BSc in Computer Science

Learning Analytics

Discovering and understanding different learning patterns

Benjamín Aage B. Birgisson
Þorgeir Kristján Þorgeirsson

May 2020

Learning Analytics

Discovering and understanding different learning patterns

Thesis of 12 ECTS credits submitted to the School of Computer Science
at Reykjavik University in partial fulfilment
of the requirements for the degree of
Bachelor of Science (B.Sc.) in Computer Science

Examining Committee:

Dr. Anna Sigríður Islind, Supervisor
Assistant Professor, Reykjavik University, Iceland

Dr. María Óskarsdóttir, Examiner
Assistant Professor, Reykjavik University, Iceland

Prologue

This is a 12 ECTS credited final project, submitted to the Department of Computer Science at Reykjavik University, in partial fulfilment of the requirements for the degree of Bachelor of Science (B.Sc.) in Computer Science. The project is a joint project with the Department of Computer Science at Reykjavik University, Iceland. Code is expected to be released as open source, with the intent of facilitating other research in this area, unless a clear commercial opportunity presents itself that the developers wish to pursue.

The main subject of this thesis is to discover and understand the different learning patterns and behaviours of students studying for their bachelor's degree within the Department of Computer Science at Reykjavik University, through analysing data gathered from the learning management system Canvas, as well as data gathered through a quantitative online questionnaire.

The supervisor of this project was Anna Sigríður Islind, assistant professor at the Department of Computer Science within Reykjavik University. The examiner was María Óskarsdóttir, assistant professor at Reykjavik University. A special thank you to Inga Kristín Kristjánsdóttir, who assisted with extracting the click-logs from Canvas. Also, a special thank you to Jean-Claude Lemmens, for granting permission to use parts of his previously designed questionnaire.

The authors of this thesis would also like to give acknowledgement to their friends and families, whom have given extraordinary encouragement and motivation to strive for success throughout the years, and to achieve the set goals in becoming computer scientists.

Abstract

The subject of this thesis is learning analytics in higher education, and different learning patterns and behaviours that can be seen as contributing factors when striving for good results in a course. Ultimately establishing an understanding regarding successful learning patterns and behaviour in higher education. The research questions are: 1. How do students in higher education use a learning management system as a part of their learning process, 2. Which patterns and behaviour contribute the most towards a high grade?

To answer the research questions this thesis is based on an analysis of a quantitative questionnaire (based on 273 responses), descriptive statistics as well as machine learning of click-logs generated from the learning management system Canvas, by analysing learning patterns, in connection to the final grade of students. The study included data from six undergraduate courses taught in spring and autumn of 2019 within the Department of Computer Science at Reykjavik University.

The main findings show the importance of defining the right variables when predicting final grades instead of redistributing the grade categories. The findings also indicate the importance of students remaining active throughout their courses and that through continuous engagement, high grades can be achieved.

Table of Content

Prologue ... 3

Abstract ... 4

List of Figures ... 7

List of Tables ... 8

1 Introduction ... 9

2 Related Work ... 11

2.1 Lecture Capture as Part of Learning Management Systems ... 11

2.2 Assessment Types in Higher Education ... 12

2.3 Academic Maturity ... 13

2.4 Motivations and Results ... 14

3 Methodology ... 16

3.1 Quantitative Study ... 16

3.1.1 Instruments ... 16

3.1.2 Participants ... 18

3.1.3 Data Processing ... 18

3.2 Data Analysis from Canvas ... 19

3.2.1 Data Processing ... 19

4 Results ... 20

4.1 Questionnaire Results ... 20

4.1.1 Institutional Support ... 20

4.1.2 Educational Values ... 22

4.1.3 Goals ... 23

4.1.4 Self-Efficacy ... 28

4.1.5 Academic Apathy ... 33

4.1.6 University Related ... 35

4.1.7 Social Networks ... 37

4.1.8 Other Interesting Factors ... 39

4.1.9 Main Findings from the Questionnaire ... 43

4.2 Descriptive Statistics ... 45

4.2.1 Discrete Mathematics I ... 45

4.2.2 Software Requirements and Design ... 50

4.2.3 Calculus and Statistics ... 54

4.2.4 Computer Networks ... 59

4.2.5 Data Structures ... 62

4.2.6 Databases ... 67

4.2.7 Main Summary of Descriptive Statistics 71
4.3 Machine Learning 73
4.3.1 Original Grade Categories 80
4.3.1.1 Accuracy Results 80
4.3.1.2 Feature Importance 81
4.3.1.3 Summary Results 83
4.3.2 Modified Grade Categories 85
4.3.2.1 Accuracy Results 85
4.3.2.2 Feature Importance 86
4.3.2.3 Summary Results 89
5 Discussion 91
5.1 Limitations 92
5.2 Future Work 93
6 Conclusion 94
Bibliography 95
Appendix I – Quantitative Research Questionnaire 99

List of Figures

Figure 1. Attitudes and motivations in study orientations 15
Figure 2. Discrete Mathematics - Canvas frequency 46
Figure 3. Discrete Mathematics I - Usage of Canvas 47
Figure 4. Discrete Mathematics I - Time of usage 48
Figure 5. Discrete Mathematics - Usage per week 49
Figure 6. Discrete Mathematics - Time of usage per week 50
Figure 7. Software Requirements and Design - Canvas frequency 51
Figure 8. Software Requirements and Design - Usage of Canvas 51
Figure 9. Software Requirements and Design - Time of usage 52
Figure 10. Software Requirements and Design - Usage per week 53
Figure 11. Software Requirements and Design - Time of usage per week 54
Figure 12. Calculus and Statistics - Canvas frequency 55
Figure 13. Calculus and Statistics - Usage of Canvas 56
Figure 14. Calculus and Statistics - Time of usage 56
Figure 15. Calculus and Statistics - Usage per week 57
Figure 16. Calculus and Statistics - Time of usage per week 58
Figure 17. Computer Networks - Canvas frequency 59
Figure 18. Computer Networks - Usage of Canvas 60
Figure 19. Computer Networks - Time of usage 60
Figure 20.Computer Networks - Usage per week 61
Figure 21. Computer Networks - Time of usage per week 62
Figure 22. Data Structure - Canvas frequency 63
Figure 23. Data Structures - Usage of Canvas 64
Figure 24. Data Structures - Time of usage 65
Figure 25. Data Structures - Usage per week 65
Figure 26. Data Structures - Time of usage per week 66
Figure 27. Databases - Canvas frequency 67
Figure 28. Databases - Usage of Canvas 68
Figure 29. Databases - Time of usage 69
Figure 30. Databases - Usage per week 69
Figure 31. Databases - Time of usage per week 70
Figure 32. Comparison between course frequency of usage 71
Figure 33. Comparison between different usage of Canvas within courses 72
Figure 34. Comparison between time of usage within courses 72
Figure 35. Comparison between time of usage per week 73
Figure 36. Data preparation process 74
Figure 37. Depicting the Machine Learning Process. 79
Figure 38. Discrete Mathematics I (Original) - Positive Feature Importance Scores 81
Figure 39. Software Requirements and Design (Original) - Positive Feature Importance Scores 82
Figure 40. Calculus and Statistics (Original) - Positive Feature Importance Scores 82
Figure 41. Computer Networks (Original) - Positive Feature Importance Scores 83
Figure 42. Databases (Original) - Positive Feature Importance Scores 83
Figure 43. Total Positive Feature Importance Scores (Original) 84
Figure 44. Discrete Mathematics I (Modified) - Positive Feature Importance Scores 87
Figure 45. Software Requirements and Design (Modified) - Positive Feature Importance Scores 87
Figure 46. Calculus and Statistics (Modified) - Positive Feature Importance Scores 88
Figure 47. Computer Networks (Modified) - Positive Feature Importance Scores 88
Figure 48. Databases (Modified) - Positive Feature Importance Scores 89
Figure 49. Total Positive Feature Importance Scores (Modified) 90

List of Tables

Table 1. Categories of the questionnaire and number of questions in each 17
Table 2. Frequency – Institutional support 21
Table 3. Correlation - Institutional support 22
Table 4. Frequency - Educational values 22
Table 5. Correlation - Educational values 23
Table 6. Frequency - Goals 24
Table 7. Correlation – Goals 27
Table 8. Frequency - Self-Efficacy 28
Table 9. Correlation - Self-Efficacy 31
Table 10. Frequency - Academic Apathy 33
Table 11. Correlation - Academic Apathy 34
Table 12. Frequency - Reykjavik University related 35
Table 13. Correlation - Reykjavik University related 37
Table 14. Frequency - Social Network 38
Table 15. Correlation - Social Network 39
Table 16. Correlation between question 8 and 11 40
Table 17. Correlation between question 8 and 38 40
Table 18. Correlation between question 11 and 38 40
Table 19. Correlation between questions 9, 13 and 27 41
Table 20. Correlation between questions 9, 27, 30 and 32 42
Table 21. Correlation between questions 19, 21 and 31 42
Table 22. Correlation between questions 20 and 34 43
Table 23. Discrete Mathematics I - Total usage per weekday 49
Table 24. Software Requirements and Design - Total usage per weekday 53
Table 25. Calculus and Statistics - Total usage per weekday 58
Table 26. Computer Networks - Total usage per weekday 61
Table 27. Data Structures - Total usage per weekday 66
Table 28. Databases - Total usage per weekday 70
Table 29. Detailed description of course dataset variables 75
Table 30. Original Grade Categories in Discrete Mathematics I 78
Table 31. Modified Grade Categories in Discrete Mathematics I 78
Table 32. Original Grade Categories - Comparison and Accuracy Results 80
Table 33. Original Grade Categories - Best Accuracy Scores 80
Table 34. Modified Grade Categories - Comparison and Accuracy Results 85
Table 35. Modified Grade Categories - Best Accuracy Scores 86

1 Introduction

Preparing for higher education through university studies can be a complicated and stressful thing to do. Especially when you have no background in such studies. Students also display different learning behaviours which they have developed over time through their previous studies and work. This can of course be to their beneficial gain, as they can use these experiences to their advantage in their studies in higher education through applying useful learning patterns and behaviour. However, this can also be to their unprofitable loss, which can both be measured financially and in wasted time, as they fail to make use of available resources and continue to exhibit ineffective learning patterns and behaviours.

Since learning management systems (LMS) started entering higher education in the 1990s, their success factors have been examined in lengths (Al-Busaidi and Kamla, 2012). The discussion has both revolved around the non-technical and technical factors in learning management systems, while also discussing the way knowledge is shared through learning management systems, how students are affected by the use of LMS and on how teachers practice and overall pedagogy are affected by the use of LMS (Islind, Norström, Vallo Hult, Ramadani Olsson, 2020).

Recent studies have emphasized that due to the advances in the way data can now be captured and analysed, there are emerging possibilities of using massive amounts of educational data for the purpose of mapping different aspects of the way learning management systems affect students learning behaviour (Nguyen, Gardner and Sheridan, 2020). Due to these advances, there is an emerging area of research interest called learning analytics. The area of learning analytics revolves around personalization, adaptive learning, predictive analysis and user behaviour profiling (Brown, 2012; Peña-Ayala, 2018) and learning analytics are dependent on vast amount of data derived from learning management systems.

The potentials of learning analytics offer benefits both for lecturers, and for students to further understand their own perception of the way they approach their learning, while also letting data speak to the way they actually approach their learning (Nguyen et al., 2020). Teachers can then use the profiles, in order to adapt their teaching, so that it better fits with the student cohort that they have. Independent of the potentials of learning analytics, there is still a need to understand the motivation behind certain learning patterns that can be found in higher education.

The main purpose of this thesis is to analyse different learning patterns and behaviour of undergraduate students within the Department of Computer Science at Reykjavik University, in order to identify patterns that are more likely to result in higher grades. The main focus will therefore be to answer the following two research questions: 1. How do students in higher education use a learning management system as a part of their learning process? 2. Which patterns and behaviour contribute the most towards a high grade?

The structure of the thesis is as follows: Chapter two covers related work that is relevant for the topic of this thesis. In chapter three the data gathering methods used and data analysis conducted are discussed in detail, as well as the results of these analysis. In the end, the main subject of chapter four is the conclusion made by the authors of this thesis, where the main results are lead through discussions on the different styles of learning patterns, the limitations of the project, and further analysis that would be interesting to do in the near future.

2 Related Work

The main focus of this thesis is to research the learning behaviours of students within Reykjavik University in Iceland, to learn about the learning patterns of students in regard to their final grades, as well as their preparation for becoming scholars at university level. Therefore, the authors of this thesis explored researches made by others in this particular field of study. This chapter follows some of these researches and the conclusion of their studies.

2.1 Lecture Capture as Part of Learning Management Systems

Learning management systems enable another trend, that has been growing in interest within higher education, the trend of video-taping lectures. There are different reasons behind selecting to use lecture capture as a part of the study path. For instance, some students use lecture captures, rather than attending live lectures, to be able to spend more time with their families, work outside of school or to be able to focus on other courses in their study program (Chester, Buntine, Hammond and Atkinson, 2011). Students also use lecture captures as a backup for live lectures that were missed by accident or inevitable circumstances, for example due to illness or because of transportation issues (Yeung, Raju and Sharma, 2016). Students also use lecture capture to revisit material, for clarification on specific matters, or for complex or confusing topics (Yeung et al., 2016), and to study and prepare for exams (Mallinson and Baumann, 2015).

Edwards and Clinton (2018) discuss lecture capture and note that students may use lecture capture for different purposes: i) as a substitute for attending live lectures, ii) as a supplement to live classes, both attending live classes and using video lecture captures. Therefore, the question remains of whether attending live lectures has benefits that watching video lectures does not capture. Edwards and Clinton (2018) report a significant positive relationship between students attending live classes and higher final grades. In other words, the correlation means that the more a student attended live classes, the higher their final grades score were. Additionally, they found that the students who did well in their first year of study, attended more live classes during their second year of study, compared to those who did not do as well in their first year of study. The same correlation could be found with those who did well in their second year of study and their attendance in their third year of study (they were more likely to attend more live classes the next year), compared to those who did not do as well in their second year (Edwards and Clinton, 2018). Consequently, they found that lecture capture usage did not benefit students in achieving higher grades. Additionally, the students who did not attend live classes were more likely to get lower grades, regardless of their usage of lecture captures. However, the students who attended live lectures were more likely to achieve higher grades, regardless of their usage of lecture captures.

O ́Callaghan, Neumann, Jones and Creed (2017) similarly conducted a review on the use of lecture recordings in higher education. Their main focus was on lecture recordings used alongside on-campus face-to-face teaching, and not as a substitute of any kind. In their study they found that the advantages of lecture recordings outweigh the identified disadvantages found through their study and provide their recommendation of continued use of lecture capture in higher education as a supplement to live lectures and highlight the importance of lecture capture not being seen as a substitute.

Overall, students prefer and value face-to-face lectures. The reason being more interaction, better engagement, and to maintain better attention (Jensen, 2011). Thereby, students should be advised to attend live lectures while in parallel utilizing recordings of lectures, providing students with the option to choose, instead of just offering one option or the other (Nordmann and McGeorge, 2018). Additionally, students should be advised to practice the skills of studying with consistent pace throughout the term, instead of watching a collection of lectures in a short sprint of time, which can lead to a downfall in grades (Nordmann and McGeorge, 2018). As discussed, the focus should not be on whether or not lectures should be recorded, but rather how they can be used to better benefit students in their learning behaviours and techniques (Nordmann and McGeorge, 2018).

2.2 Assessment Types in Higher Education

Assessment is considered to be amongst the most important factors of higher education as students' learning outcomes are deeply influenced by the assessment types and overall structure (Gibbs and Simpson, 2004). Furthermore, the assessment types may be divided into formative assessment and summative assessment. The former contributes to student learning by providing information regarding performance (Yorke, 2003).

In regard to higher education, formative assessment is generally conducted through frequent low-stakes assignments, contributing little to no points towards a student's actual final grade score. The main purpose is to give students continuous feedback which can be used to identify strengths and weaknesses. As opposed to formative assessment, summative assessments are generally high-stakes and often less frequent (Dixson and Worrell, 2016). Examples of summative assessments are final exams, mid-term exams and term papers, with a high point value towards a student's final grade score. The main goal of summative assessments is to obtain a final assessment of how much learning has taken place (Gardner, 2010), evaluating student learning by comparing it against relevant standards or benchmarks.

2.3 Academic Maturity

Academic maturity is defined by Althoff (2010) as one's ability to understand his/her academic strengths and weaknesses and apply "effective strategies" in order to fully utilise learning opportunities. This can be divided into two categories, subject based maturity and non-subject based maturity. Yani, Harding and Engelbrecht (2019) strived to provide measurements of changes in academic maturity by using an adapted version of the Student Academic Readiness Questionnaire (STARS), developed by Lemmens (2010). As a part of measuring academic maturity, Yani et al. (2019) more specifically discuss non-subject based maturity.

Non-subject based maturity refers to personality trades and learning discipline. Non-subject based maturity is further divided into i) organisational maturity, which focuses on a person's planning, engagement and test taking skills, and ii) personal maturity, which focuses on a person's locus of control, self-efficacy, sociability, leadership, family support and general well-being (Yani et al. 2019).

The breakdown of non-subject based maturity, discussed by Yani et al. (2019), played an important role in the creation of the questionnaire conducted in this thesis, with its goal being to measure non-subject based maturity of students studying computer science at Reykjavik University. Relating to the results reported by Yani et al. (2019) they revealed a "surprising decline" in how students perceived their abilities over the course of the first term.

These results may well be the consequences of students becoming increasingly aware of the reality that is university level studying and coming to terms with their actual abilities. Yani et al. (2019) argued that this realisation was in fact a sign of growth in academic maturity.

2.4 Motivations and Results

The grade a student gets, is highly dependent on understanding the motivations behind that particular person's attitude towards their topic. Neroni et al. (2018) explores the relationship between goal orientation and academic performance. This topic has been researched extensively and the results show a direct relationship between a person's goal orientation and their respective academic performance (Huang, 2012). It is however stated that this topic has not yet been adequately explored within the confines of understanding the way students in higher education learn, and how the use of learning management system influences their grades (Neroni et al., 2018). In addition to that, the changes in learning management systems, and the support that students nowadays can get through learning management systems has changed extensively in the recent years.

A persons' goal orientation can be described as their motives to reach a certain goal (Dweck, 1986). In the context of education, goal orientation is further categorized as mastery orientation and performance orientation (Dweck, 1986; Nicholls, 1984). Students who display mastery orientation focus on developing knowledge and skill whereas students displaying performance orientation focus on doing better than others (Dweck, 1986; Nicholls, 1984). As a contrast, in performance orientation, people can either focus on doing better than others, displaying performance approach, or avoid doing worse than others, displaying performance avoidance (Elliot and McGregor, 2001). This approach-avoidance distinction can also be applied to mastery orientation where people either focus on developing their skill, displaying mastery approach, or focus on avoiding loss of knowledge and skills (Elliot and McGregor, 2001). More recently a fifth concept called work avoidance orientation was introduced to the construct of goal orientation, where people strive to maximize success through minimum effort (Harackiewicz, Durik, Barron, Linnenbrink-Garcia, and Tauer, 2008).

Distinctions have been made in the case of approach and avoidance within the performance goal orientation. The findings clarified that performance avoidance goal orientation relates negatively towards learning outcomes while performance approach goal orientation displayed positive relation. However, even though the way the mastery approach effects academic excellence is a notion that has been discussed in lengths, there is no consensus yet in whether there is a positive relationship (Chen, 2015; Abd-El-Fattah and Patrick, 2011), and in fact, most studies fail to prove significance in regard to the relationship (Neroni et al., 2018). In addition to that, these studies are not from within higher education. However, the fifth and final category shows the most significance when correlated with final grades. Students that apply the work avoidance mindset to their studies, generally get lower grades as a result (Brdar, Rijavec and Loncaric, 2006; Harackiewicz et al., 2008; King and McInerney, 2014). However, the understanding on the effects within higher education is less explored (Neroni, et al., 2018).

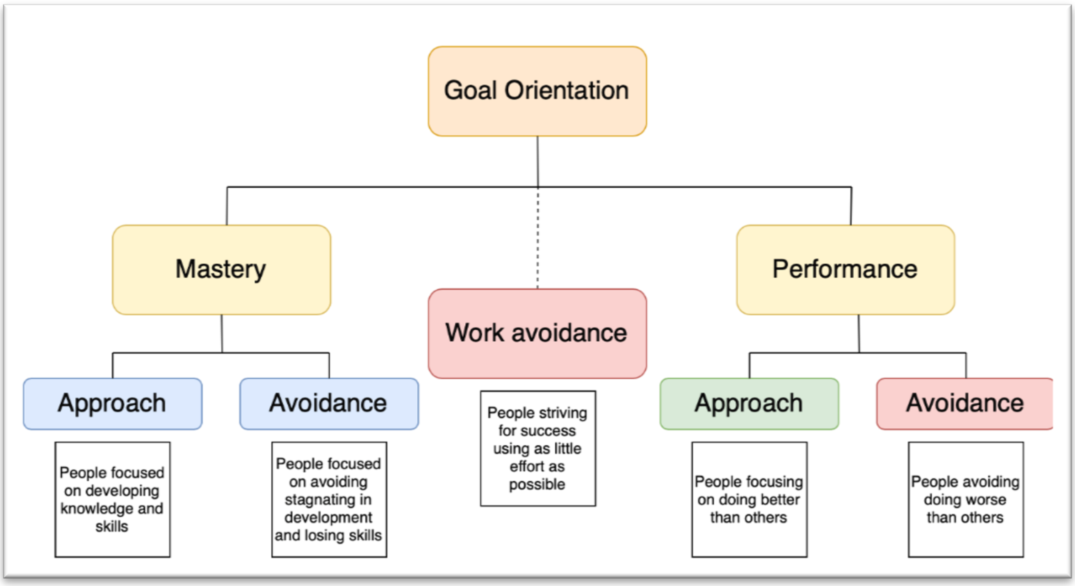

Figure 1. Attitudes and motivations in study orientations

There is thereby a gap in the literature in regard to higher education, and in regard to understanding the relationship between the different attitudes and how they affect grades. The relationship between these different attitudes is illustrated in Figure 1 here above, where it is shown how the different approaches relate to each other and the difference between them.

The following chapter discusses the methodology of the research methods used within this project, to assess the learning behaviours of students in higher education, as well as the results that these methods provided.

3 Methodology

The research of this thesis is divided into two parts, in order to assess learning behaviours and its relation to grades. As a first step, a quantitative online questionnaire was created to research the learning patterns and behaviours of students. For the second step, data was gathered through the learning management system Canvas, where click-logs were gathered from six undergraduate courses within the Department of Computer Science, to see through actual data what the learning patterns and behaviour of different students are. Through machine learning approach these click-logs were analysed in order to identify defining factors that play an important role towards achieving higher grades.

3.1 Quantitative Study

In a quantitative study, the research focuses on the use of numerical information and measurements, which then can be tested with hypotheses. The main focus is then to analyse the frequency of data gathered, the distribution of data and the reason behind these results (Blumberg, Cooper and Schindler, 2005).

The design of the questionnaire used for the quantitative study can be found in Appendix I. The parts used from the questionnaire of Lemmens (2010) were: Institutional Support, Educational Values, Goals, Self-Efficacy, and Academic Apathy. Some of the questions were as well adapted to Reykjavik University in particular, as this study focuses on the learning behaviours of students within Reykjavik University.

3.1.1 Instruments

The questionnaire was designed electronically using Google Docs. The final questionnaire can be found in Appendix I. The questions where 42 in total, which were divided into eight categories (see Table 1).

Table 1. Categories of the questionnaire and number of questions in each

Category	No. of questions
Individual background	2
Institutional support	3
Educational values	3
Goals	10
Self-efficacy	10
Academic apathy	4
University related	7
Social network	3

The first category contained two individual background questions inquiring participants on their age and gender. The second category contained questions about institutional support, three questions in total inquiring participants about information gathering regarding their planned study within Reykjavik University, before enrolling at the university in their desired field. The third category contained three questions about educational values, gathering information about the perceived view of higher education. The fourth category contained questions about goals, 10 questions in total focusing on the personality, learning behaviours and goal setting of participants. The fifth category of the questionnaire contained questions regarding self-efficacy, with 10 questions in total focusing on the perceived view of the participants regarding confidence, personal intelligence, and perceived abilities towards their field of studies within Reykjavik University. The sixth category of the questionnaire included questions about academic apathy, with four questions in regard to individual drive to succeed and their desired need to put in the work to achieve the best education possible or to not put in the work, and only do the minimal. The seventh category of the questionnaire contained questions about Reykjavik University. This category included seven questions in total gathering information about the view of participants in regard to expectation and reality with learning within the university, the teaching methods used, how the student participates in class, their attendance of lectures and study sessions of courses, and if participant has achieved honours in their study at the university. The eighth and final category of the questionnaire focused on the social networks of participants, three questions acquiring about if the participants prefer to work alone or within groups, and if their social network within the university was large or not (if they have the support needed or wanted within the university).

All the questions, excluding questions 1 and 2, which inquired about age and gender along with question 39, inquiring if the participant had reach the dean's list within Reykjavik University, an honour for exceptional grades, were in the form of scalable answer options: Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree. Participants could choose only one answer option.

3.1.2 Participants

The quantitative study was sent out to all undergraduate students within the Department of Computer Science at Reykjavik University via email notification. The email list includes a total of 678 individuals. The incentive being that participants of the questionnaire would have the opportunity to win a brand-new smartwatch. Altogether, there were 273 students who answered the questionnaire, 178 male students and 92 female students, as well as three students who did not want to disclose their gender as either male or female (non-binary). Therefore, the response rate of the questionnaire was 40.27% approximately.

The youngest participant of the questionnaire was 19 years old, the eldest one was 51 years old and the mean value of participants were 26.22 years old with the median value being 24 years old. Most participants were within the age group of 20-29 years old, or 71,43% participants (195 individuals). Participants 19 years old (or younger) where 15 individuals (5,49%), 41 participants were within the age group of 30-39 years old (15,02%), 15 participants were within the age group of 40-49 years old (5,49%), and two participants were within the age group of 50-59 years old. Five participants did not disclose their age (1,83%).

3.1.3 Data Processing

To analyse the data gathered through the online questionnaire, the data analysis tool of R and RStudio were used. Statistical calculations were made, where descriptive statistics were examined regarding frequency of responses, mean values, standard deviations, correlation between variables (questions), and cross tables to explore possible connections between named variables. Microsoft Excel was used as well, encoding responses from Google Docs before eventually transferring them into RStudio, and for making data tables with frequency of responses.

3.2 Data Analysis from Canvas

In order to analyse data from the learning management system Canvas, click-log information was required. These logs were gathered from Reykjavik University, for the six aforementioned courses, that were being taught in the terms of autumn and fall 2019. These courses, where the logs were gathered from, are the following: Discrete Mathematics I (first year, first term), Software Requirements and Design (first year, first term), Data Structures (first year, second term), Databases (first year, second term), Calculus and Statistics (second year, first term), and Computer Networks (third year, first term).

In total there were 940,523 lines of click-logs, where each and every log included the following: student ID, session ID, timestamp, IP address, and URL. Data was also gathered from the instructors of the six courses, containing the final score of students within the course. In so doing, machine learning could be conducted to analyse the data further in terms of predicting a student final score within a given course by their activity.

3.2.1 Data Processing

As the data retrieved from Canvas was all encrypted and password protected, due to personal information, the encryption software of VeraCrypt was used to access the encrypted data. For analysing and cleaning up the then decrypted data, the software environment for statistical computing of RStudio and R was used, where RStudio is an integrated development environment (IDE), and R is a programming language for statistical calculations.

To analyse further the activity of each individual for each course, the source-code editor of Visual Studio Code was used along with the programming language Python. Finally, Azure Machine Learning was used to apply machine learning to the course data sets, in order to predict the final grades of students through their activities within Canvas.

4 Results

The following chapters contain the results of analysis through quantitative study and data analysis with click-logs from the learning management system Canvas. The results are divided as follows: Questionnaire Results, Descriptive Statistics, Machine Learning.

4.1 Questionnaire Results

When analysing the results of the online questionnaire, the main focus was to dissect the frequency of responses, the mean value of responses (the sum of values divided by the number of values), median value of responses (the middle value of responses, value that separates the higher half from the lower half of responses), the standard deviation of each question (variation amongst the set of responses), the correlation (statistical relationship between two variables) between questions of each section of the questionnaire, and analyse some interesting cross tables (frequency distribution of variables, their interrelation between the two variables).

Following are the results of these above-mentioned factors, divided into seven different sections, excluding the background questions. These sections are in accordance with the questionnaire sections, which are: Institutional Support, Educational Values, Goals, Self-Efficacy, Academic Apathy, University Related, and Social Network.

4.1.1 Institutional Support

The first section covered questions regarding "Institutional Support" and included three questions inquiring participants about information gathered about their study within Reykjavik University, before enrolling. The frequency of answers can be seen in Table 2 here below.

Table 2. Frequency – Institutional support

3. I had enough information about Reykjavik University before enrolling.		
Strongly Agree	76	27,84%
Agree	153	56,04%
Neutral	40	14,65%
Disagree	3	1,10%
Strongly Disagree	1	0,37%

4. I acquired information about my degree programme before I enrolled at Reykjavik University.		
Strongly Agree	105	38,46%
Agree	135	49,45%
Neutral	24	8,79%
Disagree	8	2,93%
Strongly Disagree	1	0,37%

5. I was informed about the career possibilities for a specific degree programme.		
Strongly Agree	72	26,37%
Agree	122	44,69%
Neutral	56	20,51%
Disagree	21	7,69%
Strongly Disagree	2	0,73%

In question three, a majority of the participants (83.88% or 229 individuals out of the 273 participants) perceived to have enough information about the university before enrolling, versus 1.47% (4 individuals) who did not feel as if they had acquired sufficient information about the university before enrolling. About 14.65% or 40 individuals were neutral about the question, where the mean value of responses was 4.099, the median value being 4.0 (agree), and the standard deviation being 0.70275.

In question four, most of the participants had also acquired information about the degree programme before enrolling at the university (87.91% or 240 individuals) versus 3.30% (9 individuals) who had not acquired information about their degree before enrolment within the university. Around 8.79% or 24 individuals were neutral about the statement of the question, with the mean value of responses being 4.227, the median value being 4.0 (agree), and a standard deviation of 0.757.

In question five, in regard to informing themselves about possible careers after their specified learning programme, 71.06% of participants agreed or strongly agreed that they had done so (194 individuals), while 8.43% did not (23 individuals) and 56 participants answered the question as neutral (or 20.51%), with mean value being 3.883, the median value being 4.0 (agree), and the standard deviation being 0.912.

Table 3. Correlation - Institutional support

	Q3	Q4	Q5
Q3	1.000	0.3860	0.2820
Q4	0.3860	1.000	0.3741
Q5	0.2820	0.3741	1.000

In regard to the correlation between the three questions within this section (see Table 3), in accordance to Rumsey (ed.), they display a moderately weak positive linear relationship, with correlation being ±0.2820 and ±0.3860, meaning that the answers between participants are scattered between the possible answer options with little or no relevance.

4.1.2 Educational Values

Section number three of the questionnaire focused on "Educational Values" and contained three questions (question 6-8) about the perceived view of higher education amongst participants. The frequency of answers for the three questions can be found in Table 4 here below.

Table 4. Frequency - Educational values

6. It is important to always be prepared for class.		
Strongly agree	61	22,34%
Agree	137	50,18%
Neutral	61	22,34%
Disagree	13	4,76%
Strongly disagree	1	0,37%
7. It is important to have a good university education to be successful in life.		
Strongly agree	27	9,89%
Agree	71	26,01%
Neutral	86	31,50%
Disagree	56	20,51%
Strongly disagree	33	12,09%
8. Getting good grades is important to me.		
Strongly agree	70	25,64%
Agree	149	54,58%
Neutral	37	13,55%
Disagree	13	4,76%
Strongly disagree	4	1,47%

In question six, most of the participants, or 198 individuals out of the 273 in total, were in an agreement of the importance of always being prepared for classes (72.52%), versus 14 individuals (5.13%) who were in disagreement. A total of 61 individuals, or 22.34% were neutral of it. The mean value was 3.8938, the median value being 4.0 (agree), and the standard deviation being 0.8133.

In question seven, there seems to be a division between individuals who perceive a good university education as playing an important role towards being successful in life, as 35.90% (98 individuals) agreed versus 32.60% (89 individuals) who considered it irrelevant. A total of 31.50%, or 86 persons, were neutral of the statement. The mean value was 3.0111, the median value was 3.0 (neutral), and the standard deviation being 1.1615.

The third and last question in this section of the questionnaire, question eight, was about the personal importance of getting good grades, where most of the participants perceived it as important as 80.22% agreed with the statement (219 individuals), and only 6.23% were in disagreement (17 individuals). A total of 13.55% were neutral about it (37 individuals). The mean value was 3.9817, the median value was 4.0 (agree), and the standard deviation was 0.8465.

Table 5. Correlation - Educational values

	Q6	**Q7**	**Q8**
Q6	1.000	0.1297	0.1360
Q7	0.1297	1.000	0.2283
Q8	0.1360	0.2283	1.000

In regard to the correlation between the three questions (see Table 5), it can be seen that there is a positively weak correlation (Rumsey, ed.) amongst them, with a correlation factor between ±0.1297 and ±0.2283. This correlation can be interpreted as such, that an answer to one of the three questions has little to no effect on the answers of the other two questions in this section of the questionnaire.

4.1.3 Goals

Section number four of the questionnaire included 10 questions in total about individual "Goals" of participants when it comes to higher education. The frequency of answers can be seen in Table 6 here below.

Table 6. Frequency - Goals

9. I'm a very methodical (ísl. skipulagður/skipulögð) person.		
Strongly agree	40	14,65%
Agree	96	35,16%
Neutral	77	28,21%
Disagree	51	18,68%
Strongly disagree	8	2,93%
(blank)	1	0,37%

10. I set specific goals before I begin studying for tests/exams.		
Strongly agree	50	18,32%
Agree	112	41,03%
Neutral	71	26,01%
Disagree	37	13,55%
Strongly disagree	3	1,10%

11. Good grades provide me with an excellent goal to work towards.		
Strongly agree	61	22,34%
Agree	135	49,45%
Neutral	59	21,61%
Disagree	17	6,23%
Strongly disagree	1	0,37%

12. I like to have a routine (ísl. vanagangur / venja) to follow.		
Strongly agree	100	36,63%
Agree	125	45,79%
Neutral	31	11,36%
Disagree	17	6,23%
Strongly Disagree	0	0,00%

13. I organize my study time to best achieve my goals.		
Strongly agree	40	14,65%
Agree	120	43,96%
Neutral	74	27,11%
Disagree	37	13,55%
Strongly disagree	2	0,73%

14. I prefer to be spontaneous (ísl. hvatvís) rather than to set goals when I study for tests/exams.		
Strongly agree	16	5,86%
Agree	67	24,54%
Neutral	61	22,34%
Disagree	110	40,29%
Strongly disagree	19	6,96%

15. I usually double check things, just to make sure they are correct.		
Strongly agree	87	31,87%
Agree	139	50,92%
Neutral	30	10,99%
Disagree	16	5,86%
Strongly disagree	1	0,37%

16. I know what I want to be doing 10 years from now.		
Strongly agree	31	11,36%
Agree	78	28,57%
Neutral	56	20,51%
Disagree	69	25,27%
Strongly disagree	39	14,29%

17. I have clear and reachable/realistic goals for my studies this year.		
Strongly agree	63	23,08%
Agree	147	53,85%
Neutral	42	15,38%
Disagree	15	5,49%
Strongly disagree	5	1,83%
(blank)	1	0,37%

18. I have talked about my career goals with someone who has worked in that field.		
Strongly agree	53	19,41%
Agree	93	34,07%
Neutral	42	15,38%
Disagree	56	20,51%
Strongly disagree	28	10,26%
(blank)	1	0,37%

In question nine, about 49.81% of the participants (136 out of 273 individuals) considered themselves to be very well organized and methodical individuals, versus 21.61% (59 individuals) who did not consider themselves to be methodical and 28.21% (77 individuals) were neutral. One individual did not want to give an answer to the question. The mean value of answers being 3.401, the median value 3.5 (neutral/agree), and standard deviation being 1.0439.

In question 10, a larger half of the participants do set specific goals regarding their study for tests and/or exams, or 162 out of the 273 participants of the questionnaire (59.35%) versus those who do not set specific goals for tests and/or exams, a total of 40 participants (14.65%). There were then 71 participants (26.01%) who were neutral of the statement. The mean value for this particular question was 3.619, the median value was 4.0 (agree), and the standard deviation was 0.9709.

Question number 11 focused on good grades as a motivation or goal to work towards in the participants study, were 93.41% of participants were either positive or neutral of it (196 individuals agreeing or strongly agreeing, and 59 being neutral), with only 6.6% of participants (18 individuals) disagreeing with the statement. The mean value of the question was 3.878, the median value being 4.0 (agree), and the standard deviation being 0.8413.

Topic of question 12 covered if the participants like following a prepared routine. A majority of participants answered positively (225 individuals, or 82.42%), stating that they prefer to have a routine, versus participants whom answered negatively (17 individuals, or 6.23%). A total of 11.36% of the participants were neutral of the statement (31 individuals). The mean value for the question was 4.1282, the median value 4.0 (agree), and standard deviation being 0.8457.

Question number 13, which stated that the participants organized their study time to best achieve their goals, had a positive answer rate (strongly agreeing or agreeing) of 58.61% or 160 individuals out of the 273 participants in total. About 14.28% of participants, or 39 individuals, did not agree or strongly disagreed on the statement, stating that they did not organize their study time. The rest of the participants where neutral towards being organized, or 27.11% of participants (74 individuals). The mean value of the question was 3.5824, the median value being 4.0 (agree), and the standard deviation being 0.9243.

Question number 14 stated the preference of participants to be spontaneous in their studies, rather than setting goals in preparation for tests/exams. Here, the answers given show that 30.40% of participants (83 individuals) prefer to be spontaneous in their studies for tests and/or exams, versus 47.25% (129 individuals) who do not like this approach. Around 22.34% of participants (61 individuals) where indifferent with the statement. The mean value was 2.8205, the median value was 3.0 (neutral), and the standard deviation was 1.0645.

Question number 15 inquired if participants double checks things in their studies, to make sure they are correct. A high majority of participants, or 82.79% (226 individuals) were in agreement with the statement, while only 6.23% (17 individuals) were in disagreement. Those who answered the question neutral were 10.99% of the total number of participants, or 30 individuals. The mean value was 4.0806, the median value was 4.0 (agree), and the standard deviation was 0.8319.

In question 16, regarding future plans of participants in the next coming 10 years, about 39.93% of participants (109 individuals) agreed that they knew what they wanted to be doing in 10 years from answering the questionnaire, and 39.56% (108 individuals) did not know what they wanted to be doing in 10 years' time. Neutrals were 20.51%, or 56 individuals. The mean value for this question was 2.9744, the median value was 3.0 (neutral), and the standard deviation was 1.2527.

Question number 17, asking participants if their goals for their studies this year where clear and reachable/realistic, had 76.93% (210 individuals) in agreement that their goals where achievable this coming study year, versus 7.32% (20 individuals) of whom did not see their goals as clear/reachable/realistic for the coming study year. 42 participants were neutral of the question, and one participant did not want to answer the question, a total of 15.75%

of the total participants. The mean value here was 3.9118, the median value was 4.0 (agree), and the standard deviation was 0.8758.

Question 18, which was also the last question for this fourth section of the questionnaire, asked if the participants had talked about their career goals with someone who has worked in that particular field. There were 146 individuals, or 53.48% of participants that had done so, and 84 individuals (30.77% of participants) had not talked to anyone within the working field of their career goals. There were 42 persons (15.38% of participants) who answered the question as neutral, and one individual (0.37% of participants) did not want to answer the question. The mean value was 3.3199, the median value was 4.0 (agree), and the standard deviation was 1.2817.

Table 7. Correlation – Goals

	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18
Q9	1.000	0.4664	0.1974	0.3239	0.5632	-0.4311	0.2465	0.1568	0.0852	0.0808
Q10	0.4664	1.000	0.3187	0.3025	0.4499	-0.4611	0.1501	0.1079	0.1765	0.1272
Q11	0.1974	0.3187	1.000	0.2878	0.2318	-0.2119	0.1948	0.1492	0.1163	0.0563
Q12	0.3239	0.3025	0.2878	1.000	0.3644	-0.2388	0.1325	0.1159	0.0574	0.0531
Q13	0.5632	0.4499	0.2318	0.3644	1.000	-0.4319	0.0705	0.1255	0.2174	0.1895
Q14	-0.4311	-0.4611	-0.2119	-0.2388	-0.4319	1.000	-0.1665	0.0225	-0.0936	0.0156
Q15	0.2465	0.1501	0.1948	0.1325	0.0705	-0.1665	1.000	-0.0127	0.0665	0.0477
Q16	0.1568	0.1079	0.1492	0.1159	0.1255	0.0225	-0.0127	1.000	0.3577	0.4558
Q17	0.0852	0.1765	0.1163	0.0574	0.2174	-0.0936	0.0665	0.3577	1.000	0.2598
Q18	0.0808	0.1272	0.0563	0.0531	0.1895	0.0156	0.0477	0.4558	0.2598	1.000

With regards to the correlation between questions, there were some interesting things to be found (see Table 7). For example, there was moderate correlation to be found (Rumsey, ed.) between questions 9, 10, 13 (positive correlation amongst each other) and question 14 (negative correlation to the other three questions). This makes sense, as all of these questions relate to the mindset of the person answering. If a person is very methodical (question 9), the same person is likely to set specific goals before studying (question 10), and organizes his/her study time to achieve these goals (question 13), therefore not being spontaneous in his/her studies (question 14).

The same goes with moderately positive correlation between question 16 and question 18, where if the individual answering the questionnaire knows what he/she wants to be doing in 10 years' time, he/she has most likely talked about possible career goals with someone who has worked in that field.

4.1.4 Self-Efficacy

Section five of the questionnaire included 10 questions about the self-efficacy of participants, where the frequency of answers can be seen in Table 8 here below. In regard to self-efficacy, the main goal was to inquire individuals on their personally perceived ability of reaching their goals.

Table 8. Frequency - Self-Efficacy

19. I expect to have a harder time to perform academically (ísl. fræðilega) than most students here.		
Strongly agree	26	9,52%
Agree	56	20,51%
Neutral	73	26,74%
Disagree	79	28,94%
Strongly disagree	39	14,29%

20. I can easily adapt to different styles of teaching.		
Strongly agree	35	12,82%
Agree	141	51,65%
Neutral	67	24,54%
Disagree	29	10,62%
Strongly disagree	1	0,37%

21. I am as skilled academically (ísl fræðilega) as the best students here.		
Strongly agree	18	6,59%
Agree	70	25,64%
Neutral	84	30,77%
Disagree	85	31,14%
Strongly disagree	16	5,86%

22. I enjoy working on complicated, intellectually-demanding (ísl. vitsmunalega krefjandi) problems.		
Strongly agree	74	27,11%
Agree	141	51,65%
Neutral	50	18,32%
Disagree	8	2,93%
Strongly disagree	0	0,00%

23. I know what I want and I usually make sure that I get it.		
Strongly agree	48	17,58%
Agree	132	48,35%
Neutral	74	27,11%
Disagree	17	6,23%
Strongly disagree	1	0,37%
(blank)	1	0,37%

24. I have the ability to plan my work (study time).		
Strongly agree	41	15,02%
Agree	139	50,92%
Neutral	53	19,41%
Disagree	35	12,82%
Strongly disagree	5	1,83%

25. I expect to do very well in my degree.		
Strongly agree	55	20,15%
Agree	136	49,82%
Neutral	73	26,74%
Disagree	5	1,83%
Strongly disagree	2	0,73%
(blank)	2	0,73%

26. I am quick to understand new concepts (ísl. hugtök) and ideas.		
Strongly agree	53	19,41%
Agree	145	53,11%
Neutral	56	20,51%
Disagree	18	6,59%
Strongly disagree	0	0,00%
(blank)	1	0,37%

27. I can motivate (ísl. hvetja) myself to study when I need to.		
Strongly agree	53	19,41%
Agree	110	40,29%
Neutral	57	20,88%
Disagree	45	16,48%
Strongly disagree	8	2,93%

28. I learn things more quickly than most people.		
Strongly agree	28	10,26%
Agree	80	29,30%
Neutral	108	39,56%
Disagree	41	15,02%
Strongly disagree	16	5,86%

In question 19, the frequency of participants believing that they would have as equally a hard time as other students in their upcoming studies were 43.23% (118 out of 273 individuals) versus 30.03% (82 individuals) who did not believe they would have as hard of a time. A total of 26.74% (73 individuals) were indifferent to the question. The mean value of answers was 2.8205, the median value was 3.0 (neutral), and the standard deviation was 1.1916.

In question 20, the inquiry was made if participants are able to adapt easily to different teaching methods. Here, a large part of participants believe that they have the capability to adapt easily to different teaching styles, with 64.47% (176 out of 273 individuals) agreeing with the statement, versus 10.99% (30 individuals) who believe that they are not capable of doing so, and 24.54% (67 individuals) were neutral about the statement. The mean value of answers was 3.6593, the median value was 4.0 (agree), and the standard deviation was 0.8474.

In question 21 of the questionnaire, participants were asked about their personal judgement of their own academical skills in comparison with other students. Around one-third of participants believe themselves to be as skilled academically as the best students in their chosen field, or 32.23% (88 out of 273 participants). However, 37.00% (101 individuals) of the participants found themselves to be not as skilled as the best students around, and

30.77% (84 individuals) were neutral of the statement. The mean value of question 21 was 2.9597, the median value was 3.0 (neutral), and the standard deviation was 1.0336.

In question 22, the inquiry was about participants enjoyment in taking on complicated and intellectually demanding tasks. Significantly many participants agreed upon the statement, or 78.76% of participants (215 out of 273 individuals). And only 2.93% (8 individuals) do not enjoy taking on such tasks. Around 18.32% (50 individuals) were indifferent on the statement. The mean value of answers for question 22 was 4.0293, the median value was 4.0 (agree), and the standard deviation was 0.7568.

In question 23, the inquiry was regarding participants resilience in knowing what they want and making sure they achieve it. Here, as well, a significant amount, or 65.93% of participants (180 out of 273 individuals) agreed with the statement, versus 6.59% (18 individuals) who did not agree with the statement. About 27.11% of participants (74 individuals) were neutral, and one individual did not want to disclose an answer to the question. The mean value was 3.7684, the median value was 4.0 (agree), and the standard deviation was 0.8249.

For question 24, the inquiry was on planning, that is if the participants believed themselves to be able to plan their work (study time) in advance. Here, the ratio of answers was as follows: 65.94% (180 individuals) agreed upon being able to plan their work, 14.65% (40 individuals) did not agree upon being able to plan their work in advance, and 19.41% (53 individuals) were neutral of the statement. The mean value of answers was 3.6447, the median value was 4.0 (agree), and the standard deviation was 0.9482.

In question 25, the inquiry was about the participant's expectation within their study-field of interest. Here, a large number of participants expected to do very well in their degree, or 69.97% (191 out of 273 individuals), and only 2.56% (7 individuals) did not expect to do very well. Many of the participants were also neutral of the statement, or 26.74% (73 individuals), as well as 0.73% (2 individuals) who did not want to answer the question in hand. The mean value of the question was 3.8745, the median value was 4.0 (agree), and the standard deviation was 0.7740.

In question 26, the question regarded how quickly participants believe themselves to be able to understand new concepts and ideas. Again, a significantly large amount of people believes themselves to be quick in understanding new concepts and ideas, or 72.53% (198 out of 273 individuals) of participants, versus 6.59% (18 individuals) who believe themselves to need more time to adjust to new concepts and ideas. One individual (0.37% of participants) did not want to disclose an answer to the question, and 20.51% of participants (56 individuals) were neutral of the inquiry. The mean value was 3.8566, the median value was 4.0 (agree), and the standard deviation was 0.8045.

In question 27 of the questionnaire, the inquiry was on the self-motivation of participants, that is the self-discipline in studying when needing to do so, instead of postponing it. The frequency of participants agreeing with being able to motivate themselves to study when needed were 59.71% (163 out of 273 individuals), versus 19.41% (53 individuals) disagreeing with the statement. Around 20.88% (57 individuals) were neutral of the statement. The mean value of answers was 3.5678, the median value was 4.0 (agree), and the standard deviation was 1.0693.

For the last question in this section five of the questionnaire, question number 28, the statement was "I learn things more quickly than most people". The frequency of positive answers to the statement was 39.56% (108 individuals), versus 20.88% (57 individuals) who were negative towards the statement. About 39.56% (108 individuals) were indifferent. The mean value of answers was 3.2308, the median value was 3.0 (neutral), and the standard deviation was 1.0191.

Table 9. Correlation - Self-Efficacy

	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28
Q19	1.000	-0.3889	-0.3617	-0.2653	-0.1109	-0.2288	-0.3437	-0.3413	-0.1279	-0.4153
Q20	-0.3889	1.000	0.2679	0.2513	0.1999	0.2515	0.2697	0.3772	0.2558	0.3535
Q21	-0.3617	0.2679	1.000	0.2980	0.1981	0.1833	0.2722	0.3885	0.1809	0.4189
Q22	-0.2653	0.2513	0.2980	1.000	0.3183	0.1332	0.2651	0.4003	0.1896	0.3165
Q23	-0.1109	0.1999	0.1981	0.3183	1.000	0.3756	0.3978	0.2341	0.3544	0.1213
Q24	-0.2288	0.2515	0.1833	0.1332	0.3756	1.000	0.3721	0.1809	0.4348	0.1547
Q25	-0.3437	0.2697	0.2722	0.2651	0.3978	0.3721	1.000	0.3287	0.3749	0.2812
Q26	-0.3413	0.3772	0.3885	0.4003	0.2341	0.1809	0.3287	1.000	0.2533	0.6186
Q27	-0.1279	0.2558	0.1809	0.1896	0.3544	0.4348	0.3749	0.2533	1.000	0.1280
Q28	-0.4153	0.3535	0.4189	0.3165	0.1213	0.1547	0.2812	0.6186	0.1280	1.000

Table 9 here above shows all calculation made in regard to correlation between questions within the section of "Self-Efficacy". Evidently, there seems to be a moderate positive correlation between question 21 and question 28, as well as questions 22 and 26. Additionally, there seems to be a considerably high positive correlation between question 26 and question 28, as well as a moderately negative correlation between question 19 and question 28, and moderately positive correlation between question 24 and question 27.

In regard to the correlation between questions 21 and question 28, this makes much sense as the first question states that the participant is as academically skilled as the best students around, and the second question states that the participant learns things at a faster rate than most people. Therefore, if a participant believes himself/herself to be academically skilled as the best students around, then he/she is likely to learn things more quickly than most people, in terms of personal opinion.

The same can be said about the correlation between questions 22 and question 26, which do not come as much of a surprise as the first question states that the participant enjoys working on complicated and intellectually-demanding problems, and the second question states that the participant is quick to understand new concepts and ideas. Therefore, those who tend to learn new concepts quickly, also like taking on complicated and intellectually demanding problems.

As visible correlation is displayed between questions 21 and 28, as well as in questions 22 and 26, there is little surprise to see moderately positive correlation between questions 26 and 28 as well. Question 26 states that the participant is quick to understand new concepts and ideas, and question 28 states that the participant learns things more quickly than most people. It is however more of a surprise to see weak positive correlation between questions 21 and 22 in this regard, were question 21 states that the participant believes himself/herself to be skilled academically as the best students around, and question 22 states that the participant enjoys working on complicated and intellectually demanding problems. It would have been thought that as an addition to being skilled academically, you would want intellectually demanding problems. However, the results show that the correlation is only positively weak, meaning that this is indeed not the case with most of the participants who answered the questionnaire.

A moderately negative correlation is between questions 19 and question 28 of the questionnaire, which does not come as any surprise as the first question states that the participant expects himself/herself to have a harder time to perform academically than most students, and the second question states that the participant learns things more quickly than most people. As the correlation between the two questions is negative, the students who believe themselves to struggle academically, do not like taking on demanding and highly challenging problems.

At last we have the moderately positive correlation between question 24 and question 27, where question 24 states that the participant has the ability to plan his/her study time, and question 27 states that the participant can motivate himself/herself in studying when needed to. The correlation between the two questions would have considered as going hand-in-hand before answers to the questionnaire were gathered, and it does so with a moderately positive correlation factor.

4.1.5 Academic Apathy

Section six of the questionnaire, which included four questions regarding academic apathy, how participants organize their study and how much work they put into their study within Reykjavik University. The frequency of answers can be found in Table 10 here below.

Table 10. Frequency - Academic Apathy

29. I tend to study in spurts (ísl. sprettum) rather than at a regular steady pace.		
Strongly Agree	53	19,41%
Agree	97	35,53%
Neutral	66	24,18%
Disagree	51	18,68%
Strongly Disagree	6	2,20%

30. My goal is to get the best grade I can without making a lot of effort on my course work.		
Strongly Agree	13	4,76%
Agree	67	24,54%
Neutral	60	21,98%
Disagree	106	38,83%
Strongly Disagree	27	9,89%

31. I often don't see things through to the end.		
Strongly Agree	14	5,13%
Agree	70	25,64%
Neutral	56	20,51%
Disagree	102	37,36%
Strongly Disagree	31	11,36%

32. I plan my study sessions in advance and almost always follow my study plan.		
Strongly Agree	15	5,49%
Agree	64	23,44%
Neutral	80	29,30%
Disagree	89	32,60%
Strongly Disagree	25	9,16%

In the first question of this section, question 29, approximately 54.94% (150 individuals) agreed that they study in spurts rather than at a regular steady pace over the term, versus 20.88% (57 individuals) who disagreed and tend to study at a more consistent pace over the course of the term. Neutral answers were 24.18%, or 66 individuals out of the 273 participants of the questionnaire. The mean value was 3.5128, the median value was 4.0 (agree), and the standard deviation was 1.0714.

In question 30, it was stated that the goal of the student was to get the best grade without making a lot of effort in the course work, a total of 80 individuals (29.30% of participants) agreed, and 48.72% of participants (133 individuals) did not agree with the statement. Neutral answers were 21.98% in total, or 60 answers out of the total 273 gathered. The mean value was 2.7546, the median value was 3.0 (neutral), and the standard deviation was 1.0788.

For question 31, stating that the student seldom sees things through to the end, 48.72% of participants (133 individuals) did not agree with the statement, and 30.77% (84 individuals) agreed with the statement. About 20.51% (56 individuals) were neutral with the statement of the question. The mean value was 2.7582, the median value was 3.0 (neutral), and the standard deviation was 1.1115.

At last, for section six, was the statement of "I plan my study sessions in advance and almost always follow my study plan". Here, only 28.93% (79 individuals) agreed with the statement, versus 41.76% (114 individuals) who did not agree with the statement. There were 29.30% (80 individuals) neutral. The mean value was 2.8352, the median value was 3.0 (neutral), and standard deviation was 1.0599.

Table 11. Correlation - Academic Apathy

	Q29	Q30	Q31	Q32
Q29	1.000	0.3415	0.2867	-0.3203
Q30	0.3415	1.000	0.2600	-0.1802
Q31	0.2867	0.2600	1.000	-0.3023
Q31	-0.3203	-0.1802	-0.3023	1.000

As can be seen in Table 11 here above, the correlation between questions in this section of the questionnaire were not strong, and mostly with negative correlation between themselves as well. The reason most likely being that participants differentiate highly when it comes to planning their study sessions as well as setting and working on their set goals.

4.1.6 University Related

Section seven of the questionnaire covers seven questions specifically relating to Reykjavik University. This is done in regard to the teaching methods used at the university, how students are using the teaching methods and different services, the overall attitude towards the university, as well as student's accomplishments and goals in their studies at the university. The frequency of answers to each question in this section is shown in Table 12 here below.

Table 12. Frequency - Reykjavik University related

33. Reykjavik University is meeting my standards/expectations of higher education.		
Strongly agree	79	28,94%
Agree	149	54,58%
Neutral	31	11,36%
Disagree	13	4,76%
Strongly disagree	1	0,37%

34. I like the way of teaching (the methods) used at Reykjavik University.		
Strongly agree	62	22,71%
Agree	158	57,88%
Neutral	44	16,12%
Disagree	8	2,93%
Strongly disagree	1	0,37%

35. I do not usually attend lectures (ísl. fyrirlestra) at Reykavik University.		
Strongly agree	41	15,02%
Agree	53	19,41%
Neutral	52	19,05%
Disagree	66	24,18%
Strongly disagree	61	22,34%

36. I watch the lectures (ísl . fyrirlestra) online, on Echo360 in Canvas, rather then attend class.		
Strongly agree	75	27,47%
Agree	62	22,71%
Neutral	49	17,95%
Disagree	50	18,32%
Strongly disagree	37	13,55%

37. I always attend practical classes (ísl. dæmatímar).		
Strongly agree	67	24,54%
Agree	90	32,97%
Neutral	61	22,34%
Disagree	40	14,65%
Strongly disagree	15	5,49%

38. Reaching the Dean's list (isl. forsetalisti) is one of my study goals.		
Strongly Agree	23	8,42%
Agree	46	16,85%
Neutral	53	19,41%
Disagree	74	27,11%
Strongly Disagree	77	28,21%

39. I have been on the deans list (ísl. forsetalisti).		
Yes	20	7,33%
No	253	92,67%

Question 33 states the following: "Reykjavik University is meeting my standards/expectations of higher education", where participants agreeing with the statement were 83.52% (228 out of 273 individuals), versus 5.13% (14 individuals) not agreeing with the statement. There were also 11.36% (31 individuals) that were neutral about the statement. The mean value of answers given where 4.0696, the median value was 4.0 (agree), and the standard deviation was 0.7898.

Next question, question number 34, stated “I like the way of teaching (the methods) used at Reykjavik University”. Here, an overwhelmingly large part of participants agreed with the statement, or 80.59% (220 individuals) of total participants. Participants disagreeing with the statement were only 3,30% (9 individuals), and 16.12% (44 individuals) where indifferent on the statement. The mean value was 3.9963, the median value was 4.0 (agree), and the standard deviation was 0.7351.

Question 35 inquired participants if they attended live lectures at the university. Here, about half of the students do seem to attend live lecture or 46.52% (127 individuals), but 34.43% (94 individuals) do however not attend live lectures, on a regular basis. Around 19.05% (52 individuals) then seem to both attend and not attend live lectures, as they answered the statement as being neutral about it. The mean value of the answers to this particular question was 2.8059, the median value being 3.0 (neutral), and the standard deviation being 1.3783.

For question number 36, participants were given the statement “I watch the lectures online, rather than attend classes”. About half of total participants answer this statement positively, with 50.18% (137 individuals) agreeing with it, versus the 31.87% (87 individuals) not agreeing with it. Then there were 49 individuals (17.95% of participants) that were neutral about it. The mean value was 3.3223, the median value was 4.0 (agree), and the standard deviation was 1.3981.

Question 37 asked participants about their attendance during practical classes, where 57.51% (157 out of 273 participants) answered that they do attend practical classes, and 20.15% (55 individuals) do not attend practical classes, as well as 22.34% (61 individual) who were neutral of the statement. The mean value was 3.5641 for this question, the median value being 4.0 (agree), and the standard deviation being 1.1681.

Next question, question 38, regarded goals set by participants to reach the dean’s list of Reykjavik University, an annual list of students who excel in their studies at the university with the reward being exemption from expenses for the next term. Around one-quarter of participants set their goals to reach the dean's list, or 25.27% (69 individuals) of total participants, and 55.32% (151 individuals) do not do so. About 19.41% (53 individuals) were neutral of it. The mean value was 2.5018, the median value was 2.0 (disagree), and the standard deviation was 1.2895.

For the last question of this section, the participants were asked if they had reached the dean's list in their studies at Reykjavik University, with a possible binary reply of either "yes" or "no". Twenty individuals answered positively, or 7.33% or participants, having reached the dean's list. And 92.67% (253 individuals) had not done so, yet.

Table 13. Correlation - Reykjavik University related

	Q33	Q34	Q35	Q36	Q37	Q38
Q33	1.000	0.5640	-0.0483	-0.0737	0.0808	0.0955
Q34	0.5640	1.000	0.0138	0.0083	0.1180	0.1222
Q35	-0.0483	0.0138	1.000	0.8168	-0.3062	-0.2491
Q36	-0.0737	0.0083	0.8168	1.000	-0.2581	-0.2410
Q37	0.0808	0.1180	-0.3062	-0.2581	1.000	0.1824
Q38	0.0955	0.1222	-0.2491	-0.2410	0.1824	1.000

In regard to correlation between questions in this seventh section of the questionnaire, there is not much correlation to be found (see Table 13). However, between question 33 and question 34 there is a moderately positive correlation, where the first question states that the university is meeting the expectation of higher education made by the participant, and the second question states that the participant likes the way of teaching at the university.

Additionally, there is high positive correlation between question 35 and question 36, where the first question states that the participant does not attend live lecture, and the second question states that the participant watch lectures online instead. This confirms then that those who do not attend lectures, will rather watch lectures online and vice versa as those who do attend lectures, are unlikely to watch the lectures online.

4.1.7 Social Networks

The last part of the questionnaire, section eight, focuses on social networks. The frequency of answers within each question can be seen in Table 14 here below.

Table 14. Frequency - Social Network

40. I prefer to work in groups (arranged by the teacher), rather than work on my own.		
Strongly Agree	13	4,76%
Agree	36	13,19%
Neutral	73	26,74%
Disagree	90	32,97%
Strongly Disagree	61	22,34%

41. I prefer to work in groups (chosen by students), rather than work on my own.		
Strongly agree	75	27,47%
Agree	76	27,84%
Neutral	64	23,44%
Disagree	39	14,29%
Strongly disagree	19	6,96%

42. I have a large social network (10 persons or more) within Reykjavik University.		
Strongly Agree	50	18,32%
Agree	47	17,22%
Neutral	41	15,02%
Disagree	82	30,04%
Strongly Disagree	53	19,41%

Question 40 asked if the person answering prefers to work as part of a group arranged by the teacher of the course, rather than working on his/her own. Here, 17.95% of participants (49 individuals) preferred working within a group that was arranged by the teacher, and 55.31% (151 individuals) did not prefer to work within a group arranged by the teacher. Neutral answers were 26.74% (73 individuals). The mean value was 2.4506, the median value was 2.0 (disagree), and the standard deviation was 1.1174.

In regard to question 41, which inquired about if the student likes to work within groups selected by the students themselves rather than working alone, 55.31% (151 individuals) agreed with the statement, versus 21.25% (58 individuals) who preferred working alone. Neutral answers were 23.44% (64 individuals). The mean value was 3.5458, the median value was 4.0 (agree), and the standard deviation was 1.2273.

The last question of the questionnaire, question 42, asked the participants if they had large social networks (10 people or more) within Reykjavik University. Here, 35.53% (97 individuals) confirmed that they did have large social networks within the university, and 49.45% (135 individuals) answered that they did not have such large social networks. Around 15.02% (41 individuals) answered the question neutrally. The mean value in this question was 2.8498, the median value was 3.0 (neutral), and the standard deviation was 1.4023.

Table 15. Correlation - Social Network

	Q40	Q41	Q42
Q40	1.000	0.1632	0.0527
Q41	0.1632	1.000	0.3170
Q42	0.0527	0.3170	1.000

In regard to correlation between the three questions (see Table 15), there was no proven linear relationship between the questions, or a weak positive linear relationship at best between the answers to these three questions.

As correlations between answers to the questions within each individual section of the questionnaire has been discussed, it is also interesting to see what kind of correlation is between questions across the sections of the questionnaire. This will be the talking point in the following chapter.

4.1.8 Other Interesting Factors

The questionnaire designed along with the answers gathered, provided a lot of interesting and peculiar information. The questionnaire was originally categorized into eight separate sections and questions within each respective section were examined further as well as their relationship to other questions of the same section. However, even more information was to be gained through exploring the answers to each question without regarding their respective section. That is, different correlation amongst questions within different sections.

To start with, correlation was established between question 8 (Educational Values) and question 11 (Goals), where high positive correlation was found (see Table 16). In question 8, the following statement was given: "Getting good grades is important to me.", and in question 11 the following statement was given: "Good grades provide me with an excellent goal to work towards.". By this high correlation, for this particular questionnaire, it can be seen that when grades are important to a student, it becomes their goal to work towards, to excel in their field of study, in this case computer science.

Table 16. Correlation between question 8 and 11

	Q8	Q11
Q8	1.000	0.6007
Q11	0.6007	1.000

When looking into question 8 (Educational Values) and question 38 (University Related) of the questionnaire, as shown in Table 17 here below, a moderately strong correlation can be found. As mentioned before, question 8 states the following: "Getting good grades is important to me.", and in question 38 the following is stated: "Reaching the Dean's list is one of my study goals.", which again makes sense as getting good grades goes hand in hand with both study goals and reaching the achievement list of Reykjavik University's Dean's list.

Table 17. Correlation between question 8 and 38

	Q8	Q38
Q8	1.000	0.4867
Q38	0.4867	1.000

It is also interesting to look into question 11 (Goals) and question 38 (University Related), as the goal for participants trying to reach the dean's list (question 38) would most likely be highly correlated with the goal of good grades to work towards (question 11). The results do show some moderately positive correlation between answers to the two questions considered, but not as high as would have been thought beforehand, see correlation results in Table 18 here below.

Table 18. Correlation between question 11 and 38

	Q11	Q38
Q11	1.000	0.3917
Q38	0.3917	1.000

It is interesting to see that there is a moderately positive correlation between question 9 (Goals) and question 13 (Goals), where in question 9 the following statement is given: "I'm a very methodical person.", and in question 13 the following statement is given: "I organize my study time to best achieve my goals”. This means that those who organize their study time, are also able to motivate themselves in studying, thereby keeping the study time organized to achieve their set goals. This should then also mean that the person should be able to motivate himself/herself to study when needed, which is the statement given in question 27 (Self-Efficacy) in the questionnaire, but the correlation between these three questions vary though, as can be seen in Table 19 here below. By the given correlation, it seems that students that are able to organize their work, are mostly also very methodical, but they vary however when it comes to motivation in their study.

Table 19. Correlation between questions 9, 13 and 27

	Q9	Q13	Q27
Q9	1.000	0.5677	0.3442
Q13	0.5677	1.000	0.4414
Q27	0.3442	0.4414	1.000

In regard to questions 9 (Goals) and 27 (Self-Efficacy), where the inquiry is about the participants methodical thinking and self-motivation, it is also interesting to see the motivations of the participants. Is a person who is methodical also driven to get the best grades possible (question 30 – Academic Apathy) and plan study sessions in advance to achieve these goals (question 32 – academic apathy)? According to the results of the questionnaire, there is a moderately negative correlation between a person being methodical and not making an effort to get as good grades as possible. The same goes for motivation as there is a weak negative correlation between a person who is motivated to study when needed, and not willing to make an effort to get as good grades as possible (see Table 20). There is also a strong positive correlation between a person who is very methodical and is also able to plan study sessions in advance and follow a study plan, which all makes sense. A person who is methodical, plans his/her study in advance and follows this plan, to achieve the highest possible grade by putting in the necessary work to achieve set goals.

Table 20. Correlation between questions 9, 27, 30 and 32

	Q9	Q27	Q30	Q32
Q9	1.000	0.3442	-0.3357	0.4852
Q27	0.3442	1.000	-0.2286	0.4230
Q30	-0.3357	-0.2286	1.000	-0.1780
Q32	0.4852	0.4230	-0.1780	1.000

It is also interesting to see the expectation of individuals participating in the online questionnaire, in accordance to the perception of individuals in regard to their academical skills and seeing things through to the end, for example why learn computer science and how to personally achieve the best learning experience. We can take a look into question 19 (Self-Efficacy), stating the following: "I expect to have a harder time to perform academically than most students here.", and question 31 (Academic Apathy) which states: "I often don't see things through to the end.", as well as question 21 (Self-Efficacy) which states: "I am as skilled academically as the best students here.", see correlation results below in Table 21.

Table 21. Correlation between questions 19, 21 and 31

	Q19	Q21	Q31
Q19	1.000	-0.3641	0.3419
Q21	-0.3641	1.000	-0.2200
Q31	0.3419	-0.2200	1.000

We see a moderately negative correlation between question 19 and 21 (see Table 21), which shows that a person whom is skilled academically, also expects to not have difficulties academically than other students around, and vice versa. There is also a negatively moderate correlation between participants that are skilled academically, and not seeing things through to the end, which can be interpreted as such that individuals who are not certain about their choice in field of study believe themselves to be not as skilled academically. This correlates also into questions 19 and 31, where individuals who expect to have a harder time to perform academically, also don't see things through to the end.

At last, it is intriguing to see the correlation between question 20 (Self-Efficacy) and question 34 (University Related), where in question 20 the following statement is given: "I can easily adapt to different styles of teaching.", and in question 34 where the following statement is given: “I like the way of teaching methods used at Reykjavik University.”. The correlation between these two questions are moderately positive, as can be seen in Table 22 here below.

Table 22. Correlation between questions 20 and 34

	Q20	Q34
Q20	1.000	0.3580
Q34	0.3580	1.000

As the correlation between these two questions (see Table 22) are moderately positive, it is can be said that individuals who can easily adapt to different styles of teaching, also like the teaching methods used at Reykjavik University. This does not come as a big surprise, as there are many different styles of teaching at the Department of Computer Science within Reykjavik University, and the ability of adapting to these styles of teaching, as well as adapting to learning and developing different computer programs, are important traits to have and/or to learn in this field of study.

4.1.9 Main Findings from the Questionnaire

After statistical analysis of the questionnaire results, and calculation of the correlation between different variables, how do students in higher education perceive their own learning patterns and what can be learned from these results?

First of all, the majority of students seem to gather themselves the information they need when making the decision of pursuing higher education, with the acquirement of information about the degree and university of, as well as inquiring about possible career paths after graduation.

The majority also values the importance of being prepared for classes, with the motivation getting good grades, thereby having realistic goals for their individual study. However, the difference between students lies in their methodology for organizing things, whereas half of participants pride themselves as organized and being able to prepare their study time in advance, as well as keeping up with their set schedule, while the other half does not.

A large part of participants in the questionnaire also show high self-esteem and belief in themselves and their capabilities. For example, being able to adjust to different teaching methods, perform as well academically as most other students and even as well as the perceived best students around, taking on difficult tasks that are highly intellectually demanding and complicated, being quick to grasp new concepts, being capable of learning new things quickly, and believing that they will do very well in their field of study. This can be perceived as the mindset of the average student in higher education, being highly motivated and unstoppable in achieving their goals.

When it comes to Reykjavik University itself, the majority of participants believe that their standards and expectations are being met and the teaching methods within the university are enjoyable. When looking further into these matters, the results of the questionnaire show that just under 50% of participants do attend live lectures, while around 50% watch the lecture online rather than attending classes. When it comes to practical classes, a large part of participants seems to attend, or around 50-75% of the total sample.

Participants of the questionnaire also seem to prefer choosing the groups they work within, when it comes to group projects, rather than the teacher assigning the groups. Participants also prefer to work within such groups instead of working alone. At last, the social network size amongst participants of the questionnaire varies between students, as about 35% believe themselves to have a large social network within the facilities of Reykjavik University, and just under 50% do not have such a large network.

In regard to correlation between different variables / questions in the results of the questionnaire, there is not much to grasp as a large part of the result show only low correlation (below ±0.29) or moderate correlation (between ±0.30 and ±0.49) amongst variables. The variables recorded with high positive correlation (±0.50) within the questionnaire results are four in total, and they are: Firstly, those individuals whom are very methodical persons (Q9) and those who organize their study times to best achieve their goals (Q13), with a correlation factor of +0.5632, which does not come as a surprise as methodical persons tend to organize themselves very well in general. Secondly, those who are quick to understand new concepts and ideas (Q26) and those who learn things more quickly than most people (Q28) with a correlation factor of +0.6186, which again does not come to a surprise, as those who interpret new things quickly tend to have the capability to learn things more quickly as well. Thirdly, those who rate their expectation and standards towards Reykjavik University are being met (Q33), and those who like the way of teaching methods used at Reykjavik University (Q34) with a correlation factor of +0.5640, as (once again) these two factors would be expected to have coherence with each other. Fourthly, those who do not usually attend lectures at Reykjavik University (Q35), and those whom watch lectures online rather than attending classes, have the highest scored correlation factor amongst variables in the questionnaire with correlation of +0.8168, which is very interesting and shows that those who are not attend live lectures in classrooms, are very likely to watch online lectures instead, and vice versa. Other variables within the questionnaire scored lower than ±0.50 in correlation amongst each other and are therefore not discussed in more details here.

The subject of the next chapter is the data analysis performed with data gathered from the learning management system of Canvas, which is the teaching system used at Reykjavik University. The focus was originally set on data analysis of six courses being taught at the Department of Computer Science within Reykjavik University. Those six courses are: Discrete Mathematics I (first year, first term), Software Requirements and Design (first year, first term), Data Structures (first year, second term), Databases (first year, second term), Calculus and Statistics (second year, first term), and Computer Networks (third year, first term).

4.2 Descriptive Statistics

Data from Canvas was first analysed using the statistical computing of RStudio and R, where the courses as a whole were analysed, and the data from Canvas was adapted to the needs of this thesis.

The Canvas data was gathered from six courses that where taught in two terms, spring and autumn in the year of 2019, at the Department of Computer Science within Reykjavik University. Following are the results from data analysis in the statistical program of R, which are divided into six sections, one section for each of the six courses.

4.2.1 Discrete Mathematics I

The first course analysed from RStudio was the course of “Discrete Mathematics I” (ísl. Strjál stærðfræði I), which is taught to undergraduate students in their first term within the

Department of Computer Science at Reykjavik University. According to the data from Canvas, there were 223 students registered for the course, and was being taught throughout the time period of 10.August 2019 - 31.December 2019.

In Figure 2 here below can be seen the frequency of usage by students for the course on a daily basis, where the highest usage recorded 3,481 mouse-clicks and the lowest recorded 8 mouse-clicks over a 24-hour time period, with an average of approximately 986 mouse-clicks per day. Therefore, there is high variance between days of usage for the course, especially shown with eight large spikes within the graph of Figure 2. These spikes do not come as a particular surprise, as the students within the course had to turn in five small assignments and three larger assignments throughout the teaching period to pass the course, as well as passing a final exam which was held in November 2019 which explains the spike that occurs at the right end of the graph in Figure 2. The final spike then occurs in late November 2019, which was predicted to be the release of the final grades in the course. After this last spike, we can see a rapid drop in the usage of Canvas for the course.

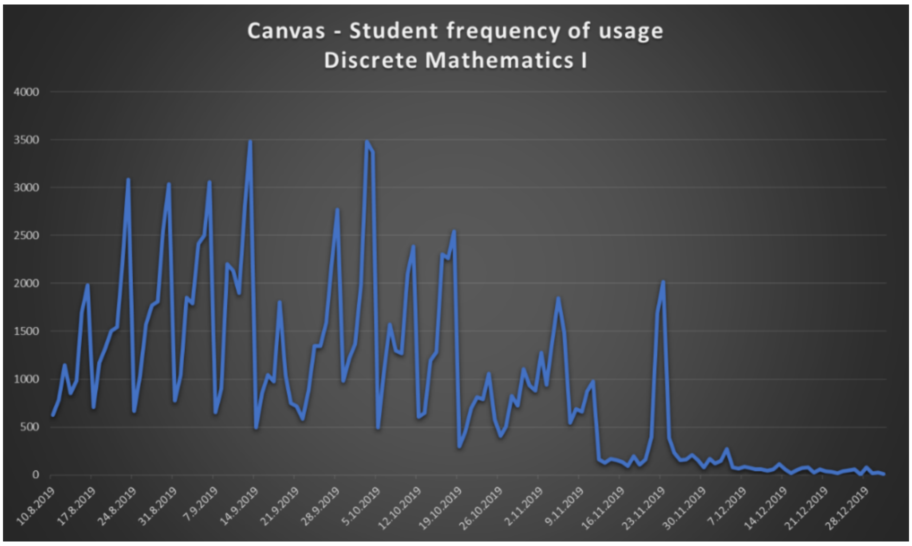

Figure 2. Discrete Mathematics - Canvas frequency

It is also interesting to see the different usage of Canvas that the students display within the course, but the main three pages are Assignments, Grades, and Modules. In Figure 3 here below is shown the total usage of students for these three pages over the time period for the course. As can be seen within the graph in Figure 3, the page of Grades (orange line) does not have much activity, until after the first spike in Assignments (blue line), where students are getting feedback of the previously turned in assignments within the grades page.

The same pattern happens with these two pages throughout the course, as well as a large spike in the activity of the grades page in late November 2019, which is assumed to be when the final score grades of students within the course were released.

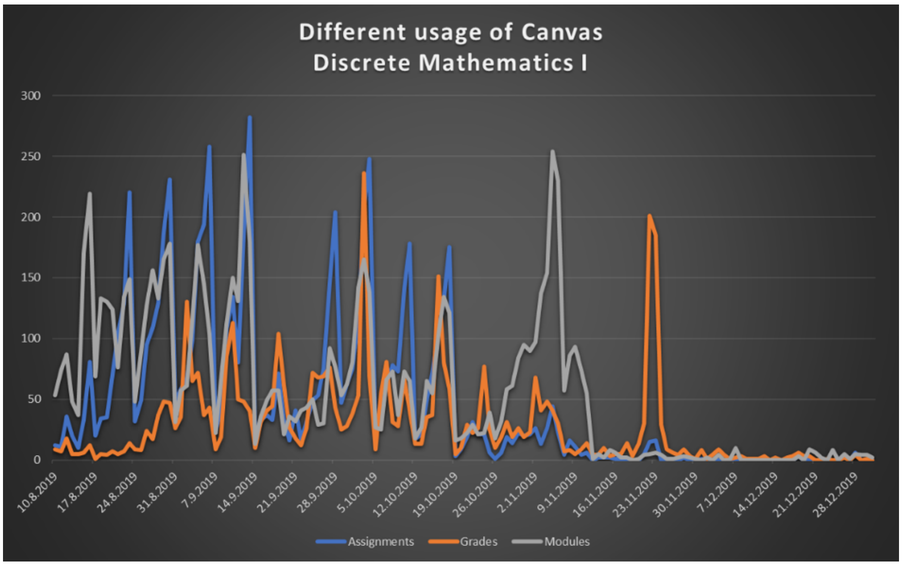

Figure 3. Discrete Mathematics I - Usage of Canvas

In regard to the page of Modules (the grey line within the chart of Figure 3 here above), where all of the course materials are gathered on behalf of the teachers for the course and students can download, the same pattern seems to be occurring, with large spikes throughout the course and small activity in-between, as well as a large grey spike over a larger time period in the beginning of November 2019 (preparation for the final exam). This shows that students are using the course webpage as a preparation before classes held, preparation and review before assignment deadlines, and most likely some kind of mixture of the two.

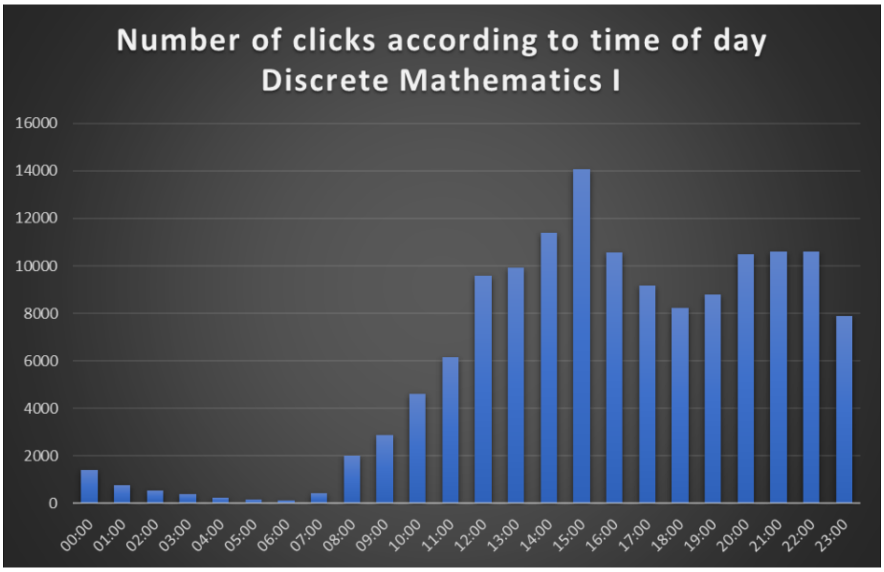

Figure 4. Discrete Mathematics I - Time of usage

Another interesting thing to analyse is when students are actually using the pages of the course within Canvas. In Figure 4 here above has been put together the number of mouse-clicks according to what time of the hour it is within the 24-hour time period of a day. As can be seen in the graph in Figure 4, there is activity almost all the time, even at night time between 00:00 o´clock and 06:00 o'clock. But the highest activity is still between 08:00 o´clock and 00:00 o'clock, where the activity increases throughout the day, until after 16:00 o'clock when it drops a bit and then holds its course until midnight. Even though the usage of Canvas drops significantly after 00:00 o'clock, there is still a large amount of usage between 00:00 o'clock and 01:00 o'clock, so some students seem to be working longer hours.

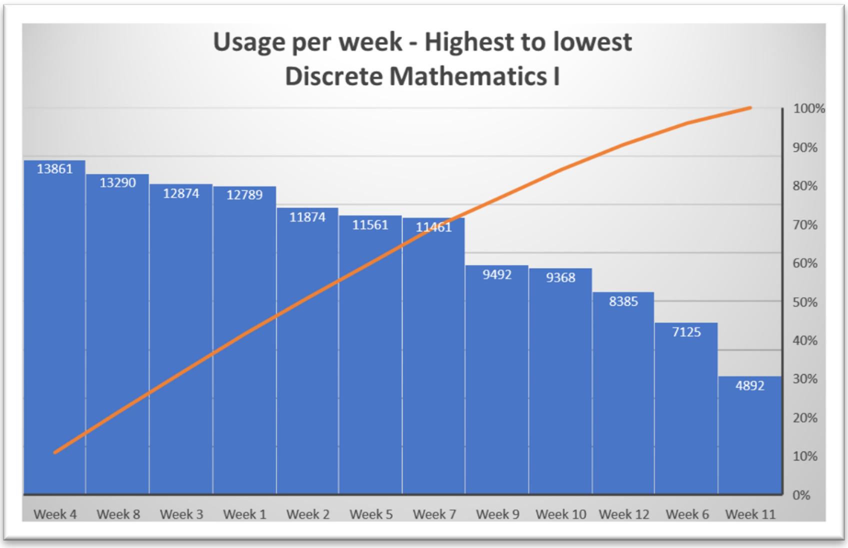

Figure 5. Discrete Mathematics - Usage per week

In Figure 5 here above is shown the total usage of Canvas within the course per week, from the highest frequency (to the left of the graph) with 13,861 mouse-clicks during week 4, to the lowest frequency with 4,892 mouse-clicks in week 11. By this, it can be seen that the usage of Canvas lies mostly in the first weeks of the course (week 1-5), while at the end of the course the frequency of usage drops significantly (week 9-12).

Table 23. Discrete Mathematics I - Total usage per weekday

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Total:
Wednesday	986	1545	1812	2416	1898	1807	1346	1990	1270	2303	794	1104	19271
Thursday	1696	2199	2552	2501	2815	1036	1591	3479	2097	2263	1056	933	24218
Friday	1983	3084	3035	3058	3481	751	2188	3369	2388	2546	579	877	27339
Saturday	1337	666	781	656	496	713	2766	499	606	302	408	1279	10509
Sunday	1958	1036	1047	898	853	588	986	1082	651	446	505	940	10990
Monday	2472	1573	1855	2198	1044	883	1216	1570	1195	698	827	1405	16936
Tuesday	2357	1771	1792	2134	974	1347	1368	1301	1285	810	723	1847	17709
Total:	12789	11874	12874	13861	11561	7125	11461	13290	9492	9368	4892	8385	

In Table 23 here above is shown the total frequency of usage for the course, according to weekdays and weeks. Here, we can see that most frequency of usage is occurring Monday-Friday, but much less during the weekends. This may be due to assignments and homework that students have to turn in during weekdays for the course, or due to other courses that take more time from students during the weekends.

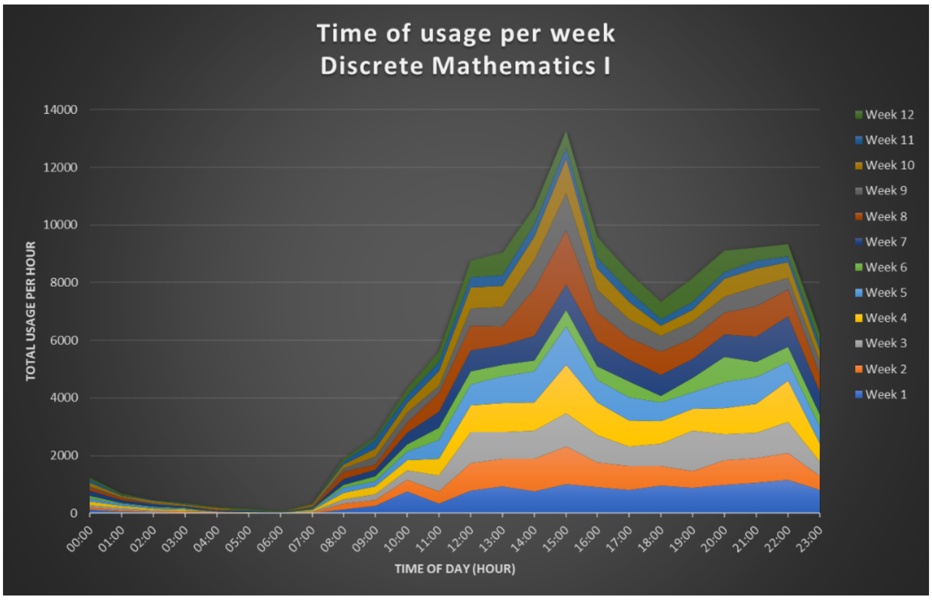

Figure 6. Discrete Mathematics - Time of usage per week

In Figure 6 here above is then shown the time and amount of usage per week, that is at what time of hour are students using Canvas over the 12 weeks of teaching (before final exam period). As shown within the graph of Figure 6, there is high usage of Canvas from 08:00 o'clock until 00:00 o'clock over the whole course, with a peak of around 13,000 mouse-clicks at about 15:00 o'clock in the afternoon, as to be expected in accordance to Figure 4 here above. It also can be seen that in the first weeks of the course (week 1-5) there is recorded higher usage than for the last couple of weeks (week 11-12). However, there seems to be fair resemblance between weeks over the course period, from week 1 to week 12.

4.2.2 Software Requirements and Design

The second course analysed in RStudio was "Software Requirements and Design" (ísl. Greining og hönnun hugbúnaðar), a course that is taught on the first term. According to the data from Canvas, there were 314 students registered for the course, and was being taught the time period of 08.August 2019 - 31.December 2019.

In Figure 7 here below can be seen the frequency of usage by students for the course on a daily basis, where the highest peak of usage recorded is 4,659 mouse-clicks and the lowest was only 14 mouse-clicks over a 24-hour time period, with an average of approximately 1,269 mouse-clicks per day. As shown within the graph of Figure 7, there are several spikes within the time period of the course, which may be because students had to turn in weekly assignments over the teaching period of the course in addition to four large group assignments, as well as having a final exam which was held in November 2019.

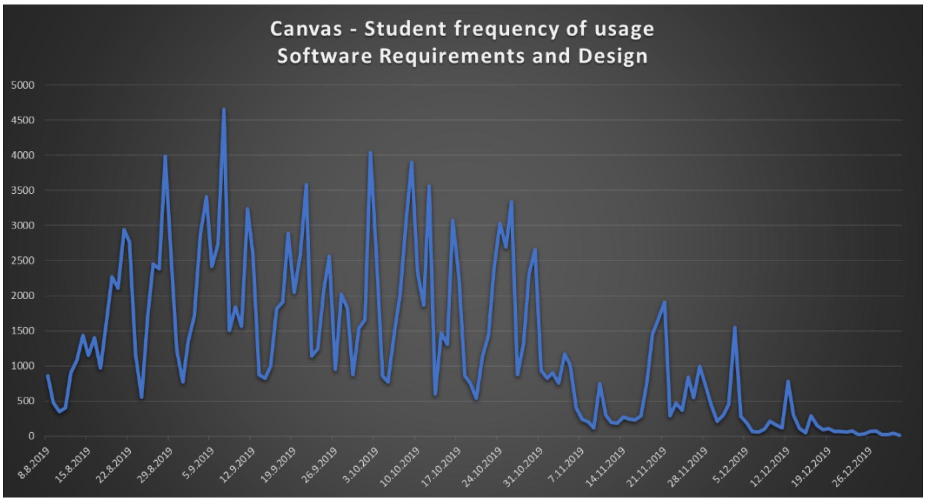

Figure 7. Software Requirements and Design - Canvas frequency

It is interesting to see in Figure 7 here above that there are a couple of small spikes as well after the exam period (late November, beginning of December 2019). This can be explained by the release of gradings, both for the final exam as well as assignments that had been turned in by students over the course of the term This is confirmed in the graph of Figure 8 here below, where the page for Grades is shown in orange colour.

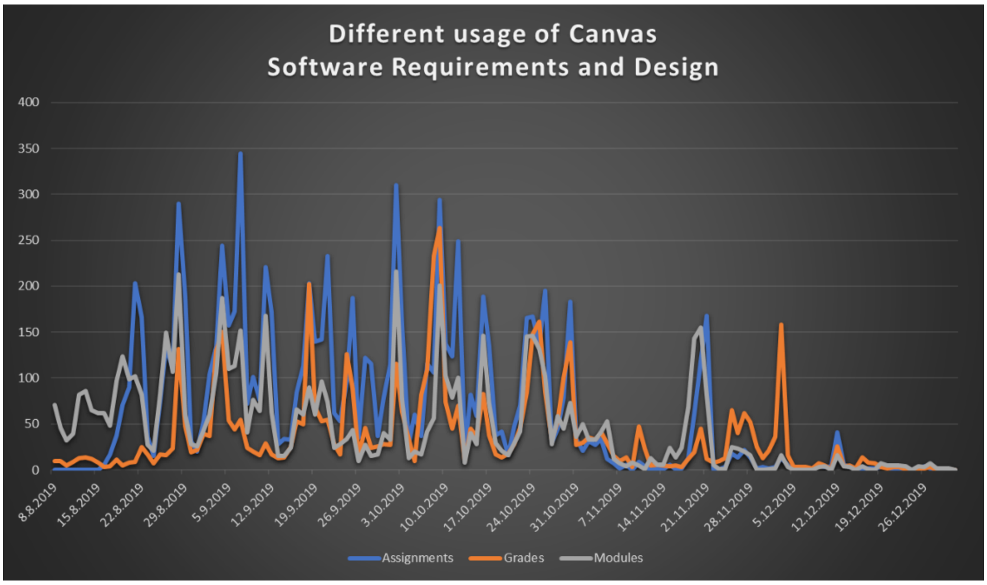

Figure 8. Software Requirements and Design - Usage of Canvas

It can also be seen within Figure 8 here above that the Modules page, where all the material is posted from instructors of the course, marked with a grey line is steadily going up and down over the time period of the course. In particular it is detected that spikes occur within the Assignments page (marked with a blue line), at the same time that spikes occur in the Modules page, which may be due to assignments and/or homework material for the course.

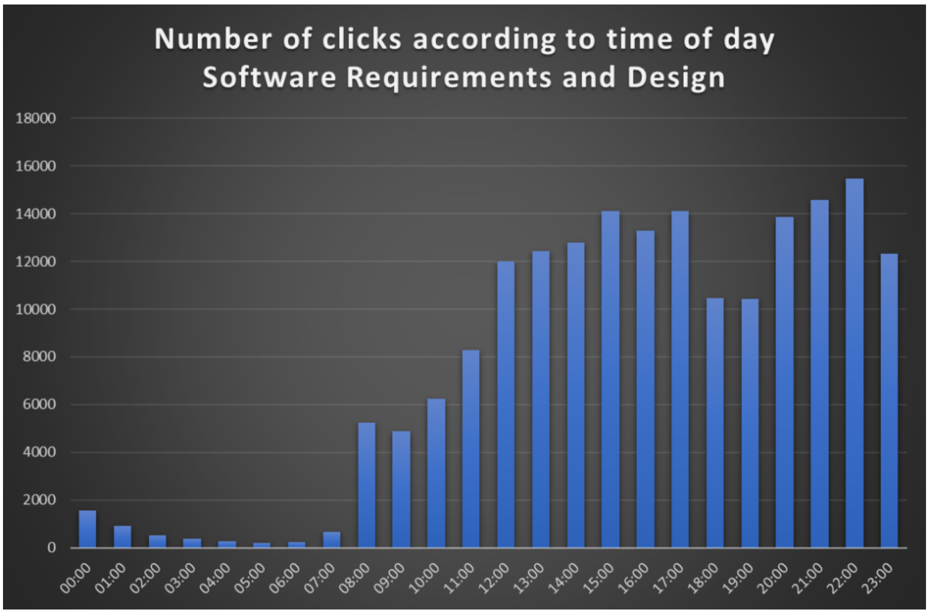

Figure 9. Software Requirements and Design - Time of usage

Another factor that was analysed is at what time of day students are mostly using Canvas for the course. This is shown in Figure 9 here above, where it can be seen that usage is mostly during the time period 08:00-00:00, with increasing usage from 08:00 o'clock until 18:00 o'clock, and then again peaking between 20:00 o'clock and 23:00 o'clock. It can also be seen that even though not much activity occurs during night time (between 00:00 o'clock and 06:00 o'clock) in comparison to usage in the daytime, there are consistently some students that are active during these late hours.

In Figure 10 here below is shown the different usage in frequency of mouse-clicks per week, from the highest frequency of 18,126 clicks in week 4, to the lowest frequency of 8,278 clicks in week 12. Here, it can be seen that at the early- and mid-stages of the course there is more usage, in comparison to the later stages of the course (week 10 and week 12). Amongst the week from 1 to week 9, there seems to be fairly well distributed usage amongst weeks, with the highest peak occurring in week 4 and short drops within week 5 and week 7. At the later stages of the course (week 10 and week 12), the usage seems to drop though.

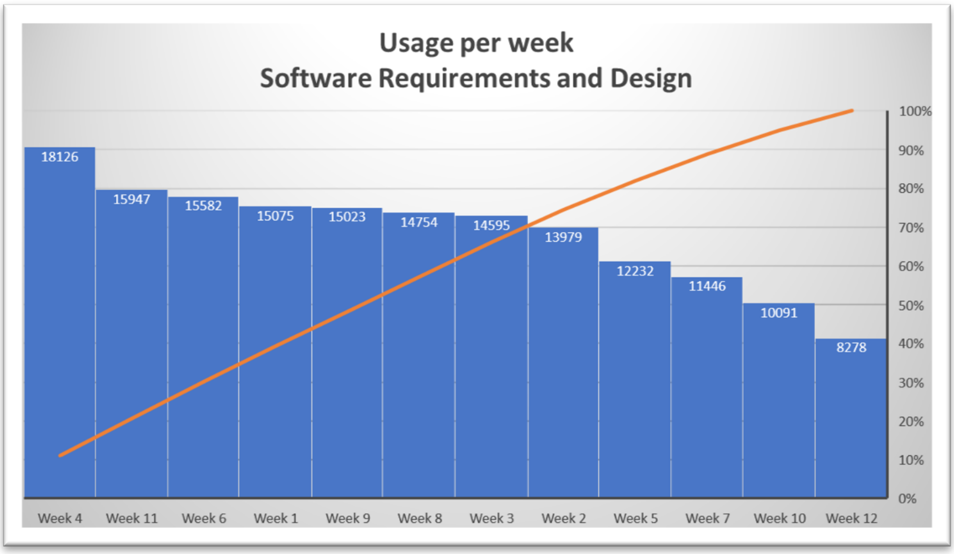

Figure 10. Software Requirements and Design - Usage per week

In Table 24 here below is the frequency of usage within Canvas during weekdays (and weeks). Here, it can be seen that the highest usage is during Sundays (total usage 37,118 mouse-clicks), which may be relating deadlines of assignments mostly being on Sundays. The very lowest frequency was however on Thursdays.

Table 24. Software Requirements and Design - Total usage per weekday

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Total:
Sunday	2046	2945	3983	3409	3236	2894	2560	4039	3898	3075	2375	2658	37118
Monday	3173	2765	2660	2419	2572	2048	957	2550	2319	2289	3027	939	27718
Tuesday	3189	1159	1231	2728	876	2579	2021	856	1870	863	2691	826	20889
Wednesday	1440	559	775	4659	825	3581	1827	771	3559	753	3342	906	22997
Thursday	2016	1715	1362	1503	1000	1145	876	1442	598	540	879	759	13835
Friday	1881	2452	1713	1844	1817	1246	1542	2024	1474	1133	1321	1172	19619
Saturday	1330	2384	2871	1564	1906	2089	1663	3072	1305	1438	2312	1018	22952
Total:	15075	13979	14595	18126	12232	15582	11446	14754	15023	10091	15947	8278	

At last, the time and amount of usage per week was analysed for this course, where the usage per week is quite similar over the teaching period for the course but does minimize a bit during the last weeks (weeks 10-12) it seems – see Figure 11 here below.

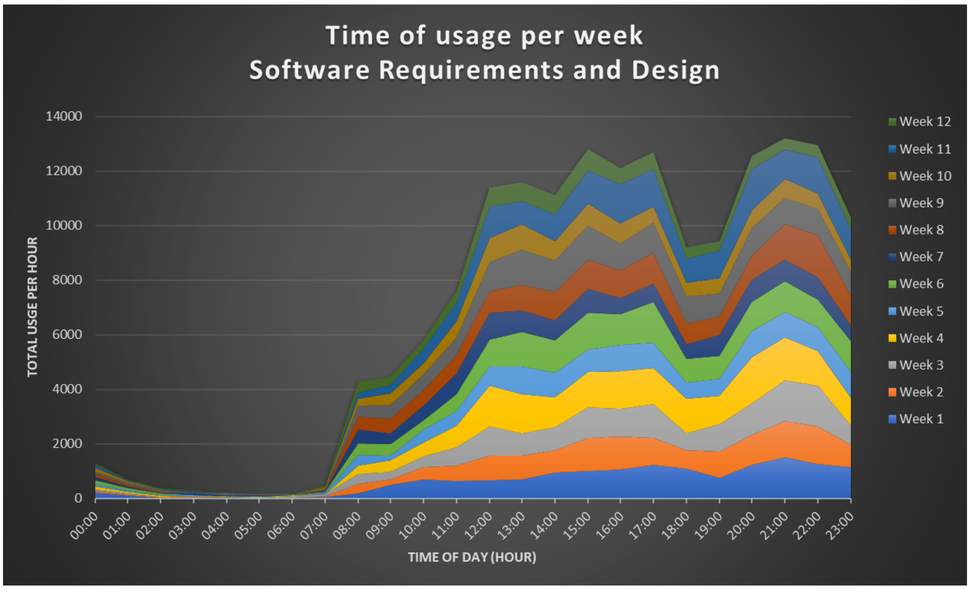

Figure 11. Software Requirements and Design - Time of usage per week

Why this distribution between usage per week throughout the course (Figure 11), may be explained by the weekly assignments that students had to turn in throughout the teaching period, and therefore the recorded amount in usage of Canvas for the whole time period is found, with little discrepancies between weeks (excluding week 10 and 12).

4.2.3 Calculus and Statistics

Third course analysed was "Calculus and Statistics" (ísl. Stærðfræðigreining og tölfræði), which is taught to undergraduate students on their third term within the university. According to the data from Canvas, there were 131 students registered for the course, which was taught during the time period of 13.August – 31.December 2019.

In Figure 12 here below is the frequency of student usage within Canvas for the course on a daily basis, where the highest frequency of usage recorded on a single day was 1,737 mouse-clicks, and the lowest was 8 mouse-clicks over a 24-hour time period, with an average of approximately 421 mouse-clicks per day, which is very low in comparison to other courses analysed.

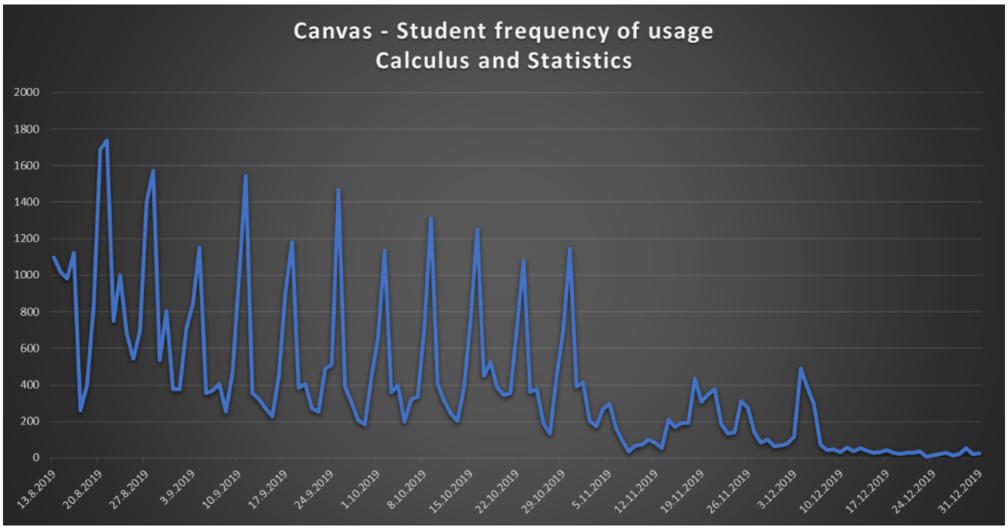

Figure 12. Calculus and Statistics - Canvas frequency

As shown in Figure 12 here above, there are several high spikes between the time period 16.August 2019 - 30.October 2019. This can be attributed to the 11 different assignments students had to turn in through Canvas over the course period, where students showed high usage of the Assignments page and Modules page. There are also small spikes during mid-and-late November 2019, which can be explained by the final exam period of the course. At last, there is a high spike in the beginning of December 2019, which is due to the release of final grades for the course.

In Figure 13 here below are shown the frequency of mouse-clicks between the three most used pages within the course, which are the Assignments page (shown in blue on the graph), the Modules page (shown in grey), and the Grades page (shown in orange). This graph verifies what has been written here above, that the pages of Modules and Assignments are being used throughout the teaching part of the course. There is then a high spike in the Modules page during the final exam period, and at last there is a high spike in the Grades page in the beginning of December 2019, when the final grades for the course were released.

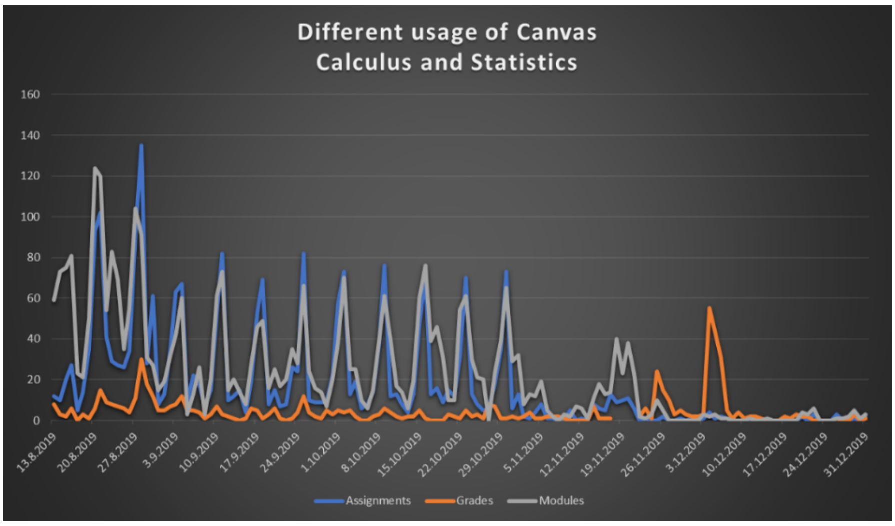

Figure 13. Calculus and Statistics - Usage of Canvas

Another interesting thing to look into is when students are actually active on Canvas, that is what time of hour are students mostly active within the course pages in Canvas. This can be seen in Figure 14 here below, where number of clicks are shown per hour according to the time of day. As shown within the graph of Figure 14, there is low activity during night time (between 00:00 o'clock and 07:00 o'clock), but increases at a high rate and peaks around 14:00 o'clock, afterwards keeping the usage fairly moderate until midnight when it drops completely. But, as stated before the amount of usage (shown on the vertical axis) is very small compared to other courses analysed in this thesis.

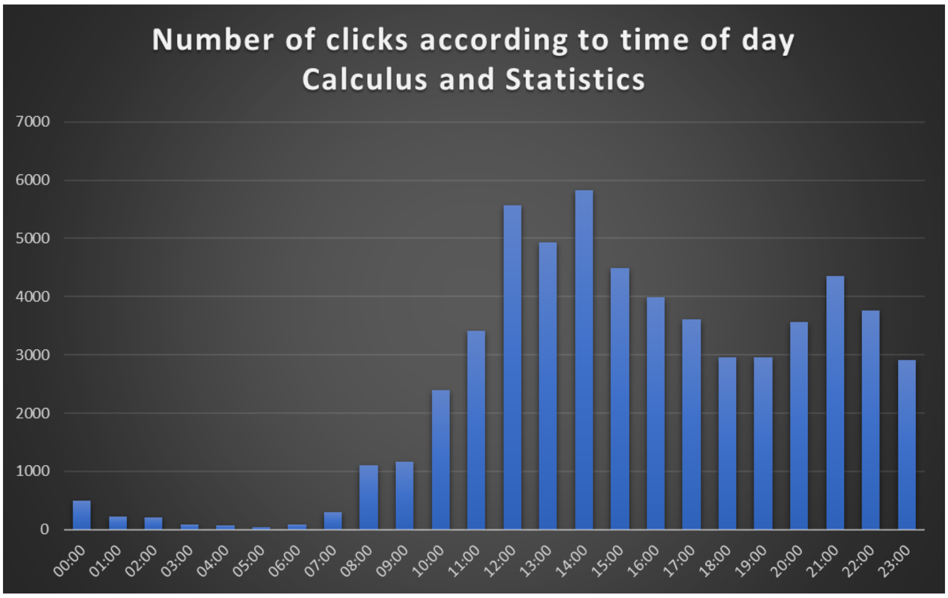

Figure 14. Calculus and Statistics - Time of usage

Next to notice is the total usage per week for the course, which can be seen in Figure 15 here below. As shown in the figure, there is moderately high frequency of usage during the very first weeks of the course (week 1-2), recording 7,391 clicks in week 1 and 6,801 clicks in week 2, but then drops about 25-40% thereafter and keeping its pace from week 5 until week 12, where the lowest frequency of usage is recorded to be 2,890 mouse-clicks over the course of a whole week.

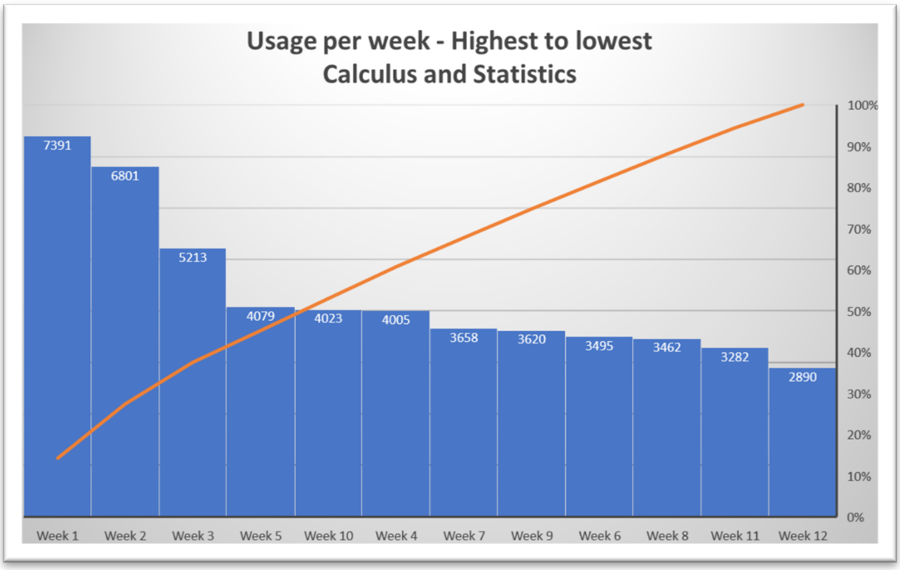

Figure 15. Calculus and Statistics - Usage per week

It is also interesting to note the frequency of usage per weekday, which can be seen in Table 25 here below. As shown, the total frequency of usage is significantly lower in comparison to other courses analysed within this thesis. As shown in Table 25, the highest frequency occurred during Tuesdays and Wednesday (11,280 and 15,588 clicks respectively), while other days recorded significantly lower usage (from 3,408 up to 6,354 clicks).

Table 25. Calculus and Statistics - Total usage per weekday

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Total:
Wednesday	1016	1737	1573	1153	1543	1179	1468	1135	1311	1248	1079	1146	15588
Thursday	982	750	537	356	356	385	390	360	407	450	362	391	5726
Friday	1121	1000	803	369	321	407	305	400	313	525	376	414	6354
Saturday	262	679	375	405	273	270	208	200	241	386	189	202	3690
Sunday	398	544	375	255	229	254	185	318	201	343	133	173	3408
Monday	828	695	706	467	449	490	443	336	387	356	449	267	5873
Tuesday	2784	1396	844	1000	908	510	659	713	760	715	694	297	11280
Total:	7391	6801	5213	4005	4079	3495	3658	3462	3620	4023	3282	2890	

In Figure 16 here below is then shown the time and amount of usage dated per week. As shown, the frequency of usage is recorded to be high during the first couple of weeks (week 1-2), and thereafter keeping a steady pace in the time of usage per week. It also shows, as was displayed in Figure 14 here above, that the frequency of usage occurs mostly between 08:00 o'clock and 00:00 'clock.

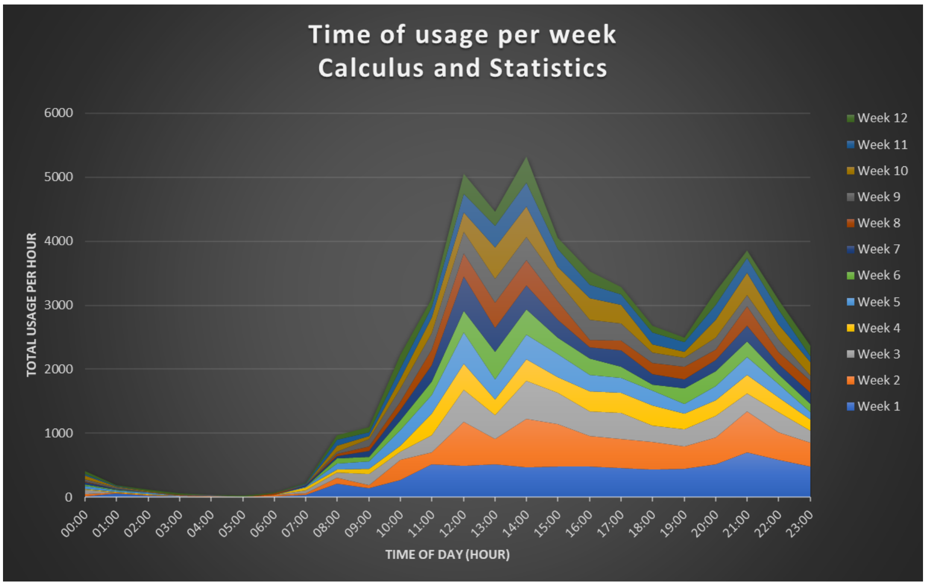

Figure 16. Calculus and Statistics - Time of usage per week

4.2.4 Computer Networks

The fourth course analysed was "Computer Networks" (ísl. Tölvusamskipti), a course taught for undergraduate students on their fifth term (third year, second term). In total, 205 registered students for the course and the course period was 13.August 2019 - 31.December 2019.

Figure 17 shows the frequency of usage per day. The highest frequency of usage recorded was 4,749 mouse-clicks and the lowest was 23 mouse-clicks during the 24-hour time period of a day, with an average of approximately 1,131 mouse-clicks per day. There are some high spikes during the period of the course, which is explained by the six assignments/projects that students turned in during the course, while another peak is related to the final exam that was held in November 2019.

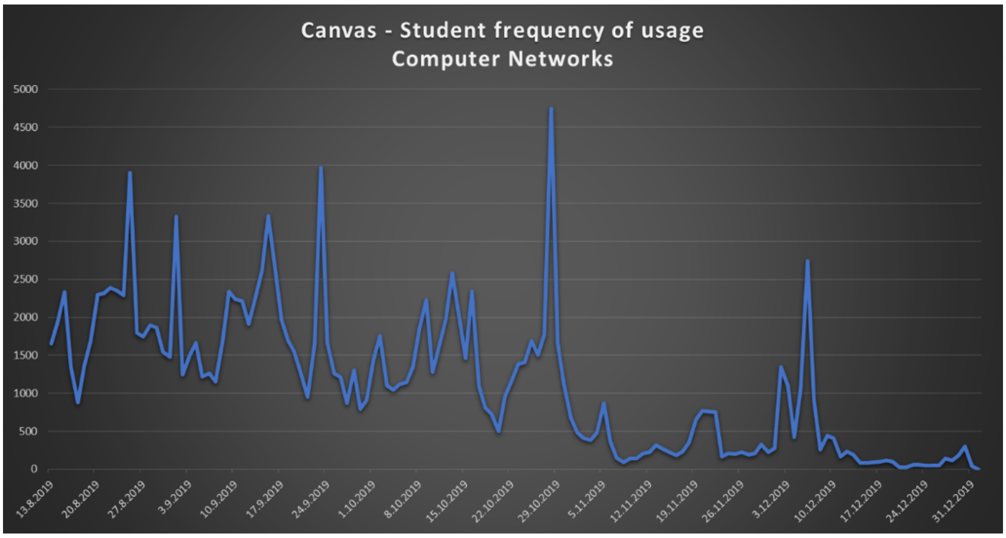

Figure 17. Computer Networks - Canvas frequency

The three main pages that students interacted with during the course, were the pages of Assignments, Grades, and Modules. To show the different use patterns related to these three pages, a graph was made to show the frequency of use for each of these pages (see Figure 18). Here, the big spikes in usage are related to the Assignments page (the blue line on the graph). The Modules page, represented by the grey line in the graph, shows when material from teachers is being interacted with, and that line shows a steady use over the course with a few spikes around the period when assignments were posted. With the Grades page, which is represented by an orange line in the graph, also shows spikes, but at a later stage compared to assignments/modules, which makes sense as grades are released after assignments have been turned in and graded. We can also see a large spike in the frequency of usage for the Grades page in the beginning of December 2019, which is when the final grades were released.

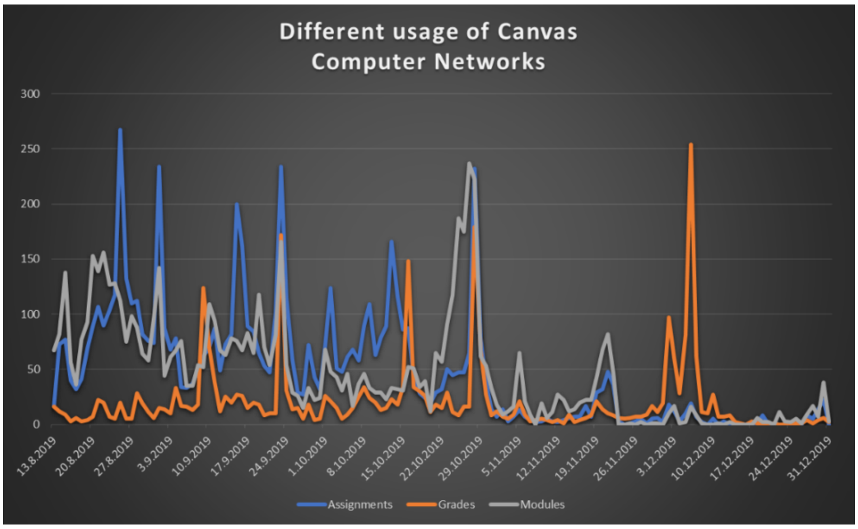

Figure 18. Computer Networks - Usage of Canvas

Interestingly, there is more activity in the course of “Computer Networks” after midnight, compared to the other courses, which can be explained by the allowed late hand-in for the assignments of the course (but with the consequence of a lower grade). In figure 19, there also seems to be high activity during the normal waking hours, from 08:00 o´clock until 00:00 o’clock, with a small drop between 18:00 o’clock and 20:00 o’clock, which is unusual compared with other courses analysed herein.

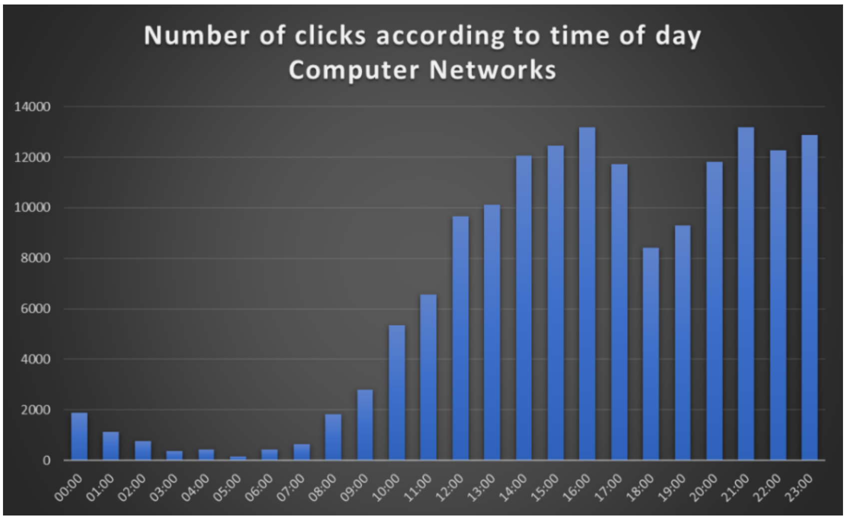

Figure 19. Computer Networks - Time of usage

Figure 20 shows the frequency of usage per week in the course. As shown, the frequency of usage varies a lot between weeks, with the highest frequency recorded during weeks 5 (16,939 clicks) and week 2 (16,780 clicks), and the lowest frequency of usage occurring in the later part of the course, more specifically in week 10 (7,584 clicks) and in week 12 (only 4,435 clicks).

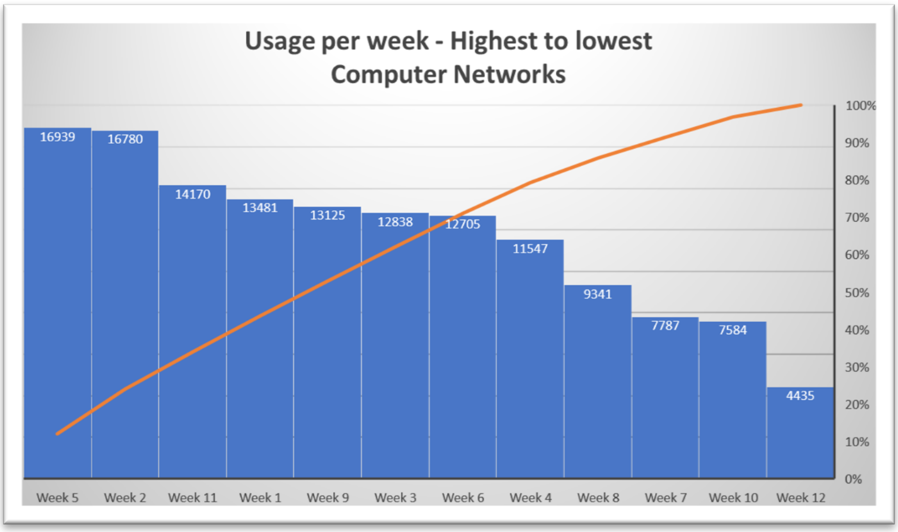

Figure 20.Computer Networks - Usage per week

Table 26 also shows the total usage per weekday (and week), regarding day that show the most frequency of mouse-clicks recorded in Canvas for the course. As shown in the table, the frequency of usage is fairly similar throughout the week, where the highest frequency of 24,170 clicks occurs on Mondays, and the lowest frequency of 16,403 clicks occurs on Saturdays. The small difference between days, yet high overall frequency is unusual in comparison to the other courses that have been analysed for the purpose of this thesis.

Table 26. Computer Networks - Total usage per weekday

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Total:
Wednesday	1938	2311	1892	1663	2215	1693	1258	1751	2227	2335	1383	1119	21785
Thursday	2327	2387	1860	1217	1911	1546	1207	1100	1277	1098	1406	675	18011
Friday	1343	2350	1541	1259	2227	1241	871	1044	1609	807	1687	487	16466
Saturday	873	2292	1482	1152	2603	949	1305	1121	1981	729	1507	409	16403
Sunday	1360	3899	3321	1675	3335	1645	795	1147	2579	505	1774	388	22423
Monday	1688	1794	1245	2342	2688	3972	907	1344	1993	960	4749	488	24170
Tuesday	3952	1747	1497	2239	1960	1659	1444	1834	1459	1150	1664	869	21474
Total:	13481	16780	12838	11547	16939	12705	7787	9341	13125	7584	14170	4435	

Finally, in Figure 21 the time and amount of usage per week in the course can be seen. Similarly to Figure 19, the frequency of usage seen in Figure 21 is mostly concentrated to the hours between 07:00 o'clock and 00:00 o'clock, with a small extension to about 02:00 o'clock when the frequency drops close to zero afterwards. Additionally, it can be seen in Figure 21 that the usage is fairly consistent over the time period included, with a higher frequency during the first parts of the term.

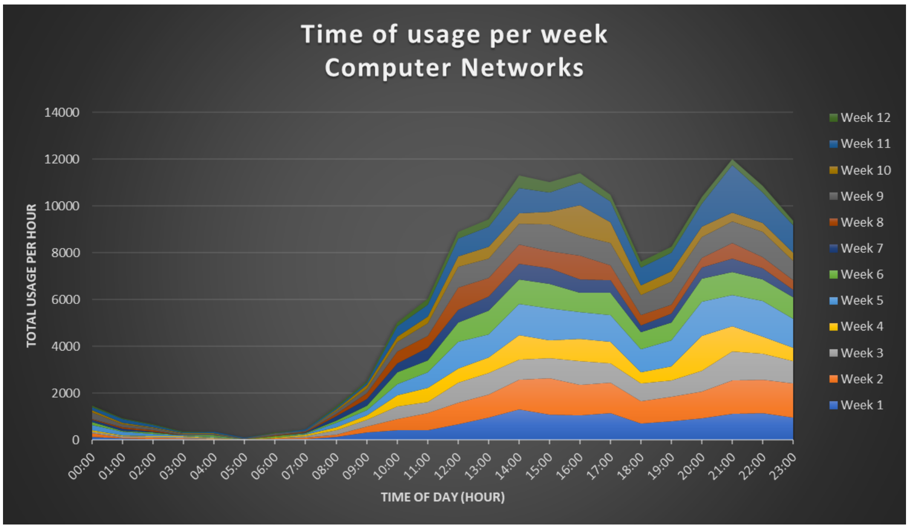

Figure 21. Computer Networks - Time of usage per week

4.2.5 Data Structures

The fifth course was "Data Structures" (ísl. Gagnaskipan), a course taught to undergraduate students on their second term of the program. There were 191 students registered for the course, and it ran from 6.January – 31.May 2019.

Figure 22 shows the frequency of usage by students of the course pages. The highest frequency recorded for a one day was 6,185 mouse-clicks, and the lowest was 5 mouse-clicks, with an average frequency of 1,471 mouse-clicks per day. During the days between 9.January 2019 and 29.March 2019, there are two spikes visible for each week. This correlates with quizzes held as a part of the course, and these quizzes were held in Canvas. There were also 6 assignments that were handed in during the course period, and three partial exams held as well. The course was finalized with a final exam in mid-April 2019, which can be seen in the spike in the graph of Figure 22 here below.

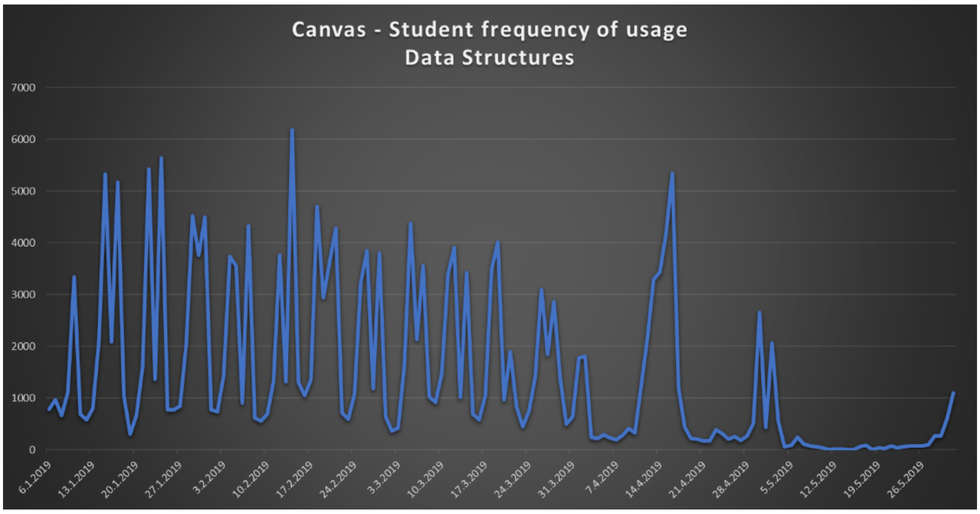

Figure 22. Data Structure - Canvas frequency

To further illustrate this, a graph was made with the frequency of the three main pages of the course. The Assignment page, the Grades page, and the Modules page. The graph can be seen in Figure 23, where the frequency of mouse-clicks for the Assignments page is seen in blue colour, the frequency of mouse-clicks for the Grades page can be seen in orange colour, and the frequency of mouse-clicks for the Modules page can be seen in grey colour within the graph.

As can be seen in the graph of Figure 23 here below, the frequency of the Modules page is high compared to the other two pages. This can be explained, because the lectures each week was not live lectures, but were instead lecture captures that were posted on Canvas under the Modules page. Additionally, the Assignments page and the Grades page are fairly similar, because the quiz was placed on the Assignments page, and the grades for these quizzes were posted under the Grades page afterwards. There are also some spikes in the frequency of mouse-clicks for the Assignments page (the blue line in the graph), which are when the assignments were posted.

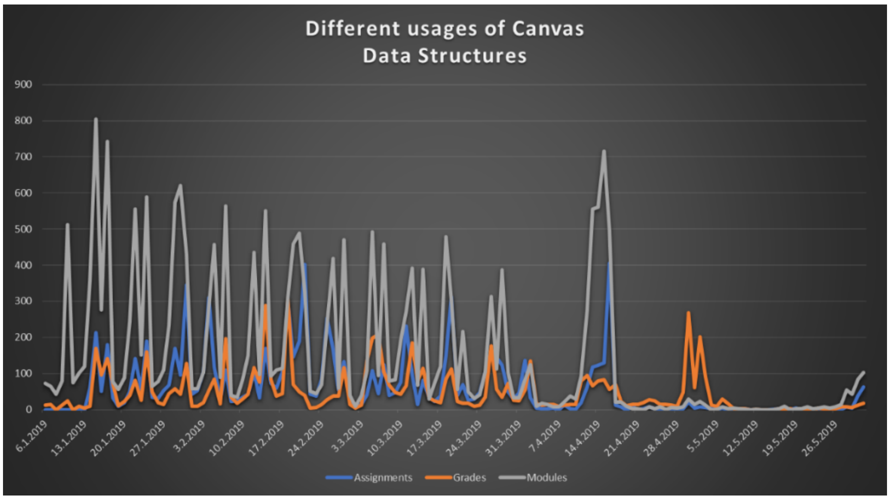

Figure 23. Data Structures - Usage of Canvas

The graph in Figure 23 shows a high spike in the frequency of mouse-clicks of the Modules page in mid-April 2019, as well as a fairly high spike in the frequency of the Assignments page around the same time. This relates to the final exam, which was held in mid-April 2019. Additionally, there is a spike in the frequency of mouse-clicks for the Grades page in the beginning of May 2019, explained by the release of final grades for the course. Furthest to the right in the graph, in late May 2019, there is a small increase for the frequency of mouse-clicks for the Modules page, as well as the Assignments page, related to re-take exam held late May 2019.

In regard to the time of usage, the number of mouse-clicks recorded can be seen in Figure 24. Here, the largest frequency of usage occurs between 08:00 o´clock and 00:00 o'clock, with low frequency (but still some activity) during the night time, which is similar to the other courses analysed in this thesis.

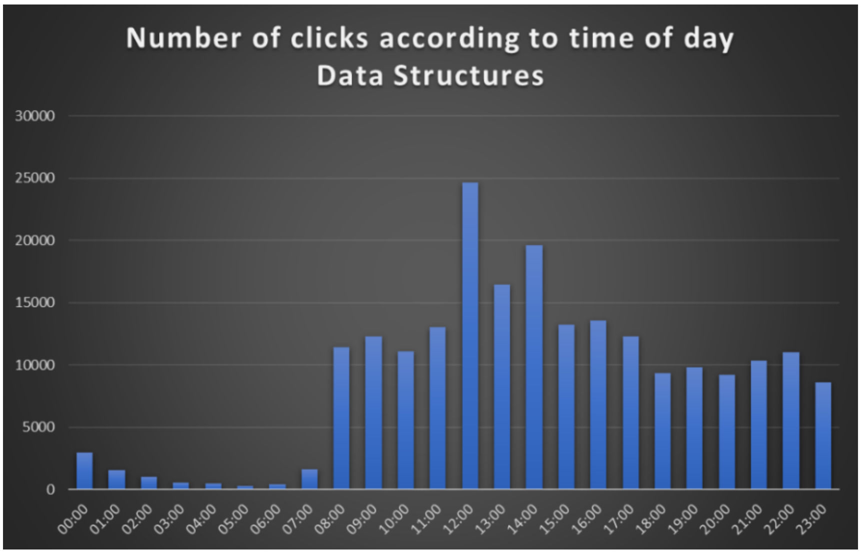

Figure 24. Data Structures - Time of usage

Figure 25 shows the usage of Canvas per week during the course. The most frequency in usage occurs during the first few weeks of the term, in weeks 1-9 (excluding week 5 and week 8), with the highest frequency of usage taking place in week 6 with 18,850 mouse-clicks. The lowest frequency is recorded to be at the end of the term, in weeks 11 and 12, where the lowest frequency of usage is 9,410 mouse-clicks in week 11.

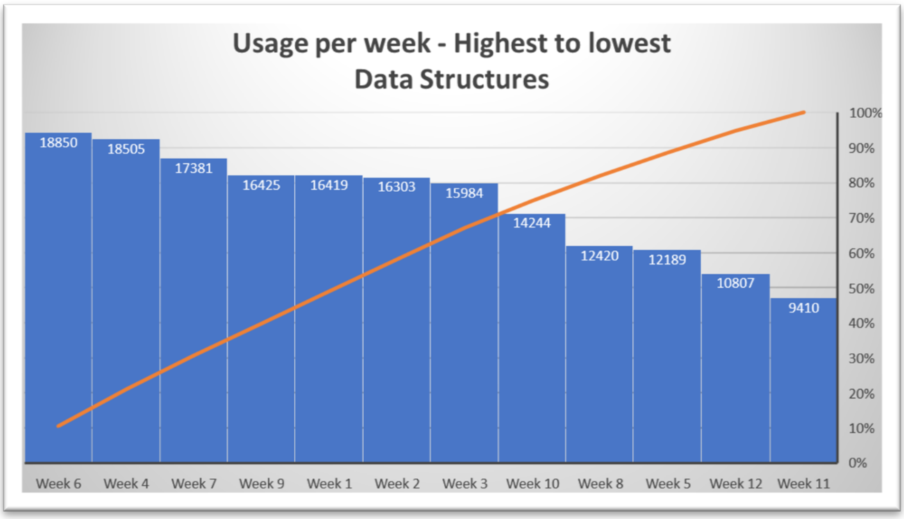

Figure 25. Data Structures - Usage per week

Table 27 shows the total usage per weekday, as well as weeks, were the highest frequency recorded is 48,965 clicks on Thursdays, and close second is Tuesdays with 47,268 mouse-clicks. The lowest frequency however occurs on Saturdays, with only 7,414 mouse-clicks.

Table 27. Data Structures - Total usage per weekday

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Total:
Wednesday	1117	2091	1369	3763	909	1315	3633	1184	2141	1013	971	1843	21349
Thursday	3336	5168	5643	4492	4322	6185	4279	3794	3561	3426	1897	2862	48965
Friday	696	1042	777	773	618	1303	716	641	1027	688	838	1385	10504
Saturday	582	310	770	743	560	1053	600	364	917	577	445	493	7414
Sunday	1590	664	853	1440	692	1348	1071	430	1479	1047	750	645	12009
Monday	3104	1607	2049	3733	1321	4700	3235	1629	3390	3482	1409	1769	31428
Tuesday	5994	5421	4523	3561	3767	2946	3847	4378	3910	4011	3100	1810	47268
Total:	16419	16303	15984	18505	12189	18850	17381	12420	16425	14244	9410	10807	

The time and amount of usage per week can be seen in Figure 26 here below. As expected, in accordance to Figure 24 here above, the frequency of usage is largely between 07:00 o'clock and 00:00 o'clock, with its highest peak at around 12:00 o'clock.

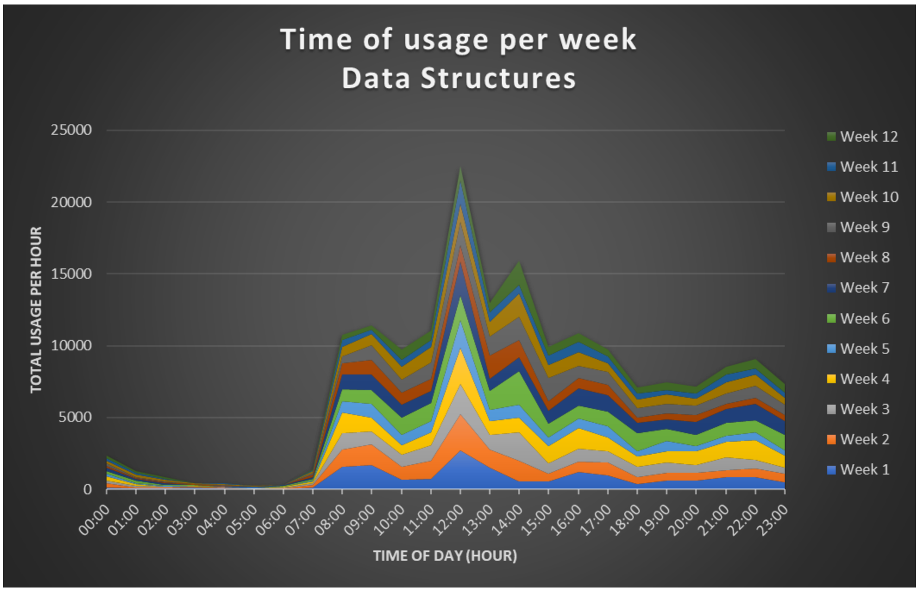

Figure 26. Data Structures - Time of usage per week

The difference in usage between weeks (shown in Figure 26 here above) is however insignificant, expect for week 1 and 2 where the recorded frequency increases compared to the usual usage.

4.2.6 Databases

The last course analysed was “Databases” (ísl. Gagnasafnsfræði), which is taught to undergraduate students on their second term. According to the data from Canvas, there were 215 students registered for the course, which was taught the time period of 08.January – 31.May 2019.

In Figure 27 here below can be seen the frequency of usage by students for the course on a daily basis, where the highest usage recorded 6,080 mouse-clicks and the lowest recorded 7 mouse-clicks over a 24-hour time period, with an average of approximately 1,259 mouse-clicks per day. As shown in the graph in Figure 27, there are around five large spikes, which can be explained by the five large group projects that students had to turn in during the course. There were also homework assignments that students had to turn in on a weekly basis, which could explain the smaller spikes seen within the graph. At the end of the course, students had to take a final exam, which explains the big spike in the beginning of April 2019, when students were preparing for the exam. In late April 2019 we can see another spike, which may be explained by the release of final grades.

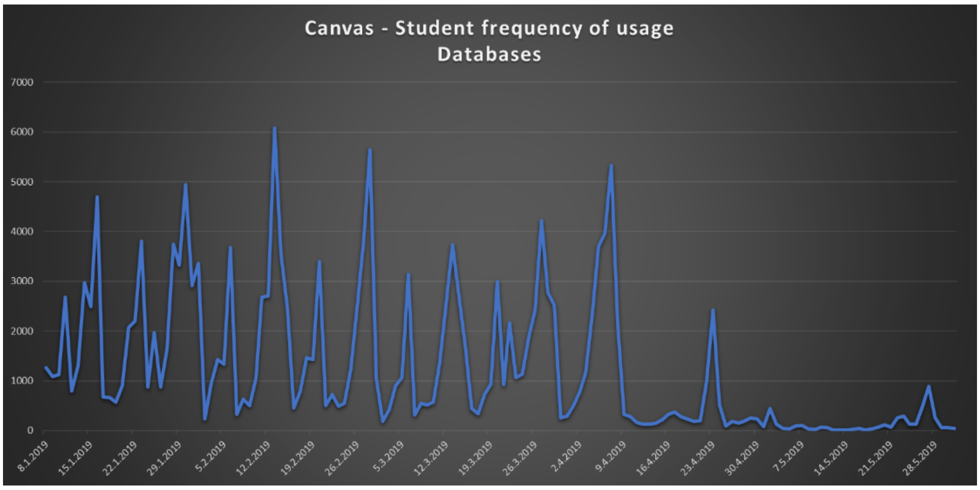

Figure 27. Databases - Canvas frequency

In Figure 28 here below can be seen the different usage of the course, within the three main pages: Assignments page (marked with a blue line), Grades page (marked with an orange line), and Modules page (marked with a grey line). As shown in the graph, the spike that occurred in late April 2019 in Figure 27, can be explained here in Figure 28 with students clicking on the Grades page of the course, most likely checking their final grade from the final exam. What is strange to see is the last spike in late May 2019, where students seem to be using the Modules page. This could be linked to retake exams, that took place around this time in 2019. Other interesting things seen within the graph is that the spikes of each of the three page seem to go hand in hand, that is first students check the Modules page, shortly after (a few days) they check the Assignment page more frequently, and thereafter an increase occurs on mouse-clicks for the Grades page. This may not come as a surprise, as materials from the instructors of the course are posted within the Modules page, then assignments that students have to do are posted and handed in within the Assignments page, and finally the grades for these assignments are posted within the Grades page.

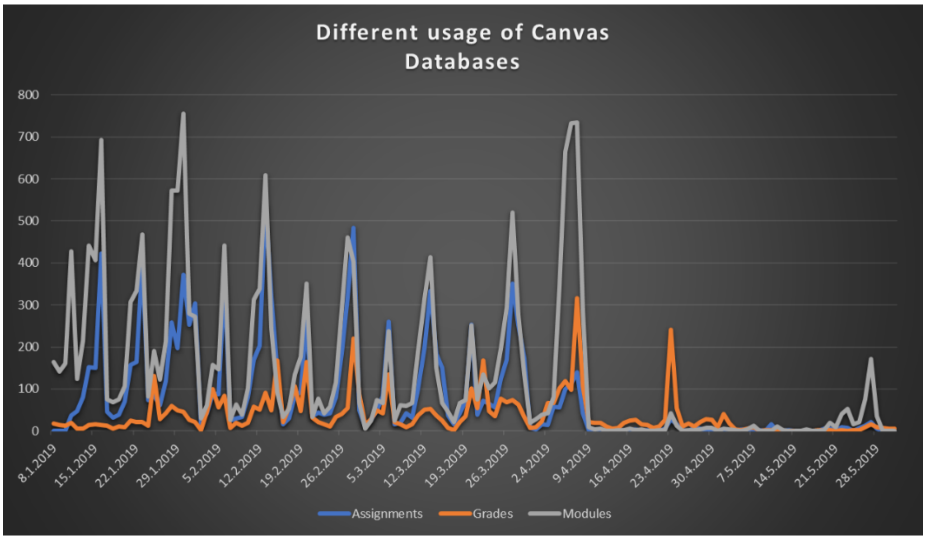

Figure 28. Databases - Usage of Canvas

In Figure 29 here below is shown the number of mouse-clicks according the time of day the clicks occurred within the 24-hour time period per day. As with the other aforementioned courses, there seems to be a trend between students using Canvas pages steadily over the course of 08:00 o'clock and 00:00 o'clock, and during the night time between 00:00 o'clock and 07:00 o'clock the activities drop significantly, though there are always some late night students using Canvas regularly.

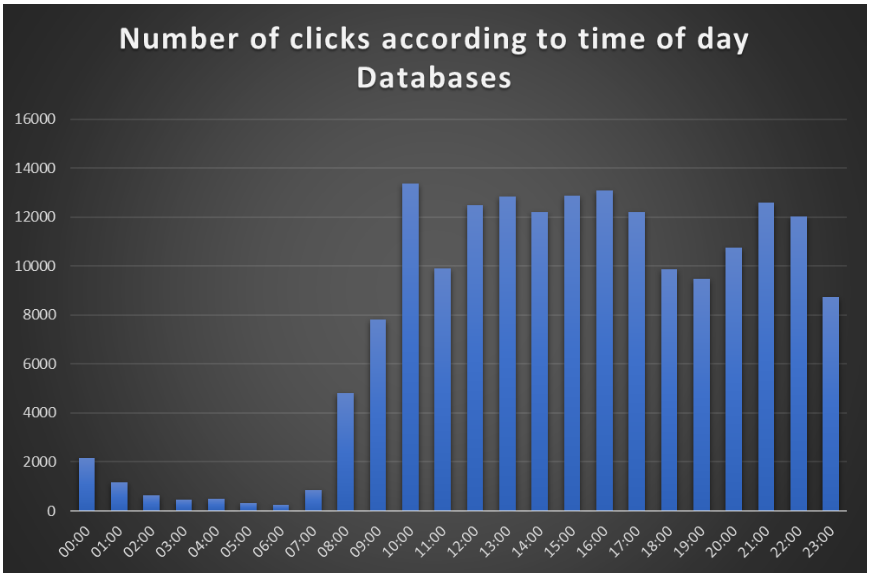

Figure 29. Databases - Time of usage

The total usage per week for the course can be seen in Figure 30 here below, where the highest frequency recorded for a single week was 16,238 mouse-clicks in week 3, and the lowest frequency recorded for a single week was 9,023 mouse-clicks in week 9. There are no abnormal sequences in regard to the usage per week, other than the large difference between highest and lowest frequency amongst weeks. This could be due to workload during different time periods within the course.

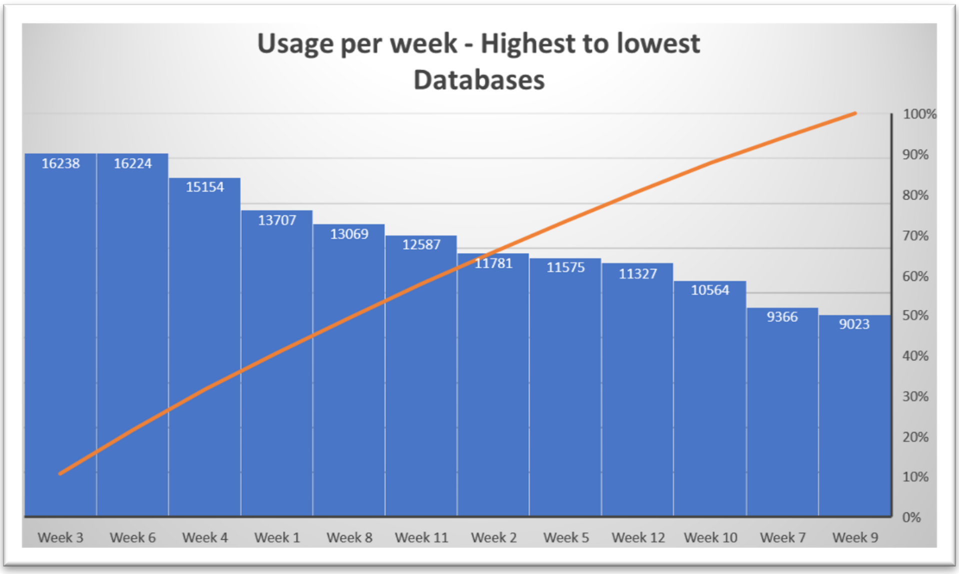

Figure 30. Databases - Usage per week

In Table 28 here below is shown the total frequency of usage per weekday and weeks, where the highest usage is recorded on Wednesdays with a total of 45,549 mouse-clicks, whereas the lowest usage occurs on Saturdays with only 6,384 mouse-clicks. It should also be noted that Thursdays, Fridays, Mondays, and Tuesdays have a high frequency of mouse-clicks as well, each day with more than 20,000 total clicks. Whereas on Sundays the frequency is quite low, similar to Saturdays, with under 10,000 clicks in total. This high difference between days is most likely due to deadlines of assignments and homework.

Table 28. Databases - Total usage per weekday

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Total:
Wednesday	1087	4693	3807	4942	3672	6080	3397	3798	3139	3730	2994	4210	45549
Thursday	1122	684	880	2910	326	3552	502	5646	319	2700	926	2770	22337
Friday	2680	665	1969	3360	636	2471	725	1038	545	1693	2166	2518	20466
Saturday	798	569	874	228	505	453	487	191	520	439	1067	253	6384
Sunday	1298	897	1637	946	1048	784	554	425	571	335	1133	288	9916
Monday	2968	2075	3747	1432	2682	1459	1223	902	1374	730	1877	503	20972
Tuesday	3754	2198	3324	1336	2706	1425	2478	1069	2555	937	2424	785	24991
Total:	13707	11781	16238	15154	11575	16224	9366	13069	9023	10564	12587	11327	

At last, in Figure 31 here below, can be seen the time and amount of usage per week for the course. As shown within the graph, the amount of usage is higher for the first few weeks of the course, and little by little seems to diminish as weeks go by. It can also be seen that the time of usage is, like was shown in Figure 29 here above, mostly occurring between 07:00 o'clock in the morning until around 00:00 in the night.

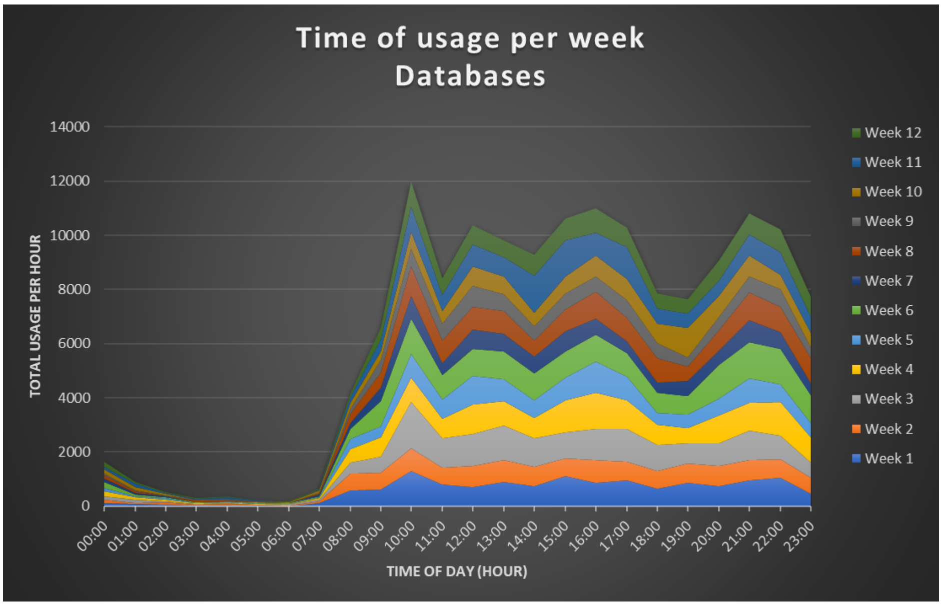

Figure 31. Databases - Time of usage per week

Putting it all together, the following chapter covers the main findings within the analysis done in R, in accordance to the data analysis covering the six courses which has been described in the previous chapters.

4.2.7 Main Summary of Descriptive Statistics

From reviewing the descriptive statistics in general, a few similarities and differences can be outlined. To start with, when comparing the six courses and the student frequency of usage per day (see Figure 32) there are not significant differences. All courses show peaks, and the number of peaks can be related to the amount of assignments included in each course. Thereby, a conclusion of this is that the usage of Canvas increases during assignments and projects periods and in relation to the deadlines of these.

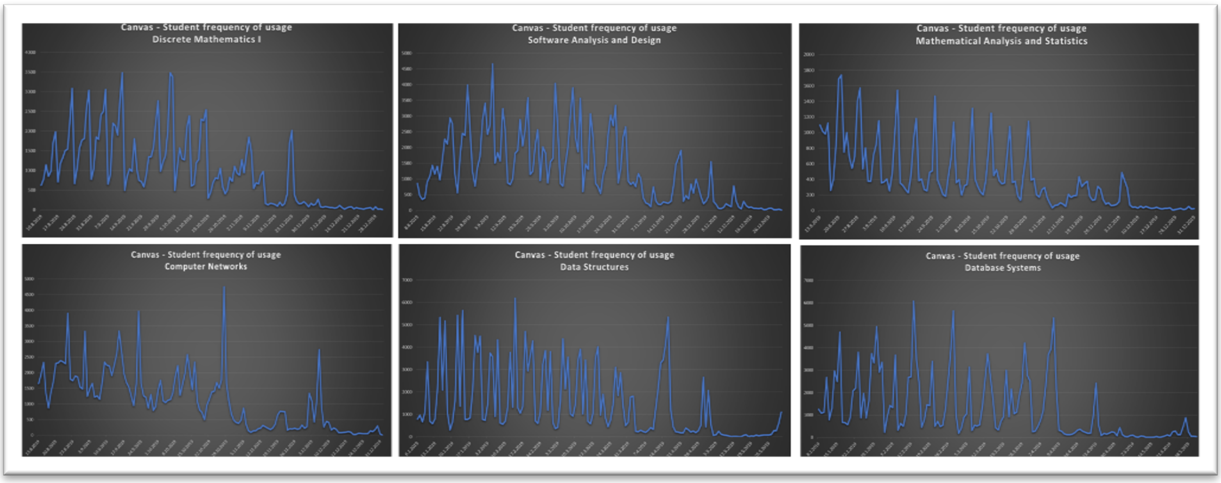

Figure 32. Comparison between course frequency of usage

Figure 33 shows that students are using similar methods in regard to using Canvas while comparing these six courses. The only difference being that some pages in a specific course may have more peaks than in other courses, for example the page Modules in the course “Data Structures”, which can be explained by the lecture captures, that could be watched by the students from Modules in Canvas, thereby explaining more frequent intervals shown in usage for this particular course and the Module page therein.

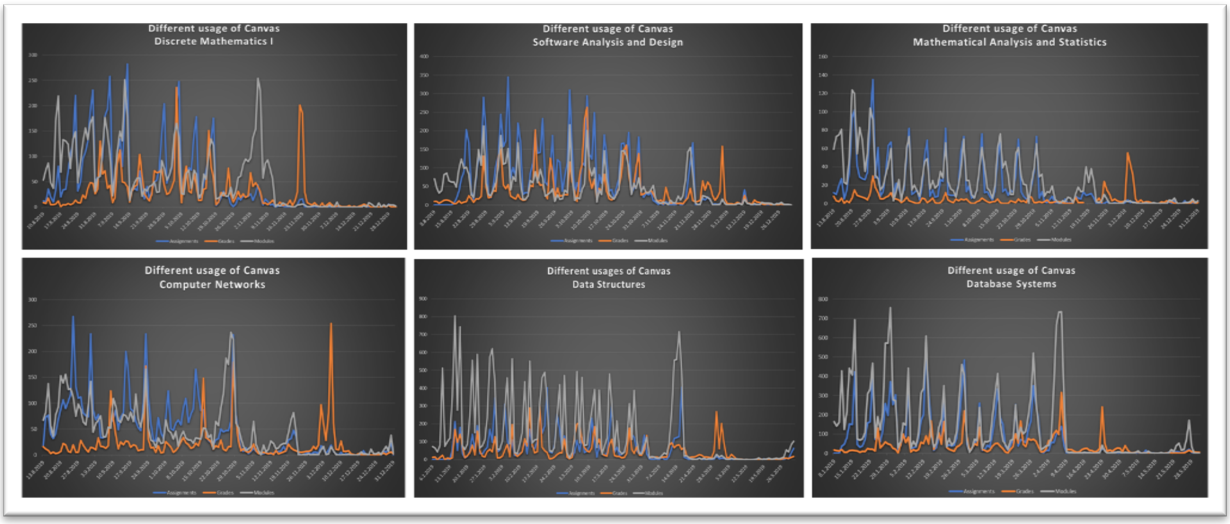

Figure 33. Comparison between different usage of Canvas within courses

In regard to what time of day the students display the most activity on Canvas, all the courses show similar trends (see Figure 34). The activity starts around 08:00 o'clock, increases throughout the day until reaching its peak around mid-day and afterwards holding its frequency of usage until around 00:00 o'clock, when the usage decreases significantly throughout the night. The learning point here being that the usage spans over the course of the whole day, until late night, but even after midnight there is some recorded activity (even though very small activity).

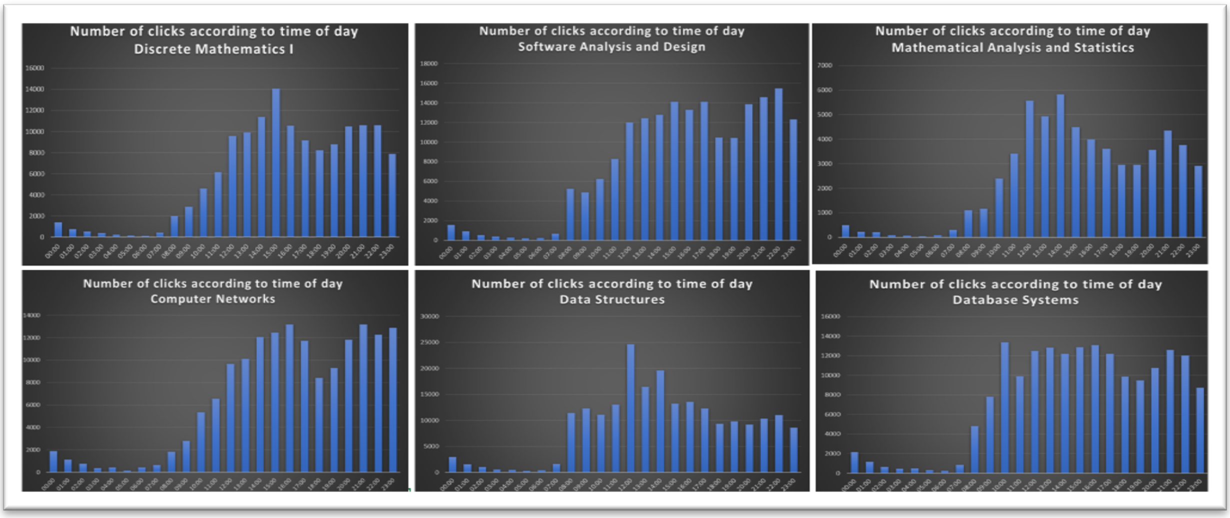

Figure 34. Comparison between time of usage within courses

When comparing the six courses, we see the frequency of use in the course "Data Structures" to be significantly higher overall than in the other five courses. Also, the similarity can be derived (though excluding Data Structures), about the usage per week that covers the whole day, excluding a down-swing around evening time, and then picking up again until midnight. However, in the course of Data Structures, the peak is around 12:00 o'clock, then reduces, until it drops completely around 00:00 o'clock in the night, thereby not showing an increase in usage during the day, nor the evening as is noticeable in the other courses.

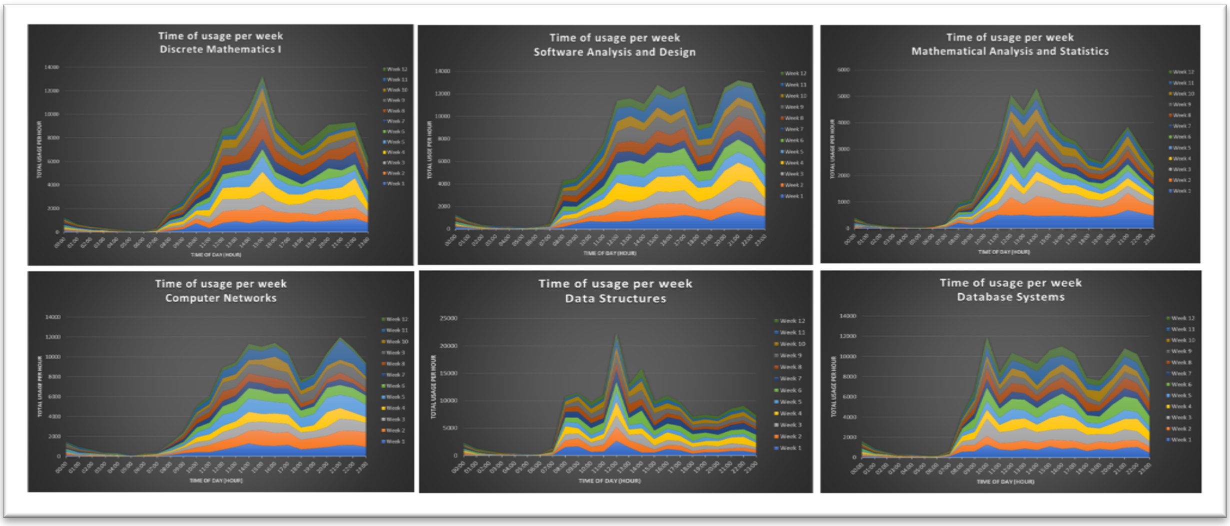

Figure 35. Comparison between time of usage per week

Next chapter will report in-depth on what the data can really tell us, by applying machine learning to analyse the learning patterns and behaviours that can be derived from the use of Canvas, to better understand what type of patterns are beneficial towards a high final grade.

4.3 Machine Learning

The data used for the analysis done in the following chapter is based on five courses: i) Discrete Mathematics I, ii) Software Requirements and Design, iii) Calculus and Statistics, iv) Computer Networks, v) Data Structures. Unfortunately, the course Data Structures was not used herein due to insufficient data regarding final grades of students in the course. Due to that limitation, and because the aim of the following analysis is to predict final grade scores as a complete dataset with final grades was needed for that analysis, the course was disregarded.

Before the process of machine learning could take place, the Canvas data from all five courses had to be extensively explored in order to establish both independent and dependent variables worth pursuing and analysing. The goal being to create a single data set consisting of relevant variables that could be imported to a machine learning environment for further analysis. This was done through Python programming using the Pandas software library for data manipulation and analysis.

By extracting the raw data from each of the five courses separately, a series of smaller datasets were constructed. Each of these smaller datasets contained different variables offering various sorts of information. The data sets provided some clarity to the raw data and presented the possibility of exploratory data analysis. The process of constructing these smaller data sets involved a lot of trial and error as there where many different combinations of variables and possible angles that seemed to be worth exploring through machine learning. Unfortunately, that was not always the case as further testing in machine learning did not deliver the desired results and the whole process had to be repeated. This iterative data preparation process (see Figure 36) consisted of defining new variables and creating possible data sets through Python programming, testing them in through machine learning and observing the results.

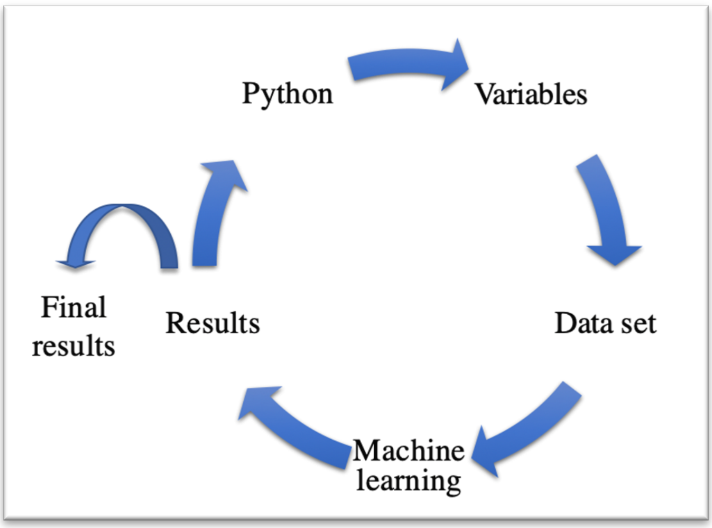

Figure 36. Data preparation process

After going through several cycles of the data preparation process, the data sets perceived to be of most relevance towards the desired end result were ultimately merged into a single data set. The variables presented in this singular dataset were the same for each respective course and the following table contains further details of each one.

Table 29. Detailed description of course dataset variables

Variable	Type	Description	Used in Machine Learning
user_id	Numeric (integer)	The personal university id number of every student registered in the respective course.	No
date_count	Numeric (integer)	Total number of different dates were a respective student interacted with Canvas in a respective course. From the beginning of the course until the end of the official final exam period.	Yes
total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the respective course on Canvas. From the beginning of the course until the end of the official final exam period.	Yes
assignment_clicks	Numeric (integer)	Total number of clicks made on the Canvas module "Assignments" by a respective student in a respective course.	Yes
modules_clicks	Numeric (integer)	Total number of clicks made on the Canvas module "Modules" by a respective student in a respective course.	Yes
grades_clicks	Numeric (integer)	Total number of clicks made on the Canvas module "Grades" by a respective student in a respective course.	Yes
week_1_total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the first week of the respective course on Canvas.	Yes
week_2_total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the second week of the respective course on Canvas.	Yes
week_3_total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the third week of the respective course on Canvas.	Yes
week_4_total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the fourth week of the respective course on Canvas.	Yes
week_5_total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the fifth week of the respective course on Canvas.	Yes
week_6_total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the sixth week of the respective course on Canvas.	Yes
week_7_total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the seventh week of the respective course on Canvas.	Yes
week_8_total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the eighth week of the respective course on Canvas.	Yes
week_9_total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the ninth week of the respective course on Canvas.	Yes
week_10_total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the tenth week of the respective course on Canvas.	Yes
week_11_total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the eleventh week of the respective course on Canvas.	Yes
week_12_total_clicks	Numeric (integer)	Total number of clicks made by a respective student in the twelfth week of the respective course on Canvas.	Yes

average_clicks_per_week	Numeric (non-integer)	Average number of total clicks made in week by a respective student in a respective course. From the beginning of the course until the beginning of the final exam period (weeks 1 - 12).	Yes
max_clicks_per_week	Numeric (integer)	The maximum value of clicks made in a week by a respective student in a respective course. From the beginning of the course until the beginning of the final exam period (weeks 1 - 12).	Yes
min_clicks_per_week	Numeric (integer)	The minimum value of clicks made in a week by a respective student in a respective course. From the beginning of the course until the beginning of the final exam period (weeks 1 - 12).	Yes
stdev_clicks_per_week	Numeric (non-integer)	The standard deviation of total clicks made in a week by a respective student in a respective course. From the beginning of the course until the beginning of the final exam period (weeks 1 - 12).	Yes
weekend_use_ratio	Numeric (non-integer)	The ratio in which a respective student interacted with Canvas in a respective course during a weekend. From the beginning of the course until the beginning of the final exam period (weeks 1 - 12).	Yes
business_day_use_ratio	Numeric (non-integer)	The ratio in which a respective student interacted with Canvas in a respective course on a normal business day. From the beginning of the course until the beginning of the final exam period (weeks 1 - 12).	Yes
night_use_ratio	Numeric (non-integer)	The ratio in which a respective student interacted with Canvas in a respective course from 00:00 – 05:59. From the beginning of the course until the beginning of the final exam period (weeks 1 - 12).	Yes
morning_use_ratio	Numeric (non-integer)	The ratio in which a respective student interacted with Canvas in a respective course from 06:00 – 11:59. From the beginning of the course until the beginning of the final exam period (weeks 1 - 12).	Yes
afternoon_use_ratio	Numeric (non-integer	The ratio in which a respective student interacted with Canvas in a respective course from 12:00 – 18:59. From the beginning of the course until the beginning of the final exam period (weeks 1 - 12).	Yes
evening_use_ratio	Numeric (non-integer	The ratio in which a respective student interacted with Canvas in a respective course from 19:00 – 23:59. From the beginning of the course until the beginning of the final exam period (weeks 1 - 12).	Yes
on_campus_use_ratio	Numeric (non-integer	The ratio in which a respective student interacted with Canvas in a respective course while connected to the university's wireless network. From the beginning of the course until the beginning of the final exam period (weeks 1 - 12).	Yes
off_campus_use_ratio	Numeric (non-integer	The ratio in which a respective student interacted with Canvas in a respective course while not connected the university's wireless network. From the beginning of the course until the beginning of the final exam period (weeks 1 - 12).	Yes
final_score	Numeric (non-integer)	The numerical final grade value of a respective student in a respective course.	No

course_passed	Boolean value	True/False value depending on whether or not a respective student got a passing final grade in a respective course.	No
grade_placement	String	A value of either A, B, C or D, depending on a respective student's final grade in a respective course.	Yes

As stated above in Table 29, not all variables in the final course data set were used in the following machine learning process. The variable *user_id* was excluded from the process as it contains the personal university identification number of a respective student. Since no two students studying at Reykjavik University can have the same identification number, the variable offers no further value. The *final_score* variable was excluded as well, but on different grounds. Initially, the *final_score* variable was used as the dependent variable with the idea being to create a trained regression-based machine learning model that could predict the final score of a test set with sufficient accuracy. However, as further testing was conducted it became evident the task of predicting the exact numerical final grade of a respective student in the test set was too great compared to the small size of the actual course data set. It was then decided that the machine learning model should be classification-based with the dependent variable being *grade_placement*. In turn the variable *course_passed* was excluded since it is in direct correlation with one of the possible values of the *grade_placement* variable and would, if included, skew the final results from the machine learning process.

As mentioned above in Table 29 the *grade_placement* variable for each respective student, in a respective course, could only display one of four different values A, B, C or D, where each possible value represented a predefined range of final grade scores. Each possible value thus represented a possible grade category. The same definition of categories was applied to all five courses and will hereinafter be referred to as the Original Grade Categories.

When researched further it soon became apparent that it might not be best to unilaterally apply the same grade categories to all five courses as they may vary in levels of difficulty. In turn, the number and overall ratio of students in each category can vary greatly between courses. It was then decided that in addition to conducting machine learning from the perspective of Original Grade Categories another version should be created and analysed, containing a more equal distribution between grade categories. This meant redefining the final score range for each category A, B, C, D with the ratio of students in each one being as close to 25% as possible. The exception being that grade category D would be left unchanged, continuing to exclusively represent a failing final score in a respective course. This redefinition of grade categories was specifically tailored to each and every course and will hereinafter be referred to as the Modified Grade Categories. The following tables show how the ratio between categories in the Original Grade Categories could be adjusted in order to create the Modified Grade Categories in the course Discrete Mathematics I.

Table 30. Original Grade Categories in Discrete Mathematics I

Grade Category	Final Score range (0-10)	Rounded Final Score range (0-10)	Number of students	Ratio
A	> 8.74	>= 9	25	11%
B	6.75 – 8.74	7 – 8.5	81	36%
C	4.75 – 6.74	4.5 – 6.5	65	29%
D	< 4.75	< 4.5	52	24%

Table 31. Modified Grade Categories in Discrete Mathematics I

Grade Category	Final Score range (0-10)	Rounded Final Score range (0-10)	Number of students	Ratio
A	> 8.24	>= 8.5	49	22%
B	6.75 – 8.24	7 - 8	57	25%
C	4.75 – 6.74	4.5 – 6.5	65	29%
D	< 4.75	< 4.5	52	24%

The machine learning process was conducted using Azure Machine Learning Studio. Two datasets where imported for each of the five courses, one containing Original Grade Categories, the other containing Modified Grade Categories. In accordance with Table 29 here above, variables considered to be of no further use where excluded from the course dataset. The dataset was randomly split into two parts with 70% of the overall data being used for training the model and 30% being used for testing the model. The split was stratified by the *grade_placement* variable to prevent under sampling one or more grade categories. As mentioned before, the ultimate goal of the machine learning process was to create and train a machine learning model that could sufficiently predict the *grade_placement* values of students in a respective course, as well as computing the permutation feature importance of feature scores. This would reveal the variables that contribute the most positive effects towards the accuracy of the model.

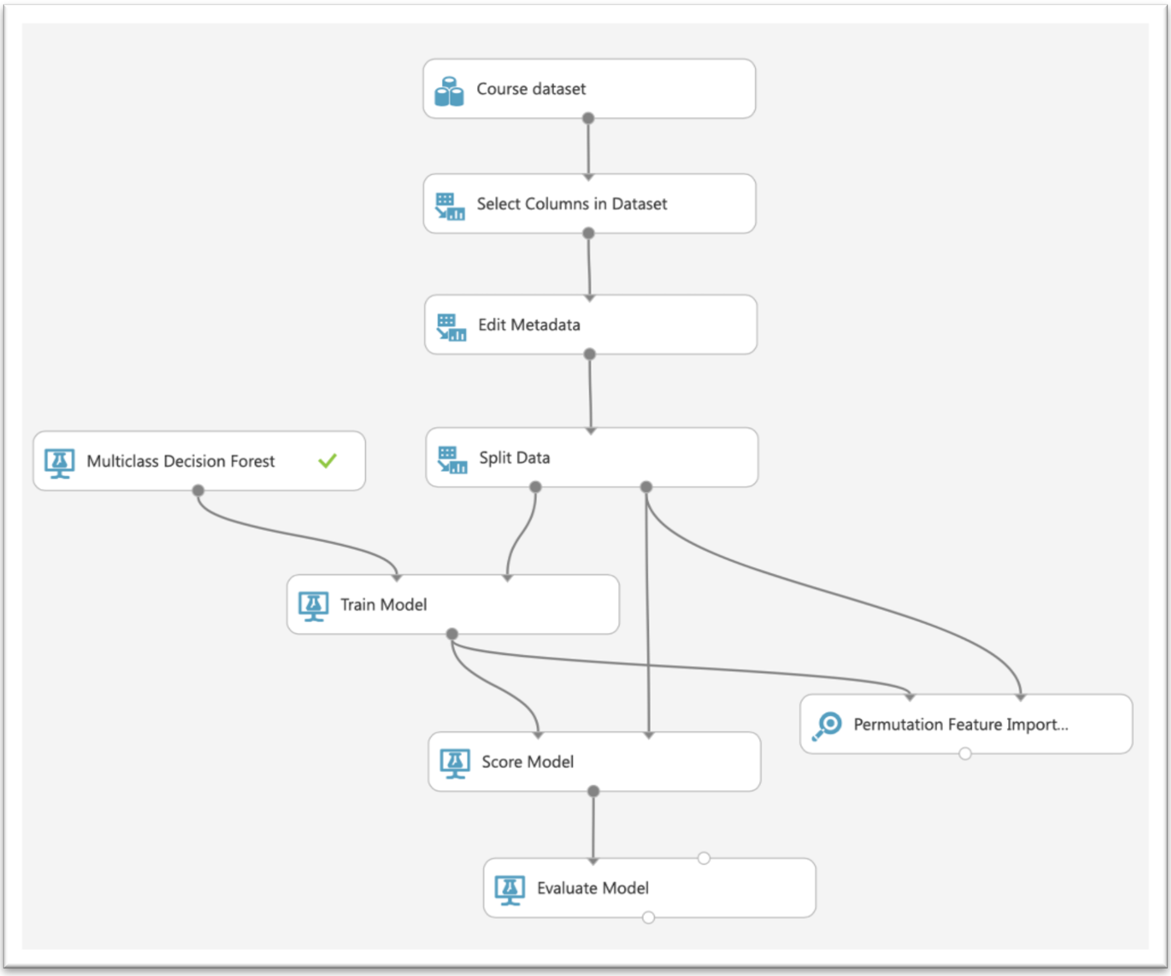

Figure 37. Depicting the Machine Learning Process.

Since *grade_placement* has four possible values, different multiclass algorithms were explored and contemplated. Eventually, two algorithms were selected to go head to head, with further analysis being conducted on the results from the superior one for each individual course. These chosen algorithms were Multiclass Decision Forest and Multiclass Decision Jungle (see example in Figure 37). Additionally, as two versions of datasets existed for each of the five courses, one using Original Grade Categories and the other Modified Grade Categories, there were in turn two separate versions of results. The following two chapters provide a detailed breakdown of the final results gathered from both versions.

4.3.1 *Original Grade Categories*

The Original Grade Categories represent the non-course specific definition of final score ranges for each category A, B, C, D, in the *grade_placement* variable. This definition was the same across all five courses and can be seen in Table 30 here above. It focused solely on the final score range in each category with D representing a failing score, C being a barely passing or average score, B being a good or an above average score and A representing an exceptional score.

4.3.1.1 *Accuracy Results*

The accuracy results from the machine learning analysis, using the Original Grade Categories version of course datasets, are as follows:

Table 32. Original Grade Categories - Comparison and Accuracy Results

	Multiclass Decision Jungle		Multiclass Decision Forest	
Course	**Overall Accuracy**	**Average Accuracy**	**Overall Accuracy**	**Average Accuracy**
Discrete Mathematics I	0.485	0.742	0.394	0.697
Software Requirements and Design	0.523	0.761	0.521	0.761
Calculus and Statistics	0.333	0.667	0.344	0.671
Computer Networks	0.459	0.729	0.377	0.689
Databases	0.453	0.727	0.484	0.742

Table 33. Original Grade Categories - Best Accuracy Scores

Course	Overall Accuracy	Average Accuracy	Better Fitting Multiclass Algorithm
Discrete Mathematics I	0.485	0.742	Decision Jungle
Software Requirements and Design	0.523	0.761	Decision Jungle
Calculus and Statistics	0.344	0.671	Decision Forest
Computer Networks	0.459	0.729	Decision Jungle
Databases	0.453	0.727	Decision Forest
Average:	**0.453**	**0.726**	

In terms of the algorithm debate (see Table 32), there is no clear winner since the algorithm that provides the better accuracy for one course does not necessarily do so for the other. The model is scored on the basis of overall accuracy and average accuracy. Overall accuracy (OA) represents the ratio of correctly predicted items in the test dataset while average accuracy (AA) displays the average of each accuracy per class, each class being one of the four values in the *grade_placement* variable. Specifically, the grade categories A, B, C and D. Accuracy results from the machine learning process (see Table 33) provided a decent average OA of 0.458 or 45.8% with an average AA of 0.726 or 72.6% for all five courses.

4.3.1.2 Feature Importance

The next step of the machine learning process was to compute the permutation feature importance of feature scores. This meant analysing the variables that had the most positive effect towards the accuracy of the better fitting machine learning model for each course. Theoretically, this step would shed light on the most important factors of every course individually. The results of these computation can be seen in the following Figures 38-42.

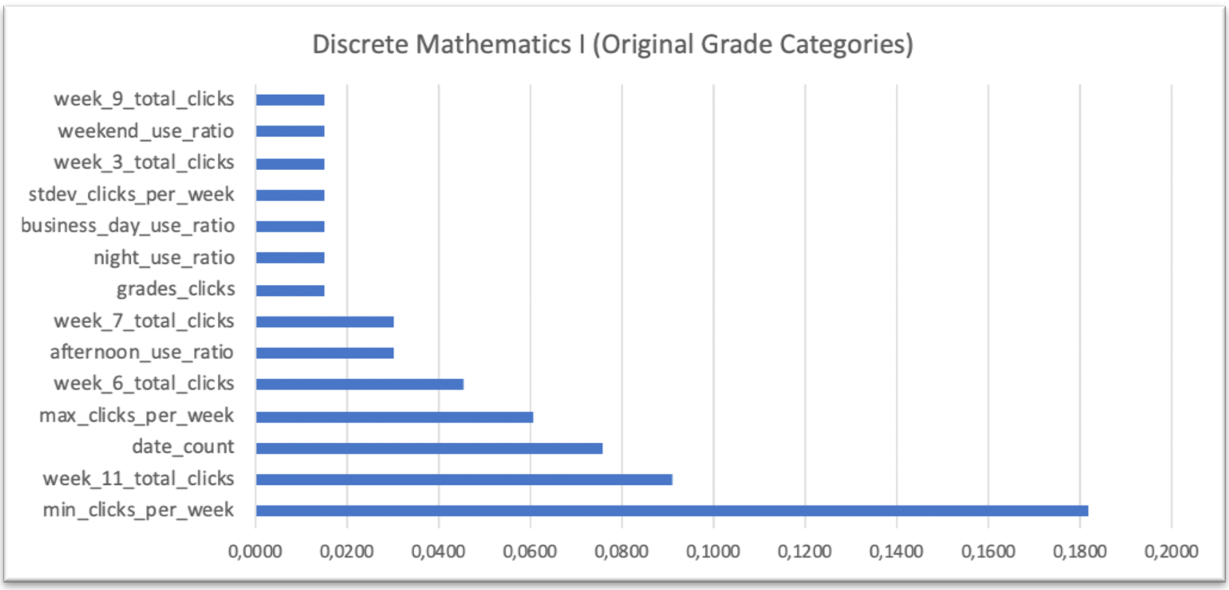

Figure 38. Discrete Mathematics I (Original) - Positive Feature Importance Scores

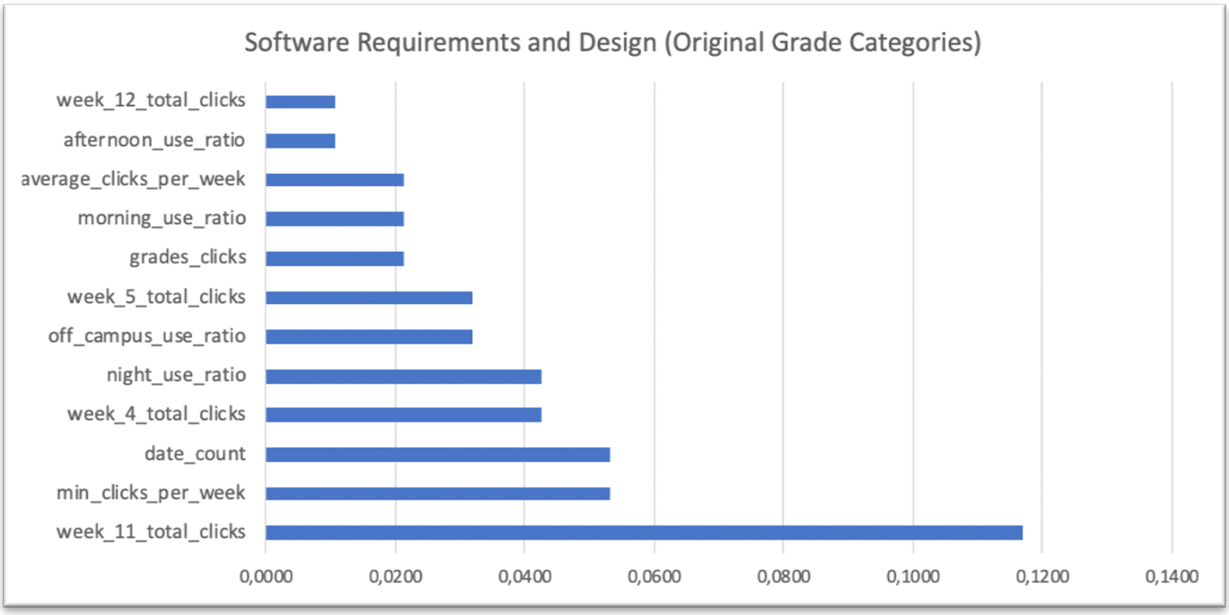

Figure 39. Software Requirements and Design (Original) - Positive Feature Importance Scores

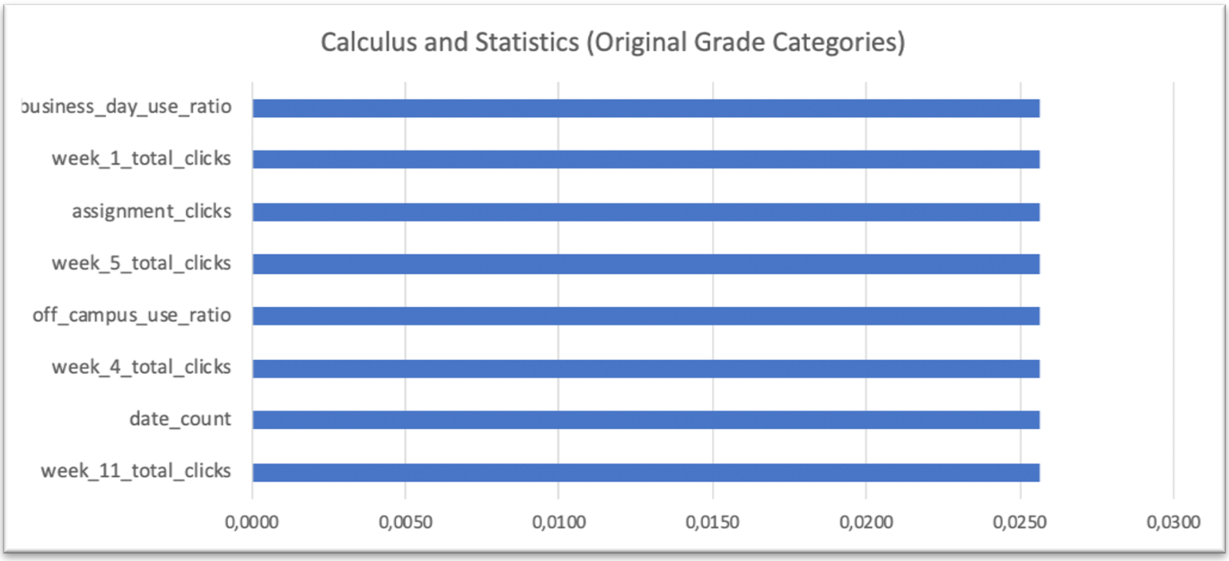

Figure 40. Calculus and Statistics (Original) - Positive Feature Importance Scores

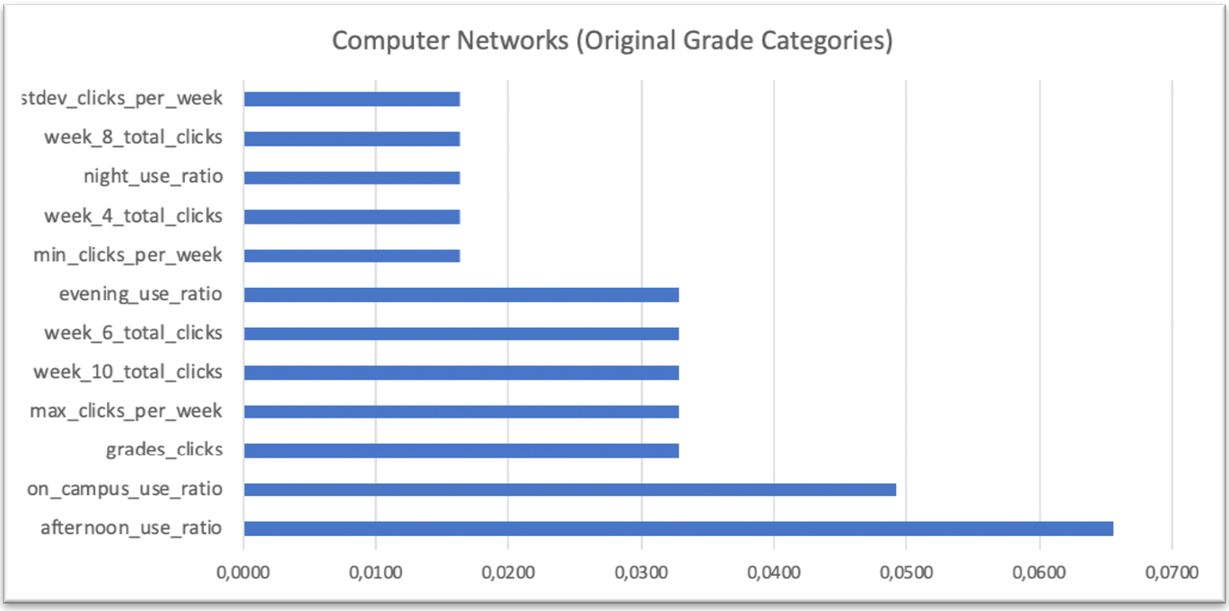

Figure 41. Computer Networks (Original) - Positive Feature Importance Scores

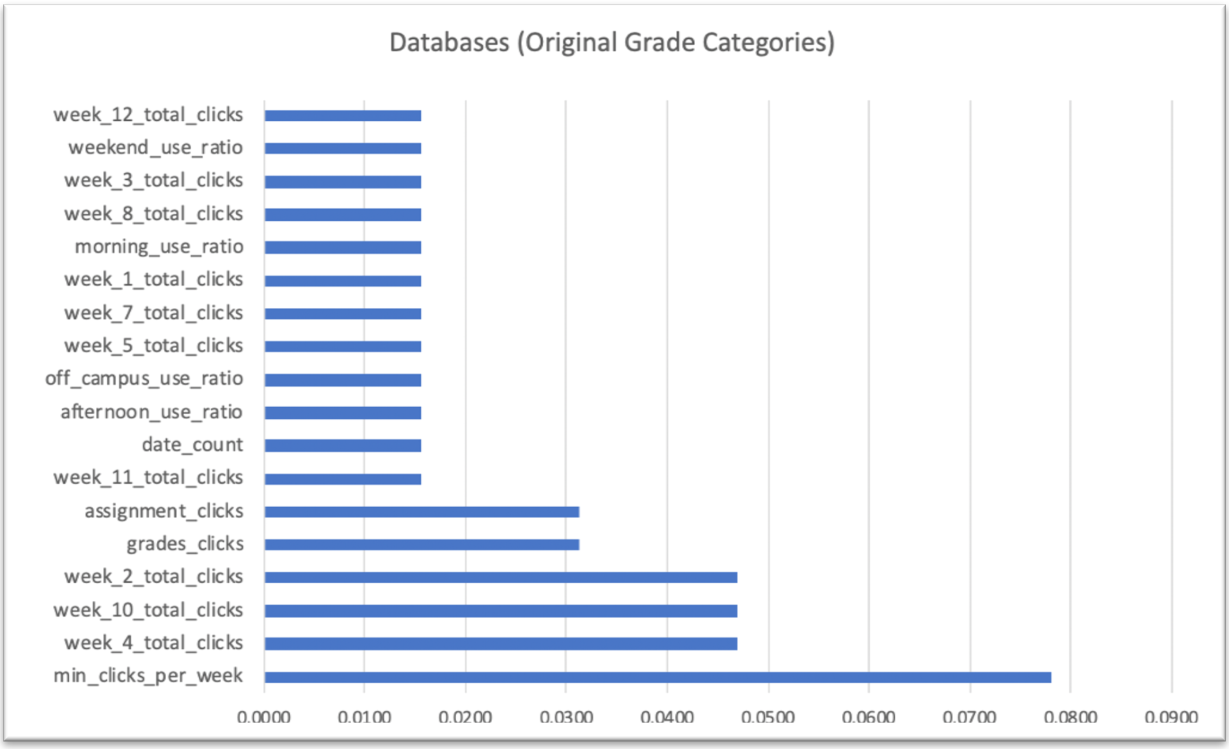

Figure 42. Databases (Original) - Positive Feature Importance Scores

4.3.1.3 Summary Results

Further analysis and comparison of positive feature scores in each of the five courses (see Figures 38-42), made it apparent that there are certain variables that consistently provide a higher positive value score. Similarly, there are variables that have no or next to no effect on the machine learning model. When looking at the accuracy results and feature importance scores for every individual course side by side there seems to be a certain unpredictability present in the Calculus and Statistics course. The accuracy scores are visibly lower than in the other courses (see Table 33) and feature importance scores quite distinct as listed variables display the same miniscule importance score (see Figure 40). It is also interesting to see the feature importance scores for Computer Networks, as the higher rating variables (see Figure 41) are dissimilar to those found in the other four courses.

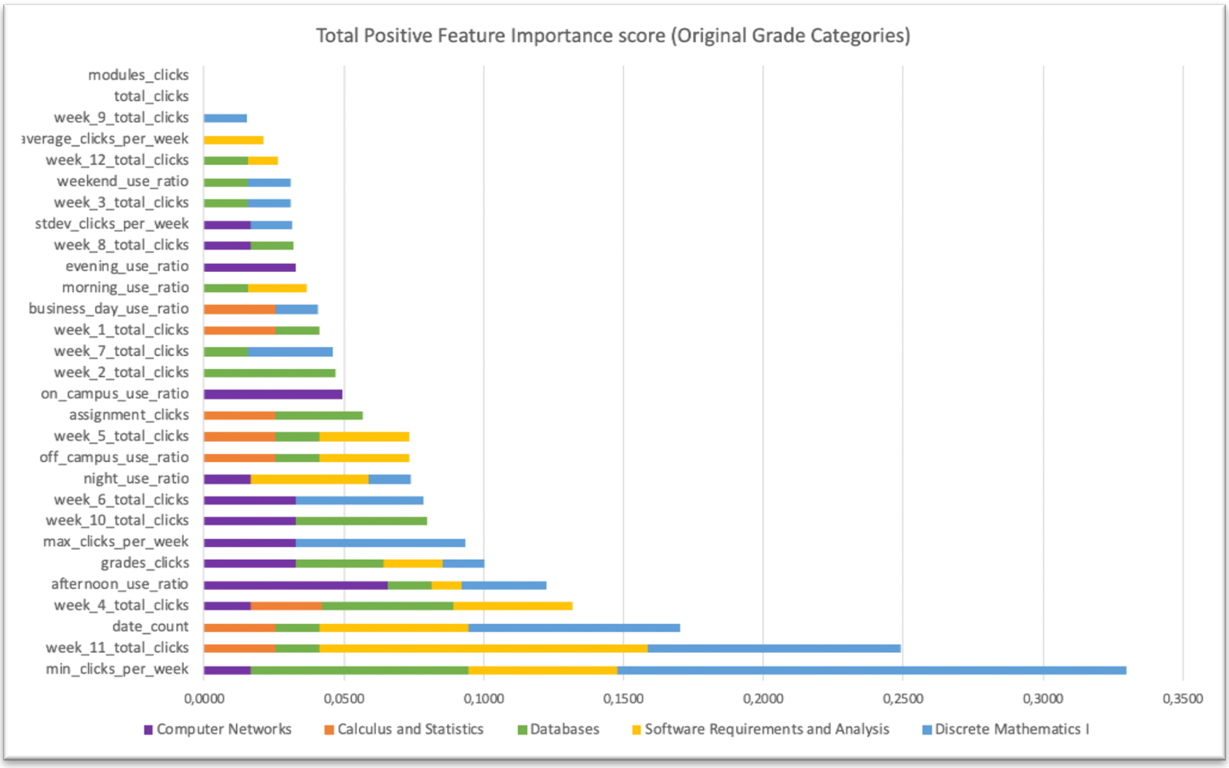

Figure 43. Total Positive Feature Importance Scores (Original)

As expected, there are quite a few variables out of the 29 used in the machine learning model, that have significantly low impact across all courses (see Figure 43). These variables are thus considered to be of no further importance in terms of identifying deciding factors. However, the five variables that show the most overall positive effect are *min_clicks_per_week*, *week_11_total_clicks*, *date_count*, *week_4_total_clicks* and *aternoon_use_ratio.* Details of each variable can be found above in Table 29.

Although the positive effect for each of these variables differs between courses, it is consistently high. These distinctively positive variables are therefore the main defining factors that the model takes into account when attempting to accurately predict a student's final grade category. It can then be said that the activity displayed by a respective student in relation to these key variables will have a significant effect towards his/her final grade category.

4.3.2 Modified Grade Categories

As opposed to the Original Grade Categories, the Modified Categories version was course specific, with the aim of equalizing the amount and ratio of students in each grade category A, B, C, and D. One exception was made in regard to grade category D which was excluded from the equalization process. Consistently throughout the five courses, grade category D already contained around 25% of the total number of students. Another incentive was that grade category D explicitly represented students with a failing final grade score in their respective course. Equalization for the D category was thereby deemed unnecessary and remained unchanged, directing all focus towards adjusting the ratio of the other three A, B, and C, so that they would each reflect around 25% of registered students as well (see Tables 30 and 31).

4.3.2.1 Accuracy Results

The accuracy results from the machine learning analysis using the Modified Grade Categories version of course datasets, are as follows:

Table 34. Modified Grade Categories - Comparison and Accuracy Results

	Multiclass Decision Jungle		Multiclass Decision Forest	
Course	**Overall Accuracy**	**Average Accuracy**	**Overall Accuracy**	**Average Accuracy**
Discrete Mathematics I	0.433	0.716	0.329	0.664
Software Requirements and Design	0.347	0.674	0.421	0.711
Calculus and Statistics	0.282	0.641	0.341	0.669
Computer Networks	0.311	0.658	0.344	0.672
Databases	0.347	0.677	0.375	0.688

Table 35. Modified Grade Categories - Best Accuracy Scores

Course	Overall Accuracy	Average Accuracy	Better Fitting Multiclass Algorithm
Discrete Mathematics I	0.433	0.716	Decision Jungle
Software Requirements and Design	0.421	0.711	Decision Forest
Calculus and Statistics	0.341	0.669	Decision Forest
Computer Networks	0.344	0.672	Decision Forest
Databases	0.375	0.688	Decision Forest
Average:	**0.383**	**0.691**	

In this Modified Grade Categories version of machine learning tests, the Multiclass Decision Forest algorithm provides superior accuracy for all courses but one, which can relate to the added level of difficulty in predictions. After all, the ratios for each category where adjusted to be more equalized making it harder for the model to pinpoint defining qualities between students, which was expected. However, the model still returns a fair score for both Overall Accuracy (OA) and Average Accuracy (AA). For all five courses using the Modified Grade Categories the average OA is 0.383 or 38.3% while the average AA is 0.691 or 69.1% (see Table 35).

4.3.2.2 Feature Importance

As before, the next step of the process was to compute the permutation feature importance of feature scores. It must however be stated that as the categories become more equalized, the harder it is for the model to make distinctive predictions. The results of feature importance scores for the course data sets using Modified Grade Categories are then to be taken with a grain of salt. The main purpose of this feature importance exploration was to identify possible similarities with the feature scores presented in the Original Grade Categories version, that are considerably more reliable towards reflecting actual course data. The results of these computation can be seen in the following Figures 44-48.

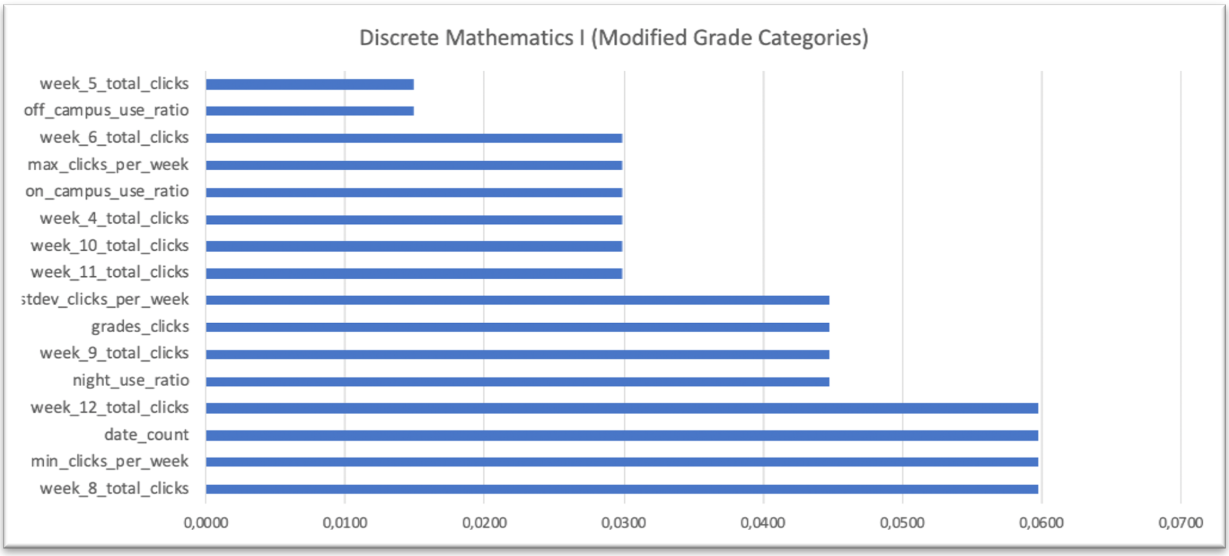

Figure 44. Discrete Mathematics I (Modified) - Positive Feature Importance Scores

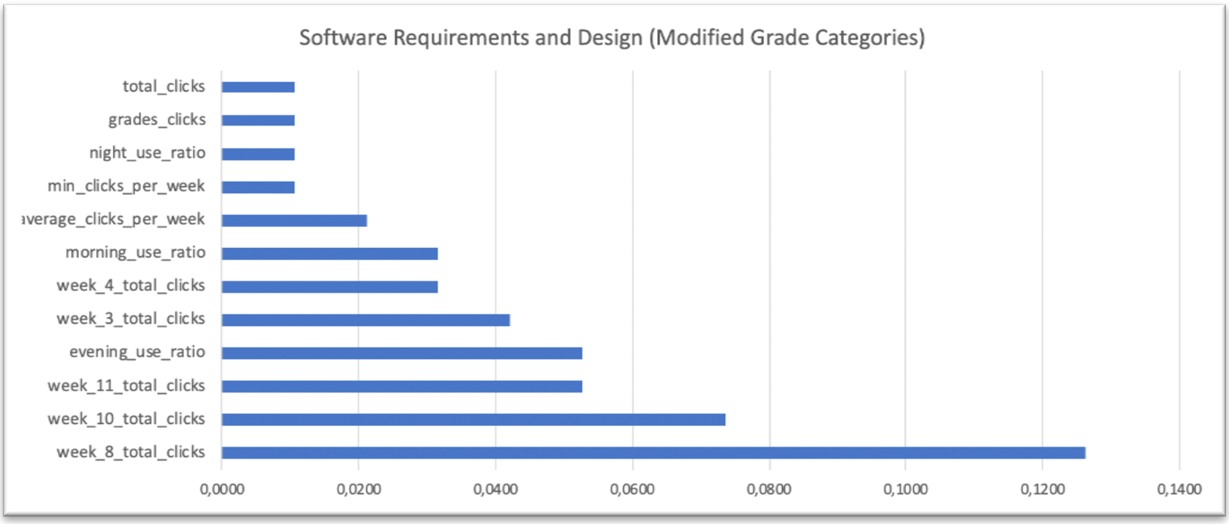

Figure 45. Software Requirements and Design (Modified) - Positive Feature Importance Scores

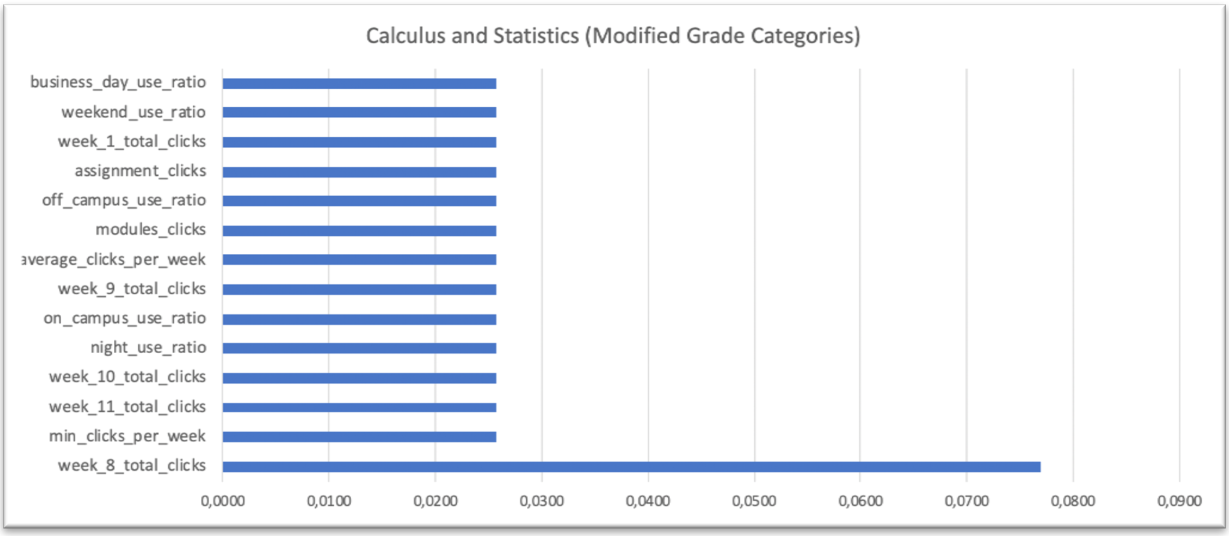

Figure 46. Calculus and Statistics (Modified) - Positive Feature Importance Scores

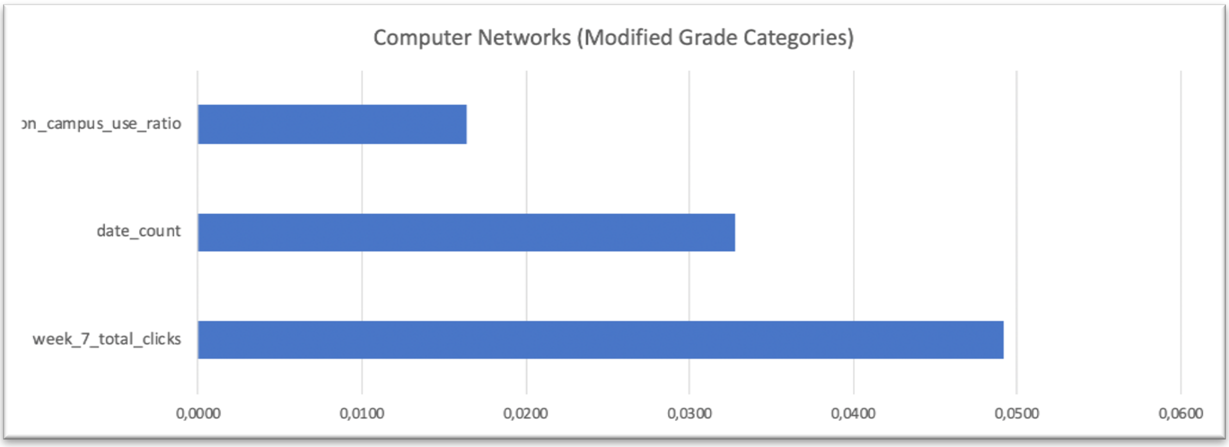

Figure 47. Computer Networks (Modified) - Positive Feature Importance Scores

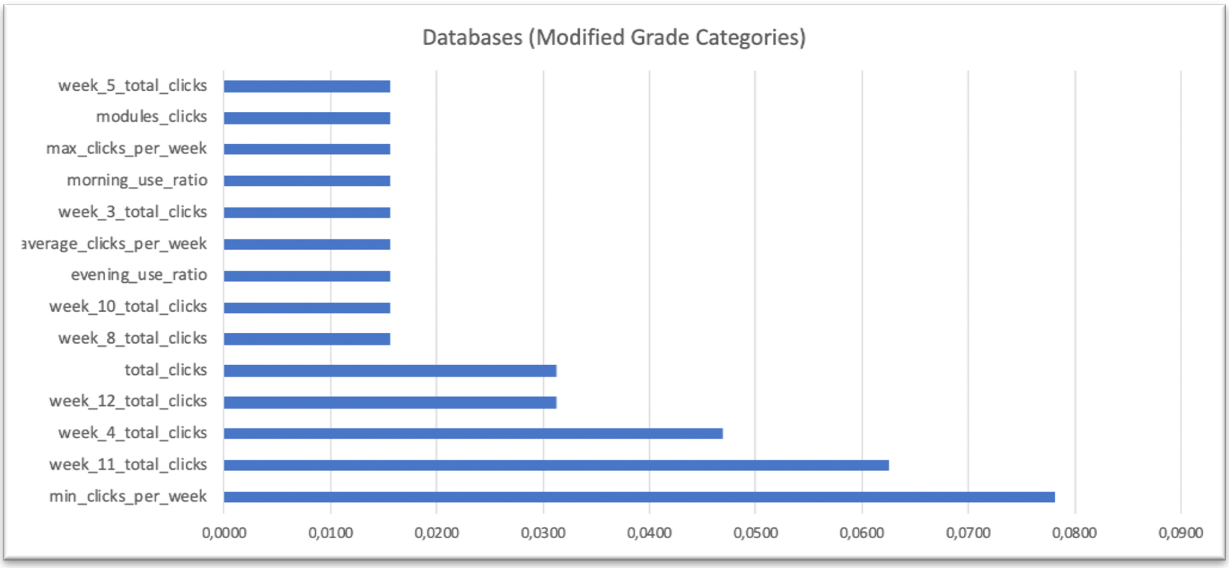

Figure 48. Databases (Modified) - Positive Feature Importance Scores

4.3.2.3 Summary Results

As stated before, the model experienced increased difficulty in figuring out distinctive variables in terms of grade categories, due to the more equalized basis. However, despite the drastic adjustments in ratio displayed by the Modified Grade Categories version, the average accuracy scores (see Table 35) remain fair and indeed comparable to those presented in the Original Grade Categories (see Table 33).

As with feature importance scores for each course, the model especially struggles in defining variables for the course Computer Networks, listing only three possibly defining variables (see Figure 47). Additionally, in the Calculus and Statistics course, the majority of listed variables depicts the same miniscule positive score (see Figure 46) due to the unpredictable nature of the course itself. This is innkeeping with previous observations from the Original Grade Categories version.

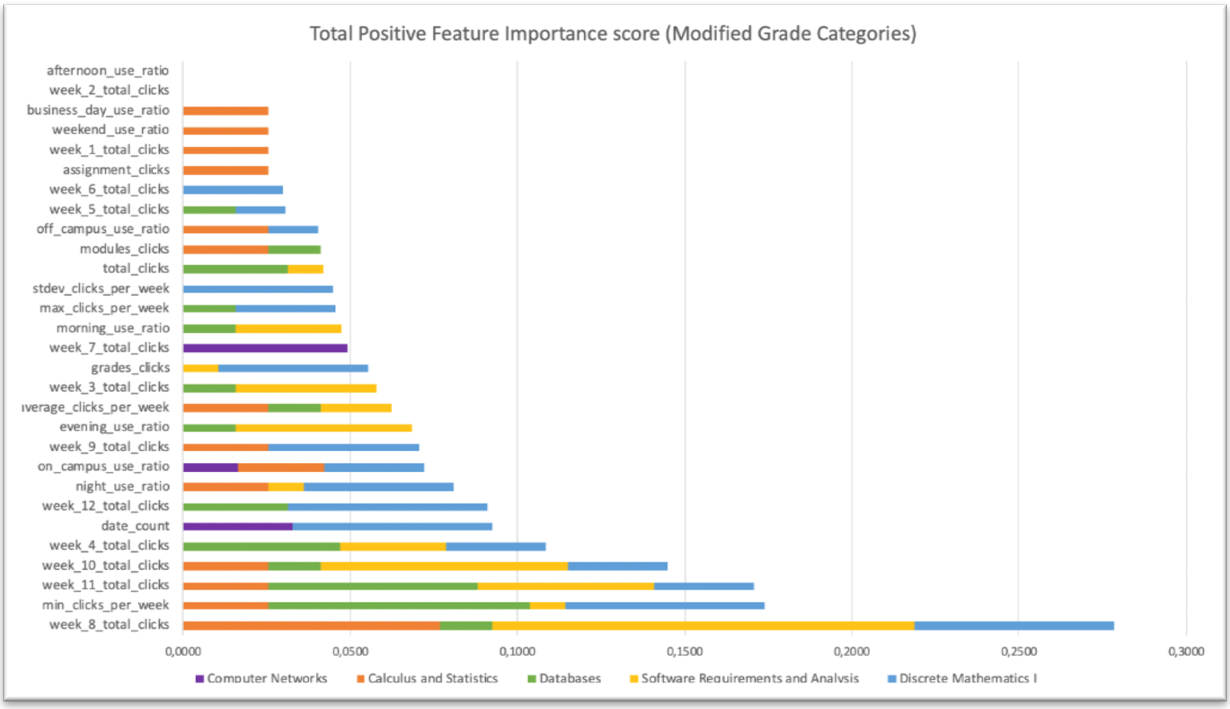

Figure 49. Total Positive Feature Importance Scores (Modified)

When all the variables used throughout the machine learning process have been gathered with their sum of positive scores across all five courses (see Figure 49), there are some interesting observations to be made. There are clearly some variables that have little to no overall effect towards the model as well as certain variables that exhibit consistently higher scores, having a greater positive effect on the model.

The variables representing the latter are *week_8_total_clicks*, *min_clicks_per_week*, *week_11_total_clicks*, *week_10_total_clicks*, *week_4_total clicks* and *date_count*. For further details on each one, see Table 29 above. These particularly positive variables may therefore be acknowledged as being the most defining factors for the Modified Grade Categories version as the values of each one, displayed by students may play the most important role towards their final grade category.

Again, it should be stated that the main purpose of the analysis conducted on the Modified Grade Categories version was to identify possible similarities with corresponding results from the Original Grade Categories version, presented earlier in this chapter.

5 Discussion

The importance of understanding learning patterns and behaviour has never been more vital. The pedagogical structure of higher education has changed drastically. Moving from face-to-face interactions towards increased lecture capture and use of learning management systems that revolutionize the learning structures in higher education. With that trend emerge the possibilities of using massive amounts of educational data for the purpose of mapping different aspects of the way learning management systems affect students learning behaviour (Nguyen, Gardner and Sheridan, 2020). In turn, this massive amount of data can be extensively analysed through advanced methods such as descriptive statistics and machine learning, which has been the focus point of this thesis. Moreover, in order to understand how students in higher education use a learning management system as a part of their learning process we first sent out an online questionnaire in order to map student characteristics and behaviour. The breakdown of non-subject based maturity done in the research of Yani et al. (2019) played an important role in the creation of the questionnaire.

The main results from the questionnaire shows that a large part of participants show high self-esteem and belief in their academic capabilities, which when compared to the research made by Yani et al. (2019) would show that the majority of participants are in the early stages of their education. The results also show that around 46.5% of participants to attend live classes within Reykjavik University, and approximately 50% of participants watch the lecture online rather than attending the classes. Therefore, it seems that students are divided into two categories, those who attend live classes and forego watching lectures online, and those who do not attend classes but watch lectures online instead. This is in line with the first category, described by Edwards and Clinton (2018), as students use lecture capture as a substitute for attending live lectures. However, this is in contrast to the second category that they define where students used lecture captures as a supplement to live classes.

In regard to which patterns and behaviour contribute the most towards a high grade, there is little understanding on the effects of work avoidance within higher education (Neroni, et al., 2018). In light of that the authors of this thesis studied how many clicks a student makes throughout the course and found that students who are not continuously active get lower grades than their counterparts. This is in line with other studies that have concluded that students that apply the work avoidance mindset to their studies, generally get lower grades as a result (Brdar, Rijavec and Loncaric, 2006; Harackiewicz et al., 2008; King and McInerney, 2014).

In the case of higher education, formative assessment is generally conducted through frequent low-stakes assignments, contributing little to no points towards a student's actual final grade score. The main purpose is to give students continuous feedback which can be used to identify strengths and weaknesses. What can be derived from the data is that courses that include frequent formative assignments trigger the above-mentioned continuous work ethic. As opposed to formative assessment, summative assessments are generally high-stakes and often less frequent (Dixson and Worrell, 2016; Gardner, 2010). Courses that rely more heavily on summative assessments, i.e. with a final exam that outline the majority of the final grade, prove less predictable.

5.1 Limitations

As with any research study, there are certain limiting factors that should be acknowledged. The first thing is regarding the online questionnaire as when looking back, it would have been better to include within the background questions regarding what term the participants were currently in, such that a comparison could between Academic Maturity and different years of study, to see if there are any major changes in the mindset over time.

The second thing would have been to have data on lecture capture, which can be acquired by gaining access to click-logs from the third-party application Echo360, used at Reykjavik University for lecture capture. Due to time constraints, the data could not be included. It would have been very interesting to see, how frequently students are using these online lectures throughout their respective course. Effectively contributing towards the discussion of lecture captures versus live lectures and their effect on final grades (in line with Yeung et al., 2016; Mallinson & Baumann, 2015; Edwards & Clinton, 2018).

The third thing is that the click-logs only included clicks made in the web interface of Canvas, excluding all data from the mobile application which is available to all operating systems. Analysing the data from the mobile application would have enabled the possibility of drawing more generalized conclusions regarding the overall usage of Canvas, whereas most students typically use the web interface when handing in assignments and working on actual course work, while the mobile application is mainly used on the fly, for example to check the course schedule, announcements and/or assignment grades.

5.2 Future Work

There are a few things to be mentioned for future analysis. First of all, would be to continue using the questionnaire, and develop it further in order to compare student cohorts. It would also be interesting to look into students in other programs as they might display other characteristics and tendencies. It might also prove valuable to conduct a side-by-side comparison of assignment grades and Canvas usage at certain points in time during the course. Additionally, it would be interesting to pursue qualitative study with students from different grade categories A, B, C, and D. The reason being to explore in further detail what factors the students perceive to be the reason for their final grades, their personal circumstances, their motivation, their personal qualities, etc. versus the patterns and behaviour they actually display through their Canvas usage. Furthermore, it could be exciting to explore the usage of online lectures. More specifically how they are being used by students, at what time periods and how often. Eventually comparing grades and overall use of online lectures.

6 Conclusion

The main finding of the questionnaire can be summarized to the majority of the students showing high self-esteem and belief in their capabilities. For example, being able to adjust to different teaching methods, perform at least as well academically as most other students, taking on difficult tasks that are highly intellectually demanding and complicated, being quick to grasp new concepts, being capable of learning new things quickly, and believing that they will do well in their studies. The result therefore being that most students are motivated and goal oriented. Some students do however indicate that it is difficult to keep up and organize their learning efforts. When it comes to Reykjavik University, the majority of participants believe that their standards and expectations are being met by the school when it comes to higher education, and the teaching methods within the university are enjoyable. Looking further into these matters, the results of the questionnaire show that just under 50% of participants prefer to attend live lectures, while around 50% prefer to watch lectures online.

In addition to the questionnaire, the authors of this thesis reported on the findings from descriptive statistics regarding course data along with the machine learning phase. Regarding what variables are considered to be most predicting in terms of achieving a high grade, the continuous activity of students throughout the course is essential. Overall weekly usage should be consistent and significant drops in usage between weeks should be avoided at all cost. Students who regularly put in the extra work with afternoon usage, often reap the benefits. Additionally, usage in week 4 and week 11 are noticeably important differentiators. The main contributions from the machine learning process are derived from the results of the two versions of course data sets. On one hand the Original Grade Categories version, a non-course specific version with actual distribution between grade categories, the definitions of which were the same for all individual courses. On the other hand, the Modified Grade Categories version, a course specific version with equalized distribution in each grade category (25% in each), tailored specifically to every individual course. Despite this drastic difference in distribution, the resulting accuracy scores provided by the machine learning model for both versions are similar. This underlines the validity and reliability of the 29 variables defined for machine learning purposes and in turn indicates that variables are vastly more important than alterations made to the grade category distribution. Defining the right variables is the most important factor when predicting final grades through machine learning.

Bibliography

Abd-El-Fattah, S. M., and Patrick, R. R. (2011). The relationship among achievement motivation orientations, achievement goals, and academic achievement and interest: A multiple mediation analysis. *Australian Journal of Educational and Developmental Psychology*, 11, 91–110.

Al-Busaidi and Kamla Ali (2012). Learners’ Perspective on Critical Factors to LMS Success in Blended Learning: An Empirical Investigation. *Communications of the Association for Information Systems:* Vol. 30 , Article 2. DOI: 10.17705/1CAIS.03002

Althoff, R. W. (2010). *Eastern Illinois University: The big five personality traits as predictors of academic maturity.* Gathered 18.January 2020 from http://thekeep.eiu.edu/theses/605.

Blumberg, B., Cooper, D.R. and Schindler, P.S. (2005). *Business research methods* (p.124-126). London: McGraw-Hill.

Brdar, I., Rijavec, M., and Loncaric, D. (2006). Goal orientations, coping with school failure and school achievement. *European Journal of Psychology of Education*, *21*, 53–70. https://doi.org/10.1007/BF03173569

Brown, M. (2012). Learning analytics: Moving from concept to practice. *EDUCAUSE Learning Initiative (ELI)*. Gathered 10.May 2020 from: https://library.educause.edu/-/media/files/library/2012/7/elib1203-pdf.pdf

Chen, W. W. (2015). The relations between perceived parenting styles and academic achievement in Hong Kong: The mediating role of students’ goal orientations. *Learning and Individual Differences*, *37*, 48–54. https://doi.org/10.1016/j.lindif.2014.11.021

Chester, A., Buntine, A., Hammond, K. and Atkinson, L. (2011). Podcasting in Education: Student Attitudes, Behaviour and Self-Efficacy. *Educational Technology & Society*, 14 (2), 236–247.

Dixson, Dante D. and Worrell, Frank C. (2016) *Formative and Summative Assessment in the Classroom, Theory Into Practice*, 55:2, 153-159, DOI: 10.1080/00405841.2016.1148989

Dweck, C.S. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040-1048. https://doi.org/10.1037/0003-066X.41.10.1040

Edwards, M.R. and Clinton, M.E. (2018). A study exploring the impact of lecture capture availability and lecture capture usage on student attendance and attainment. *Higher Education – The International Journal of Higher Education Research*, 77, 403-421.

Elliot, A. J., and McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80, 501–519. https://doi.org/10.1037/0022-3514.80.3.501

Gardner, J. (2010). *Developing teacher assessments: An introduction*. In J. Gardner, W. Harlen, L. Hayward, G. Stobart, & M. Montgomery (Eds.), Developing teacher assessment (pp. 1 – 11). New York, NY: Open University Press.

Gibbs, G., and C. Simpson (2004). Does Your Assessment Support Your Students' Learning. *Journal of Teaching and Learning in Higher Education* 1 (1): 1–30.

Harackiewicz, J. M., Durik, A. M., Barron, K. E., Linnenbrink-Garcia, L., and Tauer, J. M. (2008). The role of achievement goals in the development of interest: Reciprocal relations between achievement goals, interest, and performance. *Journal of Educational Psychology,* 100, 105–122. https://doi.org/10.1037/0022-0663.100.1.105

Huang, C. (2012). Discriminant and criterion-related validity of achievement goals in predicting academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104, 48–73. https://doi.org/10.1037/a0026223

Islind, A. S., Norström, L., Vallo Hult, H. and Ramadani Olsson, S. (2020). Socio-technical interplay in a two-sided market: The case of learning platforms. *Digital Transformation and Human Behaviour. Innovation for People and Organizations*.

Jensen, S. A. (2011). In-Class Versus Online Video Lectures: Similar Learning Outcomes, but a Preference for In-Class. *Teaching of Psychology*, 38(4) 298-302.

King, R. B., and McInerney, D. M. (2014). The work avoidance goal construct: Examining its structure, antecedents, and consequences. *Contemporary Educational Psychology*, *39*, 42–58. https://doi.org/10.1016/j.cedpsych.2013.12.002

Lemmens, J.-C. (2010). *Students' readiness for university education.* PhD thesis at the University of Pretoria, Pretoria, South Africa.

Mallinson, D. and Baumann, Z. (2015). Lights, Camera, Learn: Understanding the Role of Lecture Capture in Undergraduate Education. *PS: Political Science & Politics*, 48(3), 478-482.

Neroni, J. Meijs, C., Leontjevas, R., Kirschner, P., and Groot, R. (2018). Goal Orientation and Academic Performance in Adult Distance Education. *The International Review of Research in Open and Distributed Learning*, 19(2). https://doi.org/10.19173/irrodl.v19i2.3440

Nguyen, A., Gardner, L., and Sheridan, D. (2020, January). A Design Methodology for Learning Analytics Information Systems: Informing Learning Analytics Development with Learning Design. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*. http://hdl.handle.net/10125/63753

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91, 328–346. https://doi.org/10.1037/0033-295X.91.3.328

Nordmann, E. and McGeorge, P. (2018, 4.May). Lecture capture in higher education: time to learn from the learners. Gathered 17.January 2020 from https://www.researchgate

.net/publication/324943672_Lecture_capture_in_higher_education_time_to_learn_from_the_learners.pdf.

Ó Callaghan, F.V, Neumann, D.L., Jones, L. and Creed P.A. (2017). The use of lecture recordings in higher education: A review of institutional, student and lecturer issues. *Education and Information Technologies*, 22(1), 399-415.

Peña-Ayala, A., Learning analytics: A glance of evolution, status, and trends according to a proposed taxonomy, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(3), 2018.

Rumsey, Deborah J. (ed.). *How to interpret a correlation coefficient r*. Gathered 18.February 2020 from https://www.dummies.com/education/math/statistics/how-to-interpret-a-correlation-coefficient-r/.

Yani, B., Harding, A. and Engelbrecht, J. (2019). Academic maturity of students in an extended programme in mathematics. *International Journal of Mathematical Education in Science and Technology*. Gathered 18.January 2020 from https://repository.up.ac.za/handle/2263/72631.

Yeung, A., Raju, S. and Sharma, M. D. (2016). Online lecture recordings and lecture attendance: Investigating student preferences in a large first year psychology course. *Journal of Learning Design*, 9(1), 55 – 71.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education,* 45**,** 477–501. https://doi.org/10.1023/A:1023967026413

Appendix I – Quantitative Research Questionnaire

ACADEMIC QUESTIONNAIRE FOR REYKJAVIK UNIVERSITY

The aim of this questionnaire is to establish your preferences and feelings related to your readiness to benefit from university education.
All questions are in relation to the department of computer science at Reykjavik University.
There are no correct or incorrect responses to these statements.
There is no time limit; however, most people take about 10-12 minutes to complete the questionnaire.

Individual background:

1. Age	***Dropdown menu, age from "16 or younger" up to "65 or older" and all ages between***

	Female	**Male**	**Non binary**
2. Gender			

Institutional support:

Statement	**Definently disagree**	**Disagree**	**Neutral**	**Agree**	**Definently Agree**
3. I had enough information about Reykjavik University before enrolling (ísl. skrá mig í nám).					
4. I acquired information about my degree programme before I enrolled at Reykjavik University.					
5. I was informed about the career possibilities for a specific degree programme.					

Educational values:

Statement	**Definently disagree**	**Disagree**	**Neutral**	**Agree**	**Definently Agree**
6. It is important to always be prepared for class.					
7. It is important to have a good university education to be successful in life.					
8. Getting good grades is important to me.					

Goals:

Statement	Definently disagree	Disagree	Neutral	Agree	Definently Agree
9. I'm a very methodical (ísl. skipulagður/skipulögð) person.					
10. I set specific goals before I begin studying for tests/exams.					
11. Good grades provide me with an excellent goal to work towards.					
12. I like to have a routine (ísl. vanagangur / venja) to follow.					
13. I organize my study time to best achieve my goals.					
14. I prefer to be spontaneous (ísl. hvatvís) rather than to set goals when I study for tests/exams.					
15. I usually double check things, just to make sure they are correct.					
16. I know what I want to be doing 10 years from now.					
17. I have clear and reachable/realistic goals for my studies this year.					
18. I have talked about my career goals with someone who has worked in that field.					

Self-efficacy:

Statement	Definently disagree	Disagree	Neutral	Agree	Definently Agree
19. I expect to have a harder time to perform academically (ísl. fræðilega) than most students here.					
20. I can easily adapt to different styles of teaching.					
21. I am as skilled academically (ísl fræðilega) as the best students here.					
22. I enjoy working on complicated, intellectually-demanding (ísl. vitsmunalega krefjandi) problems.					
23. I know what I want and I usually make sure that I get it.					
24. I have the ability to plan my work (study time).					
25. I expect to do very well in my degree.					
26. I am quick to understand new concepts (ísl. hugtök) and ideas.					
27. I can motivate (ísl. hvetja) myself to study when I need to.					
28. I learn things more quickly than most people.					

Academic apathy:

Statement	Definently disagree	Disagree	Neutral	Agree	Definently Agree
29. I tend to study in spurts (ísl. sprettum) rather than at a regular steady pace.					
30. My goal is to get the best grade I can without making a lot of effort on my course work.					
31. I often don't see things through to the end.					
32. I plan my study sessions in advance and almost always follow my study plan.					

Reykjavik University related:

Statement	Definently disagree	Disagree	Neutral	Agree	Definently Agree
33. Reykjavik University is meeting my standards/expectations of higher education.					
34. I like the way of teaching (the methods) used at Reykjavik University.					
35. I do not usually attend lectures (ísl. fyrirlestra) at Reykavik University.					
36. I watch the lectures (ísl . fyrirlestra) online, on Echo360 in Canvas, rather then attend class.					
37. I always attend practical classes (ísl. dæmatímar).					
38. Reaching the Dean's list (isl. forsetalisti) is one of my study goals.					
39. I have been on the deans list (ísl. forsetalisti).					

Social network:

Statement	Definently disagree	Disagree	Neutral	Agree	Definently Agree
40. I prefer to work in groups (arranged by the teacher), rather than work on my own.					
41. I prefer to work in groups (chosen by students), rather than work on my own.					
42. I have a large social network (10 persons or more) within Reykjavik University.					

Please list the names (max 10) of the students who you communicate most with at Reykjavik University (please provide either full name or RU-email).	