Learning Machine Learning:
<A Comparative Study of Aerospace Engineering and Computer Science Students>

Junwon Yoon

Supervisor(s): Gosia Migut, Ilinca Rențea

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
January 25, 2025

Name of the student: Junwon Yoon
Final project course: CSE3000 Research Project
Thesis committee: Gosia Migut, Ilinca Rențea, Examiner

Abstract

Machine learning (ML) is increasingly integrated across diverse academic disciplines, necessitating effective teaching strategies tailored to varied student backgrounds. This study investigates the influence of prior mathematical knowledge on the learning outcomes of ML topics among Computer Science (CS) and Aerospace Engineering (AE) students. Employing a mixed-methods approach, the research involved initial mathematical assessments, interactive tutorials on key ML topics (Bayes Rule, Perceptrons, ML Pipelines), and subsequent evaluations of ML comprehension.

The results reveal significant differences in performance between the two groups. CS students, with their integrated programming and mathematical preparation, consistently outperformed AE students, who demonstrated variability despite their strong quantitative foundations. Probability and linear algebra emerged as key contributors to ML learning, showing stronger correlations with outcomes than calculus. Qualitative analysis highlighted the need for tailored instructional approaches: AE students preferred application-driven and interactive learning, while CS students valued structured and technically detailed resources.

These findings underscore the importance of interdisciplinary teaching strategies that bridge gaps in programming and mathematical competencies. The study's insights have implications for designing inclusive ML curricula, emphasizing real-world applications, adaptive learning technologies, and frameworks to support diverse learner needs. Future research should explore broader ML topics, larger participant groups, and long-term skill retention to further enhance ML education across disciplines.

1 Introduction

Machine learning (ML) is being acknowledged as an important tool in different fields.[1] As a result, there is a growing need for efficient machine learning instruction. Traditional courses are usually taught in computer science (CS) degrees. Meanwhile, the teaching of machine learning (ML) has gained popularity across various disciplines, spanning both STEM (science, technology, engineering, and mathematics) fields and non-STEM areas. However, teaching machine learning to students with different academic backgrounds presents a number of difficulties, especially because of the difference in mathematical preparation and expertise. [2]

As ML continues to make an impact in fields such as business, healthcare, and beyond, professionals in these sectors need a solid understanding of ML and its applications. We hope that better knowing how students from various academic backgrounds learn ML will help us build more focused teaching approaches. This study builds upon previous work, such as Amy Ko's (2017)[3], that highlighted the need for better ML teaching methodologies, and aims to provide practical ideas for more effective ML instruction across different fields.

This study aims to explore how students' academic backgrounds impact their understanding of machine learning concepts and skills. Specifically, our objective is to compare the learning outcomes of students from aerospace engineering faculty who may have different strengths with those of computer science majors, since they have a strong foundation in mathematics but limited exposure to machine learning concepts. For instance, TU Delft CS students are introduced to Calculus, Probability and Statistics, and Linear Algebra in their first year, alongside programming courses. This integrated foundation enables them to grasp ML concepts more easily and to apply and test these concepts through coding.[4] In contrast, TU Delft AE students also study Calculus, Probability and Statistics, and Linear Algebra, in addition to Physics and Dynamics courses, which strengthens their familiarity with STEM-related material. However, AE students typically lack experience with coding in the context of ML, and this distinction is the key focus of our investigation.

Aerospace engineering provides a particularly compelling context for this research due to its reliance on advanced mathematics and physics, alongside its growing adoption of ML in real-world applications. Fields like physics and economics also involve quantitative reasoning, but aerospace engineering uniquely integrates these skills into practical, high-stakes problem-solving scenarios such as predictive maintenance, aerodynamic optimization, and trajectory planning. These applications require engineers to bridge the gap between theoretical ML concepts and highly specialized domain-specific knowledge. By focusing on aerospace engineering, this study aims to shed light on how students with strong quantitative backgrounds but limited exposure to programming or data science engage with ML concepts.[5]

Furthermore, recent research, such as the work by Sundberg et al. (2023), [6] has explored the use of no-code AI tools to lower the barrier for non-CS students learning ML concepts. Such approaches highlight the importance of tailoring teaching strategies to students' specific needs and strengths.

We will first use a math assessment test to assess students' mathematical knowledge to gain a better understanding of these variations. Following that, we will offer machine learning tutorials on different topics, including classification error and Bayes' theorem. Finally, to determine the connection between prior math knowledge and ML comprehension, students will be required to take an ML test covering the topics covered in the tutorials.

The research question of this paper is as follows:

Research Question

The main research question for this project is the following.

> *"How does prior mathematical knowledge influence the learning of different Machine Learning topics among Computer Science and Aerospace Engineering students?"*

To answer this question, the study will explore several sub-questions:

- How do mathematical foundations in calculus, linear algebra, and probability correlate with the understanding of specific ML topics (e.g., Bayes' Theorem, Perceptrons, ML Pipelines)?
- What differences exist in the learning outcomes of ML topics between Computer Science and Aerospace Engineering students?
- What instructional adjustments could enhance the teaching of ML to students with strong mathematical backgrounds but limited programming experience?

This paper is organized as follows: Section 2 reviews existing work on teaching ML to diverse academic audiences. Section 3 describes the methodology used, including the assessment tests and tutorial design. Section 4 presents the results of the comparative analysis between aerospace engineering and computer science students. Section 5 discusses the implications of the findings, and Section 6 concludes with recommendations for future research and instructional strategies.

2 Related Work

The rapid growth of machine learning (ML) applications has underscored the necessity of effective ML education across diverse academic disciplines. While ML is traditionally taught within computer science (CS) programs, its wide-ranging relevance has prompted efforts to integrate it into curricula for non-CS majors. Despite these efforts, structured methodologies for teaching ML concepts to students with diverse academic backgrounds—particularly across STEM and non-STEM domains—remain underexplored.

Existing research on ML education offers insight into teaching techniques, but there is a lack of comparative analysis of which strategies work best in diverse educational contexts. Amy J. Ko's statement "We need to learn how to teach machine learning" highlights these ongoing challenges and the need for better, more structured approaches(2017)[3]. This project aims to address this gap by investigating effective ways to teach ML to students with various educational backgrounds, focusing particularly on the role of prior math knowledge in learning ML.

For example, Shannon Wongvibulsin (2019) advocate for inclusive educational models that introduce ML without prerequisites, enabling students from diverse fields to gain foundational ML skills.[7] Their work suggests that designing courses accessible to non-CS majors can expand the reach of ML education and address existing disparities.

Similarly, Tenorio and Romeike (2023) proposed a framework of AI competencies tailored for non-CS students in interdisciplinary settings.[8] By identifying core competencies, they offer a structured way to design ML curricula that align with the specific needs of students from fields like aerospace engineering.

Moreover, integrating domain-specific applications into ML education has been shown to improve engagement and comprehension. For example, Kasinidou et al. (2023) demonstrated how tailoring assignments to students' fields of study can help bridge the gap between theoretical ML concepts and their practical applications.[9]

The integration of machine learning (ML) into educational research has highlighted its potential to enhance learning outcomes across diverse disciplines. Recent studies, such as Ayanwale et al. (2024)[10], emphasize the importance of tailoring ML methods to specific educational contexts, fostering personalized and data-driven approaches. This aligns with our study's focus on addressing the unique challenges faced by students from varied academic backgrounds in learning ML concepts.

In summary, the existing body of work underscores the growing importance of developing tailored ML education strategies for students from diverse academic backgrounds. While progress has been made in creating inclusive and domain-specific learning models, significant gaps remain in understanding how prior knowledge, particularly in mathematics, impacts ML learning outcomes. This chapter has highlighted key studies that inform the foundation of this research, emphasizing the need for comparative analyses and structured methodologies. By focusing on aerospace engineering and leveraging insights from prior work, this study aims to contribute to the ongoing discourse on effective and inclusive ML education.

3 Responsible Research

This study adhered to the highest standards of ethical research, ensuring participant privacy, data security, and transparency throughout all stages of the research process. The following measures were implemented to ensure responsible and ethical research practices:

3.1 Human Research Ethics Checklist

The study followed institutional guidelines for ethical research, as outlined in the Human Research Ethics Committee (HREC) checklist. This process involved identifying potential ethical risks, such as participant discomfort or data misuse, and implementing strategies to mitigate these risks. The research design and methodology were reviewed and approved by the relevant ethics board to ensure compliance with institutional and professional standards.

3.2 Informed Consent

All participants were provided with a detailed informed consent form outlining the study's objectives, procedures, potential risks, and benefits. Participants were informed of their right to withdraw at any time without consequences. To maintain transparency, the consent form included information about the anonymity of responses, how data would be stored and used, and contact information for the researchers and ethics committee. Only those who provided explicit consent were included in the study.

3.3 Data Management Plan

A data management plan (DMP) was developed to outline how data would be collected, stored, secured, and used during this project. The plan details the types of data being saved, their purpose, storage duration, and measures taken to ensure confidentiality and quality.

3.4 Use of Generative AI Tool

This study employed generative AI tools to assist in structuring sentences, refining language, and enhancing clarity in written communication. These tools were used solely for language improvement and did not influence the study's methodology, data analysis, or interpretation of results. The final manuscript was reviewed and validated by the researchers to ensure accuracy and integrity.

4 Methodology

This study employs a mixed-methods approach to explore the relationship between prior mathematical knowledge and the learning outcomes of machine learning (ML) concepts among students from the Computer Science (CS) and Aerospace Engineering (AE) faculties. The methodology is structured into four interconnected phases: participant recruitment, tutorial delivery, knowledge assessment, and data analysis. Each phase is carefully designed to gather comprehensive information about the learning process and its influencing factors.

In addition,this section reflects collaborative efforts with two colleagues conducting parallel research on industrial design and mathematics education. The design and distribution of the tests and surveys were shared responsibilities between the researcher and the colleagues. The specific tasks undertaken by the researcher are explicitly detailed at the end of this section.

4.1 Participant recruitment

To ensure a balanced comparison, the study recruited 20 students: 10 from the CS faculty and 10 from the AE faculty. Participants were selected through purpose sampling to include students with varying academic backgrounds in STEM subjects. Inclusion criteria include:

- Current enrollment in programs within CS or AE faculties.
- No prior exposure to formal ML courses.
- Availability for all study sessions.

Recruitment involved personal visits to lecture halls and lab sessions within the Computer Science and Aerospace Engineering faculties. During these visits, the study objectives and timeline were explained to students directly, and interested participants were provided with an information sheet for further details. This approach ensured a more engaging and personal interaction, fostering greater interest and participation in the study. Following the recruitment phase, all research materials, including the survey, tutorials, and assessments, were distributed and conducted online. This approach was chosen to ensure accessibility, flexibility, and convenience for participants. By delivering the content online, students could engage with the materials at their own pace and from any location, reducing logistical barriers such as scheduling conflicts or the need to physically attend sessions. Additionally, this method also facilitated efficient data collection and analysis, as responses were automatically recorded and securely stored.

4.2 Initial Survey, Tutorial Delivery, and ML Concept Assessment

Before starting the research, ethical concerns were carefully addressed to ensure the integrity of the study. The researcher decided not to collect any personal identifying information, such as names, ages, or genders, to respect participants' privacy and eliminate any potential biases associated with personal data. Participants were provided with detailed information sheets outlining the study objectives, procedures, and their rights as participants, including the right to withdraw at any time without consequence.
To further ensure privacy, participants were prompted to create a unique, anonymous code name in the initial survey. This code name, formed by combining their favorite color, dessert, and animal (e.g., "Blue Cake Tiger"), was used consistently throughout the research. This method enabled data to be matched and analyzed while preserving anonymity. The study adhered to institutional ethical standards, and all data were anonymized and securely stored to maintain confidentiality.
The assessment phase involves an initial educational background survey, followed by a series of tutorials to teach key ML concepts. Lastly, the participants are requested to complete a Machine Learning test. These steps are outlined below:

Initial Background Assessment
Participants will complete:

- **Educational Background Survey:** A questionnaire to collect prior exposure to STEM subjects, learning preferences, and their confidence in STEM subjects.
- **Mathematics Assessment:** A short test evaluating foundational mathematical knowledge in calculus, linear algebra, and probability. This test is designed to measure competencies directly relevant to ML topics, listed below:
  1. Calculus
  2. Probability and Statistics
  3. Linear Algebra

  It includes a combination of multiple choice questions, short answer questions, and one open question, ensuring a well-balanced distribution in terms of both content and difficulty levels. All the math questions had the same weight, and the students' score were calculated as the number of questions answered correctly. The complete versions of the educational background survey and mathematical assessment are available in Appendix A.

ML Tutorials
Participants will engage in three interactive tutorials, each focusing on a key machine learning topic:

- **Bayes' Theorem:** An exploration of probabilistic reasoning, showcasing its relevance and applications in machine learning models such as classification algorithms.
- **Perceptrons:** A foundational concept in neural networks, demonstrating the use of linear algebra and calculus to perform binary classification through linear decision boundaries.
- **ML Pipeline:** A comprehensive overview of the machine learning workflow, including concepts like overfitting, underfitting, and the importance of training, validation, and testing datasets in building reliable models.

The entire tutorial was made through Notion, to ensure structured content delivery, visual integration, and accessibility. The researcher was responsible for making the part for the Bayes' Theorem. The full version of ML tutorial can be found on the website link. The learning objectives for each topics could also be found within each section of the tutorial.

Post Tutorial ML Assessment After completing the tutorials, the participants took an ML Assessment Test. This test evaluated the participants' understanding of the topics covered in the tutorials, emphasizing concepts that required mathematical reasoning. The assessment consists of 15 multiple choice questions that cover both theoretical concepts and mathematical calculations. Care was taken to ensure that all questions could be solved using the material covered in the tutorial. The results of this test were used to measure the learning outcomes for each group of participants. The complete version of the Machine Learning Assessment is available in Appendix B.

To ensure a cohesive learning experience, the math assessment and ML tutorials were designed in parallel, aligning specific sections of the math assessment with corresponding sections of the ML tutorial. This alignment aimed to reinforce the mathematical foundations necessary for understanding key ML concepts:

- **ML Pipeline – Calculus:** The tutorial on the ML pipeline introduced topics like overfitting, underfitting, and the role of optimization in model training. These concepts rely heavily on calculus, particularly derivatives and their applications, which were tested in the math assessment.
- **Bayes' Rule – Probability:** The tutorial on Bayes' rule focused on probabilistic reasoning and its applications in machine learning models, such as classification algorithms. The associated math assessment section evaluated participants' understanding of probability and statistics, directly supporting their comprehension of this tutorial.
- **Perceptrons – Linear Algebra:** The tutorial on perceptrons highlighted their reliance on linear algebra, including vector operations and matrix calculations, essential for understanding the role of perceptrons in neural networks. The corresponding math assessment section tested participants' proficiency in linear algebra.

This parallel design was intentional and rooted in the study's research question: to investigate whether prior mathematical knowledge influences the learning of machine learning topics. By structuring the tutorials and assessments in this manner, participants were provided with a clear connection between mathematical theory and its practical application in ML. This approach not only supported a more comprehensive evaluation of the participants' learning outcomes but also emphasized the interdisciplinary nature of ML, bridging abstract mathematical concepts with real-world applications. This alignment underscores the importance of tailoring ML education to the mathematical competencies of learners, ultimately aiding in the study's goal of improving instructional methodologies for diverse academic backgrounds.

Here is the revised section with paragraphs instead of bullet points:

4.3 Data Analysis

The collected data will be analyzed using both quantitative and qualitative methods to address the research questions.

Quantitative Analysis To assess the impact of the tutorials, descriptive statistics such as mean, median, and standard deviation will be calculated for the math and machine learning test scores of both groups. Additionally, Pearson correlation coefficients will be used to examine the relationships between participants' prior math knowledge and their performance on machine learning tests, providing insights into potential connections between foundational knowledge and new learning outcomes.

Qualitative Analysis Qualitative data from the survey responses will be analyzed through thematic analysis to identify key trends in students' perceptions of the tutorials and their self-reported learning experiences. The open feedback provided by the participants will also be reviewed to identify the instructional strategies that were particularly effective or challenging, offering valuable insights into how the tutorials can be improved in future iterations.

The researcher's contributions extended to the design and implementation of the assessment and instructional materials. Specifically, the researcher was responsible for developing the 'Probability and Statistics' questions in the math assessment, ensuring that they aligned with foundational ML concepts. Additionally, the researcher took the lead in creating the Bayes rule tutorial, which introduced participants to probabilistic reasoning and its applications in machine learning. The accompanying tutorial assessments were also designed to evaluate participants' understanding of this critical concept. By focusing on these areas, the researcher ensured that the materials were tailored to bridge the gap between theoretical mathematical knowledge and its practical applications in ML, while also addressing the study's core research objectives.

Furthermore, the researcher played a key role in designing the post-tutorial ML assessment, focusing on the alignment of questions with the tutorial content. This ensured that the assessment not only measured theoretical understanding, but

1	Aerospace Engineering	ML Pipeline (ML Score)	3.22	3.05	1.0
2	Aerospace Engineering	Calculus Score	0.99	1.35	0.89
3	Aerospace Engineering	Perceptron (ML Score)	2.93	3.0	1.11
4	Aerospace Engineering	Linear Algebra Score	1.04	0.8	1.11
5	Aerospace Engineering	Bayes Rule (ML Score)	3.28	2.95	1.24
6	Aerospace Engineering	Probability Score	2.61	2.85	1.13
7	Computer Science and Engineering	ML Pipeline (ML Score)	4.04	3.95	0.82
8	Computer Science and Engineering	Calculus Score	1.28	0.9	0.46
9	Computer Science and Engineering	Perceptron (ML Score)	3.51	3.7	0.5
10	Computer Science and Engineering	Linear Algebra Score	1.1	0.8	0.95
11	Computer Science and Engineering	Bayes Rule (ML Score)	3.55	3.1	1.01
12	Computer Science and Engineering	Probability Score	2.65	2.85	1.08

Figure 1: Average Score Values

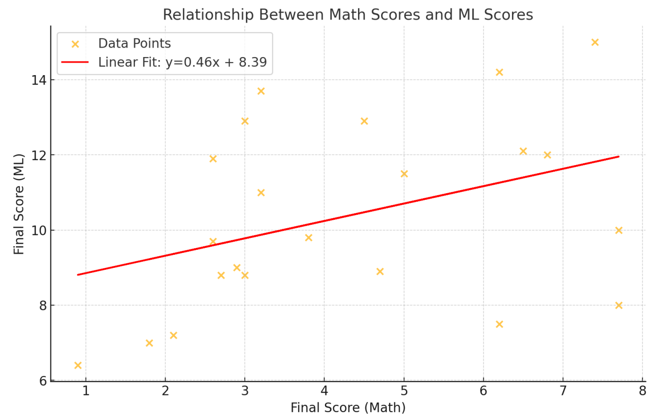

Figure 2: Relationship Between Math and ML Scores

also provided insight into students' ability to apply ML concepts. Through these contributions, the researcher aimed to create a coherent and effective learning experience while maintaining consistency with the overarching objectives of the study.

In summary, the methodology of this study was carefully designed to investigate the relationship between prior mathematical knowledge and learning outcomes in machine learning. By integrating participant recruitment, structured assessments, interactive tutorials, and comprehensive data analysis, the approach ensures a holistic evaluation of the research questions. The alignment of the math assessment with the ML tutorials highlights the interdisciplinary nature of the study, bridging theoretical mathematical concepts with practical applications in machine learning. This methodological framework provides a solid foundation for analyzing the influence of mathematical competencies on ML education and contributes valuable insights to the development of effective instructional strategies for diverse academic audiences.

5 Results

This section presents the findings of the study, focusing on the relationship between prior mathematical knowledge and the learning outcomes of machine learning (ML) topics. The analyses include visual representation of the data, correlation assessments, and interpretation of statistical significance.

5.1 Relationship Between Math and ML Scores

The analysis of average scores and their relationships reveals notable differences between Aerospace Engineering (AE) and Computer Science (CS) students. As illustrated in Figure 1 and Figure 2, CS students consistently outperformed their AE counterparts across all ML topics. In the ML Pipeline, CS students had higher mean (4.04 vs. 3.22) and median (4.00 vs. 3.25) scores, with a slightly larger standard deviation (Std: 0.82 vs. 0.77), suggesting a broader range of high-performing individuals. For Perceptrons, CS students again demonstrated stronger performance (Mean: 3.51, Median: 3.50, Std: 0.89) compared to AE students (Mean: 2.93, Median: 3.00, Std: 0.85), reflecting greater consistency in understanding neural network concepts. In Bayes Rule, the pattern persisted, with CS students showing both higher averages (Mean: 3.55 vs. 3.28) and lower variability (Std: 0.70 vs. 0.76).

Similarly, in mathematical topics, CS students demonstrated stronger and more consistent foundational knowledge. For Calculus, CS students achieved a higher mean (1.28 vs. 0.99) and median (1.30 vs. 1.00) with slightly larger variability (Std: 0.25 vs. 0.21). In Linear Algebra, their mean (1.10) and median (1.10) were slightly higher than AE students (Mean: 1.04, Median: 1.05), with comparable standard deviations (Std: 0.30 vs. 0.27). For Probability, CS students again outperformed AE students (Mean: 2.65, Median: 2.70, Std: 0.33 vs. Mean: 2.61, Median: 2.60, Std: 0.30).

This variability among AE students highlights the need for tailored instructional strategies to address gaps in their foundational preparation. For example, in ML Pipeline tasks, AE students' slightly lower median (3.25 vs. 4.00) and higher standard deviation (0.77) suggest difficulty in applying calculus-based optimization concepts. Similarly, the broader range of scores in Perceptrons and Bayes Rule among AE students underscores the challenge of leveraging linear algebra and probability skills in ML contexts. Furthermore, the scatter plot in Figure 2 shows a generally positive but weak relationship between mathematics and ML scores overall. This indicates that while mathematical proficiency is important, other factors such as study habits, learning styles, and prior programming exposure likely play a significant role in determining ML learning outcomes. These findings emphasize the importance of creating interdisciplinary learning environments that address diverse academic backgrounds and provide targeted support where needed.

		Correlation	p-value
1	Calculus - ML Pipeline	0.284	0.2
2	Probability - Bayes Rule	0.561	0.007
3	Linear Algebra - Perceptrons	0.44	0.04

Figure 3: Correlation among subjects

5.2 Correlation among subjects

The analysis of Figure 3 reveals notable relationships between the selected mathematical topics and their corresponding machine learning (ML) concepts. A moderate positive correlation (r = 0.561, p = 0.007) between Probability scores and the Bayes rule suggests that students with stronger probability skills tend to perform better on tasks involving the Bayes rule, a concept that is heavily reliant on probabilistic reasoning. Similarly, a moderate positive correlation (r = 0.440, p = 0.040) was observed between Linear Algebra and Perceptron scores, indicating that linear algebraic proficiency, essential to understanding vectors, matrices, and decision boundaries, contributes significantly to the mastery of perceptrons in neural networks. In contrast, the relationship between Calculus and ML pipeline scores was weaker (r = 0.284, p = 0.200), suggesting that while calculus is relevant for optimization and gradient-based methods, its immediate impact on understanding ML pipeline tasks may not be as pronounced. In general, the findings emphasize that mathematical foundations play varying roles in understanding different concepts of ML, with probability and linear algebra showing stronger direct associations compared to calculus.

5.3 Qualitative Analysis among Faculties

This subsection explores faculty-specific patterns in learning experiences, preferences, and challenges related to machine learning (ML) education. The analysis highlights key themes and differences between Aerospace Engineering (AE) and Computer Science (CS) students, providing insights to guide instructional strategies.

1. Perceived Difficulty of ML Topics AE students generally found ML topics like Perceptrons and Bayes Rule more challenging than their CS counterparts. This difficulty is reflected in the wider variability of their responses, suggesting a broader range of preparedness within the AE cohort. In contrast, CS students provided more consistent ratings, aligning with their stronger backgrounds in mathematics and programming. These differences highlight the need for tailored support for AE students, such as breaking down complex concepts or using domain-specific examples. Figure 4 presents detailed data on the perceived difficulty of ML topics for both faculties.

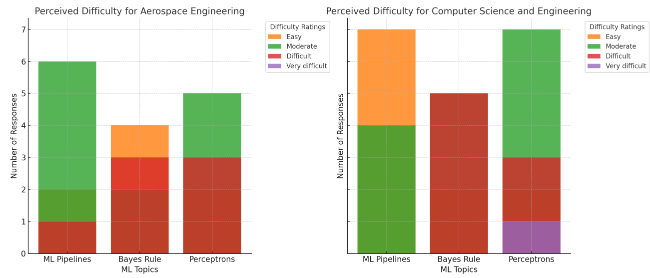

Figure 4: Perceived Difficulty for Both Faculties

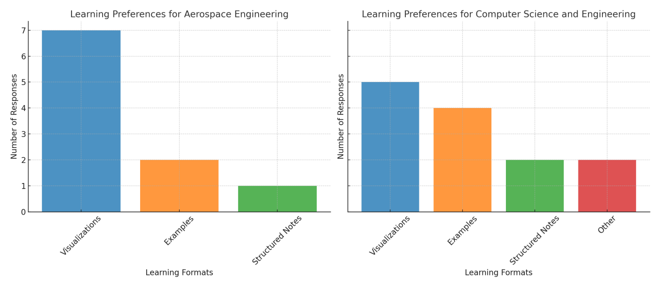

Figure 5: Learning Preferences for Both Faculties

2. Learning Preferences Learning preferences varied between the two faculties. CS students strongly preferred structured notes and visualizations, emphasizing clarity and logical organization in study materials. AE students, while also valuing visual aids, expressed a greater need for real-world examples and practical applications. These preferences suggest that tailoring instructional methods to match the practical orientation of AE students and the theoretical focus of CS students can improve learning outcomes. Figure 5 illustrates the learning preferences for both faculties.

3. Time Investment AE students reported a broader distribution of time spent on tutorials, including more instances of "More than 2 hours" responses. This pattern suggests that AE students may face a steeper learning curve, requiring more time to grasp ML topics. On the other hand, CS students demonstrated a tighter clustering of responses around "Between 1 hour and 2 hours," likely reflecting their familiarity with the subject matter. Figure 6 provides the time investment data for both faculties.

4. Teaching Mediums AE students showed a strong preference for interactive formats, such as video tutorials and lectures, which align with their desire for engaging, dynamic content. CS students, in addition to valuing videos, preferred slides and textbooks, reflecting their comfort with self-guided learning materials. This suggests that a blended teaching approach, combining interactive and structured content, could meet the diverse needs of both groups. The preferences for teaching mediums are shown in Figure 7.

5. Insights for Instructional Design The qualitative data un-

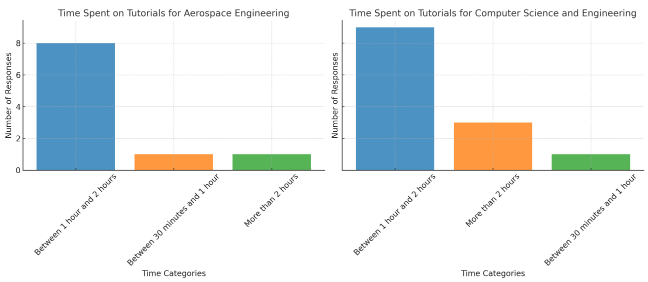

Figure 6: Time Spent on Tutorials for Both Faculties

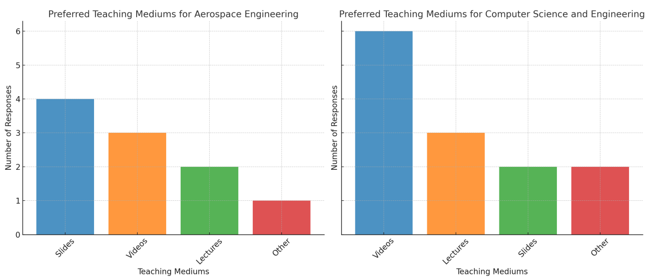

Figure 7: Preferred Teaching Mediums for Both Faculties

derscores the importance of an interdisciplinary approach to ML education. For AE students, integrating practical applications, real-world examples, and interactive content can help address gaps in understanding. For CS students, challenging their strengths in mathematics and programming with advanced tasks and structured resources can encourage deeper learning. By aligning instructional strategies with the unique needs of each faculty, educators can create more effective and engaging learning environments.

6 Discussions

This chapter looks at the study's findings, focusing on the differences in learning outcomes between Computer Science (CS) and Aerospace Engineering (AE) students. It explores how factors like math knowledge and programming experience influenced their understanding of machine learning (ML). The chapter also discusses the study's limitations and suggests ways the research could be improved or expanded in the future.

6.1 Interpretation of findings

The results demonstrate clear differences in the performance and learning patterns of Computer Science (CS) and Aerospace Engineering (AE) students. CS students consistently outperformed their AE counterparts across all machine learning (ML) topics, with higher mean scores and more symmetric distributions (as indicated by aligned means and medians). This consistency highlights the advantage of an academic background that integrates programming and mathematical foundations, which appears to facilitate the learning of ML concepts.

For example, in the ML Pipeline task, the higher mean (4.04) and lower variability (standard deviation: 0.82) among CS students suggest a strong ability to connect calculus-based optimization methods to practical ML workflows. In contrast, AE students displayed slightly lower performance (mean: 3.22, standard deviation: 0.77), likely due to limited exposure to programming, which is essential for applying ML concepts effectively.

Similarly, in Perceptrons, which rely heavily on linear algebra, CS students again outperformed AE students (mean: 3.51 vs. 2.93). The stronger foundation in programming and ML concepts among CS students likely enabled them to grasp the computational aspects of neural networks more readily. However, AE students, despite their strong mathematical preparation, faced challenges in translating their knowledge of linear algebra into the coding and application required for this topic.

Bayes Rule, a concept grounded in probability, revealed a similar trend. The higher mean score (3.55) and lower variability (standard deviation: 0.70) among CS students indicate that their combined mathematical and programming skills enabled more consistent comprehension. AE students, while slightly less consistent (mean: 3.28, standard deviation: 0.76), benefited from their quantitative background in probability, though their relative unfamiliarity with ML applications likely hindered their performance.

The correlation analysis further supports these observations, with moderate positive correlations between mathematical topics and their respective ML concepts. For instance, the correlation between probability and Bayes Rule scores ($r = 0.561$, $p = 0.007$) indicates that strong probability skills are crucial for mastering probabilistic reasoning in ML. Linear algebra showed a moderate correlation with Perceptron scores ($r = 0.440$, $p = 0.040$), emphasizing the importance of vector and matrix operations for understanding neural networks. However, the weaker correlation between calculus and ML Pipeline scores ($r = 0.284$, $p = 0.200$) suggests that while calculus is relevant, it may not be as immediately impactful for learning optimization concepts in ML.

6.2 Limitations of the study

Despite the valuable insights generated by this research, several limitations should be acknowledged.

First, the study included a relatively small sample of 20 participants (10 from each faculty). This limited sample size reduces the generalizability of the findings. A larger and more diverse group of participants would provide a stronger basis for identifying patterns and drawing conclusions across a broader range of academic backgrounds.

Second, the scope of the study was narrow, focusing on three specific ML topics: Bayes Rule, Perceptrons, and ML Pipelines. While these are foundational concepts, the inclusion of additional topics, such as clustering, deep learning, or reinforcement learning, could provide a more comprehensive understanding of how different mathematical skills influence ML learning.

Third, the analysis was limited to students from the Computer Science (CS) and Aerospace Engineering (AE) faculties. Expanding the study to include students from other disciplines, such as biology, economics, or social sciences, would offer broader insights into the influence of diverse academic backgrounds on ML comprehension.

Fourth, the study relied on short-term assessments conducted immediately after the tutorials. While these evaluations capture immediate learning outcomes, they do not address long-term retention or the ability to apply ML concepts in practical settings. Future research could incorporate longitudinal studies to explore these aspects.

Fifth, the study did not directly assess participants' prior programming experience, which likely contributed significantly to the observed differences in performance. Including a programming skills assessment in future studies could help isolate the impact of programming proficiency on ML learning outcomes.

Finally, the delivery format of the tutorials and assessments was entirely online. While this approach ensured accessibility and convenience, it may have influenced engagement and performance. In-person sessions could provide richer opportunities for interaction, collaboration, and immediate clarification of complex concepts, which might impact the results.

7 Conclusion

This study explored the influence of prior mathematical knowledge on the learning outcomes of machine learning (ML) topics among Computer Science (CS) and Aerospace Engineering (AE) students. Through quantitative and qualitative analyses, key differences were identified between these two academic groups, shedding light on how foundational skills and academic backgrounds shape ML comprehension.

The results revealed notable differences in performance and learning patterns between the two groups. Students with a background that included significant programming exposure and an integrated approach to mathematical preparation demonstrated more consistent success across ML topics. Conversely, those with strong quantitative skills but limited programming experience exhibited greater variability in their performance, underscoring the challenges of bridging theoretical knowledge with its application in ML tasks. Notably, probability and linear algebra emerged as key contributors to ML comprehension, showing stronger correlations with learning outcomes compared to calculus, which played a more supporting role in this context.

Qualitative findings further emphasized the need for tailored instructional strategies. AE students expressed a preference for interactive, application-driven learning, while CS students valued structured and technically detailed resources. These differences highlight the importance of interdisciplinary approaches to ML education, blending theoretical rigor with practical applications to address the diverse needs of learners.

Despite its contributions, the study had limitations, including a small sample size and a narrow focus on specific ML topics. Future research should include larger, more diverse participant groups and explore a broader range of ML concepts. Longitudinal studies could also provide insights into the long-term effectiveness of different instructional strategies and their impact on skill retention.

The findings of this research have significant implications for the design of ML curricula. Educators should consider incorporating adaptive learning technologies, real-world applications, and interdisciplinary teaching frameworks to bridge gaps in mathematical and programming competencies. By tailoring instruction to the strengths and challenges of students from various academic backgrounds, ML education can be made more inclusive and effective, equipping learners with the skills needed to apply ML in diverse professional and academic contexts.

As machine learning continues to impact a wide range of fields, fostering a deeper understanding of how students learn ML across disciplines is essential. This study contributes to that effort by providing actionable insights into the relationship between mathematical knowledge, programming skills, and ML comprehension. With further research and innovation in teaching methodologies, ML education can evolve to meet the growing demands of a diverse and rapidly changing world.

References

[1] Alagar. Machine learning applications and examples. 1 2025. URL https://iabac.org/blog/machine-learning-applications-and-examples.

[2] Omar Shouman, Simon Fuchs, and Holger Wittges. Experiences from Teaching Practical Machine Learning Courses to Master's Students with Mixed Backgrounds. URL https://openreview.net/forum?id=ijuM-MVwVEk.

[3] Amy J. Ko. We need to learn how to teach machine learning - Bits and Behavior - Medium. 5 2018. URL https://medium.com/bits-and-behavior/we-need-to-learn-how-to-teach-machine-learning-acc78bac3ff8.

[4] Paul J. Atzberger. Importance of the mathematical foundations of machine learning methods for scientific and engineering applications, 8 2018. URL https://arxiv.org/abs/1808.02213?utm.

[5] Weicheng Wang and Jinye Ma. A review: Applications of machine learning and deep learning in aerospace engineering and aero-engine engineering. *Advances in Engineering Innovation*, 6(1):54–72, 2 2024. doi: 10.54254/2977-3903/6/2024060. URL https://doi.org/10.54254/2977-3903/6/2024060.

[6] Leif Sundberg and Jonny Holmstrom. Teaching tip Using No-Code AI to teach machine learning in higher education. *Deleted Journal*, pages 56–66, 1 2024. doi: 10.62273/cypl2902. URL https://doi.org/10.62273/cypl2902.

[7] Shannon Wongvibulsin. Educational strategies to foster diversity and inclusion in machine intelligence. *Nature Machine Intelligence*, 1(2):70–71, 1 2019. doi: 10.1038/s42256-019-0021-8. URL https://doi.org/10.1038/s42256-019-0021-8.

[8] Kamilla Tenório and Ralf Romeike. AI Competencies for non-computer science students in undergraduate education: Towards a competency framework. *AI Competencies for non-computer science students in undergraduate education: Towards a competency framework*, 11 2023. doi: 10.1145/3631802.3631829. URL https://doi.org/10.1145/3631802.3631829.

[9] Maria Kasinidou, Styliani Kleanthous, Matteo Busso, Marcelo Rodas, Jahna Otterbacher, and Fausto Giunchiglia. Artificial Intelligence in Everyday Life 2.0: Educating University Students from Different Majors. *Artificial Intelligence in Everyday Life 2.0: Educating University Students from Different Majors*, pages 24–30, 7 2024. doi: 10.1145/3649217.3653542. URL https://doi.org/10.1145/3649217.3653542.

[10] Musa Adekunle Ayanwale, Rethabile Rosemary Molefi, and Saheed Oyeniran. Analyzing the evolution of machine learning integration in educational research: a bibliometric perspective. *Discover Education*, 3(1), 5 2024. doi: 10.1007/s44217-024-00119-5. URL https://doi.org/10.1007/s44217-024-00119-5.

A Initial Survey and Math Assessment

The following appendix contains the full details of the survey and math assessment administered during the study. These materials were designed to collect essential background information about participants and to evaluate their foundational mathematical knowledge, which is directly relevant to understanding machine learning (ML) concepts.

Learning Machine Learning - Survey on Mathematics Background (9점)

Welcome, and thank you for participating in this research! This study investigates how prior mathematics knowledge influences learning outcomes when students are introduced to foundational machine learning (ML) topics.
You will be asked to provide information about your past maths-related coursework and activities by asking you questions about your experience and with mathematics questions. Afterwards, you can continue with the tutorials for the ML topics. After you study it and become familiar with it, we will test your learning outcome with a set of quiz in Week 7. The information you provide in this form, combined with quiz results, will help us analyze trends across students with different academic backgrounds.

Important Notes:

- Participation is entirely voluntary, and you may withdraw at any time.
- Your responses will remain anonymous, and no personal identifiers will be collected.
- Data will be used only for academic purposes and published in aggregated form in a master's thesis and the university repository.
- Please answer the questions honestly to ensure the reliability of the research.

2

Please generate your unique participant id using the following formula:
Your favorite color + favorite dessert + favorite animal.(e.g. Orange Stroopwafel Cat)
Please remember this id, as you will use it again for the quiz in Week 7. *

답변을 입력하세요.

3

What is your current degree program? *

- ○ Computer Science and Engineering
- ○ Industrial Design
- ○ Applied Mathematics
- ○ Aerospace Engineering
- ○ Other

4

What is your current academic level? *

- ○ BSc
- ○ MSc
- ○ Other

5

Please list any maths-related courses you took in high school (e.g., IB Mathematics HL, AP Mathematics, Dutch Mathematics A). *

답변을 입력하세요.

6

Please list any maths-related courses you took Bachelor's degree (if applicable).

답변을 입력하세요.

7

Please list any maths-related courses you took during your Master's degree (if applicable).

답변을 입력하세요.

8

Please list any extracurricular maths-related activities or courses you have participated in (e.g., online courses, maths clubs).

답변을 입력하세요.

9

Please list any prior experience in machine learning (e.g., highschool course, online course, YouTube tutorial).If you have any, please list the topics you have learnt.

답변을 입력하세요.

10

How confident do you feel in your mathematics skills? Rate from 1 being not confident at all to 5 being very confident. *

1	2	3	4	5

11

Following questions are mathematics questions. Please give solutions to them without using any extra materials (e.g., calculators, textbooks, etc). You may use pen and paper. If you are unsure, you can write "Unsure" and move on to the next question. *

○ I understand

12

Q1 (1점)

The gradient of a function $f(x) = x^2 + 3x + 1$ at a point is given by the derivative of the function at that point. What is the gradient of $f(x)$ at $x = 3$?

답변을 입력하세요.

정답: 9

13

Q2 (1점)

Given the function $f(x, y) = 3x^2 + 2x + 2y^2$, what is the gradient vector ∇f at the point $(x, y) = (1, 2)$?

답변을 입력하세요.

정답: (8,8) | (8, 8) | [8,8] | [8, 8]

14

Q3 (1점)

Question 3. Suppose we have matrices $A = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 4 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ -2 & 1 \\ 3 & 0 \end{bmatrix}$. Calculate matrix $C = AB$

답변을 입력하세요.

정답: [-3 2 ₩₩ 1 4] | [-3 2 / 1 4]

15

Q4a (1점)

Question 4a. Suppose we have matrix $T = \begin{bmatrix} 1 & 3 & 1 \\ 0 & 2 & 5 \\ 0 & 0 & 4 \end{bmatrix}$, calculate its eigenvalues λ_1, λ_2 and λ_3

답변을 입력하세요.

16

Q4b (1점)

Question 4b. and calculate their corresponding eigenvectors $\vec{x_1}, \vec{x_2}$ and $\vec{x_3}$

답변을 입력하세요.

17

Q5 (1점)

Question 5. Given invertible matrices A, B and C, as wel as their product $D = ABC$, define $D^\top$ in terms of A, B and C.

답변을 입력하세요.

정답: D^T = C^T B^T A^T

18

Q6 (1점)

A bag contains 5 red balls, 3 blue balls, and 2 green balls. A ball is randomly drawn from the bag.

1. What is the probability that the ball is red?
2. What is the probability that the ball is blue after taking out the red ball?

답변을 입력하세요.

정답: 1/2, 1/3 | 5/10, 3/9 | 0.5, 0.33 | 50%, 33% | 1/2, 3/5 | 5/10, 3/5 | 0.5, 0.6 | 50%, 60%

19

Q7 (1점)

In a survey of people's beverage preferences:

- 60% of people like coffee (this includes those who like both coffee and tea).
- 40% of people like tea (this includes those who like both coffee and tea).
- 20% of people like both coffee and tea.

If a randomly selected person is known to like tea, what is the probability that this person also likes coffee?

답변을 입력하세요.

정답: 0.5 | 50% | 1/2

20

Q8 (1점)

Two six-sided dice are rolled. Let event A be that the sum of the numbers rolled is 7, and let event B be that the first die shows an odd number. Are A and B independent? Show your calculations.

답변을 입력하세요.

B Machine Learning Assessment

This Machine Learning Test was designed to evaluate participants' comprehension of the key topics covered in the tutorials, including Bayes' theorem, perceptrons, and the ML pipeline. The assessment emphasized the application of mathematical reasoning to solve problems, bridging theoretical knowledge and practical understanding.

Learning Machine Learning: Test Your Knowledge (15점)

Welcome to the final step of our journey! Here, you will solve questions that are related to the topics that you have learned for the past weeks. This is not an assessment, and you will receive no penalties for your scores. If you are not sure about how to solve a question, choose "Not sure"; please do not refer to any external resource or try to guess an answer. You will solve 15 multiple choice questions, and at the end we prepared some questions for you to tell us about your experience during the learning program. Thank you for your time and cooperation in this study, and good luck!

1

Enter the same id you used for the math test. If you did not write it down, it was of the following format:
Your favorite color + favorite dessert + favorite animal. (e.g. Orange Stroopwafel Cat) *

답변을 입력하세요.

2

What is your current degree program? *

- Computer Science and Engineering
- Industrial Design
- Applied Mathematics
- Aerospace Engineering
- Applied Physics
- Other

3

Why do we split data into training and test sets in a machine learning pipeline? * (1점)

- To evaluate the model's performance on unseen data. ✓
- To improve the speed of the model.
- To reduce the size of the dataset.
- To optimize the hyperparameters of the model.
- Not sure

4

You are training a machine learning model with a dataset of 10,000 samples. If you use 80% of the data for training and 20% for testing, which of the following might happen? * (1점)

- The model may not have enough data to learn patterns properly.
- The model may suffer from overfitting during training.
- The test set might not fully represent the variability in the data. ✓
- The model's performance on the test set might overestimate its real-world accuracy.
- Not sure

5

What happens when a machine learning model overfits? * (1점)

- The model becomes more efficient in processing data.
- The model performs well on training data but poorly on new data. ✓
- The model fails to learn the patterns in the training data.
- The model performs equally well on training and test data.
- Not sure

6

A model achieves 95% accuracy on the training set but only 60% accuracy on the test set. What does this indicate, and how can it be resolved? * (1점)

- The test set is not representative; adjust the train-test split.
- Underfitting; increase model complexity or use a larger test set.
- Balanced performance; no changes are needed.
- Overfitting; reduce model complexity or gather more training data. ✓
- Not sure

7

Why is it important to include a validation set when training a model? * (1점)

- It ensures the test set remains untouched until final evaluation. ✓
- It reduces the size of the training set, preventing overfitting.
- It simplifies the pipeline by eliminating the need for a test set.
- It allows the model to learn more patterns from the data.
- Not sure

8

Which of the following best describes the role of the **prior probability** in Bayes' Rule? * (1점)

- It measures the compatibility of evidence with a hypothesis.
- It is the initial belief about an event before new evidence is considered. ✓
- It is the total probability of the evidence occurring.
- It represents the updated belief after considering new evidence.
- Not sure

9

In Bayesian classification, which of the following describes a **Type I error**? * (1점)

- Incorrectly classifying an item as belonging to a class when it does not. ✓
- Failing to classify an item into a class when it belongs there.
- Miscalculating the prior probability of a class.
- Minimizing the posterior probability of an incorrect classification.
- Not sure

10

Spam Email Detection:

- 90% of spam emails are correctly identified as spam
- 10% of legitimate emails are incorrectly identified as spam
- 40% of all emails are spam
- 60% of all emails are legitimate

What is the Bayes error rate for this spam filter? * (1점)

- 5%
- 10% ✓
- 15%
- 20%
- Not sure

11

A company screens applicants for a job using a test. The test is designed such that:

- 80% of qualified applicants pass the test
- 30% of unqualified applicants pass the test
- 60% of applicants are unqualified
- 40% of applicants are qualified

If an applicant passes the test, what is the probability that they are actually qualified? * (1점)

- ○ 56%
- ○ 64% ✓
- ○ 72%
- ○ 82%
- ○ Not sure

12

A factory uses a machine to sort defective items. The sorting system is imperfect:

- P(Detected Defective|Defective) = 0.9
- P(Not Detected Defective|Not Defective) = 0.85
- P(Defective) = 0.05
- P(Not Defective) = 0.95

The cost of classifying a defective item as not defective is $10. The cost of classifying a non-defective item as defective is $5.

Given this information, how should the system classify an item if the system detects it as defective? * (1점)

- ○ Defective
- ○ Not Defective ✓
- ○ Not sure

13

What is **not** a disadvantage of using an Artificial Neural Network? * (1점)

- ○ ANNs are not good at finding complex patterns in datasets. ✓
- ○ It is difficult to figure out what made the model give a certain output.
- ○ The training of ANNs generally requires large amounts of data.
- ○ ANNs need a lot of computing power for training the model.
- ○ Not sure

14

Given that the formula for updating weights during training is:

$$w_i \leftarrow w_i + \eta \cdot (y - \hat{y}) \cdot x_i$$

What can we say about the learning rate η? * (1점)

- ○ The update of the weights are only dependent on whether the prediction is correct or wrong, not by how far is from the real expected output.
- ○ The learning rate needs to be positive. ✓
- ○ If the model predictions are 100% correct, the weights can still change.
- ○ Not sure

15

Given a perceptron with weight vector [3,-1,1], bias -2, and activation function f(x) = -1 if x < 0, f(x) = 1 if x >= 0. What would the perceptron output with input vector [-1,-2,3]? * (1점)

- ○ -1
- ○ 0
- ○ 1 ✓
- ○ 2
- ○ Not sure

16

Which of the following best describes a model acting as a black box? * (1점)

- ○ In a worst case scenario, the model will still function.
- ○ The model requires a lot of data for computing its output.
- ○ It is hard to find out how the model came to its output. ✓
- ○ The model has a large memory to store data.
- ○ Not sure

17

Assume we have a Multi-Layer Perceptron with 3 input nodes, two hidden layers of 4 nodes (h1 & h2), and an output layer of 2 nodes (out). What are the sizes of the weight matrices that can store this model? * (1점)

- ○ w_h1 = 3x1, w_h2 = 4x1, w_out = 2x1
- ○ w_h1 = 3x4, w_h2 = 4x4, w_out = 4x2 ✓
- ○ w_h1 = 4x1, w_h2 = 4x1, w_out = 2x1
- ○ w_h1 = 3x2, w_h2 = 4x2, w_out = 4x2
- ○ Not sure

18

How did you find the difficulty of the topics you learned? *

	Very easy	Easy	Moderate	Difficult	Very difficult
Topic 1: ML pipelines	○	○	○	○	○
Topic 2: Bayes' Rule	○	○	○	○	○
Topic 3: Perceptrons	○	○	○	○	○

19

How did you find the difficulty of the test? *

	Very easy	Easy	Moderate	Difficult	Very difficult
Difficulty	○	○	○	○	○

20

How much time did you roughly take for studying the tutorial? *

- ○ Less than 30 minutes
- ○ Between 30 minutes and 1 hour
- ○ Between 1 hour and 2 hours
- ○ More than 2 hours

21

During the learning phase, which part did you find the most comfortable to learn/easiest to understand? Why do you think that is? *

답변을 입력하세요.

22

During the learning phase, which part did you find the most difficult to learn/hardest to understand? Why do you think that is? *

22

During the learning phase, which part did you find the most difficult to learn/hardest to understand? Why do you think that is? *

답변을 입력하세요.

23

Were there any parts or formats of the tutorials that you found particularly helpful in learning and understanding new topics? If yes, what were they? *

답변을 입력하세요.

24

If you would teach the topics to students from your own study, how would you teach them? What kind of medium would you use? *

답변을 입력하세요.