# Developing Machine Learning Algorithm Literacy with Novel Plugged and Unplugged Approaches

Ruizhe Ma*
University of Massachusetts Lowell
Lowell, MA, USA
ruizhe_ma@uml.edu

Ismaila Temitayo Sanusi*
University of Eastern Finland
Joensuu, Finland
ismaila.sanusi@uef.fi
University of Massachusetts Lowell
Lowell, MA, USA
ismaila_sanusi@uml.edu

Vaishali Mahipal*
University of Massachusetts Lowell
Lowell, MA, USA
vaishali_mahipal@student.uml.edu

Joseph E. Gonzales*
University of Massachusetts Lowell
Lowell, MA, USA
Joseph_Gonzales@uml.edu

Fred G. Martin*
University of Massachusetts Lowell
Lowell, MA, USA
fred_martin@uml.edu

## ABSTRACT

Data science and machine learning should not only be research areas for scientists and researchers but should also be accessible and understandable to the general audience. Enabling students to understand the details behind the technology will support them in becoming aware consumers and encourage them to become active participants. In this paper, we present instructional materials developed for introducing students to two key machine learning algorithms: decision trees and $k$-nearest neighbors. The materials were tested in a middle school's afterschool artificial intelligence program with four participating students aged 12 to 13. A combination of hands-on activities, innovative technology, and intuitive examples facilitated student learning. With hand-drawn decision trees and penguin species classifications, students used the algorithms to solve problems and anticipate other possible applications. We present the technology used, curriculum materials developed, and classroom structure. Following the guidelines from AI4K12 and introducing foundational machine learning algorithms, we hope to foster student interest in STEM fields.

## CCS CONCEPTS

• **Social and professional topics → K-12 education**; • **Computing methodologies → Machine learning**; **Artificial intelligence**.

## KEYWORDS

data science; machine learning; plugged activities; unplugged activities; $k$-nearest neighbor; decision tree


*SIGCSE '23, March 15–18, 2023, Toronto, ON, Canada.*

ACM ISBN 978-1-4503-9431-4/23/03…$15.00
https://doi.org/10.1145/3545945.3569772

**ACM Reference Format:**
Ruizhe Ma, Ismaila Temitayo Sanusi, Vaishali Mahipal, Joseph E. Gonzales, and Fred G. Martin. 2023. Developing Machine Learning Algorithm Literacy with Novel Plugged and Unplugged Approaches. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2023), March 15–18, 2023, Toronto, ON, Canada.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3545945.3569772

## 1 INTRODUCTION

The use of data and artificial intelligence (AI) is becoming increasingly integrated into our day-to-day lives. When topics such as data science, artificial intelligence, or machine learning are discussed, they are often reserved for professionals in the technological field [37]. Meanwhile, everyday activities such as operating autonomous vacuum cleaners and lawnmowers as well as using Google, Cortana, and Siri all involve the use of interactive artificial intelligence [14, 32]. The rapid advancements in technology have resulted in an imbalance where consumers are largely unaware of what lies behind the technology, which raises some serious concerns. Privacy and safety are the most obvious issues, but the problems are much more extensive. Inequity and bias are more covert and, in many ways, more detrimental problems. The perception of AI neutrality can be used as an excuse to mislead and discriminate against uninformed consumers [8]. Furthermore, from the current standpoint, many future careers will involve the use of AI and data in some way. A lack of knowledge about AI technologies may deter many people from entering STEM or data-related careers, thereby limiting access to such careers to a small number of people. This suggests the need to disseminate AI literacy and educate students on the inner workings of technology [17, 34]. Therefore, it is imperative to educate the next generation of students about technology, allowing them to understand its potential and limitations and empowering them to create intelligent solutions [37].

Coding programs geared towards K-12 students have received much attention for a few years now [4, 36, 39]. Data science and machine learning programs in the K-12 curriculum have also gained popularity in recent years [1, 15, 29]. The effort is observed in both formal settings [25, 26, 40] and informal settings [1, 18, 24].

However, much more action is needed to fully integrate AI literacy and data science intuition into the K-12 curriculum.

The key to learning data science and machine learning at a young age is to embrace data science thinking and see technologies and algorithms as enjoyable and approachable. Most machine learning algorithms are based on intuitive concepts that can be explained to almost any age group with suitable examples. Typically, only the optimization and application of big data require advanced mathematical and programming skills. Thus, our machine learning activities are hands-on and can be taught both in formal and informal settings, as well as incorporated into a variety of data science and artificial intelligence courses. More importantly, our materials do not require any prerequisite knowledge or coding skills.

The work described in this paper was conducted as part of an eleven-session afterschool program, each session was approximately two hours. We addressed several components of data science, machine learning, AI, and ethics. We had four middle school students aged 12 to 13 participating in the AI Afterschool program. This paper focuses on the two machine learning algorithm sessions. The rest of the paper is organized as follows: In Section 2 we look at some of the existing work and the state-of-the-art activities. In Section 3 we introduce the layout of our program. Section 4 discuss the assessments. Section 5 shows the work of the student participants and Section 6 examine the activity limitations. Finally, Section 7 concludes this paper.

## 2 BACKGROUND

Data science and machine learning are traditionally taught in higher education [5, 30]. Increasingly, machine learning algorithms are being introduced to students in an attempt to democratize artificial intelligence in schools and beyond. This effort is especially necessary as the current AI development and functionalities are mostly carried out by machine learning algorithms [27]. According to Anyanwu [3], data science encompasses computer science, artificial intelligence, machine learning, and deep learning. In accordance with this, initiatives have been ongoing to specifically introduce data science and machine learning algorithm concepts in the K-12 context.

In many ways, data science thinking, machine learning thinking, and related skills are an integral part of future computational skills [9]. Therefore it is imperative that students are instructed on machine learning algorithms as well as how to manipulate data so that they develop both data and AI literacy [17]. The extant literature describes many programs and methods for teaching data science and machine learning to younger students [28, 35]. Podworny et al. [23] used data cards to explain decision trees to middle school students. Based on the nutrition values of 55 different food items, the students developed decision rules to classify the food items as being recommendable or not. Using CODAP and Jupyter Notebook, Biehler and Fleischer [6] introduced decision trees to secondary-school students by using survey data on media consumption to predict who plays online games most frequently. Priya et al. [24] proposed ML-Quest, a game to incrementally present a conceptual overview of three machine learning concepts, including $k$-nearest neighbor ($k$-NN) algorithm. Similarly, Williams et al. [38] outlined how the $k$-NN algorithm can be presented to teachers who want to introduce machine learning and AI to their students in primary and secondary school settings. These initiatives and their outcomes suggest that the data science and machine learning algorithms can be taught well without prior CS, math, or related backgrounds [12, 42].

The goal of teaching data science and machine learning in primary education and in higher education is very different. Choosing the appropriate tools, activities, and instructional methods is a challenging and critical process [11]. Inspired by prior works, we developed curriculum activities using simple, bite-sized examples to introduce machine learning algorithms, specifically decision trees and $k$-NN, along with additional data visualization activities. Pasta Land, a decision tree activity, was originally designed by Lee and Martin [16]. Students were instructed to create a decision tree with different types of pasta and then use the decision tree to classify a mystery pasta. In addition, we developed a $k$-NN activity based on the Palmer Penguin dataset [13] for demonstrating nearest neighbors when the neighbor value $k$ is equal to 1 and 3. As part of the visualization process, we used the iSENSE platform, which is known to increase students' confidence in the interpretation of scientific data [21].

The AI Afterschool program associates with two of the five AI4K12 "Big ideas" [31]; namely, *Big Idea 2: Agents maintain models/representations of the world and use them for reasoning*, and *Big Idea 3: Computers can learn from data.* In this program, we focused on middle school students for several reasons. First, at this age, most students have had some interaction with social media and AI technologies and are developing strong individual interests and opinions. Second, students' experiences at this age may affect their future educational and career paths. Third, at the middle school level, students have already learned foundational mathematical [22] and critical thinking skills [7] required for the activities in this program that introduce machine learning algorithms. The next section details the activities and gives illustrations of the exercises involved.

## 3 AI AFTERSCHOOL PROGRAM

We present in this paper an experience report on the activities of the AI Afterschool program. Our team developed the activities, presented the materials, supervised the afterschool program, and interviewed the students. In two of the sessions, we introduced two popular classification algorithms, decision tree and $k$-NN. We presented the concepts and demonstrated how each algorithm works using interactive examples and activities. Since learning occurs best when children interact with their environment and learn through experiences [10, 33], creative and intuitive instruction is more effective in K-12 settings [2, 28, 31, 41]. In the two sessions, activities are formulated using relatable objects, i.e., pasta and penguin. By doing so, it is easier to generate interest, and students can better understand the purpose of the tasks.

### 3.1 Motivation

Increasing interest in data science, machine learning, and computer science, in general, has led to increased interest in incorporating such topics into K-12 curricula. The foundations of data science and machine learning are deeply rooted in everyday life despite being popular research topics in academia and industry. In many cases, the use of structured and targeted reasoning is sufficient for the processing and analysis of smaller amounts of data. The purpose of the activities is to introduce K-12 students to foundational data science concepts and machine learning algorithms using accessible applications and unplugged activities.

## 3.2 Curriculum

The afterschool program consisted of eleven sessions in total, each session is approximately 2 hours. Day 2 was dedicated to decision trees with the Pasta Land activity [16], and Day 5 was dedicated to the $k$-nearest neighbor using the Palmer Penguin dataset [13]. The *Pasta Land-Decision Tree* activity includes a short lecture, a few examples, and an unplugged activity where students created decision trees using pasta handouts. The *k-NN-Penguin* activity involves a brief lecture, some examples, an online data exploration activity, and an unplugged activity where students applied the $k$-NN algorithm to classify penguins. We discuss the details of the two activities in the remainder of this section. The activities and curriculum are available on our GitHub repository [19].

**Pasta Land-Decision Tree:** The idea of a decision tree is straightforward. Whether explaining decision trees to elementary school children or graduate students, at the root, a decision tree is a way of asking questions to partition a group of objects or data into smaller subsets. A decision tree is a simple algorithm that is easy to understand since it is similar to the way in which a human thinks. In the absence of optimization, any order of the questions can be used to partition datasets. When we want to optimize a tree, we consider calculations such as entropy or Gini index. Aside from optimization, there are other methods of improving performance for decision trees. However, optimization is beyond our scope and does not align with the goal for this activity.

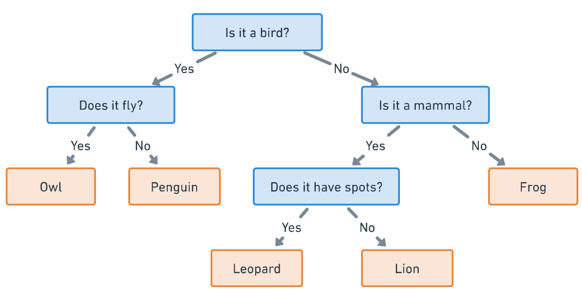



**Figure 1: 4-question game of animals.**

Our introduction begins with a smaller version of the 20 questions game. Originally appearing as a spoken parlor game in the United States during the 19th century, the 20 questions game was popularized through television in the 1940s. Today, it continues to be very popular through toys and online media such as the online game *Akinator*. The format was adapted to a smaller scale, similar to the 4-question example shown in Fig. 1. In addition, we provided examples of how to recognize popular cartoon characters through a series of questions. We also demonstrated how we could change the shape of the tree by rearranging the question order.

Through the examples of an animal-tree and a cartoon-tree, students quickly grasped the idea that by asking questions in specific orders or, in more intuitive terms, asking questions that are able to split data points into roughly similar sizes, the trees become shallower and more efficient. Students also understood that the questions are part of the classification process and that the trees are built with training data and can be used to identify new testing data. This level of understanding is sufficient to build workable decision trees.

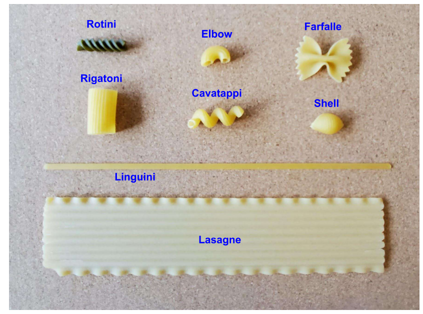



**Figure 2: Eight types of pasta handed out to students.**

The purpose of this activity is to guide students to develop their own decision trees using their own questions. The students were provided with bags of pasta, as shown in Fig. 2. Initially, all the pasta is at the root of the tree, and after the first question, pasta categories are split into two nodes in the level below. This process is repeated until each leaf node only has one pasta left. In the group discussion, we gave example questions that can discern pasta through characteristics such as *is it flat?* or *is it hollow?*. For example, when we ask the question *is it flat?*, we can split the eight types of pasta into two categories, {yes: linguini, lasagne} and {no: rotini, rigatoni, elbow, cavatappi, farfalle, shell}. Alternatively, if we ask *is it hollow?*, we can split the eight types of pasta into two categories, {yes: rigatoni, elbow, cavatappi, shell} and {no: rotini, farfalle, linguini, lasagne}. The students were then instructed to identify additional characteristics of each type of pasta in order to construct the tree-like structure displayed in Fig. 1.

Students were also given pencils and paper to draw the trees, as well as a ruler to measure the pasta if necessary. In this activity, students worked in pairs. When all groups had completed their pasta decision trees, we asked them to exchange their trees and see if the other groups were able to classify all the testing pasta successfully. All groups were successful. To conclude the activity, we provided each group with a mystery pasta as testing data, a tortellini, that was not used in the creation of the pasta decision trees. Testing data are introduced as a way to apply the model. We asked each group to classify the mystery pasta based on their decision trees and discussed if the results made sense in terms of shape. As the pasta decision trees were simple, it was easy for the models to miss some important features, resulting in high bias. With this activity, we discussed the concept of bias, which was not introduced in previous Pasta Land activities.

The discussion involving the mystery pasta focused on three questions:

- Did your tree classify the mystery pasta with a closely related pasta shape?

**Table 1: A Segment of the Palmer Penguin Dataset**

| ID | Species | Island | Bill Length (mm) | Bill Depth (mm) | Flipper Length (mm) | Body Mass (g) | Sex | Year |
|---|---|---|---|---|---|---|---|---|
| 1 | Adelie | Torgersen | 39.1 | 18.7 | 181 | 3750 | male | 2007 |
| 2 | Adelie | Torgersen | 39.5 | 17.4 | 186 | 3800 | female | 2007 |
| 3 | Adelie | Torgersen | 40.3 | 18.0 | 195 | 3250 | female | 2007 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 344 | Chinstrap | Dream | 50.2 | 18.7 | 198 | 3775 | female | 2009 |

- Does your tree capture what's important about the mystery pasta?
- If not, do you think there is bias?

The mystery pasta is the testing data that is used to determine the robustness of students' decision trees. As we only used one pasta for each pasta type, and each leaf is labeled with the pasta name, the mystery pasta would clearly not match the classes in the decision tree. However, students agreed that the mystery pasta had characteristics in common with the pasta type it was classified with based on the questions in their decision trees.

**Penguins-*k*-Nearest Neighbor:** The *k*-NN algorithm is a fundamental machine learning algorithm. It is a classification algorithm where data are assigned labels based on the majority voting by its closest neighboring data. The Palmer Penguin dataset [13] was used to introduce the concept of *k*-NN. A total of 344 penguin measurements from three different species are included in this dataset: Chinstraps, Gentoos, and Adelies. A portion of the penguin dataset is shown in Table 1. There are nine columns of information, which include: ID, species, island, bill length, bill depth, flipper length, body mass, sex, and year. To illustrate the relationship between the various measurements, we showed short videos of the different penguin species and brought plush penguin toys into class.

The students showed enthusiasm for penguins, and some were able to recognize different species' pictures. We discussed why we need algorithms to classify penguins when we can observe them directly. The first problem is that not everyone can identify and recall all the characteristics of different species, and it may be necessary to identify hundreds of penguins. It is also more time-consuming and complex to identify through pictures. Using a set of measurements for each animal, we would be able to automate the classification process with limited human involvement.

We wanted the students to acquire an overall understanding of the dataset and demonstrate how data visualization can be used to gain insight into the data itself. We utilized the iSENSE platform [20], on which the Palmer Penguin dataset has already been uploaded. We demonstrated the tabular format of the data, similar to Table 1, and how we can gather statistics on the dataset through the platform. We discussed differences between species, such as average bill depth or average flipper length, then showed how each penguin can be plotted as a point on a two-dimensional plot based on two measurements. Fig. 3 shows the scatter plot of 344 penguins based on their bill depth and body mass. The three colors, red, yellow, and blue, correspond to the three different species. Here, we deliberately chose characteristics where the three species did not separate well and asked students to plot different characteristics against each other in order to identify the characteristics that did separate the three species well.


**Figure 3: An example scatter plot of the Palmer Penguin dataset. Here different species are color-coded, bill depth is the y-axis, and body mass is the x-axis.**

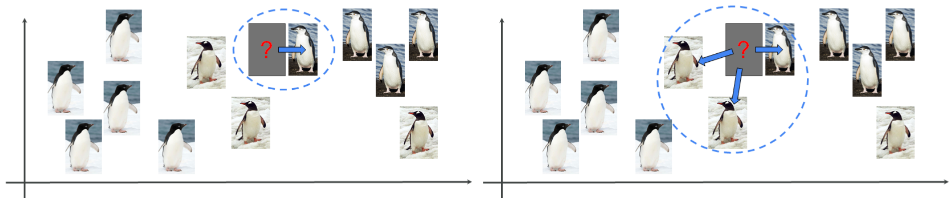
After exploring the Palmer Penguin dataset thoroughly, all students were able to identify the distinct characteristics of different penguin species. We then discussed the *k*-NN algorithm in more detail. Unlike the Pasta Land decision trees, *k*-NN does not construct a model, which is why it is known as a lazy learner. Therefore, no action is taken until new data must be classified. Based on the scatter plot from the previous activity, students were asked to identify the species of the mystery penguin, provided we only had its measurements. The next step was to discuss the reasons for *k*-NN's success. As illustrated in Fig. 4, the students are shown how choosing a different number of *k* could influence the classification result. Our discussion of the *k*-NN algorithm is limited to the Euclidean distance to be at an age-appropriate level. Lastly, students were given worksheets, where they were asked to classify the mystery penguins in accordance with two different *k* values in *k*-NN.

## 4 ASSESSMENT

Working with the four participants, we assessed the activities from three perspectives: (1) students were observed during the activities; (2) surveyed on the day of the activities; and (3) upon completion of the program, the students were asked to use what they learned to implement a small-scale project. The research team was present to teach the material, supervise and assist with the activities, as well as conducting the surveys.

For the Pasta Land-Decision Tree activity, the following survey questions were used:

- Was your tree more balanced or lop-sided?
- If your tree was balanced - why do you think this was so?
- If your tree was lop-sided - why do you think this was so?



(a) 1 Nearest Neighbor

(b) 3 Nearest Neighbor

**Figure 4: An example of how the $k$-NN algorithm assign class labels. (a) When $k = 1$ the new data is classified with the single nearest neighbor. (b) When $k = 3$ the new data is classified with the majority vote of the 3 nearest neighbors.**

- How did your tree handle the mystery pasta - do you think grouped it with other pasta shapes properly?

For the Palmer Penguin-$k$-NN activity, the following survey questions were used:

- Did you find out anything interesting from visualizing the data on iSENSE?
- Were you able to identify the species of the penguin in the $k$-NN activity?
- Where else do you think $k$-NN can be applied to?
- What did you find interesting about this activity?

In the last two sessions of the afterschool program, students developed datasets on topics of their choice and revisited the two machine learning algorithms introduced on Days 2 and 5. Each student created a scatter plot of their own data using the iSENSE platform. Using the data that they collected, students then constructed decision trees. By selecting questions to partition the data, the students demonstrated their understanding of how decision trees are constructed and how they can be used to classify new data. In addition, students also discussed how the $k$-NN algorithm might be utilized to classify new testing data.

## 5 RESULTS

In the Pasta Land activity, students were able to recognize the connection between asking questions, partitioning a dataset, and creating a decision tree. As a result of the decision tree they created, students were also able to identify new data correctly.

Since students worked in pairs for the Pasta Land activity, there was a great deal of discussion on what questions to ask to partition the pasta dataset. Fig. 5 illustrates three decision trees created. In Fig. 5 (a), the work does not have a tree structure; the group brainstormed different questions separately. If the segments of the tree are connected to the results from the layer before it, then it can be arranged into a true tree structure. In Fig. 5 (b), one of the branches was not finished, but all other leaf nodes have only one type of pasta. Lastly, in In Fig. 5 (c), the group created a legend for the pasta names and built the entire decision tree.

In the subsequent interview, students agreed that *the decision tree did a decent job in classifying the mystery pasta*. Two students commented on the difficulty of finding good questions. One stating that they *can't find enough questions for flat pasta*, and *it is hard to split the pasta into two equal groups*. Another student commented *some other dataset that is easier to ask questions on would be better.*

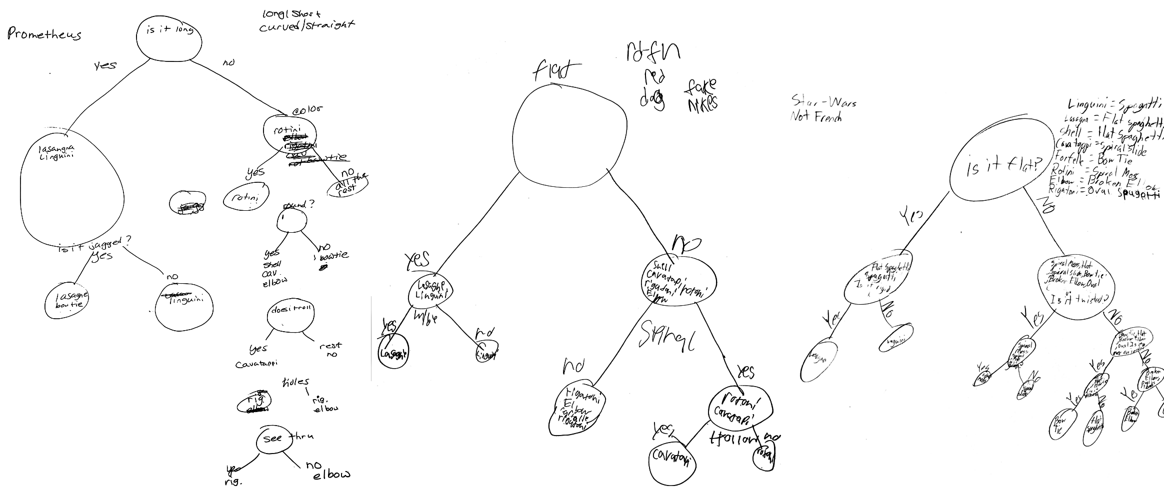
The $k$-NN activity was completed by each student individually. As mentioned in Section 3.2, the students first explored the penguin dataset on the iSENSE platform. This exploration includes finding the average, mean, and median of the Bill Length, Bill Depth, Flipper Length, and Body Mass, as well as creating various scatter plots of different pairs of measurements to identify how to differentiate between the three species. Furthermore, some students chose to color-code the data points according to the island, and the sex of the penguins in their scatter plots.

Following this, students engaged in the unplugged activity for $k$-NN. Fig. 6 show the three worksheets completed by three individuals. In Fig. 6 (a), the three votes for $k$-NN where k=3 are written. In Fig. 6 (b), the blue penguin where k=1 and the red penguin where k=3 are misclassified. In Fig. 6 (c) all penguins are correctly classified.

In the subsequent interview, one student found it *cool to use k-NN to find a penguin*. Some students preferred data exploration with iSENSE, stating that they learned visualization methods they didn't know before. And some students preferred the worksheet activity, stating *I like the k-NN on paper part the most.* Students were also able to brainstorm other application areas for $k$-NN, such as *symptom of diseased by comparing them*, then we can use it to differentiate *coronavirus, flu, and chickenpox.*

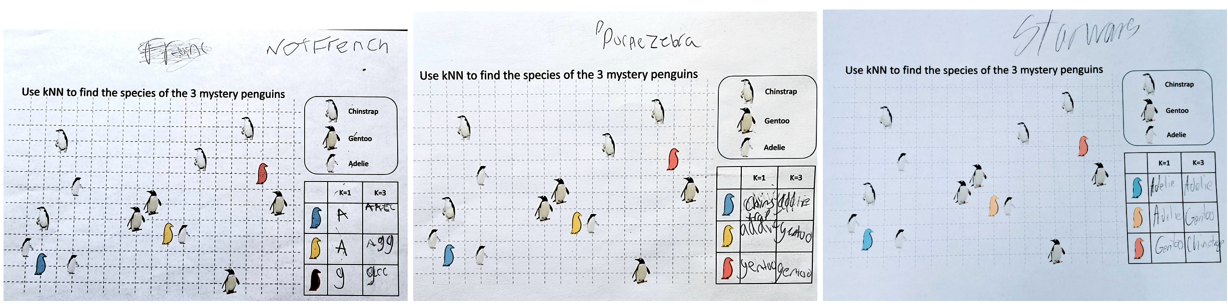
## 6 LIMITATIONS AND FUTURE DIRECTIONS

This paper is a pilot study and serves as an experience report. Our goal for the program is to see if middle school students can understand typical machine learning algorithms with the right activities. Next, we will increase the sample size so we can observe the effectiveness of our activities and further improve them. With a larger sample size, we would also be able to perform quantitative assessments. We aim to reach a wide range of students, which would require specific materials, depth, and varied delivery methods. Specifically, we intend to create a tier-structured program for all student age groups. Our conception is as follows. At the elementary level, the goal should be to develop data science thinking and to gain a better understanding of some fundamental machine learning algorithms. For middle school students, the objective should be to gain a deeper understanding of how some machine learning algorithms work, as well as to become aware of common errors and issues that may occur. At the high school level, students can systematically calculate the results to optimize algorithms on a small scale.




(a) Decision Tree 1 (b) Decision Tree 2 (c) Decision Tree 3

**Figure 5: Pasta Land decision trees created by students.**




(a) *k*-NN Worksheet 1 (b) *k*-NN Worksheet 2 (c) *k*-NN Worksheet 3

**Figure 6: Classifying penguins with the *k*-NN algorithm worksheets by students.**

For example, in the Pasta Land-Decision Tree activity, at the elementary level, students can understand that decision trees are a way to use questions to split a dataset, which can then be used to classify new data. In middle school, students can learn how to arrange the questions purposefully so that the decision tree is more efficient, as well as what can happen when some data records are missing or inaccurate. At the high school level, students can learn to calculate entropy and information gain to create and optimize decision trees systematically. For Penguins-*k*-NN activity, we can introduce 1-nearest neighbor for beginners, then advance to multiple *k* neighbors, and finally, discuss topics such as decision boundary and complex data types such as sequential data.

## 7 CONCLUSION

In this pilot study, we investigated the introduction of two popular machine learning algorithms to middle school students: the decision tree algorithm and the *k*-nearest neighbor algorithm. Students' engagement appears to be related to the subject of the data. One student was able to name and distinguish multiple penguin species before the program and was enthusiastic about helping introduce the various species to the class. Students demonstrated their understanding by building models on the training data presented, as well as using the models with testing data.

The work presented here serves as a foundation for future data science and machine learning programs. The next steps include increasing the student sample size, performing quantitative assessments, and revising the curriculum. By collecting additional data, we will be able to calibrate our curriculum further. We plan to expand our in-person program and create an online platform to reach a broader audience. The curriculum will also be tier-structured to provide age-appropriate content.

## ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2112633. We would like to thank all the students, teachers, and facilitators who made it possible for this afterschool program. We are grateful to all the reviewers for their reviews and valuable comments.

# REFERENCES


[1] Ashraf Alam. 2022. A Digital Game based Learning Approach for Effective Curriculum Transaction for Teaching-Learning of Artificial Intelligence and Machine Learning. In *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*. IEEE, 69–74.

[2] Safinah Ali, Blakeley H Payne, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2019. Constructionism, ethics, and creativity: Developing primary and middle school artificial intelligence education. In *International workshop on education in artificial intelligence k-12 (eduai'19)*. 1–4.

[3] Kele Anyanwu. 2021. Making Sense of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) in 21st Century K-12 Classrooms.. In *Society for Information Technology & Teacher Education International Conference*. Association for the Advancement of Computing in Education (AACE), 1140–1145.

[4] Michal Armoni. 2016. COMPUTING IN SCHOOLS Computer science, computational thinking, programming, coding: the anomalies of transitivity in K-12 computer science education. *ACM Inroads* 7, 4 (2016), 24–27.

[5] Ben Baumer. 2015. A data science course for undergraduates: Thinking with data. *The American Statistician* 69, 4 (2015), 334–342.

[6] Rolf Biehler and Yannik Fleischer. 2021. Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics* 43 (2021), S133–S142.

[7] Bengi Birgili. 2015. Creative and critical thinking skills in problem-based learning environments. *Journal of Gifted Education and Creativity* 2, 2 (2015), 71–80.

[8] Joy Adowaa Buolamwini. 2017. *Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[9] Peter J Denning and Matti Tedre. 2019. *Computational thinking*. Mit Press.

[10] John Dewey. 1986. Experience and education. In *The educational forum*, Vol. 50. Taylor & Francis, 241–252.

[11] Varvara Garneli, Michail N Giannakos, and Konstantinos Chorianopoulos. 2015. Computing education in K-12 schools: A review of the literature. In *2015 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 543–551.

[12] Tom Hitron, Yoav Orlev, Iddo Wald, Ariel Shamir, Hadas Erel, and Oren Zuckerman. 2019. Can children understand machine learning concepts? The effect of uncovering black boxes. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–11.

[13] Allison Marie Horst, Alison Presmanes Hill, and Kristen B Gorman. 2020. *palmerpenguins: Palmer Archipelago (Antarctica) penguin data*. https://allisonhorst.github.io/palmerpenguins/ R package version 0.1.0.

[14] Martin Kandlhofer, Gerald Steinbauer, Sabine Hirschmugl-Gaisch, and Petra Huber. 2016. Artificial intelligence and computer science in education: From kindergarten to university. In *2016 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–9.

[15] Dale Lane. 2022. *Machine Learning for Kids*. https://machinelearningforkids.co.uk

[16] I Lee and F. Martin. 2018. Investigating Bias in Machine Learning Applications. In *CSTA Conference*.

[17] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–16.

[18] Duri Long, Anthony Teachey, and Brian Magerko. 2022. Family Learning Talk in AI Literacy Learning Activities. In *CHI Conference on Human Factors in Computing Systems*. 1–20.

[19] Ruizhe Ma, Ismaila Temitayo Sanusi, Vaishali Mahipal, Joseph E. Gonzales, and Fred G. Martin. 2022. *AIML-Curriculum-Spring2021*. https://github.com/RuizheMa/AIML-Curriculum-Spring2021

[20] Fred Martin, Sarah Kuhn, Michelle Scribner-MacLean, Christopher Corcoran, James Dalphond, John Fertitta, Michael McGuiness, Sam Christy, and Ivan Rudnicki. 2010. iSENSE: A web environment and hardware platform for data sharing and citizen science. In *2010 AAAI Spring Symposium Series*.

[21] Samantha Michalka, James Dalphond, and Fred Martin. 2016. Inquiry learning with data and visualization in the STEM classroom. In *Society for Information Technology & Teacher Education International Conference*. Association for the Advancement of Computing in Education (AACE), 2204–2211.

[22] David G Moursund. 2006. *Computational thinking and math maturity: Improving math education in K-8 schools*. D. Moursund.

[23] Susanne Podworny, Yannik Fleischer, Sven Hüsing, Rolf Biehler, Daniel Frischemeier, Lukas Höper, and Carsten Schulte. 2021. Using data cards for teaching data based decision trees in middle school. In *21st Koli Calling International Conference on Computing Education Research*. 1–3.

[24] Shruti Priya, Shubhankar Bhadra, Sridhar Chimalakonda, and Akhila Sri Manasa Venigalla. 2022. ML-Quest: a game for introducing machine learning concepts to K-12 students. *Interactive Learning Environments* (2022), 1–16.

[25] Ismaila Temitayo Sanusi and Sunday Adewale Olaleye. 2022. An Insight into Cultural Competence and Ethics in K-12 Artificial Intelligence Education. In *2022 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 790–794.

[26] Ismaila Temitayo Sanusi, Sunday Adewale Olaleye, Friday Joseph Agbo, and Thomas KF Chiu. 2022. The role of learners' competencies in artificial intelligence education. *Computers and Education: Artificial Intelligence* 3 (2022), 100098.

[27] Ismaila Temitayo Sanusi, Sunday Adewale Olaleye, Solomon Sunday Oyelere, and Raymond A Dixon. 2022. Investigating learners' competencies for artificial intelligence education in an African K-12 setting. *Computers and Education Open* 3 (2022), 100083.

[28] Shashank Srikant and Varun Aggarwal. 2017. Introducing data science to school kids. In *Proceedings of the 2017 ACM SIGCSE technical symposium on computer science education*. 561–566.

[29] Jiahong Su, Yuchun Zhong, and Davy Tsz Kit Ng. 2022. A meta-review of literature on educational approaches for teaching AI at the K-12 levels in the Asia-Pacific region. *Computers and Education: Artificial Intelligence* (2022), 100065.

[30] Lisa Torrey. 2012. Teaching problem-solving in algorithms and AI. In *3rd Symposium on Educational Advances in Artificial Intelligence*.

[31] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. Envisioning AI for K-12: What should every child know about AI?. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 9795–9799.

[32] David S Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. K-12 guidelines for artificial intelligence: what students should know. In *Proc. of the ISTE Conference*, Vol. 53.

[33] Emel Ültanir. 2012. An epistemological glance at the constructivist approach: Constructivist learning in Dewey, Piaget, and Montessori. *International journal of instruction* 5, 2 (2012).

[34] Jessica Van Brummelen, Tommy Heng, and Viktoriya Tabunshchyk. 2021. Teaching tech to talk: K-12 conversational artificial intelligence literacy curriculum and development tools. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 15655–15663.

[35] Henriikka Vartiainen, Matti Tedre, and Teemu Valtonen. 2020. Learning machine learning with very young children: Who is teaching whom? *International journal of child-computer interaction* 25 (2020), 100182.

[36] F Vico, M Molina, D Orden, J Ortiz, R Garcia, and J Masa. 2019. A coding curriculum for K-12 education: the evidence-based approach. In *Proceedings of the 11th annual International Conference on Education and New Learning Technologies*. 7102–7106.

[37] Christiane Gresse von Wangenheim, Lívia S Marques, and Jean CR Hauck. 2020. Machine learning for all–Introducing machine learning in K-12. (2020).

[38] Randi Williams, Stephen P Kaputsos, and Cynthia Breazeal. 2021. Teacher Perspectives on How To Train Your Robot: A Middle School AI and Ethics Curriculum. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 15678–15686.

[39] Gary KW Wong, Ho-Yin Cheung, Edwin CC Ching, and John MH Huen. 2015. School perceptions of coding education in K-12: A large scale quantitative study to inform innovative practices. In *2015 IEEE international conference on teaching, Assessment, and learning for engineering (TALE)*. IEEE, 5–10.

[40] Qi Xia, Thomas KF Chiu, Min Lee, Ismaila Temitayo Sanusi, Yun Dai, and Ching Sing Chai. 2022. A self-determination theory (SDT) design approach for inclusive and diverse artificial intelligence (AI) education. *Computers & Education* 189 (2022), 104582.

[41] Xiaofei Zhou, Jessica Van Brummelen, and Phoebe Lin. 2020. Designing AI learning experiences for K-12: emerging works, future opportunities and a design framework. *arXiv preprint arXiv:2009.10228* (2020).

[42] Abigail Zimmermann-Niefield, R Benjamin Shapiro, and Shaun Kane. 2019. Sports and machine learning: How young people can use data from their own bodies to learn about machine learning. *XRDS: Crossroads, The ACM Magazine for Students* 25, 4 (2019), 44–49.