# Challenges and Preferences of Learning Machine Learning: A Student Perspective

Effat Farhana* Fan Wu† Hossain Shahriar‡ Shubhra Kanti Karmaker Santu§ Akond Rahman¶
* Auburn University, AL, USA
† Tuskegee University, AL, USA
‡ University of West Florida, FL, USA
§ University of Central Florida, AL, USA
¶ Auburn University, AL, USA
Email: *ezf0057@auburn.edu †fwu@tuskegee.edu ‡hshahriar@uwf.edu §santu@ucf.edu ¶akond@auburn.edu

***Abstract*—This research paper systematically identifies the perceptions of learning machine learning (ML) topics. To keep up with the ever-increasing need for professionals with ML expertise, for-profit and non-profit organizations conduct a wide range of ML-related courses at undergraduate and graduate levels. Despite the availability of ML-related education materials, there is lack of understanding how students perceive ML-related topics and the dissemination of ML-related topics. A systematic categorization of students' perceptions of these courses can aid educators in understanding the challenges that students face, and use that understanding for better dissemination of ML-related topics in courses. The goal of this paper is to help educators teach machine learning (ML) topics by providing an experience report of students' perceptions related to learning ML. We accomplish our research goal by conducting an empirical study where we deploy a survey with 83 students across five academic institutions. These students are recruited from a mixture of undergraduate and graduate courses. We apply a qualitative analysis technique called open coding to identify challenges that students encounter while studying ML-related topics. Using the same qualitative analysis technique we identify quality aspects do students prioritize ML-related topics.**

**From our survey, we identify 11 challenges that students face when learning about ML topics, amongst which data quality is the most frequent, followed by hardware-related challenges. We observe the majority of the students prefer hands-on projects over theoretical lectures. Furthermore, we find the surveyed students to consider ethics, security, privacy, correctness, and performance as essential considerations while developing ML-based systems. Based on our findings, we recommend educators who teach ML-related courses to (i) incorporate hands-on projects to teach ML-related topics, (ii) dedicate course materials related to data quality, (iii) use lightweight virtualization tools to showcase computationally intensive topics, such as deep neural networks, and (iv) empirical evaluation of how large language models can be used in ML-related education.**

***Index Terms*—artifical intelligence, empirical study, machine learning, perception**

## I. Introduction

According to the 'O'Reilly 2021 AI Adoption' survey [36], 3,500 business leaders identified a lack of skilled machine learning (ML) workforce to be a top challenge in adopting artificial intelligence (AI). To address this gap, profit and non-profit organizations are conducting courses where students are exposed to state-of-the-art ML techniques. These courses are aimed to help students in learning about ML techniques, tools, and practices. Education approaches that involve ML, such as ML-related courses have promise, e.g., artificial intelligence and ML-involved education market are expected to cross US $20 billion by 2027 [6].



The incorporation of students' perceptions can advance the science and development of ML-based courses as prior research [27], [30] shows pedagogical approaches to improve by accounting for students' perceptions. According to Shuell et al. [27], derivation of students' perceptions helps confirm or refute learning-related assumptions often held by educators. Struyven et al. [30] mentioned that collection and analysis of student perceptions are often neglected, and "*cannot be neglected if full understanding of student learning is the purpose of our educational research and practices*". Therefore, collecting and analyzing students' perceptions related to ML learning can not only improve ML-based courses but also advance knowledge related to ML education. Furthermore, we can derive perceptions on ML-related quality aspects, such as characteristics related to ML-based models' accuracy and performance. While prior research has investigated students' perceptions of bot-guided pedagogical approaches [34] and cybersecurity [18], [19], similar endeavors remain under-explored for ML-related courses.

*The goal of this paper is to help educators teach machine learning (ML) topics by providing an experience report of students' perceptions related to learning ML.*

We answer the following research questions (RQs):

- **RQ1**: What challenges do students encounter while studying ML-related topics?
- **RQ2**: What quality aspects do students prioritize while performing ML tasks?
- **RQ3**: What teaching methods do students prioritize for ML-related topics?
- **RQ4**: What are the students' preferences for ML-related topics?

We conduct a survey with 83 students to answer our RQs. We apply a qualitative analysis technique called open coding [22] to derive categories of challenges and preferences reported by students. We follow guidelines provided by the Internal Review Board (IRB) at University A prior to conducting our empirical study. We also follow recommended practices advocated by the ACM SIGSOFT [28] to conduct this empirical study. The ACM SIGSOFT community has provided guidelines on how to conduct surveys to quantify perceptions related to software engineering. We have used one their guidelines to conduct the survey for the paper.

**Contributions** We list our contributions as follows:

- A categorization of challenges reported by students that are related to learning about ML;
- A categorization of preferences reported by students, which are related to learning about ML; and
- A characterization of quality aspects prioritized by students when performing ML-related data quality tasks.

## II. Methodology

We collect responses from students by deploying an online survey. We describe the survey construction and deployment process as follows.

*1) Survey Construction:* The focus of the survey is to identify challenges that occur while learning ML and using that knowledge to develop ML-based applications. We follow the SIGSOFT Empirical Standards [28] to construct our survey: to identify the target population and sampling strategy (Table I), characterize the demographics (Results section), response management (Categorization for RQ1, RQ2, and RQ3 section), and analyzing construct validity (Threats to Validity section). Our survey includes multiple open-ended and Likert scale item questions, as described below:

- **Background-related questions**: We ask questions about the background of the students, focusing on their experience in machine learning, educational level (graduate or undergraduate), and the programming language they utilize for machine learning tasks. Except for educational level, all background-related questions are open-ended.
- **Challenges-related questions**: We ask questions where we explicitly ask about challenges that students face while collecting and curating data as well as applying ML techniques. Altogether, we ask three questions, which are all open-ended.
- **Preference-related questions**: We ask two questions related to preferences: one open-ended question asking students about their preferred teaching methods for ML-related education. The second one is a five-item Likert scale question asking students about their preferences for ML-related considerations. An ML-related consideration is an outcome of applying ML techniques for a computational task. We collect students' responses about their preferences for the following considerations: end result, security, privacy, ethics, correctness, and performance. We chose these factors since they are all integral to developing ML systems. End-result corresponds to the output of an ML model. Security corresponds to an ML technique's ability to preserve integrity, availability, or confidentiality. Privacy corresponds to an ML technique's ability to preserve users' sensitive information. Ethics corresponds to an ML technique's not violating ethical principles. Correctness corresponds to an ML technique's accuracy metrics. Performance corresponds to an ML technique's ability to consume fewer resources, such as CPU and memory.

As part of our survey, we asked the following questions:

1) "What are the problem topics you apply data science and ML techniques?"
2) "What challenges do you face?"
3) "What tools do you use?"
4) "How do you get data?"
5) "What teaching method do you prefer?"
6) "How important do you think about the following considerations in data science and ML tasks?" a) 'end result', b) 'security', c) 'privacy', d) 'ethics', e) 'correctness', f) 'performance'

Questions $1-5$ are open-ended. Question 6 has a five-point scale for each consideration: Extremely Important, Important, Somehow Important, I am neutral, Not Important. The Likert item Somehow Important expresses negative sentiment as a survey respondent is not fully convinced about the item that is being queried about.

*2) Survey Deployment:* We distribute the survey across five schools in the USA, as shown in Table I: two universities with "very high research activity" ($U-A$, $U-B$), two with "high research activity" ($U-C$, $U-D$), and one minority-serving university ($U-E$) with limited research activities according to the Carnegie Classifications of Higher Institutions [4]. The 'Level' column shows the level of the course; e.g., 'U/G' and 'G' denote undergraduate/graduate and graduate courses, respectively. We select these institutions because we want to collect student responses from diverse backgrounds. Teaching experiences might vary based on research activities, [3] and by deploying the survey we assume to capture the diversity in teaching experiences. We deploy the survey in one course for each of these five universities. Prior to the distribution of the survey, we seek approval from the IRB authority. As per IRB guidelines, we (i) do not collect any personal information, (ii) do not release students' grades, and (iii) explicitly mention that student participation is voluntary. The timeline was: Spring 2020 (U-C), Fall 2020 (U-C), Spring 2021 (U-C), Spring 2022 (U-A, U-D, U-E), and Fall 2022 (U-B). The courses taught at $U-C$ and $U-E$ is application oriented where the emphasis is on building ML-based applications, where theoretical aspects of ML are covered first. In the case of courses $U-A$, $U-B$, $U-D$ the emphasis is on the theoretical aspects of ML.

| Course Title | University | Level |
|---|---|---|
| Data Mining & Machine Learning | $U-C$ | U/G |
| Machine Learning | $U-A$ | G |
| Machine Learning | $U-B$ | G |
| Machine Learning | $U-D$ | G |
| Machine Learning | $U-E$ | U/G |

TABLE I: Course Attributes

*3) Categorization for RQ1, RQ2, and RQ3:* Our categorization approach for RQ1, RQ2, and RQ3 involves two phases: category derivation and rater verification for categorization.

**Step#1 Category Derivation.** We apply a qualitative analysis called open coding to identify themes in participants' responses [22]. First, we analyze each response and identify themes that emerged into codes. Next, we derive categories based on coding similarities. We discard blank and irrelevant answers (e.g., N/A and "Did not understand the question").
**Step#2 Rater Verification for Categorization.** The first author derived all categories, which is susceptible to bias. To mitigate bias, another author separately rated each response using derived categories. Next, we compute a Cohen's Kappa [5] between the two authors' ratings.

We repeat the above-mentioned steps for all our three RQs. A student's response can belong to multiple categories. We report the frequency of each category and student responses grouped by their prior data science experience. For RQ1, we analyze students' challenges from two perspectives: (1) what challenges do students face during the collection and curation of data? and (2) what challenges do students face while applying data science and ML techniques?



Figure 1 shows an example of category derivation for RQ1: What challenges do students face in ML courses while applying data science and ML techniques? The leftmost part shows three students' responses to the question. Next, we generate two initial coding categories from these three response texts e.g., 'choosing task specific right techniques' and 'choosing right technique'. Finally, we determine the category 'Choosing appropriate technique' by combining initial categories. We combine these two initial categories, as both correspond to a common pattern of selecting appropriate data mining techniques for the task at hand.

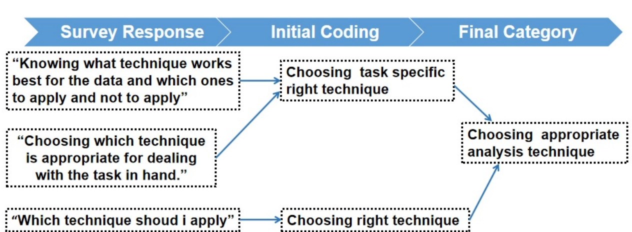


Fig. 1: Example: Category derivation using open coding.

*4) Survey Analysis for RQ4:* We answer RQ4 by analyzing the survey responses to the Likert response question 6 'How important do you think about the following considerations in data science and ML tasks' (see Survey Construction and Deployment section).

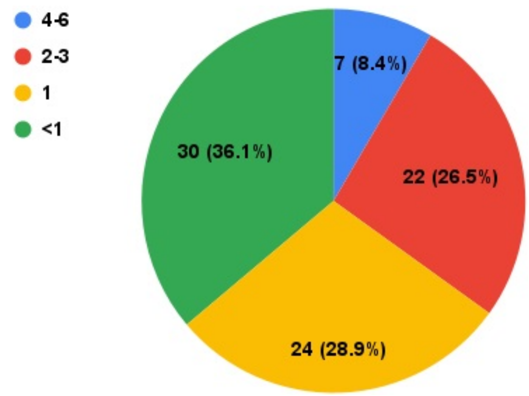


Fig. 2: Students' experience in ML-related topics (Years)

## III. Results

**Demographics.** We had 83 survey respondents, with 27 (32.5 %) undergraduate and 56 graduate students (67.5%). We asked students to report their academic or professional experience in data science in years. Figure 2 shows the count and percentage of student demographics by years of experience. We obtain 60, 12, 6, 2, and 3 respondents respectively, from $U-C$, $U-D$, $U-E$, $U-A$, and $U-B$. We refer to students as novice (N) with <= 1 year ($n=54$), intermediate (I) with 2-3 years ($n=22$), and experienced (E) with 4-6 years ($n=7$) of prior ML experience. 72 of the 83 students use Python, 33 R, 2 Julia, and 3 Matlab, respectively.

We provide answers to our research questions in the following subsections. Figure 3 is a brief summary of the findings from our survey on students' perceptions. All perceptions are grouped into three categories: 'challenge', 'quality assurance', and 'teaching preference'.

### A. Answer to RQ1

In this section, we answer **RQ1: What challenges do students encounter while studying ML-related topics?** by reporting: (1) challenges faced during data curation ('Data Curation' in Figure 3), and (2) challenges faced during applying ML techniques ('Apply Technique' in Figure 3).
*1) Data Curation Challenge:* We describe each category with examples of students' responses below. The frequency of students' responses is available in Table II.

- **Data Collection.** This category includes challenges related to data collection when accomplishing ML tasks. Students report data collection challenges due to access difficulty, insufficient data, time-intensive procedures, and tool interaction. For example, a student stated *"Since a lot of my work with data science was with malware analysis, it was difficult to obtain a dataset as the subject area is in malware."*

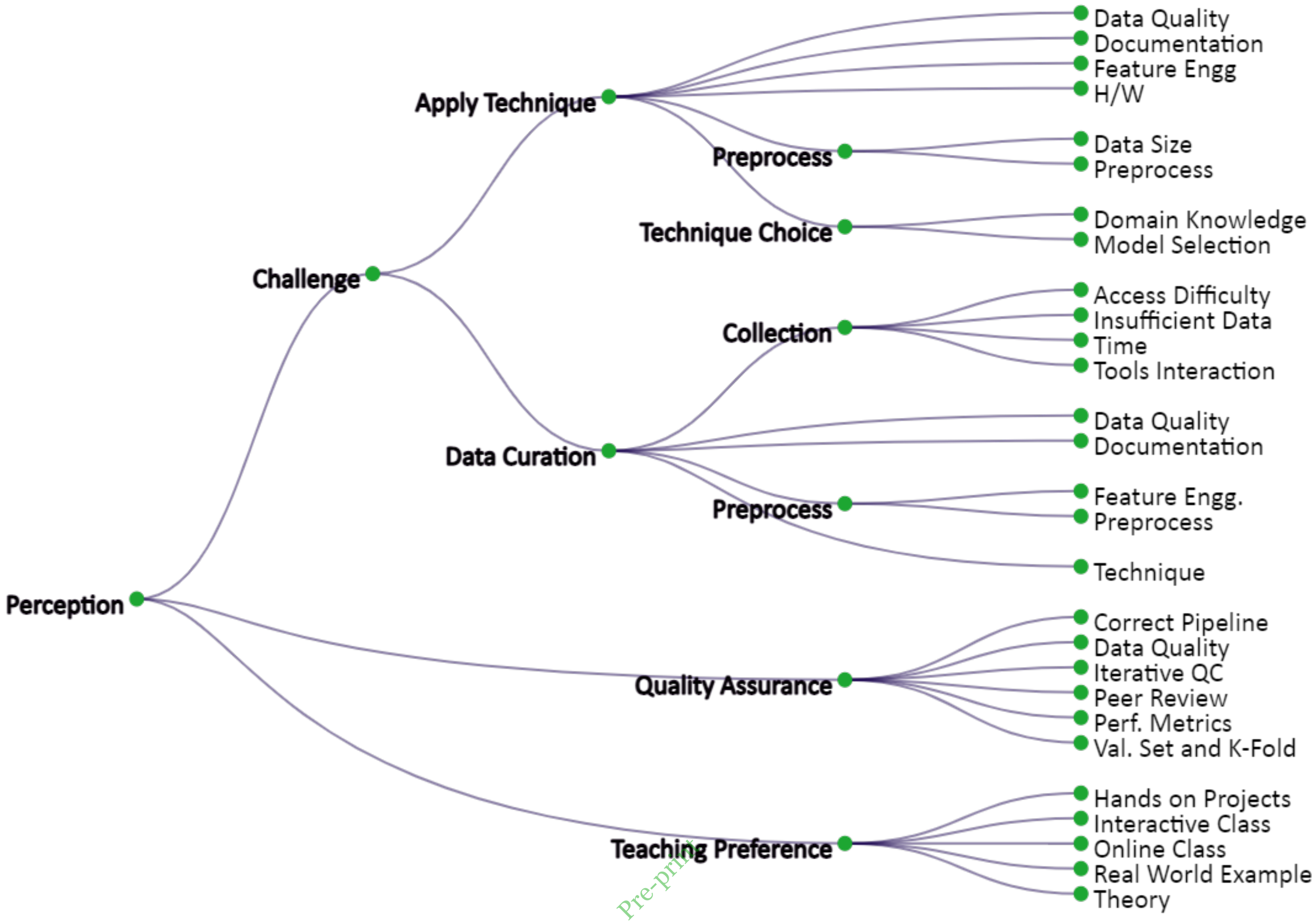


Fig. 3: A summarized overview of students' perceptions in ML.

- **Data Quality.** This category includes challenges related to data quality while accomplishing ML tasks. A typical ML task involves a data curation stage during which input data is filtered. An example of a student's response to a data quality issue during the data curation step is: *"Accuracy of data, how much data, duplicates in data"*
- **Preprocessing and Feature Engineering.** This category includes challenges related to data preprocessing and feature engineering during the data curation stage for ML tasks. A student provided a preprocessing example stating: *"the data is not as a proper format that needed to apply ML models, ..."*
- **Lack of Documentation.** This category includes challenges related to the documentation for ML data curation tasks. An example of this category is: *"Finding relevant data with adequate documentation so that I can understand the dataset"*.
- **Selection of Adequate Techniques.** This category includes challenges related to correct technique during the data curation stage for ML tasks. One student attributed a lack of *"domain knowledge"* to describe this challenge.

*2) Technique Application Challenges:* We derive six challenges that students face when applying data mining techniques. Table II presents students' response frequency for each category.

- **Data Quality.** This category includes challenges related to handling data quality when accomplishing ML tasks. One student described the data quality challenge by stating *"missing, inconsistent, incorrect data entrys"*.
- **Choosing Appropriate Technique.** This category includes challenges related to model selection and results interpretation when accomplishing ML tasks. Students reported unfamiliarity with programming libraries and a lack of domain knowledge for applying ML techniques and interpreting results. For example, one student mentioned *"Knowing what technique works best for the data and which ones to apply and not to apply."*
- **Lack of documentation.** This category includes challenges related to documentation when accomplishing ML tasks. An example response is: *"Unfamiliarity with necessary libraries and tools, low quality documentation for third party tools"*

- **Preprocessing.** This category includes data cleaning, handling missing values, inconsistent data, and handling large datasets when accomplishing ML tasks. An example response is: *“sorting or cleaning, removing of unwanted data (emojis, url and missing values)”*.
- **Feature Engineering and Parameter Tuning.** This category includes feature selection, dimensionality reduction, and hyperparameter tuning for ML tasks. We observe such an example from this student response: *“...trying graph mining techniques to model time series data. So, it will be challenging to embed high-dimensional features into low dimensional features by preserving the most of the properties of original graph network.”*
- **Hardware and Computing Resource.** This category describes high computation requirements and challenges in handling large datasets. An example is: *"For big-data, the data preprocessing especially for the data involving larger volumes of text which requires text processing consumed a lot of time."*

The Cohen’s Kappa [5] between the first and last author for category derivation is 0.71 and 0.75 for the data collecting challenge and technique application challenge, indicating ’substantial’ agreement [14].

### *B. Answer to RQ2*

In this section, we answer **RQ2: What quality aspects do students prioritize while performing ML tasks?** Table III presents our six categories by applying open coding.



- **Performance Metrics.** This category includes measuring accuracy, precision, recall, and examining confusion metrics when accomplishing ML tasks. While describing how a student measures performance, a student stated *“Comparing with existing work in respect with accuracy, scalability, and performance."*
- **Data Quality.** This category includes ensuring the quality of the dataset when accomplishing ML tasks. For example, one student stated: *“[I was] searching for expected mistakes in the dataset and ensuring they do not occur."*;
- **Subset Iteration.** This category describes the generation, collection, and testing of subsets of data, and document changes for analysis in ML tasks. One student stated *“[they] run scripts on many subsets of data to ensure correct behavior before applying to large dataset, look for errors/strange results after doing any data transformations"*
- **Correct Pipeline.** This category includes ensuring correct pipeline implementation, following guidelines, and reviewing the state-of-the-art (SoTA) when implementing ML tasks. One student stated that *“[they] check every step of the code to validate proper output for many different inputs."*
- **Validation Set and K-Fold Performance.** This category includes performance assessment on the validation set and K-fold technique when accomplishing ML tasks. A student mentioned that *"through test and validation"* they assess performance.
- **Peer Review.** This category includes peer review and code review techniques when accomplishing ML tasks. A student mentioned that they *“ask another person on the team to double check and view the graphs or findings, and make sure that they are in line with the goal of the project.”*

The Cohen’s Kappa [5] between the first and last author is 0.70, indicating ‘substantial’ agreement [14].

### *C. Answer to RQ3*

In this section, we answer **RQ3: What teaching methods do students prioritize for ML-related topics?** We derived five categories. Table III presents categories related to RQ3 and the response count associated with each category.

- **Hands-on projects and Coding assignments.** This category mentions preferences for hands-on projects and coding assignments when learning ML topics. An example of student-reported perception is: *“Hands on example with step by step dictation for the beginners”*.
- **Real World Example.** This category corresponds to preferences for real-world dataset projects when learning ML topics. Such preference was expressed by one student’s preference of *“lecture coupled with experiments using real life data.”*
- **Theoretical Lecture.** This category corresponds to preferences for lecture materials on theoretical aspects of data science when learning ML topics. For example, one student stated *“Going over theoretical aspects first is important to understand the logic and reasoning before we dive into utilizing any techniques. Ability to engage in a hands-on project where we can perform implementation for the discussed methodology/theory in class."*
- **Interactive Class.** This category corresponds to preferences for interactive teaching methods preference when learning ML topics. An example of a student’s preference for an interactive class is *“...providing assignments and collecting feedback”*.
- **Online Class.** This category corresponds to preferences for the online teaching preference of students when learning ML topics. For example, one student stated their preference of *“MOOC with hands-on exercises"*

The agreement rate between the first and last author had Cohen’s Kappa [5] of 0.65, indicating ‘substantial’ agreement [14].

| | **Category** | Count | |
|---|---|---|---|
| | | Total | By Experience |
| Data Curation | Data Collection | 38. Time Intensive (9), Insufficient data (16) Access difficulty (10), Tools Interaction (3) | N (24), I (10), E (4) |
| | Quality | 30 | N (19), I (9), E (2) |
| | Preprocessing & Feature Engineering | 12 | N (8), I (2), E (2) |
| | Lack of Documentation | 6 | N (3), I (3), E (0) |
| | Selection of Adequate Techniques | 6 | N (4), I (2), E (0) |
| Technique Application | Choosing Appropriate Technique | 31. Model Selection (19), Domain Knowledge(12) | N (21), I (7), E (3) |
| | Data Quality | 8 | N (5), I (2), E (1) |
| | Lack of documentation | 6 | N (5), I (1), E (0) |
| | Preprocessing and Dataset Size | 16 | N (12), I (3), E (1) |
| | Feature Engineering | 7 | N (2), I (4), E (1) |
| | Hardware and Computing Resource | 12 | N (3), I (7), E (2) |

TABLE II: Categories and Frequencies: Data Curation and Technique Application. **N**: Novice $<= 1$, **I**: Intermediate 2-3, **E**: Expert 4-6 years

| | **Category** | Count | |
|---|---|---|---|
| | | Total | By Experience |
| Quality Assurance | Data Quality | 23 | N (16), I (5), E (2) |
| | Performance Metrics | 20 | N (8), I (9), E (3) |
| | Correct Pipeline | 19 | N (14), I (5), E (0) |
| | Subset Iteration | 16 | N (12), I (3), E (1) |
| | Validation Set & K-Fold | 7 | N (2), I (2), E (3) |
| | Peer Review | 5 | N (4), I (0), E (1) |
| Preferred Teaching | Hands-on projects | 59 | N (38), I (17), E (4) |
| | Theoretical Lecture | 17 | N (7), I (7), E (3) |
| | Real-world Example | 10 | N (4), I (3), E (3) |
| | Online Class | 6 | N (3), I (3), E (0) |
| | Interactive Class | 5 | N (3), I (2), E (0) |

TABLE III: Categories and frequencies: ML Quality Aspects and Preferred Teaching Methods. **N**: Novice $<= 1$, **I**: Intermediate 2-3, **E**: Expert 4-6 years

### D. Answer to RQ4

In this section, we answer our **RQ4: "What are the student preferences for ML-related considerations?"**. Figure 4 presents students' perceptions of the six considerations related to ML-related implementation tasks.

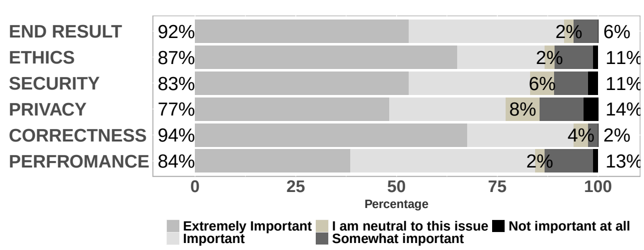


Fig. 4: Students' preferences for ML-related considerations.

The percentage value on the left-hand side of each barplot corresponds to the student percentage which responded EXTREMELY IMPORTANT and IMPORTANT. For example, END RESULT is either EXTREMELY IMPORTANT or IMPORTANT for 92% of the students.

The percentages of EXTREMELY IMPORTANT and IMPORTANT in the case of *Model's Performance* are: end result: 92%, performance: 84%, and correctness 94%. None of the participants selected NOT IMPORTANT AT ALL for end result and correctness.

For *Ethical Considerations*, we observe a relatively lower percentage who respond EXTREMELY IMPORTANT and IMPORTANT compared to *Model's Performance*. The percentages are: 87% for ethics, 83% for security, and 77% for privacy. The percentage of students who reported SOMEWHAT IMPORTANT and NOT IMPORTANT AT ALL are: ethics: 11%, security: 11%, and privacy: 14%. In contrast to *Model's Performance*, participants selected NOT IMPORTANT AT ALL for all three *Ethical Considerations*.

## IV. DISCUSSION



We discuss the findings of our experience report below.

**Data Quality-related Implications for Educators:** According to Kim et al. [9], ensuring data quality is one of the many challenges that professional data scientists face. From Table III, we also observe challenges related to data quality that students face. One implication of this finding is that while designing ML-related courses, educators should include materials and activities so that students are aware of challenges related to data quality. For example, before discussing how an ML algorithm, such as decision trees, can be applied, educators should discuss and showcase what data quality issues arise, and how such issues can be accounted for through techniques, such as removal of `NaN`s, min-max normalization [32], and z-score normalization [32]. In this manner, students pursuing ML-related courses will be better equipped to handle data quality tasks when they are professionals—an area highlighted by the Data Science Association's Code of Professional Conduct [1], which states:

- "*A data scientist shall rate the quality of evidence and disclose such rating to client to enable client to make informed decisions*"; and
- "*A data scientist shall not knowingly cherry-pick data and data science evidence.*"

**Teaching Method-related Implications:** Table III shows that 71.1% of the surveyed students (43 graduate and 16 undergraduate students) prefer hands-on projects when learning ML. The implication of this finding is that educators who teach ML-related courses should incorporate hands-on projects using established teaching approaches, such as authentic learning [12]. In other domains, such as securing configuration management [20], authentic learning experiences have shown to be effective. Authentic learning emphasizes hands-on experiences by demonstrating real-world examples [12].

**Implications related to Theoretical Foundations:** We observe a rarity of challenges related to the theoretical aspects of ML. One possible explanation is the experience of survey respondents. Perhaps, the experience of survey respondents biased the survey responses from which we identify any challenges learning the theory and algorithms of ML. Another possible explanation is the availability of ML APIs and tutorials that students can easily use in their programming environments to implement ML-based applications. Also, it is also possible that majority of the courses focused on the theoretical aspects of ML in forms of exercises and exam questions. As a result, majority of the students do not report any challenges.

Despite the rarity of challenges related to theoretical aspects, from Table II, we observe students struggle in selecting the correct analysis techniques while using ML techniques. This particular issue can be broken down into two categories: (i) challenges in selecting the correct ML technique and (ii) lack of domain knowledge. We recommend instructors design courses focusing on theoretical ML aspects with appropriate application scenarios to address these issues. For example, while introducing decision trees, educators can describe the business cases in decision trees. When communicating theoretical concepts, educators also need to consider the usage of mathematical symbols because mathematical symbols may pose challenges as described in the RQ1 analysis. As mathematical symbols are integral to ML-related discourse, we advocate future education researchers to systematically explore methods to effectively disseminate ML-related mathematics without posing challenges to students. Prior research [35] on mathematics symbols and their correlation with learning abilities could be of interest to education researchers.

**Implications for Toolsmiths:** Researchers have documented that the community focuses on ML-based model accuracy at the expense of increased hardware resources [25]. In particular, resource consumption can aggravate in the case of deep learning-based applications [8]. Our findings in Table II also show that students face challenges in ML tasks for computing resource constraints.

Accordingly, we suggest the following recommendations for toolsmiths who develop education tools:

- Development of lightweight virtualization tools, such as Docker images that can be shared with educators; and
- Use lightweight ML models, e.g., lightweight deep learning models [10].

**Ethics, Security, and Privacy for ML: Promises and Future Directions:** Our experience report shows students perceive ethics, security, and privacy to be important considerations while accomplishing ML-related tasks. According to Figure 4, 77%, 83%, and 87% of the survey respondents respectively, consider privacy, security, and ethics as important or extremely important for ML-related tasks. Our findings show promise but also showcase room for improvement so that all students acknowledge and are aware of the importance of ethics, security, and privacy for ML-related tasks. For the integration of privacy and ethics, we recommend educators follow the guidelines of Dana et al. [26] and Saltz et al. [23].

**Large Language Models (LLM) and ML Education: Promises and Future Directions:** Tu and colleagues [33] shed light on how data science and ML education will look like in the era of LLMs, such as ChatGPT by OpenAI [16]. We recommend the following adaptations in ML education:

- LLMs, such as ChatGPT can perform different stages of the data science pipeline, including data cleaning, exploration, and choosing appropriate ML models for the tasks in hand [33] – that students reported as challenging tasks in our survey. We advocate for empirical evaluation of how LLMs can be used in traditional ML teaching.
- We recommend educators adopt LLM-specific topics in ML teaching and learning. Educators can design class projects with the requirements to explore LLMs' capabilities. From students' perspective, LLMs can act as teaching assistants for ML education. Our recommendations for educators to use LLMs are in line with Tu et al.: make students aware of the potential of using LLMs including its limitations, plagiarism, and bias.

## V. Related Work

We organize this section by discussing prior research related to ML-based education and softwqare quality assurance.

### A. Research Related to ML-based Education

Researchers have investigated ML-based education approaches for different age groups including K-12, high school, and college levels. Our paper is closely related to existing research on ML-based education approaches.

Sanusi et al. identified eight pedagogical approaches for teaching ML K-12, recommending student-centric approaches like active learning and design-oriented learning [24]. Norouzi and colleagues conducted a one-month course on ML and natural language processing for high school students. [15]. They report revisiting programming concepts in group projects is necessary to ensure effective learning.

At the college level, Sulmont's study reveals that non-CS and non-statistics major students face barriers in learning machine learning (ML) due to a lack of math and programming prerequisites, with students reporting math, particularly linear algebra and probability, more frequently than programming [31]. Instructors used visualizations, real-world examples, and domain-specific applications (e.g., combining psychology with computation) to clarify ML concepts. Our findings of students' perception resemble Sulmont et al., in terms of preferred teaching methodology. Skripchuk et al. identified common ML errors in 19 term projects by 63 upper-undergraduate and graduate students, primarily involving library usage, hyperparameter tuning, and test data mishandling [29]. Lau and colleagues studied teaching methodology differences between computer science and statistics instructors in college-level data science courses [11]. They found that statistics instructors focus on 'why', while computer science teachers focus on 'how', such as fitting regression with accuracy or explaining prediction by a confidence interval. Orchard and Radke collected 53 undergraduate engineering students' surveys on a hypothetical ML ethics scenario on facial recognition [17]. They found only 17% of participants identified plausible negative implications of the system. While prior research has investigated the effectiveness of teaching methods for ML-related education, we observe a lack of research that synthesizes students' perceptions of ML-related education. We address this gap in this paper by focusing on students' perspectives in undergraduate and graduate-level data science courses.

### B. Research Related to Quality Assurance

Our paper is related with prior publications that have investigated teaching frameworks to educate students on topics related to software quality assurance. Valle et al. [7] found game-based learning to be helpful for learning software testing. Aniche et al. [2] investigated how the 'pragmatic' technique can be helpful for instructors to teach software testing by analyzing feedback reports and survey responses. Cybersecurity has garnered interest amongst researchers as well. Lukowiak et al. [13] found incremental presentation of lecture materials to aid students to learn about cybersecurity. Rahman et al. [20] used a hands-on exercise to educate students on secure configuration management. In other papers, Rahman et al. in separate publications synthesized perceptions related to static source code analysis [18], [19] and white-box testing [21].



## VI. Threats to Validity

We discuss the limitations of our paper as follows:

- Conclusion Validity: Our derived categories related to perceptions and challenges is susceptible to rater bias, as all categories are derived by the first author. We mitigate this limitation by assigning another rater, i.e., the last author of the paper who performed rater verification. The challenges that are being reported are biased by our sample of survey participants, As such, the derived categories is susceptible to bias. We mitigate this limitation by conducting the survey across five institutions. Also, the count of survey respondents is 83, which is not reflective of all students and may bias the obtained results.
- External Validity: We acknowledge our empirical results may not generalize to other students who are enrolled in ML-related courses in other universities. We mitigate this limitation by deploying the survey across four institutions across the US.
- Construct Validity: Our empirical findings are susceptible to construct validity as the survey respondents may have implicit expectations of the survey's considerations, which in turn can influence their responses. In the case of the Likert survey, the Likert item SOMEWHAT IMPORTANT maybe perceived a s a positive statement even if we think this item to have a negative impression.

## VII. Conclusion

Due to the prevalence of ML applications, the dissemination of ML-related topics and techniques in undergraduate and graduate courses has become commonplace in academic institutions. The purpose of conducting these courses is to help students learn about ML techniques. To make these courses successful, educators should account for student perceptions so that students can benefit from these courses. To that end, we have surveyed students about their perceptions of ML-related courses. From our survey with 83 students, we observed students prefer hands-on projects when learning about ML topics. We observe students encounter challenges while pursuing ML-related courses, such as challenges related to data quality and hardware. Furthermore, students perceive crucial topics relevant to ML, such as ethics, security, and privacy to be important, which is promising. Based on our findings, we recommend educators incorporate authentic learning-based hands-on projects and disseminate the theoretical foundations of ML. As LLMs continue to reshape the field of data science, we also discussed how educators can embrace the potential LLMs into their pedagogy while addressing LLMs' limitations.

## Acknowledgments

We thank the PASER group at Auburn University for their valuable feedback. This research was partially funded by the U.S. National Science Foundation (NSF) Award # 2247141, Award # 2310179, Award # 2312321, and the U.S. National Security Agency (NSA) Award # H98230-21-1-0175. This work has benefitted from Dagstuhl Seminar 23181 "Empirical Evaluation of Secure Development Processes."

## References

[1] "Data science code of conduct," Data Science Association. http://www.datascienceassn.org/code-of-conduct.html, 2022, online: Accesses on Sept 4, 2022. [Online]. Available: http://www.datascienceassn.org/code-of-conduct.html

[2] M. Aniche, F. Hermans, and A. Van Deursen, "Pragmatic software testing education," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 2019, pp. 414–420.

[3] S. Buchheit, A. B. Collins, D. L. Collins *et al.*, "Intra-institutional factors that influence accounting research productivity," *Journal of Applied Business Research (JABR)*, vol. 17, no. 2, 2001.

[4] C. Classification, "Carnegie classification of institutions of higher education," https://carnegieclassifications.acenet.edu/, 2023. [Online]. Available: 10.48550/arxiv.2303.08774

[5] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[6] A. Hegde, "Ai in education market revenue to cross $20b by 2027," June 2021, online: Accessed on Sept 7, 2022. [Online]. Available: https://www.prnewswire.com/news-releases/ai-in-education-market-revenue-to-cross-20b-by-2027-global-market-insights-inc-301318889.html

[7] P. Henrique Dias Valle, A. M. Toda, E. F. Barbosa, and J. C. Maldonado, "Educational games: A contribution to software testing education," in *2017 IEEE Frontiers in Education Conference (FIE)*, 2017, pp. 1–8.

[8] J. Hestness, N. Ardalani, and G. Diamos, "Beyond human-level accuracy: Computational challenges in deep learning," in *Proceedings of the 24th Symposium on Principles and Practice of Parallel Programming*, 2019, pp. 1–14.

[9] M. Kim, T. Zimmermann, R. DeLine, and A. Begel, "Data scientists in software teams: State of the art and challenges," *IEEE Transactions on Software Engineering*, vol. 44, no. 11, pp. 1024–1038, 2018.

[10] X. Kong and Z. Ge, "Deep pls: A lightweight deep learning model for interpretable and efficient data analytics," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2022.

[11] S. Lau, D. Nolan, J. Gonzalez, and P. J. Guo, "How computer science and statistics instructors approach data science pedagogy differently: Three case studies," in *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1*, 2022, pp. 29–35.

[12] M. M. Lombardi and D. G. Oblinger, "Authentic learning for the 21st century: An overview," *Educause learning initiative*, vol. 1, no. 2007, pp. 1–12, 2007.

[13] M. Lukowiak, S. Radziszowski, J. Vallino, and C. Wood, "Cybersecurity education: Bridging the gap between hardware and software domains," *ACM Trans. Comput. Educ.*, vol. 14, no. 1, Mar. 2014. [Online]. Available: https://doi.org/10.1145/2538029

[14] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[15] N. Norouzi, S. Chaturvedi, and M. Rutledge, "Lessons learned from teaching machine learning and natural language processing to high school students," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 09, 2020, pp. 13 397–13 403.

[16] OpenAI, "Gpt-4 technical report," arXiv (2023). DOI: 10.48550/arxiv.2303.08774, 2023. [Online]. Available: 10.48550/arxiv.2303.08774

[17] A. Orchard and D. Radke, "An analysis of engineering students' responses to an ai ethics scenario," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 03, 2023, pp. 15 834–15 842.

[18] A. Rahman, S. Hossain, and D. B. Bose, "Exercise perceptions: Experience report from a secure software development course," in *Quality of Information and Communications Technology*, A. C. R. Paiva, A. R. Cavalli, P. Ventura Martins, and R. Pérez-Castillo, Eds. Cham: Springer International Publishing, 2021, pp. 521–535.

[19] A. Rahman, H. Shahriar, and D. B. Bose, "How do students feel about automated security static analysis exercises?" in *2021 IEEE Frontiers in Education Conference (FIE)*, 2021, pp. 1–4.

[20] A. Rahman, S. I. Shamim, H. Shahriar, and F. Wu, "Can we use authentic learning to educate students about secure infrastructure as code development?" in *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 2*, ser. ITiCSE '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 631. [Online]. Available: https://doi.org/10.1145/3502717.3532125

[21] A. Rahman, Y. Zhang, F. Wu, and H. Shahriar, "Student perceptions of authentic learning to learn white-box testing," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2*, ser. SIGCSE 2024. New York, NY, USA: Association for Computing Machinery, 2024. [Online]. Available: https://doi.org/10.1145/3626253.3635584

[22] J. Saldana, *The coding manual for qualitative researchers*, 2nd ed. Los Angeles, CA, USA: Sage, 2015.

[23] J. Saltz, M. Skirpan, C. Fiesler, M. Gorelick, T. Yeh, R. Heckman, N. Dewar, and N. Beard, "Integrating ethics within machine learning courses," *ACM Trans. Comput. Educ.*, vol. 19, no. 4, aug 2019. [Online]. Available: https://doi.org/10.1145/3341164

[24] I. T. Sanusi and S. S. Oyelere, "Pedagogies of machine learning in k-12 context," in *2020 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2020, pp. 1–8.

[25] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green ai," *Communications of the ACM*, vol. 63, no. 12, pp. 54–63, 2020.

[26] D. Schwieger and C. Ladwig, "Protecting privacy in big data: A layered approach for curriculum integration," *Information Systems Education Journal*, vol. 14, no. 3, p. 45, 2016.

[27] T. J. Shuell and S. L. Farber, "Students' perceptions of technology use in college courses," *Journal of Educational Computing Research*, vol. 24, no. 2, pp. 119–138, 2001.

[28] A. SIGSOFT, "Standards | empirical standards," Sep 2023, online: Accessed on Sep 09, 2023. [Online]. Available: https://sigsoft.org/EmpiricalStandards/docs/?standard=GeneralStandard

[29] J. Skripchuk, Y. Shi, and T. Price, "Identifying common errors in open-ended machine learning projects," in *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education-Volume 1*, 2022, pp. 216–222.

[30] K. Struyven, F. Dochy, and S. Janssens, "Students' perceptions about evaluation and assessment in higher education: A review," *Assessment & evaluation in higher education*, vol. 30, no. 4, pp. 325–341, 2005.

[31] E. Sulmont, E. Patitsas, and J. R. Cooperstock, "Can you teach me to machine learn?" in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 2019, pp. 948–954.

[32] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining, (First Edition)*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.

[33] X. Tu, J. Zou, W. J. Su, and L. Zhang, "What should data science education do with large language models?" *arXiv preprint arXiv:2307.02792*, 2023.

[34] J. Van Brummelen, V. Tabunshchyk, and T. Heng, ""alexa, can i program you?": Student perceptions of conversational artificial intelligence before and after programming alexa," in *Interaction Design and Children*, ser. IDC '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 305–313. [Online]. Available: https://doi.org/10.1145/3459990.3460730

[35] A. Veloo, H. N. Krishnasamy, and W. S. Wan Abdullah, "Types of student errors in mathematical symbols, graphs and problem-solving," *Asian Social Science*, vol. 11, no. 15, pp. 324–334, 2015.

[36] K. Wiggers, "Survey finds talent gap is slowing enterprise ai adoption," April 2021, online: Accessed on Jan 17, 2023. [Online]. Available: https://venturebeat.com/business/survey-finds-talent-gap-is-slowing-enterprise-ai-adoption/