1. 

**Contents**

[**1.**](#_heading=h.1fob9te) **SUMMARY 6**

[**2.**](#_heading=h.3znysh7) **INTRODUCTION 7**

[2.1 Outline of the Project 7](#_heading=h.2et92p0)

[2.2 Target and Expected Outputs 7](#_heading=h.tyjcwt)

[2.3 Methods 7](#_heading=h.3dy6vkm)

[**3.**](#_heading=h.4d34og8) **PROJECT ACTIVITIES 8**

[3.1 Gathering data 8](#_heading=h.2s8eyo1)

[3.2 Data pre-processing 8](#_heading=h.17dp8vu)

[3.3 Researching the model that will be best for the type of data 8](#_heading=h.3rdcrjn)

[3.4 Training and testing the model 8](#_heading=h.26in1rg)

[3.5 Evaluation 8](#_heading=h.lnxbz9)

[**4.**](#_heading=h.35nkun2) **RESULTS and DISCUSSION 9**

[4.1](#_heading=h.1ksv4uv) Expected vs. Achieved Results 9

[4.2](#_heading=h.44sinio) Overall Project Results 9

[4.3](#_heading=h.2jxsxqh) Variances in Planned versus Actual Results 9

[4.4](#_heading=h.z337ya) A list of factors that influenced the project results 9

[***5.***](#_heading=h.3j2qqm3) **Lessons Learned and Recommendations 10**

[**6.**](#_heading=h.1y810tw) **Future Plans 11**

[**7.**](#_heading=h.4i7ojhp) **References 12**

[**8.**](#_heading=h.3whwml4) **Appendix 13**

**ABBREVIATIONS**

1. DUO
2. DEEPLO AI
3. **List of Appendix**

**Acknowledgement**

# SUMMARY

**ALI DURSUN**

In this project, an artificial intelligence application will be developed for the education sector. The project outcomes will guide school administrators and their clients. Firstly, Dutch primary school achievements will be studied. At the end of the project, it is aimed to be transformed into an application that can be applied to all schools.

Measuring school achievements correctly is very important for both education service providers and customers. It is directly related to the school's ability to increase funding resources. Prediction models will be developed with the Machine Learning approach so that schools can understand what to do to increase their success. With Machine Learning linear regression, it will be revealed which factor and how much to success.

Train and test inputs will be obtained by using open-source data of the Dutch Ministry of Education and other open sources. Features will be related not only to numerical values but also to social life. What can be done to increase the success of a school and how much success will increase as a result. This question will be answered. Managers can predict the cost of increasing success with this project. Also, the unnecessary investment will be prevented for variables that have no direct impact on success. This project will be carried out within Deeplo AI and will be carried out by a total of 5 trainees.

*.*

# INTRODUCTION

**BEYZA**

Machine learning is one of the most effective AI applications. How various processes in the world affect each other reveals with very complex mathematical models. Manual calculations that cannot be done can be modeled with machine learning. In many areas from health to automotive, problems are solved with the help of machine learning. This can sometimes be an unresolved problem to increase efficiency. The adequate practice has not yet been developed in the field of education. This project will fill this gap.

## Outline of the Project

## Target and Expected Outputs

There is a success-oriented approach to the education system in every country. To measure this, besides the course success, the schools where the student is placed after graduation are also considered. This project will be based on these success criteria. First, let's summarize why the project need arises.

1. Students and their families prefer more successful schools.

2. Increasing the success of schools, in general, will prevent accumulation.

3. Factors affecting success are required for the correct use of identified resources.

4. To increase competitiveness and contribute to the country, each school should increase their success.

1. A machine learning application, which can predict the success of the school, is aimed to be implemented primarily for these needs. Historical datasets are needed to train the machine learning model. Weights of features should be determined in model determination and training

To create and make at least two datasets available.

• To develop 90% performance prediction algorithm.

## Methods ORGANIZE EDIYORUM HALA BITMEDI

# MATERIALS AND METHODS

# Student Data

# In the Netherlands, primary education consists of 8 years, preceding three different types of secondary school (*middelbare school*) at age 12. These are VMBO (preparatory secondary vocational education), HAVO (senior general secondary education) and VWO (university preparatory education). VMBO is a four year vocationally-oriented education that aims to continue with MBO (vocational training). HAVO is a five year education prepares students for HBO (university of applied sciences-hogescholen). VWO (aka atheneum & gymnasium) is a six year long theoretical education prepares students for WO (research university). During

uyarla

# highly diversified lower secondary education

# two out of six streams lead to higher education

# four out of six streams lead to upper vocational education

# several routes through system

* <https://eacea.ec.europa.eu/national-policies/eurydice/content/netherlands_en>

# Two third of education system is private and held by religious groups. <https://link.springer.com/chapter/10.1007/978-94-007-0926-3_3>

# Most of the students join private schools.

http://www.cesifo-group.de/DocDL/dicereport211-rm1.pdf

# Government equally funds private and public education. Families required to pay some amount and for international schools education is required to be paid by parents. Although there are weekly and monthly exams, the flow of education for each child is to be determined by the exams done at the end of 8th grade. Like several other countries students have a test at the end of the 8th grade (e.g. …….. ), …….. point grading scale is used, where 0 is the lowest grade and ... is the perfect score. During the 8th grade, students are evaluated with exams ……., ……., ……... (G3 of Table 1) corresponds to the final grade.

# This study will consider data collected during the 2018-2019 school year from

# ….. (?)public schools, from all the regions of The Netherland. Although afther the year 2008 (?) Dutch government began sharing huge amount of data regarding many fields including education, it took a lot of effort and time to search, find and bring together the data used in the study. Workers from data overheid, DUO and CBS are being contacted in person to find detailed and relevant data. . …. and …. methods are used to bring the data together. (kisaltmalari girise ekle) . While schoolbased data were outside on the web, it required a lot of time and enery to derive the related data from these sources. Final version of dataset build from DUO … and ….. and ….. data.

# We have included ……., …….., and ……… exams and other school related variables that thought to expect the success of student. (?).

# In the preprocessing stage some variables are discarded hence they did not discriminate anything in the dependent variable. For instance, there was lack of information in the variable …, and there were variable … and …. same for almost all of the students. Remaining features are shown below in Table 1.

# *Data Mining Models uyarla ya da kaldir*

# *Classification and regression are two important DM goals. Both require a supervised learning, where a model is adjusted to a dataset made up of k 2 {1, ...,N} examples,*

# *each mapping an input vector (xk 1 , . . . , xk I ) to a given target yk. The main difference is set in terms of the output representation, (i.e. discrete for classification*

# *and continuous for regression). In classification, models are often evaluated using the Percentage of Correct Classifications (PCC), while in regression the Root*

# *Table 1: The preprocessed student related variables*

*Computational Environment*

-----------------

The data derived from DUO website. Data collected depending on their predictability on the sucess rate. Data has been trained and classified. The classification (6 classes?) has been done depending on the features …, … and … . We have done decision tree and random forest. Last version of decision tree suggested that… This requires more of interpretation whihch will be in conclusion. To .. and .. we applied backward elimination and other models such as… and … being used. The results was ……………..

When linear regression applied the result was …………………..

Polynomial regression on the other hand showed that ……………...

From the computations done on the related data below

According to Lasso regression…………….

According to support vector machine ………………….

To … and … we applied feature selection Feature Selection

When correlation matrix has been done we

Information gathered from the results suggested that ……… That is why principal component analysis has been used.

We standardized our data using max min …….. When we try data here it has been found that the results are significantly .. to the previous ones. That is why we quit this method.

Feature engineering, when we look at area specific and education level……..

# PROJECT ACTIVITIES

1. OSMAN

## Gathering data

## Data pre-processing

## Researching the model that will be best for the type of data

## Training and testing the model

## Evaluation

# RESULTS and DISCUSSION

## Expected vs. Achieved Results

## Overall Project Results

## Variances in Planned versus Actual Results

## A list of factors that influenced the project results

1. **ALI TEVFIK**

# Lessons Learned and Recommendations on working

* 1. We found out that in a classification problem like this, where one class (supernova images) is well defined, but the other class (non-supernova images) is defined only as not being the well-defined class, simpler/linear classifiers like kNN, logistic regression, etc. perform poorly. It is only by using a classifier that utilizes the features of the defined class that satisfactory performance can be achieved.
  2. • It’s better to use a distributed algorithm for neural nets, such as caffe, than to run all the code on one machine, such as scikit-learn’s multilayer perceptron.
  3. • Standardization is important. Before standardization, neural nets, logistic regression, and linear svm all performed very poorly.
  4. to search and gather data: it is important to find datas that are powerful and useful.

# Future Plans

1. ***RASIN***
2. .

# References

1. ALI DURSUN

# 

# Appendix