Matriculation exams and university student performance

Antti Sieppi

Jyväskylä School of Business and Economics

April 16, 2018

Background

- Who would have been chosen if we had chosen university students solely based on matriculation exams?
- Are entrance exams, matriculation exams or other variables potentially available to universities good signals of student performance?
- What would be the optimal selection criteria?

Background

- Selection process is a basic signaling game
  - Universities want the best students, but have very limited information
  - Applicants can give signals that the university may interpret
- Ideally both school and applicant preferences would be known and we could create a Pareto-efficient allocation.

Background

- Problem 1:
  - Two-sided markets like these may result in applicants employing strategies where they do divulge their preferences (Pathak, 2011)
  - For example, it could be optimal to prepare for the entrance exam of your second favorite university if the threshold of getting into the primary favourite is too high.
  - Removing entrance exams might make students more likely to reveal their preferences (but something else might still encourage not revealing preferences).

background

- Problem 2:
  - While a good mechanism can make students reveal their preferences, it may be hard for schools to determine who are the best students for them.
  - According to prior literature, entrance examinations do not seem to be a very good signal of student performance (for example, Huttunen (2010) and Pietilä (2012))
  - Finding a good predictor for student performance might help with student selection process.

Data

- University of Jyväskylä student transcript, selection and degree databases
  - Data starting from the 1970's
  - Selection data only since year 2000
- Statistics Finland population level microdata

Variables

- 41 explanatory variables
  - Matriculation exam grades, highest scores and points received for the application
  - Personal characteristics (gender, age, previous degree, income and location on previous year etc)
  - Parental characteristics (income, degree, location etc)
  - Previous study rights at University of Jyväskylä, unfulfilled military service

Variables

- Explained variables
  - Graduation within 5, 6 or 7 years from the acceptance to the subject itself
  - Graduation within 5, 6 or 7 years from the first acceptance to University of Jyväskylä
  - Mean grade of university studies
  - Being above the mean grade of other students of the same subject
  - (Credits earned within first 5-6 years?)

Procedure

- 1016 psychology and 1933 JSBE students that were chosen between 2000-2014
- For each year, the applicants who would have been chosen if selection criteria were purely the 2017 background score based on matriculation exams.
- Example: If 80 individuals were chosen to Economics in certain year, choose the 80 best scoring people.

Psychology selection criteria

- Psychology selection criteria: Psychology or best non-language non-math subject + native language + math or other language
  - Psychology or best non-language non-math subject
  - Native language
  - Math or other language
- Maximum 65 (60 without psychology exam)

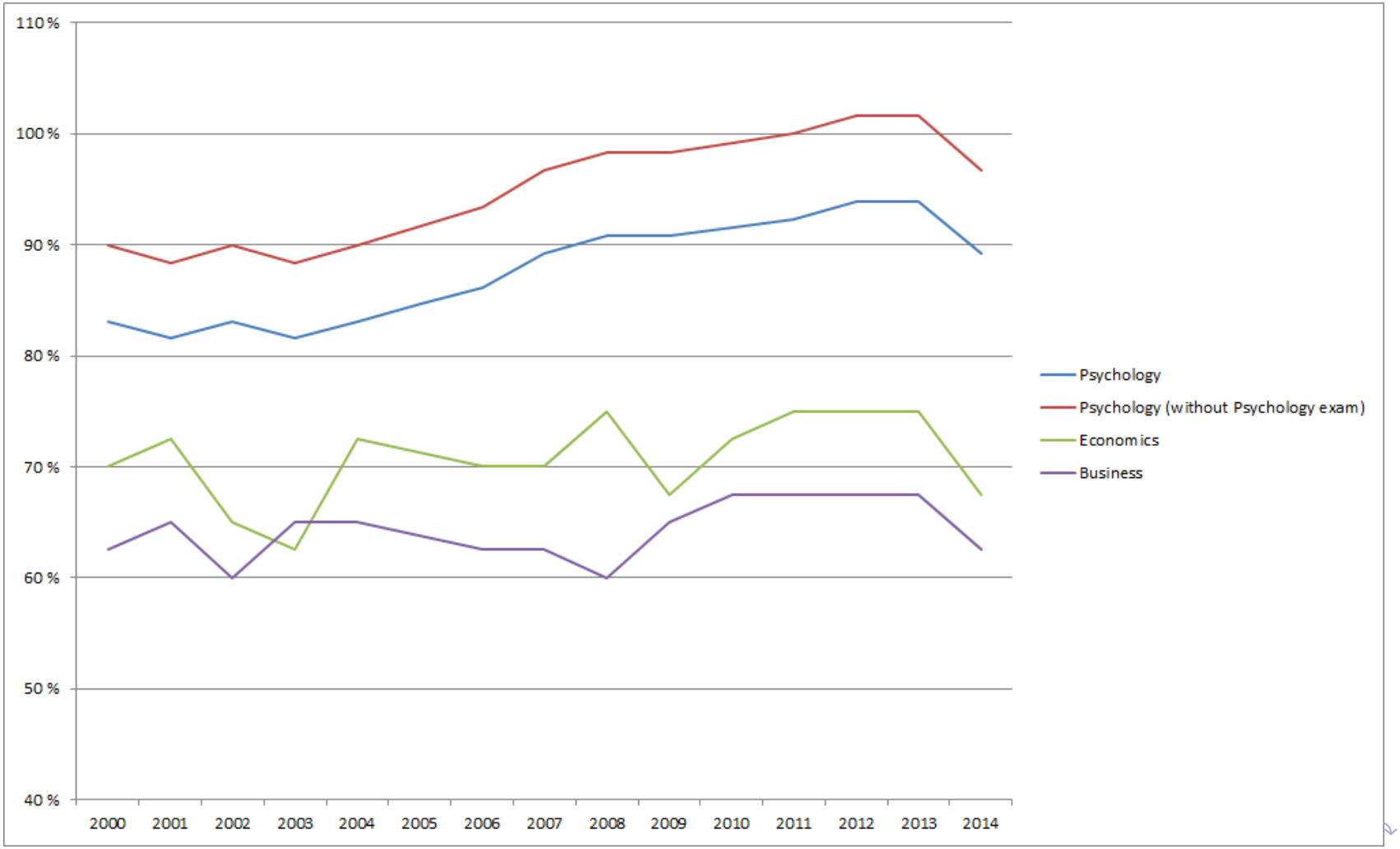
JSBE selection criteria

- For JSBE: Math + native language + three other best examinations, two of which may be non-languages
  - Math
  - Native language
  - Three other best examinations, two of which may be non-languages
- Maximum 40

Selection by matriculation exam

- For JSBE, 42.8% of the students that were actually chosen would have been chosen based on the matriculation exams
  - Economics 45.6%, Business 41.7%
- For psychology, 31.0% of the students that were actually chosen would have been chosen based on the matriculation exams

Percentage of maximum points required to be selected



Grade statistics

	Psychology				Business and economics			
	Current	Estimated	P-value		Current	Estimated	P-value	
Application points	48,44	59,28	0,000	***	20,36	29,42	0,000	***
Matriculation mean	4,36	4,95	0,000	***	3,85	4,62	0,000	***
Finnish	4,59	5,44	0,000	***	4,02	4,80	0,000	***
Swedish	4,49	4,98	0,000	***	3,95	4,75	0,000	***
English	4,16	4,99	0,000	***	3,85	4,65	0,000	***
Math (long), grade	3,34	3,78	0,000	***	3,17	4,07	0,000	***
Math (short), grade	4,17	4,56	0,000	***	4,05	4,89	0,000	***
Other subjects (before 2006)	4,77	5,42	0,000	***	4,12	4,79	0,000	***
Other subject mean	4,86	5,35	0,000	***	4,09	4,80	0,000	***
Psychology (after 2005)	5,16	5,57	0,000	***	4,31	4,97	0,000	***
Took math exam (long)	0,40	0,51	0,000	***	0,51	0,60	0,000	***
Took math exam (short)	0,42	0,37	0,019	*	0,39	0,39	0,837	

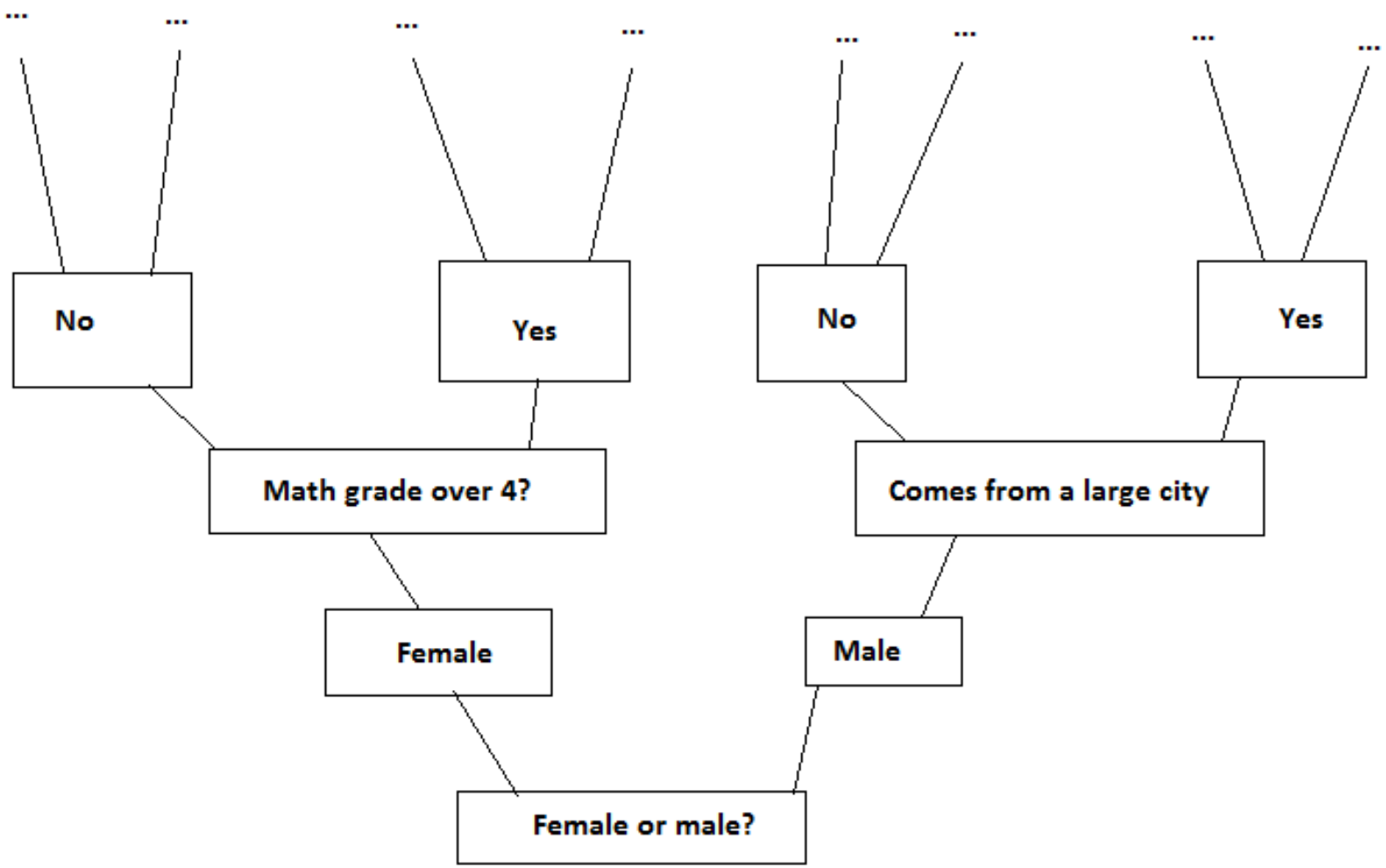
Demographic statistics

	Psychology				Business and economics			
Female	0,91	0,88	0,070		0,53	0,61	0,000	***
Age	22,3	20,7	0,000	***	22,5	21,1	0,000	***
Had a job previous year	0,47	0,31	0,000	***	0,44	0,36	0,000	***
Income previous year	6819	4129	0,000	***	7368	5424	0,000	***
Father's income	28064	26690	0,194		31767	32500	0,625	
Mother's income	20680	21411	0,258		20752	21185	0,417	
Unfinished military service	0,01	0,01	0,344		0,04	0,02	0,039	*
Previous study right	0,15	0,11	0,003	**	0,19	0,18	0,669	
Vocational degree	0,03	0,01	0,001	**	0,04	0,02	0,000	***
From Jyväskylä	0,25	0,18	0,000	***	0,41	0,32	0,000	***
From top 20 city	0,64	0,55	0,000	***	0,69	0,59	0,000	***
From top 10 city	0,53	0,44	0,000	***	0,59	0,50	0,000	***
Prior university degree	0,13	0,08	0,000	***	0,16	0,11	0,000	***
Father's university degree	0,51	0,54	0,270		0,52	0,56	0,010	*
Mother's university degree	0,59	0,61	0,325		0,57	0,64	0,000	***

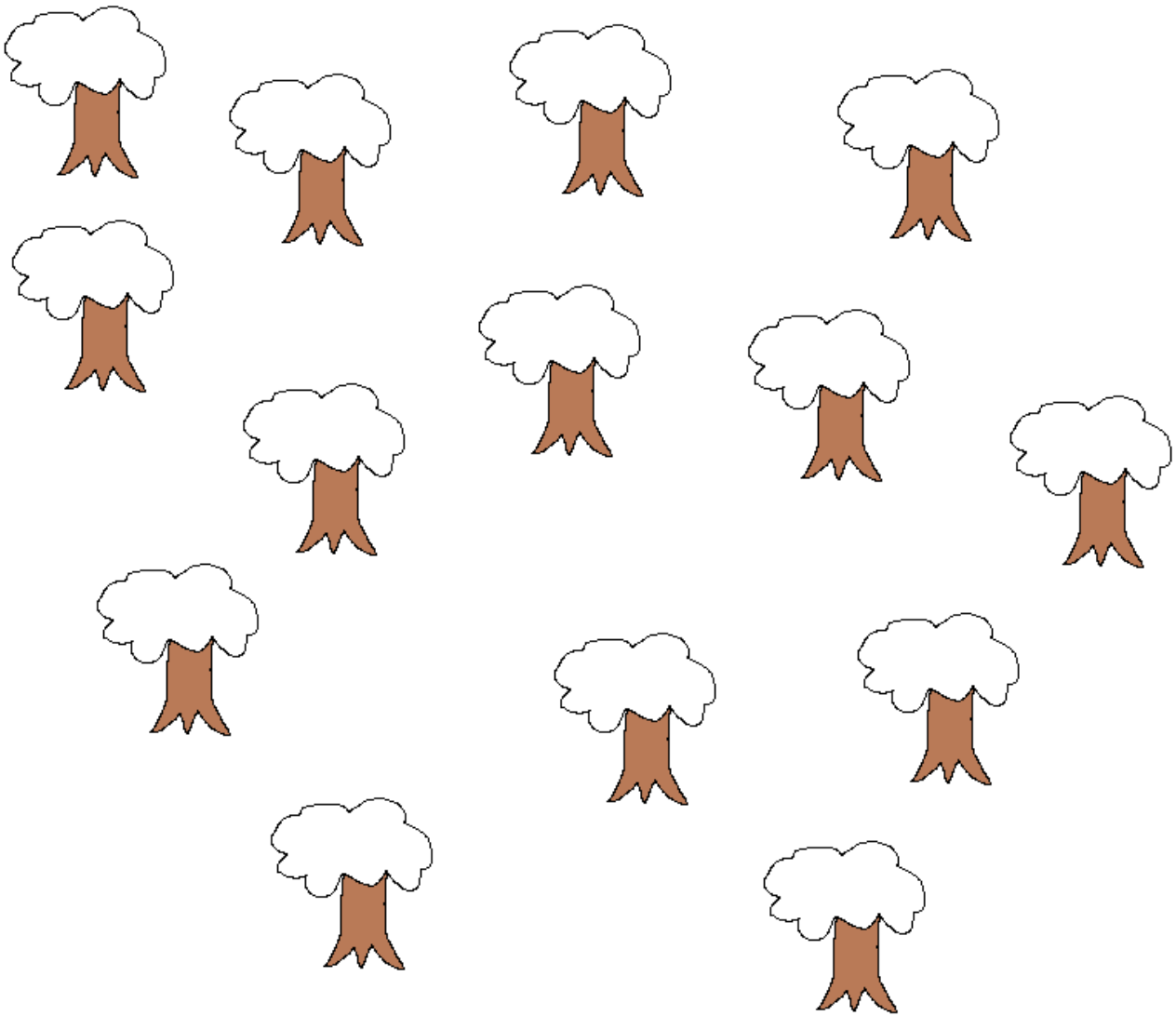
Prediction

- No counterfactual for current system is available: We cannot observe student performance of those who were not accepted in real life.
- Solution: Prediction
- Method: Random Forest

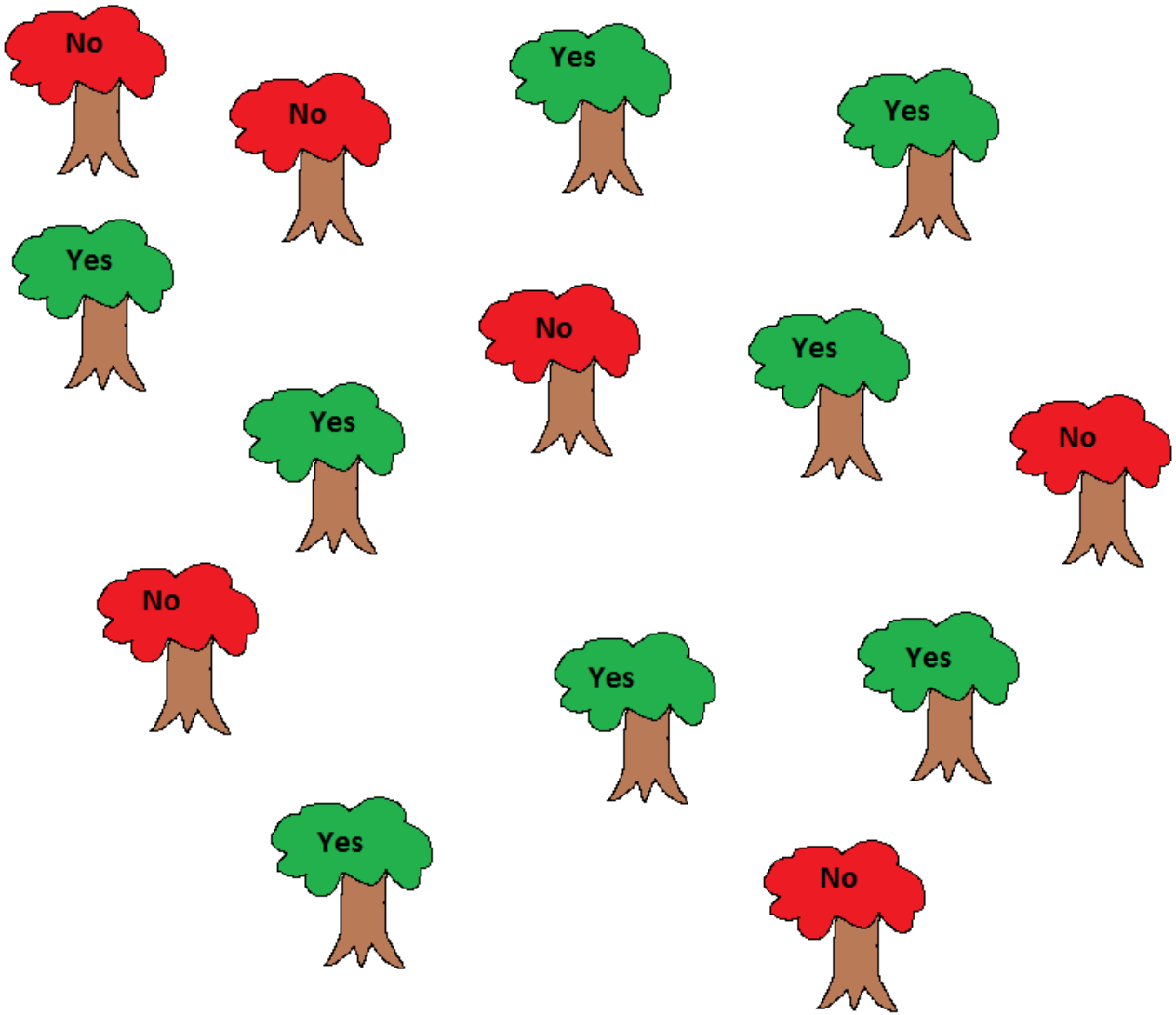
Decision tree



Random Forest



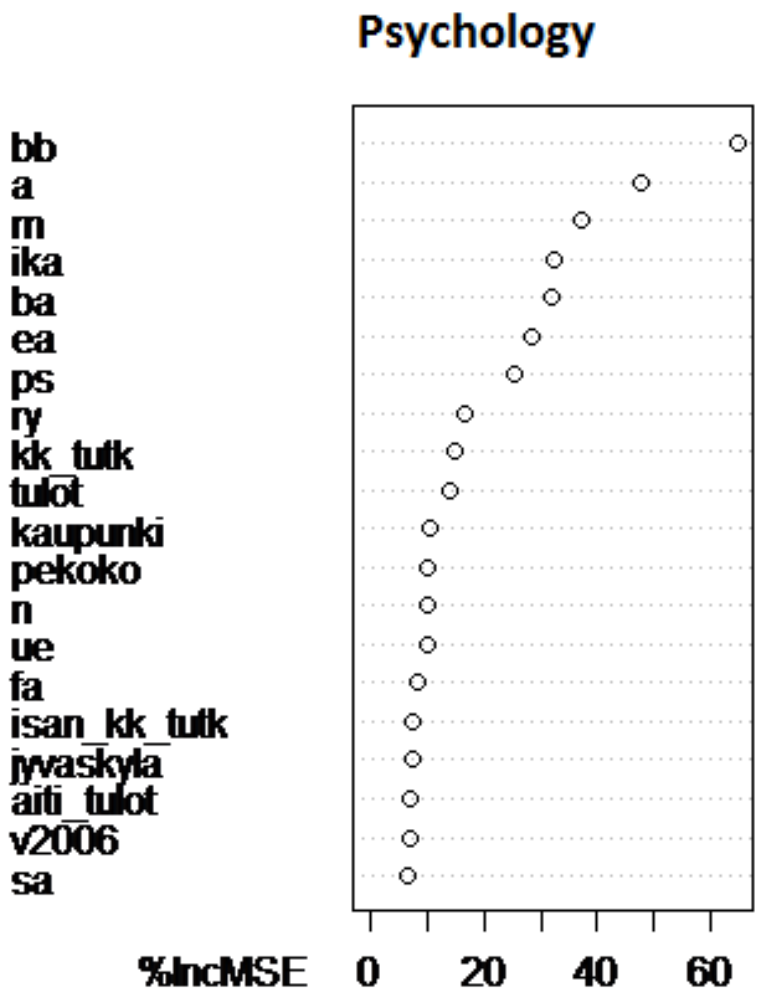
Random Forest

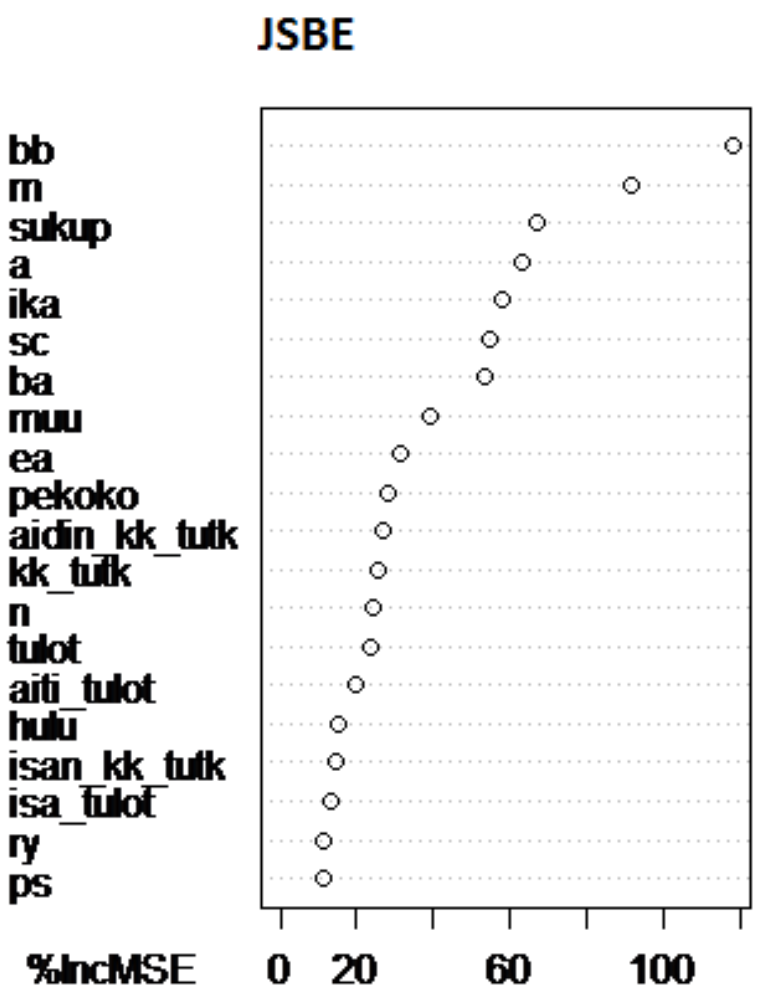


Prediction: Grade

	Psychology	Business and economics
Data	3,70	3,25
Prediction for currently chosen	3,70	3,24
Prediction for matriculation selection	3,86	3,42
Variance explained	13,93 %	22,38 %
N (data)	722	1369
N (prediction)	949	1777

Prediction: Grade





Prediction: Graduation within 6 years

	Psychology	Business and economics
Data	0,71	0,49
Prediction for currently chosen	0,92	0,51
Prediction for matriculation selection	0,81	0,45
OOB Error rate	29,42 %	40,40 %
N (data)	656	1240
N (prediction)	949	1777

Prediction: Graduation within 6 years

Business and economics

	FALSE	TRUE	Class error
FALSE	371	256	40,83 %
TRUE	245	368	39,97 %

Psychology

	FALSE	TRUE	Class error
FALSE	25	166	86,91 %
TRUE	27	438	5,81 %

Spring 2018 selection

- What kind of scoring criteria would maximize mean course grade of students?
- Criteria for business and economics, Spring 2018
  - High score from extensive math and native language
  - Medium score from long language and short math
  - Low score from two best other examinations

	Grade			
	L	**E**	**M**	**C**
3: High	10	9	7	5
2: Medium	8	6	4	2
1: Low	6	4	2	0
0: None	0	0	0	0

Quick and dirty solution

- Use permutations of the Spring 2018 criteria: Allow scoring criteria for different subjects to only be high, medium, low or none
- Limit possible subjects to extensive or short math, English, Finnish, Swedish, two best other non-language subjects and best other language
- Calculate z for $4^8 = 65536$ permutations of scoring criteria one by one and select the permutation with the highest z
- Benefit: Results are extremely easy to comprehend

Results: Optimal selection criteria

Finnish	Swedish	Long math	Short math	English	First other	Second other	Other language	Mean grade
3	2	1	0	0	0	0	1	3,484
3	1	1	0	0	0	0	1	3,484
2	2	1	0	0	0	0	1	3,483
1	2	1	0	0	0	0	1	3,483
3	2	2	0	0	0	0	1	3,483
2	1	1	0	0	0	0	1	3,483
3	3	1	0	0	0	0	1	3,482
2	2	2	0	0	0	0	1	3,482
1	2	2	0	0	0	0	1	3,482
1	1	1	0	0	0	0	1	3,482

Results: Worst "feasible" selection criteria

Finnish	Swedish	Long math	Short math	English	First other	Second other	Other language	Mean grade
0	0	2	2	1	0	0	0	3,320
0	0	3	3	3	3	3	0	3,320
0	0	2	2	3	3	3	0	3,319
0	0	3	3	2	3	3	0	3,318
0	0	2	2	2	3	3	0	3,317
0	0	1	1	2	3	3	0	3,316
0	0	1	1	3	3	3	0	3,316
0	0	1	1	1	3	3	0	3,314
0	0	2	2	1	3	3	0	3,313
0	0	3	3	1	3	3	0	3,312

Results: Demographic differences

	Current	Estimated	Optimal	P-value	
Female	0,53	0,61	0,71	0,000	***
Age	22,5	21,1	21,6	0,000	***
Had a job previous year	0,44	0,36	0,41	0,000	***
Income previous year	7368	5424	6252	0,000	***
Father's income	31767	32500	31802	0,883	
Mother's income	20752	21185	21361	0,570	
Unfinished military service	0,04	0,02	0,02	0,287	
Previous study right	0,19	0,18	0,20	0,421	
Vocational degree	0,04	0,02	0,01	0,784	
From Jyväskylä	0,41	0,32	0,35	0,166	
From top 20 city	0,69	0,59	0,62	0,080	
From top 10 city	0,59	0,50	0,53	0,087	
Prior university degree	0,16	0,11	0,15	0,000	***
Father's university degree	0,52	0,56	0,55	0,280	
Mother's university degree	0,57	0,64	0,63	0,235	

Conclusion

- Predicting student performance is hard
- Predictions tell that students who would've been chosen with matriculation degrees might have studied slightly longer and probably would have gotten better grades
- Especially for psychology, one would need extremely high scores in the relevant subjects to get accepted in
- Demographic changes would seem to be different for both subjects

Conclusion

- Setting matriculation exam scoring that maximizes mean grade for business and economics students results in emphasis on non-English languages and extensive math
  - Largest contrast to current system: English and non-language subjects give no score
- Predicted increase on mean grade provided by optimization is fairly small
- Student performance being so hard to predict implies that it might be optimal for universities to have only very loose preferences on who to choose.