KTH ROYAL INSTITUTE
OF TECHNOLOGY

DEGREE PROJECT IN TECHNOLOGY AND LEARNING

SECOND CYCLE, 30 CREDITS

Early Warning System of students failing a course

A binary classification modelling approach at upper secondary school level

NIKLAS KARLSSON & ALBIN LUNDELL

Stockholm, Sweden 2022

TRITA-ITM-EX 2022:242

Early Warning System of students failing a course:

A binary classification modelling approach at upper secondary school level

NIKLAS KARLSSON & ALBIN LUNDELL

Master of science in engineering and education
Date: June 5, 2022
Supervisor: Eva Hartell, KTH Royal Institute of Technology
Supervisor: Olga Viberg, KTH Royal Institute of Technology
Examiner: Cecilia Kozma, KTH Royal Institute of Technology
School of Industrial Engineering and Management
Host organization: K-ULF
Swedish title: Förebyggande Varningssystem av elever med icke godkänt betyg: Genom applicering av binär klassificeringsmodell inom gymnasieskolan

Abstract

Only 70% of the Swedish students graduate from upper secondary school within the given time frame. Earlier research has shown that unfinished degrees disadvantage the individual student, policymakers and society. A first step for preventing dropouts is to indicate students about to fail courses. Thus the purpose is to identify tendencies whether a student will pass or not pass a course. In addition, the thesis accounts for the development of an Early Warning System to be applied to signal which students need additional support from a professional teacher.

The used algorithm Random Forest functioned as a binary classification model of a failed grade against a passing grade. Data in the study are in samples of approximately 700 students from an upper secondary school within the Stockholm municipality. The chosen method originates from a Design Science Research Methodology that allows the stakeholders to be involved in the process.

The results showed that the most dominant indicators for classifying correct were Absence, Previous grades and Mathematics diagnosis. Furthermore, were variables from the Learning Management System predominant indicators when the system also was utilised by teachers. The prediction accuracy of the algorithm indicates a positive tendency for classifying correctly. On the other hand, the small number of data points imply doubt if a Early Warning System can be applied in its current state. Thus, one conclusion is in further studies, it is necessary to increase the number of data points. Suggestions to address the problem are mentioned in the Discussion. Moreover, the results are analysed together with a review of the potential Early Warning System from a didactic perspective. Furthermore, the ethical aspects of the thesis are discussed thoroughly.

Keywords: Machine learning, Random Forest, Early Warning System, Drop out, Algorithm, Binary classification model, Upper secondary school

Sammanfattning

Endast 70% av svenska gymnasieelever tar examen inom den givna tidsramen. Tidigare forskning har visat att en oavslutad gymnasieutbildning missgynnar både eleven och samhället i stort. Ett första steg mot att förebygga att elever avviker från gymnasiet är att indikera vilka studenter som är på väg mot ett underkänt betyg i kurser. Därmed är syftet med rapporten att identifiera vilka trender som bäst indikerar att en elev kommer klara en kurs eller inte. Dessutom redogör rapporten för utvecklandet av ett förebyggande varningssystem som kan appliceras för att signalera vilka studenter som behöver ytterligare stöd från läraren och skolan.

Algoritmen som användes var Random Forest och fungerar som en binär klassificeringsmodell av ett underkänt betyg mot ett godkänt. Den data som använts i studien är datapunkter för ungefär 700 elever från en gymnasieskola i Stockholmsområdet. Den valda metoden utgår från en Design Science Research metodik vilket möjliggör för intressenter att vara involverade i processen.

Resultaten visade att de viktigaste variablerna var frånvaro, tidigare betyg och resultat från Stockholmsprovet (kommunal matematikdiagnos). Vidare var variabler från lärplattformen en viktig indikator ifall lärplattformen användes av läraren. Algoritmens noggrannhet indikerade en positiv trend för att klassificeringen gjordes korrekt. Å andra sidan är det tveksamt ifall det förebyggande systemet kan användas i sitt nuvarande tillstånd då mängden data som användes för att träna algoritmen var liten. Därav är en slutsats att det är nödvändigt för vidare studier att öka mängden datapunkter som används. I Diskussionen nämns förslag på hur problemet ska åtgärdas. Dessutom analyseras resultaten tillsammans med en utvärdering av systemet från ett didaktiskt perspektiv. Vidare diskuteras rapportens etiska aspekter genomgående.

Nyckelord: Maskininlärning, Random Forest, Förebyggande Varningssystem, Algoritm, Binär klassificeringsmodell, Gymnasieskola

Acknowledgments

Throughout the work of the master thesis, we have received a large magnitude of feedback and support.

First of we would like to thank our supervisors, university lecturers Eva Hartell and Olga Viberg whose expertise was invaluable in formulating purpose, research questions and conducting the method. The feedback taught us to write a more scientific thesis.

We would also like to acknowledge our examiner Cecilia Kozma, project manager for the K-ULF project. She provided insights into how we would include both teaching and engineering in the thesis. Moreover was she responsible for hosting the regular K-ULF meetings. The meetings provided possibilities for us to discuss the thesis with both teachers and other thesis writers.

Furthermore, would we like to thank the upper secondary school for providing data making this thesis possible. Moreover, would we like to acknowledge teachers at the school who were also members of K-ULF. Their drive for Learning Analytics to succeed and being an enthusiast resulted in rubbing off on us in the motivation of writing this thesis.

Lastly, we would like to acknowledge our dear friends Leonardo, Raphael, Donatello and Michelangelo. When reading through the thesis they discovered mistakes we did not notice ourselves.

Glossary

Accuracy, Sensitivity, Specificity = Measurements for how well an algorithm performs

Classification = The task the algorithm performs while deciding if a student will pass or fail.

Database = An organized collection of data stored and available on a server

Data pipeline = Data transport between different systems, servers or software

Data point = One unique point of data

Data sample = One row of data points with a theme. In this thesis the correlated theme is the unique student

Data set = A collection of data

Decision Tree = A classifier algorithm

Hyperparameters = Hypertuned parameters i.e parameters that have been optimised over iteration

Indicator = In this thesis it is used to summarise the most important variables related to answering research question one

Label = A targeted goal the algorithm will learn to classify for

Overfitting = When the algorithm/model learns the patterns of the training dataset too well and thus the prediction for the test set becomes worse

Parameter = A variable defined by a programming method/function

Random Forest = A classifier algorithm that is a collection of many Decision Trees

Table = A table within a database is a folder in which similar data is stored

Variable = A quantity that varies. In this thesis, each variable is presented by data points with different values

Variable importance = A measurement of how important a variable is for classification

Contents

Chapter 1

Introduction

The chapter introduces the challenges of upper secondary education and how Learning Analytics could be used to address them. Firstly a background is given before the research questions are covered. Moreover, the Context of the study is set out and lastly, delimitations of the thesis are highlighted.

1.1 Background

Education paves the way to personal and societal prosperity. Poverty, unemployment, decreased health and increased crime are all factors equated to failed education. Across the world, there is a decreasing trend of literacy in school and dropout rates are increasing across countries, (McPhail, n.d.). Moreover, the Global Goals (n.d.) mention that by 2030 all individuals should be ensured a relevant education. Explicitly examining Sweden, Konstenius and Schillaci (2010) state that having an upper secondary school degree doubles the student's probability of having a job at the age of 24. (In this thesis, the expression Upper secondary school is applied to refer to both the Swedish gymnasium and the US high school.) In addition, Konstenius and Schillaci (2010) mention students with a low school presence miss improvement opportunities of knowledge necessary for further studies, interpersonal skills and increasing confidence. Meanwhile, there is an increasing problem with the small number of authorised teachers. For example, the decreased retention of upper secondary school teachers in Sweden is an emergent problem. There is a need for an increase of 4 100 full-time teachers by the year 2033 compared to the year 2018, (Skolverket A, 2019). A solution to the problem apart from the trivial solution of recruiting additional teachers is proposed by Skolverket A

(2019). The suggested solution is that the school systems have to start working more systematically and long term to improve the organisation and solve the challenges coming ahead.

The most important factor for young adults to establish themselves within the Swedish society and labour market is to graduate from upper secondary school (Utbildningsdepartementet, 2015). The Swedish National Agency for Education has emphasised that adolescents who have dropped out of secondary school are at a greater risk of neither finding work nor continuing studying in upcoming years (Skolverket B, 2016). Reviews by the Swedish School Inspectorate have likewise displayed that initiatives and measures in place are often more reactive than proactive. Skolverket mentions multiple areas that can be more proactive, where decreasing drop-outs is one of many, (Skolverket B, 2016).

With the Covid-19 pandemic, there has been an accelerated increase in using digital technologies within schools. The reason was the need for social distancing and authorities strongly recommending upper secondary school transition to teach online (Folkhälsomyndigheten, 2021). An outcome of the Covid-19 pandemic is the quickened application of digital learning environments and further increased knowledge of using digital tools (Kovanovic et al., 2021). The increase in digital footprints together with expertise from the field of Learning Analytics can be used to support teachers and learners for improved student success (Varanasi et al., 2018). For example, three out of ten students at Swedish upper secondary school do not graduate from their education within the intended three years (SCB, 2017). What further illustrates a problem is the stagnation of students examined from upper secondary school, which according to the Skolverket C (n.d.) is just below 80 % of the students each year.

With accelerated digitalisation and increased quantities of available student data, the interest in Learning Analytics has flourished (Kovanovic et al., 2021). Learning Analytics is recapitulated as exploration and analysis of learners' data and applied to optimise and understand further learning, (Long & Siemens, 2011). Learning Analytics constitutes a possible aid in addressing the rising retention problems in upper secondary schools. It can be used to address problems such as school dropouts (Khalil & Ebner, 2015) and to identify students who will not graduate (Aguiar et al., 2015).

Despite the possibilities Learning Analytics offers to address the teachers' and students' needs in the context of K-12 education, only a small amount of research has so far been conducted in upper secondary schools (Ifenthaler, 2021).

There are several reasons for the slow adoption of Learning Analytics in school contexts. López-Zambrano et al. (2021) mention that there are less data available for K-12 education compared to tertiary levels. In this thesis Tertiary Education refers to education on College and University level. One reason is that K-12 education doesn't have access to their data. However, there are positive amendments occurring. In an investigation by the Swedish National Agency for Education, they proposed technical guidelines to facilitate K-12 education which would include recommendations for data gathering systems (Skolverket D, 2021). According to Tsai et al. (2021), another challenge at the upper secondary level is that stakeholder involvement is necessary to succeed in the Learning Analytics deployment. However, the combination of interested stakeholders, available data and clear ideas of Learning Analytics applications makes it possible to apply Early Warning System concepts to identify students who are at risk of not passing classes (Lakkaraju et al., 2015). Identifying students who are not passing classes would contribute to counteracting the larger problem of students not graduating on time and worst-case dropping out.

1.2 Purpose

The purpose of this study is twofold ensuing both an examining and a developing section of the thesis. First, this thesis aims to study which tendencies of upper secondary school students' learning progress can predict if they will pass a course or not. Second, based on this understanding, propose, construct and evaluate machine learning algorithms for an Early Warning System supposed to assist professional teachers in their work. Moreover, the second purpose is discussed through the didactic framework, formative assesment and the Pygmalion effect.

1.3 Research Questions

To fulfil the purpose the following questions were examined:

- What tendencies are dominant indicators for upper secondary school students not passing a course?
- To what extent are machine learning algorithms possible to use to identify a student who will not pass a course?

1.4 Context of Study

The thesis was authored in collaboration with K-ULF, which is a project initiated by the Swedish government (KTH, 2022). It consists of researchers from KTH and teachers in the Stockholm region. The background of the research project is to develop and test models between academia and schools by performing practitioner based research and research-informed teacher education. To further fulfil the aim, this is done by incorporating research with education. More specifically K-ULF works as a platform of scientific foundation to unite students' thesis projects with practical research with the ambition of providing school development (KTH, 2022).

Through K-ULF the case study school (which refers to the school examinied in this study) expressed their interest in Learning Analytics. The interest of the school came from a group of teachers that proposed an interest in examining the possibilities of applying Learning Analytics due to their availability of data. The school's ambitions were broad but a starting point was to examine if datapoints could be applied to gain more insight into the student's current status in courses. Before the main study was carried out, a pilot study took place at the case study school. The pilot study was conducted through interviews with a teacher and the principal of the school and presented in section 1.4.1 - 1.4.6. Further information on how the pilot study was conducted can be found in section 4.2.1, Awareness of The Problem. Transcripts of the interviews are presented in Appendix A and the two interviews are referred to as interview A1 and A2.

1.4.1 Learning Analytics History at The Case Study School

The first steps towards initiating a Learning Analytics project on the school were taken in 2018 during conversations between the school's personnel and shortly after data of interest began to be collected. By today, the use of gathered data has so far been applied to a small extension. For example, there is a system applied where students self-grade themself and their ambitions. The teachers may respond to the self-evaluation of whether the student is on track compared to their ambitions and the grade the student is heading towards. The data generated by the self-evaluating system can be used by the mentor during performance reviews. In addition, an early analysis of the data was conducted by a teacher at the school in order to see if any trends were possible to detect.

The results were that a few patterns from the Learning Management Systems data could be seen when analysing the variables Participation and Page views, interviewee (Appendix A1).

1.4.2 Prerequisites for Applying Learning Analytics

The case study school is based on entrepreneurship and has a technical profile. Compared to other municipal schools the unique possibilities provide great prerequisites for the application of a new technical system. A more technical proficient teacher has a better understanding of a Learning Analytics project according to the interviewee (Appendix A2). Applying and understanding the product a Learning Analytics project would produce is thus easier. Further, improved technical understanding enhance communication between included stakeholders. The interviewee (Appendix A2) further mentions that the school has an employee dedicated to saving data who has access to all school platforms providing data. Compared to other schools this creates opportunities for projects analysing the data.

1.4.3 Outlooks and Ambitions of the Stakeholders

The interest and ambitions of the case study school management in applying Learning Analytics are broad. They contemplate Learning Analytics would contribute to providing more tools to be used by teachers. Any tendencies or trends based on the student data would be of interest, mentioned by the interviewee (Appendix A2).

Since this thesis is the first major analysis of the school's collected data, the expectations of the stakeholders are broad since there are no concrete predictions on what the analysis might result in, another teacher mentioned. There are however aspirations from the interviewee (Appendix A2) that the outcome could be used by teachers in performance reviews. During a performance review, a system could provide a holistic overview of the student's progression in different courses. Moreover, the interviewee (Appendix A1) mentioned that getting an early detection of students who seem to be slipping off track could help mentors to get in contact with these students before they start falling behind.

Lastly, the interviewee (Appendix A1) mentioned another reason for using Learning Analytics would be to prevent stress among the students and faculty by having a system that would aid the teacher in his or her mentoring role. Furthermore, the system's main purpose would not be to enhance student grades, but that a benefit of countering stress could be that grades improve.

1.4.4 Potential Risks and Problems

In discussions on the risks and downsides of the usage of a Learning Analytics system, the main focus is that the data must be protected, interviewee (Appendix A2). Another aspect brought up concerned the possibility of labelling students, interviewee (Appendix A1). Namely, a student that is labelled as high performing by an algorithm may also be perceived the same by the teacher and graded accordingly, leading to biassed assessments from variables provided by the system.

Even though the technical understanding among the teachers is high at the school according to the interviewee (Appendix A2), some of them might argue that Learning Analytics and data-driven decisions would take over the professional teachers' role. A further risk, according to a teacher at the school, is that applying algorithms that are too abstract could imply less trust in the Early Warning System since the teachers would have little understanding of how it works.

1.4.5 Systems Used at The Case Study School

At the case study school, the data is gathered from various systems. The Absence data is collected from a system the whole Stockholm municipality is using. Teachers are alleged to register the Absence of students on the platform and data is at a later stage pipelined (Data transport between different systems, servers or software). The grades are stored in both a separate system and in the self-grading system. The Learning Management System used at the school is Canvas. The diagnoses are conducted by the students when they are enrolled at the school. All the diagnosis data is either stored on a hard drive or as a CSV file, interviewee (Appendix A1). Further information on the diagnoses is provided in section 2.6.6. What all the systems have in common is that the data is manually extracted and saved into a SQL database (An organized collection of data stored and available on a server).

1.4.6 Diagnoses

The mathematics diagnosis is conducted across schools within Stockholm municipality and for upper secondary schools to request. The English and Swedish tests are conducted on the initiative of the case study school. The score of the English tests are presented by percentage and the Swedish tests are represented as a ratio scale interviewee (Appendix A1). The Swedish diagnosis consists of five separate tests: dictation, spelling, letter chains, word chains and sentence chains. Letter, word and sentence chains test how fast a student decodes written letters, words and sentences respectively.

1.5 Delimitations

In this section, factors that limited the study will be highlighted. The effect of the delimitations on the results is further discussed and with possible implemetations by both the case study school and further thesis workers.

1.5.1 Delimitations of the Final Result

The thesis focused on one upper secondary school within Stockholm municipality which only provides the technical programme of the Swedish upper secondary school for the students. The results can therefore not be considered as general and applicable for every upper secondary school and all possible programmes to choose from within the Swedish school system. Moreover, the results, i.e. tendencies that are dominant indicators and Machine Learning algorithms (see section, 3.2.1 Machine learning) do not provide any certainty of being decisive correct regarding whether a student will pass the class or not. There are limitations within data analysis and Machine Learning and the results should therefore only be seen as supporting tools for teachers within the school. Thus the algorithms are not supposed to be used as formative assessments. The professional teacher should address the feedback with help of the insight the algorithms can contribute with.

1.5.2 Limited Data

Since the data was provided from one upper secondary school the study was limited by the quality and magnitude of their data processing. Thus fewer data points and variables were applied to train the algorithms which could affect the results negatively. The school did not have a precise pipeline system for data, i.e. no system in place that directly gathers the data in a database. Neither were there any guidelines regarding documenting data for Absence or Learning Management System teacher routines, which increase human errors related to data gathering. Course Absence was only available as the total sum for students and not continuous while Learning Management System was used to a different extent in courses. These factors imply that analysis over time would not provide any considerable results for the thesis in contrast to stakeholders' ambitions. Lastly, the GDPR legislation for how data has to be handled puts limitations on what kind of data and how data may be collected and used in this thesis.

1.5.3 Limited Time Frame

The limited time frame of 3-4 months for the thesis, implied merely one Machine Learning algorithm was applied. A project with a greater time frame could include different Machine Learning algorithms to be compared with each other and therefore potentially improve the final prediction accuracy. The framework Design Science Research Methodology recommends multiple circumscriptions before a final product can be put to use. With the time frame creating limits an extensive circumscription was not able to be done based on the evaluation. An extensive circumscription could have resulted in further insight with a better result as the outcome.

1.6 Outline of thesis

The thesis combines two research fields and professions, namely teaching and engineering, in line with the authors' master's education. Namely, a master in science and education, resulting in both a master's degree in science and a teachers degree for upper secondary school. This implies the thesis can be interpreted with the eyes of a professional teacher (Formative assessment, performance reviews and didactical theory) and from a more technical point of view as an engineer (Data analysis and Machine Learning). Moreover, Learning Analytics can be interpreted as a combination of teaching and engineering, see figure 1.1.

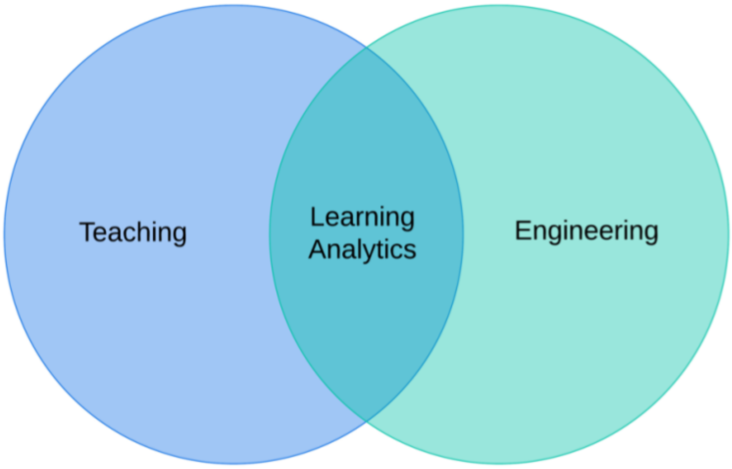

Figure 1.1: *The distinction between the two fields, teaching and engineering, and how they correlate to Learning Analytics.*

The thesis follows the IMRAD structure (Introduction, Method, Results, Analysis, Discussion) with two exceptions, chapter 3 and 5. Chapter 3 addresses both a didactic framework and a part of technical theory. The didactic framework is recommended for the professional teacher to read and refer to the teaching field. The technical part provides a more in-depth understanding of the concepts and relates to the engineering field. Chapter 5 is unique for this thesis and visualises how the data was processed before being used in the algorithms. In the same way, the didactical aspects in chapter 2 together with the discussion of these in chapter 7 relates mostly to a teaching perspective.

Chapter 2

Literature review

The chapter provides information on Learning Analyticsand Early Warning Systems together with the complexities of implementing Learning Analytics. Furthermore, the chapter brings up earlier research on Early Warning System. The earlier research is used to illustrate algorithms, variables and methods proven to be applicable for Learning Analytics research and Early Warning System.

2.1 Learning Analytics

Learning Analytics as a research field has emerged from the rapid technical advancements during the last decades. Resulting in a vast amount of big data within different institutions of education. A popular definition of Learning Analytics is “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs” (Long & Siemens, 2011, p 34) The challenge of Learning Analytics is to find ways to use the data in order to support education and enhance teaching (Larusson & White, 2014). Moreover, Learning Analytics is a very broad field of research as it concerns everything from collecting the data, how to model it, which statistical approaches are most sufficient, how to present and visualise the data and results. The purpose of using Learning Analytics may also vary, for example analysing if pedagogical methods should be changed, monitoring student performance, students’ engagement with the material or finding students that are at risk of failing a course or dropping out (Larusson & White, 2014).

2.2 Early Warning Systems

This section gives a definition and description of Early Warning Systems along with examples of instances where Early Warning Systems have been used. Previous research with the aim of developing Early Warning Systems or dropout algorithms is presented as well.

2.2.1 Definition

An Early Warning System can be expressed as a system that uses variables or information gathered from students to make predictions as early as possible if they are failing a course or is at risk of dropping out (López-Zambrano et al., 2021).

Early Warning Systems are being used in various ways across different levels of education and for different purposes (Larusson & White, 2014) such as keeping track of students' progress in courses or finding and aiding students at risk of failing a course (López-Zambrano et al., 2021). A recent overview conducted by López-Zambrano et al. (2021) shows that a large majority of studies concerning the implementation of Early Warning Systems so far has focused on tertiary education. On the other hand, secondary education has been far less researched (López-Zambrano et al., 2021). This indicates that further inquiries on secondary education level are of importance since it is an unexplored area of research.

López-Zambrano et al. (2021) mention that the definition of early differs depending on which form of education it concerns. Naturally, the length of courses varies between different forms of education but also within the same field. The results of the same study further showed that the choice of relevant variables differs between studies on Early Warning Systems, indicating that every implementation of an Early Warning System requires an individual selection of data and variables, depending on availability and the purpose of implementing an Early Warning System. The most frequently used variables for systems designed for classroom teaching education can however be sorted into three subcategories which are demographics, activity and performance (López-Zambrano et al., 2021). The most frequently used variables in these categories are presented in Table 2.1.

Table 2.1: *According to (López-Zambrano et. al., 2021) the most customary variables to apply for Learning Analytics*

Category	Variables
Demographics	*Age, nationality, sex, city, family income level, having a scholarship, having a job or baby, living with parents, legal guardians educational attainment*
Activity	*Homework grade, homework clicks, attendance, discussion, positive valence, negative valence, neutral valence, an average of valence, ePortfolio engagement variables*
Performance	*Total credits, credits gained, failing credits, passing rate, arithmetic mean score, weighted average credit score, average credit score point, credit score point, failing score*

2.2.2 Examples of Early Warning Systems

Earlier studies on the implementation of Early Warning Systems are presented in this section. The studies covered were conducted in both tertiary and upper secondary education.

Course signals is a system that was developed to aid students to reach their full potential in courses and alert institutions of students that were at risk of failing courses. It was developed at Purdue University in the United States, where data from the Learning Management System was used, (Arnold & Pistilli, 2012). Results of the performed study indicated that courses in which Course Signals were applied had increased degrees of satisfactory grades, decreased degrees of unsatisfactory grades and fewer dropouts from courses (Arnold & Pistilli, 2012). The system alerted institutions and instructors of students that were underperforming and they could then intervene to help the students. Examples of interventions, in this case, were the posting of a traffic signal indicator, with either red, yellow or green on the Learning Management System, messages sent with reminders either using e-mail or text messages. And also referrals to meetings with academic advisors or the instructor.

Additionally, the study included responses from students on how they perceived the Course Signals system. The feedback was positive as the students felt that they were not seen as only a number in the statistics but addressed personally by the instructors (Arnold & Pistilli, 2012). The response from instructors and faculty members was also positive and one person stated that the system helped identify which students needed assistance. In addition, students became more proactive due to the direct feedback of the stoplights on the Learning Management System and started working on projects earlier. However, the system and article by Arnold and Pistilli received some critique both on ethical grounds and their presented results. The main issue with the result was that they presented an increased retention rate of 21% but other data analysis showed that this might not be the case and that the results were misleading (Caulfield, 2013)). The ethical dilemma concerns Purdue licensing the product to a company on the research claims, while the two researchers are not hired tenure. This implies the researchers are held responsible for the work but the institution holds power over what they are allowed to say about the research (Feldstein, 2013).

Student Explorer is another Early Warning System that has been in use since 2011 and was developed for academic advisors. The purpose of the software is to monitor students' progress and to know when further engagement, such as counselling and tutoring is necessary in order for the student to improve their progress within the course (Lonn & Teasley, 2014), (Larusson & White, 2014). The system was substantially based on students' results on assignments and how much they used the Learning Management System as this information was reported on a weekly basis. The advisors could follow the progression from week to week and also compare individual students to the average performance of the class.

Early Warning Intervention and Monitoring Systems is the name of an Early Warning System designed for upper secondary schools in the USA and based on a seven-step program. In an article by Davis et al. (2019), it is mentioned that there are several Early Warning Systems capable of detecting students falling off track. But they also state that a majority of the systems do not provide a strategy on how to get the students back on track. They argue that the implementation of an Early Warning System is not enough and that a strategy for how to use the information to aid the student is needed. Therefore, the first step of the system is a distribution of responsibilities among the school's personnel. Steps 2-3 describe how to implement the warning system and how to review the results of the Early Warning System. Step 4-5 concerns the choice of appropriate interventions and how to monitor them. The final two steps account for an evaluation of the warning system and how to refine it (Davis et al., 2019). Variables used in the system are attendance, behaviour (such as suspensions) and course performance measured by grades and credits. In the study by Davis et al. (2019), they examined 20 schools using the warning system. These schools were matched against schools that did not implement the warning system. Results showed that students at schools where the system was implemented the most were three times more likely to pass their courses than in schools that did not implement it at all.

A study conducted in the US between 2009 and 2012 used Machine Learning to detect students that were prone to dropping out of upper secondary school. The researchers examined data that was available during the 9th grade (14-15 years) and then followed it up in the 11th grade (16-17 years) (Sansone, 2019). In the study, the variables that were the most important in training the Decision Tree are presented. The variables' importance in order were:

- The student's grade point average in the 9th-grade
- The year the student was born
- The students' maths test scores from a national survey
- If the students had previously been suspended or expelled
- If the parents had been contacted more than 4 times or not concerning poor attendance

Moreover, the study mentions that drop-outs from upper secondary school are not homogeneously represented, with an overrepresentation of male students. Further, the reasons for dropping out varies between different students and therefore the measures taken to aid the students differs too (Sansone, 2019). However, the results of the study presented that Machine Learning algorithms could be used to assist in identifying students who were at risk of dropping out (Sansone, 2019).

The aforementioned examples of Early Warning Systems employed in upper secondary education originate from the USA. Even though the educational systems of the USA and Sweden are comparable in many ways there are some aspects where they differ. For example, the United States does not have a national educational system, but instead each state is in charge of creating its own. It results in a large variety of school forms (Loo, 2018). Another difference is that upper secondary schools span over four years compared to three years in Sweden (Loo, 2018). Furthermore, the student is in charge of putting together his or her own curriculum depending on which courses they are interested in (Loo, 2018). Compared to Sweden, there are 18 national programmes with different orientations the student can choose from and only a few courses that are voluntary (Skolverket E, 2021). A further difference is the broad use of standardised tests in the USA that are machine graded and used for admissions, placement and counselling decisions (U.S. Department of Education, 2022).

2.2.3 Results of earlier research

Earlier studies within Learning Analytics presents varying algorithm accuracy prediction, using different algorithms while the number of data points differs. The algorithms with the highest accuracy applied within Early Warning System were found in the following studies: Miguéis et al. (2018), with an algorithm prediction accuracy of 96.1% by using Random Forest. Razak et al. (2018) achieved their result using linear regression which had a prediction accuracy of 96.2%. Chung and Lee (2019) who also applied Random Forest, did instead achieve a prediction accuracy of 95%. Costa et al. (2017) achieved 92% accuracy by using Decision Trees and Naive Bayes. Lastly, Natek and Zwilling (2014) achieved a prediction accuracy of 97% by applying Random Forest.

The result of Natek and Zwilling (2014) can be questioned based on the number of data points. They conducted the algorithm on two samples of student groups, n=32 and n=42 students. Compared with Chung and Lee (2019) who had data points from 12 000 students and Miguéis et al. (2018) with a total of 2459 data points. Meanwhile, Costa et al. (2017) had two groups of 262 and 161 students generating data points over the duration of 16 weeks. Miguéis et al. (2018) mention that a small number of data points are possible to be used for providing insight, even though the general opinion is that more data points generate better algorithms (Hastie et al., 2009).

Costa et al. (2017) mention two major problems that may arise during an analysis of student-driven data, i.e., the high dimensionality of variables and imbalanced data. High dimensionality, implies using too many variables may hinder the prediction accuracy an algorithm is able to reach. Therefore it is recommended by Costa et al. (2017) to perform an analysis of how much the variables bring value to the algorithm. Costa examined the variables by using the library WEKA, developed using Java. From there, an algorithm based on information gain was used to sort the variables. The analysis was performed parallel with running the algorithm. Thereafter the algorithm was run again with only optimised variables. Chung and Lee (2019) and Miguéis et al. (2018) used similar methods while examining the variable's importance. Miguéis et al. (2018) used Gini Index to evaluate how much the algorithm was promoted by each variable after the initial algorithm was performed. Chung and Lee (2019) evaluated the importance of the variables by analysing the out of bag errors contributed by the specific algorithm.

An imbalanced data set is a data set (a collection of data) where one label (targeted goal the algorithm will learn to classify for) is much more represented in numbers compared to the second label as described by Costa et al. (2017). An imbalanced data set is problematic since algorithms tend to focus their prediction on learning how to categorise the label that is of the majority numbers. In the report by Costa et al. (2017) the number of students that passed and did not pass were not equally many and therefore they applied the Synthetic Minority Over-sampling Technique (SMOTE) algorithm. The algorithm creates a balanced data set where the number of samples by each label now are equal. Chung and Lee (2019) did likewise use the SMOTE algorithm since they had an equivalent problem.

2.3 Learning Management Systems

The digitalisation of educational environments has enabled the use of Learning Management Systems. By Turnbull et al. (2019, p.1) who presents an overview of Learning Management System and define it as, "web-based software platforms that provide an interactive online learning environment and automate the administration, organisation, delivery, and reporting of educational content and learner outcomes". Learning Management Systems offers possibilities for enhancing teaching and learning by storing course-related content like documents and multimedia files. Moreover, the teachers can construct quizzes for the students to perform while they can submit their assignments directly on the platform. A Learning Management System further includes tools for communication, for instance, message functions, discussion forums and the possibility as a teacher to make announcements. Lastly, Learning Management System provides possibilities for the teacher to oversee students' progression and mark assignments, while the students can follow their own process (Macfadyen & Dawson, 2010).

Learning Management Systems have the variable of recording and storing the users' actions on the platform. Examples of user data stored are: pages visited, messages opened and posted along with the number of visits on the platform and their duration Macfadyen and Dawson (2010). The information offers an opportunity to research aspects of education and student behaviours that would otherwise be difficult to analyse Macfadyen and Dawson (2010). Educational institutions use a variety of information systems alongside Learning Management System including library systems, digital repositories and administrative systems. The multitude of systems results in an increasing amount of data that can be used to enhance education (Sousa et al., 2021).

2.4 Complexities in Learning Analytics adaptation

The adaptation of Learning Analytics presents a number of concerns that have to be dealt with in order for it to be successful. This section brings up the following aspects: data ethics, resources, level of education and the inclusion of stakeholders.

2.4.1 Data ethics

The digital footprints of students in educational environments can be used in a wide scope of research. An example is examining the patterns students generate when solving problems (Murchan & Siddiq, 2021). The data provided by educational platforms and digital tools offer new ways of analysing students' learning strategies. This information gives new insight into how teaching and learning might be enhanced (Murchan & Siddiq, 2021). However, the focus on technology and on how to process data in Learning Analytics is far ahead of the ethics focus. Even though the processing of student-based data is making promising progress, related ethical and regulatory implications are often not taken into adequate account (Murchan & Siddiq, 2021). Much analysis starts from complex files containing information on how students are engaging with assignments. Regarding documentation within files, there is an Absence of ethical and legal knowledge in relation to the processing of the data, (Siddiq et al., 2017). Meanwhile, an increasing magnitude of data is making the situation more challenging. Despite the possibilities to enhance learning and teaching the situation can be summarised with risks concerning intrusion of pupils' privacy (Murchan & Siddiq, 2021).

In 2018, the EU initiated General Data Processing Regulation (European Union, 2018). The regulations include a statement of how user data is allowed to be processed and applied to organisations such as school institutions. Principles the organisations apply to are stated in European Union (2018) and summarised by Murchan and Siddiq (2021, p.8) as

- Process an individual's personal data lawfully and fairly, providing transparency about its specific purpose
- Collect data for a legitimate, limited purpose
- Collect from an individual no more data than is necessary for the purpose for which it will be used
- Ensure that data are accurate and up to date and erase inaccurate data
- Store the data for no longer than is necessary for the intended purpose
- Keep data confidential and secure from loss or unauthorised processing.

Moreover, Murchan and Siddiq (2021) conclude that the GDPR guidelines are considerable but can be further improved. More precisely, they state that research within Learning Analytics on processing data has to be better at addressing issues of ethics, privacy and regulation compared to GDPR statements. Murchan and Siddiq (2021) suggests research using an ethics framework Learning Analytics can build upon. One example according to Murchan and Siddiq (2021) is the Sclater (2016) framework which addresses 8 variables (responsibility, transparency, consent, privacy, validity, access, minimising adverse impacts and stewardship of data) that has to be analysed.

2.4.2 Resources

Resource challenges include funding, people and data, where data is seen as the major concern (Tsai et al., 2021). Returning issues concern the quality and the scope of data. Due to digitalisation, there has been an increase in data, (Tsai et al., 2021). However, according to Siemens (2013), the process of obtaining and integrating data from different sources has been reported as a difficulty. Further challenges related to data is the purpose of Learning Analytics to inform learning and teaching with evidence-based data. To succeed there is a necessity to develop data literacy among key users, that is stakeholders and teachers (Tsai et al., 2021). Regarding funding the attention of Learning Analytics has to compete with other institutional factors, which results in prioritisation within institutions where the financial support of Learning Analytics often falls (Tsai et al., 2021).

2.4.3 Education levels and education system

Learning Analytics and more specifically early prediction can be applied at various educational levels and different systems of education. There is traditional education, referred to as education practised in schools and in-person. Further, there is E-learning, which is online learning done through digital channels. Moreover, a mixture of the two can be included, that is Blended learning (Romero & Ventura, 2013). The different educational levels are primary education (initial stages of basic education), secondary education (final stage of basic education) and tertiary education (education provided at universities leading to academic degrees).

López-Zambrano et al. (2021) investigated the distribution of Learning Analytics applied at different levels and systems by examining the magnitude of research papers related to the fields and levels. The results are presented in figure 2.1. The general theme is that research is mostly done at a tertiary level and within E-learning, which is in accordance with the accessibility of data (López-Zambrano et al., 2021). The figure, therefore, illustrates gaps where the magnitude of research and application of Learning Analytics can be increased. Ifenthaler (2021) further states that despite the proven advantages of applying Learning Analytics at upper secondary school, few publications have explored the benefits. One reason why there is a small level of research and application of Learning Analytics in upper secondary school is due to the difficulties of including stakeholders (Tsai et al., 2021). According to López-Zambrano et al. (2021), another reason is due to the accessibility of data. At a tertiary education level, there is a larger number of learning environments which results in more opportunities for data management and analysis.

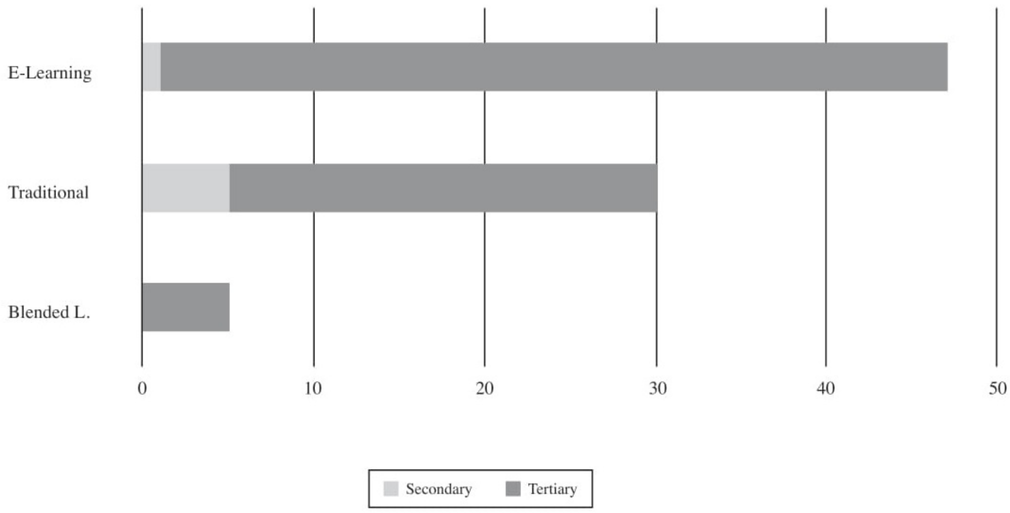

Figure 2.1: *Education level data by type of learning environment* Note: Adapted from Early prediction of student learning performance through data mining: A systematic review, by López-Zambrano et. al, 2021. Copyright © 2021 Psicothema

2.4.4 Inclusion of Stakeholders

Stakeholder involvement is crucial for the success of Learning Analytics deployment. For Learning Analytics to be utilised successfully, the area of concern must be expressed and illustrated clearly by the stakeholders (Howell et al., 2018). For teachers as stakeholders to adopt Learning Analytics, there have to be intentions to make it pedagogically useful (Tsai et al., 2021). Discrepancies in experience and data literacy among stakeholders increase the challenge of finding common solutions where every included stakeholders' expectations are met. To address this, it is necessary to bridge the gap between stakeholders and the benefits of technological capacity in learning environments (Tsai et al., 2021). According to Sousa et al. (2021), it is recommended to proceed from a framework for investigating the upper secondary schools' need for Learning Analytics while stakeholders are included. There exist multiple frameworks that can be used but for this thesis, a framework that takes both stakeholders and the product to be developed into account is of interest. One example of such a framework is Design Science Research Methodology. See chapter 4. Method for further details on Design Science Research.

2.5 UN's global goals

Quality Education is one of seventeen global goals drawn up for 2030 by the United Nations to ensure inclusive and lifelong learning for everyone. One of the subgoals explicitly states: ”By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes.”(Global Goals, n.d.). Through the perspective of Learning Analytics and an Early Warning System tool, the subgoal is addressed due to the possibilities of directing students towards a passing grade.

Chapter 3

Theoretical review

In this chapter, the theoretical frameworks used in the study is presented and explained in depth. First didactic perspectives are presented, creating the didactic framework through which Learning Analytics is examined. Secondly, the theory related to engineering is presented which consists of theory on more advanced technical level.

3.1 Didactive framework

First the section unfolds the didactive framework of formative assessment and how it relates to Learning Analytics. Secondly, theory of the Pygmalion effect is presented and lastly, grades and performance reviews of the Swedish school system are presented.

3.1.1 Formative assessment

The section's purpose is to provide an overview of formative assessment and to present to the reader how it is related to data collection and Learning Analytics. Moreover, this section uncovers how formative assessment could activate students as owners of their own learning and how this could be motivated by the support of Learning Analytics. Reconnecting to the purpose of the study, the following section is not supposed to give the professional teacher a recommendation on how to use formative assessment or how the feedback should be addressed. Providing feedback is instead seen as a task for the professional teacher.

Definition of formative assessment

The purpose of formative assessment is to inform the students about their learning progress and classroom actions (Wiliam & Leahy, 2015). In the field of formative assessment, a definition is difficult to draw up according to Black and Wiliam (1998). Hence the definition may vary depending on the context. In the research article, however, formative assessment is described as: "it is to be interpreted as encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged." Black and Wiliam (1998, p.7-8).

Within formative assessment a reaction from the student (recipient) upon feedback given by teachers is preferable. No matter how good the communicated feedback is unless the student acts on or responds to it, it is a waste of effort Wiliam and Leahy (2015). The word feedback originates from the labour market. Thus it can be seemed hijacked while applied in the context of formative assessment. In the labour market feedback is distinguished by it being a loop where recommendations and comments should imply an initiative of change, (Wiliam & Leahy, 2015). Meanwhile, formative assessment has interpreted feedback more as the actual recommendation given and not the whole loop. A potential problem could thus be that the definitions differ.

Wiliam and Leahy (2015) bring forward an important takeaway based on research of formative assessment, which is the time eliciting between the gathering of evidence and using it to provide the student with feedback. Moreover Black and Wiliam (1998) makes a partition into three cycles depending on the time factor for the applied feedback, long-cycle, medium-cycle and short-cycle. Long-cycle is defined as instructional adjustments on a month-to-month basis, medium-cycle on a week-to-week basis and short-cycle imply a response within minutes. Wiliam and Leahy (2015) imply all feedback is for a good cause, but mention that the shorter the time interval is, the bigger the impact of potential learning. With this in mind, short cycle feedback would be the most beneficial for the student.

Student independencies

The most important key strategy regarding formative assessment is activating students as owners of their own learning (Wiliam & Leahy, 2015). Students control their own learning and therefore they dependably regulate how they will respond to the feedback (Tempelaar et al., 2013). The most beneficial student behavioural change with a positive outcome is feedback resulting in students straining harder or increasing their ambitions (Wiliam & Leahy, 2015). According to Tempelaar et al. (2013), Learning Analytics could be applied to stimulate students as owners of their own learning since it is providing a multitude of information the student could use to improve personal learning habits. Moreover, it creates an opportunity for the student to start reflecting on her strengths and weaknesses.

Learning Analytics and formative assessment

An important aspect of formative assessment is to gather data supporting conclusions teachers have outlined (Wiliam & Leahy, 2015). Wiliam and Leahy mention that the purpose of the data collection has to be stated before the analysis of data to provide meaningful feedback. The question remains whether datapoints for formative assessment purposes and Learning Analytics can be interpreted as the same.

Learning Analytics can however provide tools to stimulate the communication represented as feedback by using visualisations based on learner activity-related data to improve the possibilities of students reflecting on their behaviour (Tempelaar et al., 2013). According to a study performed by Buckingham Shum and Crick (2012), visual effects based on data enable reflection on learning and taking responsibility for their own learning. But the purpose is twofold since it also enables the teacher to compare learning characteristics as a group compared to the individual.

3.1.2 Pygmalion effect in the classroom

In an experiment conducted and presented by Rosenthal and Jacobson (1968), there was evidence that a person's expectations of another individual may imply biased opinions. This is known as the Pygmalion effect, or Teacher expectancy effect (Szumski & Karwowski, 2019). 50 years ago Rosenthal released his initial study and since then continuous studies have been done within social science to further evaluate the Pygmalion effect (Szumski & Karwowski, 2019). What the Pygmalion effect in detail describe is that, if teachers expect students to show certain behaviour or intellect, the same students did likewise show the promoted behaviour or intellect (Rosenthal & Jacobson, 1968). However, evidence shows that the effect is more common to occur among younger children (Rosenthal & Jacobson, 1968). Szumski and Karwowski (2019) mention that conditions of biased expectations by teachers in a majority of cases refer to stereotypes associated with students' race, socioeconomic status or gender. A further dimension of the Pygmalion effect and the outcome of it is how the expectations are expressed to the students and thus how it would be transformed into student achievement or not (Szumski & Karwowski, 2019).

3.1.3 Performance review

Teachers conducting the performance review [utvecklingssamtal] are representatives of the school and therefore responsible to inform both the student and the custodians of the students situation. It is important for a teacher to be provided with information regarding all the classes a student is enrolled for. Thereafter, the mentor provide the student with information to be used as development for both the indivudal student and the class. Performance reviews within the Swedish school are conducted twice a year and unless the student is 18 years old custodians have to be informed. (Skolverket F, 2022)

3.1.4 Grades

The Swedish grading system for upper secondary education contains six grades, A-F. The grades A to E are all passing grades where A is the highest and E is the lowest. Grade F means that the student has failed any or all of the requirements for grade E (Skolverket G, 2017).

3.2 Review of technical theory

In this section, the theoretical technical knowledge used in the study is presented and explained in more detail. The theory consists of algorithms, data processing tools, visualisation tools and prediction concepts. All of the concepts presented correlate to the technical part of the thesis.

3.2.1 Machine learning

Machine learning is explained as the automation and improvement of computers' learning process based on experiences. The process is initiated by feeding data that is used to train machines. To support the cause, models and algorithms are built to improve the machine's possibilities to learn. Machine learning can be categorised into three subcategories. That is, Supervised learning (algorithm learns on labelled input data), Unsupervised learning (data for the algorithm is unlabeled) and Reinforcement learning (the algorithm learns by making mistakes. For every new mistake the algorithm tries again with the past mistakes taken into account). The choice of algorithm depends on what the dedicated reason for applying Machine Learning is Chung and Lee (2019).

In Supervised learning, the Machine Learning algorithm is trained to learn relations between predictors (descriptives) and outcomes (targets). With the trained algorithm the purpose is to predict and categorise further observations. Predicting what grade (label) a student is heading towards based upon student data, such as attendance and activity (descriptives) is an example of a supervised learning algorithm. The most common fields for applying supervised learning algorithms are regression and classification, (Chung & Lee, 2019).

3.2.2 Classification algorithms

A survey done by López-Zambrano et al. (2021) displays multiple options for analysing data in a Learning Analytics project. Among the options is the classification technique the most standardised. Classification is defined as a process of recognition and grouping of objects into predetermined categories (Chung & Lee, 2019). The basis of the performance measure of a classification algorithm is to analyse how many classifications are done correctly. For a test sample, the algorithm makes a prediction and the prediction outcome is compared to the actual label. This is repeated for the entire group of test samples and the average number of predictions done correctly is given by the accuracy (Chung & Lee, 2019). López-Zambrano et al. (2021) mention the best performing classifications algorithms were: Decision Trees, Random Forest, Support Vector Machine and Naive Bayes.

3.2.3 Decision Trees

A Decision Tree is a classifier designed to find patterns that sort data points into two different categories from a number of variables through an iterative process. The process is binary, starting with the value of an attribute as a threshold and splits the data points into two different nodes. The primary node is referred to as the parent node and the nodes constructed after the split are referred to as child nodes. The process is continued until all data points have been ideally sorted. At first, the tree is maximised, i.e a supreme number of levels are created. The tree is then reduced by removing nodes that are ineffective at sorting the data points. This process continues until the tree is minimal with only the most efficient nodes left, (Wu et al., 2008).

The method for evaluating the classifier strength of a node (n) is the Gini measure (G) of impurity which is:

$$G(n) = 1 - p(n)^2 - (1 - p(n))^2$$

p(n) is the fraction of correctly sorted data points into a child node. This indicates that a low Gini value corresponds to an efficient data split and thus a Gini value equal to 0 implies that the data in a child node is homogenous and therefore no further splits are necessary, (Wu et al., 2008). The Decision Tree further evaluates which nodes to prune by an Improvement gain (I) defined as

$$I(P) = G(P) - qG(L) - (1 - q)G(R)$$

P is the parent node, L and R are the left and right child nodes respectively. q stands for the quota of data points that are placed in the left child node. In this way the algorithm decides on which splits that are most efficient to use.

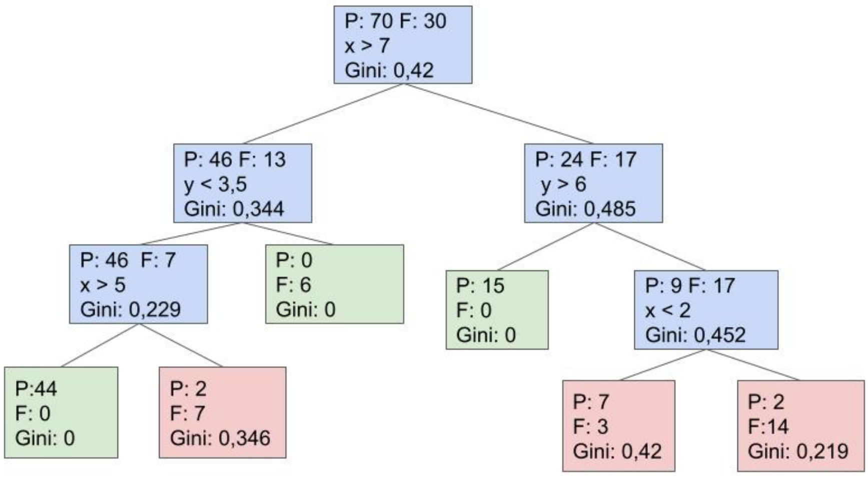

Figure 3.1: *Example of a Decision Tree*

The Decision Tree illustrated in Figure 3.1 has been trained with 100 data samples (rows of student data), 70 with P and 30 with F. Using the variables x and y from the data, the algorithm has managed to separate a part of the samples correctly.

3.2.4 Random Forest

Random Forest is an algorithm built upon the understanding of Decision Trees. The strength of Random Forest in comparison with Decision Tree is that a Random Forest uses multiple trees to make a prediction. Thus one advantage of Random Forest is that it reduces overfitting (the algorithm learns the patterns of the training dataset too well) of the data (Hastie et al., 2009). During the training phase of the algorithm, multiple Decision Trees are constructed. The idea of constructing multiple trees is based on the concept of bootstrapping. Bootstrap implies, that instead of training the trees with all the data samples, every Decision Tree is trained with a subset of data samples, (Breiman & Cutler, 2004). The unique subset of samples for the unique Decision Tree is called its bag and the remaining data samples are called the out of bag points. Each Decision Tree is trained as before, based on the Gini index and Information gain. To summarise, every Decision Tree looks different since the subsets and the out of bag samples are unique. Figure 3.2 is an example of a Random Forest and its Decision Trees. About 63% of the total data is used within each subset (Chernick & LaBudde, 2011).

The out of bag samples are not squandered (not used). Each data sample is used to test the Decision Trees where it is not part of the bag. Every data sample is run through all Decision Trees and a majority vote is registered on what classifier that unique data sample is being registered as. This is repeated for all data samples, (Breiman & Cutler, 2004). The proportion of out of bag samples that were incorrectly classified is the out of bag error. The out of bag error is thereafter summarised as an average for all data samples and is used to describe the Random Forest algorithm's internal accuracy. This process is summarised by the name aggregation, (Breiman & Cutler, 2004).

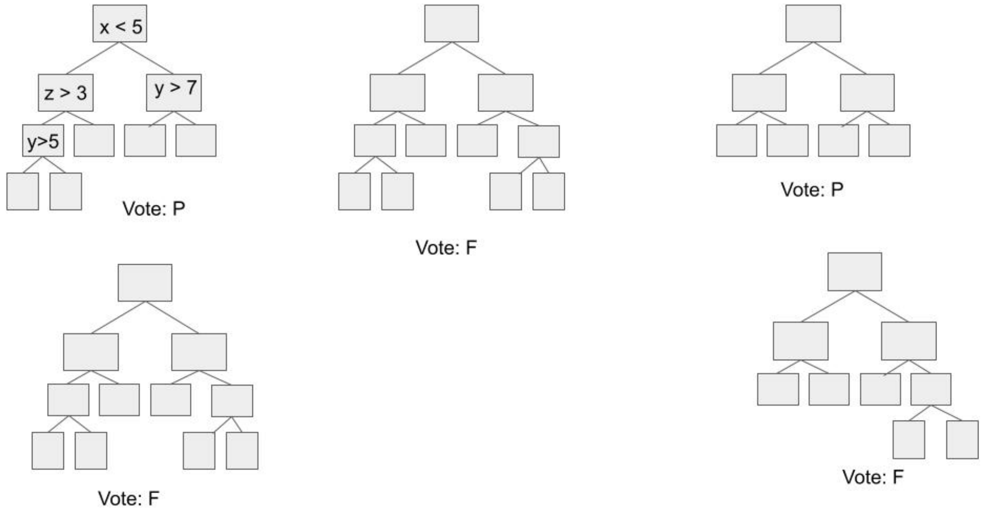

Figure 3.2: *An example of an Random Forest algorithm based on five Decision Trees*

Figure 3.2 displays a Random Forest consisting of five Decision Trees. x,y, and z are the variables used to train the trees as described in section 3.2.3. In this figure all five trees are uniquely constructed using the method Bootstraping implies that all trees classify differently. The variables x, y and z for a student are then used in the Random Forest to classify the student as either passing (P) or failing (F). For example, x could be the variable Absence while y and z are represent Diagnosis results. The outcome in the figure is that two trees classify the student as passing the course (marked with Vote: P), while three trees classify the student as failing it (marked with Vote: F). Therefore, the algorithm classifies the student as not passing the course since the majority of the trees classify the student with a grade of F.

3.2.5 Confusion matrix

A confusion matrix is applicable for binary classifications with the purpose of achieving a wider perspective regarding the quality of the algorithm compared to merely testing the accuracy. First of, predicted values are sorted into four different groups related to the true value as in figure 3.3. True positive (TP) indicates the number of positive examples classified accurately and True negative (TN) is the number of negative examples classified accurately, (Kulkarni et al., 2020). By convention, the class label of the minority class is positive, and the class label of the majority class is negative, (Han et al., 2005). False positive (FP) presents the number of negative values classified as positive while False negative (FN) are positive values classified as negative, (Kulkarni et al., 2020). According to Han et al. (2005) the accuracy with values from a confusion matrix is given as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

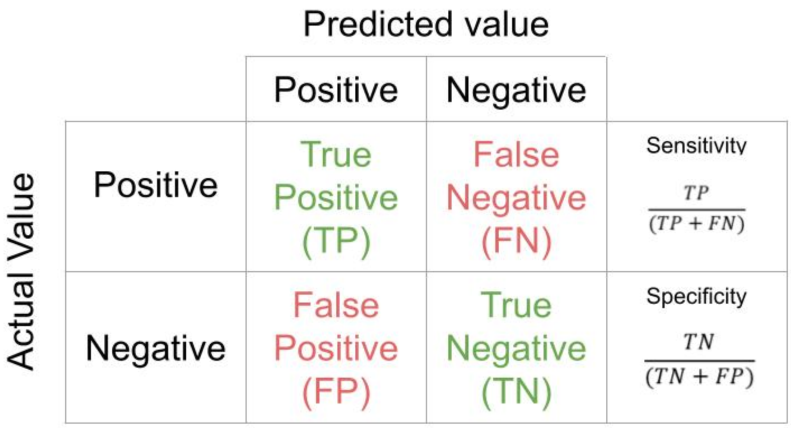

Figure 3.3: *Confusion Matrix and formulas for sensitivity and specificity*

Only examining by accuracy can be misleading for imbalanced data sets due to the accuracy paradox. The paradox states that for an imbalanced data set the classification accuracy is based on measuring the majority of the major data sample label (Kulkarni et al., 2020). Thus, more reasonable evaluation metrics and formulas are needed to complement the potential of accuracy to be misleading (Kulkarni et al., 2020). Two evaluation criteria that can be used in addition to accuracy are sensitivity and specificity. These criteria handle the prediction of the classification groups separately and thus imbalanced data sets are not equally affected (Márquez-Vera et al., 2016). The formulas are displayed in Figure 3.3

3.2.6 Imbalanced Data sets

An imbalanced data set will in the thesis use the definition of Han et al. (2005). Their definition of a binary imbalanced data set is a data set where the two labels of the two classes are not represented equally. Han et al. (2005) mention two methods for dealing with imbalanced data sets. Either the distribution of the imbalanced changed or existing data algorithms are being modified.

Changing the distribution of the data set can be done in two ways. Either by under-sampling the data, meaning that data from the major classifier is randomly deleted to match the two pools of data. However, deleting data implies fewer data points to feed the algorithm with. Instead oversampling (increase the number of data points for the minority classifier) is preferred. Han et al. (2005) have created a method called Synthetic Minority Over-sampling Technique, (SMOTE). The method generates new samples among the minority pool. SMOTE makes decision regions larger and less specific. To summarise, the data set is getting more evenly distributed and the method is proven to increase the accuracy of the classification algorithm used.

Modifying the existing data mining algorithm can be done in two separate ways (Han et al., 2005). Either by boosting the algorithm, that is training many individual models within a sequence and thus the model improves from its earlier mistakes in the sequence chain. The customary algorithm to apply for the mentioned case is Adabosting and it is mostly applicable for the SVM. Bagging is another modification that can be applied. Bagging is the sum of the two concepts of Bootstrap and Aggregation. It implies that multiple models are trained in parallel and by a random subset out of the total data points. A frequently used algorithm for bagging is Random Forest (Han et al., 2005).

3.2.7 Hyperparameter tuning and cross-validation

Hyperparameters or Hypertuned parameters are parameters that has been optimised over iteration. Theses hyperparameters are used to tune the algorithm settings. Machine Learning algorithms are represented within different programming languages but they have in common that they consist of parameters that change the outcome of the algorithm. Many of the hyperparameters are related to the quantity and magnitude of the data but generally speaking, the user can decide how to apply the tuning settings. The final accuracy of a model is linked to the quality of the fine-tuning done on the hyperparameters of the algorithm. Adjusted hyperparameters may result in more accurate models (Queiroga et al., 2020). To find the best hyperparameters the optimal case is to iterate all possible combinations of them (Chung & Lee, 2019).

To avoid overfitting, Cross-validation can be applied. With the method of k-fold cross-validation, the training data set is divided into k number of sub-data sets. k-1 number of sub-data sets are used to train the model and the k:th data set, known as the validation data set, is applied to test the model. The cross-validation iterations continue over every k data set. Thus the average accuracy can be calculated for every iteration of a data set (Chung & Lee, 2019).

Chapter 4

Method

This chapter starts with a contextualisation followed by a definition of the applied research process, Design Science Research. Secondly, the chapter describes how this thesis applied the phases of Design Science Research and interpretations of the framework unique for the thesis, including a walkthrough of the pilot study. Moreover, both a timeline and the data workflow is visualised. Lastly, ethical considerations are presented.

4.1 Design Science Research

In this thesis Design Science Research Methodology has been used. Design Science Research emphasises the configuration of artefacts, that is systems, methods or algorithms (Peffers et al., 2018) and is often applied in engineering and computer science (Vaishnavi & Kuechler, 2019). Design Science Research projects frequently start with a research problem stated by a user (Peffers et al., 2018). However, there are no distinct classifications of who is supposed to take initiative within a Design Science Research project, as Vaishnavi and Kuechler (2019) mention that multiple sources can be included from start. A Design Science Research application consists of five process steps: awareness of the problem, suggestion, development, evaluation and conclusion (Vaishnavi & Kuechler, 2019).

4.1.1 Adaptations in The Present Thesis

Design Science Research is a methodological framework and therefore a proposal of how to set up the work process. Vaishnavi and Kuechler (2019) mention that Design Science Research is interpreted and applied differently by different researchers and therefore a multiplicity of variants exist. This thesis adds to the large variety of practised Design Science Research methodologies. The differences between how this thesis applies Design Science Research and the description of the methodology by Vaishnavi and Kuechler (2019) are found in the phases of Awareness of the problem (Section 4.2.1), Evaluation (Section 4.2.4) and Conclusion (Section 4.2.5). Continuous evaluations were done resulting in circumscriptions and the final evaluation did not result in returning to Awareness of the problem due to the limited time frame. Further, a pilot study (Section 1.4, Context of study) was conducted to illustrate the context of the study at the case study school. The pilot study contextualises through a didactic framework the benefits of applying Learning Analytics. Thus it creates an equivalence next to Design Science Research being more of a technical appliable framework.

4.2 Design Science Research Methodology

The five phases that pose the Design Science Research Methodology process will be theoretically described followed by an explanation of how each phase was applied. In figure 4.1 the Design Science Research model visualises the process steps related to workflow, actions applied and outputs of each phase. The process steps go in order from top to bottom. The arrows indicate the possibilities of reformulating the initial idea based on knowledge, insights and impediments achieved during the thesis process.

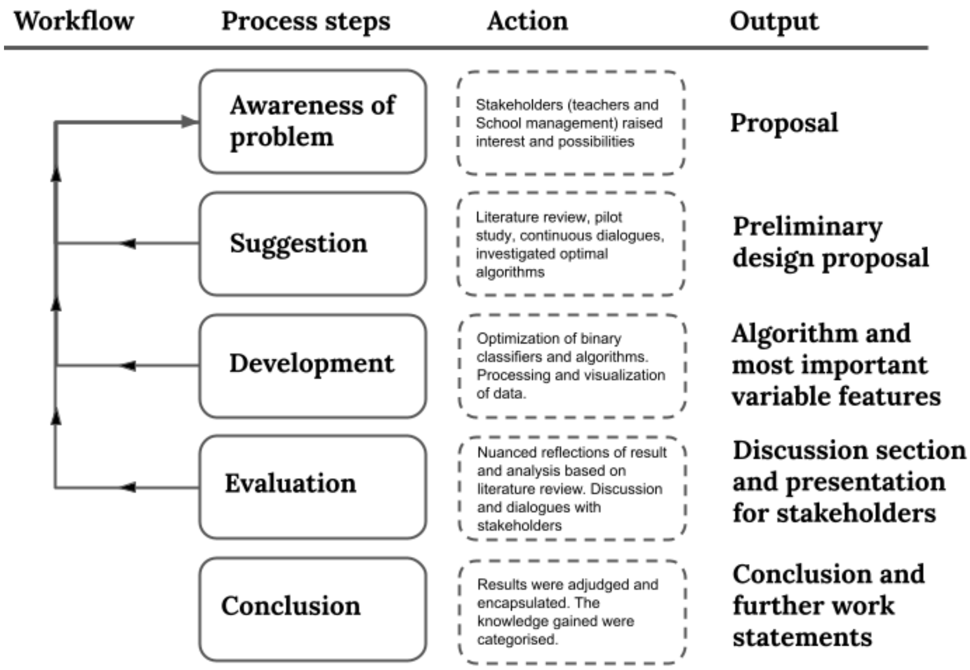

Figure 4.1: *Design Science Research Methodology cycle: The current study*

4.2.1 Awareness of the problem

Awareness of a research problem to solve may involve different stakeholders but a returning theme is the stakeholder's need for a product that will fulfil a purpose. Design Science Research Methodology tends to be effective when a problem-solving focused approach is possible to address the research problem. After researchers have gathered considerable insight into the problem, the most important aspect is to consider the criteria to be applied when evaluating the final product. The output of the phase should be a formal or informal proposal addressing the research effort (Vaishnavi & Kuechler, 2019).

The involvement of stakeholders is proven to be a necessity for a Learning Analytics project to succeed (Howell et al., 2018). Stakeholders are in this case teachers and school management. The experience gap between included stakeholders' experience and knowledge has to be bridged. To address stakeholder inclusion, Sousa et al. (2021) mention that an optimal starting point is as simple as dialogues.

In the thesis, *the awareness of the problem* was initiated when a group of teachers and the IT-pedagogue at the case study school in 2017 initiated data gathering with ambitions to perform Learning Analytics in the future. The idea of applying Learning Analytics was thereafter presented by one of the teachers to K-ULF (section 1.4, Context of study). Hence a first meeting was conducted between the stakeholders (thesis writers, K-ULF representative and teachers) in November 2021. The meeting was followed by a visit to the case study school in December 2021, where two teachers, the IT-pedagogue and thesis writers attended. During the visit to the school in December and followed by another in January 2022 the initial ideas and possibilities of Learning Analytics were drafted. During the visit in January two encrypted USB flash drives of data were handed over. Based on the school's needs and interests the proposal was to create an Early Warning System built upon algorithms based on the following criteria: applicable in conjunction with performance reviews with students, technically understandable for teachers and possibilities to further improve and expand upon (Appendix A1, A2).

To achieve stakeholder inclusion to benefit development of both the thesis and algorithm a pilot study along with continuous dialogues throughout the project were conducted. The pilot study was carried out through semi-structured interviews that were performed in Swedish to further evaluate the stakeholder's needs, thoughts and reasoning. They were conducted in early February and contributed to gain insight of the research problem and the criterias to be tested were further distinguished. Semi-structured interviews are applicable for gathering facts, opinions and reasoning in an interview (Dyckhoff et al., 2012). A formal questionnaire with open answers was used during the interviews to simplify the comparison of answers between different interviewees (Björndal, 2005). The use of a semi-structured interview further provides the advantage of spontaneously investigating details of interest (Dyckhoff et al., 2012). In total two semi structured interviews were conducted. One with a teacher and one with the principal. The interviews were thereafter transcribed and referred to both in section 1.4, Context of study and section 7, Discussion and presented in Appendix A, Transcription of interviews. In addition to the semi structured interviews continuous dialogues with one teacher were taking place every second week during regular recurring K-ULF meetings. During the meetings, the teacher was informed of how the project was progressing while results were provided and discussed. The feedback from the stakeholder was taken into consideration when determining how to proceed with the project for the upcoming two weeks. The dialogues concluded helpful information for the thesis and are thus also referred through notes taken from the meetings. In this thesis those dialogues are referred to as, teacher at the case study school.

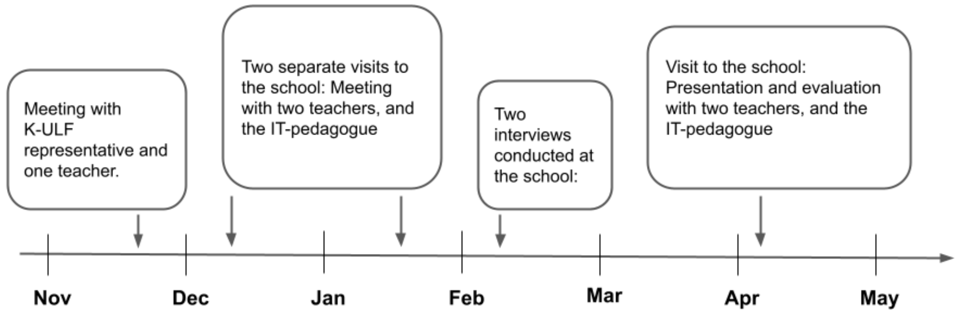

Figure 4.2: *Timeline of when meetings and interviews occurred with stakeholders, apart from the regularly occuring K-ULF meetings.*

4.2.2 Suggestion

Based on the proposal defined in Awareness of a problem, is the phase of Suggestion. In the phase, functionality and variables to potentially implement are discussed based on what is to be changed or applied according to the proposal. The goal is to establish a preliminary design suggestion (*Tentative design*), answering the proposal. If the design won't fulfil the proposal, the idea is discarded and the proposal has to be redefined (Vaishnavi & Kuechler, 2019). The *suggestion* phase is a creative part of the project and can include workshops and dialogues with stakeholders to gather a greater magnitude of design ideas (Peffers et al., 2018). To summarise, the goal is to establish a hypothetical design answering the demand of the proposal to be developed in the next stages of the project.

Leading up to the *Suggestion*, the theory of algorithms had to be enriched. Thus a literature review was done, researching optimal algorithms to apply and interpret Learning Analytics projects aiming at design and evaluation of Early Warning Systems at upper secondary schools. Moreover, stakeholders' thoughts and opinions were taken into consideration, by applying the aforementioned semi-structured interviews and continuous dialogues. The phase resulted in an initial idea of the preliminary design, which in this thesis was to apply Random Forest.

4.2.3 Development

In the development phase, the tool is developed to meet the requirements of the preliminary design. Techniques used for implementation can vary depending on the type of artefact to implement. An algorithm can for example require construction to be made to test its correctness, for example, prediction accuracy. As the research process develops new knowledge and insights it entails circumscriptions of the proposal or the preliminary design (Vaishnavi & Kuechler, 2019).

The Development phase of the thesis was the most time-consuming since it included multiple stages. The major stages were data cleaning and data transformation, examining optimal variables, applying machine learning algorithms and moreover improving the same algorithms with hypertuning. All the stages consisted of programming using Python. Each major step of the data processing is covered in a separate heading explained more in detail (section 5, Data processing). Figure 4.3 illustrate the work and data flow of the method.

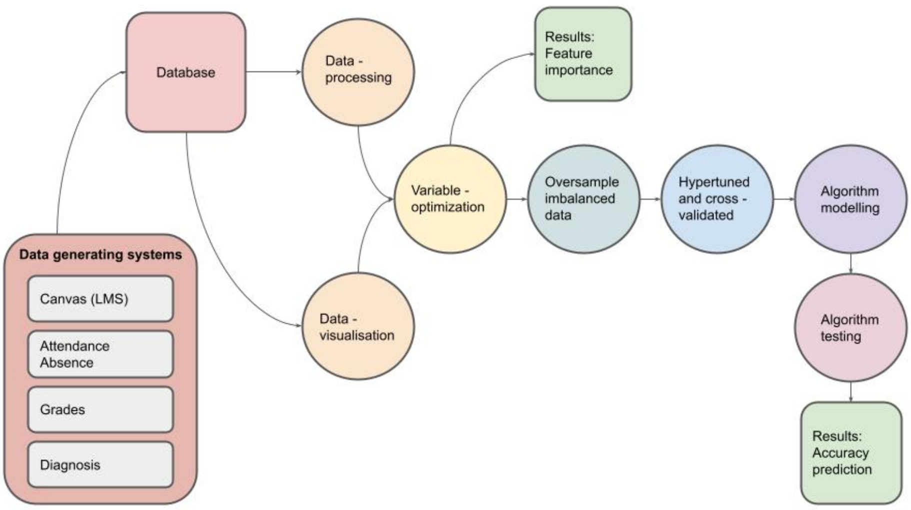

Figure 4.3: *Illustration of the work and data flow of the method. From data in systems to answers of the two research questions. The circle shape illustrate programming in Python.*

The programming fulfilled was inspired by KTHs guidelines of Pair programming to optimize the programming sessions but applied differently. KTH describe the pair programming process as using one driver (the person coding) and one navigator (the person observing and supposedly detects mistakes)

(KTH, n.d.). In this thesis, the two drivers instead programmed at the same time and then compared their code (navigator discussion) to explore improvements and nuanced opportunities. Continous testing of the code was done and the progress was continuously updated on a shared GitHub repository. Hence both partners had the most updated code accessible. Moreover, the code can be found at (GitHub, 2022).

Before training the Random Forest algorithms, the binary classifiers had to be decided. The options were a classification between grade F and grade E or the classification of F against all passing grades, namely (E, D, C, B, A). Depending on what the binary classification is it results in different magnitudes of data samples. In the thesis, the F and E binary classifier was chosen eventhough it resulted in fewer data samples. The reasoning was, by uniting six classifiers into one the variance is increased (Hastie et al., 2009). Moreover, the imbalance of data would be further increased. A final benefit of the decided binary classifier is that the unused data samples (D, C, B, A) can be applied to test the specificity of the algorithms.

As stated in section 4.2.2, Suggestion, a selection of algorithms to use resulted in Random Forest as the optimal choice based on a review of earlier studies where Random Forest was the most prominent algorithm. Moreover, Random Forest is an algorithm less abstract which decreases the technical gap between developers and stakeholders. The Random Forest algorithm was applied from the SciKit library in Python (Scikit-learn A, n.d.). The algorithm was employed in four sequential instances through the development of the final algorithm. The available variables were compared and the four most prominent were selected by variable importance. variable importance is a function that uses Gini index to determine which variables in the trees are the most efficient at sorting data. The function used was likewise extracted from the SciKit library (Scikit-learn B, n.d.). In the subsequent instances, the top four variables were employed. In the third instance, the SMOTE-algorithm was applied to enhance the imbalanced data (Imbalanced-learn, n.d.). Lastly, hypertuning and cross-validation of the variables were done to improve the algorithm further. The functions applied to combine hypertuning and cross-validation originates from the SciKit library (Scikit-learn C, n.d.).

4.2.4 Evaluation

Once the artefact is constructed it is to be evaluated according to Design Science Research. The evaluation is made aligned with the proposal defined in

Awareness of the problem. The additional information gained in the development phase and testing of artefact (algorithm) is summarised. The summarisation is concluded with a circumscription to improve the artefact further. The results of evaluation often imply recommendations for new designs or variables. But it can similarly consist of the insight of additional research that is necessary and thereafter appended to the research project (Vaishnavi & Kuechler, 2019).

Evaluation and circumscriptions are two concepts that share common ground. As interpreted in this thesis circumscriptions are the action to keep improving the algorithm based on conclusions in any phase. Namely circumscriptions were made in phases before the evaluation, but also during this phase. Examples were that Artifical Neural Network would not work as an algorithm due to the small amount of data or hypertuning the algorithm would improve it further. Both the continuous dialogues with teachers, (section 4.2.1, Awareness of the problem) and thesis mentors who provided feedback and opinions resulted in circumscriptions.

When there was a final result, i.e an algorithm and tendencies proved as dominant indicators the evaluation phase was initialised. In early April a meeting was conducted with two teachers and the It-pedagogue of the case study school. During the meeting the results of the study were discussed. Moreover, the dialogues covered improvements the school could establish and necessary adjustments for a large scale Learning Analytics project to be a reality. Many improvements are similarly mentioned in chapter 7, Discussion, where the results are processed in a nuanced way.

4.2.5 Conclusion

The concluding phase is described as the finale of a certain research effort or the end of a research cycle (Vaishnavi & Kuechler, 2019). The phase is initiated when the product is deemed to be functioning adequately. Furthermore, the phase summarises the results and documents the knowledge gained. The knowledge gained is divided into two subcategories: firm and loose ends. The firm category includes facts that have been learned during the process that can be applied in future projects. The loose ends category consists of the results that need further explanation and poses a need for future research (Vaishnavi & Kuechler, 2019). In chapter 8, Conclusion the firm and loose ends knowledge is outlined. However in this study, they are referred to as Summary and Further work.

4.3 Research Ethics

An important aspect of research ethics concerns questions such as how participants should be treated (Vetenskapsrådet A, 2017). Moreover, the article mentions there is a necessity to protect the individual as much as possible. In this thesis research ethics concern both how the pilot study was conducted and the data ethics of how data was processed.

According to Vetenskapsrådet B (2021) there are ethical considerations that have to be taken into account to achieve a balance. The organisation mentions research has to be trustworthy, respectful, honest and responsible. For example, an important aspect when an interview is conducted is to assure anonymity (Vetenskapsrådet A, 2017). For the thesis the interviews were therefore anonymised. Moreover were the interviews recorded to be further analysed afterwards, with the purpose to achieve trustworthiness. Before initiating the interview, the participants were asked to provide their oral consent to recording the interview. The recordings were deleted when the thesis was finished. Transcriptions of the interviews were made and are presented in Appendix A, Transcription of interviews.

Learning Analytics bring further concern to research ethics, namely data ethics, which is not fully covered by Vetenskapsrådet. Thus the data ethics approach for the thesis is based on the article by Sclater (2016) and guidelines according to GDPR, see also (section, 2.4.1 Data ethics). In the article by Sclater (2016), eight areas of ethics concerns are mentioned regarding data. They are responsibility, transparency, consent, privacy, validity, access, minimising adverse impacts and stewardship. These aspects were considered during the study and are further evaluated in section 7.6, Review of Ethics. For example, the data was provided from the case study school on two seperate encrypted USB flash drives and the students were anonymised by an id. When the research project was finalised the flash drives were given back. Before the analysis of the data could be conducted there was general consent given by an attorney, interviewee (Appendix A2). To clarify, the validity of the results is discussed in section 7.3, Reliability and validity, while the validity of data is discussed in section 7.6, Review of Ethics.

Chapter 5

Data processing

Before applying the Random Forest algorithm to the various courses data processing had to be performed. This chapter describes the process of how the data was handled, the enhancement of certain variables and how variables were selected for the algorithm. Furthermore, the chapter presents which courses were selected together for this study with a motivation of the selection.

5.1 Data

The data provided from the case study school included attendance, grades, diagnosis results, national test grades, Learning Management System data and teacher predictions from all the courses given at the case study school. The data is stored on various databases at the school. The entire set of data was provided by the school and the students were anonymised by identifying each student with a random code. In total, the number of students was approximately 700. However, all data points were not gathered of all students, which means that the number of data samples varied depending on which course were examined.

5.2 Collection of data

The data was processed by an employee of the case study school and supplied as an encrypted SQL database. The database was split into different tables depending on the origin of the data, see section 1.4.5 and figure 4.3 for more information on the partition of data.

Regarding Learning Management System data, the earliest records were from 2019 and the latest were from the end of 2021. Grades were available for students enrolled in the school as of 2017. The grades data are added consecutively. Thus students enrolled in their third year provide more data samples since they have taken more courses. 'Diagnosis' data were available for all students enrolled between 2017 and 2021. The diagnoses consisted of the students' knowledge in mathematics, Swedish and English. The diagnoses are performed by the students when they are enrolled in the school. The data about student Absence from the lessons was available from the semester 2016/2017 and up until the end of 2021.

5.3 Selection of courses

An important aspect when choosing courses to examine was to confirm there were enough data of students with a failed grade (F). Figure 5.1 illustrates in which courses the magnitude of F students are the most extensive. The y-axis displays the courses abbreviations in Swedish and the x-axis number of students with F. The total number of students for each course was roughly 700. Another aspect when choosing courses was the hypothesis that some variables had a stronger correlation to a specific course. For example, the score on mathematics diagnosis data has a stronger correlation with the mathematics grade and grades in Mathematics 1 would be a beneficial variable for Mathematics 3 and Mathematics 4. The two aspects resulted in examining the eight following courses; Physics 1, Chemistry 1, English 5, English 7, Mathematics 1, Mathematics 3, Mathematics 4 and Technology 1.

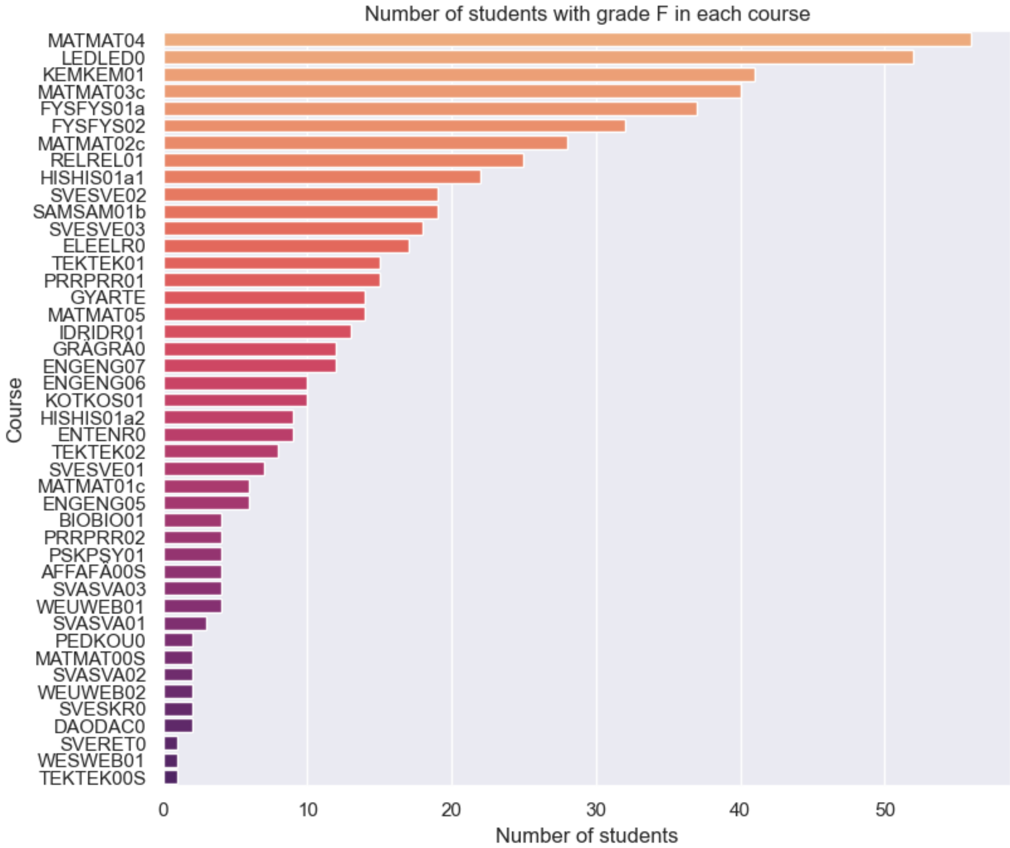

Figure 5.1: *The number of students with grade F for all courses.*

5.4 Selection of date for Learning Management System variables

Aforementioned in delimitations, an early ambition for the study was to examine if the prediction of the algorithms would improve over the duration of the semester. Due to limited data points that were updated over time, this goal was considered to be challenging to achieve. Therefore, the Learning Management System variables were only collected once for each course. The chosen time was adjacent to the first performance review if the course was only given during one term. In case the courses were given over the entire semester, the date was instead chosen adjacent to the second performance review.

5.5 Variable enhancement

Teachers use the Learning Management System platform differently depending on course and class. Mentioned by a teacher at the case study school it implies a non-correct comparison between data points. It would be more accurate to create a variable defining the student's Learning Management System activity compared to the class average for each student. With classes consisting of approximately 30 students, it would be appropriate to use the median measure. The median is better applied for smaller sample groups and is not affected by extreme values to the same extent (Denscombe, 2017, p.354). Thus based on the Canvas data and preprocessing the variables page view factor, participation factor, missing factor, on-time factor and late factor were constructed from the original data from Canvas. The math behind the page view factor variables was constructed as the fraction of the unique student's number of page views divided by the class median page views. This was repeated for all of the five variables.

5.6 Selection of variables

Parallel with data transformation and data cleaning a screening of variables was performed. The selection of the final variables consisted of two steps. Firstly, removing variables with low magnitude. Thereafter removing variables that proved to have low importance for the algorithm. The reasoning was that processing of the optimal variables improves the results of an algorithm.

5.6.1 Removal of samples with missing data points

In the first step, the purpose was to detect variables with a great magnitude of missing data points. The variables were removed since data samples (rows of student data) was prefered to not miss data points. Moreover, data samples were removed in correlation to what SQL table the data was read from. The different tables are illustrated with specific colours. This is due to results illustrating both the accuracy of an algorithm with and without Learning Management Systems (Canvas) data. Figure 5.2 illustrate variables before any were removed. To note is that English total was the only English variable kept since it is a summarisation of the three variables English Vocabulary, English grammar and English reading.

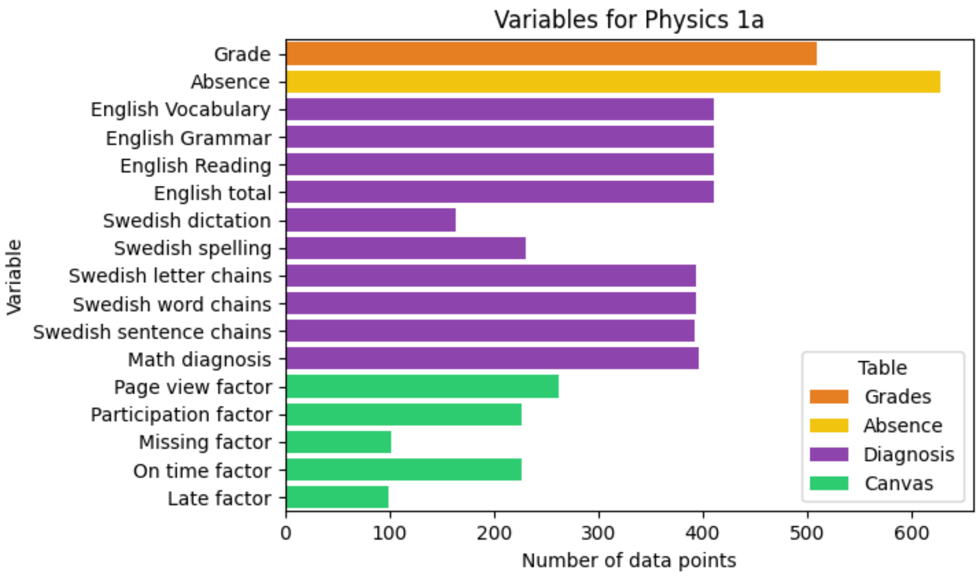

Figure 5.2: *Number of students for each variable in Physics 1a*

In addition, the diagnosis Swedish dictation and Swedish spelling contained considerably fewer data points than the other diagnosis and was therefore removed. A remark is that Canvas data for different courses vary and therefore would the remaining variables differ for every course. In Physics 1a were for example the variables Swedish dictation, Swedish spelling, Missing factor and Late factor removed before applying the algorithm (see figure 5.2). When merging data from different tables the outcome is a decrease in samples. Moreover, optimisations like using the binary classification E-F and applying the SMOTE-algorithm, (see section 4.2.3, Development) change the number of data samples. Therefore the number of data samples both decreases and increases during the work process. Figure 5.3 illustrates how the number of data samples varied in the course Physics 1a depending on which variables that were used in the algorithm. The figure illustrate how the number of data samples decrease when different tables are combined. For example, the total number of data samples decrease by one when the tables Grades and Absence are combined. But when Canvas data is applied to the other tables the total number of data samples drops from 386 to 203.

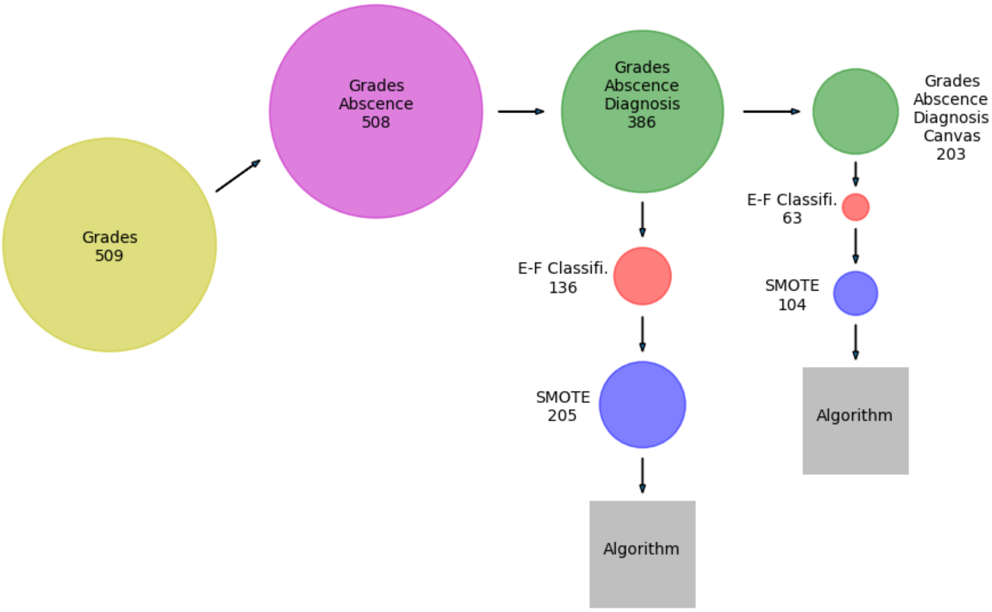

Figure 5.3: *Visualisation of how the number of data samples varied during the work process, due to merging tables and applying SMOTE*

5.6.2 Removal of samples due to variable importance

In the second step, the remaining variables were applied to the algorithm. Thereafter was the function variable importance of the Random Forest algorithm from the Scikit-learn library used (Scikit-learn B, n.d.). The function returns the variable importance for all variables. The top four variables were chosen due to the risks of high dimensionality in the algorithms. The variables' variable importance for Physics 1a is illustrated in figure 5.4. The figure shows the variable importance for the variables in relation to each other on a percentage scale. For example, math diagnosis has a value around 0,55 coresponding to being the most important variable with an importance of 55% for the whole algorithm.

The same procedure was repeated for all courses examined.

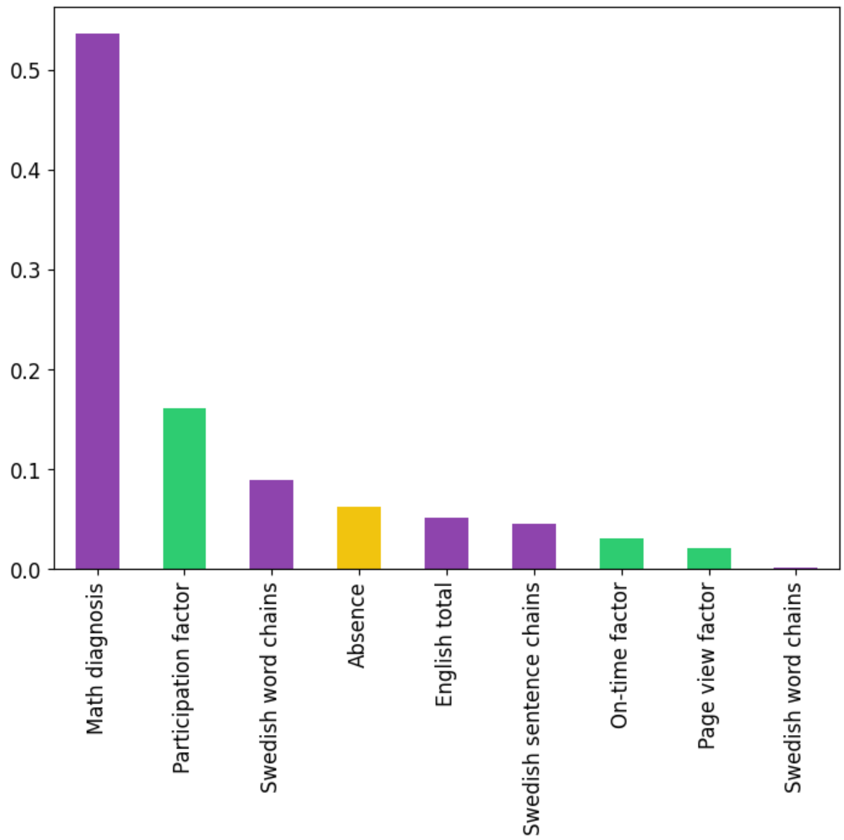

Figure 5.4: *Sorted variable importance for Physics 1a. The different colours indicate the original database table the variable was selected from. (yellow = absence, purple = diagnoses, orange = grades and green = Canvas - Learning Management System.)*

Chapter 6

Results and Analysis

This Chapter is divided into two sections where each section addresses one research question respectively. Thus, the chapter includes the result of prominent tendencies and a presentation of the achieved Random Forest algorithms and further an analysis.

6.1 Tendencies with most dominant indicators for students not passing a course

The section presents the answers to the first research question: What tendencies are dominant indicators for upper secondary school students not passing a course? Figures 6.1 and 6.2 describe the average variable importance for the courses examined (Physics 1a, Chemistry 1, English 5, English 7, Mathematics 1c, Mathematics 3c, Mathematics 4 and Technology 1). The y-axis represent the percentage of importance the variable had in relation to the other variables used when applied in the Random Forest algorithm. All the variables together add up to 1. The mean value for each variable is calculated over the courses that it was used for and is then referred to as the average variable importance. Few data points can imply misleading variable importance and therefore Mathematics 1 was removed from all plots. The same reasoning was applied for English 5 and English 7 when they were removed before the average was illustrated.

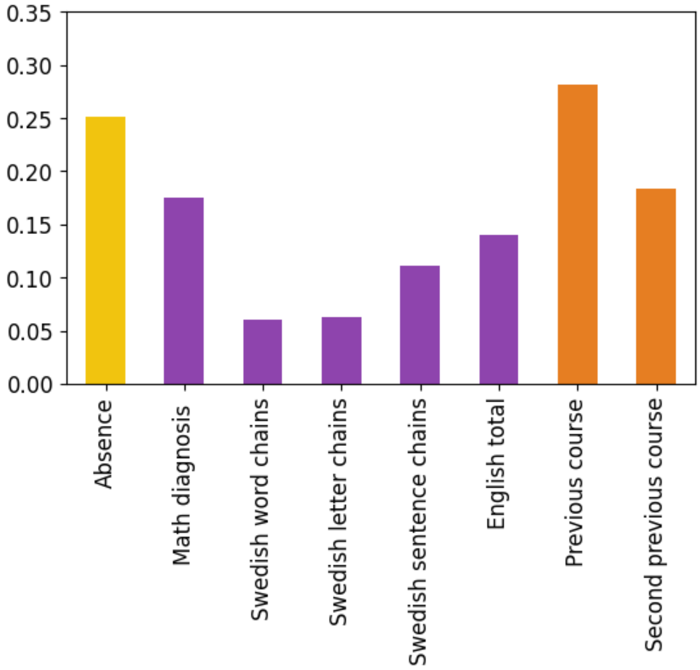

Figure 6.1: *The average variable importance as mean across all courses. Note; The orange bars are calculated as the mean of the courses Mathematics 3, 4 and English 7. The different colours indicate the original database table the variable was selected from. (yellow = absence, purple = diagnoses, orange = grades and green = Canvas - Learning Management System.)*

The courses English 7, Mathematics 3 and Mathematics 4 provided the possibility to examine if grades in previous courses are important indicators. When observing figure 6.1, It is evident that the student's grades from previous courses in the subject is the variable with the highest importance for the algorithms used for these courses. Figure 6.1 also shows that for the courses given during the first year, Absence is the single most dominant variable. Math diagnosis is the second most important variable and English total is third. An important trend is that the variables, Swedish word chains, Swedish letter chains and Swedish sentence chains shows the worst results. One reason could be that knowledge of Swedish pays less dividend compared to other variables. Another reason could be that there are three Swedish diagnosis variables while the other core subjects English and Mathematics are combined into one variable. Furthermore, the diagnosis variables and previous grades are variables available at the start of the course which is promising for a system supposed to do an early prediction. This is compared to Absence which has data registered as the course is given.

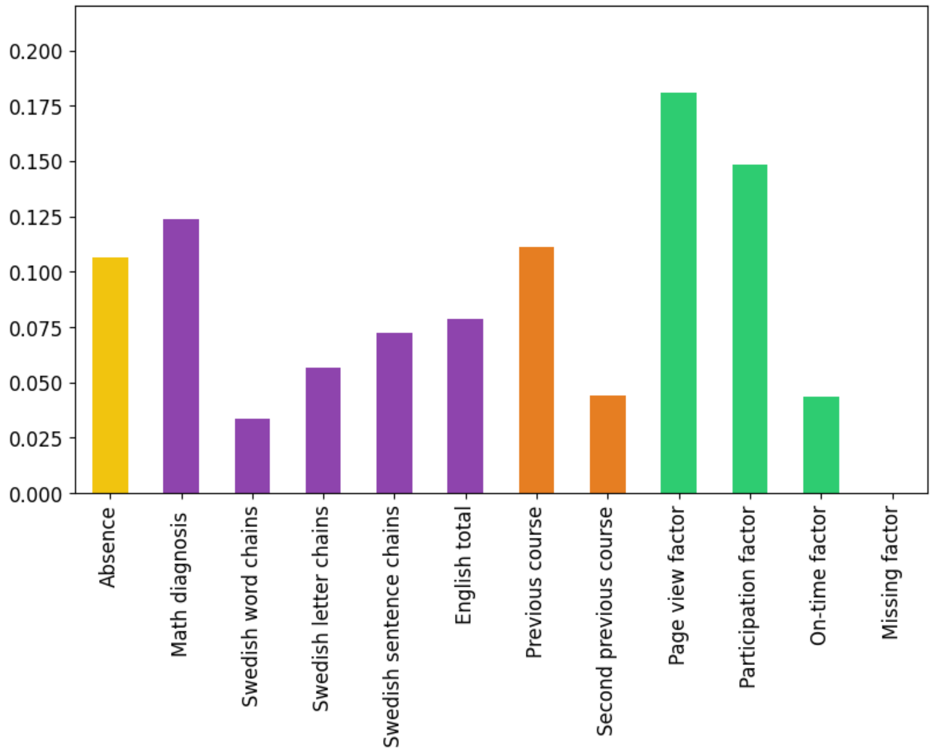

Figure 6.2: *The average variable importance as mean across all courses, with Learning Management System variables. The different colours indicate the original database table the variable was selected from. (yellow = absence, purple = diagnoses, orange = grades and green = Canvas - Learning Management System.)*

Figure 6.2 describes the variable importance when Canvas data was included in the Random Forest algorithm. As in figure 6.1, the y-axis displays the portion in percentage of importance the variable had in relation to the other variables used. The variables, summed up is the total of 1 (100 %). The mean value for each variable has been calculated for the courses that the Random Forest was trained for. The variable importance of the Page view factor and Participation factor is the highest indicating they are valuable variables for training the Random Forest algorithm. Note that the variable Participation factor was only used in three classifications. Comparing figure 6.1 with figure 6.2 many of the variables follow the same trend even though Learning Management System variables now are included. However, when Learning Management System variables are used, Absence is less important as a variable. One theory is that high Absence might imply you spend more time on the Learning Management System platform.

Table 6.1 is meant to complement figure 6.1 and 6.2 by illustrating the variable importance uniquely for two courses, namely Physics 1a and Chemistry 1. A table with the variable importance for all the courses can be found in Appendix B. The table present the percentage of variable importance for each variable in relation to the other variables. The variable importances marked in bold are the ones selected to be used in the final version of the algorithm.

Table 6.1: *Variables choosen for algorithms with Learning Management System data*

variable \ Course	Physics 1a	Chemistry 1
Absence	**6,3%**	**18%**
Math diagnosis	**54%**	11%
Swedish letter chains	0,17%	0,75%
Swedish word chains	**9,0%**	1,3%
Swedish sentence chains	4,6%	**12%**
English total	5,1%	5,0%
Previous course	-	-
Second previous course	-	-
Page view factor	2,1%	**34%**
Participation factor	**16%**	**15%**
On-time factor	3,1%	2,7%
Missing factor	-	-

By comparing the two courses Physics 1a and Chemistry 1 further trends are illustrated. In Physics 1a, Math diagnosis is the most dominant variable. Physics 1a is an abstract course which could be one reason. However, more essential is to question, why is it so high? It might be due to the lack of data and therefore the mathematics diagnosis is overfitted in relation to other variables. In Chemistry 1 the variable Participation factor has similar variable importance compared to Physics 1a. However, the Page view factor differs a lot. A final trend is that Absence constitutes a larger role as a variable in Chemistry 1.

Revised that the variable Math diagnosis had been overfitted may affect the analysis of comparing the variable importance since it hinders the other variables to be correctly represented.
An important remark to mention is that the importance of the variables for every course is different (see Appendix B).

6.2 The possibilities to identify students not about to pass a course

The section aims to answer Research question two: To what extent are machine learning algorithms possible to use to identify a student who will not pass a course? Results and plots from the used algorithms are displayed together with an analysis.

Table 6.2 and Table 6.3 illustrates the accuracy, specificity and sensitivity of the binary classifier, Random Forest algorithm, for Physics 1a and Chemistry 1, both with and without Learning Management System data. (See Appendix B for all courses). The value for each meassure is between 0 and 1 where 1 imply the algorithm classified all students correct and thus is the optimal value. The tables also shows the number of data samples (Students) used while testing the algorithm and in detail the distribution of students with grade E and grade F respectively, labeled as test data category.

Table 6.2: *Prediction accuracy, specificity and sensitivity for algorithms without Learning Management System data for the courses Physics 1a and Chemistry 1*

Course	Meassure	Value	Test data category	Students
	Accuracy	0,74	Total	31
Physics 1a	Specificity	0,78	Grade E	27
	Sensitivity	0,5	Grade F	4
	Accuracy	0,68	Total	19
Chemistry 1	Specificity	0,86	Grade E	14
	Sensitivity	0,2	Grade F	5

Table 6.3: *Prediction accuracy, specificity and sensitivity for algorithms with Learning Management System data for the course Physics 1a and Chemistry 1*

Course	Meassure	Value	Test data category	Students
	Accuracy	0,85	Total	13
Physics 1a	Specificity	1	Grade E	11
	Sensitivity	0	Grade F	2
	Accuracy	0,75	Total	12
Chemistry 1	Specificity	0,71	Grade E	7
	Sensitivity	0,8	Grade F	5

Prefered is an Accuracy above 0.5. An accuracy equal to 0.5 indicates a probability equal to guessing randomly. Thus by examining Table 6.2 and Table 6.3 the Random Forest classifier for Chemistry 1 without Learning Management System data can be questioned if it even should be applied, since the meassurements are poor. However, other Random Forest classifiers perform better (See Appendix B). The classifier for Physics 1a does however have poor sensitivity in both cases.

Apart from Chemistry 1 with Learning Management System data, all other sensitivities are equal to or below 0,5 in table 6.2 and 6.3. This implies that the majority of students failing the course are not detected by the system. However, for Chemistry 1 the sensitivity is 0,8 due to detecting 4 out of 5 students correctly.

Figure 6.3 and figure 6.4 are confusion matrices that visualise the same data presented in tables 6.2 and 6.3. The top-left position of each matrix represents the number of students with F classified correctly (that is with failing). The top right position corresponds to students with F classified as passing the course and thus the most critical position. The bottom left position is the number of students who passed the course but were classified as failing and the bottom right corner position represent students passing the course and classified correctly (that is with passing).

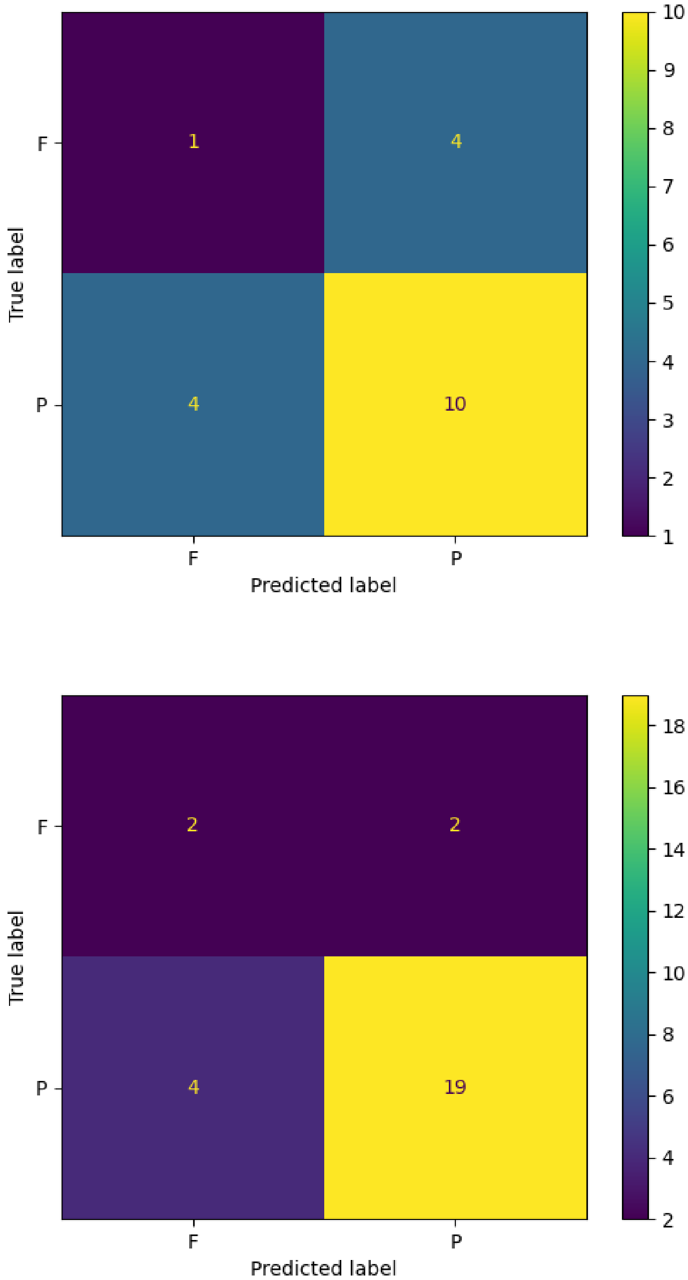

Figure 6.3: *Confusion matrices without Learning Management System data. The top matrix is Chemistry 1, the bottom matrix is Physics 1a. The numbers in each square represents the number of students in each category (true positive, false positive, true negative, false negative).*

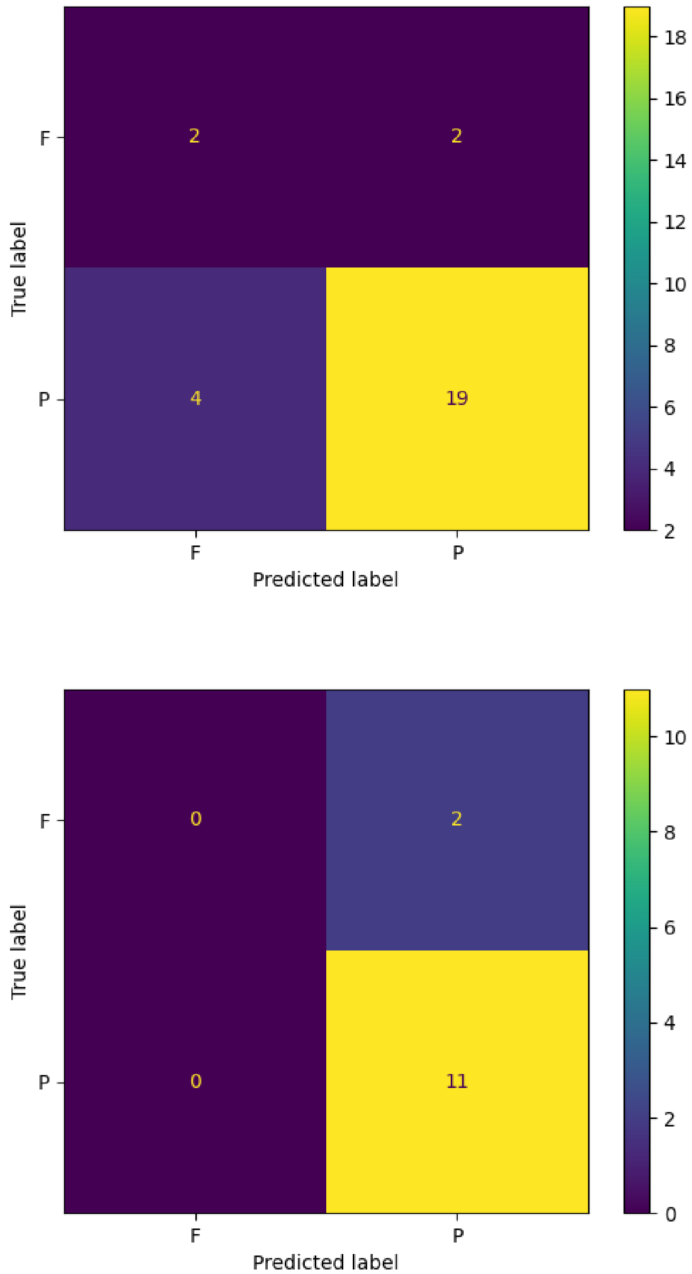

Figure 6.4: *Confusion matrices with Learning Management System data. The top matrix is Chemistry 1, the bottom matrix is Physics 1a. The numbers in each square represents the number of students in each category (true positive, false positive, true negative, false negative).*

Figure 6.5 displays a confusion matrix of the students with grades D, C, B and A (i.e all students passed). The columns named P represent students classified as passing the course. Columns named F represent students classified as not passing the course. The confusion matrix is similar to the ones in figures 6.3 and 6.4 with the exception that there are no students who received F since the data samples only contained students with grades D-A.

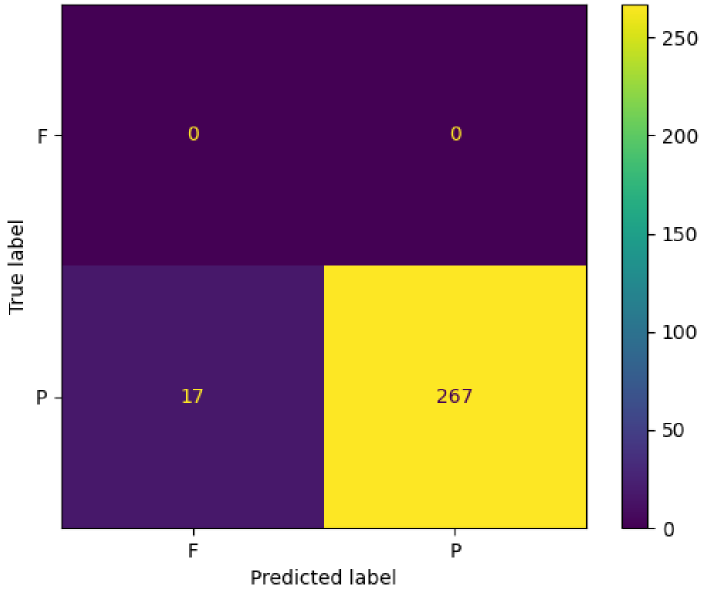

Figure 6.5: *Students with grades A, B, C, D in Chemistry 1 and whether they got classified as passing or failing.*

While examining the confusion matrix in figure 6.5, 17 of the total students (283) in Chemistry 1 with grades A, B, C and D got classified wrong, namely as they would fail the course. That is equal to a specificity of 94 %, which can be argued to be very good. It is better than the specificity of just E students. That gives a tell about the higher the grade is, there exists a correlation in how the classifier works. This is visualised further in the plot (figure 6.6). Figure 6.6 describe how many students of A, B, C, D got classified with pass or fail where the y-axis is the number of students. Students classified with F (fail) is thus classified wrong.

Since the plot illustrates a concise decrease in the number of students getting classified as failing the course when the student has a higher degree it could be argued indicating possibilties for Early Warning Systems to be applied.

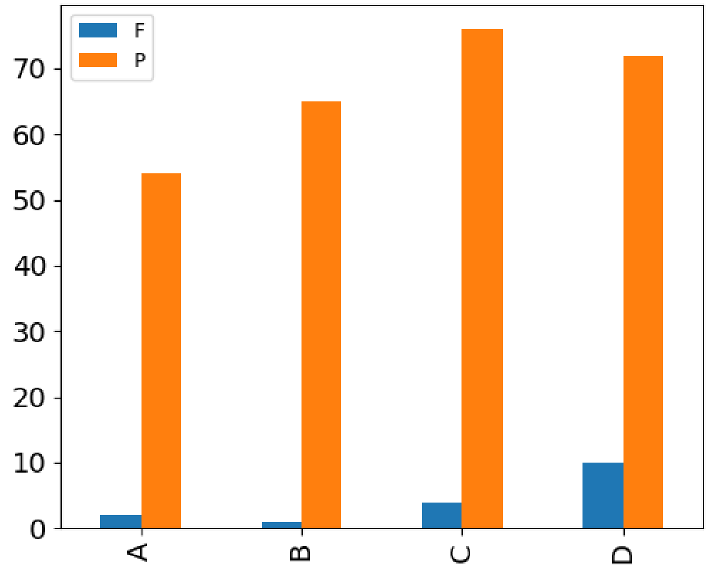

Figure 6.6: *Whether students with grade A, B, C, D got classifed with passing or failing.*

Final remarks are that all four classifiers present an accuracy above 0.6 but the sensitivity can be questioned. Even though the trends exist, the small number of data can be misleading. A decreasing number of students getting classified wrongly when examining students with higher grades does however indicate a potential for the classifier. Therefore the possibilities to identify students not passing a course are promising. However, there are many improvements that can be implemented that would imply an even better algorithm for doing so, see chapter 7, Discussion.

Chapter 7

Discussion

In this chapter, the analysis and results of the research questions are discussed, including the aspects of reliability and validity. Furthermore, the results are discussed from the didactic framework, that is how formative assessment and the Pygmalion effect are correlated to a potential application of Early Warning Systems. Lastly, the thesis is evaluated from an ethical point of view.

7.1 What tendencies are dominant indicators for upper secondary school students not passing a course?

The results of Research Question 1 presented in section 6.1, Tendencies with most dominant indicators for students not passing a course, provide answers to what tendencies are the most significant even though the data points are few. An advantage of Random Forest is that it can handle overfitting well. But, due to the number of data points being minimal the variable importance method has an increased risk of overfitting a certain variable. Thus the safety caution of choosing the top four variables was applied, to make sure Random Forest did not overfit certain variables due to over dimensionality (each variable is the axis in one dimension). Nonetheless, this occurred, for example, the variable Mathematics diagnosis reached above 50 % variable importance for Physics 1a, while the number of F students was only four. This led to the variable not being taken into account while presenting the average variable importance. With more data, this would be less of a problem. In figure 6.1 it is presented that diagnosis variables of Swedish are poor (Swedish word chains, Swedish letter chains, Swedish sentence chains). One could therefore propose the variables are unnecessary.

In dialogues with a teacher at the case study school, a belief that the Swedish diagnoses would have a higher significance was expressed. This hypothesis was based on earlier experiences where reading comprehension in Swedish had shown to be essential in a number of courses apart from Swedish courses alone. A reflection is that the Swedish diagnosis variables instead would be combined into one variable, likewise English total, instead of having multiple Swedish diagnosis variables. Furthermore, the teacher explained that the variable Swedish letter chains are used as a calibration for how to interpret the results of the other diagnosis. Therefore, if a single Swedish diagnosis variable were to be merged, the difference between the Swedish letter chain and the other two diagnoses should be taken into consideration. It was further discussed that a measure of the students reading comprehension in Swedish could be used as a variable in the algorithm.

The Learning Management System variables have a great potential of being used as a classifier based on the variable importance in figure 6.2. First, the variable importance Page view is the highest compared to other variable importances. However, the result and analysis is based upon three courses only and thus the analysis should be done with reservations. In the Chemistry 1a course, the teachers at the case study school apply Learning Management System a great amount according to a teacher at the case study school. Moreover is the variable importance of Learning Management System variables based data on Chemistry 1a great. Therefore correlations can be concluded between applying Learning Management System on a greater scale as a teacher and improved variables importance. Thus if Learning Management System is applied well it creates data points and variables that can be used for classification. The possibility to use additional variables from Canvas to further improve the algorithm was discussed with the teachers. A variable that could be possible to analyse is the time elapsed from when an assignment is posted on Canvas until the students begin working on it. Another variable to analyse could be how much time elapsed between the student opening the assignment and handing it in. According to the teachers, such information would be obtainable given that the assignments are created on Canvas with explicit starting dates and deadlines.

If the variables were to be split into two types. i.e, not time-dependent variables, diagnosis and previous grades and time-dependent variables, Learning

Management System variables and Absence, remarks are concluded. Throughout a course, more Learning Management System and Absence data is generated as time pass. Thus a hypothesis is that time-dependent variables' significance would increase over time. Therefore diagnosis and previous grades variable are overrepresented at earlier stages of an analysis of all the variables. A further remark is in the data used in the thesis, the Absence data are merged together in the database and thus do not illustrate the change over the span of the course. A recommendation is therefore that data should be stored continuously, for example, every week if potential change of variable importance would be analysed for the variable Absence.

7.2 To what extent are machine learning algorithms possible to use to identify a student who will not pass a course?

Considering Research question 2, it is evident that all accuracies in appendix B range from 0,6 to 0,9. Hence all algorithms provide a better prediction than random guessing would and proves potential for a working Early Warning System. However, the accuracies might still be considered too low for a finalised Early Warning System since many students would be classified incorrectly. In earlier studies, Early Warning System have proven to have much higher accuracies, 96.1 %, (Miguéis et al., 2018), 95 % (Chung & Lee, 2019) and 92 % (Costa et al., 2017)

Considering the sensitivities (section 6.2 and appendix B) they range from 0 to 1. An explanation for the large variation is the low amount of data samples with F used in the testing process. For the algorithm to be used in a finalised Early Warning System, a consistent sensitivity would be optimal to ensure the systems dependability. Compared to Chung and Lee (2019) whose system had a sensitivity of 0.85, (specificity of 0.95) the sensitivity is poor. The specificities for the algorithms are more consistent, ranging between 0,7-0,9. Thus, providing higher dependability. The reason might be that the data sets for testing contain a larger portion of students with E than F. This further strengthens the hypothesis that more data points would improve the sensitivity. This could further be identified in earlier applied systems, i.e Course Signals (section 2.2.2, Examples of Early Warning Systems) which proved to have an Early Warning System beneficial for their students.

Even though the quality of a supposed Early Warning System applied today is questionable there are reasons to continue the research.

The sensitivity of the classifier could be considered more important than the specificity in this case. If students who are failing a course are classified as passing, it results in them not being detected and given guidance. In contrast, specificity is not of the same importance since it results in students who are about to pass being classified as failing. This is not as problematic since classifying these students might result in them getting more support. However, it results in more work for the teacher, which is not ideal since the intention of the system is to reduce the teachers' already high workload. A possible approach to improving the sensitivity would be to weight the algorithm in favour of sensitivity. If this is possible, it would mean that the algorithm prioritises sensitivity over specificity when being constructed.

7.3 Error Sources

A large source of error in the thesis is the limitation of data points used in the algorithm. A greater amount of data points would possibly mean that the results of the algorithm would be more stable and less sensitive to changes. For example, in the algorithm for Physics 1a with Canvas data, there were only two students with F in the test data.

The small amount of data points results in a high risk of errors in both the variable importance and the three measurements accuracy, specificity and sensitivity. For the variable importance, variations of which data points are used as train and test data could lead to a difference in the variable importance. With a larger amount of data, such differences would most likely decrease.

For the accuracy, specificity and sensitivity the same uncertainty applies. Small variations of the algorithm or the data sets for train and test data could result in large changes in the values of all three measurements.

7.4 Reliability and Validity

The Random Forest algorithm is built on statistical probability which means that the algorithm changes each time it is applied with different data. Simply one data sample different would change the logic of the Machine Learning algorithm. There is moreover a randomised train-test split function built within Python, that split the total amount of data samples into 80% training data and

20% testing data. If the data split would be completely randomised next time different data would be used to both train and test the algorithm. The prediction accuracy, sensitivity and specificity would therefore be different if the next person tried to do the same experiment with a randomised train-test split, which increases the variance and therefore decrease the reliability of the study. However, Python does have a parameter called random state which could be set to the same number and result in the same test-train split. But it is not a scientific method. From a statistical perspective, an increase in the number of data points would imply that results become more equal to each other and therefore a lower variance. Thus more data points would increase the reliability of the study. Cross-validation was applied to find the best hyperparameters. The downside of using cross-validation is however that no final algorithm to be used within an Early Warning System is created.

Research question one is visualised by a plot of which variables are the most important for the Random Forest algorithm (see figure. 6.1 and 6.2). What is supposed to be tested is tested which argues the validity is high. Research question one could have been answered by a different method. For example, other algorithms to illustrate what variables are important, like Support Vector Machine or simply a Decision Tree. However, since Random Forest was also applied in Research question two it seemed more logical to also apply Random Forest when testing variables. The most dominant variables could, however, be tested by using p-value or a combination of p-value and a variable importance function (Costa et al., 2017). An advantage of examining the p-value would have been that the dominant variables would have been illustrated outside the context of a Machine Learning algorithm. A disadvantage of mixing two measurements could imply increased complexity in the method.

Research question two was to examine to what extent it is possible to find the students not passing a grade by the help of Machine learning. In the analysis of Research question two, it was illustrated that students were found by applying a Random Forest algorithm. Thus Research question two tested what was supposed to be tested and proved good validity. However, the Research question can be further elaborated by examining different Machine Learning algorithms and comparing their outcomes. By testing more algorithms the validity would have increased. But due to delimitations of the time frame it was not possible in this study. In section 8.2, Future research, a suggestion is to examine if other predictive models are more accurate.

A final remark on reliability and validity is that this study did not have access to the diagnoses used as variables. This means that an evaluation of the diagnoses from this perspective is not possible to do. However, such an evaluation would be of interest in future research.

7.5 Didactic Review

This section aims to provide a discussion of the result and thesis as a whole from the didactic framework. The didactic concepts the discussion uses are formative assessment and Pygmalion effect in relation to performance reviews, Early Warning System and the results.

7.5.1 Formative assessment

In the literature review regarding formative assessment, it is stated that long cyclic feedback is seen as less efficient than medium and short cycles (Wiliam & Leahy, 2015). From this point of view, it is of interest to evaluate if the Early Warning System is best used in relation to performance reviews or if the system should be used in shorter cycles. One could argue that the limited amount of data that is updated continuously implies that the classification prediction would not change considerably over shorter periods. However, with an increased amount of data points from the Learning Management System and the Absence system continuously analogous implementations could be possible. In the current state, short cyclic feedback is not as adjustable for Learning Analytics since the outcome of the classifier might not change much over a week. Moreover, if an algorithm were applied every second week and would give the student different predictions without reasonable cause, trust in the system would arguably decrease.

As mentioned earlier, Early Warning System is based on the idea of being a tool the teacher can use to find the students to whom the formative assessment and feedback will be addressed. Both the Research questions provide help for the teacher if they would be further developed. Knowing what variable importance is most important could motivate the professional teacher to improve their Learning Management System content and their precision in the registration of Absence. Moreover, could the algorithms be a help for the teacher to provide substantiated feedback, where feedback is a part of formative assessment.

One of the results from the study of Course Signals was that the students became more proactive in dealing with their assignments when they received direct feedback in the Learning Management System (Arnold & Pistilli, 2012). In section 3.1.1, Formative assessment it is stated that a key strategy of formative assessment is to make the students independent and be responsible for their own learning. Therefore, it can be argued that Course Signals was beneficial in achieving student independence. Regarding this study, the possibilities for extended use of Learning Management System variables and a focus on adapting the system to shorter cycles of feedback could prove valuable in increasing student independence. The aforementioned possibility to extract data on when a student opens an assignment and hands it in would be interesting to evaluate from this aspect. Hence it offers an opportunity for shorter cycles of feedback in order for the student to engage in assignments earlier.

7.5.2 Performance reviews

One of the proposals of the thesis, (section 4.2.1, Awareness of the problem) aforementioned the algorithms were to be applied in conjunction with performance reviews. Moreover, in the pilot study, it is mentioned that performance reviews are held in October and in March, interviewee (Appendix A1). Thus, if algorithms were to be applied in conjunction with performance reviews, the student data (Learning Management System and Absence) has to be pipelined so that the data is available and applied for the algorithm. A different aspect to reflect upon is that the amount of Learning Management System and Absence data differ for courses at the time of a performance review. Mathematics courses are given across half a year while Physics 1a and Chemistry 1 are given over the duration of one year. Therefore it can be argued whether the algorithm should be applied for Chemistry 1 and Physics 1a during the first performance review and if so with or without Learning Management System and Absence data. An argument against this is that variables will show less significance and there is a higher risk of incorrect classification. A remark is that the performance review where mathematics data can be applied is based on half of the course content. Compared to the latest possible performance review for Chemistry 1 and Physics 1a being conducted after three-quarters of their courses.

Another aspect to mention of how to improve the algorithm would be to push the performance reviews further ahead. However, a reason for not pushing performance reviews forward is that it does not give the student enough time to change their behaviour. An optimal time frame could be investigated further by analysis of data. The algorithms can be further improved (section 8.2, Further Research). As of now, the algorithm is not connected to the software's used by the case study school, which is a must if an algorithm were to be used. Thus it can be stated that the proposal, algorithms were to be applied in conjunction with performance reviews, is fulfilled but there the aforementioned improvements can be conducted.

7.5.3 Risks of Implementation

During an interview with one of the stakeholders, the need for a program that does not put a negative label on the students was mentioned interviewee (Appendix A1). The stakeholder stated that a system potentially could provide a faulty interpretation of the student's progress. Related to this argumentation is the Pygmalion effect (Rosenthal & Jacobson, 1968). The Pygmalion effect argues that a Early Warning system could lead to the students that are labelled at risk of not passing a course being perceived this way by the teacher. Therefore, the expectations of the student could differ from the expectations of the other students in the same class.

With the Pygmalion effect taken into consideration, the short cyclic use of an Early Warning System can be further questioned. With a prediction early on in the course, labelling the student as either passing or at risk of failure could mean that the Pygmalion effect is reinforced. Therefore, it might be better for the teachers to receive the prediction later on in the course when they have already established an opinion of the students' progress.

If a Learning Analytics system such as Early Warning System would be applied, according to Tsai et al. (2021) there is a necessity to develop data literacy among key users, i.e teachers and school management. The interviewee (Appendix A2) mentioned that the technical understandability is higher than average at the case study school. However, Machine Learning and the algorithm Random Forest build upon complex mathematics. Moreover was one of the proposals for the algorithm that it would be technically understandable for teachers. Among algorithms, Random Forest was thus chosen as it decreases the gap of technical understandability between stakeholders. If Artifical Neural Network would be applied as an algorithm the technical gap would be bigger but the results could have perhaps been better. This is left as a concern for the stakeholders of the case study school to discuss regarding the technical understandability or performance of the algorithm is to prefer. However, it would be beneficial for the teacher to understand the holistic view of the Early Warning System. This could be achieved by initialising a workshop for teachers so they could learn the system from a technical point of view. It would be time-consuming, but it would also develop data literacy and therefore interesting to consider in future studies.

7.6 Data improvements

A necessity for Learning Analytics is to have enough data (Sousa et al., 2021). Learning Management System are often great since they provide large magnitudes of data, but the applicable data often gets reduced due to processing and errors in the data. The prediction accuracy of the algorithms presented in the results is inferior compared to the accuracy of earlier studies. However, there is a correlation that may depend on the number of data points. Both the studies of Chung and Lee (2019) and Miguéis et al. (2018) had great prediction accuracies while the number of data samples (students examined) were 12 000 respectively 2459. A higher number of data points is a trivial necessity to achieve a more accurate algorithm. A concern for the case study school is thus how the number of data points can be increased to provide a more accurate algorithm.

With time more data points will be generated. Therefore the algorithms will presumably improve over time. A possible way to accelerate the number of data points generated for the school is to increase the number of students. Demographic data could also be applied, but there are ethical concerns to take into consideration, see section 7.7, Review of Ethics. An option would be to conduct a study combining data points from different schools. From the understanding of a teacher at the case study school, collaboration with schools within Stockholm municipality would be complex, due to all schools not having access to all of their data. A further problem for collaboration to be possible is that schools use different systems. Not only Learning Management System but also administrative systems including grades, Absence and diagnosis data. A question to be answered would thus be if a comparison of data would be possible between schools using different systems. A promising directive is the technical guidelines to facilitate K-12 education which include recommendations for data gathering systems (Skolverket D, 2021). When schools have the same systems and data gathering process the data would be similar and could thus result in better algorithms.

7.7 Review of Ethics

The section accounts for the ethical aspects of the thesis, including a discussion on data ethics for the data points used in this thesis. In section 4.3, Research ethics, the 8 categories from the article written by Sclater (2016) are presented. In this section, the thesis is evaluated in relation to these categories.

Responsibility for the project both concerns the case study school and the authors of the thesis. However, the major responsibility of data processing is being done at the school. The data was handed over on an encrypted flash drive with the password sent by email which implies responsibility was considered. Thereafter the flash drive and laptops with access to the data were handled carefully. There could have been a problem if a laptop would have been stolen. Discussions to solve such a situation occurred but did not result in any decisions.

Transparency and consent concern to what extent students should be told about data collection, preprocessing and analytics. According to Sclater (2016) it is a strong recommendation this information is given to the students. The concept of access answers many of the same ethical questions. Regarding data gathering, each platform has a consent option which strengthens the student's ethical rights. However, to the knowledge gained from interviewees (Appendix A1, A2) there was no general consent or information given or asked students that data of them would be analysed. If it in the future would be the reality that algorithms were used to analyse students and Early Warning System would be applied to communicate with a student, general consent would be necessary.

European Union (2018) states that an individual has a legal right to know how the data points collected on them are used. Absence, diagnosis and grades are data points that is already used by teachers and schools to assess students during performance reviews for example. With data accessible from Learning Management Systems the topic of consent becomes more important. A question to raise is if the students or their legal guardians have to approve that Learning Management System data is used to assess their current progress in the course. A further aspect is that with Learning Management System data the student's behaviour outside school can be analysed since data from Learning Management System do not take into concern whether the student is sitting at home or in the school environment. The interviewee (Appendix A1) confirms this by stating, that students are used to getting assessed at school but not on their behaviour at home.

Privacy is an aspect that is important to make sure individuals are kept anonymous in the study. In the study, each student was only represented by a randomised id number and thus motivates that privacy was taken into consideration. Moreover, there was no demographic data about the student registered. However, demographic data could contribute to better tendencies and thus better algorithms. If demographic data would be applied in a further study it could transgress students' ethical rights.

Validity is an important aspect of what can be implemented in the field of Learning Analytics even though it stands in contrast to privacy, (Sclater, 2016). Bad validity in the data results in worse algorithms or more preprocessing has to be done. Much of the validity data ethics concern the organisation, i.e the case study school since all the data was provided by them. Parallel with examining the research questions, the thesis has resulted in insight (section 7.6, Data improvements) about data processing and data collection that can be transformed and would increase the validity of the data.

Minimising adverse impacts is a recommendation by Sclater (2016) for analysis since students and teachers may alter their actions when they know Learning Analytics is applied. The ethics concern the aspect of data analysis may affect a change in human behaviour and thus if Learning Analytics still should be applied. From the teacher's point of view is the Pygmalion effect a consideration. Meanwhile, students can change their behaviour on the Learning Management System if they know that the data collected is analysed. For example, could a behaviour such as a student entering Canvas more often be motivated by the student knowing Learning Management System data is registered. Student behaviours like these would decrease the trustworthiness of the data.

Stewardship of data is primarily a legal responsibility. Therefore, the thesis has very little effect or change in any legal decision. For the thesis, according to the interviewee (Appendix A2) all the legal concerns were solved before the data could be distributed.

Chapter 8

Conclusion

This project can be seen as the starting point for implementing Learning Analytics at the case study school and in particular the development of an Early Warning System. Moreover it could be expressed as a timely report for educational stakeholders at upper secondary level with a interest of increasing their knowledge in Learning Analytics. As a result, valuable facts have been learned and areas in need of further research have been established. These are disclosed in two sections, Summary and Future Research respectively.

8.1 Summary

Dominant tendencies for why students would not pass a course were found by sorting of variable importance. Absence, Previous grades and the Math diagnosis were dominant indicators that a student would fail a course. When Learning Management System data were included in the analysis and if teachers of the course used the system to a great amount then Page view and Participation factor were also proven to be good indicators. Moreover, variables closer in time to the application of the algorithm showed better results. Namely, Math diagnosis and previous grades decreased as a dominant indicator as the time passed.

The results prove that the developed algorithms could classify students moderately correctly which implies it is possible to use data to identify students that are failing a course. The accuracy for each algorithm was above 0,6 which indicates a positive tendency for classifying correctly. However, the accuracies are low in comparison to earlier studies that were successful, thus demonstrating a need for further enhancement. From a more holistic perspective, the development of a final Early Warning System could help students from failing courses and furthermore prevent them from dropping out. If an operating Early Warning System is constructed, it would be in alignment with The global goals to provide every adolescent with education by the year 2030.

8.2 Future Research

The section will briefly cover what the results of the thesis can build further upon or areas of research that independently can be examined. One of the proposals of the Design Science Research Methodology (section 4.2.1, Awareness of the problem) was the possibility for the algorithm to be further developed. With the code being uploaded and documented on GitHub (2022), (see section 4.2.3 for further details) there are possibilities to develop or improve the current algorithms.

During dialogues at the ending phase of the project with teachers at the case study school, they expressed an interest in further developing the Early Warning System in the future. As a result, proposals for additional master theses on the project will be added within the K-ULF project.

How an algorithm improves over time could be of interest to illustrate the variable importance of Learning Management System and Absence data. For this to be applied two improvements would be optimal. First, the Absence data has to be registered at least weekly. A side point is that with Learning Management System data only being applied for a number of dates a recommendation is that Learning Management System data only have to be registered weekly, decreasing the size of the SQL file by almost a fifth. Secondly, the magnitude and quality of the Learning Management System data have to be improved, see Section 7.6, Data improvements. By examining the variable importance over time it could be illustrative of how much the variables contribute and how much the algorithms are improved. A further result that could be presented is that of when it is appropriate to start training the algorithms with Learning Management System and Absence data.

An idea for further work is the proposal of using more algorithms to optimise the results. There are multiple classification algorithms and by no means Random Forest is the most optimal algorithm. Other classification algorithms that could have been applied were Supper Vector Machine, Naive Bayes and Artifical Neural Network. By testing multiple algorithms the best could have been chosen for the cause. However, a possibility is that different algorithms are best applied for different courses. If it in the future would be a reality to have a large scale Early Warning System applying different algorithms for different courses the system would have to be at a larger scale. Another option would be to compare all algorithms and choose the best one among the options and apply it to all courses.

The study by Davis et al. (2019) argues that the use of an Early Warning System alone might not be enough to help the students back on track. Instead, the schools need to have a strategy for how to use the information gained from the system to aid the students who are struggling. Therefore, if an Early Warning System were to be used in the future it would be necessary to develop a strategy for how the information should be used to assist the students by the professional teacher.

Bibliography

Aguiar, E., Lakkaraju, H., Bhanpuri, N., Miller, D., Yuhas, B., & Addison, K. L. (2015). Who, when, and why: A machine learning approach to prioritizing students at risk of not graduating high school on time. *LAK '15: Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, 93–102. https://doi.org/doi:10.1145/2723576.2723619

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at purdue: Using learning analytics to increase student success, 267–270. https://doi.org/https://doi.org/10.1145/2330601.2330666

Björndal, C. R. (2005). *Det värderande ögat*. LIBER.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Breiman, L., & Cutler, A. (2004, June 15). *Random forests* [Stat Berkeley]. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

Buckingham Shum, S., & Crick, R. D. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 92–101. https://doi.org/10.1145/2330601.2330629

Caulfield, M. (2013, February 12). *Purdue course signals data issue explainer* [eLiterate]. https://eliterate.us/purdue-course-signals-data-issue-explainer/

Chernick, M. R., & LaBudde, R. A. (2011). *An introduction to bootstrap methods with applications to r* (first editon). Wiley.

Chung, J. Y., & Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review*, *96*, 346–353. https://doi.org/https://doi.org/10.1016/j.childyouth.2018.11.030

Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, *73*, 247–256. https://doi.org/https://doi.org/10.1016/j.chb.2017.01.047

Davis, M. H., Iver, M. A. M., Balfanz, R. W., Stein, M. L., & Fox, J. H. (2019). Implementation of an early warning indicator and intervention system. *Preventing School Failure: Alternative Education for Children and Youth*, *63*(1), 77–88. https://doi.org/10.1080/1045988X.2018.1506977

Denscombe, M. (2017). *Forskningshandboken*. Studentlitteratur.

Dyckhoff, A., Zielke, D., Bültmann, M., Chatti, M., & Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Educational Technology and Society*, *15*.

European Union. (2018, May 25). *Art. 5 gdpr principles relating to processing of personal data*. European Commission. https://gdpr-info.eu/art-5-gdpr/

Feldstein, M. (2013, February 25). *Purdue university has an ethics problem* [eLiterate]. http://mfeldstein.wpengine.com/purdue-university-ethics-problem/

Folkhälsomyndigheten. (2021, January 21). *Gradvis återgång till närundervisning på gymnasiet*. https://www.folkhalsomyndigheten.se/nyheter-och-press/nyhetsarkiv/2021/januari/gradvis-atergang-till-narundervisning-pa-gymnasiet/

GitHub. (2022, May 26). *Master thesis - early warning system of students failing a course*. https://github.com/AlbinLundell/Exjobb

Global Goals. (n.d.). *Quality education*. Retrieved March 5, 2022, from https://www.globalgoals.org/goals/4-quality-education/

Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-smote: A new over-sampling method in imbalanced data sets learning. In D.-S. Huang, X.-P. Zhang, & G.-B. Huang (Eds.), *Advances in intelligent computing* (pp. 878–887). Springer Berlin Heidelberg.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Elements of statistical learning* (second editon). Springer.

Howell, J. A., Roberts, L. D., Seaman, K., & Gibson, D. C. (2018). Are we on our way to becoming a "helicopter university"? academics' views on learning analytics. *Technology, Knowledge and Learning*, *23*, 1–20. https://doi.org/10.1007/s10758-017-9329-9

Ifenthaler, D. (2021). Learning analytics for school and system management. *Pushing the Frontiers with Artificial Intelligence, Blockchain and Robotsl Conference on Learning Analytics And Knowledge*, 93–102. https://doi.org/https://doi.org/10.1787/d535b828-en

Imbalanced-learn. (n.d.). *Smote*. Retrieved March 17, 2022, from https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html

Khalil, M., & Ebner, M. (2015). A stem mooc for school children—whatdoes learning analytics tell us? *International Conference onInteractive Collaborative Learning (ICL) (Springer)*, 1217~1221. https://doi.org/doi:10.1109/ICL.2015.7318212

Konstenius, V., & Schillaci, M. (2010). *Skolfrånvaro-kbt-baserat kartläggnings-och åtgärdsarbete*. Studentlitteratur.

Kovanovic, V., Mazziotti, C., & Lodge, J. (2021). Learning analytics for primary and secondary schools. *Journal of Learning Analytics*, *8*.

KTH. (2022, January 24). *K-ulf - compensatory teaching for learning and research*. https://www.kth.se/en/2.83768/projekt/k-ulf-kompensatorisk-undervisning-for-larande-och-forskning-1.1048897

KTH. (n.d.). *Riktlinjer för parprogrammering*. Retrieved May 25, 2022, from https://www.csc.kth.se/tcs/projects/cerise/parprogrammering/riktlinjer.html

Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). 5 - foundations of data imbalance and solutions for a data democracy. In F. A. Batarseh & R. Yang (Eds.), *Data democracy* (pp. 83–106). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-818366-3.00005-8

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. https://doi.org/10.1145/2783258.2788620

Larusson, J. A., & White, B. (2014). *Earning analytics from research to practice*. Springer.

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*, *46*(5), 31–40.

Lonn, S., & Teasley, S. D. (2014). Student explorer: A tool for supporting academic advising at scale. *Proceedings of the First ACM Conference on Learning @ Scale Conference*, 175–176. https://doi.org/https://doi.org/10.1145/2556325.2567867

Loo, B. (2018, July 12). *Education in the united states of america* [World Education News + Research]. https://wenr.wes.org/2018/06/education-in-the-united-states-of-america

López-Zambrano, J., Torralbo, J. A. L., & Romero, C. (2021). Early prediction of student learning performance through data mining: A systematic review. *Psicothema*, *33*(3), 456–465. https://doi.org/https://doi.org/10.7334/psicothema2021.62

Macfadyen, L. P., & Dawson, S. (2010). Mining lms data to develop an "early warning system" for educators: A proof of concept. *Computers Education*, *54*(2), 588–599. https://doi.org/https://doi.org/10.1016/j.compedu.2009.09.008

Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H., & Ventura, S. (2016). Early dropout prediction using data mining: A case study with high school students. *Expert Systems*, *33*(1), 107–124. https://doi.org/https://doi.org/10.1111/exsy.12135

McPhail, S. (n.d.). Learning analytics for primary and secondary schools. http://www.nytimes.com/2010/03/07/magazine/07Teachers-t.html?_r=1&scp=1&#38

Miguéis, V., Freitas, A., Garcia, P. J., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, *115*, 36–51. https://doi.org/https://doi.org/10.1016/j.dss.2018.09.001

Murchan, D., & Siddiq, F. (2021). A call to action: A systematic review of ethical and regulatory issues in using process data in educational assessment. *Large-scale Assessments in Education*, *9*. https://doi.org/UR-https://doi.org/10.1186/s40536-021-00115-3

Natek, S., & Zwilling, M. (2014). Student data mining solution–knowledge management system related to higher education institutions. *Expert Systems with Applications*, *41*(14), 6400–6407. https://doi.org/https://doi.org/10.1016/j.eswa.2014.04.024

Peffers, K., Tuunanen, T., & Niehaves, B. (2018). Design science research genres: Introduction to the special issue on exemplars and criteria for applicable design science research. *European Journal of Information Systems*, *27*(2), 129–139. https://doi.org/10.1080/0960085X.2018.1458066

Queiroga, E. M., Lopes, J. L., Kappel, K., Aguiar, M., Araújo, R. M., Munoz, R., Villarroel, R., & Cechinel, C. (2020). A learning analytics approach to identify students at risk of dropout: A case study with a technical distance education course. *Applied Sciences*, *10*(11). https://doi.org/10.3390/app10113998

Razak, R. A., Omar, M., Ahmad, M., & Mara, P. (2018). A student performance prediction model using data mining technique. *International Journal of Engineering Technology*, *7*(2.15), 61–63. https://doi.org/DOI:10.14419/ijet.v7i2.15.11214

Romero, C., & Ventura, S. (2013). Data mining in education. *WIREs*, *3*(1), 12–27. https://doi.org/https://doi.org/10.1002/widm.1075

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, *3*, 16–20. https://doi.org/10.1007/s10758-017-9329-9

Sansone, D. (2019). Beyond early warning indicators: High school dropout and machine learning. *Oxford Bulletin of Economics and Statistics*, *81*(2), 456–485. https://doi.org/https://doi.org/10.1111/obes.12277

SCB. (2017, October 4). *3 av 10 tar inte examen från gymnasieskolan*. https://www.scb.se/hitta-statistik/artiklar/2017/3-av-10-tar-inte-examen-fran-gymnasieskolan/

Scikit-learn A. (n.d.). *Sklearn.ensemble.randomforestclassifier*. Retrieved March 17, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Scikit-learn B. (n.d.). *Feature importances with a forest of trees*. Retrieved March 17, 2022, from https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html

Scikit-learn C. (n.d.). *Tuning the hyper-parameters of an estimator*. Retrieved March 17, 2022, from https://scikit-learn.org/stable/modules/grid_search.html

Sclater, N. (2016). Developing a code of practice for learning analytics. *Journal of Learning Analytics*, *3*(1), 16–42. https://doi.org/10.18608/jla.2016.31.3

Siddiq, F., Gochyyev, P., & Wilson, M. (2017). Learning in digital networks – ict literacy: A novel assessment of students' 21st century skills. *Computers Education*, *109*, 11–37. https://doi.org/https://doi.org/10.1016/j.compedu.2017.01.014

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Skolverket A. (2019, December 5). *Lärarprognos 2019 redovisning av uppdrag att ta fram återkommande prognoser över behovet av förskollärare och olika lärarkategorier* (No. 220). https://www.skolverket.se/publikationsserier/regeringsuppdrag/2019/lararprognos-2019---uppdrag-att-ta-fram-aterkommande-prognoser-over-behovet-av-forskollarare-och-olika-lararkategorier?id=5394

Skolverket B. (2016, March 23). *Redovisning av genomförandeplan av uppdraget att genomföra verksamhetsnära insatser för huvudmän och gymnasieskolor för att förebygga avhopp från gymnasieskolan givet i regleringsbrevet 2016* (No. 13). https://www.skolverket.se/publikationsserier/regeringsuppdrag/2016/uppdrag-att-forebygga-avhopp-fran-gymnasieskolan

Skolverket C. (n.d.). *Gymnasieskolan - genomströmning*. https://siris.skolverket.se/reports/rwservlet?cmdkey=common&geo=1&report=gy_betyg_genomstromning&p_flik=G&p_programkod=&p_ar=2018&p_lankod=01&p_kommunkod=0180&p_hmantyp=&p_hmankod=&p_skolkod=

Skolverket D. (2021, November 9). *Skolverket rekommenderar en gemensam teknisk standard – ska underlätta för skolor*. https://www.skolverket.se/om-oss/press/pressmeddelanden/pressmeddelanden/2021-11-09-skolverket-rekommenderar-en-gemensam-teknisk-standard---ska-underlatta-for-skolor

Skolverket E. (2021, December 21). *Gymnasieprogrammen*. https://www.skolverket.se/undervisning/gymnasieskolan/laroplan-program-och-amnen-i-gymnasieskolan/gymnasieprogrammen

Skolverket F. (2022, March 9). *Utvecklingssamtal*. https://www.skolverket.se/regler-och-ansvar/ansvar-i-skolfragor/utvecklingssamtal

Skolverket G. (2017). *Swedish grades*. https://www.skolverket.se/andra-sprak-other-languages/english-engelska

Sousa, E. B. G. d., Alexandre, B., Ferreira Mello, R., Pontual Falcão, T., Vesin, B., & Gašević, D. (2021). Applications of learning analytics in high schools: A systematic literature review. *Frontiers in Artificial Intelligence*, *4*. https://doi.org/10.3389/frai.2021.737891

Szumski, G., & Karwowski, M. (2019). Exploring the pygmalion effect: The role of teacher expectations, academic self-concept, and class context in students' math achievement. *59*, 101787.

Tempelaar, D. T., Heck, A., Cuypers, H., van der Kooij, H., & van de Vrie, E. (2013). Formative assessment and learning analytics, 205–209. https://doi.org/10.1145/2460296.2460337

Tsai, Y.-S., Kovanović, V., & Gašević, D. (2021). Connecting the dots: An exploratory study on learning analytics adoption factors, experience, and priorities. *The Internet and Higher Education*, *50*. https://doi.org/https://doi.org/10.1016/j.iheduc.2021.100794

Turnbull, D., Chugh, R., & Luck, J. (2019). Learning management systems: An overview. https://doi.org/10.1007/978-3-319-60013-0_248-1

U.S. Department of Education. (2022, February 16). *Structure of u.s. education: Evaluation and assessment*. https://www2.ed.gov/about/offices/list/ous/international/usnei/us/edlite-evaluation.html

Utbildningsdepartementet. (2015, November 19). *Vägar framåt - strategi för unga som varken arbetar eller studerar*. https://www.regeringen.se/rapporter/2015/11/vagar-framat---strategi-for-unga-som-varken-arbetar-eller-studerar/

Vaishnavi, V., & Kuechler, B. (2019). Design science research in information systems. *Association for Information Systems*. http://desrist.org/design-research-in-information-systems

Varanasi, M. R., Fischetti, J. C., & Smith, M. W. (2018). Analytics framework for k-12 school systems. *Data Leadership for K-12 Schools in a Time of Accountability*, 206–233. https://doi.org/doi:10.4018/978-1-5225-3188-3.ch011

Vetenskapsrådet A. (2017). God forskningssed. https://www.vr.se/analys/rapporter/vara-rapporter/2017-08-29-god-forskningssed.html

Vetenskapsrådet B. (2021, December 16). *Etik i forskningen*. https://www.vr.se/uppdrag/etik/etik-i-forskningen.html

Wiliam, D., & Leahy, S. (2015). *Embedding formative assessment: Practical techniques for k–12 classrooms*. by Learning Sciences International.

Wu, X., Kumar, V., & Ross Quinlan, J. e. a. (2008). Top 10 algorithms in data mining. *14*, 1–37. https://doi.org/https://doi.org/10.1007/s10115-007-0114-2

Appendix A

Transcription of interviews

A.1 Interview 1

Vad vi hoppas få ut med det här samtalet är väl lite tankar, idéer, förslag men lite bakgrund också. Hur kom ni på det här? Vad är ert syfte? Vad vill ni att det här ska leda fram till? Så lite bakgrund och så. Lite kontext.

Ja exakt, är det okej att vi spelar in detta samtal? Det är okej

Toppen, känner du till konceptet LA? Ja.

Och vad betyder det för dig? För mig, asså, LA handlar för mig handlar ju just om att man försöker med hjälp av den data man kan på nåt sätt komma åt om elever eller studenter och hur de interagerar med i första rummet med digitala system för det är den datan man oftast kan titta på då. Gör analyser och ser hur det kopplar ihop med elevens lärande progression och i slutändan kanske resultat. Och att man försöker hitta mönster som gör att man kan förbättra de resultaten så småningom.

Juste, Vi och internet håller med om den definitionen. Men om vi hoppar vidare då, vad är det bakomliggande syftet till att börja använda LA här på skolan och hur gick tankarna där? Tankarna där, i ja, 2018 när vi började egentligen diskutera det här ordentligt, för tankarna har ju funnits där tidigare egentligen. Men vi var kanske lite upptagna med att starta skolan och bygga upp den och så. 2018 var väl ett tillfälle när vi såhäOkej, men nu är vi igång, nu kanske vi kan börja titta på andra saker. Sen så kom det en pandemi. Men då gick tankarna lite sådär, vad har vi för stora utmaningar egentligen? Då tänkte vi okej, ett ständigt närvarande problem är stress till exempel. För lärare till exempel, vi vet att läraryrket överlag är väldigt stressigt. Det är många sjukskrivningar till följd av stressrelaterade sjukdomar. Vi vet att stressen hos eleverna också är hög och ökar mer och mer. Man har sett en tydlig trend de senaste åren med stress och psykisk ohälsa hos elever. Vi vet också att vi har haft en trend med fler elever, mer undervisning, effektiviseringskrav i skolan. Då behöver man titta på såhär, finns det nåt sätt vi skulle kunna använda liksom tekniken. Det här är ändå en skola som grundades på teknik och på IT och på att använda IT för att göra undervisningen bättre på nåt sätt, göra skolan bättre. Och så här, vad finns det för verktyg vi skulle kunna jobba med för att kanske hitta nån ingång att jobba med de här frågorna, Det var iallafall min ingång framförallt. Jag började titta bland annat då på mentorskapet för det är en erfarenhet som jag har personligen som lärare men jag vet också att många delar den med mig: Att mentorskapet är tungt som lärare. Du har väldigt många elever att undervisa och ett antal elever som du är mentor för och där du ska försöka ha någon form av helhetsbild över deras utbildning och det är jättesvårt att ha den helhetsbilden. Och då så började vi fundera på: Okej men alla digitala verktyg, kan vi inte på något sätt använda datat från de verktygen för att kunna ge mentorer ett bättre underlag för att utföra den delen av arbetet, och EHT, elevhälsoteamet, för att utföra sitt arbete. Just för att vi vet att den psykiska ohälsan både bland elever och personal är ett problem som har vuxit. Det var egentligen min första ingångsvinkel. Kan vi använda det här, inte i första ledet för att öka just studieresultaten utan minska stressen, göra så att både elever och lärare är bättre informerade. elever bättre informerade om sin egen studiegång och sin egen utveckling. Lärarna bättre informerade om vilka problem eleverna får. Kan vi få till bättre signaler på att en elev är på väg åt fel håll. Kan vi hitta korrelationer mellan liksom beteenden och resultat som gör att vi kanske kan agera tidigare, kan koppla in Elevhälsovården tidigare. Jag tror att det var ingången.

Och i de här tidiga tankarna, var det här dina idéer eller bollade ni dem tillsammans med andra? Till en början bollade jag ju mycket med Donatello och med dåvarande rektor. Så där bollade vi förstås en del. Sen kontaktade vi också en kille på SU i den här frågan. Nu kommer jag inte ihåg exakt hur vi kom fram till att vi skulle ta kontakt med just SU kontakten. Han sitter här i huset. Han jobbar ju med LA, det är hans forskningsområde. Teknikstött lärande med hjälp av LA, och AI. Då tog vi kontakt med honom och tanken var att vi skulle få igång något litet projekt där. Men sen så, det blev aldrig riktigt någonting där på grund av... Han hade för mycket att göra, vi hade för mycket att göra och sen fick vi en pandemi på köpet. Det stannade av där, men vi började ju där. Så började vi bolla där. Lite som vi gjorde nu [refererar till innan intervjun när vi gick igenom datan som var tillgänglig] , vi hade en genomgång av vilken data som vi har tillgänglig. Och började prata om hur kan vi göra. Och sen så började vi ta fram en leverans av data som han kunde titta på. Men i väntan på att det skulle rulla igång så började vi också prata lite mer själva och titta på den datan vi hade och där under våren 2019 så började IT-pedagogen också exportera data och spara ner den data som vi hade tillgänglig. Så under sommarlovet 2018 satt jag och gjorde lite analyser själv bara på amatörnivå.

Spännande och hittade du någon rolig trend och i era tidiga dialoger hade ni några konkreta idéer på vad de kunde användas till? Ja, Om vi börjar med andra frågan, konkreta idéer, för det vi pratade om var ju bland annat sådana saker som kan man hitta liksom trender i det här, kan man då göra så att vi automatiserar en analys och presenterar resultatet i det systemet som vi har för utvecklingsamtal t.ex. så att man kan hitta en visualisering av det här datat som gör att dels elever får en bättre förståelse för sitt eget beteende kopplat till skolresultat. men också .då till lärare, mentor och EHT att kunna i det systemet få någon visualisering som gör att man kan agera tidigare. Så det var egentligen det som vi tänkte skulle var målet med det här. Att få till en enkel visualisering för skolpersonal. Och en liten del av det har vi också implementerat. Så att vi har vissa parametrar som vi då identifierade där det fanns en stark korrelation har vi då tagit in i det systemet och presenterar. Och då är det närvaro förstås, den är så uppenbar att den påverkar. Men sen har vi också tagit in framförallt participation level i kurserna och saknade uppgifter. Vi har page views med också. Det vi kunde se när vi började titta, men då är det här ett för litet underlag för att egentligen kunna dra några generella slutsatser om, men det vi kunde ana i alla fall i de data bearbetningarna var ju då dels att närvaron korrelerade väldigt starkt, Antal saknade uppgifter korrelerade extremt starkt med F-betyget och med varningar. Jag körde framförallt korrelationer med underkända betyg, e-betyget också. Jag kollade på F och E betyget, framförallt just för att även låga betyg är en bra varningssignal. Och varningar. Och då så såg jag ju också att det fanns lite olika trender när det gäller page views och participations. Det fanns ingen enkel korrelation där. Det var inte så att fler page views var bättre, t.ex och det var inte nödvändigtvis så att mer participation var bättre heller alltid. Men jag kunde se vissa trender, att det fanns folk som hade höga page views men låg participation t.ex. De var ofta en riskfaktor som jag kunde identifiera. Och rent, väldigt icke säkerställd förklaring på det skulle kunna var att det här är en elevgrupp som är väldigt osäkra. Går in och kollar väldigt mycket men har väldigt svårt för agerandet. Tittar väldigt mycket, försöker förstå men kommer aldrig loss och lyckas göra uppgiften. Elever med mycket prestationsångest eller den typen av problematik. Sen har vi dom som har låg participation och låg page views och det var också en stark riskfaktor för det tenderade att vara elever som inte engagerade sig överhuvudtaget i skolarbetet. Elever med låga page views och höga participation tenderade vara elever som klarade sig väldigt bra. De behövde bara kolla på informationen en gång, dom gör sakerna klart, liksom.

Skulle man vilja att det här resulterade i ett verktyg, eller som du sa kopplat till utveckling? Jag tycker ju att det vore, kan man liksom hitta stabil, bra korrelationer här så skulle ju det vara värdefullt om man på nåt sätt sammanställer det här till någon form av varningssystem. Så tror jag att det skulle vara väldigt användbart.

Och när du satt och programmerade eller data analyserade lite på kammaren, vad använde du för verktyg? Använde du några moduler, språk? Python för att det enklast, iallafall för min del. Det är också så att en annan person som jag faktiskt bollade en del med om metodik är min fru. Min fru är doktor i bioteknik, hon jobbar som bioinformatiker och analyserar mycket större datamängder än det här på daglig basis, men i hennes fall så är det ju då framförallt DNA-sekvenser som hon analyserar, så det är ju väldigt annan typ av data egentligen koppla till mycket frågeformulär och sådär. Men det var också en anledning till att använda Python för hon jobbar ju i det, hon jobbar ju i R också förstås men med Python kunde jag ju också få hjälp när jag körde fast för jag jobbar ju inte med det här. Jag har läst programmering och webbutveckling och har en grund i det men jag har ju inte läst statistik på högre nivå, jag har läst en statiskt kurs på universitetet och det var 2002 så det var ganska längesen.

Använde du någon specifik AI-algoritm? Nej, så långt kom jag inte, jag gjorde bara, tittade bara efter korrelationer i det här fallet och gjorde ju egentligen en analys i taget med olika datapunkter. Satt i Jupiter notebook, Python.

Framtidsambitioner då? Tänk om du fick drömma stort. Om jag fick drömma stort! Får jag drömma stort skulle jag ju verkligen försöka få till, vilja få det här till Machine learning hållet i längden, om man kunde skapa ett verktyg som kör den här analysen och förbättrar algoritmen, där man liksom kan ta den här datan löpande och kunna få en initial, ja en initial varning eller något sånt för, ganska tidigt, vi gör ju dom här diagnoserna tidigt för att kunna upptäcka saker tidigt. Men det vi märker är ju också så att visst, vi försöker liksom se, vi gör diagnoser i matte, Svenska Engelska, och så försöker vi kolla på om vi får in någon information och sen har vi något litet möte där vi går igenom dom här sakerna. Men det är ganska mycket att titta på och det är till viss del personberoende. Jag är ju till exempel ganska van vid dom här, jag har jobbat ganska länge med dom här stanine värdena på svenskan till exempel. Jag vet ungefär, om det här är högt, det här är lågt så betyder det det här. om det däremot är tvärtom så betyder det någonting annat. Men det är ju en kunskap som sitter i min erfarenhet. Och försvinner jag, så försvinner den kunskapen med mig på skolan, eller i det här fallet. Och det kanske det kommer in någon annan som besitter den kunskapen, eller jag kanske hinner lära upp någon av mina kollegor att se de här sakerna. Men kan man, det här är ju som den typen av kunskap som går att automatisera, till skillnad från den kunskapen som är i klassrummet, som faktiskt inte går att automatisera på det sättet. Det är en träningssak, iallafall med dagens teknik så går det inte. Men det här är ju sånt som skulle gå att bygga in i system.

Vad finns det för risker med LA? Dels finns det ju en del etiska komponenter. Vi vet sedan tidigare att om en lärare får en grupp elever och får information om att det här är en E-elev, det här är en C-elev så oavsett om det stämmer med verkligheten så ökar sannolikheten för att det är just det betyget som eleven får i slutändan. Det finns ju den risken, att om man ger tidig information också cementerar vissa saker så det är ju en tydlig risk, som behöver diskuteras. För man vill ju att det ska vara en positiv sak att vi får den här informationen, att vi kan agera på den snarare än att det ska vara en information som gör att vi begränsar en elevs utveckling så det ser jag som en ganska tydlig risk i det här fallet. Finns väl också vissa integritetsrisker med all den här dataanlaysen och hur den hanteras och presenteras.
Har ni tänkt i de här vägarna om hur man skulle kunna lösa de eventuella.... Har tänkt men har inga bra lösningar än så länge. En del av det är också att fundera över vilken visualisering man ska presentera för vissa personer. t.ex. att presentera saker för EHT som inte är betygssättande och inte undervisande kanske är, de kan få en helt annan nivå av den här analysen, för deras jobb är specifikt att jobba med elevernas hälsa. Jag kan också tänka mig att det finns en hel del skolor där mentorskapet är frikopplat från lärarskapet. Där mentorer skulle kunna få någon information som inte riskerar att på samma sätt påverka undervisningen och betygsättningen. Och det är väl där jag ser att den stora risken är att presentera det här för undervisande lärare eftersom vi också är myndighetsutövande personer. Det finns sätt men i vissa fall kanske det skulle kräva en omstrukturering av skolans verksamhet.

Vi pratade förutom också om psykisk ohälsa och stress. Tror du att det finns någon risk att eleverna känner någon stress om man har för mycket data på eleverna, att man känner sig övervakad som elev? Så är det ju också, Det finns nog också en sådan risk. Min erfarenhet är att elever är så vana vid det här att de inte tänker på det men det finns ju definitivt något man behöver tänka på där, vilken nivå av övervakning är det rimligt att ha? Samtidigt, vi gör redan allt det här, vi samlar ju in all den här datan, den finns ju nånstans. Det är bara det att just nu är det leverantörer som sitter på den, det är leverantörer som sitter på den datan. Och visst, då har de inte hela bilden utan bara fragmenten, men datan är ju redan insamlad på ett sätt. så frågan är om det så mycket värre om man faktiskt använder datan istället för att bara samla in den på olika ställen. Men det är definitivt en fråga man behöver titta på också.

Något som vi är inne på analyserat också är ju det här, när man samtalar med eleven, hur ska man framföra feedbacken, men just också beroende på om man säger att vi har analyserat era Canvas data, om det medför att eleven får ett förändrat beteende på Canvas plattformen, "Nu är jag inne och klickar här en massa böra för att". Där måste man vara försiktig med vad man presenterar och hur man presenterar för det tror jag också att det finns en studie på att elever, även för yngre åldrar om jag inte missminner mig, just att det finns en tendens att: Okej, om det är det här att man fokuserar på beteendet och inte det som ligger bakom beteendet, att det är det här med att jag klickar runt mycket betygsätts, men det är ju inte det som betygsätts. Man måste presentera det på ett sätt där man presenterar just vad, motivationen till beteendet och inte beteendet i sig.

Nu, har vi fått uppfattningen kring att du är mer tekniskt intresserad än gemene lärare på den här skolan. Det kan nog stämma ja.

Och då tänker vi att när vi ska presentera den här analysen som eventuellt kommer så kanske du har ett försprång eftersom du har det tekniska kunnandet. Vad är viktigt för att andra lärare ska kunna använda just de här verktygen och att man ska försöka få någon slags transparens i det som sker i någon typ av data analys och eventuell AI-algortimer. Framför allt är det ju viktigt att presentera det och visualisera det på ett lättillgängligt sätt. Vilket vi är ganska dåliga på i skolans värld överhuvudtaget. Den lilla förändringen vi gjorde med att lyfta några parametrar i underlaget för utvecklingssamtalet, det var ju väldigt uppskattat av väldigt många men det var ju även väldigt lätt att förstå vad dom handlade om. I det fallet så var det inte heller någon vidare analys av olika parametrar eftersom det var väldigt enkelt.

Det enda vi gjorde var en färgkodning, ligger du över det här gränsvärdet så blir det rött för att det här är någon man behöver kolla närmare på. Så det handlar ju mycket om hur man visualiserar och att man gör det på ett sätt som inte belastar läraren. För kräver det att läraren måste lära sig verktyget eller lära sig förstå, då kommer det inte funka enligt min erfarenhet. Är det så att det är intuitivt att förstå, "det här betyder det här och det är baserat på det här." Där är väl en av de största utmaningarna för att det finns en skepticism här hos många. I många fall en sund sådan och kanske även en överdriven sådan. Det är också så att det kommer ju aldrig komma, eller iallafall inte din och min livstid så långt i den här typen av automatiska analyser att vi fångar allt. Det finns så många olika förklaringar till vissa saker. Och då är det ju också viktigt att man, eh, jag kan ta ett exempel: Min erfarenhet från, ni känner till Urkund?
mmhm eller original som de heter nu för tiden. Där får man bara en procentsats från Urkund i första ledet. Och i ganska många fall så ser jag bara lärare som gör den väldigt enkla analysen att: Okej, men om det är över den här procentsiffran så kommer jag räkna det som fusk och den analysen är för trubbig, det funkar inte. Utan du måste gå steget vidare, vad beror de här procentsatserna på? Och gå in och titta på analysen och undersöka saken ordentligt, men gränssnittet inbjuder inte till det, gränssnittet är så enkelt att se den här procentsatsen. Särskilt om integrationen av Urkund i Canvas också är färgkodad. Då blir det så lätt att bara; Okej det är grönt, det här är gult och det där är rött. Att man bara går på det. Det gäller utmaningen att göra gränssnittet så enkelt att förstå som möjligt utan att göra det så pass enkelt att man undviker att undersöka vad de underliggande faktorerna är.

Så en typ av trafikljus skulle vara att rekommendera då säger du eller? På ett sätt ja, men med den brasklappen att det finns en risk att det inbjuder till en förenklad analys av situationen.

Skulle man vilja komplettera med, skulle det vara intressant att ha grafer och liknande saker? Ja, Grafer ger på en gång en lite mer, Okej nu måste jag faktiskt tänka efter vad det här betyder. Man anstränger sig lite mer där, men det är fortfarande något som alla är vana vid egentligen.

Finns det möjligheter att integrera grafer kopplat till andra betygssystem till exempel, eller de plattformarna som ni har? Säg att du har elev 123, om du klickar på dess sida skulle man kunna se att här har vi en graf över klassens snitt kanske? Och sen har vi elevens snitt och att det är lägre eller någonting? Det är ju fullt möjligt att göra, i de plattformar som vi kontrollerar själva, på den här skolan har vi möjlighet att göra det, på en annan skola kan det vara svårare.

Är det någonting som du skulle vilja lägga till som du anser att vi missat? En viktig aspekt i det hela. Tänk brett. Nej jag vet inte. Det är inget som dyker upp så, en sak som jag kan tänka mig bara på just det som vi pratade om nu är att det vore bra om man nu har den typen av visualisering så kan det vara bra att ha ett enkelt sätt för lärare att ge feedback på den visualiseringen för att ibland så kan lärare sitta på förklaringsmodeller som man inte har tänkt på. Att den inte bara är statisk utan det finns en möjlighet att, ett enkelt sätt att säga: Nej men det här blir missvisande.

Det finns ju, LA har ju tidigare kunnat, det finns höga ambitioner om att det ska gynna konkretiserad feedback. Det är svårt att säga vad LA kommer kunna ge. Än så länge har det ju inte visat sig ge särskilt mycket. Men jag tror ju att det kan göra det. Jag tror att en anledning till att man inte har sett särskilt mycket är för att man inte listat ut hur man ska använda det än. Och att man behöver båda, dels programmeringen men även statistiker som ska jobba med det statistiska analyserna och att de faktiskt är statistiska och riktiga. Och sen behöver du UX-designers. Du behöver folk som jobbar med den biten, som man har missat ganska mycket.

Och samarbetet mellan inblandade aktörer? Så är det ju, det här är ju en av kritikerna som vi har mot Stockholms stad och arbetet med skolornas system är att man har varit, projektet har varit för långt bort ifrån verksamheten. Jag är väl en av de som varit mest inkopplad i skolplattformen, specifikt i utvecklingen av den, i egenskap av att jag är förstelärare på den här skolan med uppdrag inom digitalisering och jag var med i nån grupp av utvecklingslärare inom staden, så jag är en av de som varit inkopplad mest och jag har knappt sett någonting av den utvecklingen. Och då känner man ju lite att nej, då har man ju kommit för långt bort från aktörer som faktiskt ska använda systemet. Jag fick inte se någon del av systemet under någon fas av utvecklingen trots att jag var en av dem som skulle utbilda resten av personalen på skolan. Jag fick tillgång till det samtidigt som personalen på skolan. Där har man ju missat någonting. Man måste göra den här utvecklingen iterativt i samröre med olika skolor, olika lärare med olika bakgrund. Som vi var inne på här, jag har en bakgrund och ett intresse som gör att det här är en utav de frågor som jag tycker är väldigt intressant. Man måste ju utveckla sådana här saker tillsammans med de lärare som hatar allt vad, datorer och LA är. De ryser vid tanken. De måste man ju också ha med.

Har ni lyckats med det här på skolan? Eller hur går de konservativa

tankarna? Till viss del, vi har ju lyckats på så sätt att de lärarna har tyckt att iallafall de informationspunkter som vi förde in har gett någonting. Frågan är hur man får med dem längre. Där får man väl jobba lite stegvis. Och det är väl också en viktigt punkt att man kan ge möjligheten för lite individanpassade gränssnitt när man ska titta på datan. Att det finns ett enkelt gränssnitt för de som inte vill behöva sätta sig in i det. Men att det finns möjligheter att titta lite djupare i datat för de som vill det. Då kan man också få, man kan börja med bara det här och stegvis komma längre.

Inget annat? Inget som jag kommer på just nu iallafall.

Men som jag förstod det så huvudmålet från din sida skulle vara mer åt att man inte just höjer betygen utan man siktar på mentorskap? Jag tror att betygen skulle vara en konsekvens av det ändå. Det jag känner att man behöver sikta in sig på är det som jag upplever är det enskilt största problemet. Kan man komma åt det och förbättra det, då kommer det andra ge synergieffekter i form av bättre betyg och bättre kunskapsutveckling hos eleverna. Det är svårt att lära sig någonting när man sitter och har ångestattacker på lektionerna. Så för min del, jag tror också att ett bättre mentorskap skulle hjälpa den progressionen. Kan man använda det här för att komma åt det här enorma problemet i skolan så vore det bra.

När du pratar mentorskap, är det framförallt individuellt då eller är det också de här mentorslektionerna? Det skulle ju teoretiskt sett kunna vara båda två. Det jag tänker mig främst är ju det individuella i det här fallet. Sen kan man ju också tänka att kan man analysera och identifiera behov på gruppnivå då kan det också vara underlag för den typen av arbete.

Det är bra så, ska vi säga så? Ni får höra av er om ni har följdfrågor.

A.2 Interview 2

Okej att spela in? Ja!

Hur delaktig har du varit i detta initiativet, som började utifrån Raphael och Donatello? Ja ganska lite, har varit i det direkt, vetat om det hela, generellt från mitt perspektiv viktigt som möjlighet att vara med eftersom vi har förutsättningarna att kunna vara med. och att Michelangelo har ett uppdrag som ligger inom ramen. På detaljnivå är jag inte delaktig.

Är du bekant med begrepp som Learning Analytics? Lite granna. Vi definierar det som att upptäcka mönster och trender

Är ditt ansvar på en högre struktur eller hur skiljer sig ditt ansvar? Mitt ansvar här är, vi har en annan ledningsstruktur, så det skiljer sig. Man kan vara med och initiera, men görandet och själva detaljerna är inte jag. Det är grupper i personal eller enskilda individer. Sen tar jag hemskt gärna del av hur det löper på, följer resultat. men ett sådant här projekt som ska vara öppet och ligga mer på personer i verksamheten, jag är inte så nära verksamheten och ska inte vara heller.

Men du anförtror mycket ansvar? Ja det måste man göra. En person, en VD ska inte gör allting. Så funkar det inte. Det är ett förtroende man ger. Du önskar att de kommunicerar? Under resans gång vill man ha avstämningar.

Vad skulle du säga att dina förhoppningar är med projektet? Det är spännande om man går in på IT sidan. Det är så att vi har en jävla massa system som pratar med varandra. Det ser ut olika på olika skolor. Vi har lite mer professionella system än staden har upphandlat. Det är viktigt att kunna sammaställa data och få ett resultat av det. Få skilda sektorer att kommunicera med varandra. Upptäcka saker, dra slutsatser.

Är det några slutsatser dom du som rektor skulle anse vara extra viktiga och extra intressanta att applicera? Det pratas litegranna och i något visst öra har man haft med. Kring det här med Michelangelo och Donatello osv. De kanske inte sagt detta men att följa elevers väg till examen så nu vet jag inte om de har nämnt detta, resultatet från grundskolan det kanske inte har kopplat in. Som ledare här så kan jag tycka att resultatet från grundskolan kontra det som händer här är ett dike. Du har ett dike mellan alla skolsystem. Jättedike mellan gymnasieskolan till KTH. Ännu större dike mellan KTH och arbetslivet sedan

Så ett dike är att elev 123 behöver jobba mer på det här? Ja det kan vara så här, jag kopplar det till glädjebetyg och till de tester och diagnoser vi gör i början. För vi ser variation beroende på vilken skola du har lämnat. Men det kanske är ganska stort, men det kan vara intressant att följa våra elever och säga vilka förutsättningar de kommer in med, och vilken output man har i slutändan. Vilka förädlingsvärden egentligen. Hur ser förädlingsvärdet ut i poäng. Tyvärr så visar det sig att det inte är så där jättepositiva värden på förädlingen i våra skolor, kring vad man kommer in med och vad man kommer ut med. Där hade det kunnat vara bättre. Vi har duktiga elever här

Men då skulle det behövas stöd? Det problemet ligger ännu högre upp i organisationen? Så skulle det kunna vara, men det är inte vad man tittar på. Inte så Leonardo och Michelangelo tänker. Det är bättre att ta något som finns konkret på den nivå så att säga.

Med dem två har vi pratat om EWS och mycket kring mentorskapsbiten och elevhälsoteamet. Det har vi pratat om. Det är löpande här. Jag pratande om EWS med Michelangelo igår. Den är central, men man fångar och hur man fångar det är faktiskt kopplat till vad Raphael har lyft. Mentorskapet hänger ihop. Kan man som mentor på ett enklare sätt få också en varning när det håller på och…

Om man tänker så här, dem två är ju väldigt tekniskt drivande lärare, hur ska man kunna göra för att det här skiftet mellan teknik och någon visualisering ska nå och vara applicerbart på alla lärare? De flesta är väl ändå lite tekniskt drivna här, annars ska man inte jobba på SSiS. Vi väl gärna vara en förebild. Jag tycker vi kan utgå från det tekniska här, Vi har ju hela skolans, sitt andra ben är att vi ska vara ledande inom ICT. Så det är ett av våra ben. Så det kan jag förvänta mig att all personal då kan ta del av.

Så, man kan säga att förutsättningar är bättre? De är bättre här än om man går ut på någon annan skola, så är det. Det är betydligt bättre förutsättningar här. Det gäller att jobba digitalt och med teknik överlag? Ja det är bättre även fast det finns en och annan som, det finns ju på alla arbetsplatser där man kan göra bättre ifrån sig, eller vilja lära sig mer. Jag själv kan tex skriva under på det

Finns det några risker med allt det här? Ja det är klart att det gör, det är ju att datat måste hållas inom stängda dörrar. Spårbarheten är viktig, ur det perspektivet att det ska inte vara spårbart så att det går att identifiera individer. Men det känns så i just det här men det finns alltid en risk med datat. Jag ringde in och frågade om juridiskt godkännande och man såg inte bekymmer på det. Jag frågade vår jurist. Men det är ju ett ansvar man har, att det inte finns personnummer och namn i tid och otid. Det kan finnas annan spårbarhet också i och med system som kopplas ihop, det kanske går att spåra individer ändå om man är insatt, jag vet inte men det är viktigt att det hålls anonymt.

Om du fick drömma stort? Vi ska in och pilla litegrann men om man ser långt in i framtiden? När jag tänker framåt så tänker jag liksom mer i stort och egentligen har vi en språngbräda just nu när Coviden har gjort sitt. Så har vi blivit bättre på mycket av det digitala. Så tänker man stort ur det perspektivet så när det gäller utbildning generellt framåt, kommer skolan se likadan ut även efter Coviden? Eller var det här den där Kickoffen man fick? Skolan ser ju tyvärr likadan ut som när jag började som lärare 1984 och det ser ganska likt som det gjorde för mina föräldrars skolgång. Och det vill man väl hoppas att sådan här verktyg, om man kollar på EWS, att det kan underlätta för att använda det här på något sätt i utbildningen på ett bra sätt sådant att det underlättar för lärare. Så om man kan få trender tidigt för att kunna hjälpa och stötta men också kan det vara just den här språngbrädan, måste utbildningssystem se likadana ut om tio år? Dvs, generellt inom skolutvecklingen, vad ska vara digitalt och avd ska vara fysiskt på plats? Vi har faktiskt elever här som uttrycker att de pluggar mycket bättre hemma

Det är väldigt individuellt? Ja det är väldigt individuellt, hur kan man öppna sådana system, då kan man förbättra dom. Då kommer ju sådana saker in som EWS. Om du har en elev som inte är på plats, hur ska man följa upp den eleven? Och hur kan man se om det inte går bra då man inte är på plats? Vi har ju bilder av hur skolan ska vara, tyvärr får jag säga.

Du som jobbat som lärare och rektor ett tag, ser du några skillnader sedan du började som lärare? Jag ser alldeles för många likheter. Det är klart jag ser skillnader. Men jag ser för många likheter i förhållande till en generations servicetid. Eller ett arbetsliv, 40 år eller vad det kan vara. Alltså, det har hänt för lite

På alla nivåer? Nej på alla nivåer! Ni går på KTH, hela utbildningsvärlden är konserverande system på något sätt. Det är inte likadant i bilindustrin. Vi kan ju prata om Elon Musk och sådär så får man en bild av vad som hänt. Jag kommer ihåg när jag var liten pojke. Dom där Saab 93. De hade en affärsmodel där bakrutan skulle bli lite större för varje år. Utvecklingen var att göra större och större bakrutor varje år. Och det levde man ganska länge på hos Saab. Men man kan ju bara börja fundera lite då, hur många år har man haft dem där? (Håller upp sin mobil)
10, 15? Hur snabbt går det när saker och ting tar fart? och sedan tittar man på skolutveckling, hur skolor ser ut? Även från er sida som är så pass unga just

nu jämfört med mig, så ser skolor lika ut. Ni hade inte mobiltelefoner när ni gick på lågstadiet.
Nokia 33.10. Men inte smartphones.
Nej de kom på högstadiet. Och nu har man fått någonting som även gamlingar använder, som slagit igenom hela samhället. Vad har slagit igenom av samma kraft i utbildningsvärlden?

Men skulle du säga att ni på SSiS är bättre eller har bättre förutsättningar för att lyckas eftersom ni har en teknisk nisch? Ja det tror jag kan säga att man har, för att man behöver ha en bra teknisk utrustning. Och det ju visat sig nu under pandemin att vi har bättre förutsättningar. Inspektionen var ju här och gick tillbaka och skrev i rapporten att de har lärt sig saker här. Det är ju positivt. Vi har konferensutrustning i varje sal. En projektor kostar 8-10 tusen kronor i vanliga fall, här kostar de 40. Vi har haft tur på det sättet att kunna investera i högtalar och mikrofon system för konferensanläggningen. Och det har vi i varje sal. Och det gör ju också att när man haft en pandemi och jobbat på distans till och från, 100 eller 150 % så har kvalitén i det man levererat varit bra. Vi har personal som kan hantera de här system på ett bättre sätt tror jag.

Om vi går tillbaka lite, för jag tycker det är intressant. Initativet till gymnasieskolan kom det från Stockholms stad? Det drog igång egentligen med tankar 2011 och sedan pågick uppstarten i 2 års tid. Och jag hade ju i uppdrag att starta, så jag har varit med i från början. Så det här huset fanns ju inte. Vi lever kvar på de fyra benen vi hade då. Vi ska ha en lärandemiljö som är modern. Vi ligger ju också i affärshus, här ute i Kista. ICT fokus var ju uppdrag. Sedan skulle vi då också ha nära kontakt med näringsliv och högskole, Universitet och det har vi. Och det fjärde benet är arbetsformerna. Och det jobbar vi lite mer med nu. Alltså att föra in mer projekt. Föra in mer företag i projekt när eleverna jobbar med det. Så vi inte bara har elev, lärare, klassrum. Och det är lite det jag menar med utveckling av skola liksom. Vi måste här förbereda våra ungdomar till det arbetsliv de kommer ut i. När de kommer ut efter KTH eller DVS, här nere eller kanske får jobb direkt efter gymnasiet så man ändå ha lite koll på designthinking på IBM. Man måste vara lite van att jobba på det sättet. Och det här är ju en stor fråga nu, att vi ska försöka driva på lite kring företag och arbetsformer.

Hoppar tillbaka lite och vissa lärare kanske vill ha grafer eller typ av visualisering just kopplat till EWS. Skulle det vara intresse att som skolledning att få grafer just kopplat till hur elever ligger till? Att få en överblick? Ja det är klart att det är, Man vill ju se. På individnivå poppar det alltid upp. Man vill gärna se helheten. Om det går att få trender och tendenser på skolnivå eller årskursnivå. Så man kan kika och titta. och dra slutsatser. Individerna, alltså om man pratar om rena elevhälsobekymmer eller problem förövrigt med studierna de poppar upp ändå. Men i sådant där system kan det ju vara bra att läraren lättare kan upptäcka eleverna när det börjar knaka. Men sedan när det går riktigt dåligt, då kommer det ändå upp via elevhälsan, ofta.

Man pratar mycket om school drop out. Har ni många som hoppar av? Det har vi väl inte, inte jämfört med andra skolor. ganska lite dropouts.

Och det som hoppar av, vet ni varför? Ahh, det kan vara lite olika. En del hoppar av för att det var fel program. Det sker väl ganska tidigt, en del upptäcker efter en tid att de inte vill gå i sådan här liten pluttskola, jag vill gå i sådan där tegelhög inne i stan. Det kan ju vara ett skäl. De som lämnar generellt sätt så finns det ofta av dem man skrivit ut och som kommit in i systemet, då är det elevhälsoärenden, till 99% skulle jag säga. Det finns något annat som inte har med skolan att göra, som genererar det. En grej som är jobbig idag på gymnasieskolan och kommer vara några år till är och som ni säkert också upplevt är ju stressen, det är koppla till systemet vi har idag. Vi har ett helt galet skolsystem med olika kurser. Jag har aldrig förstått varför man ska ha tre betyg i svenska och tre, fyra eller fem i matte. Det var aldrig så förr och det har man lyckats utveckla från politiskt håll men till det sämre. För det har genererat en extrem stress hos ungdomar, så man tappar kopplingen och helhetssynen på sin utbildning där för att fokuserar på att få betyget A i matte 1 och sedan A i matte 2. Och sedan är det Nationella prov som är gud fadern själv som droppar ned. Systemet har gjort en björntjänst till alla som går på gymnasiet idag. Liknande som på högstadie kommer troligtvis 2024. Det är beslut fattat om att försöka gå över till ämnesbetyg. Du kommer få ditt betyg i slutet av trean. Då har du möjlighet att utvecklas och bli bättre. Det är helt vansinnighet att kursindela gymnasiet

Är det något du skulle vilja lägga till kring Learning Analytics eller data analys biten? Det är kul att ni är här, hoppas att det blir ett bra samarbete. Ska bli kul att ta del av resultatet.

Appendix B

Tables

Table B.1: *Variable importance for all eight courses applied in figure 6.1 and figure 6.2*

variable \ Course	Physics 1a	Chemistry 1	English 5
Absence	0,063	0,18	0.26
Math diagnosis	0,54	0,11	0.20
Swedish letter chains	0,0017	0,0075	0.19
Swedish word chains	0,090	0,013	0.090
Swedish sentence chains	0,046	0,12	0.080
English total	0,051	0,050	0.19
Previous course	-	-	-
Second previous course	-	-	-
Page view factor	0,021	0,34	-
Participation factor	0,16	0,15	-
On-time factor	0,031	0,027	-
Missing factor	-	-	-

variable \ Course	English 7	Mathematics 1c	Mathematics 3c
Absence	0.040	0.058	0.16
Math diagnosis	0.036	0.18	0.22
Swedish letter chains	0.069	0.054	0.062
Swedish word chains	0.025	0.58	0.089
Swedish sentence chains	0.021	0.076	0.12
English total	0.043	0.048	0.073
Previous course	0.079	-	0.069
Second previous course	0.50	-	0.045
Page view factor	0.040	-	0.17
Participation factor	0.11	-	-
On-time factor	0.00	-	-
Missing factor	0.035	-	-

variable \ Course	Mathematics 4c	Technology 1
Absence	0.067	0.034
Math diagnosis	0.089	0.12
Swedish letter chains	0.14	0.12
Swedish word chains	0.025	0.049
Swedish sentence chains	0.084	0.094
English total	0.093	0.038
Previous course	0.19	-
Second previous course	0.08	-
Page view factor	0.19	0.33
Participation factor	-	0.15
On-time factor	-	0.071
Missing factor	-	-

Table B.2: *Variables choosen for algorithms with Learning Management System data*

variable \ Course	Physics 1a	Chemistry 1	English 5
Absence	**0,063**	**0,18**	**-**
Math diagnosis	**0,54**	0,11	**-**
Swedish letter chains	0,0017	0,0075	**-**
Swedish word chains	**0,090**	0,013	-
Swedish sentence chains	0,046	**0,12**	-
English total	0,051	0,050	**-**
Previous course	-	-	-
Second previous course	-	-	-
Page view factor	0,021	**0,34**	-
Participation factor	**0,16**	**0,15**	-
On-time factor	0,031	0,027	-
Missing factor	-	-	-

variable \ Course	English 7	Mathematics 1c	Mathematics 3c
Absence	0.040	**-**	**0.16**
Math diagnosis	0.034	**-**	**0.22**
Swedish letter chains	**0.069**	-	0.062
Swedish word chains	0.025	**-**	0.089
Swedish sentence chains	0.021	**-**	**0.12**
English total	0.043	-	0.073
Previous course	**0.080**	-	0.069
Second previous course	**0.50**	-	0.045
Page view factor	0.040	-	**0.17**
Participation factor	**0.11**	-	-
On-time factor	0.00	-	-
Missing factor	0.035	-	-

variable \ Course	Mathematics 4c	Technology 1
Absence	0.066	0.034
Math diagnosis	0.089	**0.12**
Swedish letter chains	**0.14**	**0.12**
Swedish word chains	0.025	0.049
Swedish sentence chains	0.084	0.094
English total	**0.093**	0.038
Previous course	**0.19**	-
Second previous course	0.081	-
Page view factor	**0.19**	**0.33**
Participation factor	-	**0.15**
On-time factor	-	0.071
Missing factor	-	-

Table B.3: *Variables choosen for algorithms with out Learning Management System data*

variable \Course	Physics 1a	Chemistry 1	English 5
Absence	**0.35**	**0.50**	**0.26**
Math diagnosis	**0.25**	**0.12**	**0.20**
Swedish letter chains	0.065	0.062	**0.19**
Swedish word chains	0.088	**0.078**	0.090
Swedish sentence chains	**0.12**	0.063	0.080
English total	**0.12**	**0.17**	**0.19**
Previous course	-	-	-
Second previous course	-	-	-

variable \Course	English 7	Mathematics 1c	Mathematics 3c
Absence	**0.14**	**0.058**	**0.16**
Math diagnosis	**0.083**	**0.18**	**0.22**
Swedish letter chains	0.0075	0.054	0.062
Swedish word chains	0.011	**0.58**	**0.089**
Swedish sentence chains	0.0052	**0.076**	**0.12**
English total	0.068	0.048	0.073
Previous course	**0.22**	-	0.069
Second previous course	**0.47**	-	0.045

variable \Course	Mathematics 4c	Technology 1
Absence	**0.15**	**0.39**
Math diagnosis	0.087	**0.18**
Swedish letter chains	**0.095**	0.059
Swedish word chains	0.045	0.079
Swedish sentence chains	0.051	**0.20**
English total	**0.18**	**0.096**
Previous course	**0.19**	-
Second previous course	0.081	-

Table B.4: *Prediction accuracy, specificity and sensitivity for algorithms without Learning Management System data*

Course	Meassure	Value	Test data category	# data samples
	Accuracy	0,74	Total	31
Physics 1a	Specificity	0,78	Grade E	27
	Sensitivity	0,5	Grade F	4
	Accuracy	0,68	Total	19
Chemistry 1	Specificity	0,86	Grade E	14
	Sensitivity	0,2	Grade F	5
	Accuracy	0,75	Total	4
English 5	Specificity	0,67	Grade E	1
	Sensitivity	1	Grade F	3
	Accuracy	0,62	Total	8
English 7	Specificity	0,67	Grade E	2
	Sensitivity	0,5	Grade F	6
	Accuracy	0,9	Total	10
Mathematics 1c	Specificity	1	Grade E	9
	Sensitivity	0	Grade F	1
	Accuracy	0,77	Total	35
Mathematics 3c	Specificity	0,86	Grade E	28
	Sensitivity	0,43	Grade F	7
	Accuracy	0,62	Total	26
Mathematics 4c	Specificity	0,64	Grade E	22
	Sensitivity	0,5	Grade F	4
	Accuracy	0,75	Total	8
Technology 1	Specificity	0,71	Grade E	7
	Sensitivity	1	Grade F	1

Table B.5: *Prediction accuracy, specificity and sensitivity for algorithms with Learning Management System data*

Course	Meassure	Value	Test data category	# data samples
	Accuracy	0,85	Total	13
Physics 1a	Specificity	1	Grade E	11
	Sensitivity	0	Grade F	2
	Accuracy	0,75	Total	12
Chemistry 1	Specificity	0,71	Grade E	7
	Sensitivity	0,8	Grade F	5
	Accuracy			
English 5	Specificity		Not enough DP	
	Sensitivity			
	Accuracy	0,75	Total	4
English 7	Specificity	0,67	Grade E	3
	Sensitivity	1	Grade F	1
	Accuracy			
Mathematics 1c	Specificity		Not enough DP	
	Sensitivity			
	Accuracy	0,65	Total	20
Mathematics 3c	Specificity	0,81	Grade E	17
	Sensitivity	0	Grade F	3
	Accuracy	0,67	Total	12
Mathematics 4c	Specificity	0,78	Grade E	9
	Sensitivity	0,33	Grade F	3
	Accuracy	1	Total	2
Technology 1	Specificity	1	Grade E	1
	Sensitivity	1	Grade F	1