

On Transdisciplinary Research through Data Science and Engineering Education

Junwhan Kim

Computer Science and Information Technology
University of the District of Columbia
Washington, DC, USA
junwhan.kim@udc.edu

Abstract

The paper highlights the challenges non-computer science professionals encounter when managing and analyzing heterogeneous data sources. To address these issues, it details the innovative learning methodologies employed in Data Science and Engineering (DSE) courses at the University of the District of Columbia. These courses are specifically designed to equip students from diverse disciplines with the skills needed to effectively apply DSE techniques within their respective fields. The outcomes underscore the transformative power of DSE education in fostering transdisciplinary research, enhancing the research capabilities of both computer science and non-computer science students, and driving innovation and scientific discovery across a broad spectrum of domains.

CCS Concepts: • Computing methodologies → Machine learning approaches; • Applied computing → Collaborative learning.

Keywords: Transdisciplinary Research, Data Science and Engineering Education

ACM Reference Format:

Junwhan Kim. 2024. On Transdisciplinary Research through Data Science and Engineering Education. In *Proceedings of ICECT24*. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 Introduction

Data science is an interdisciplinary field that integrates statistical, computer science, and mathematical principles to acquire, analyze, and extract valuable insights from large

datasets [6]. Closely related to this is data engineering, which focuses on building the infrastructure, interfaces, and mechanisms that enable the seamless flow and access of data [8]. While these two disciplines are distinct, they are deeply interconnected and often collaborate on complex projects. Data engineers are responsible for preparing and organizing data, creating the foundational systems that data scientists rely on to analyze and interpret the information, ultimately driving informed decision-making and innovation.

The expansive scope of data science and engineering (DSE) has led to the proliferation of data-driven applications across various fields. In non-DSE domains like finance and agriculture, DSE techniques serve as essential tools for making data-based decisions. However, the process of collecting valuable data from multiple, heterogeneous sources poses significant challenges and costs for researchers. More fundamentally, individuals and groups outside the realms of computer science and engineering may lack the expertise to effectively handle these challenges. Applying DSE techniques to other data-intensive research fields offers a promising opportunity to accelerate scientific discovery and technological innovation. Consequently, DSE education emphasizes research opportunities and student training to address these interdisciplinary challenges [4].

Despite the potential benefits, the contributions of DSE students to scientific discoveries and technological innovations in other research fields may be limited without relevant and domain-specific expertise. To support and enhance DSE initiatives, various sectors—government, private, and public—encourage education and direct experience to equip individuals with the necessary qualifications and knowledge for success. Students trained solely in traditional DSE curricula may lack the interdisciplinary knowledge needed to address challenges in other domains. Motivated by our experiences, we have developed four DSE-related courses focusing on transdisciplinary fields.

These courses—data science, data engineering, advanced DSE, and vision AI—are offered by the Department of Computer Science and Information Technology at the University of the District of Columbia. We have observed the performance of students from diverse disciplines, including mechanical, biomedical, civil, and computer engineering, agriculture, law and business, who have enrolled in these courses.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org. ICECT24, 979-8-4007-1781-9

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

This paper highlights the transdisciplinary research conducted by these students and demonstrates the impact of DSE in bridging disciplines to enhance research skills for both computer science and non-computer science students.

The organization of this paper is as follows: Section 2 reviews related work regarding DSE education; Section 3 presents our learning methodologies for DSE-related courses; Section 4 discusses 36 transdisciplinary research topics conducted in the courses; and Section 5 concludes with our findings and future work.

2 Related Work

Biehler *et al.* focus on addressing the challenges of incorporating data science into secondary school curricula [2]. They explore the necessary competencies for teaching data science, the ethical and societal implications of big data, and the gap between traditional computer science education and the interdisciplinary demands of data science. The paper questions how current educational models in computer science need to evolve to include essential elements such as data management, visualization, and machine learning. Additionally, it emphasizes the importance of discussing societal issues like data privacy and the ethical use of algorithms within the curriculum. The authors argue that while technical skills, such as algorithmic thinking, are important, data science education must also encompass broader discussions about the societal impact of data. Ultimately, they conclude that integrating data science at the school level should not solely aim at producing data scientists, but should also equip students with the ability to critically engage with data-driven decision-making in society.

Similar to this paper [2], we place a strong emphasis on the societal implications of data use, especially because students outside of computer science often have limited opportunities to explore the ethical and societal dimensions of working with data. Our curriculum also incorporates content that focuses on the societal impact of data-driven decision-making and how scientific insights affect our society.

The author, Nicolescu, highlights key learning methods essential for transdisciplinary education, focusing primarily on the concept of "learning to know," which emphasizes the ability to discern reality from illusion through a scientific spirit grounded in constant questioning and open-minded inquiry [7]. Nicolescu stresses that education should not merely involve the accumulation of facts, but should also foster the development of students' capacities to create connections between different disciplines and bridge the gap between external knowledge and their internal abilities, integrating intellect, emotion, and physical experience. This approach seeks to cultivate flexibility, creativity, and adaptability, allowing individuals to navigate a rapidly changing world. He proposes practical steps to implement transdisciplinary learning, such as dedicating 10% of teaching time to

transdisciplinary studies, creating transdisciplinary research ateliers, and encouraging dialogues between disciplines, including the sciences and the arts. Ultimately, these methods aim to form individuals who can dynamically adjust to new challenges and integrate their learning into creative problem-solving and innovative thinking.

Based on this educational research [7], we combined multiple existing learning methods to accelerate transdisciplinary education, as our curriculum does not require specific prerequisites. This approach enables non-computer science students to engage in and advance their research more effectively. By fostering an inclusive environment, we encourage synergetic research collaborations that enhance problem-solving capabilities, allowing students from diverse backgrounds to contribute meaningfully to complex, multidisciplinary challenges.

Hazzan and Mike address the emerging need for a dedicated journal to cover interdisciplinary data science education [3]. The authors aim to solve the problem of fragmentation in data science education, where current research and resources are scattered across various disciplines such as computer science, mathematics, statistics, and domain-specific fields like the natural and social sciences. They argue that no existing journal solely focuses on data science education from an interdisciplinary standpoint. Through the proposed journal, the authors seek to unite the diverse facets of data science education under one roof, fostering collaboration among educators across disciplines. We strongly support this approach with a comprehensive understanding of data science's interdisciplinary nature, which is crucial for data-oriented applications. In addition, we emphasize the importance of transdisciplinary education for enabling deeper research activities, a point further highlighted in this paper.

Sahneh *et al.* emphasizes the importance of fostering transdisciplinary collaboration in data science by cultivating what has been termed "collaboratory cultures." These cultures prioritize the interpersonal dynamics between researchers and data scientists, recognizing that successful collaboration goes beyond technical expertise [10]. The study introduces ten simple rules designed to enhance transdisciplinary collaboration, with a focus on nontechnical aspects such as reflexive practices, effective communication, and project design that ensures equitable benefits for all participants. These guidelines are essential for creating environments where diverse disciplinary perspectives can converge, ultimately leading to novel insights and accelerating scientific discovery.

Building on this foundation [10], our work develops a new program centered on DSE research across diverse fields. While inspired by prior studies on collaboration, our approach shifts the focus from fostering collaborative research to promoting collaborative learning. We emphasize the development of leadership skills, encouraging students to take ownership of their research and lead as independent scholars.

Rather than merely refining collaboration skills or enhancing the quality of research, this program equips students to lead transdisciplinary research efforts, preparing them to navigate and integrate diverse perspectives while driving innovation in their respective fields.

Unlike the aforementioned efforts, our approach integrates data science and data engineering courses with a strong emphasis on fostering transdisciplinary research in DSE. By offering learning methodologies and research topics that span a broader range of disciplines, we enable students to engage with diverse fields and address complex, real-world problems. This paper demonstrates how our methodologies impact DSE education, highlighting the effectiveness of transdisciplinary learning in equipping students with the skills needed to lead research across multiple domains and contribute to advancements in both data science and engineering.

3 Learning Methodologies for DSE Courses

The learning methodology for DSE education is multifaceted, integrating theoretical knowledge, practical skills, and transdisciplinary approaches to prepare students for the complex challenges of modern data-driven environments. This section outlines a comprehensive learning methodology that incorporates project-based learning, experiential learning, peer teaching learning, and direct lectures.

Each DSE course employs four primary learning methods:

1. **Direct Lectures:** The cornerstone of traditional classroom teaching, direct lectures provide students with essential theoretical knowledge and practical skills. Recognizing that some students, particularly those from non-computer science backgrounds, may lack programming experience, these lectures are supplemented with online resources to help students acquire foundational programming skills. This ensures that all students, regardless of their initial proficiency, can grasp the essential concepts of DSE.
2. **Peer Teaching [9]:** An effective method for reinforcing learning is through peer teaching, where students teach concepts and skills to their classmates. This not only strengthens their own understanding but also enhances their communication and pedagogical abilities. Given that many DSE students come from diverse academic backgrounds, peer teaching facilitates the sharing of domain-specific knowledge. For instance, students specializing in agriculture can teach others about analysing ammonia datasets critical for agricultural automation, such as controlling real-time ammonia levels in aquaponics systems. This method underscores how DSE bridges various disciplines, enriching the learning experience for all students.
3. **Experiential Learning [5]:** Experiential learning involves hands-on activities that simulate real-world

tasks and challenges. Following lectures and peer teaching, students engage in lab sessions where they apply DSE techniques using data science tools and technologies. These practical experiences enable students to deepen their understanding of datasets and refine their technical skills, preparing them for real-world applications.

4. **Project-Based Learning (PBL) [1]:** PBL is a core component of DSE education, requiring students to work on real-world projects that integrate theoretical knowledge and practical skills. Students undertake final projects based on their own research data or knowledge, navigating the entire DSE process from data collection and processing to exploratory analysis, data visualization, hypothesis testing, and deriving insights for policy decisions. While these projects are typically conducted individually to ensure comprehensive understanding, they still reflect the interdisciplinary nature of DSE by addressing real-world problems.

The integration of these learning methods ensures that students are well-equipped to tackle the complexities of DSE. Figure 1 illustrates the learning methodologies employed in each course, highlighting the balanced approach that combines traditional and innovative teaching strategies to foster a robust educational experience.

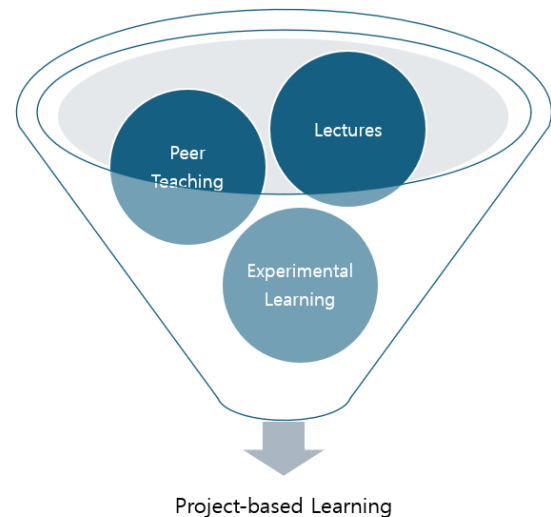


Figure 1. Learning Methodologies for DSE

4 Transdisciplinary Research in DSE

This section highlights the transdisciplinary research carried out within the DSE courses. The following research topics have emerged from our DSE courses, divided into two categories: those undertaken by computer science students and those by non-computer science students.

4.1 DSE Research by Computer Science Students

4.1.1 Sentiment Analysis Based on Social Media Data.

- Real-time Sentiment Analysis for Customer Feedback Using Social Media Data
Dataset: Social media posts, customer reviews, and feedback data
Techniques Used: Natural Language Processing (NLP) techniques such as tokenization, sentiment lexicons, and machine learning classifiers – SVM, Decision Tree, LSTM to analyze customer sentiment and identify trends in feedback
- Predicting Customer Churn Using Sentiment Analysis on Social Media Data
Dataset: Social media data, customer reviews, and transactional data related to product usage
Techniques Used: Sentiment analysis combined with predictive modeling techniques like logistic regression, decision trees, and random forests to identify customers at risk of churn based on their sentiments expressed on social media
- Assessing the Risk of High-Profile Campaigns Through Sentiment Analysis
Dataset: Social media mentions, campaign-specific hash-tags, and public opinion data
Techniques Used: Sentiment analysis using NLP, topic modeling (e.g., LDA), and risk assessment models to gauge public sentiment and potential risks associated with high-profile campaigns
- Predicting Presidential Election Outcomes Through Social Media Sentiment Analysis
Dataset: Social media data and election-related hash-tags from Twitter
Techniques Used: Sentiment analysis, time series analysis, and predictive modeling techniques such as support vector machines (SVM) and neural networks to predict election outcomes based on public sentiment trends
- Impact of Financial News on Market Trends: A Sentiment Analysis Approach
Dataset: Financial news articles, stock market data, and social media mentions
Techniques Used: NLP techniques for sentiment extraction, coupled with econometric models and machine learning algorithms to correlate news sentiment with stock market movements
- Sentiment Analysis of Social Media Reactions to Movie Trailers and Box Office Prediction
Dataset: Social media reactions, YouTube comments, and box office revenue data
Techniques Used: Sentiment analysis, emotion detection using NLP, and regression models to predict box office success based on trailer reactions

4.1.2 Natural Language Processing (NLP).

- Fine-Tuning Large Language Models (LLMs) for Domain-Specific Text Generation
Dataset: Domain-specific text corpora relevant to the target application
Techniques Used: Transfer learning, supervised fine-tuning of LLMs (e.g., Llama3, BERT), and optimization techniques like hyperparameter tuning to enhance model performance in specialized tasks
- Assessing and Modifying Personality Traits in LLMs
Dataset: Text data annotated with personality traits, conversational data.
Techniques Used: NLP techniques for personality assessment using the Big Five personality traits, model interpretability techniques, and fine-tuning to adjust model outputs based on desired personality traits
- Enhancing Large Language Models with Retrieval-Augmented Generation (RAG)
Dataset: Large text corpora, relevant external databases for retrieval
Techniques Used: Combining retrieval mechanisms with generative models, using RAG architecture, and incorporating knowledge retrieval with LLMs to improve accuracy and relevance of generated content

4.1.3 Anomaly Detection.

- Real-Time Anomaly Detection for Fraud Detection in Financial Transactions
Dataset: Financial transaction data, labeled datasets with fraudulent and non-fraudulent transactions.
Techniques Used: Machine learning algorithms such as decision trees, random forests, and deep learning models (e.g., autoencoders) for real-time anomaly detection, along with unsupervised learning techniques for identifying new fraud patterns.
- Securing Systems Through Anomaly Detection in System Calls
Dataset: System logs, process call data, and network traffic logs.
Techniques Used: Sequence-based anomaly detection methods, Hidden Markov Models, and deep learning approaches like LSTM/Transformer for detecting deviations in system call sequences indicative of attacks
- Network Traffic Analysis for Real-Time Anomaly Detection in Cybersecurity
Dataset: Network traffic data, labeled data with normal and abnormal traffic patterns
Techniques Used: Clustering algorithms (e.g., K-means), deep learning models like CNNs for feature extraction, and anomaly detection techniques (e.g., Isolation Forests) for identifying suspicious network activity

4.1.4 Disinformation.

- Automated Fact-Checking Using Natural Language Processing

Dataset: News articles, social media posts, fact-checked statements, and knowledge bases

Techniques Used: NLP techniques for information extraction, claim detection, and fact-checking using knowledge bases and pretrained LLMs to verify claims against reliable sources

- Detecting Disinformation on Social Media Platforms Using Machine Learning

Dataset: Social media posts, news articles, and labeled datasets with disinformation and factual content.

Techniques Used: Machine learning classifiers, text classification techniques, and deep learning models like BERT for detecting and classifying disinformation

- Assessing the Risk of Disinformation in Online Information Ecosystems

Dataset: Social media data, network analysis data, and labeled datasets of disinformation campaigns.

Techniques Used: Network analysis, sentiment analysis, and predictive modeling techniques to assess the spread and impact of disinformation and identify high-risk information networks

4.2 DSE Research by Non-Computer Science Students

4.2.1 Healthcare.

- Predictive Modeling for Early Disease Detection Using Patient Data
Dataset: Electronic health records (EHR), clinical datasets, and genetic data.
Techniques Used: Deep learning models (CNN and Vision Transformer) for predictive analytics, feature selection, and model interpretability to identify early indicators of diseases
- Personalized Medicine Through Genomic Data Analysis
Dataset: Genomic datasets, patient medical history, and treatment outcome data
Techniques Used: Machine learning models for clustering patients based on genetic profiles, predictive modeling for treatment response, and precision medicine techniques to tailor treatments to individual patients.
- Reducing Hospital Readmissions Using Predictive Analytics
Dataset: Hospital readmission records, patient EHR, and demographic data
Techniques Used: Logistic regression, decision trees, and deep learning models for predicting readmission risk, with feature engineering to identify key predictors of readmission.
- Building Scalable Data Infrastructure for Genomic Data Processing
Dataset: Large-scale genomic datasets, cloud storage systems, and computational resources.

Techniques Used: Distributed computing frameworks (e.g., Apache Spark), cloud-based storage solutions, and parallel processing techniques to handle and analyze large volumes of genomic data

- Brain Tumor Detection in MRI Images Using YOLO
Dataset: MRI images of brain tumors, annotated image datasets

Techniques Used: Convolutional Neural Networks (CNNs), specifically the YOLO (You Only Look Once) architecture, for real-time object detection, image segmentation, and classification to accurately detect and localize brain tumors

4.2.2 Agriculture and Environment.

- Optimizing Water Usage in Agriculture Through Precision Irrigation Using IoT and Data Analytics
Dataset: Sensor data from soil moisture sensors, weather data, and crop growth datasets
Techniques Used: IoT for real-time data collection, machine learning models for predictive irrigation scheduling, and data analytics for optimizing water usage based on crop needs and environmental conditions
- Analyzing the Nitrogen Cycle in Aquaponics Systems Using Data Science Techniques
Dataset: Sensor data from aquaponics systems, water quality data, and plant growth metrics.
Techniques Used: Time series analysis, statistical modeling, and machine learning techniques to model the nitrogen cycle and optimize nutrient management in aquaponics systems
- Predictive Modeling and Data Analysis for Climate Change Impact and Weather Forecasting
Dataset: Climate datasets, historical weather data, and remote sensing data
Techniques Used: Time series forecasting, deep learning models and statistical analysis to predict climate change impacts, weather patterns, and their effects on agriculture
- Enhancing Food Supply Chain Efficiency Through Data Analytics
Dataset: Supply chain data, transportation logs, and inventory data
Techniques Used: Machine learning models for demand forecasting, optimization algorithms for supply chain management, and data analytics to reduce waste, optimize logistics, and improve supply chain efficiency

4.2.3 Law.

- Using Predictive Analytics to Forecast Legal Case Outcomes
Dataset: Historical case data, legal documents, and court rulings
Techniques Used: Machine learning classifiers, natural language processing for document analysis, and

predictive modeling to estimate the likelihood of case outcomes based on historical data

- **Automating Contract Generation Using Natural Language Processing and Machine Learning**
Dataset: Legal contract templates, annotated legal clauses, and case law
Techniques Used: NLP for contract clause generation, machine learning models for template selection and customization, and rule-based systems for ensuring legal compliance
- **Data-Driven Analysis of Mergers, Acquisitions, and Investments Using Due Diligence Data**
Dataset: Due diligence reports, financial data, and market analysis reports
Techniques Used: Text mining, sentiment analysis, and machine learning models for risk assessment, valuation, and decision support in mergers and acquisitions
- **Crime Trend Analysis and Geospatial Mapping Using Data Science Techniques**
Dataset: Crime reports, geospatial data, and demographic data
Techniques Used: Geospatial analysis, clustering algorithms, and machine learning models for trend analysis and hot spot detection, with a focus on identifying patterns and predicting future crime occurrences
- **Strategizing Litigation Using Data-Driven Analysis of Historical Legal Cases**
Dataset: Historical legal case data, court rulings, and legal documents
Techniques Used: NLP for text analysis, machine learning for pattern recognition, and predictive analytics to develop litigation strategies based on historical case patterns

4.2.4 Mechanical Engineering.

- **Advanced Control Systems for Robotics and Autonomous Vehicles Using Machine Learning**
Dataset: Sensor data from robotics systems, simulation data, and real-time control data
Techniques Used: Reinforcement learning, model predictive control, and deep learning algorithms to enhance the performance of robotic systems and autonomous vehicles
- **Data-Driven Approaches for Sustainable Energy Systems and Thermal-Fluid Management**
Dataset: Energy consumption data, fluid dynamics simulation data, and environmental data
Techniques Used: Computational fluid dynamics (CFD), machine learning models for energy optimization, and sustainability metrics analysis to develop and improve sustainable energy systems
- **Integrating AI in Design and Advanced Manufacturing for Intelligent Systems**
Dataset: CAD models, manufacturing process data,

and material property datasets

Techniques Used: AI-driven design optimization, machine learning for predictive maintenance, and advanced manufacturing techniques and additive manufacturing for intelligent systems integration

4.2.5 Civil Engineering.

- **Structural Health Monitoring Using Data from Embedded Sensors in Civil Infrastructure**
Dataset: Sensor data from bridges and buildings, structural models, and environmental data
Techniques Used: Time series analysis, machine learning models for anomaly detection, and predictive maintenance algorithms to ensure the safety and longevity of civil infrastructure
- **Optimizing Urban Infrastructure Management Using Data-Driven Approaches**
Dataset: Urban infrastructure data, traffic data, and environmental impact data
Techniques Used: Data analytics, optimization algorithms, and machine learning models to improve the efficiency, sustainability, and resilience of urban infrastructure.
- **Environmental Impact Assessment of Construction Projects Using Data Analytics**
Dataset: Environmental impact data, construction project data, and regulatory compliance data.
Techniques Used: Statistical analysis, environmental modeling, and machine learning techniques for impact assessment and mitigation planning in construction projects
- **Geotechnical Data Analysis for Improved Foundation Design and Risk Assessment**
Dataset: Soil and rock sample data, geotechnical survey data, and environmental conditions data.
Techniques Used: Statistical modeling, geospatial analysis, and machine learning techniques to analyze geotechnical data and optimize foundation design for construction projects

5 Findings and Future Work

DSE courses engage with diverse datasets from various domains, emphasizing the critical role of domain expertise in extracting meaningful insights. This paper highlights that, unlike traditional computer science education, DSE research significantly benefits from the integration of multidisciplinary expertise. Our approach—blending lectures, peer teaching, experiential learning, and project-based learning—has proven effective in equipping students with the necessary skills for DSE.

While traditional computer science education has focused primarily on programming, education for non-computer science students has centered on using tools and languages to implement ideas. As LLMs increasingly take over routine

coding tasks, a fundamental understanding of programming remains essential for refining and evaluating AI-generated code. DSE education addresses this challenge by laying a strong foundation in data-driven research and fostering collaboration between computer science and non-computer science students, encouraging a more synergistic approach to problem-solving.

Through transdisciplinary research activities in DSE courses, non-computer science students can effectively learn and engage with advanced programming concepts by utilizing prompt-based development environments powered by LLMs. This approach is gaining traction across various fields, as prompt-based programming simplifies complex coding tasks and makes advanced programming more accessible. Thus, while programming languages remain a core component of computer science education, the emphasis should increasingly shift toward fostering creative problem-solving, building reliable and scalable systems, and ensuring that human expertise continues to guide and enhance the development of AI-driven technologies.

Looking ahead, we plan to further bridge the gap between academic learning and industry practices by incorporating more practical, discipline-specific datasets, such as those from biology and chemistry, into our curriculum. By expanding the use of LLM-based programming and integrating a wider variety of datasets, we aim to accelerate transdisciplinary research and better prepare students for the evolving demands of DSE.

Acknowledgments

This work has been supported by the National Science Foundation under Grant No. 2131269, through the CISE-MSI Program (RCBP-ED: CNS: Data Science and Engineering for Agriculture Automation)

References

- [1] Füsün Alacapınar. 2008. Effectiveness of project-based learning. *Eurasian Journal of Educational Research* 32, 1 (2008), 17–34.
- [2] Rolf Biehler, Lea Budde, Daniel Frischmeier, Birte Heinemann, Susanne Podworny, Carsten Schulte, and Thomas Wassong. 2018. *Paderborn Symposium on Data Science Education at School Level 2017: The Collected Extended Abstracts*. <https://doi.org/10.17619/UNIPB/1-374>
- [3] Orit Hazzan and Koby Mike. 2021. A journal for interdisciplinary data science education. *Commun. ACM* 64, 8 (jul 2021), 10–11. <https://doi.org/10.1145/3469281>
- [4] Mark G. Lawrence, Stephen Williams, Patrizia Nanz, and Ortwin Renn. 2022. Characteristics, potentials, and challenges of transdisciplinary research. *One Earth* 5, 1 (2022), 44–61. <https://doi.org/10.1016/j.oneear.2021.12.010>
- [5] Linda H Lewis and Carol J Williams. 1994. Experiential learning: Past and present. *New directions for adult and continuing education* 1994, 62 (1994), 5–16.
- [6] Koby Mike, Benny Kimelfeld, and Orit Hazzan. 2023. 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>.
- [7] Basarab Nicolescu. 1999. The transdisciplinary evolution of learning. In *Symposium on overcoming the underdevelopment of learning at the annual meeting of the American educational research association, Montreal, Canada*.
- [8] Vijay Janapa Reddi, Greg Diamos, Pete Warden, Peter Mattson, and David Kanter. 2021. Data Engineering for Everyone. arXiv:2102.11447 [cs.LG] <https://arxiv.org/abs/2102.11447>
- [9] Lois Rubin and Catherine Hebert. 1998. Model for Active Learning: Collaborative Peer Teaching. *College Teaching* 46, 1 (1998), 26–30. <http://www.jstor.org/stable/27558871>
- [10] Faryad Sahneh, Meghan A. Balk, Marina Kisley, Chi-kwan Chan, Mercury Fox, Brian Nord, Eric Lyons, Tyson Swetnam, Daniela Huppenkothen, Will Sutherland, Ramona L. Walls, Daven P. Quinn, Tonantzin Tarin, David LeBauer, David Ribes, Dunbar P. Birnie, III, Carol Lushbough, Eric Carr, Grey Nearing, Jeremy Fischer, Kevin Tyle, Luis Carrasco, Meagan Lang, Peter W. Rose, Richard R. Rushforth, Samapriya Roy, Thomas Matheson, Tina Lee, C. Titus Brown, Tracy K. Teal, Monica Papeş, Stephen Kobourov, and Nirav Merchant. 2021. Ten simple rules to cultivate transdisciplinary collaboration in data science. *PLOS Computational Biology* 17, 5 (05 2021), 1–12. <https://doi.org/10.1371/journal.pcbi.1008879>