



UNIVERSITEIT VAN AMSTERDAM
Economics & Business

Data Challenge BAN



Introduction

Fred Pope

Developmental Psychology

Student Advisor > Student Advice System STAP

Web/Application Developer

Databases

Data Analysis

Data Based Management Decisions

Student & Learning Analytics

Machine Learning



BSA: Binding Study Advice

Many norms over the years:

30, 35, 42, 48 > Covid > 42

Percentages do not vary much at different norms

Due to Covid able to gain insights on BSA

A very good selection tool

How can we help students by creating a early warning system?

If we can signal at an early stage, many students could profit by adjusting Learning Methods



BSA Data Sets

The data sets are real-world

Studentnumbers, Programs and courses have been anonymised

BSA norm is 42 credits

Dispensations, Exemptions and Returning after 3 years have been removed

The training sets are comprised of “true” Negative or Positive BSA’s

A studentnumber can appear twice if registered for more than one program



Data Sets

(will run thru the data with you in a moment)

The Training set has four programs for two academic years each (21-22 & 22-23)

Each excel Tab has a program

The Prediction set has the same four programs for the running academic year (23-24)

B1, B2 & Resit results can be used to predict BSA, based on the trained model



Deliverables

- BSA prediction based on Block 1 & 2 + (Resits) data ++ extra personal data

Explore earlier predictors, such as only B1, or specific courses

Drill down into the data: which factors have the greatest influence?:
T-test, ANOVA, Factor Analysis PCA

Consider:

- Predicting BSA or acquired Credits? >> Different ML Models
- What happens to the model if a program changes its BSA credit-norm?
- What happens to the model if a new course is added to a program?
- Model Evaluation (Confusion Matrix, MSE, MAPE, R2)
- Uncertainty Quantification
(40-45?) individual evaluation
- Explainability
Dig into the prediction, what is the cause of the ML result?



Deliverables Product

- A dashboard that provides an early warning signal for students and advisors so that students can profit by adjusting Learning Methods
- Python code for the model
- JASP implementation would be appreciated



Course Codes

code-1 is the first exam grade of a course

code-R is the resit grade of a course

code is the last registered (final) grade of a course



Excel Files

Lets go thru the two excel files:

- Train
- Predict



Supervision

In general: follow and contact your supervisor

You may contact me if there are issues/questions about the data:

- there is an error due to the dataset
- you would like additional data (*maybe*)
- you need clarification on data definitions / meaning etc
- the BSA procedure and/or rules
- How specific data was collected

Vacation thru April 25th, but if there is a major issue you can contact me 😊



Considerations

Ethical questions:

There are always false negatives.

Is it ethical to warn a student at an early stage if perhaps more is going on ?

Is it possible that warning a student can worsen a situation?



Privacy

Even anonymised data is privacy sensitive

Be professional

Treat the data with care for privacy!