Bridging the Knowledge Gap

Identifying Essential Machine Learning Concepts for Effective Progression in Follow-Up Courses

Liselotte Jongejans

Supervisor: Gosia Migut

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 25, 2023

Name of the student: Liselotte Jongejans
Final project course: CSE3000 Research Project
Thesis committee: Gosia Migut, Christoph Lofi

Abstract

This research paper aims to investigate the adequacy of concepts taught during an introductory machine learning course in preparing students for subsequent courses and their professional careers. The study adopts a comprehensive approach, including a literature review, interviews with teaching staff of follow-up courses, and a survey administered to students. The findings of the research indicate a homogeneity in the results, with no significant knowledge gaps identified in the concepts covered by the ML course. However, the study highlights the importance of emphasizing the underlying mathematical foundations more prominently, to enhance understanding and application in real-world scenarios.

1 Introduction

Machine Learning (ML) is one of the most popular and sought-after modern technologies. In 2021, the market size of ML was valued at USD 15.44 billion, with an expected market value of USD 209.91 billion by the year 2029 [7]. Presently, ML empowers an extensive array of applications, ranging from speech recognition systems to search engines, self-driving cars, and many other examples [12]. Numerous applications that were previously designed and coded by humans, now integrate human-written elements with behaviours acquired through data-driven learning. This increasing significance of ML generates challenging questions for education in the field of Computer Science and Engineering (CSE) [12]. While ML has traditionally been confined to a specialized domain within CSE, it is now gaining growing relevance across the core disciplines of CSE. University programs need to explore the potential impact of ML on the perception of fundamental CSE knowledge and skills, and consider how this should influence the design of ML courses, as well as the broader CSE curriculum. Furthermore, exposure to ML concepts and practical applications in CSE programs will contribute to the development of experts capable of addressing challenges associated with this rapidly evolving field [13]. Nevertheless, ML is comprised of complex concepts, and introducing these concepts without adequate attention can potentially dampen students' enthusiasm for this highly significant subject [13]. It is therefore unquestionable that the topic of ML must be effectively and efficiently taught within (undergraduate / Bachelor) CSE programs [10].

However, considering the teaching of ML concepts is often scattered throughout various CSE courses, it can be challenging to determine whether students have a sufficient understanding of fundamental ML concepts prior to taking advanced ML courses and entering the workforce. To address this issue, the objective of this research is to determine the position of ML within CSE programs, and to assess the extent introductory ML courses adequately prepare students for subsequent ML-based courses in both the Bachelor's and Master's programs.

1.1 Research Question

More specifically, the purpose of this research is to assist the current professors and teaching staff of the Bachelor CSE course Machine Learning (CSE2510) at Delft University of Technology (TU Delft) in evaluating whether the concepts covered in the course prepare students for subsequent courses. The study aims to identify if any adjustments are necessary in the course manual to enhance the students' understanding and preparedness for future ML-based courses.

Accordingly, the main question of my research project is:

Which (basic) knowledge concepts of Machine Learning need to be taught during CSE2510, to prepare students for follow-up courses?

In order to address this question, the research will create a roadmap of courses in the CSE programs at TU Delft teaching ML concepts in order to determine ML's current position within the studies, and reflect on concepts that teaching staff of follow-up courses and students suggest to be covered in the curriculum of CSE2510.

1.2 Research Objectives and Contributions

The research makes contributions in multiple areas. Firstly, it sheds light on the effectiveness of CSE2510's curriculum in providing students with a solid foundation in ML concepts. By assessing whether the concepts taught during the course are adequate in preparing students for follow-up courses, this paper provides valuable insights into the strengths and potential gaps in the curriculum. Additionally, as later discussed, the current body of research regarding the essential concepts to be included in an introductory ML course is relatively scarce. Consequently, the findings of this research can serve as valuable insights for educators and curriculum developers, aiding them in enhancing ML course content to align more effectively with the requirements of subsequent courses and industry demands, and thus contribute to the limited research available.

Furthermore, the research study contributes to the broader academic community by expanding the understanding of the relationship between introductory ML courses and subsequent courses or professional paths. By investigating the strengths and weaknesses of introductory courses in bridging the gap between foundational knowledge and advanced concepts, the study provides valuable evidence to inform curriculum design in the field of ML education. Ultimately, the research study's contributions can drive improvements in undergraduate ML education, ensuring that students receive a strong foundation and are better prepared for further studies and careers in this rapidly evolving field.

1.3 Research Paper Overview

Section 2 provides a review of related research and offers an overview of industry standards for teaching ML courses. The third section of this paper outlines the research process, which involves conducting interviews with teaching staff of subsequent courses and surveying students to gather their feedback and experiences following CSE2510 and follow-up courses. Section 4 discusses the ethical aspects of the research. The fifth section reports the results of the research, providing

a summary of the interview and survey results. Section 6 demonstrates recommendations for the course CSE2510, based on the results discussed in the prior section. The seventh section explains improvements to be made and suggestions for future research. Finally, a conclusion is made in section 8.

2 Literature Review

Firstly, it should be noted that due to the relatively nascent nature of research in this field, the literature review, as discussed below, revealed a scarcity of prior research directly addressing the preparation of introductory ML courses for follow-up courses.

The main focus of this literature review considers the Curricula Recommendations provided by the Association for Computing Machinery (ACM) for undergraduate students [1]. ACM's report was consulted to ensure alignment with best practices in the field. By examining the ACM recommendations, the research aims to evaluate how well CSE2510 adheres to the established benchmarks and guidelines for undergraduate education in CSE, providing an objective framework for assessing the course's effectiveness in preparing students for future studies. Incorporating the ACM recommendations into the research methodology also adds credibility and robustness to the evaluation of CSE2510's preparation for follow-up courses. The recommendations put forth by the ACM do not explicitly focus on ML as a standalone curriculum. However, they do recognize the transformative impact of ML on the field of computing. With the maturing of ML, computing has embarked on a new path driven by empirical data analysis and predictive modelling. Recognizing the significance of this shift, the ACM acknowledges that ML stands among the top ten emerging computing trends [1]. This recognition highlights the growing importance and relevance of incorporating ML concepts and methodologies into CSE education, aligning with the evolving needs and demands of the computing industry. As part of this recognition, the ACM has drafted the Data Science Curricula report in 2021 [2], providing a comprehensive exploration of the specific concepts that undergraduate students should be acquainted with in the field of data science. Given the extensive nature of this report, it is not feasible to present all the findings. However, this report provides a summary of the key concepts highlighted by the ACM that are deemed essential for undergraduate students to be acquainted with. ACM's Data Science report emphasizes the criticality of students comprehending ML algorithms and being able to make well-informed decisions regarding the algorithms' applications [2]. The report therefore suggests that undergraduate students should have a solid understanding of various ML approaches, including supervised learning (both classification and regression) and unsupervised learning (clustering and dimensionality reduction). They should be able to comprehend and analyse these methods by comparing and contrasting them. Additionally, students should be aware of common challenges in ML, such as overfitting, and possess knowledge of techniques to mitigate these challenges. It is crucial for students to be familiar with performance metrics in order to select the most appropriate one for evaluating ML algorithms. Further, understanding the methodology for training and testing ML models is deemed essential for students. Lastly, the report highlights the significance of deep learning, emphasizing its status as a new and influential technique for large-scale learning. It emphasizes that students should have a clear understanding of multilayer neural networks (MLP) and be familiar with prominent deep learning architectures, including deep feedforward networks, convolutional neural networks, and recurrent neural networks.

In addition to the recommendations put forth by the ACM, several previous studies have explored the subject matter of this research.

As per the findings of Sun and Gao [14], early ML courses are crucial within undergraduate programs in order to provide foundational support for students. The authors explicitly outline the essential concepts that should be taught in these initial ML courses to ensure students derive maximum benefit in their advanced coursework. They argue that ML, as a field, encompasses various disciplines involving computer science, statistics, and intelligent science, among others. While there may be content overlaps with these related fields, ML possesses its own distinctive characteristics and continually advances with the formulation of new theories and methodologies. Consequently, when teaching ML at the undergraduate level, it is vital to carefully curate and optimize the teaching content. This approach enables students to grasp the fundamental concepts of ML, gain familiarity with prevalent ML models and algorithms, and empower them to utilize key ML techniques for practical problem analysis. Furthermore, it allows students to stay updated with the ongoing developments in the field of ML. The authors therefore suggest to teach the distinction between supervised (statistical classification and regression analysis) and unsupervised learning (clustering and association rules). Besides, the implementation process of ML needs to be taught, including feature extraction from the original sample set, pre-processing of the feature sample set in order to obtain both train and validation sets, training of the model, and validation for evaluation purposes. Additionally, they propose to teach the concept of deep learning, with a focus on MLP and Artificial Neural Network (ANN), explaining the learning strategy and loss function. In terms of classification models, they suggest teaching K-nearest neighbour (k-NN), Decision Tree, Bayes Classifier, Support Vector Machines (SVM), and AdaBoost boosting method. For unsupervised learning the article suggests introducing the Hidden Markov model.

In line with my research regarding the construction of a road map of all courses teaching ML concepts, another article highlights the need to incorporate ML into the undergraduate curriculum to prepare students for the evolving demands of the field [12]. The authors emphasize that professors must grapple with questions surrounding how ML fits into the existing curriculum, which topics should be covered, and how to effectively teach these concepts to students. They mention the importance of striking a balance between theoretical understanding and practical application of ML algorithms. The article underscores the necessity of adapting the computing curriculum to encompass ML, specifically considering its

growing impact on various industries and domains.

Although my research is limited to listing the concepts that are to be taught in an introductory ML course, and does not span the topic of how to teach these concepts, an interesting article to include in the literature review is a paper about the inclusion of practical assignments in an undergraduate ML course. According to literature, the primary difficulty in teaching ML in undergraduate programs lies in the task of helping students understand and appreciate the connections between complex concepts in linear algebra, statistics, and optimization [13]. Although students will have encountered individual concepts within these topics, the challenge lies in compelling them to integrate and apply these concepts together. Hence, to facilitate students in acquiring a deeper understanding of these concepts, it is advised to offer hands-on activities that provide immersive experiences as this will support students' appreciation for the practical use of ML. In a study conducted by the authors of this literature article, the preliminary assessments indicated that integrating these hands-on activities alongside, and synchronized with, regular course content were favoured by 90% of the participating students, as it provides valuable context to the presented content and reinforcement.

As stated earlier, this literature review indicates that the current body of research is limited, highlighting the need for further research. Therefore, this paper aims to fill in this gap by investigating which concepts need to be covered in an introductory ML course, and thus contributing to the existing body of knowledge in a meaningful way. My research will be discussed next.

3 Methodology

The methodology employed in this research involves a combination of techniques, as discussed below. By employing a mixed-method approach, this research aims to provide a comprehensive understanding of the preparation provided by CSE2510 and its impact on students' readiness for subsequent courses in the field.

3.1 Interviews

The primary methodology employed in this study was conducting semi-structured interviews with teaching staff of follow-up courses, in order to gain insights into their experience regarding the preparation for subsequent courses in a ML context. After thorough investigation of the CSE study guide [4], a selection was made of courses teaching ML concepts. Email invitations were dispatched to the course staff of these follow-up courses, requesting their potential participation in an interview and assessing their availability and willingness for the same. Following this, three members of the TU Delft staff, who are responsible for four courses that build upon the foundational ML course, namely courses in Data Mining, Computational Intelligence, Image Processing, and Artificial Intelligence Techniques, were able to schedule interview sessions. By engaging teaching staff from both the Bachelor and Master's levels, a holistic understanding of the curriculum continuity and progression was achieved. These interviews aimed to explore their perspectives on the effectiveness of the CSE2510 in preparing students for their advanced coursework, as well as any modifications or enhancements they deemed necessary. A list of predetermined questions (as mentioned in section 5) was sent to the interviewees prior to the interview, yet the interviews had the flexibility to probe further and subsequent questions were asked based on the participant's responses [9]. This approach allowed for a certain level of structure and consistency across the interviews, while also facilitated in-depth discussions and exploration of the participants' viewpoints regarding CSE2510's content, structure, and its efficacy in preparing students for future learning. The questions focused on topics such as the alignment between the content covered in CSE2510 and the follow-up courses, the strengths and weaknesses observed in students transitioning from the introductory course, and recommendations for improving the preparation of students for subsequent courses. The interviews were audio-recorded to ensure accurate data capture and subsequently transcribed for analysis. Valuable insights, as discussed in a later section, were gathered regarding the curriculum's cohesiveness across the Bachelor and Master's phases [8].

While only one Master's course, namely Artificial Intelligence Techniques, was examined, the course is the first course in the CSE Master's program to incorporate ML concepts. Consequently, it assumes paramount significance within this research as all subsequent Master's courses primarily rely on the foundations established in the aforementioned course.

3.2 Surveys

Comparable to the interviews, to gain insights into the experience of students in terms of the preparation of CSE2510 for subsequent courses, a survey was administered to CSE students who have taken either CSE2510 or at least one follow-up course or both, to gather quantitative data, providing a broader understanding of their experiences and perceptions. The survey sought to capture their feedback on the course's relevance, the skills acquired, and their confidence in pursuing further studies.

The questionnaire was designed to delve into several key areas, and the questions were carefully drafted to be clear, concise, and unbiased, in order to ensure the reliability and validity of the collected responses. Questions focused on how well the concepts, skills, and knowledge acquired in CSE2510 prepared them for subsequent courses, ensuring continuity and smooth progression in their academic journey. Additionally, the survey aimed to identify specific concepts or areas in which students felt they lacked sufficient preparation upon commencing a follow-up course, and reflect on any perceived gaps in their knowledge or understanding. Thus, the survey aimed to assess the perceived relevance of CSE2510 in relation to their follow-up studies. The decision to survey students regarding their experiences taking CSE2510 and its adequacy in preparing them for subsequent courses was driven by the need to gather firsthand insights and perspectives. By directly engaging with the students who had undergone the course, valuable information regarding their learning experiences, challenges encountered, and perceived benefits could be obtained. Understanding the students' perspectives

on the course's effectiveness in preparing them for follow-up courses is crucial for assessing CSE2510's overall efficacy and identifying potential areas for improvement. By leveraging survey responses, this research aims to capture a comprehensive understanding of students' experiences, enabling a thorough evaluation of the course's impact on their preparedness for more advanced topics in the field. This analysis will provide valuable insights for educators and curriculum developers to refine and enhance the course structure, ensuring that future students are adequately equipped for subsequent learning endeavours.

The survey was shared digitally through Qualtrics[11], an online survey platform, which in this instance reached 48 students who participated in the survey. The obtained data was carefully reviewed, cleaned, and analysed (discussed in section 5).

Yet, obtaining an adequate number of respondents for the surveys proved challenging. Despite efforts to reach out to participants, via course specific communication channels, the sample size may have been limited, impacting the generalizability of the findings. How statistically significant, and thus representative, the survey results are, is based on the population and sample size and the level of tolerance for inaccuracy [6]. In this study, the population is all students who have (recently) taken CSE2510 and at least one subsequent course. Accurately ascertaining the population size for this research presents a challenge, considering it is difficult to determine the exact number of individuals enrolled in ML follow-up courses. However, for the purposes of this study, an estimated population size of 1,918 students was used. This amount was determined as follows: according to the latest figures [5], there are currently (June 2023) 1,638 Bachelor CSE students, with 464 being first-years who have not taken any ML related courses yet, so effectively 1,174 students, and 744 Master Computer Science students. Considering a confidence level of 95% and an allowable margin of error of 10%, a minimum sample size of 84 was required [6]. However, it is evident that the research fell short of this target with only 48 respondents, and thus the representativeness of the respondents must be considered when drawing conclusions. Nevertheless, it is crucial to acknowledge the challenge of reaching the entire population, as not all individuals within the target group are easily accessible or reachable. Furthermore, given that not all students are enrolled in advanced ML courses, the figure of 1,918 will be a conservative estimate on the higher end.

4 Responsible Research

With a robust methodology in place, it is equally essential to emphasize responsible research practices to uphold the integrity of the research and safeguard the rights and well-being of the participating respondents.

Firstly, when conducting interviews with teaching staff to gather their opinions on CSE2510, responsible research practices play a vital role in maintaining integrity and ethical standards. To begin with, it is essential that informed consent from the participants is obtained, in which the purpose of the interview is clearly explained. Confidentiality must be ensured, with all information provided by the teaching staff treated with utmost respect and anonymity maintained when reporting their answers. During the interviews, it was very important that the participants were able to express their opinions freely. Ensuring this allows participants to provide honest and unbiased feedback without the fear of repercussions or their opinions being linked to specific results. Therefore, in the reporting of the findings, I present the results without any identifiable information that could potentially be linked back to the specific individuals. Furthermore, it was crucial to avoid leading or biased questions, ensuring that the interview process remains objective and unbiased. Besides, it was necessary to exercise caution when interpreting the participants' views accurately and to avoid misinterpretation.

Secondly, conducting a survey amongst students also necessitates a keen adherence to responsible research practices to ensure the integrity and ethical considerations of the research. As part of this, it was also crucial to obtain informed consent from the students. Besides, respecting the privacy and confidentiality of respondents was also important, and as such no personal information was obtained. The survey was designed in a way to avoid leading or biased questions. Additionally, the survey was distributed amongst students of diverse backgrounds, in order to ensure inclusivity. Furthermore, for conducting the survey, it was crucial to be mindful of the General Data Protection Regulation (GDPR) [15]. As per GDPR guidelines, the survey obtained explicit and informed consent, the survey's invitation stated the purpose of the data collection, and the collected data was stored securely. During the survey design phase, a considerable amount of time was dedicated to finding a suitable survey hosting platform that aligns with GDPR requirements. Despite exceeding the initially allocated time, this effort proved valuable as it was crucial to ensure the survey's compliance with responsible research standards.

5 Results

In this section the findings of the research will be presented, by giving a summary of the gathered insights, highlighting key themes and patterns that emerged from the literature review and the participants' responses.

5.1 Road Map

As stated before, the CSE study guide [4] was explored to determine which courses teach ML concepts. For the Bachelor program, this led to the following road map, shown in the figure below.

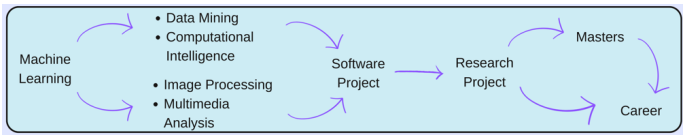

Figure 1: Current road map of courses covering ML concepts in their curriculum.

5.2 Machine Learning Concepts

The following sections present the results derived from the literature, interviews, and survey, focusing on the ML concepts

that need to be included throughout CSE2510 to adequately prepare students for subsequent courses.

Before presenting the results, it is good to have a brief look at an overview of the concepts currently being taught in the course. However, due to space limitations in this report, it is not possible to delve into a comprehensive discussion of all the concepts. Hence, this report will provide only a broad overview of the topics currently addressed in CSE2510 (please also refer to the figures in Appendix A outlining the topics).

The course begins with an introduction to ML during the first week. This is followed by discussions on generative parametric models (Linear Discriminant Analysis (LDA), Nearest Mean, Quadratic Discriminant Analysis (QDA), Naïve Bayes), as well as non-parametric generative models (k-NN, Parzen). Discriminative linear models (Logistic Regression and SVM) are then explored. Subsequently, discriminative non-linear models (Decision Trees and MLP) are covered. Time is also dedicated for responsible ML. Finally, the course concludes with a focus on unsupervised learning (clustering, with especially a focus on K-means, and dimensionality reduction in which Principal Component Analysis (PCA) is explained). Hence, CSE2510 introduces the students to the basic concepts of ML.

Interviews

In this section, I will provide further details on the responses obtained during the interviews. It is important to note that while the interviews were semi-structured and permitted additional inquiries, this report will focus on presenting the findings based on the initial interview questions.

Q1. What specific concepts and skills are most important for students to have mastered before taking your course?

According to the interviews conducted, it was highlighted that students should have a solid grasp of certain concepts and skills before embarking on follow-up courses. Firstly, it is crucial for students to possess a theoretical understanding of ML algorithms (especially decision trees, SVM) and be able to comprehend their functionality and purpose. Furthermore, a thorough comprehension of parameters, including their definition and significance, is deemed essential. Proficiency in working with datasets and understanding the role of data sets in ML processes was also emphasized. Moreover, a strong understanding of PCA was identified as particularly important. In addition, an understanding of loss functions was also highlighted as important by the interviewees. Familiarity with different types of loss functions and their role in ML models is crucial for effectively optimizing the model's performance. A solid grasp of how loss functions measure the discrepancy between predicted and actual values enables students to make informed decisions regarding model selection and parameter tuning. Thus, a comprehensive understanding of loss functions complements the essential knowledge and skills needed to excel in subsequent ML courses. Furthermore, it is necessary to have a solid grasp of the distinctions between the training, testing, and validation sets, as well as a clear understanding of their respective roles. Besides, it would be advantageous if students are well-prepared to comprehend the mathematical notations and formulas employed throughout ML. Notably, it was emphasized that a thorough understanding of Bayes' rule is highly important before advancing to subsequent courses. Moreover, a comprehensive understanding of matrix manipulation is crucial.

These specific concepts and skills were mentioned as prerequisites to ensure a solid foundation for students undertaking subsequent courses in the field of ML.

According to the interviewed participants, it was noted that the practical applications, such as handling unclean datasets and applying ML techniques in real-world scenarios, do not necessarily need to be extensively covered in CSE2510. They acknowledged that these practical aspects are addressed in subsequent courses like Data Mining. These follow-up courses provide students with the necessary skills and knowledge to effectively handle real datasets, including data cleaning, preprocessing, and applying ML algorithms in practical settings. The practical approach taken in subsequent courses complements the theoretical foundation provided by CS2510 and equips students with the necessary skills to tackle real-world data challenges.

Q2. Are there any areas where you find that students typically struggle or have knowledge gaps when they come into your course?

In general, the consensus among the interviewed participants was that students are typically adequately prepared for their subsequent courses. However, based on the interviews, the instructors noted certain areas in which students commonly face challenges when entering their course. Yet, it is challenging, especially in the Masters' courses, to attribute this solely to CSE210 as approximately half of the Master's students originate from other universities.

Notably, students often struggle with the concept of PCA. Despite Linear Algebra and CSE2510 being prerequisites to their courses, students tend to be unfamiliar with fundamental operations such as matrix transposing and multiplication. Additionally, understanding the covariance matrix, a key component in various statistical analyses, is often insufficient. Overall, the instructors observed that students' mathematical skills are not up to par, particularly when it comes to logarithms, probability, and statistics. These areas require additional attention and support to bridge the gap between the prerequisite knowledge and the mathematical proficiency necessary or successful comprehension and application in the course. Moreover, it was observed that students often have encountered Bayes' rule but lack a comprehensive understanding of its underlying principles. Given the significance of Bayes' rule, it is essential for students to strive for a solid comprehension of this topic.

One interviewee specifically highlighted that students frequently encounter difficulties in distinguishing between the train, test, and validation sets. Occasionally, students inquire about the discrepancies, indicating a lack of complete comprehension regarding the ML pipeline.

Q3. Are you familiar with the curriculum and content of the machine learning course? If so, How could the current

machine learning course be adjusted to better prepare students for your course?

All participants who were interviewed exhibited a reasonable familiarity with the curriculum and content of CSE2510. Therefore, the participants were able to provide their insights in terms of adjusting the current ML course.

During the interviews, one participant pointed out that the current structure of CSE2510 is influenced by the background and expertise of the individual teaching staff involved in the course. This approach has its merits, as the instructors bring enthusiasm and passion to their respective areas of expertise. However, it was also highlighted that this method has the potential downside of diverting attention to aspects that may be considered niche sections of machine learning, which might not be essential in an introductory ML course. QDA was mentioned as an example. Consequently, there is a risk of certain topics receiving undue emphasis while other foundational concepts and principles may not receive adequate coverage. Striking a balance between leveraging the instructors' enthusiasm and ensuring comprehensive coverage of essential ML concepts is crucial for optimizing the learning experience in CSE2510.

Additionally, the interviewed individuals expressed their perspectives on specific techniques that should receive more attention in CSE2510. It was suggested that deep learning and random forest, as widely used techniques in the field, should be given greater emphasis within the course. In particular, there was a call for providing more comprehensive coverage of gradient descent, a fundamental optimization algorithm used in various ML models. Additionally, it was highlighted that special attention should be given to ensemble methods, particularly in relation to the issue of overfitting. On the other hand, one interviewee proposed that unsupervised learning might not require significant attention in CSE2510. Considering the limited time available during the course (10 weeks), it was argued that unsupervised learning could be omitted entirely from the introductory course and reserved for more advanced courses. The rationale behind this recommendation was that unsupervised learning is a distinct and complex topic within ML, warranting dedicated focus in subsequent courses rather than being superficially covered in CSE2510. Thus, the suggestion was to primarily concentrate on classifiers in CSE2510 and defer unsupervised learning to follow-up courses.

Most of the interviewees shared a common view that the concepts they teach in their respective courses are typically not extensively covered in CSE2510. However, they regarded this as a positive aspect because it allows their own courses to delve into these concepts in greater detail. By not duplicating the content covered in CSE2510, their courses can provide a more comprehensive and in-depth understanding of these concepts. This approach enables students to explore the intricacies and nuances of the subject matter, equipping them with a deeper knowledge that can be valuable for advanced applications and research in the field.

The interviewed individuals noted that the programming skills of students are generally strong and adequately sufficient for their course, particularly in terms of Python proficiency.

Q4. How do you see the field of machine learning evolving in the coming years, and what skills do you think will be most important for students to have as they enter this field?

The interviewed participants did not provide extensive details on this matter as they expressed the belief that in-depth coverage of real-world applications should not be a primary focus in CSE2510. They suggested that, at most, applications could be mentioned briefly without demonstrating the actual implementation, as this topic is more extensively addressed in subsequent courses. Specifically, the application aspect is often reserved for Master's courses where the emphasis is placed more on practical application rather than theoretical foundations. This approach is justified by the understanding that not all students taking CSE2510 necessarily require a deep understanding of ML applications, as not everyone will pursue further specialization in the ML field. However, it is important for students to comprehend the ML pipeline and have an ethical understanding of what can and cannot be done within the field, albeit without delving into comprehensive application scenarios in the CSE2510 course.

The consensus among the interviewees is that the Bachelor's program should prioritize theoretical understanding. Nevertheless, they acknowledge that numerous students encounter difficulties in this area, as they tend to gravitate towards practical assignments and building ML applications. However, it is essential not to overlook the fact that universities have the responsibility to impart theoretical comprehension and guide students on the significance of this knowledge.

Survey

Out of the total 48 survey responses received, 38 proved to be valuable for the analysis of this study (one response was excluded due to lack of consent, three responses indicated being below 18 years of age, two voluntary pulled out midway the survey, and four mentioned not majoring in CSE leading to their exclusion). Interestingly, it was observed that only 20 students had taken the CSE2510 course, leaving a surprising 18 students who reported not having taken an introductory course at all. Among this latter group, however, seven students had pursued a follow-up course that delved into more advanced concepts in ML. These students were included in the research, as their experience can shed light on the challenges encountered and whether they faced difficulties in grasping the advanced concepts without the foundation provided by the introductory course. Below, I present the insights shared by the survey participants, categorized based those who took CSE2510 and those who did not.

Have Taken CSE2510 - Out of the 20 students who took CSE2510, 13 of them also enrolled in a follow-up course. Interestingly, the majority of these students expressed positive views regarding the preparation provided by CSE2510 for the subsequent courses. Specifically, a significant number of respondents reported that the course adequately prepared them for the follow-up courses, indicating that they felt somewhat or very well equipped to tackle the material covered in subsequent courses. The respondents predominantly emphasized their appreciation for the perceptron and PCA components of the course, along with the simple examples

given when introducing new topics that provided clear and concise explanations, especially considering that topics can easily get very complex. Besides, the vast majority of respondents agreed that CSE2510 adequately covers foundational concepts and knowledge they believe are important for success in future courses, with only one participant disagreeing. Upon inquiring about any perceived gaps or areas that were not sufficiently addressed in the course, the respondents expressed a desire for more hands-on exploration and improved explanations of mathematical concepts, as they felt they never really understood these concepts and had to play catch-up later. According to them, implementation makes the theory more easily understood. In general, the majority of participants expressed a sense of confidence in their ability to apply the knowledge and skills acquired from CSE2510, indicating their belief that they are well-equipped to succeed in future courses or pursue related academic endeavors. Out of the remaining seven students (those who did take CSE2510, but not a follow-up course), two reported being currently employed and working in positions that involve ML. One stated that the introductory ML course somewhat adequately prepared them for their job, whilst the other indicated that the course somewhat unprepared them. Additionally, these respondents were neutral towards how effective the course covered practical aspects of ML, such as real-world applications, hands-on projects, or industry case studies. However, they noted that the course did not sufficiently provide an understanding of common ML algorithms, techniques, and their applications in real-world scenarios.

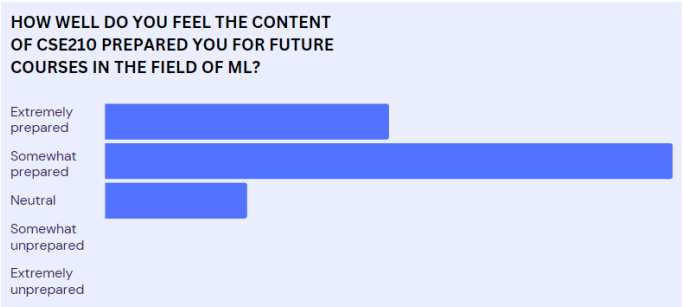

Figure 2: Extent to which students believe that CSE2510 adequately equipped them for subsequent courses.

Have Not Taken CSE2510 - Among the 18 students who did not take CS5210 but enrolled in a follow-up course, there was a prevailing sense of optimism. However, a greater number of respondents indicated feeling neutrally prepared when it came to understanding and actively engaging with the content of the follow-up courses, despite not having taken CSE2510. Several indicated that they found it challenging to grasp certain concepts or topics during the subsequent course due to a lack of background. A number of participants expressed difficulties in comprehending specific concepts or topics during the follow-up course, attributing these challenges to a lack of background knowledge. The main areas identified as causing difficulties were the understanding of mathematical formula notations, matrix transformations, and other linear algebra concepts.

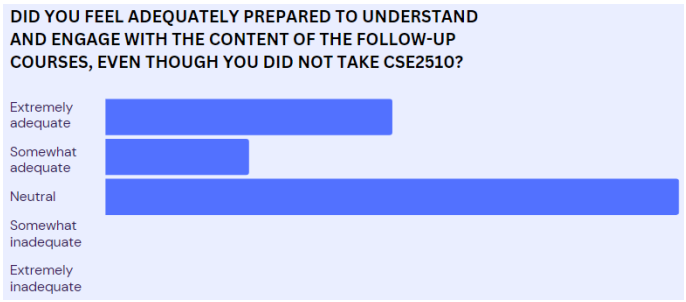

Figure 3: Extent to which students believe that they were prepared for a follow-up course, despite not taking CSE2510.

Prior to delving into the recommendations for CSE2510, this paper will first present an overview of the general results derived from a literature review, interviews, and surveys in the form of a figure.

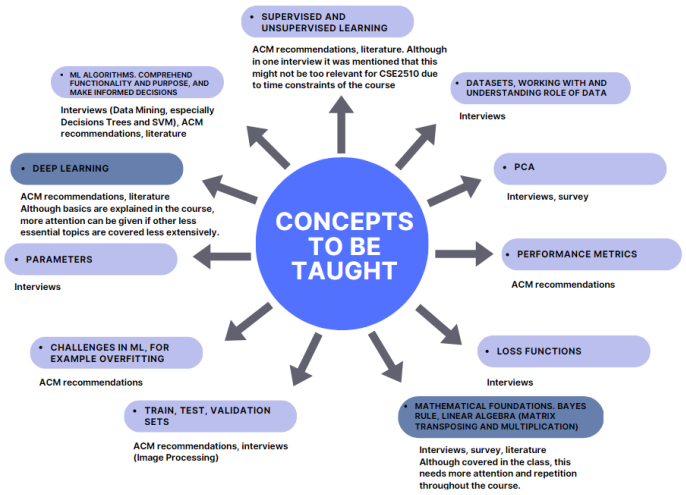

Figure 4: Concepts recommended by literature, interviews, and the survey. Concepts in light blue are recommended and currently covered. Concepts in dark blue are recommended and only slightly covered in the course, but require more attention.

6 Recommendations for CS2510

After conducting a literature review, interviews, and surveys, the findings indicate a notable trend of homogeneity in the responses obtained, with no pronounced contradictory opinions. The participants' views generally align when it comes to the preparation of CSE2510 for follow-up courses, with no glaring gaps identified in the content or structure of the course.

However, suggestions can be made regarding the context in which the course is taught. Firstly, participants express a need for further emphasis on the importance of understanding the mathematical foundations required in ML. As indicated by the student demographic statistics [5], a significant proportion of Bachelor (71.8%) and Master's (55.3%) are international students. This diverse composition suggests that students in the program come from various backgrounds, re-

sulting in varying levels of mathematical foundations, despite having to adhere to admission requirements including prerequisites in terms of mathematics level and grade [3]. It is therefore crucial for students to recognize that ML techniques heavily rely on mathematical concepts, and consequently it becomes vital for them to establish a robust mathematical foundation. Understanding the fundamentals of, for example, linear equations is imperative for grasping the foundations of ML methodologies. The need for students to grasp the underlying mathematical concepts and appreciate the significance of mathematics in ML extends beyond the scope of CSE2510. This need resonates across three foundational courses in the freshman year of the program, namely Linear Algebra (CSE1205), Probability Theory and Statistics (CSE1210), and Calculus (CSE1200). As a result, it is highly recommended that CSE2510 collaborates closely with the Mathematics Department to help facilitate these mathematical courses. By doing so, the collaboration can effectively highlight the importance of mathematics at an early stage of the Bachelor's program. It will ensure that students comprehend the rationale behind taking these math courses and thoroughly understand the concepts taught within them. This collaborative effort aims to establish a strong foundation and foster a deep understanding of mathematics, emphasizing its crucial role in the field of ML. According to findings from both the interviews and surveys, it is evident that students often face difficulties in grasping the fundamental mathematical concepts. Unfortunately, there is limited opportunity in subsequent courses to revisit the basics, such as linear equations and Bayes' theorem. This highlights the significance of emphasizing these concepts as early as possible to ensure students have a solid understanding before progressing further in their studies.

Another notable suggestion that emerged from the feedback gathered during the interviews is to consider limiting the number of lecturers for CSE2510. Participants expressed the opinion that there are currently many lecturers involved, resulting in a lack of cohesion and a fragmented learning experience. To address this issue, it is recommended to synchronize the efforts of the teaching staff, forming a unified and focused approach towards delivering a concise and well-rounded ML course. By streamlining the teaching process, students can benefit from a more cohesive curriculum that delves deeper into the core concepts of ML, rather than covering a broad range of topics in a disjointed manner. This approach can enhance the overall learning experience and ensure a more effective transfer of knowledge to the students.

7 Limitations and Future Work

In this section, I discuss the limitations of the research. Besides, I explore potential opportunities for future work in order to enhance the comprehensiveness and validity of the findings.

7.1 Limitations

The conducted research faced several limitations. Thus, in this section I elaborate on these limitations.

One notable constraint was the tight schedule, hindering the ability to thoroughly explore all subsequent courses and potentially uncover more nuanced insights. In line with this, I encountered difficulty in scheduling interviews with teaching staff from subsequent courses and the Data coordinator of the program for a ML road map interview. The busy schedules of the instructors posed a challenge in arranging interviews, leading to a potential limitation in gathering comprehensive perspectives from all relevant courses.

Additionally, as mentioned previously, the survey's response rate was insufficient to achieve significant statistical significance.

7.2 Future Work

An area of research worth pursuing in the future would be to focus specifically on Master's courses that encompass ML, as the current study was unable to comprehensively cover many such courses. While the present research provides valuable insights into the experiences and perspectives of students in introductory and subsequent ML courses, it is essential to acknowledge that Master's courses, which delve deeper into advanced ML concepts, were underrepresented. Exploring the specific challenges and areas of improvement for preparation of Master's-level ML courses would provide a more holistic understanding of the overall educational journey and enhance the relevance and applicability of future research findings. This is especially important, as not all students follow the Data or Multimedia Variant within CSE, and thus potentially have CSE2510 as their sole ML course.

Another important area for future research would be to conduct a comprehensive investigation into the effectiveness of CSE2510 in preparing students for their careers in ML. Building upon the insights gained from the survey results, numerous participants expressed that CSE2510 constituted their sole ML preparation within their university studies, but found it insufficient for their ML career needs. Therefore, exploring the specific concepts and skills that should be incorporated into the curriculum to adequately equip students for ML careers would be a valuable area of inquiry.

8 Conclusion

In conclusion, this research aimed to assess the adequacy of the course CSE2510 in preparing students for follow-up courses. Through a comprehensive approach that involved a literature review, interviews with teaching staff of follow-up courses, and a survey distributed to students, valuable insights were gained. The results indicate that the current structure of CSE2510 effectively imparts the necessary concepts, leading both teaching staff and students to feel well-prepared for subsequent courses. However, it is worth noting that there is a general recognition of limited mathematical foundations among students. Therefore, it is imperative to allocate more attention to assisting students in developing a solid mathematical foundation. This effort should be undertaken in collaboration with the first-year mathematics courses, considering the time constraints imposed by CSE2510. By addressing these concerns, the overall effectiveness of CSE2510 can be further enhanced, ultimately benefiting students in their academic journey.

A Current Course Overview

Overview provided by the course

Week	Topic	Lecturer
1	Introduction to ML	Jesse Krijthe & David Tax
2	Generative Parametric Models	David Tax
3	Non-parametric generative models	Gosia Migut & David Tax
4	Discriminative Linear Models	Jesse Krijthe
5	Responsible machine learning	Gosia Migut & Mark Theunissen
6	Discriminative Non-Linear Models	Jesse Krijthe
7	Unsupervised learning	Gosia Migut
8-9	Guest Lectures + Q&A	Multiple

Figure 5: High-level overview of concepts currently taught in the curriculum.

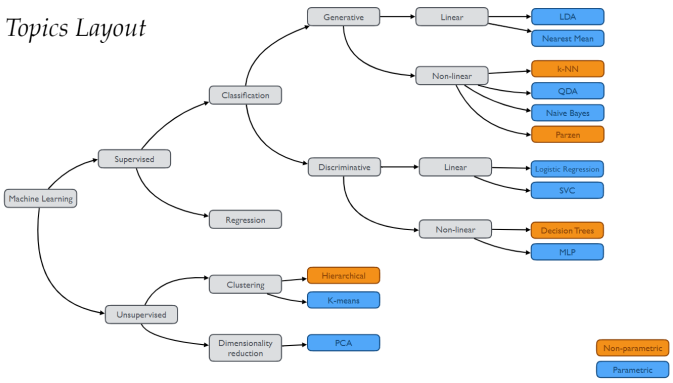

Figure 6: Layout of topics covered in the current curriculum.

B Survey Results and Interview Transcriptions

To view the survey results and to request access to the interview transcriptions, kindly visit the following link: https://github.com/liselottejongejans/BachelorThesis

References

[1] Alison Clear and Allen Parrish. Computing curricula 2020, 2020.

[2] Andrea Danyluk and Paul Leidig. Computing curricula 2020, 2021.

[3] TU Delft. Admission requirements. Accessed on June 22, 2023.

[4] TU Delft. Study guide. Accessed on May 17, 2023.

[5] TU Delft. Figures influx and population bachelor technische informatica, 2023. Accessed on June 14, 2023.

[6] Diego Graglia. How many survey responses do i need to be statistically valid? find your sample size. Accessed on June 18, 2023.

[7] Fortune Business Insights. Machine learning (ml) market size, 2022. Accessed on April 04, 2023.

[8] Emily Jacobs. Why you should transcribe interviews for a better qualitative research. Accessed on June 02, 2023.

[9] Hanna Kallio, Anna-Maija Pietilä, Martin Johnson, and Mari Kangasniemi. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, 72(12):2954–2965, 2016.

[10] Francesca Lazzeri. The importance of teaching machine learning, 2021. Accessed on April 05, 2023.

[11] Qualtrics. Qualtrics, 2023. Accessed on April 15, 2023.

[12] Rebecca Fiebrink R. Benjamin Shapiro and Peter Norvig. How machine learning impacts the undergraduate computing curriculum. pages 1–5, 2018.

[13] Chinmay Sahu, Blaine Ayotte, and Mahesh K. Banavar. Integrating machine learning concepts into undergraduate classes. In *2021 IEEE Frontiers in Education Conference (FIE)*, pages 1–5, 2021.

[14] Wenqian Sun and Xing Gao. The construction of undergraduate machine learning course in the artificial intelligence era. In *2018 13th International Conference on Computer Science Education (ICCSE)*, pages 1–5, 2018.

[15] European Union. General data protection regulation, 2016. Accessed on May 27, 2023.