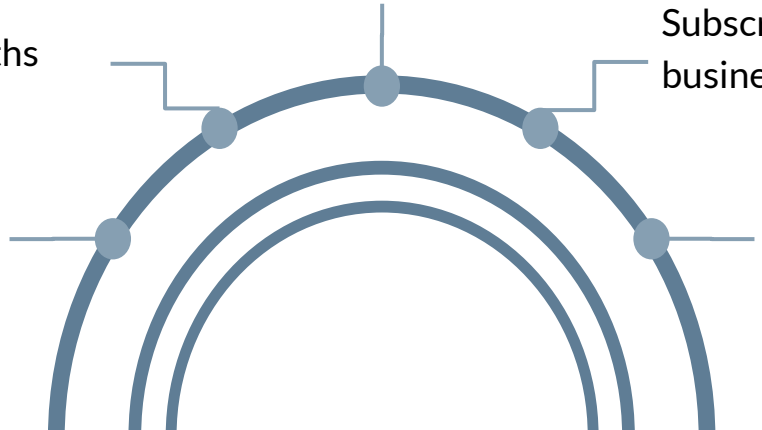
ML MODEL FOR PREDICTING EXAM GRADES

Final project by Konstantin Leube

TABLE OF CONTENTS

INTRODUCTION

Build a ML-model to predict the exam scores of students



Specific improvement recommendations

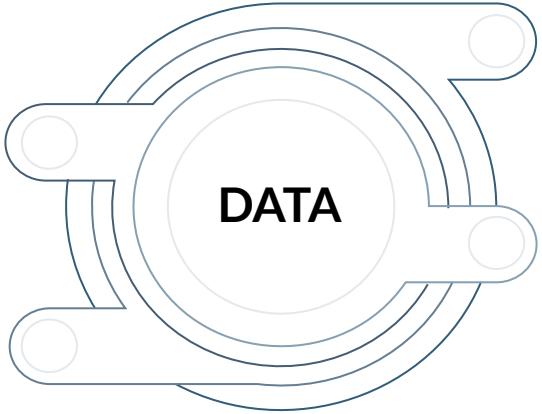
Subscription based business model

Aid for students to help them achieve the best possible grades during their studies

Study platform

Personalized learning paths

INTRODUCTION



Information on student demographics and their surrounding

>6000 rows

19 feature and 1 target column

Categorical and numerical columns

DATA TRANSFOR MATION

DATA TRANSFORMATION

OVERVIEW

Checking for NaN-values and for categorical and numerical columns

CLEANING

Minority of NaN-values

Able to drop them

WRANGLING

Transformation of categorical columns to numeric columns

HYPOTHESIS TESTS

HYPOTHESIS TESTS

TESTING

Ran **Hypothesis test** on **transformed columns**

T-test for binomial columns
Anova for the rest

EVALUATION

Out of the 13 transformed columns **11 proved to be statistically significant**

2 **non-significant** ones were **dropped**

ML-APPLICATION

COMPARISON BEFORE

	Lin. Reg.	Rndm. For.	Grad. Boost.
R2	0.69	0.60	0.36
MAE	0.70	1.21	1.61
RMSE	2.21	2.49	3.13

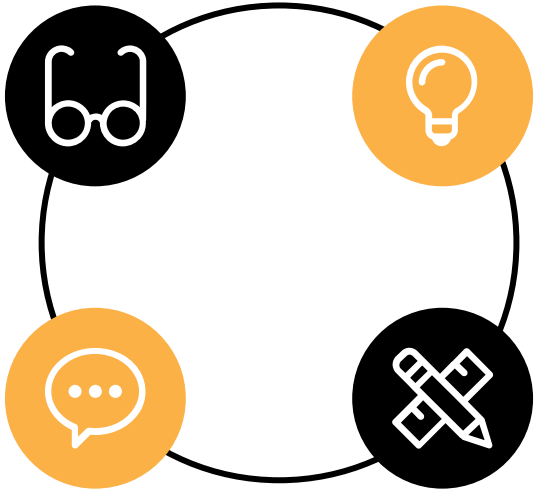
PROBLEMS I FACED

CORRELATION

Indicated that lin. Reg. might not be optimal

HYPERPARAMETER AND CROSS VALIDATION

Helped improve model performance but still did not outperform lin. reg.



SKEW IN DATA

Data normally distributed with a slight right-skew due to outliers

POWER TRANSFORMER

Wasn't able to outperform previous R2 score

COMPARISON AFTER

	Lin. Reg.	Rndm. For.	Grad. Boost.
R2	0.69	0.64	0.68
MAE	0.70	1.07	0.65
RMSE	2.21	2.36	2.20

CONCLUSIONS

PREDICTION

Able to confidently predict student grades in approx. 70% of cases

DATA

In order to increase R2-value think about additional data columns

MODEL

Lin. Reg. proved to be the best performing model
Hyperparameter tuning improved others as well

APPLICATION TO THE BUSINESS MODEL

CATEGORIZATION

Distinction between “At Risk” and “Not At Risk” students

Usage of ML-model to categorize students into categories

Able to provide tailored to help to each student in every situation early on

Still a regression problem not a classification problem in order to assess gravity of situation

THANKS!

I’m happy to answer any questions