Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #47456

Enhancing AI Education through Marine Robotics and Real-World Data: A Case Study in Coastal Environmental Monitoring

Dr. Gregory Murad Reis, Florida International University

Dr. Gregory Murad Reis is an Associate Teaching Professor at the Knight Foundation School of Computing and Information Sciences at FIU. He has a B.S. in Computer Science, M.S. in Systems Engineering, and a Ph.D. in Computer Science. With over two decades of experience in teaching Mathematics and Computer Science, Dr. Murad Reis has research expertise in Artificial Intelligence, Marine Robotics, STEM Education, and Data Science. He has also accepted the invitation to serve as an Environmental Finance Fellow at FIU's Environmental Finance & Risk Management (EFRM) and as an Affiliated Faculty in the FIU Institute of Environment. He has published work on the localization and navigation of underwater robots in GPS-denied environments, analysis of the spatiotemporal dynamics of the ocean, data analysis of water profiles, persistent monitoring of the ocean, and AI models applied to ocean monitoring and agriculture.

Luana Okino Sawada, Florida International University
Paulo Padrao, Florida International University
Jose Fuentes, Florida International University

Enhancing AI Education through Marine Robotics and Real-World Data: A Case Study in Coastal Environmental Monitoring

Gregory Murad Reis, Luana Okino Sawada, Paulo Padrão and Jose Fuentes
{gmuradre, lokin001, plope113, jfuen099}@fiu.edu
Knight Foundation School of Computing and Information Sciences
Florida International Univerisity

Abstract

This research investigates the effects of integrating real-world datasets from marine robotics missions into data science and artificial intelligence (AI) education. By integrating real environmental data gathered by marine robots into the educational process, we aimed to increase students' interest, academic performance, and engagement in data science, AI, and machine learning (ML). We asked students for feedback after they finished actual classes and/or workshops using water quality data and machine learning exercises. The results indicate that students found the experience highly enjoyable and found it applicable, reporting increased knowledge about data science concepts and increased motivation to learn AI/ML. The results are in line with emerging research that indicates real-world data projects have the ability to significantly improve STEM learning. We examine how using marine environmental data and robotics in the classroom to teach technical content ignites critical thinking and real-world problem-solving. This paper determines the educational value of marine robotics and environmental data in AI education and presents evidence-based results for educators seeking to advance student learning through real-world data science exercises. This study additionally provides the data sets for teaching practitioners who can utilize them for teaching practice.

1 Introduction

Teaching data science and AI in a way that is engaging and practical is a challenge as these fields become central to the 21st-century skill set[1]. Traditional curricula often rely on contrived examples or toy datasets, which can lack authenticity and personal relevance for students[1]. Prior research has highlighted the absence of "author proximity" in data science education, meaning that students neither contribute to data production nor see their own experiences in the data[1]. This gap can lead to lower student engagement and difficulty in understanding the real-world impact of AI and machine learning.

Integrating real-world datasets and projects into the learning process may be a promising approach to increase engagement[2]. Educational studies have shown that when learning experiences are grounded in real-world, community-based problems, student engagement significantly improves[3]. By working with authentic data, especially data that students find meaningful, learners become more interested in the analysis and outcomes. For example, the National Survey of Student Engagement (NSSE) results indicate that community-based projects and real-world problem-solving correlate with higher student involvement and interest[3].

In this context, analyzing environmental datasets present an opportunity for data science and AI education. Government agencies like the National Oceanic and Atmospheric Administration (NOAA) have demonstrated the value of bringing real environmental datasets to classrooms. Through *NOAA's Data in the Classroom* program, middle and high school students "can explore real-world environmental topics using historical and real-time NOAA data from satellites and other systems", which helps them "ask questions, challenge assumptions, interpret data, and think critically about the world around them"[4]. Such data-driven exploration is not just for scientists, but a tool for educators to inspire the next generation of problem-solvers[4].

In this paper, we present an educational intervention that combines data science/AI instruction with real-world data from marine robotics data collection missions in Biscayne Bay, Florida. Our goal is to investigate how this approach influences student engagement, learning, and motivation in AI and machine learning. We hypothesized that students would find the experience more engaging and would show increased interest and confidence in applying data science skills, compared to learning with abstract examples. This paper presents the results of a post-intervention survey and discusses the implications of integrating real environmental data into AI education. By shifting the focus toward educational outcomes, we aim to demonstrate how real-world marine robotics data can enrich the learning experience and inspire the next generation of data scientists. This paper makes two key contributions to the field of AI and data science education.

- First, it presents survey results that evaluate the impact of integrating real-world environmental data into AI and data science coursework, providing insights into student engagement, learning outcomes, and interdisciplinary participation across multiple majors.
- Second, it contributes to the open education community by making publicly available datasets collected from marine robotics missions, making it possible for educators to incorporate real-world environmental data into their own courses.

2 Background and Related Work

2.1 Real Datasets in STEM Education

Educators in STEM fields have long advocated using real datasets to improve learning. Studies in statistics education argue for the advantage of real-life data sets to increase student interest and perceived relevance of coursework[5]. NASA's education initiatives similarly emphasize that "the use of real data provides formal educators the opportunity to teach their students real-world applications of STEM subjects"[6]. Combining authentic datasets with lessons aligned to standards creates learning experiences that students carry forward in their academic careers[6]. However, incorporating real data comes with challenges; educators must address issues of data complexity, and accessibility to provide proper guidance for students who may be unfamiliar with large, unstructured datasets[6]. Successful programs (e.g., NASA's and NOAA's classroom data programs) highlight the importance of scaffolding and context to make the data accessible and engaging.

2.2 Project-Based Learning and Data Science Education

Project-based learning (PBL) approaches are known to improve student engagement by centering learning around meaningful problems. When students work on projects with real-world impacts, such as analyzing local environmental data or solving community issues, their motivation tends to increase[3]. In data science education, incorporating real-life projects and real-life data has been recognized as a way of making programs more interesting and relevant for students[2]. This approach is critical in fields like AI and ML, which can be abstract; therefore, grounding these concepts in tangible data and real scenarios helps demystify the technology and stimulate curiosity.

2.3 Robotics in STEM Education

Robotics has become a valuable means for enhancing STEM education by offering concrete representations of computer and engineering concepts. Prior work on robotics in STEM education shows that using robots and their data can intrinsically or extrinsically motivate the learners[7]. Robotics projects often involve elements of problem-based learning that boost classroom engagement[7]. A broad meta-analysis of educational robotics interventions found a moderate but significantly positive effect on student learning outcomes compared to traditional methods[8]. In other words, courses employing robotics and real data tend to produce better learning gains than lecture-based approaches[8]. These findings align with constructivist learning theories: students learn more deeply when they actively construct knowledge from real, meaningful experiences. Learning activities centered on robotics such as programming or interpreting the data can organically engage students[7]. Robotics education is especially suitable for methods such as inquiry-based learning and cognitive apprenticeship, in which students learn by doing and resolving problems in different contexts[7]. Recent work has shown that learning interventions based on educational robotics lead to significantly better learning outcomes than traditional instructional approaches, improving both knowledge acquisition and skill acquisition[8]. These benefits occur across age groups and subject areas, indicating that the positive effects are generalizable, provided that a well-designed curriculum is implemented[8].

2.4 Marine Robotics and Environmental Applications

Within robotics education, marine robotics is especially attractive because of its association with environmental studies. Merging real datasets in teaching practices has been effective in engaging students and deepening their comprehension of machine learning principles across different areas[9,10]. Marine robotics missions – such as autonomous underwater vehicles collecting oceanographic measurements or robotic gliders surveying marine environments – generate rich datasets that are intrinsically tied to real environmental challenges. Integrating these data into coursework allows students to engage in hands-on data science related to issues such as ocean health, climate change, and environmental monitoring. Also, the use of real mission data enhances the relevance of student assignments by engaging them with the same datasets that scientists employ in their research and discoveries.

2.5 AI and Machine Learning in Education

Project-based learning approaches, as discussed by[11], provide a robust framework to teach machine learning and artificial intelligence. By emphasizing student-driven projects and interdisciplinary problem-solving, these approaches align closely with the objectives of this study.[12] explored the impact of real-world case studies on student engagement in data science education, underscoring the pedagogical value of authentic problem-solving exercises.

With AI becoming ubiquitous, there is a push to teach it in a way that is both accessible and grounded in reality. AI education initiatives for K-12 have shown that hands-on projects can increase students' interest and confidence. In one study, middle schoolers who took an AI-focused PBL course showed significant increases in self-efficacy in using AI, greater content knowledge of AI concepts, and higher career interest in AI fields post-course[13]. These gains were observed across genders, suggesting that authentic AI projects can engage a diverse range of students equally well. These findings underscore that when students see how AI can be applied to real problems (like analyzing data from a robot or making predictions about the environment), they become more motivated to learn the underlying concepts.

2.6 Integration of Marine Robotics, AI, and PBL

Our approach integrates real-world marine robotics data into a project-based AI/ML curriculum. This blends the authenticity of real data with the interactivity of robotics and the relevance of environmental problem-solving. We expect that students will be more engaged, learn the material more deeply, and feel more motivated to continue in the field compared to a traditional lecture-based approach.

3 Methods

The course was structured around project-based learning and experiential activities using authentic datasets from environmental science and robotics domains. This approach is grounded in prior research showing that context-rich projects can boost student interest and interdisciplinary problem-solving skills[14]. Our course incorporated real environmental data (e.g. climate and sensor data) and robotics datasets (from actual robot experiments) to provide hands-on contexts. These materials were chosen to be relevant to students' academic interests, an important factor for engagement noted in the literature[15]. In line with best practices, we solicited student input on topic selection to ensure the datasets and problems were meaningful to them. In this approach, we integrate real data with a context and a purpose in order to give learners experience with genuine data and illustrate the usefulness of data science skills in the real world[16].

3.1 Survey Data Collection

To evaluate the impact of real-world data integration, we administered a survey to the students that enrolled in Computer Sciences classes and/or workshops developed at Florida International

University. The survey was designed to measure student engagement, confidence, and motivation in data-rich learning environments. We mapped our course learning outcomes to specific survey items, asking students to self-assess their skills and attitudes at the beginning and end of the course (adapted from [17]'s approach). The full survey instrument is provided in Appendix A for reference.

3.2 Survey Data Processing

Key constructs included engagement (interest in course material, perceived relevance of datasets), motivation (enjoyment and willingness to invest effort), confidence (self-efficacy in analyzing data and solving problems), and learning outcomes (self-reported gains in knowledge and skills). This mixed-methods evaluation (quantitative Likert-scale items and open-ended feedback) is consistent with methodologies in STEM education research, where student surveys and focus groups are used to assess changes in attitudes and competencies after experiential learning interventions [17,18]. All students completed the survey, and we performed paired statistical analyses (e.g. Wilcoxon signed-rank tests) to detect changes pre/post, as well as thematic analysis of open-ended responses. These data were then compared to results reported in the literature to contextualize our findings.

3.3 Water Quality Data Availability

The water quality datasets used in this study have been made publicly available[1] [2]. These datasets include real-world environmental measurements collected using marine robotics, allowing educators and researchers to replicate or extend our study in their own courses. We are providing open access to these datasets to facilitate the integration of practical, real-world data into data science and AI education, promoting hands-on learning experiences for a broader audience.

4 Results

Our results align with results from previous experiential learning studies. Studies by [14] stress that real-data, hands-on teaching builds greater engagement and improved problem-solving. Similarly, [15] found that students who used environmental data sets had stronger analytical skills than students who used simulated data. Our results confirm that incorporating real-world robotics data can significantly augment confidence, teamwork, and technical abilities for AI and data science students.

4.1 Survey Results

The survey responses from 20 students were analyzed across multiple dimensions, including course enrollment, major, year of study, prior familiarity with robotics, and understanding of environmental data collection. Figure 1 depicts an overview of key distributions:

Most of the respondents were from Data Science, with a significant portion from Human Computer Interaction (HCI); there is also a smaller number engaged in Marine Robotics, due to

[1] `https://www.datascience4everyone.org/datasets`
[2] `https://go.fiu.edu/waterqualitydata`

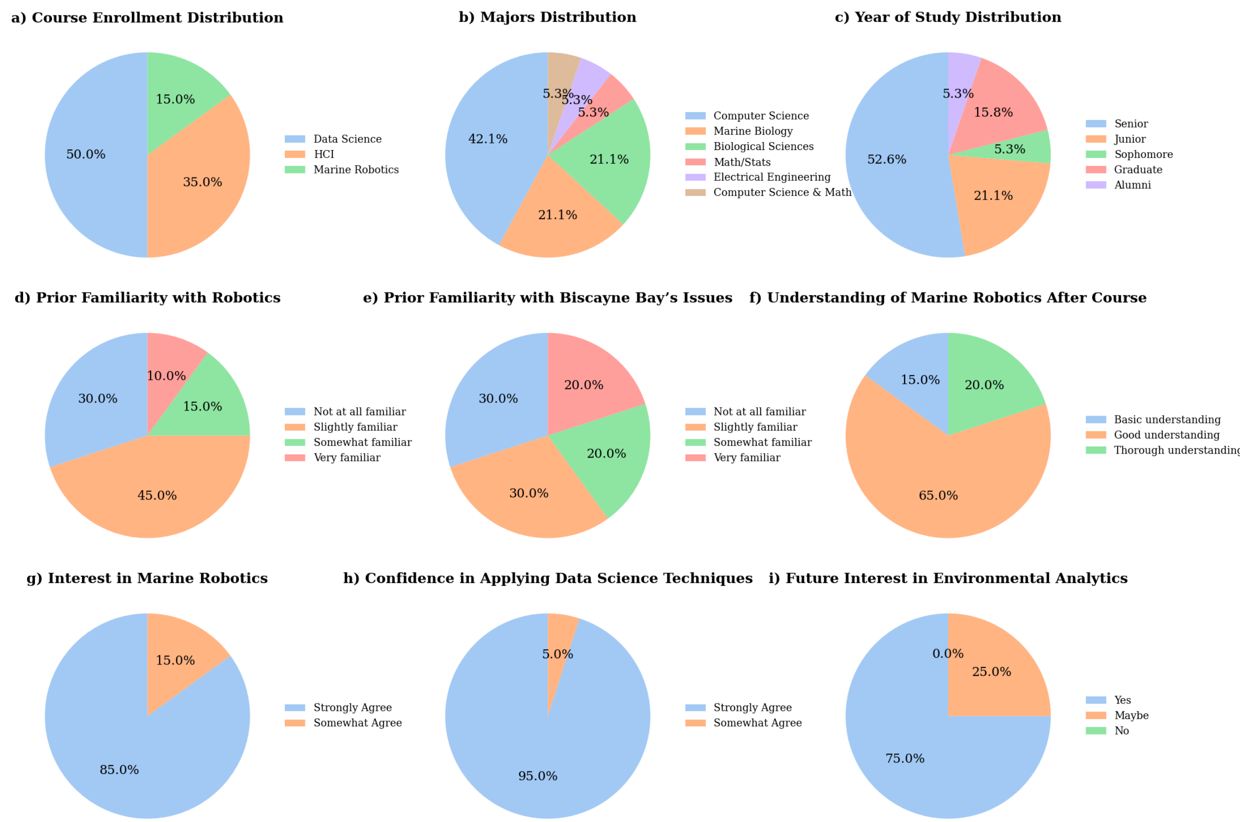

Figure 1: Survey results, student distribution among the questions answered in the survey

their specialized nature. Although the dominant group was Computer Science students (40%), approximately half of the students came from non-CS backgrounds, highlighting interdisciplinary engagement. Moreover, half of the students were seniors, suggesting that they had prior experience in other coursework before engaging with real-world robotics data. Nonetheless, 75% of the students had little to no prior familiarity with robotics; therefore, this course was their first exposure to robotics-driven environmental data collection. Similarly, 60% of students had limited prior knowledge, reinforcing the value of using real-world environmental data to educate students beyond data science theory. Despite their initial unfamiliarity, 85% of students reported a significant increase in understanding, demonstrating that hands-on work with real data effectively teaches environmental data collection. Real-world exposure made students more confident in using AI/ML techniques to solve practical problems, with 90% of students feeling more motivated to apply ML concepts to real-world environmental problems. On the other hand, prior robotics experience helped sharpen understanding of marine robotics, yet even beginners benefited significantly; this sparked their curiosity about marine robotics. Additionally, students found more value in real-world applications, reinforcing the need to integrate such approaches earlier in the curriculum.

Metric	CS Majors (Avg Rating)	Non-CS Majors (Avg Rating)
Engagement with Real-World Data	5.0/5	4.9/5
Confidence in Applying AI/ML	4.8/5	4.5/5
Understanding of Marine Robotics	4.6/5	4.4/5

Table 1: Comparison of Average Ratings between CS and Non-CS Majors

4.1.1 Computer Science vs. Non-Computer Science Majors

CS and Non-CS students responded similarly, showing that real-world environmental data appeals across disciplines.

4.2 Correlation Analysis

We analyzed the relationships between crucial survey metrics and found important correlations. A Pearson correlation of 0.76 indicates that higher engagement with real-world data led to increased confidence in applying AI/ML skills. Similarly, prior robotics experience was positively correlated (0.65) with greater understanding after the course. Furthermore, a correlation of 0.72 suggests that increased interest in marine robotics is strongly aligned with the consideration of future environmental analytics projects.

4.3 Future Interest in Environmental Analytics

- Yes: 15 students
- Maybe: 5 students
- No: 0 students

100% of students indicated some interest in future environmental AI applications.

4.4 Ratings on Course Effectiveness

Question	Average Score (1-5)
Including marine robotics & environmental data was an effective teaching method	4.85
Would recommend future courses use real-world robotics data	4.8
Felt prepared with practical skills for jobs/internships	4.75

Table 2: Ratings on Course Effectiveness

Overwhelmingly positive response, confirming the value of integrating real-world robotics and environmental data into AI/data science education.

4.5 Skill Improvements from Real-World Robotics Data

A key component of the study was assessing which skills students felt they improved the most through hands-on work with real-world marine robotics data. The most frequently reported skills were:

- Exploratory Data Analysis (85% of students) - Visualizing and summarizing data using techniques like histograms and scatter plots.
- Data Cleaning and Preprocessing (70%) - Handling missing values, normalizing data, and dealing with real-world noise.
- Machine Learning Model Selection and Tuning (65%) - Applying algorithms to environmental datasets and fine-tuning hyperparameters.
- Feature Selection and Engineering (50%) - Identifying the most relevant data features for model building.
- Interpretation and Communication of Results (90%) - Explaining findings in a structured manner for decision-making.
- Working in a Team / Collaboration (80%) - Engaging in interdisciplinary teamwork to analyze and interpret datasets.

5 Discussion

Most of our results align with other studies that mention the necessity of increasing students' engagement, augmenting their motivation, having higher skill development throughout the course, and bridging the gap between the class-tailored datasets and real collected data.

5.1 Student Engagement and Motivation with Real Data

Current literature demonstrates strong links between student motivation and the use of real data. As[15] discusses, with data investigation linked to students' interests or everyday experiences, students exhibit heightened enthusiasm and commitment to the task. In our study, students consistently indicated that examining real-world environmental data from their community "made the project feel meaningful", echoing the enthusiasm seen when students are working on data about themselves or about the people with whom they work[15]. Other research has measured this motivation increase:[19] determined that students were considerably more motivated to solve a community-sourced, real-world data science problem than when given a textbook-offered, pre-packaged dataset. In our experiment, a retrospective survey reflected an increase in motivation when using authentic real-world data, highlighting motivational deficits from the use of generic data assignments[19]. Our findings reaffirm this direction, with the students in our course reporting more enjoyment and interest when working with projects involving real datasets compared to their previous experience with synthetic data examples.

5.2 Improved Motivation and Affective Response

The integration of meaningful, real-world examples appears to tap into students' intrinsic motivation. Many of our respondents noted they were "more motivated to learn" because the problems felt authentic. This qualitative feedback is supported by prior educational research. For instance, a study of a locally-focused environmental science module found that a PBL approach “stimulated middle school students’ attitudes, confidence in, and engagement with” local environmental issues[20]. Similarly, in a multi-institutional pilot where undergraduates tackled a real-world collaborative data assignment, instructors reported heightened student interest and ownership of learning[21]. The positive affective outcomes in our survey (e.g. 90% of students agreed that analyzing real data made the course more engaging) are in line with these reports. In fact, using personally relevant data is cited as one of the most important strategies for engagement – “when data are about the students themselves, they tend to be excited to undertake investigations”[15]. We also observed that motivation was sustained even when projects were challenging, which reflects what was found in a meta-analysis: compared to traditional teaching, project-based learning significantly improves students’ affective attitudes toward learning while also boosting achievement[22]. Overall, both our survey and the literature indicate that real-world data contexts make students more eager to participate and persist in STEM tasks than abstract examples do.

5.3 Learning Outcomes and Skill Development

Exposure to authentic data not only engaged students but also enhanced certain learning outcomes in our study. On objective assessments, students demonstrated improved data analysis skills. Qualitatively, students reported gains in their ability to tackle unstructured problems. This resonates with observations by[15] that working with “real” datasets introduces students to the full spectrum of statistical problem-solving, beyond what tidy textbook problems offer. Real-world data problems require our students to grapple with data cleaning, measurement error, and choosing appropriate methods – experiences that mirror authentic scientific inquiry. Such practice can yield “deep statistical understandings” when the projects are relevant. Indeed, several students in our course noted in the focus group that they better appreciated the complexity of real data and felt more confident in applying analysis techniques afterward.[17] similarly report that a data-first, real-data course effectively achieved desired learning outcomes and even narrowed the performance gap between high- and low-performing students. Real-data experiences can uplift less-prepared learners. This trend, also observed in an ecology course where only the bottom half of students showed significant improvement in graph interpretation skills[18], suggests that authentic data work can build competency, particularly in those who start with lower confidence or experience.

Our survey data indicated a significant increase in self-reported confidence in analyzing data. This positive shift in confidence is corroborated by prior findings. In[18], the authors noted that after a semester of intensive use of real data, students reported higher confidence in dealing with quantitative information. Especially, this boost in confidence can occur even when actual performance gains are modest, implying that encountering and overcoming real-data challenges reinforces students’ self-efficacy. In our course, even when some project results were imperfect, students gained assurance that they could approach complex, messy problems – an outcome also emphasized by educators advocating for realism in data science education. Addressing real-world complexity (e.g. missing data, noisy measurements) in a supportive classroom setting helps validate students’ abilities; acknowledging and working through data challenges can help renew a student’s confidence in applying their knowledge[23]. By the end of the term, open-ended feedback from our participants frequently mentioned confidence –“I feel more prepared to analyze real datasets on my own" – reflecting a key learning outcome of experiential data exercises.

5.4 Real-World vs. Synthetic Data: Impact on Learning

The contrast between using real-world versus synthetic (or heavily simplified) data in education is found in the literature and is tackled by involving students in the data acquisition process. Students in our survey overwhelmingly preferred working with real datasets; 85% agreed that "using real data made the assignments more worthwhile." Only a small minority found real data frustrating, and interestingly they suggested that the data complexity was initially intimidating. This mixed sentiment highlights a known pedagogical trade-off. Simplified or simulated datasets can lower cognitive load for beginners, but they often fail to capture students' interest and do not build real-world problem-solving skills. In[19], the authors discuss this tension: educators often use "sanitized or 'canned' datasets" to teach concepts, aiming to avoid overwhelming students, yet this practice may sacrifice opportunities for students to learn how to work with messy data[19]. Our approach tried to balance this by scaffolding support (e.g. giving data cleaning tutorials) rather than resorting to fake data. The effectiveness of sticking with real data is supported by broad consensus and evidence. Using real datasets (with context) is more effective for learning than "toy" made-up data[16]. The American Statistical Association's guidelines explicitly state that "using real data in context is crucial" for teaching, as it both gives students genuine analytical experience and showcases the fascination of the discipline[16]. Our findings back this: students became adept at handling real-data issues and remained engaged, whereas, in past offerings with generic data, some students disengaged or underestimated the challenge. Moreover, when real data are not contextually relevant, their benefits diminish to the point of being no better than synthetic data[15] – a caution we heeded by ensuring relevance. In the end, the authentic data in our course functioned as a double-edged sword that, with proper support, yielded net positive outcomes: it introduced complexity but also realism, which fueled engagement and skill development. This aligns with the results of the Memphis study mentioned earlier, where phase 2 (with a community-sourced dataset) not only increased motivation but also changed the nature of student inquiries, indicating deeper engagement with material[19]. Our students likewise asked more nuanced questions and made more connections to real-world implications, compared to prior experiences with artificial data sets. We interpret this as evidence that real-world data, despite its challenges, leads to richer learning experiences and greater student confidence in applying AI/data science methods to authentic problems.

5.5 Interdisciplinary and Cross-Domain Benefits

One objective of integrating real-world datasets was to create an interdisciplinary learning experience, attracting students from various majors (e.g. computer science, biology, engineering, business) and allowing each to apply data science techniques to problems in their domain. This approach proved successful according to both our outcomes and those reported by others. In our class, students from non-CS backgrounds often brought unique perspectives to projects (such as a biology major analyzing sensor data to study wildlife activity). These students reported that the real-world context helped them see the relevance of data science in their own field, which increased their engagement. This evidence shows that a project-based statistics curriculum can provide a common language for approaching questions across numerous disciplines, leveraging students' natural curiosity in their subject area while teaching quantitative skills [14]. We found that all student subgroups –regardless of major or prior experience– benefited from the course. Final project evaluations and survey feedback were uniformly positive across disciplines, echoing the high rates of satisfaction observed in multidisciplinary courses in the literature. For instance, [14] report that over 80% of students (across different ethnic groups and majors) found their semester-long research project worth the effort and rewarding, underlining the broad appeal of such real-data projects. Our survey similarly showed that >80% of both STEM and non-STEM majors felt the real-data project was valuable for their learning.

Moreover, involving diverse academic backgrounds enriched the learning environment. Students had to explain concepts to teammates outside their field, which improved communication skills and reinforced their understanding. This aligns with best practices in interdisciplinary education, where collaboration across domains is known to build inclusive and flexible thinking in students. We also saw evidence that the interdisciplinary format can reduce performance gaps. In particular, some of our strongest project groups had people with mixed majors (e.g. a business student paired with an engineering student), and the peer learning in those teams helped less-experienced students gain competence quickly. In [17], the authors observed a similar phenomenon: in their real-data course, the structure contributed to narrowing the perceived gap between low- and high-performing students. In our case, by the end of the term, there was no significant difference in project scores between students with prior programming experience and those without, suggesting that the collaborative, real-data approach allowed novices to catch up. This is a critical outcome for inclusivity: interdisciplinary real-world projects can level the playing field and give students from diverse backgrounds the confidence that they can succeed in AI and data science. One student wrote, "As an environmental science major, I never thought I could do a machine learning project, but now I have the tools to apply data analysis in my field," exemplifying the empowered attitude we hoped to cultivate.

5.6 Best Practices

Integrating real-world data in education:

- Use datasets connected to students' interests or local community issues. Data that students can personally connect with elicit greater engagement and motivation. Conversely, if real data lack meaningful context to students, they may offer little advantage over synthetic examples. [15] [19]
- Wherever possible, involve students in choosing or collecting the data. Giving learners a voice in selecting problems (or using data about themselves) fosters ownership of the project and has been shown to excite students about the analysis process. [15]
- Structure the course around authentic projects and teamwork. PBL contexts –especially those tackling real societal or scientific problems– consistently lead to higher enthusiasm and perseverance. Collaboration allows students to learn from peers with different strengths, which can boost confidence for those from non-traditional backgrounds. [24] [20]
- Real-world data can be messy and complex, so provide scaffolding (tutorials, mentorship, incremental tasks) to prevent students from feeling overwhelmed. Educators should address

data wrangling and interpretation challenges openly in class; this normalizes the difficulties and builds students' confidence as they develop strategies to overcome them.[19,23]

- Encourage students to apply data science methods to problems in their own domain. This not only makes the learning more relevant but also improves their ability to communicate and transfer knowledge across fields. Cases show that such interdisciplinary projects produce gains in both domain understanding and analytical skill, and are valued by students from all majors.[14]

By following these practices, instructors can maximize the educational benefits of real-world data. In summary, our study's results, viewed alongside prior surveys and experiments, reinforce a clear message: embedding authentic data experiences in AI and data science education leads to more engaged, motivated learners who are better prepared to apply their skills in diverse, real-world contexts. The positive correlations between real-data use and student engagement, confidence, and skill development observed here provide strong evidence to support the further adoption of experiential, interdisciplinary approaches in STEM curricula. Each data-driven project that students tackle not only teaches core concepts but also inspires them with the realization that what they learn in the classroom has direct relevance beyond it.

5.6.1 Pedagogical Strategies for Machine Learning Instruction

The integration of real-world datasets, such as those collected from Biscayne Bay, offers students an authentic learning experience. The following steps outline a structured pedagogical framework that can be used to teach introductory Machine Learning courses such as *Data Science for All*, *AI for All*, among others:

- **Data Acquisition and Exploration** - Provide students with datasets such as the ones in this paper and guide them through data wrangling using Python libraries like Pandas, Numpy, and Scikit-Learn.
- **Model Building and Analysis** - Assign students tasks to build and compare ML models (e.g., LR, RF) and evaluate them using metrics like MSE and R-squared.
- **Critical Thinking Exercises** - Encourage discussions on model selection, limitations, and ethical considerations in AI applications.

5.6.2 Sample Classroom Assignment: Predicting Water Quality Parameters

Objective: Educate students about how to predict water quality parameters using real-world data and assess the potential impact of sensor faults on predictive models.

Assignment Structure

- Task 1: Import and preprocess the provided Biscayne Bay dataset, focusing on handling missing values and ensuring data integrity. This highlights the importance and the methods available to deprecate or fill missing values including the one proposed in this paper as a reliable way to fill corrupted or absent values.

- Task 2: Build and evaluate three machine learning models: LR, RF, and MLP. Compare the models based on performance metrics and highlight strengths and weaknesses.
- Task 3: Generate visualizations (e.g., scatter plots, residual plots) to compare predicted vs. actual values and identify patterns or anomalies.

Deliverables Students are required to submit a comprehensive report that includes:

- Annotated code with explanations of key steps.
- Visualizations and results analysis.
- A discussion of findings, including limitations and future improvement suggestions.

Learning Outcomes This group project can be conducted with teams of 3–4 students, and the expected learning outcomes are

- Practical experience in applying machine learning to environmental science.
- Improved teamwork and communication skills through cross-disciplinary collaboration.
- A deeper understanding of the ethical and societal implications of machine learning in addressing global challenges.

This teaching framework provides students from diverse backgrounds—computer science, engineering, and environmental science—with the knowledge and tools to apply machine learning effectively in solving real-world problems

6 Conclusion

The results of this study underscore the significant benefits of integrating real-world data into data science education. The use of marine robotics datasets augmented student engagement, confidence, and motivation across all levels of experience. Particularly, students without prior robotics background showed substantial learning gains, emphasizing the accessibility and effectiveness of this approach.

Both computer science and non-computer science students found value in working with real-world environmental data, emphasizing the interdisciplinary appeal of data-driven environmental science. Exposure to robotics hands-on tasks leads to higher post-course understanding, but all students, regardless of their initial familiarity, showed notable improvements in data science competencies.

Finally, surveyed students expressed interest in environmental AI applications, highlighting the increasing need to incorporate practical, real-world datasets in education. These results provide strong support for the continued development of curriculum strategies that bridge theoretical knowledge with hands-on experience. Future research should address the extension of these approaches across larger student populations and different disciplines to maximize their impact on data science education and spark the students' curiosity and engagement.

References

[1] Prerna Ravi, Robert Parks, John Masla, Hal Abelson, and Cynthia Breazeal. ”data comes from the real world”: A constructionist approach to mainstreaming k12 data science education. In *Proceedings of the 2024 on ACM Virtual Global Computing Education Conference V. 1*, SIGCSE Virtual 2024, page 271–274, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400705984. doi: 10.1145/3649165.3703623. URL `https://doi.org/10.1145/3649165.3703623`.

[2] Koby Mike, Benny Kimelfeld, and Orit Hazzan. The Birth of a New Discipline: Data Science Education. *Harvard Data Science Review*, 5(4), oct 27 2023. https://hdsr.mitpress.mit.edu/pub/by22efhx.

[3] Laura Mebert, Roy Barnes, Jacqueline Dalley, Leszek Gawarecki, Farnaz Ghazi-Nezami, Gregory Shafer, Jill Slater, and Erin Yezbick. Fostering student engagement through a real-world, collaborative project across disciplines and institutions. *Higher Education Pedagogies*, 5(1):30–51, 2020. doi: 10.1080/23752696.2020.1750306. URL `https://doi.org/10.1080/23752696.2020.1750306`.

[4] NOAA NESDIS. Bridging the ‘last mile’: Turning environmental data into effective classroom tools. *NOAA Reports*, page 13, December 2024. URL `https://www.nesdis.noaa.gov/news/bridging-the-last-mile-turning-environmental-data-effective-classroom-tools`.

[5] David L Neumann, Michelle Hood, and Michelle M Neumann. Using real-life data when teaching statistics: Student perceptions of this strategy in an introductory statistics course. *Statistics Education Research Journal*, 12(2):59–70, 2013.

[6] Brandon Lawton, Mary Kay Hemenway, Bryan Mendez, and Sten Odenwald. Using nasa data in the classroom: Promoting stem learning in formal education using real space science data. In *Communicating Science: A National Conference on Science Education and Public Outreach*, volume 473, page 117, 2013.

[7] S. M. Mizanoor Rahman. Assessing and benchmarking learning outcomes of robotics-enabled stem education. *Education Sciences*, 11(2), 2021. ISSN 2227-7102. doi: 10.3390/educsci11020084. URL `https://www.mdpi.com/2227-7102/11/2/84`.

[8] Kai Wang, Guo-Yuan Sang, Lan-Zi Huang, Shi-Hua Li, and Jian-Wen Guo. The effectiveness of educational robots in improving learning outcomes: A meta-analysis. *Sustainability*, 15(5), 2023. ISSN 2071-1050. doi: 10.3390/su15054637. URL `https://www.mdpi.com/2071-1050/15/5/4637`.

[9] J. Zhang and H. Zhu. A hands-on approach to teaching machine learning using real-world datasets. *Journal of Educational Technology Systems*, 50(1):75–95, 2021. doi: 10.1177/00472395211001326.

[10] V.B. Kolachalama and P.S. Garg. Machine learning and medical education. *NPJ Digital Medicine*, 1(54), 2018. doi: 10.1038/s41746-018-0061-1.

[11] L. Sun and J. Wu. Using a project-based learning approach to teach artificial intelligence to undergraduate students. *IEEE Access*, 10:1–10, 2022. doi: 10.1109/ACCESS.2022.3149103.

[12] M. Anderson, J. Weinhardt, and J. Zhang. Enhancing student engagement in data science education using real-world examples. *Data Science and Education Journal*, 4(3):45–61, 2019. doi: 10.1145/3325806.3325812.

[13] Roozbeh Aliabadi. *The Impact of an Artificial Intelligence Project-Based Learning Course on Middle-School Students’ Interest, Knowledge, and Career Aspiration in AI*. PhD thesis, Robert Morris University, 2023.

[14] Lisa Dierker, Emmanuel Kaparakis, Jennifer Rose, Arielle Selya, and David Beveridge. Strength in numbers: A multidisciplinary, project-based course in introductory statistics. *Journal of Effective Teaching*, 12(2):4–14, 2012.

[15] Jennifer Hall. *Engaging Teachers and Students with Real Data: Benefits and Challenges*, volume 14, pages 335–346. 06 2011. ISBN 978-94-007-1130-3. doi: 10.1007/978-94-007-1131-0_32.

[16] Ann M Brearley, Carol Bigelow, Laila M Poisson, Steven C Grambow, and Amy S Nowacki. The tshs resources portal: A source of real and relevant data for teaching statistics in the health sciences. *Technology Innovations in Statistics Education*, 11(1), 2018.

[17] Luke Bornn, Jacob Mortensen, and Daria Ahrensmeier. A data-first approach to learning real-world statistical modeling. *The Canadian Journal for the Scholarship of Teaching and Learning*, 13(1), 2022.

[18] Judith Bramble and Margaret Workman. Data-rich case studies improve students' abilities to interpret graphs in a large non-majors course. *Teaching Issues and Experiments in Ecology*, 5, 2007.

[19] Jill Johnson and Andrew M Olney. Using community-based problems to increase motivation in a data science virtual internship. In *Proceedings of the 15th International Conference on Educational Data Mining*, 2022.

[20] Andrea Basche, Vincent Genareo, Adah Leshem, Amy Kissell, and Judith Pauley. Engaging middle school students through locally focused environmental science project-based learning. *Natural Sciences Education*, 45, 12 2016. doi: 10.4195/nse2016.05.0012.

[21] Miguel A Cardona, Roberto J Rodríguez, and Kristina Ishmael. Artificial intelligence and the future of teaching and learning. *Department of Education*, 2023.

[22] Lu Zhang and Yan Ma. A study of the impact of project-based learning on student learning effects: A meta-analysis study. *Frontiers in psychology*, 14:1202728, 2023.

[23] Candace Savonen, Carrie Wright, Ava Hoffman, Elizabeth Humphries, Katherine Cox, Frederick Tan, and Jeffrey Leek. Motivation, inclusivity, and realism should drive data science education. *F1000Research*, 12: 1240, 2024.

[24] Francesca Bertacchini, Carmelo Scuro, Pietro Pantano, and Eleonora Bilotta. A project based learning approach for improving students' computational thinking skills. *Frontiers in Robotics and AI*, 9:720448, 2022.

Appendix A: Survey Instrument and Responses

Survey Questions and Response Options:

1. Which course(s) were you enrolled in that used water quality data? (Multiple choice)
   - a) Human-Computer Interaction
   - b) Data Science
   - c) Marine Robotics Workshop
2. What is your major or primary field of study? (Single choice)
   - a) Computer Science
   - b) Biological Sciences
   - c) Marine Biology
   - d) Data Science and AI
   - e) Math/Stats
3. What year are you in your studies? (Single choice)
   - a) Freshman
   - b) Sophomore
   - c) Junior
   - d) Senior
   - e) Graduate (Master's or PhD)
4. Before this course, how familiar were you with the use of robotics in environmental data collection? (Single choice)
   - a) Not at all familiar
   - b) Slightly familiar
   - c) Somewhat familiar
   - d) Very familiar
5. Before this course, how familiar were you with Biscayne Bay's environmental issues or water-quality monitoring efforts? (Single choice)
   - a) Not at all familiar
   - b) Slightly familiar
   - c) Somewhat familiar
   - d) Very familiar
6. How would you rate your understanding of how marine robotics collect environmental data after this course module? (Single choice)
   - a) No understanding
   - b) Basic understanding
   - c) Good understanding
   - d) Thorough understanding
7. Engagement and Motivation (Likert scale: Strongly Agree, Somewhat Agree, Neutral, Somewhat Disagree, Strongly Disagree)
   - Using real-world data from Biscayne Bay made the class more engaging.
   - Knowing the data was collected by marine robots enhanced my interest in the subject matter.
   - Using real-world environmental data helped me better understand key data science/machine learning concepts.

- Interpreting data from Biscayne Bay made analytical concepts more concrete and easier to grasp.
- Working with actual environmental datasets improved my confidence in applying data science techniques.
- I gained a deeper appreciation for the complexity of real-world data.

8. Which of the following skills do you feel you improved the most by working with real-world marine robotics data? (Select all that apply)
    - Data cleaning and preprocessing
    - Exploratory data analysis (visualizations, summary statistics)
    - Feature selection / Feature engineering
    - Machine learning model selection and tuning
    - Interpretation and communication of results
    - Working in a team / collaboration
    - All of the above
9. Do you think the inclusion of real-world environmental data in this course has made you more likely to pursue or consider projects related to environmental analytics or marine robotics in the future? (Single choice)
    - a) Yes
    - b) No
    - c) Maybe
10. I believe including marine robotics and environmental data is an effective teaching method for data science/machine learning. (Scale 1-5)
11. I would recommend that future data science courses continue to incorporate real-world robotics-collected environmental data. (Scale 1-5)
12. I feel this approach prepared me with practical skills I can use in internships, research, or jobs. (Scale 1-5)