Empower Secondary School Teachers to Create ML-Supported Inquiry-Based Learning Activities

Xiaofei Zhou[1,3], Hanjia Lyu[1], Yuxin Sa[1], Mengfan Li[1], Advait Sarkar[2], Jiebo Luo[1], Michael Daley[1], and Zhen Bai[1]

[1] University of Rochester, Rochester NY 14627, USA
{xzhou50,zbai7}@ur.rochester.edu, {hlyu5, ysa}@u.rochester.edu, mli82@u.rochester.edu, jluo@cs.rochester.edu, mdaley@warner.rochester.edu
[2] Microsoft Research, Cambridge, UK advait@microsoft.com
[3] Shanghai Jiao Tong University

Abstract. The rapid advancement of Artificial Intelligence (AI) necessitates preparing the next generation to be AI- and data-literate citizens and problem solvers. Despite efforts to integrate AI into K-12 education, many teachers lack the expertise to create meaningful AI learning experiences. We developed ML4Inq, an innovative online authoring tool that enables secondary school teachers to create data-driven inquiry-based learning (IBL) activities with machine learning (ML) technologies such as clustering and classification. Findings of a co-design workshop with 14 secondary school teachers show that teachers successfully created a diverse range of curriculum-aligned activities using ML4Inq, in collaboration with ML experts. We discovered key connections between ML-revealed patterns and IBL behaviors, offering design implications for future tools that support ML practices in secondary schools.

Keywords: Authoring Tool · Inquiry-Based Learning · AI Literacy.

1 Introduction

In the era of Big Data and Artificial Intelligence (AI), cultivating data and computational literacy has become critical to prepare the next generation for a data-centric world. Machine learning (ML) techniques like clustering and classification uncover patterns in large datasets, thereby accelerating data-driven scientific inquiry [17]. In K-12 education, inquiry-based learning [29] encourages students to adopt practices akin to professional scientists for knowledge construction [23]. Initiatives like AI4K12 [34], CSTA [8], and AI4ALL [1] have established national guidelines related to applying ML for data analysis in K-12. These range from teaching K-2 students how computers learn from examples to engaging 9th–12th graders to train models using supervised or unsupervised learning algorithms on real-world data and evaluate the results.

However, effective implementation of ML-supported inquiry-based learning (ML-IBL) in classrooms presents significant challenges, particularly for teachers.

The complexity of ML concepts, multidimensional data, and statistical reasoning creates high knowledge barriers, often overwhelming novices before they can effectively use ML to process data, discover patterns, and derive meaningful insights [25]. This complexity, combined with students' lack of ML literacy, highlights the need for pedagogically sound ML practices that captivate students' interests [33]. However, there remains a critical gap in teaching resources that support the integration of customized ML practices into IBL, ensuring alignment with K-12 curriculum standards while addressing teachers' needs and interests [33]. Furthermore, over 60% of pre-service math teachers feel underprepared to teach statistics [24], let alone the integration of ML, which requires additional statistical expertise [25].

Our work presents ML4Inq, a web-based authoring tool [40] with low entry barriers for secondary teachers to create ML-IBL activities. ML4Inq enables teachers to explore data, adjust parameters for K-means clustering and K-nearest neighbors (KNN), interpret ML-revealed patterns, and link with IBL activities for their classrooms without requiring extensive programming or statistics expertise. Our main research question is how secondary school teachers incorporate ML into curriculum-aligned IBL activities using ML4Inq.

We invited 14 secondary school teachers to use datasets of their teaching interests and create ML-IBL activities, by co-designing with ML experts who ensured the validity of ML usage. We collected and analyzed 26 ML-IBL activities with K-12 topics ranging from biology to world history. Teachers applied ML-revealed patterns to promote various desired IBL behaviors for students, such as questioning and hypothesis formation [29]. For example, they first developed hypotheses through cluster analysis, used KNN to retrieve the nearest neighbors as specific data groups, and analyzed patterns within and between these groups to test existing hypotheses. Incorrect predictions were used as opportunities to prompt students to explore potential causes, such as low correlations between the target variable and the selected independent variables. While this study references U.S. educational standards such as NGSS and the Common Core State Standards (CCSS) [15], the findings and tools could be adaptable for integration into other national and regional educational frameworks. In summary, our main contributions include:

1. ML4Inq, a novel authoring tool for teachers to create ML-IBL activities aligned with the curriculum;
2. A key set of connections between ML-revealed patterns and IBL;
3. Design implications of tools for supporting ML practices in secondary school.

2 Data Analysis and IBL in Secondary Education

Recent advancements in statistics education underscore exploring, modeling, and understanding the computational aspects of data [13]. This approach seeks to enhance students' engagement with data and prepare them for more advanced topics, such as inferential statistics and ML. However, complex statistical techniques may be inaccessible to younger students.

To address this challenge, researchers advocate for making computation more accessible through different programming modalities [36], computational notebooks [4], etc. For example, CODAP, an online data analysis platform with embedded binary data visualization, has shown promise in engaging secondary school students in data exploration and analysis [16]. Inq-ITS provides real-time support for inquiry-based science learning [19]. Computational notebooks and block-based programming platforms apply cell grouping and labeling features [7] to improve data sensemaking.

Despite these tool advancements, teachers tend to have low confidence in teaching statistics [24] or ML-related content [33]. Research also highlights the practical need for effective instructional resources to contextualize data practices into STEM and non-STEM curricula [32]. However, a gap exists in supporting teachers to author authentic ML-IBL activities for their classrooms [21]. Studies have identified obstacles in data interpretation resulting from unexpected uses, providing insights to improve educational tools and practices [14].

Research in preparing teachers to teach ML in K-12 settings highlights the scarcity of teaching resources [25]. Professional development (PD) programs have been developed for STEM teachers to integrate computational thinking with science education using Logic Programming [37] or to conceptually create high-level lesson plans integrating basic ML methods [39]. These initiatives relied primarily on direct instruction methods, such as presentations and textbooks, which may not fully connect ML applications with classroom practice.

To address these gaps and obstacles, our design aims to integrate ML-related data practices into the curriculum through data visualization and IBL [29], a pedagogical approach familiar to teachers. We focus on three key IBL phases with direct relevance to data: conceptualization, where students propose research questions or hypotheses, investigation, where students explore, experiment, analyze, and interpret data, and conclusion, where students evaluate whether their research questions or hypotheses have been effectively addressed by their study results. Our design adheres to guidelines for teaching statistics: (1) using large, multivariate datasets, (2) maintaining contextual relevance, and (3) engaging in iterative investigation cycles [6]. Additionally, it incorporates essential capabilities of K-12 data science tools [21]: data manipulation, statistical analysis, data visualization, and dataset availability.

3 Design of ML4Inq

3.1 Design Goals

This section outlines the ML practices and IBL behaviors that ML4Inq aims to support as an authoring tool and explains the rationale behind them.

Selection of ML practices incorporates cluster analysis and predictive analysis through K-means clustering and KNN algorithms, respectively, for three reasons. **First**, with simplicity, transparency, and interpretability [10], K-means clustering groups data into a requested number of clusters, and KNN helps with

Table 1. Three categories of ML practices.

Exploration	Customization	Evaluation and Analysis
Read attributes: Examine and understand the attributes of the dataset. **Read data**: Browse the raw data that will be used for IBL. **Manual clustering**: Manually sort similar data points together.	**Feature selection**: Identify attributes of interest to build ML algorithms. **Update input**: Modify the dataset fed into the ML algorithm. **Parameter adjustment**: Fine-tune the parameters of ML algorithms.	**Automatic clustering**: Apply K-means clustering algorithm for the k value evaluation and cluster analysis. **Predict with KNN**: Apply the KNN algorithm for performance evaluation and predictive analysis.

classification or regression tasks by identifying the nearest neighbors based on distance. Both algorithms require minimal parameter tuning, allowing teachers to get started quickly without extensive knowledge of ML concepts. **Second**, these two similarity-based algorithms are particularly amenable to visualization, which provides immediate feedback that facilitates teachers to gradually adjust parameters and see visual changes in real time. **Third**, they are widely applicable across various science and social study domains [3] and are useful for diverse analysis tasks, such as segmenting data or making predictions.

Based on data science and ML life-cycle (i.e., data exploration, feature engineering, build models, evaluate models) [35], as well as key operations in K-means clustering and KNN, we define a set of ML practices (Table 1). More details are described in Section 3.2.

Selection of IBL behaviors is guided by an extensive review of established IBL frameworks [29]: **(1)** questioning (formulate investigable questions for IBL), **(2)** data analysis (analyze the data to identify patterns), **(3)** hypothesis iteration (generate and refine hypotheses based on insights gained in IBL), **(4)** pattern interpretation (make sense of collected patterns in the subject context and synthesize new knowledge), **(5)** conclusion (draw and justify inferences and conclusions).

3.2 Design Features

F1. Modular ML/SI blocks initiate pre-designed ML and IBL components (Fig. 1.a1). When teachers drag and drop them into the main workspace with a side-by-side layout (Fig. 1.a2), the corresponding ML methods or IBL behaviors are added to the learning activity.

ML components (Fig. 1 left) enable the application of ML techniques. For instance, "Build Predictor" and "Make a Prediction" (Fig. 1.b) enable predictive analysis with KNN and the evaluation of prediction accuracy; "Automatic Clustering" applies K-means clustering to analyze input data and visually represent the resulting clusters (Fig. 2); "Manual Clustering" allows users to explore patterns within a subset by manually overlaying data visualizations to compare similarities using superposition comparative visualization [18]. This enhances transparency and hands-on manipulation for students to grasp the basics of clustering, helping them understand clustering algorithms when encountered [12].

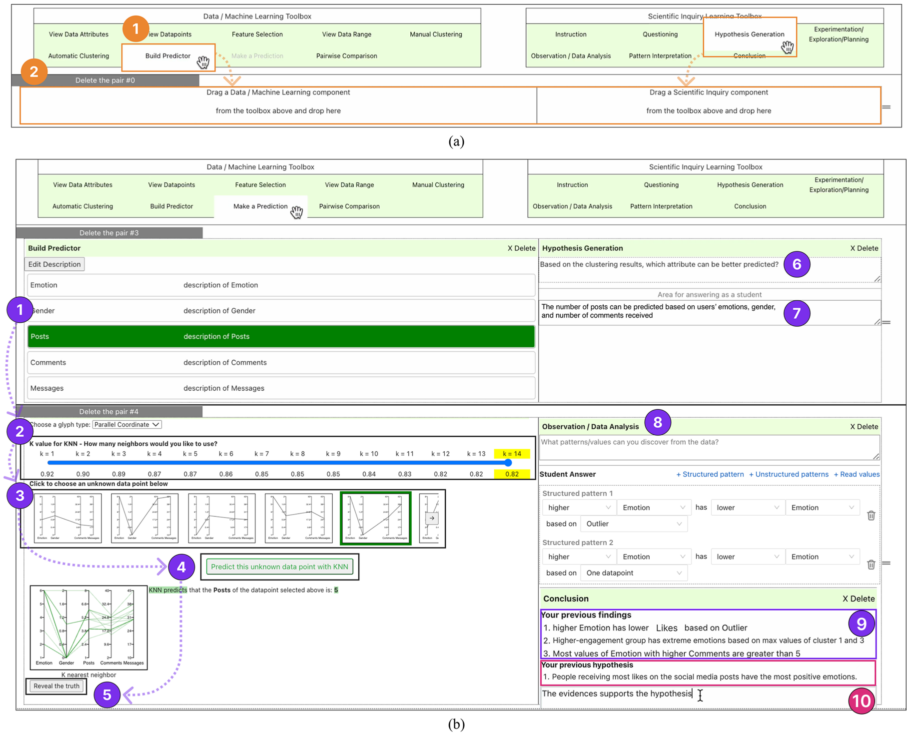

Fig. 1. (a1) Draggable blocks to initiate ML and IBL components; (a2) side-by-side layout for designing ML (e.g., build predictor) and IBL (e.g., hypothesis generation) steps; (b) a teacher-designed activity: (b1) select an attribute to predict with KNN; (b2) adjust the parameter for the number of neighbors; (b3&b4) select an unknown data point to predict, (b5) evaluate KNN output by comparing with ground truth.

SI components (Fig. 1 right) support established IBL behaviors [29] (Section 3.1). Teachers add “Questioning” or “Hypothesis Generation”, customize instructional text, and provide answers from their students’ perspective (Fig. 1.b6 & 1.b7) based on the paired ML component (Fig. 1.b left).

With “Data Analysis” (Fig. 1.b8), teachers document the ML-revealed patterns. Based on cluster analysis and predictive analysis in data science [3], seven potential patterns can be identified using K-means clustering and KNN: (1) similarities and differences between two data points, (2) intra-cluster similarities and variations, (3) inter-cluster comparisons, (4) centroid, (5) outlier analysis, (6) data range (minimum and maximum values), (7) prediction results compared to ground truth. The identification and analysis of the aforementioned patterns align with CCSS [15] for Mathematics in the U.S., particularly in reasoning abstractly and quantitatively, constructing viable arguments, and modeling with mathematics. This extends the CCSS as the high school math curriculum in the U.S. typically only covers patterns of association in bivariate data and the shapes of two- and three-dimensional data. With “Conclusion”, teachers can view a list of automatically tracked findings from "Data Analysis" (Fig. 1.b9) and compare them to their existing hypothesis (Fig. 1.b10).

F2. Side-by-side ML-IBL layout showcases the final design's visual (Fig. 1.b), which matches common templates for secondary teachers to construct lesson plans. Teachers can add, reorder, and remove individual or pairs of ML and IBL components. They can experiment with various combinations for the learning design, evaluating how to interact with the ML components to best support a specific IBL behavior, or identifying which IBL behaviors students are likely to exhibit when using a particular ML technique.

F3. Customizable ML components enable teachers to input different data, ranging from the entire dataset to a cluster formed by a previous clustering component, or a manually selected subset (Fig. 2). For instance, a teacher might save the clusters generated by K-means clustering (Fig. 2.5) and re-apply K-means clustering on a specific cluster. This supports the incremental construction of ML-related data practices throughout the iterative IBL process.

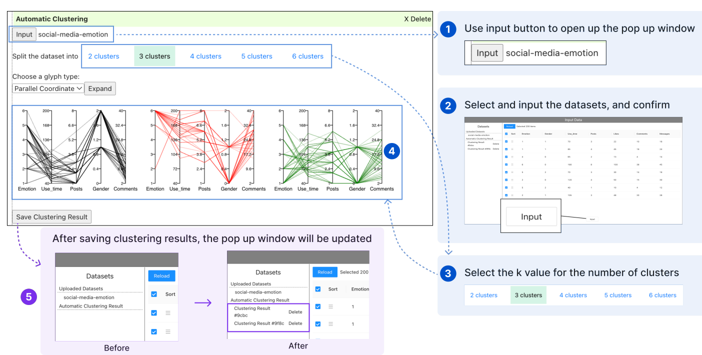

Fig. 2. (1) Update input for K-means clustering; (2) select and view the entire "social media use and emotions" dataset [5] as the input; (3) select K = 3 for the number of clusters; (4) visualize the clustering results with parallel coordinates; (5) click to save the newly-generated clusters; the new clusters saved are added to the available inputs.

4 Research Method

4.1 Participants

Through education-related mailing lists and teacher education program coordinators, we recruited 14 secondary school teachers (Table 2). The study was approved by the Institutional Research Subjects Review Board (Case Number: STUDY00003947).

4.2 Study Procedure and Data Collection

Secondary school teachers use ML4Inq to co-design ML-IBL activities with ML experts. We chose co-design approach [30], for three reasons:

Table 2. Demographics and teaching backgrounds of teachers.

PID	Gender	Types of Schools	Grade(s)	Subject(s)	Years
P1	Female	Public schools	9-12th	Algebra and AP Statistics	16
P2	Female	Public schools	9-12th	Computer Science, Business Technology	5
P3	Male	Public schools	9-12th	Math, Science	8
P4	Female	Public schools	9-12th	Biology (Science)	7
P5	Female	Public schools	7-9th, 12th	7th grade Life Science. 8th grade Accelerated Living Environment, 9th grade Living Environment, 12 grade AP Biology	5
P6	Female	Public schools	9th	World History, writing	31
P7	Male	Private schools	6-9th	Computer Science, Robotics, Pre-Algebra	27
P8	Female	Public schools	6, 8-12th	Biology, General Science, Earth Science	15
P9	Female	Public schools	K-5th, 9-10th	Science, Computer Science	12
P10	Male	Public, Charter schools	5-8th	Engineering, Computer Science	9
P11	Male	Public schools	9-12th	AP Physics, Physics, Mathematics, Computer Science, Chemistry, Earth Science	28
P12	Female	Public, Private, Charter schools	5-12th	Computer Science, Computational Thinking	23
P13	Male	Public, Private, Charter schools, Outreach programs.	6-12th	Earth Science, Chemistry, Engineering, Computer Science, Data Science	5
P14	Male	Public, Private, Charter schools	K-12th	Science, STEM/Maker Education	20

(1) Teachers' expertise is essential due to the lack of pedagogical theories on integrating ML methods into K-12 IBL [33];
(2) Despite specific training and technology enhancement, teachers often limit their instruction designs to surface-level data interpretation [6] and underestimate their ability to teach ML [25];
(3) Co-designing with teachers can promote curriculum innovation, support classroom technology integration, enhance teacher ownership and sustainability of materials, and foster teacher learning [30].

The study consists of three one-hour sessions and is conducted online via Zoom.

Session 1: Teacher-as-Learner. Before this session, teachers filled out the pre-study survey collecting demographic information and teaching backgrounds.

During this session, a researcher guided teachers to go through and get familiar with each component in ML4Inq, including (1) ML components with data visualization, (3) IBL components, and (4) the design flow in ML4Inq. In the end, the teacher and the researcher discussed which datasets the teacher wished to use for the design sessions. A minimum one-day interval is scheduled for the researcher or the teacher to acquire new datasets. The researcher pre-processed data by eliminating data points with missing values and non-numeric features, ensuring compatibility with ML4Inq.

Session 2 & 3: Teacher-as-Designer. In the two design sessions, teachers used ML4Inq to create ML-IBL activities using the datasets of their choice. They first define high-level learning objectives, explore the data using ML components, design the IBL instructions for individual steps, and conduct the specific IBL behaviors from their students' perspective. A researcher and an ML expert provided facilitation when needed. Upon completion, teachers were asked to review their process and contemplate modifications to the learning activity they created. We collected screen recordings and the log of teachers' text input and interaction with ML4Inq.

4.3 Measures and Data Analysis

To investigate how secondary school teachers incorporate ML into curriculum-aligned IBL activities using ML4Inq, we first compared the learning goals with curriculum standards by categorizing the teacher-designed ML-IBL activities based on the disciplines outlined in U.S. K-12 curriculum standards, including Next Generation Science Standards (NGSS) [9] and standards set by the National Council for the Social Studies (NCSS) [27]. We also measured the frequency of dataset subjects.

To analyze how ML and IBL were integrated in teachers' design, we measured the frequency of ML-revealed data patterns applied by teachers and examined the occurrence of the immediate subsequent IBL component that follows various data patterns to gain insights into their specific usage.

5 Results

5.1 Diversity of curriculum-aligned data topics

14 teachers created 26 learning activities. Among the 21 different datasets used by teachers, two were provided by teachers; 17 were found and retrieved by researchers based on the topics or disciplines requested by teachers; two were chosen by teachers from our example datasets. The disciplines ranged from science ($N = 10$) to social studies ($N = 11$) (Table 3). In the science-related datasets, we identified seven out of 13 Disciplinary Core Ideas in NGSS for the U.S. STEM education [9], which encompass four key disciplines: (1) Physical Sciences, (2) Life Sciences, (3) Earth and Space Sciences, and (4) Engineering, Technology, and Applications of Science. In the rest datasets, all five disciplines outlined in the U.S. social study standards [27] are identified: History, Civic and Government, Geography, Economics, and Psychology. Some are interdisciplinary, such as "Social Complexity in Ancient Civilizations," which integrates data attributes from History, Civic and Government, and Geography.

5.2 Distribution and usage of ML-revealed patterns for IBL

Teachers applied ML algorithms to discover a diverse range of patterns (Table 4) and used them to promote students' IBL behaviors (Table 5-7).

(1) *Intra-cluster similarities, inter-cluster differences, and centroids.* All of the 26 ML-IBL activities involve interpreting correlations between attributes by

Table 3. Dataset discipline used in teacher-designed ML-IBL activities.

Discipline	Dataset Subject
Environmental Science (N = 6)	Air Composition, Air Quality, Forest Coverage, Water Quality, Climate Change, Weather Pattern
Biology and Ecology (N = 4)	Fishery, Blueberry Yield, Ocean Animals, Alien Rose, Bacteria Growth
Social Studies: History, Geography, Economics, Civics, Psychology (N = 11)	Marvel Movie, Worldwide Country Development, Social Complexity in Ancient Civilization, CS Career, AI Development, Social Media Usage and Emotion, Student Stress, Teen Well-being, Worldwide Cost of Living, Adult Income, Baseball Game Logs

Table 4. The distribution of patterns covered by teacher-designed ML-IBL activities.

Pattern	Number of ML-IBL activities
Intra-cluster similarity	26
Inter-cluster comparison	26
Centroid	18
Intra-cluster variation	22
Outlier	7
Comparison of the prediction and the ground truth	18
Inter-cluster comparison of neighboring data points	3

analyzing intra-cluster similarities and inter-cluster differences. 18 of them introduced centroids for inter-cluster comparison. Such cluster analysis supported hypothesis iteration in all activities. For instance, P5 identified a counterintuitive pattern through inter-cluster comparison. She found that clone size negatively impacted blueberry yield, searched for a research paper about the underlying causes related to biodiversity, and used this as a learning opportunity for students to discuss the importance of biodiversity. P5 found it very important for students to interact with the real-world dataset and discover the patterns themselves could be more memorable. P13 and P14 utilized intra- and inter-cluster patterns to identify that most data points have higher cs credits and discussed how such data bias influenced the analysis.

(2) Analyzing variations in real-world data. 22 out of the 26 ML-IBL activities guided students in identifying and interpreting variances. Four types of applications of variance analysis in IBL are identified (see details in Table 5).

(3) Outlier analysis for further inquiry. Seven out of the 26 ML-IBL activities utilized outliers revealed by ML methods to promote IBL, with three types of IBL applications identified (see details in Table 6).

(4) Evaluation and interpretation of KNN outputs. Teachers used KNN in 18 activities. Three types of IBL applications are identified (see details in Table 7).

The analysis and interpretation of patterns (Table 5-7) align with the U.S. CCSS for Mathematics [15], specifically with reasoning abstractly and quantitatively, constructing viable arguments, and modeling with mathematics. ML-supported data analysis practices extend beyond the typical high school mathematics curriculum, which generally focuses only on patterns of association in bivariate data and the shapes of two- and three-dimensional data. The findings

Table 5. Analyzing variations in real-world data (N = 22).

IBL applications	Example
Use varying features to evaluate and refine hypotheses (N = 9)	*An activity about students' GPAs and course registration*: **(1) Hypothesis**: High CS GPA students have high math GPAs. **(2) Cluster analysis**: High CS GPA students have varying math GPAs. **(3) Conclusion**: The finding "high CS GPA students have varying math GPAs" rejects my hypothesis "high CS GPA students have high math GPAs". *An activity about blueberry yield*: **(1) Cluster analysis**: There are large variations in bumble bee intensity for all the clusters with different levels of blueberry yield. **(2) Hypothesis**: There is little correlation between bumble bee intensity and blueberry yield. **(3) New question**: What are other attributes that could be correlated with blueberry yield? **(4) Hypothesis iteration and feature selection**: Select a few new data attributes based on the updated hypothesis. **(5) Predict with KNN to test hypothesis**: Test the new set of attributes to see if they can accurately predict blueberry yield.
Interpret in real-world contexts (N = 5)	*An activity about CS career*: Identified the same finding, and further interpreted why high school CS GPA seems not to be strongly correlated with math GPA: "HS Math mostly deals with procedure fluency and has little to do with logic and abstraction. CS must do more with logic and abstraction.
Probe further question-asking with large variances (N = 4)	*An activity about Marvel movies*: **(1) Cluster analysis**: There is a large variance in the phases of Marvel movies with higher box office. **(2) New question**: Why do the phases vary for Marvel movies with higher box office? **(3) Update input**: Input a cluster with high box office and varying phases. **(4) New cluster analysis**: Apply K-means clustering on the new input.
Discuss the tolerance for variances (N = 4)	**(1) Variance analysis**: compare the variance in fish stock with the data range. **(2) Cluster analysis**: Tolerate variances to interpret the shared similarity.

Table 6. Outlier analysis for further inquiry (N = 7).

IBL applications	Example
Validate outliers with clustering (N = 4)	**(1)** Conduct clustering multiple times and analyze that the data points don't fit any clusters generated. **(2)** Interpret the potential causes of the outliers. For example, "this is a country with resources significantly higher than the rest countries for AI development, and thus, has an AI index outlier".
Multiple IBL cycles follow up the outlier analysis (N = 3)	*An activity about mental health*: **(1) Cluster analysis**: Some outliers in a high-workload cluster have low stress and low workload. **(2) New questions**: Why do some students in low-stress and high-workload clusters have low workloads? **(3) Hypothesis**: These students with low workloads but high stress may have bad living environments, such as lower safety and lower living condition indexes. **(4) Feature selection**: Add new features "safety" and "living condition" to build ML algorithms. **(5) New cluster analysis**: Apply K-means clustering with the new set of features and analyze the clusters. **(6) Conclusion**: Students with low workloads but high stress have low safety and low living conditions.
Cluster the outlier cluster for further investigation and explanations (N = 3)	*An activity about social media use*: **(1) Hypothesis**: The cluster receiving the most likes has the most positive emotions. **(2) Cluster analysis**: Identify an outlier cluster with both the most positive and the most negative emotions. **(3) Update input**: Input the cluster with the outliers. **(4) New cluster analysis**: Apply K-means clustering on the new input. **(5) Conclusion**: Extreme emotions are related to extreme engagement with social media (i.e., receiving the most likes, posting more frequently, sending the most messages).

Table 7. Evaluation and interpretation of KNN outputs (N = 18).

IBL applications	Example
Predict multiple unknown data points (N = 13)	Analyze the prediction accuracy to determine if the feature predicted is highly correlated with the independent features selected. If the accuracy is low for many trials of unknown data points, students can experiment by updating the feature selection to refine the KNN predictor.
Alternate target features for prediction (N = 5)	**(1) KNN interpretation**: Compare the predicted daily stress with the ground truth. **(2) Feature selection**: Select another set of features to predict well-being scores. **(3) KNN interpretation**.
Interpret group patterns of the k-nearest neighbors (N = 3)	*An activity about ecological field sites*: **(1) Parameter adjustment**: Select k values to predict field sites' livability based on the neighboring data points. **(2) KNN**: Predict an unknown data point with a high canopy height, and then predict a low-canopy data point. **Data analysis**: Compare the two sets of neighboring data points and their livability.

highlight teaching opportunities in science subjects to introduce advanced (1) statistical concepts (e.g., outliers, variance, and centroids) and (2) data practices (e.g., data collection and curation, evaluating the statistical power of various types of evidence, and interpreting findings with subject-specific expertise).

6 Discussion and Future Work

Design Implications for Supporting Advanced ML Practices through Novice-Centric Tools. Beyond the K-12 education context, our findings provide broader insights into designing computer-supported experiences for novices to build ML expertise incrementally through hands-on practices.

Customizable ML for incremental sensemaking in ML-IBL. The data analyst's sensemaking process [11], which involves collecting and organizing data to generate knowledge outputs, offers valuable perspectives on how customizable ML practices can support hypothesis iteration across different inquiry phases. The sensemaking process contains four major transitions, two for information foraging: (1) *filter* relevant data, (2) *search* for a larger set of data, and two for sensemaking: (3) *hypothesize* based on the relevant data, (4) *test* hypothesis with the relevant data. By looking into how teachers connected ML-revealed patterns with IBL, we identified that conducting clustering and KNN can be used to filter larger data to relevant subsets and reveal patterns for hypothesizing. ML customization practices, including updating input, parameter adjustment, and feature selection, can support hypothesis testing or enable more efficient searching of a larger dataset. These ML-IBL connections represent incremental constructions of ML and capture ML's finer-grained roles within higher-level IBL phases [29]: filtering, hypothesizing, testing, and searching. With the advances in generative AI, we can extend the sensemaking roles offered by classic ML algorithms to cover more capabilities offered by AI [38].

Modular systems have been recognized in HCI as effective for fostering flexibility and reusability in diverse contexts [26]. For instance, end-user programming environments such as Scratch and Blockly demonstrate how modularity enables users to focus on creativity and problem-solving without requiring advanced technical knowledge [31]. Modular ML components break tasks into smaller, manageable steps, reducing cognitive load and building confidence [36], allowing novices to experiment with ML and interpret the output immediately. However, we did observe an imbalanced usage of different ML components. For example, "Manual Clustering" and "Update Input" rarely appeared, even though they encourage deeper engagement by allowing teachers to group data based on their interpretations and conduct follow-up analyses. This underscores the importance of balancing usability with cognitive demand in tool design. Prior research shows that users often prefer automated over manual features when faced with time constraints or steep learning curves [28, 22]. Thus, teachers may have avoided them due to the perceived complexity or lack of immediate utility [20] compared to using the default input and automatic clustering. Future iterations could balance usability with learnability and cognitive demand [2] with contextual hints or example demonstrations.

Limitations and Future Work. While the authoring tool presents a promising approach for integrating ML into K-12 education, several limitations and the corresponding next steps should be acknowledged. First, the tool in its current version relies heavily on pre-designed modular ML components, which may limit flexibility for diverse educational scenarios. Second, the limited time for co-design constrained the coverage of subject topics, ML-IBL design possibilities, and thorough iterations to refine details. Third, there is no comparative study for teachers to design similar activities with existing conventional tools under the guidance of ML experts. We plan to conduct deeper analyses of teachers' decision-making processes and evaluate the ML-IBL activities with students. Insights from existing and future analyses on the interplays between ML practices and IBL behaviors could inform new features to enhance ML4Inq's capabilities, including a broader range of ML techniques and intelligent scaffolding to provide personalized support to users.

7 Conclusion

This work tackles the challenge K-12 teachers face in integrating ML into the curriculum due to limited preparation in statistics and ML. We created ML4Inq for teachers to customize ML and IBL components to design curriculum-aligned ML-supported IBL in their classrooms. Our study reveals how teachers applied ML practices and ML-revealed patterns to foster meaningful IBL in students.

Bibliography

[1] AI4ALL. AI4ALL: Opening Doors to Artificial Intelligence. Website, 2024. URL https://ai-4-all.org/. Accessed: 2024-11-17.

[2] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. Power to the people: The role of humans in interactive machine learning. *AI magazine*, 35(4):105–120, 2014.

[3] Michael R Anderberg. *Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks*, volume 19. Academic press, 2014.

[4] Sumon Biswas, Mohammad Wardat, and Hridesh Rajan. The art and practice of data science pipelines: A comprehensive study of data science pipelines in theory, in-the-small, and in-the-large. In *Proceedings of the 44th International Conference on Software Engineering*, pages 2091–2103, 2022.

[5] Emirhan BULUT. Social media usage and emotional well-being, 2024. URL https://www.kaggle.com/dsv/8460631.

[6] Stephanie Casey, Rick Hudson, Taylor Harrison, Heather Barker, and Jordan Draper. Preservice teachers' design of technology-enhanced statistical tasks. *Contemporary Issues in Technology and Teacher Education*, 20(2): 269–292, 2020.

[7] Souti Chattopadhyay, Zixuan Feng, Emily Arteaga, Audrey Au, Gonzalo Ramos, Titus Barik, and Anita Sarma. Make it make sense! understanding and facilitating sensemaking in computational notebooks, 2023.

[8] Computer Science Teachers Association. Computer science teachers association (csta), 2024. URL https://csteachers.org/. Accessed: 2024-11-17.

[9] National Research Council et al. Next generation science standards: For states, by states. 2013.

[10] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.

[11] Ian Drosos, Advait Sarkar, Xiaotong Xu, Carina Negreanu, Sean Rintel, and Lev Tankelevitch. "it's like a rubber duck that talks back": Understanding generative ai-assisted data analysis workflows through a participatory prompting study. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*, pages 1–21, 2024.

[12] Tim Erickson and Joachim Engel. What goes before the cart? introducing classification trees with arbor and codap. *Teaching Statistics*, 45:S104–S113, 2023.

[13] Tim Erickson, Michelle Wilkerson, William Finzer, and Frieda Reichsman. Data moves. *Technology Innovations in Statistics Education*, 12(1), 2019.

[14] Cassia Fernandez, Paulo Blikstein, and Roseli De Deus Lopes Lopes. Design failures in data visualization programming activities. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*, pages 574–586, 2024.

[15] National Governors Association Center for Best Practices. Common core science standards. national governors association center for best practices, council of chief state school, 2010.

[16] Daniel Frischemeier, Rolf Biehler, Susanne Podworny, and Lea Budde. A first introduction to data science education in secondary schools: Teaching and learning about data exploration with codap using survey data. *Teaching Statistics*, 43:S182–S189, 2021.

[17] Yolanda Gil, Mark Greaves, James Hendler, and Haym Hirsh. Amplify scientific discovery with artificial intelligence. *Science*, 346(6206):171–172, 2014.

[18] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D Hansen, and Jonathan C Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.

[19] Janice D Gobert. Inq-its: Creating rigorous assessment and real-time support of science learning. *XRDS: Crossroads, The ACM Magazine for Students*, 29(3):36–40, 2023.

[20] Eric Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 159–166, 1999.

[21] Rotem Israel-Fishelson, Peter F Moon, Rachel Tabak, and David Weintrop. Preparing students to meet their data: an evaluation of k-12 data science tools. *Behaviour & Information Technology*, pages 1–20, 2023.

[22] Victor Kaptelinin and Bonnie A Nardi. *Acting with technology: Activity theory and interaction design*. MIT press, 2009.

[23] Alla Keselman. Supporting inquiry learning by promoting normative understanding of multivariable causality. *Journal of Research in Science Teaching*, 40(9):898–921, 2003.

[24] Jennifer N Lovett and Hollylynne S Lee. New standards require teaching more statistics: Are preservice secondary mathematics teachers ready? *Journal of Teacher Education*, 68(3):299–311, 2017.

[25] Lívia S Marques, Christiane Gresse von Wangenheim, and Jean CR Hauck. Teaching machine learning in school: A systematic mapping of the state of the art. *Informatics in Education*, 19(2):283–321, 2020.

[26] Brad Myers, Scott E Hudson, and Randy Pausch. Past, present, and future of user interface software tools. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(1):3–28, 2000.

[27] Charles B Myers, Susan Adler, Allan Brandhorst, Alberta Macke Dougan, Wayne Dumas, Lewis Huffman, Pat Rossman, Donald O Schneider, Robert J Stahl, Ceola Ross Baber, et al. *National standards for social studies teachers*. National Council for the Social Studies, 2006.

[28] Don Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[29] Margus Pedaste, Mario Mäeots, Leo A Siiman, Ton De Jong, Siswa AN Van Riesen, Ellen T Kamp, Constantinos C Manoli, Zacharias C Zacharia, and Eleftheria Tsourlidaki. Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational research review*, 14:47–61, 2015.

[30] William R Penuel. Co-design as infrastructuring with attention to power: Building collective capacity for equitable teaching and learning through design-based implementation research. *Collaborative curriculum design for sustainable innovation and teacher learning*, pages 387–401, 2019.

[31] Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Amon Millner, Eric Rosenbaum, Jay Silver, Brian Silverman, et al. Scratch: programming for all. *Communications of the ACM*, 52(11):60–67, 2009.

[32] Aimee Schwab-McCoy, Catherine M Baker, and Rebecca E Gasper. Data science in 2020: Computing, curricula, and challenges for the next 10 years. *Journal of Statistics and Data Science Education*, 29(sup1):S40–S50, 2021.

[33] Matti Tedre, Tapani Toivonen, Juho Kahila, Henriikka Vartiainen, Teemu Valtonen, Ilkka Jormanainen, and Arnold Pears. Teaching machine learning in k–12 classroom: Pedagogical and technological trajectories for artificial intelligence education. *IEEE access*, 9:110558–110572, 2021.

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

[35] Dakuo Wang, Josh Andres, Justin D Weisz, Erick Oduor, and Casey Dugan. Autods: Towards human-centered automation of data science. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–12, 2021.

[36] David Weintrop and Uri Wilensky. To block or not to block, that is the question: students' perceptions of blocks-based programming. In *Proceedings of the 14th international conference on interaction design and children*, pages 199–208, 2015.

[37] Yuanlin Zhang, Jianlan Wang, Fox Bolduc, William G Murray, and Wendy Staffen. A preliminary report of integrating science and computing teaching using logic programming. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 9737–9744, 2019.

[38] Chengbo Zheng, Kangyu Yuan, Bingcan Guo, Reza Hadi Mogavi, Zhenhui Peng, Shuai Ma, and Xiaojuan Ma. Charting the future of ai in project-based learning: A co-design exploration with students. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2024.

[39] Xiaofei Zhou, Jingwan Tang, Michael Daley, Saad Ahmad, and Zhen Bai. "now, i want to teach it for real!": Introducing machine learning as a scientific discovery tool for k-12 teachers. In *Artificial Intelligence in Education: 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14–18, 2021, Proceedings, Part I 22*, pages 486–499. Springer, 2021.

[40] Xiaofei Zhou, Jingwan Tang, Hanjia Lyu, Xinyi Liu, Zhenhao Zhang, Lichen Qin, Fiona Au, Advait Sarkar, and Zhen Bai. Creating an authoring tool for k-12 teachers to design ml-supported scientific inquiry learning. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2024.