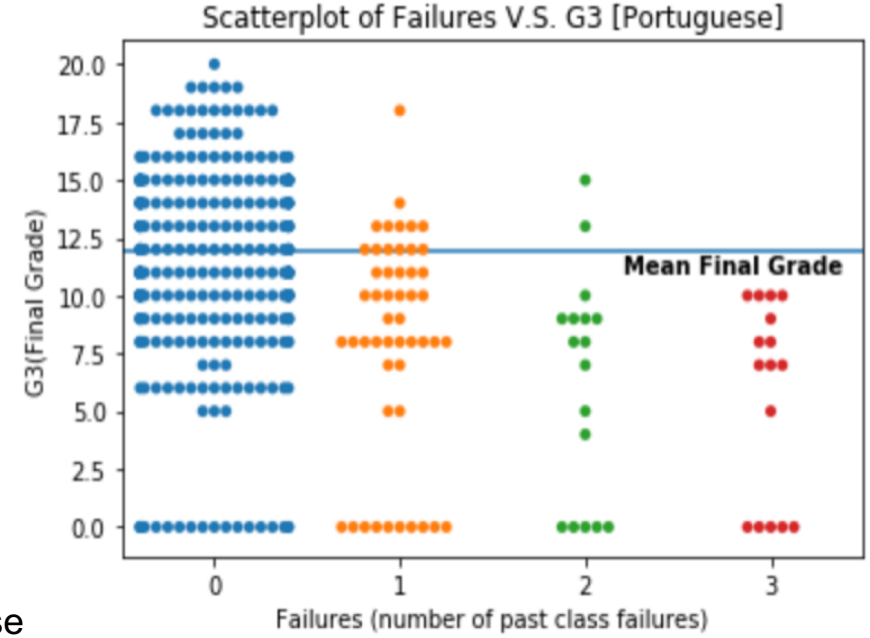
2% Students who obtained the TOP 2% final grades all have internet access.

Mean
10.6
V.S.
9.4

Proportion of Passes



→ Failure >=3 , Final Grade Below Mean.



of Past class failures

Model Tuning and Fitting

Model precision for each level compared with actual proportion

Math	SVM	Tree	RF	*Base Line*
Pass	0.74	0.73	0.69	*0.67*
Fail	0.60	0.57	0.53	*0.33*

Portuguese	SVM	Tree	RF	*Base Line*
Pass	0.88	0.86	0.88	*0.85*
Fail	0.41	0.45	0.50	*0.15*

The tables summarize the **precisions** of six final models for each level (namely Pass and Fail) of Final Grades. All the precisions are **above the baseline** given by proportion, indicating all the classifiers **perform better than blind classifier.**

These two tables are crucial since the two data sets are **imbalanced**. So accuracy alone is **not sufficient** to show performance.

Model Tuning and Fitting

Model accuracy acquired via 5-fold cross validation over whole data

	SVM(Support Vector Machine)	Tree	RF(Random Forest)
Math	0.706 (+/- 0.063)	0.646 (+/- 0.053)	0.699 (+/- 0.076)
Portuguese	**0.809 (+/- 0.155)**	**0.830 (+/- 0.044)**	**0.840 (+/- 0.035)**

The table summarizes the **accuracy** of each model on each dataset based on **a 5-fold cross validation** over the **whole data sets**.

SVM performs the best in Math case, and Random Forest performs the best in Portuguese case.

The three methods all **perform better in the Portuguese case**, since the training data size is larger, allowing the model to learn better.