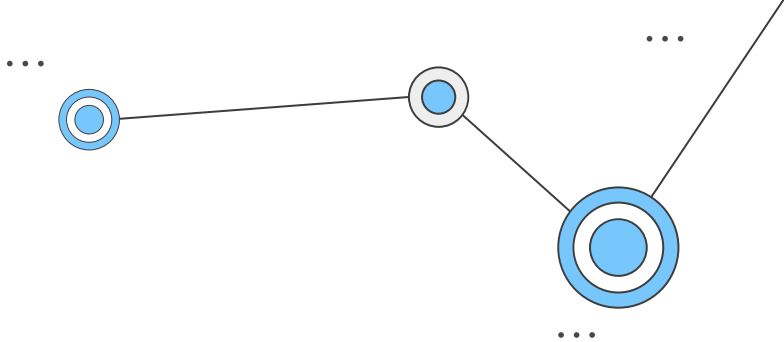


What factors contribute to educational success?

Group 56 - Alyssa Benjamin, Monisha Kapadia, Yi Chen Wu, Anh Nguyen

Motivation

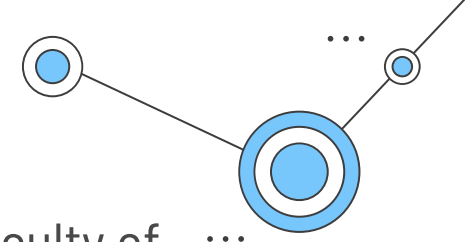
Problem:

- 75% of high school students report that they ALWAYS or OFTEN worry about not getting good grades.
- However, there are numerous factors that can affect the end determination of our grades, some in our control and some out of it.
- In our project, we plan to explore this idea of whether we have full control over the grades we receive and how we can possibly predict our grades based on a series of factors.

Motivating questions:

- Given these factors, how accurately can ML predict a student's grade?
- Which group of categories is most influential on a student's grade? (options: personal, educational, familial)

Dataset Description



- A representative sample of **145 students** in 2019 in the Faculty of Engineering and Educational Sciences at the Near East Campus in Cyprus
- Students were given a questionnaire that consists of **30** questions that provide various info. about each student
- Each student also has a final grade of **0 - 7 (0 being an F, 7 being an A+)**

STUDENTID	AGE	GENDER	HS_TYPE	SCHOLARSHIP	WORK	ACTIVITY	PARTNER	SALARY	TRANSPORT	LIVING	...	PREP_STUDY	PREP_EXAM	N
STUDENT1	2	2	3	3	1	2	2	1	1	1	...	1	1	
STUDENT2	2	2	3	3	1	2	2	1	1	1	...	1	1	
STUDENT3	2	2	2	3	2	2	2	2	4	2	...	1	1	
STUDENT4	1	1	1	3	1	2	1	2	1	2	...	1	2	
STUDENT5	2	2	1	3	2	2	1	3	1	4	...	2	1	

5 rows × 32 columns

Data Features

Educational

high school type

preparation - studied?

preparation for exam

takes notes

listens

likes to discuss

likes flipped classroom

cumulative gpa

expected gpa

weekly study hours

reading frequency

attendance to seminars

attendance to classes

preparation for midterm exams

Personal

age

gender

scholarship

work

activity

Partner

Salary

Transport

Living

regular artistic/sports activity

total salary

transportation to university

accommodation type

Familial

mother's education

father's education

number of brothers/sisters

parental status (married/divorced)

mother's occupation

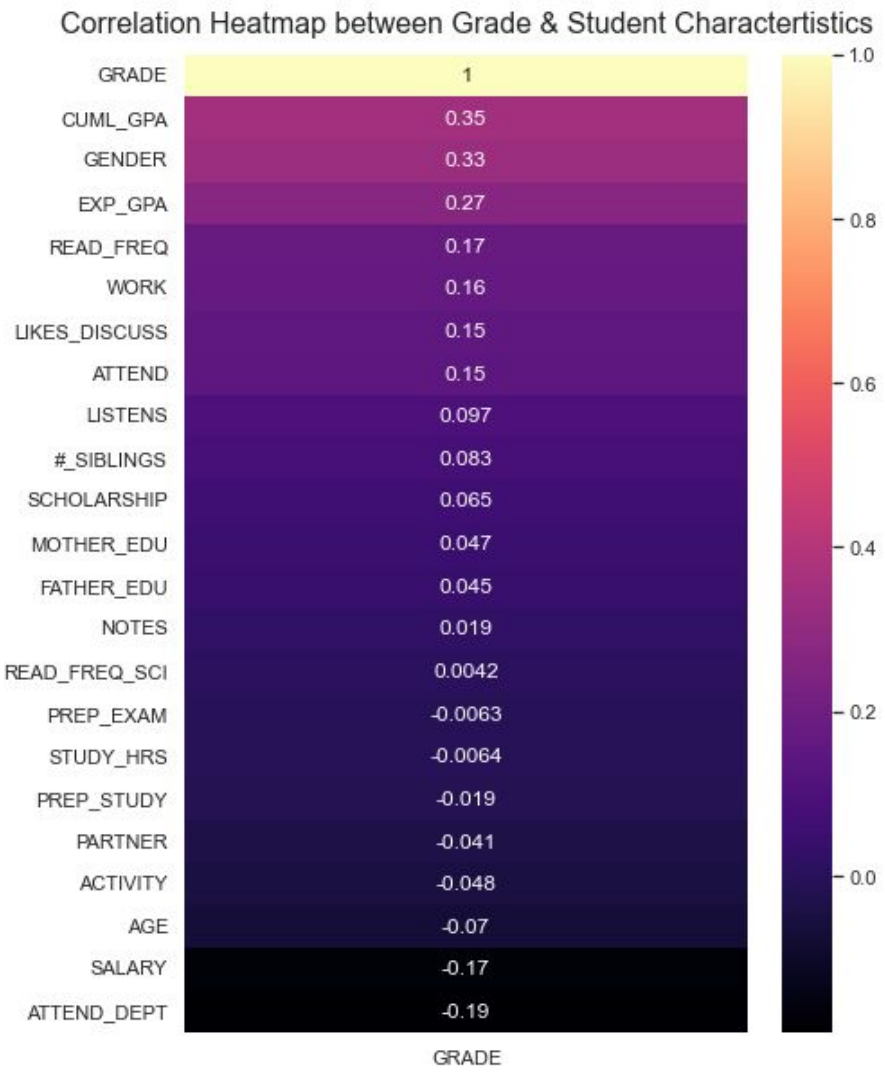
father's occupation

Correlation Heatmap

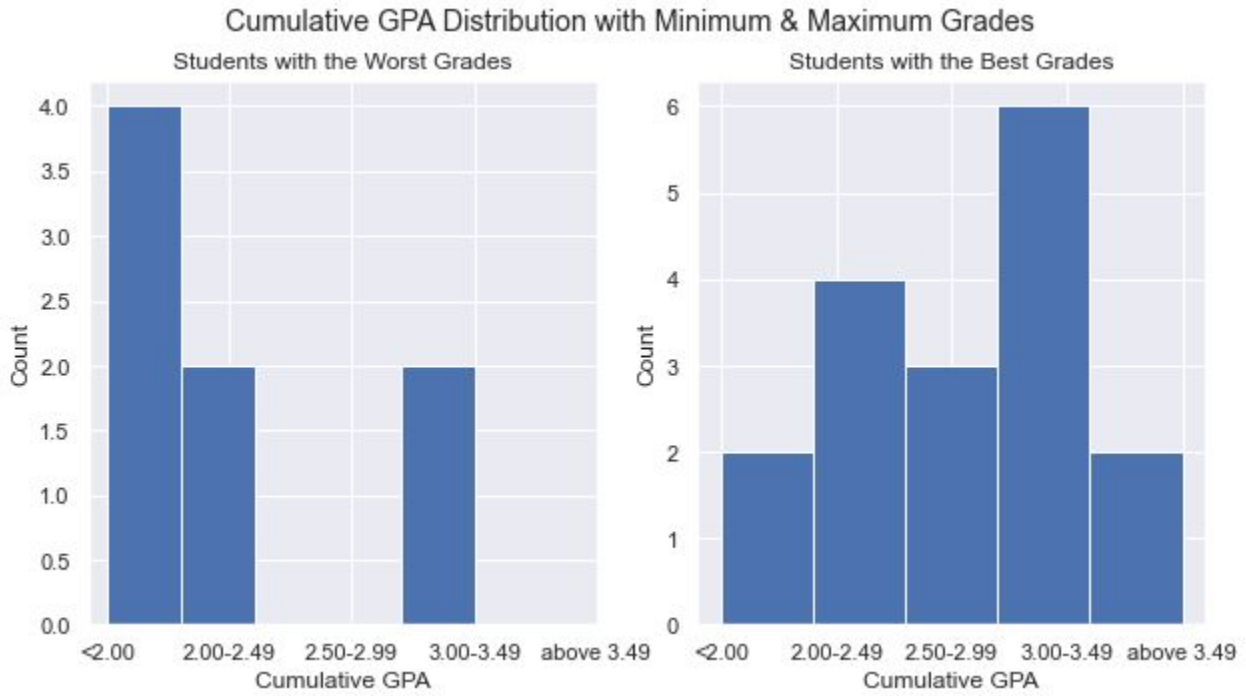
Applied to Student Characteristics with Ordinal or Binary Data

Highest POSITIVE correlation with grades : **Cumulative GPA (0.35)**

Highest NEGATIVE correlation with grades: **Attendance to Department Related Events (-0.19)**



Cumulative GPA & Grades



Machine Learning Methods

k-NN Classifier

Cross Validated & K-Optimized

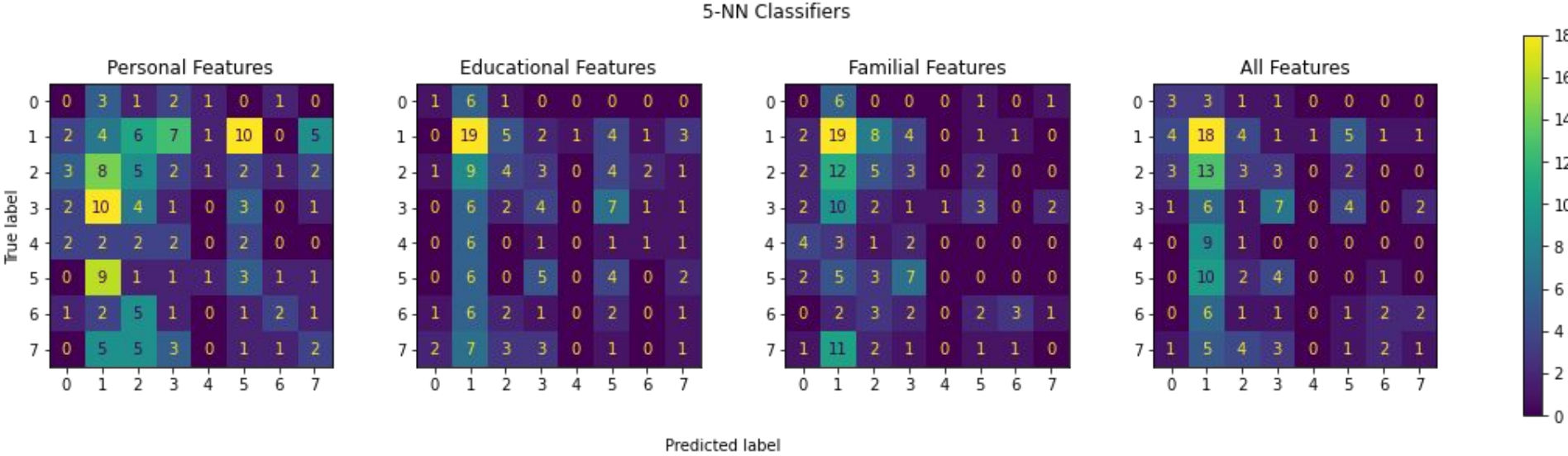
One for each category of features & One Overall

Random Forest Regressor

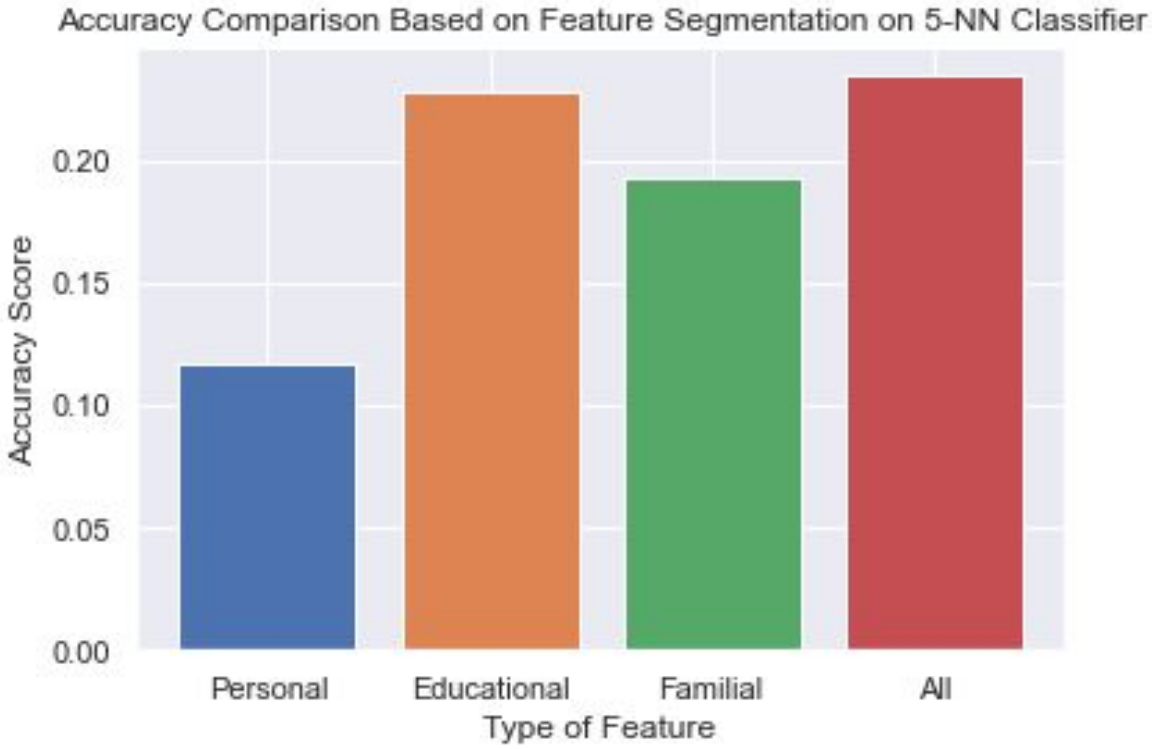
Sample weighted

Fitted on ordinal and binary data

Categorical Confusion Matrices



Comparing Accuracy Scores

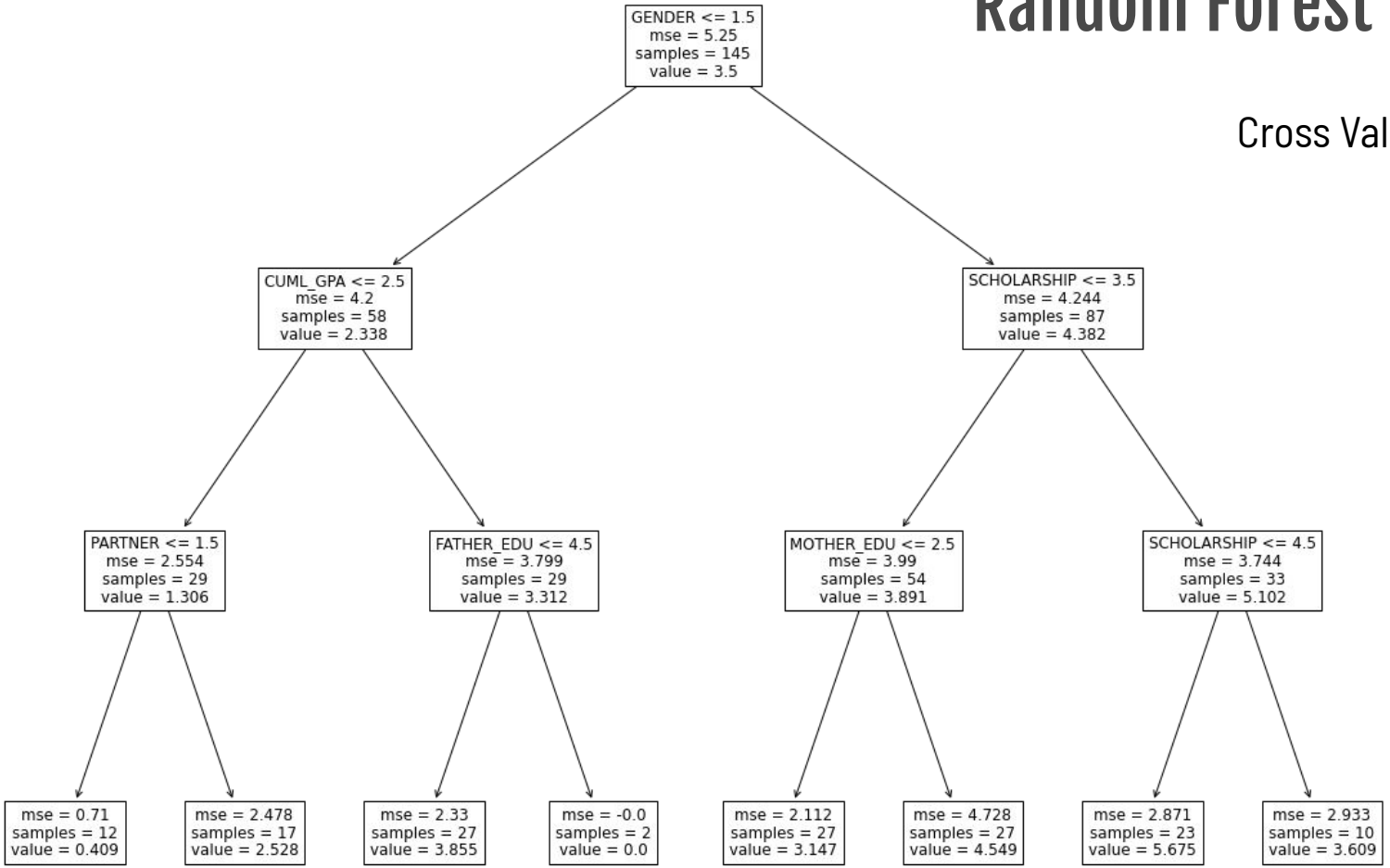


Accuracy Scores for Each Classifier:

- All Features: 0.23
- Personal : 0.11
- Familial: 0.19
- Education : 0.22

Random Forest Regressor

Cross Validated R^2: .135



Interpretation of Results

Random Forest Regressor:

- R^2 very low
  - ~13% of features (variables) helped predict grade
- No feature importance graph

k-NN Classifier:

- Low accuracy score throughout the matrices
- Education & Total Classifier had roughly equal accuracy scores
  - Education - most influential category

Limitations

Random Forest Regressor:

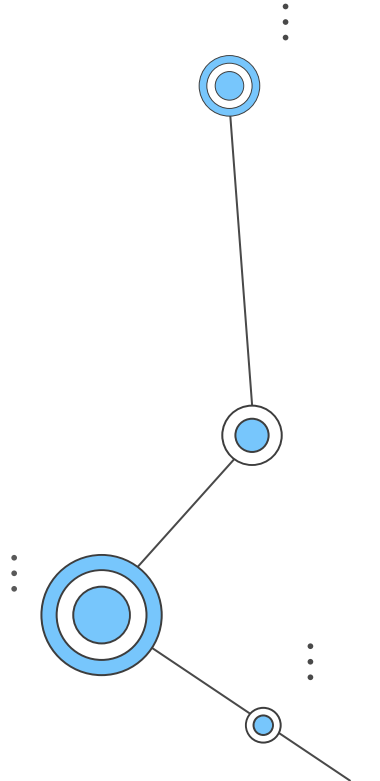
- Regressor would not output all grades
  - Our example tree only outputted grades from 0-5.

k-NN Classifier:

- Could not properly weight samples based on sample size
  - Grade 1 was the most accurately predicted (highest sample size)
- Number of features could have played a role in helping Education have a higher accuracy score
  - More information for the classifier to work with
  - Had the Highest Number of Features

Conclusions

Do not use this model



Thanks for Watching!