Discovering the misconceptions that influence learning of Machine Learning

Iulia Angelescu*
Supervisor(s): Gosia Migut, Marcus Specht †
EEMCS, Delft University of Technology, The Netherlands

June 19, 2022

A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

Abstract

The purpose of the study is to investigate and evaluate the most common misunderstandings that are associated with the subject of Machine Learning, and then to make a direct comparison between the findings of the study and the Machine Learning course that is currently being provided by the Delft University of Technology as part of their Bachelor's degree program. In order to do this, the study uses literature analysis, together with interviews conducted with teachers and students of the course in question. The paper analyses the origin of the misconceptions and proposes different methods that can be used in order to eradicate the found misconceptions and, at the same time, methods that could help teachers discover new misconceptions.

1 Introduction

Situated in a fast and rapidly changing world, with high demands and expectations, the Machine Learning domain presents one of the most promising methodologies that is required in the majority of working industries. In the article entailed "A Review of Machine Learning and Deep Learning Applications", the authors analyse multiple fields in which Machine Learning systems can be used including object recognition, object detection, and object processing; semantic analysis and natural language processing; information retrieval; predictions [1]. However, as professor Amy J. Ko points out, despite the considerable need for Machine Learning, the sector lacks engineering expertise, *"Knowledge of how to apply machine learning to products is on high demand but low supply"* [2]. Moreover, in contrast to the shortage of Machine Learning specialists, an analysis conducted within the LinkedIn[1] platform points out that starting from 2017 until 2021, the machine-learning domain was the second and the fourth fastest growing field in the UK, and respectively in US [3].

So, where does the problem begin to express itself? Why is it that there is not enough specialized work power to meet the requirements of this field? In order for a computer scientist to be classified as a Machine Learning specialist, he needs to carry high expertise within the domain, expertise that can only be obtained by setting the ground bases of Machine Learning through various courses and tutorials. For this reason, the study sets to analyse the misconceptions of the domain starting from the courses that are currently being taught, with a more in-depth analysis of TU Delft's Machine Learning course that is being taught as part of the Computer Science and Engineering Bachelor's curriculum. The reason for this is that *"Misconceptions analysis is actually necessary for teachers as well as students to improve teaching and learning process. Moreover, it is not only an objective research method, but also can evaluate the comprehensive academic achievement of students"* [4].

With the purpose of covering all the aspects of the main question of the research, it has been divided in the following sub-questions:

- "Which are the main misconceptions regarding Machine Learning?"
- "Where do these misconceptions originate from? Are these created by students or are they accidentally being thought in the Machine Learning course?"
- "How can professors identify these misconceptions?"
- "Could these misconceptions be put to rest?"

[1] https://www.linkedin.com/

The project's study aims at finding an answer to the question ***"Which are the misconceptions that influence learning of Machine Learning?"***. However, in order to unveil the main misconceptions regarding Machine Learning and how these can be eradicated from the perception of students and future computer scientists, the paper is organized as follows. Initially, the discussion focuses on identifying the most common and disadvantageous misconceptions concerning Machine Learning, with the goal of providing a response to the question ***"Which are the main misconceptions regarding Machine Learning?"***. The focus then moves towards the origin of the misconceptions in the attend of answering the questions ***"Where do these misconceptions originate from? Are these created by students or are they accidentally being thought in the Machine Learning course?"***. Finally, in order to fully address the initial question, the study explores how these assumptions may be clearly identified and whether they can be eradicated from the students' perception, therefore the concluding aspects of the report will assess ***"How can professors identify these misconceptions?"*** and ***"Could these misconceptions be put to rest?"***.

In this study, both common misconceptions and misunderstandings that arise in the subject of Machine Learning, or in relation to this field, are examined. The paper presents the definitions of these terms as they apply to this specific situation so that there is no room for interpretation and such that it may provide a better understanding of the concepts involved.

- ***"Misconception"***, in this paper, refers to particular ideas or information that students associate with a given topic that are incorrect [5]. Misconceptions may arise as a consequence of a lack of explanation about the issue or an inaccurate presentation of the subject.
- ***"Misunderstanding"***, in this particular study, refers to a specific piece of information that students wrongly understand despite having been presented correctly in a course or lecture [6].

The following constitutes the structure of the paper. The introduction part of the paper provides an overview of the particular topic that represents the focus of the subsequent discussion, as well as a brief description of the issue that is being addressed in the following paragraphs. The second section provides an explanation of the methodology that was applied during the process of carrying out the study and provides supporting reasons for the judgments that were used. In the third section of the report, which follows the process of the study, the emphasis is placed on the study that was conducted on literature that is already available. The information that was gathered is presented, along with an explanation of how this information contributes to the outcome of the report. The fourth section presents the empirical study that has been carried out as part of the process. More specifically, this section presents two sets of interviews that focused on gathering information from students as well as teaches of the course of Machine Learning that is included in the bachelor's curriculum at TU Delft. The fifth section of the report is devoted to the discussion of the research and outlines the most significant facets and strategies that may be used in order to identify or dispel misunderstandings that have been revealed in earlier sections of the report. In the sixth section, the author argues that the study was carried out responsibly and gives suitable arguments that support this conclusion. Throughout the seventh part can be found suggestions for further techniques that may be implemented in the future to enhance the research and also there are presented limitations of the current study. The eight and last part of the report offers a summary of the findings of the study and concludes the paper.

2 Methodology

In order for the carried-out study to be successful, it must first go through a series of procedures that will be explained in the up following section. A visualisation of the order of steps that have been followed throughout the development process can be seen in Figure 1.

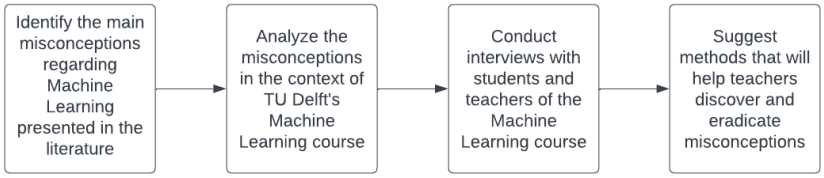

Figure 1: Order of steps followed throughout the research process

In the beginning, the focus goes towards the existing research that has been conducted within the field of Machine Learning. This initial procedure is being selected for two main purposes. First, it is selected for accurately understanding the primary objectives of the Machine Learning programs offered by various bachelor's degree programs and online courses. It is necessary to have a thorough knowledge of these courses in order to evaluate the primary and most significant components of the programs and to analyse any misconceptions that may have arisen as a result of their use. Secondly, this step is being followed in order to gather information about already discovered misconceptions that can be found within the domain. Once the most common misunderstandings have been identified, they can be categorized into a number of subgroups according to the extent to which they impact individuals' propensity to partake in these learning programs, and according to the impact that they have on the effectiveness with which certain algorithms are implemented.

The following step applies the information that was previously presented to the curriculum of the Machine Learning course that is being taught as part of the Bachelor's curriculum of the Technical University of Delft. Because the primary objective of the study is to enhance the content of the named class, it is essential to address the issues that have been discovered in relation to the program itself. This should be done with the intention of identifying and correcting any misunderstandings that may have been accidentally introduced as part of the curriculum.

The third component of this study is composed out of a series of structured and unstructured interviews conducted with professors working in the field of machine learning as well as students from the Technical University of Delft who had previously attended the abovementioned course. The interviews with former students of the course focus on detecting a lack of understanding of concepts taught as part of the curriculum, they aim at discovering topics where more explanations or practice exercises might be required. On the other side, the interviews with the professors focus on analysing students' misconceptions and misunderstandings, and gathering the teachers' perspective on these misconceptions. Also, the interviews aims at discovering other misconceptions that lectures have previously encountered during their teaching experience. These misconceptions could refer to both aspects of the curriculum that may be introduced to the students in an imprecise way, as well as aspects of the curriculum that may be often misunderstood by students.

Finally, after all the data is gathered a conclusion can be constructed. This includes the main aspects of the course where improvement is needed and also mentions specific actions

and procedures that contribute to the eradication of the discovered misconceptions.

3 Literature analysis of Machine Learning misconceptions

This section presents information that has been found throughout already existing literature. The literature classifies the various sorts of misperceptions that occur into two primary categories. The first category refers to misunderstandings regarding the discipline of Machine Learning in general as well as how students and computer scientists actually understand the principles of Machine Learning, while the second category refers to misunderstandings regarding particular models and how they are being applied [7].

3.1 Misconceptions regarding the Machine Learning domain

Based on the results of analysing the first class of existing misconceptions that were presented, it would appear that this class has a significant impact on students' motivation to learn Machine Learning and on their capacity to obtain a comprehensive understanding of the domain.

Author Hendrik Heuer examines in his paper a number of different tutorials that explain Machine Learning and reveals a very common misunderstanding that is seen in the majority of the tutorials that were reviewed. **This misconception is expressed as the idea that one may correctly apply and make use of machine learning algorithms without the requirement of having an in-depth expertise in the field.** This concept frequently occurs in tutorials as *"do not need to understand everything"* or as sustaining the statement the Machine Learning can be used *"very easily without almost any knowledge at all of how it works"* [8].

The assumption that Machine Learning represents the ability of a computer program to learn and improve its performance on its own is another misconception that has been encountered in the conducted literature study and is also part of the same group of misconceptions. However, even if at first sight this statement might seem true, the ability of learning by themselves is not clearly explained in all tutorials and can even appear as the *"learning process of computers based on their experiences without any human assistance"* [7]. In this specific case, many different tutorials and courses fail to emphasize the significant impact that the cost function has on the overall performance of the system, as well as the process of tuning parameters that is carried out by the program's developers. As a result, this misunderstanding has arisen since the participants in the courses are presented with content that is based on insufficient information, and the complete subject is not fully addressed.

The paper entitled *"Artificial Intelligence: A Clarification of Misconceptions, Myths and Desired Status"* brings to light yet another **common misunderstanding concerning this subject by focusing on the distinction between Artificial Intelligence and Machine Learning. They are sometimes considered to be two distinct technologies that attempt to accomplish different objectives, whereas this statement is incorrect for two reasons** [9]. First of all, neither Machine Learning nor Artificial Intelligence are technologies, rather they are methodologies that provide diverse strategies for addressing a variety of problems. Secondly, both of these methodologies are strongly dependent on statistical analysis of the data that is already available, therefore they do not have different

objectives, they just propose different methodologies. Moreover, Machine Learning is even defined as a subfield of Artificial Intelligence that performs nontrivial, smart tasks and it differentiates itself through the processes used in order to construct the systems [10].

3.2 Misconceptions regarding the application of Machine Learning

The second category of misunderstandings is related to the Machine Learning models themselves and their application. It would appear that these misconceptions have a negative influence on the effectiveness and implementation of the models. They have a significant impact on the overall functionality of the systems and have a tendency to reduce performance as a result of inappropriate use of existing models. Here the term "inappropriate use" refers to the practice of employing a model that is unsuitable for a particular problem as well as the practice of employing the appropriate method but adopting it in an inefficient manner. Literature examines the roots of these types of misconceptions and concludes that they are primarily brought on by an imbalance between the practical activities that are included throughout a course and the theoretical information that is covered during that course [2]. This imbalance is criticized in the literature.

While the primary objective of certain courses is to provide students with a theoretical background of the field's fundamental ideas, they frequently fail to provide sufficient knowledge regarding the more advanced facets of the subject. As a consequence of this, the majority of the models that are being used in practical exercises do not present an in-depth explanation of how and why they actually operate [2]. In these circumstances, students do not completely understand the models, therefore, it is highly common for students to apply a method to a problem that would be best handled using another approach.

On the other hand, some courses and tutorials do not provide sufficient amounts of practical activities in order to provide students with a strong comprehension of the given subject. The actual implementation of techniques might be the primary focus of courses, while the phase of data pre-processing, which has a considerable impact on the efficiency of the system, could receive relatively less attention. Limitations in Machine Learning algorithms emerge here as a result of data-drive flaws that have not been explored by the students yet. These flaws are caused by a limited grasp of both the entire process and the significance of each stage of the process [2].

4 Interviews

In order to fully understand the misconceptions that arose in courses of Machine Learning and to directly apply the findings to the course proposed by the Technical University of Delft Bachelor's curriculum, the study also includes an analysis that relies on perception of the students and teachers of the course in cause. Two primary components make up this step of the process. The first part focuses on the perceptions of students that had previously followed the Machine Learning course taught at the Technical University of Delft. The second part examines the perspectives shared by the professors who teach the same class and, more particularly, how they interpret the misconceptions that were brought to their attention by the students.

4.1 Interviews with students

The purpose of this section is to investigate the level of comprehension that students have regarding the course's curriculum. It has been decided that, in order to avoid introducing any sort of bias into the discussions, the interviews will not be based on the misconceptions that have been found in the relevant literature, but instead they went through the primary topics of the course and analyse them one at a time. The conducted interviews were constructed in a semi-structured manner, an approach that permits the interviewer to focus on different aspects that might appear throughout the discussion and personalize each interview based on the knowledge of each student. It is also possible for participants to express themselves verbally in their own unique way due to the versatility and flexibility of the semi-structured interview format [11]. Also, the interviews used the inductive approach suggested in *"The qualitative content analysis process"*, approach used to gather new data in fields that are not completely analyzed in literature [12].

In order to obtain an accurate and up to date perspective from the students, the study used 18 students who had previously participated in either the 2020-2021 or 2021-2022 edition of the course. After carrying out and aliasing the experiment, the following misunderstandings or misconceptions were the ones that came up most frequently in conversations with participants.

At the beginning of the interviews, the participants were asked whether or not they could identify any subject of the lectures which they did not fully understand, or which had not been sufficiently explained. Because this was such a broad question, not a single student was able to pinpoint a particular topic. As a result, the interviews continued with an examination of the topics covered in the syllabus in the form of broad lines, and additional information was gathered regarding the subjects in which students demonstrated a lack of understanding.

While going through each subject, students offered the concept of **cross validation** as an example of a topic that was not fully understood and yet gave some instances of misunderstanding. More specifically, most of the students were not able to recall the difference between test and validation set. Even though the distinction has been made clear during the lectures, it looks as though insufficient emphasis has been placed on this concept; as a result, the student has been prompted to indicate a lack of comprehension.

Another misconception regrades the topic of **Principal Component Analysis (PCA)**, when asked to describe the concept of PCA and how the algorithm is applied, 11 out of 18 students defined it as the method that reduces the dimension of a data set, by discarding the so called "unimportant" features of the data and only considering the features that are most relevant and from which most information can be withdrawn. This is a misconception because their statement defines the Feature Selection algorithm, while PCA is defined as the technique that maps the data space to a smaller dimensional space, without eliminating any feature, but rather compressing them [13]. This misconception seems to have appeared due to a wrong explanation of the concept that appears in the course of Machine Learning that is currently being taught. More specifically, the algorithm is described as follows: "Choose only the good distinctive features" [14].

It seems that the subject of **Gradient Descent and Stochastic Gradient Descent (SGD)**, which was covered in the Machine Learning course, was the one that students had the most trouble mastering. Here 18 out of the 18 interviewed students made it clear that they believe this subject was not given sufficient consideration and that it was only passingly discussed throughout the many classes, without a comprehensive explanation of the notion being presented. Out of the 18 participants, 6 of them mentioned that the lectures that

covered this topic where *"hard to follow as they only relied on the mathematical aspect of the method"*. Also, several students indicated that they did not consider this concept to be part of the major objectives of the course, and as a result, they did not consider it necessary to have a better comprehension of the topic.

4.2 Interviews with teachers

The second round of interviews that were carried out had the objective of discovering professors' points of view with regard to the misunderstandings that were brought up by the students, as well as any other types of misunderstandings that the lecturers had come across over the course of their careers. The participants consisted of three instructors who are currently delivering the Machine Learning course that is a required component of the bachelor's degree program at TU Delft. The interviews followed a structured form and applied one of the approaches suggested in *"The qualitative content analysis process"*, more specifically the deductive approach. The deductive approach is used to test or to get more opinions on a topic that is already studied [12]. Since the teachers had to reflect on earlier findings of the study and to improve different exercise proposals, the interviews were indeed following a deductive approach.

The participants were initially questioned about the misunderstandings brought up by the students. Also, after processing the interviews that were conducted with the students, some models of practical exercises were constructed. The teachers were required to provide feedback on those models and argue whether or not they consider the exercise to be helpful in improving the quality of study and understanding, as well as how they would change the exercises if that was the case. The concrete examples of exercises are presented in Section 5.1.

It was also requested that the participants think about the students' claim that Gradient Descent and Stochastic Gradient Descent (SGD) are not emphasised sufficiently throughout the course. While they agreed with this statement, particularly the part regarding SGD, they pointed out that the teaching time represents a barrier and that not all subjects can be taught in depth. Because the field of machine learning is so vast and incorporates so many various concepts, there is a trade-off that has to be made when it comes to deciding which topics should be emphasised more than others and which topics aren't necessary for achieving the main objectives of the course.

After that, the teachers were further asked whether, throughout the course of their careers, they have come across any recurring misconceptions in their students' thinking. Here two out of the three professors pointed out one of the misconceptions that was previously identified through literature analysis. More specifically the misconception that Machine Learning learns by itself and that as long as models are provided with enough data, the computer can solve the problem on its own. Unfortunately, as previously mentioned and sustained by literature, this misconception is a widely spread one and some courses try to tackle this problem, while others involuntary promote this misconception.TU Delft's course of Machine Learning tries to eradicate this misconception by presenting the importance of having an expert construct and enhance the model. Throughout the lectures and practical exercises teachers present the importance of having the Machine Learning models evaluated by a computer scientist and also the importance of using parameter tuning in order to obtain a model that has a high accuracy.

5 Discussion

This section of the study takes a closer look at the misunderstandings that have been presented in the above sections and proposes ways that may later be undertaken as part of courses with the aim of helping teachers to identify misconceptions that are present in the perception of the students. Students can gain a better knowledge of the concepts addressed in the lectures by using the techniques that are provided.

5.1 Proposed approaches for eradicating misconceptions

The presented approaches centre on practical methods of dispelling the many misconceptions that are currently in circulation. Because it is more effective in delivering a clear and more steady overview of the issue, the choice to develop mostly practical ways of enhancing education has been taken. The significance of practical exercises in the development of educational content was investigated in a research that was carried out in 2006 [15]. This study conducts an experiment to evaluate whether or not students believe that solving practical issues contributes to their overall knowledge and to what extent they value the opportunity of using what they have learned in the classroom. The vast majority of the students reported that this kind of instruction has a big impact in their level of comprehension, and that it also gives them the capacity to use the strategies that they have learned to solve issues that occur in the real world [15].

This section provides an exercise that might be of use in bringing clarity to any questions or concerns that may have been raised in relation to the subject of cross validation. An exercise in which students are required to utilize both the test set and the validation set has been suggested as a solution to the widespread misunderstanding that surrounds the distinction between the two types of data sets. The framework of the activity is broken down into the following steps: The students are provided with a collection of data that has already been labelled in accordance with the model, as well as an implemented model of Machine Learning that is used to classify certain data using the k-Nearest Neighbour algorithm. The students will be required to use cross validation to determine the most appropriate value for k, and they will also be responsible for dividing the collection into only two sets, namely the training set and the validation set. After students have determined the value of k that is optimal for the scenario, the following sub question of the exercise will provide another set, a test set, that will be used to verify the accuracy of the model. During this exercise, students will be tasked with analysing the performance of the model on the new set and arguing whether or not this performance is as good as their earlier estimation. It is planned to create the sets in such a way that the testing set will not yield an accuracy that is substantially equivalent to the earlier estimations. The students will be required to think about the differences in performance; as a result, they will comprehend the significance of the three sets, and they will be able to tell the validation set apart from the test set. Teachers will be able to determine whether or not a student has a complete comprehension of the information that is being covered by analysing the students' reflections. The exercise has been talked about as part of the interviews that have been carried out with the teachers, and they have mentioned that an exercise that points out the downsides of not using the three sets will be much more effective at eradicating the misconception than the construction of a problem that just applies the algorithm. This has been mentioned as part of the fact that throughout the study it have been taken into consideration to construct an exercise that only applies the algorithm and does not highlights the downsides of not using it properly.

Another exercise that is being proposed is one that refers to the topic on Principal Component Analysis (PCA), another subject that presented misconceptions in regard to its application. The paper makes a suggestion for a practical exercise that will be carried out in the following manner. The students are given a collection of data that is particularly rich in dimensionality since it has a great number of features. They are given two different strategies that they have to apply in order to bring the dimensionality of the data set down to a certain number of characteristics. For the first method they need to apply the "Features selection" algorithm, in order to preserve only the "most relevant" features of the data set, while for the second method they must use the PCA technique in order to minimize the dimensionality of the data. After obtaining the two sets, they will need to train an already constructed model using the two sets of data and test the model's performance on a set that is unseen by the model. They will need to examine both versions side by side in order to determine which of the two provides a better performance. Following the completion of the designated exercise, they will be able to recognize the difference between the two methodologies and reflect on their usability and advantages and disadvantages.

5.2 Proposed approaches for discovering misconceptions

Literature suggests that ideas are built one on top of the other and it is essential for a student to have a solid understanding of the fundamentals of a subject in order to be able to comprehend more complex aspects of the same field [16]. Because of this, it is essential for instructors to identify when students have misunderstandings and to provide assistance for them students in order to facilitate a better education.

The following approach can provide instructors the ability to identify the subject areas in which students have misunderstandings and the contexts in which those misunderstandings may have originated. According to one piece of research study [17], giving a student the responsibility of formulating a question relating to a subject can help shed light on a number of commonly held misconceptions. Based on the mentioned study, the following approach has been suggested as a viable option: throughout the course of Machine Learning, at the end of each lecture, students can be asked to write a question related to the material that was just covered, as well as the answer to the question. This exercise would not only benefit the discovering of misconceptions but also the students' attention and interest will be increased throughout the classes [18], their critical thinking will be stimulated, and they will be encouraged to reflect on the newly received information [17]. After carrying out this activity in the classroom, teachers will be able to evaluate the students' responses and determine which concepts may require further elaboration.

6 Responsible Research

In order for the search to situate itself within the standards of acceptable methods of conduct, its development carried out while closely following the five principles of responsible research presented in the *"Netherlands Code of Conduct for Research Integrity"* [19], document released in 2018.

First of all, the study conforms with the concept of **honesty** since it presents the exact information that has been found in literature and data that have been gathered throughout the experimental studies, without changing the outcome by manipulating the evidence. The findings are entirely supported by truthful material, and the discussion is endorsed by solid arguments and real-world examples.

Secondly, the concept of **scrupulousness** is exhibited within the research as a result of the research's exploitation of only scientific and scholarly techniques for the purpose of data collection and analysis as required by the study. Also, the conducted interviews follow academic approaches, that appear in multiple literature studies and they are properly motivated. To add up to this, the results are provided in an objective manner, which means that only the material that is held by reliable sources is taken into consideration, and the author does not provide any personal bias or perspective to the discussion in any way.

Besides that, the principle of **transparency** also applies to this study, as the entire process that has been followed in order to complete the research is presented within the paper. No other intermediate steps have been taken besides those that have already been mentioned. This demonstrates that the research has been conducted in a completely open and honest manner. Additionally, the collected data is maintained in an ethical manner, without any information that might be considered compromising, and all of the participants in this research have been given explicit information on the scope of the research as well as how the data will be handled and exploited in the future. All the necessary information had been communicated to the participants of the interviews though consent forms which they had to agree with in order to take part in the study.

Independence represents the fourth element that has been taken into account while conducting the study. As it was previously mentioned, all of the necessary materials that have been gathered came from certified sources such as academic papers and scientific publications, and only data sources that are relevant to the study have been take into consideration. Moreover, they were presented in a manner that is objective and does not include any indications of the author's own biases.

Last but not least, the principle of **responsibility** was also followed as the study is relevant to the problem that it is trying to solve, and it contributes to the demand that exists within society to enhance the teaching methodologies that are currently used for Machine Learning courses.

7 Limitations and Future Work

There were several constraints that prevented the research from being carried out to its best potential because of the limited amount of time that could be spent on it, namely ten weeks. When it came to conducting interviews with the students, time indeed represented a constrain. It is difficult to generalise the findings of the interview based on the fact that there were only 15 people who participated while each year there are around 400 students enrolled in the Machine Learning course. Therefore, the actual study can be further improved by conducting interviews with additional students in order to provide a larger data sample and to be able to generate more accurate generalisations of the misconceptions that arise throughout the duration of the Machine Learning course that is taught at TU Delft.

In addition, because of the restricted amount of time available for the study, it was not possible to analyse previous assignments and projects that students of the course had finished during previous editions. This also represents work that can be done in the future in order to improve the study, thus it should not be ignored.

Another potential barrier might be imposed by the timing of when the study has been carried out. Due to the fact that the research has started in April 2022 and the most recent edition of the Machine Learning course was completed in November 2021, it is possible that the information that was gathered from the students does not actually reflect their level of comprehension of the concepts at the time when the course was being taught.

8 Conclusions

After gathering information from literature and applying it to the course of Machine Learning that is being taught throughout the program presented by the Technical University of Delft as part of their Bachelor's in Computer Science and Engineering degree, it seems that most of the misconceptions and misunderstandings that students present in this field are caused by the lack of practice that is being offered by multiple courses, or the imbalanced between theory and practice. Former students of the course in question pointed out three primary topics which present misunderstandings, more specifically the topics of Principal Component Analysis, Cross Validation, and Gradient Descent and Stochastic Gradient Descent. While the for the first two topics, the paper presents methods that can be used in order to eradicate the related misconceptions, for the topic of Gradient Descent and Stochastic Gradient Descent, as students expressed in their interviews, it seams that the subject is not well addressed and they do not consider it one of the main objective of the course.

As above mentioned, it is important for teachers to discover the main misconceptions related to the course in order to be able to enhance the level of education. As the paper has its limitations and the conducted study represents only a part of the research that can be carried in this domain, Section 5.2 together with Section 7 present different approaches that can be used in order to enhance the actual study and reach its highest potential.

References

[1] Pramila P. Shinde and Seema Shah. A review of machine learning and deep learning applications. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pages 1–6, 2018.

[2] Amy J. Ko. We need to learn how to teach machine learning, August 2017. https://medium.com/bits-and-behavior/we-need-to-learn-how-to-teach-machine-learning-acc78bac3ff8.

[3] Beatrice Nolan Martin Coulter. There's a talent shortage of machine-learning engineers right now as demand for ai specialists booms, March 2022. https://www.businessinsider.com/there-is-a-machine-learning-talent-shortage-2022-2?international=truer=USIR=T.

[4] Tian-Wei Sheu, Phuoc-Hai Nguyen, Phung-Tuyen Nguyen, Duc-Hieu Pham, Ching-Pin Tsai, and Masatake Nagai. A matlab toolbox for misconceptions analysis based on sp chart, grey relational analysis and roc. *Transactions on Machine Learning and Artificial Intelligence*, 2(2):72–85, 2014.

[5] Daniel Gil-Perez and Jaime Carrascosa. What to do about science "misconceptions". *Science Education*, 74(5):531–540, 1990.

[6] Edda Weigand. Misunderstanding: The standard case. *Journal of pragmatics*, 31(6):763–785, 1999.

[7] Erik Marx, Thiemo Leonhardt, David Baberowski, and Nadine Bergner. Using matchboxes to teach the basics of machine learning: an analysis of (possible) misconceptions. 170:25–29, 08–13 Sep 2022.

[8] Hendrik Heuer, Juliane Jarke, and Andreas Breiter. Machine learning in tutorials - universal applicability, underinformed application, and other misconceptions. *Big Data & Society*, 8(1):20539517211017593, 2021.

[9] Frank Emmert-Streib, Olli Yli-Harja, and Matthias Dehmer. Artificial intelligence: A clarification of misconceptions, myths and desired status. *Frontiers in artificial intelligence*, 3:524339, 2020.

[10] Tom M Mitchell and Tom M Mitchell. *Machine learning*, volume 1. McGraw-hill New York, 1997.

[11] 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.

[12] Satu Elo and Helvi Kyngäs. The qualitative content analysis process. *Journal of advanced nursing*, 62(1):107–115, 2008.

[13] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

[14] Amira Elnouty. Unsupervised learning, 2022.

[15] Nina Katajavuori, Sari Lindblom-Ylänne, and Jouni Hirvonen. The significance of practical training in linking theoretical studies with practice. *Higher Education*, 51(3):439–464, 2006.

[16] Mark Francek. A compilation and review of over 500 geoscience misconceptions. *International Journal of Science Education*, 35(1):31–64, 2013.

[17] Marleen Olde Bekkink, ART Donders, Jan G Kooloos, Rob MW de Waal, and Dirk J Ruiter. Uncovering students' misconceptions by assessment of their written questions. *BMC medical education*, 16(1):1–7, 2016.

[18] Marleen Olde Bekkink, ART Donders, Jan G Kooloos, Rob MW de Waal, and Dirk J Ruiter. Challenging students to formulate written questions: a randomized controlled trial to assess learning effects. *BMC medical education*, 15(1):1–6, 2015.

[19] K Algra, L Bouter, A Hol, J van Kreveld, D Andriessen, C Bijleveld, R D'Alessandro, J Dankelman, and P Werkhoven. Netherlands code of conduct for research integrity 2018, 2018.