Early College Research Institute

Bard College at SIMON'S ROCK

Early College Folio

A Growth MindSTEM for Next Gen

Volume 3, Issue 1 | 2024

Early College Folio

Volume 3, Issue 1 | 2024

Humanizing Data Science Curriculum

Di Wu
Lehman College, CUNY

Abstract

This paper explores the concept of constructivism in the context of data science education, arguing that an experience-centered approach can lead to more engaging and meaningful learning experiences. Specific examples of how constructivism can be applied in the classroom, such as through reflective writing, meditation, and flexible assignments are provided. The importance of creating a supportive and inclusive learning environment and strategies for promoting collaboration and peer learning are discussed. The feedback from students and instructors is included. Overall, the paper emphasizes the need for educators to humanize the data science curriculum and prioritize the needs and interests of students. This paper aims to contribute to STEM education with the author's experience and reflections.

1 Introduction

Contemporary education has undergone significant transformations due to the advent of the Internet and computers. The increased volume, variety, and velocity of knowledge pose new challenges for both educators and learners. Constructivism is an educational theory that emphasizes the active role of learners in constructing their own knowledge and the importance of social interaction and collaboration in learning [Dollard et al., 1996, Brophy, 1999]. Constructivism is better suited for contemporary education because it engages learners in active, collaborative, and meaningful learning experiences that develop their higher-order thinking skills, self-regulation, and lifelong learning abilities in today's rapidly changing world [Steffe and Gale, 1995, Matthews, 1998, Krahenbuhl, 2016].

The main entities involved in education are teachers, learners, and subjects. While pedagogical approaches and practices take care of all three entities, they might be initiated, centralized, and emphasized differently. Teacher-centered pedagogy focuses on knowledge transmission and regards teachers as the source of knowledge. In this approach, students are expected to passively receive the information presented by the teacher and to demonstrate their learning through performance and evaluation [Cuban, 1993]. Student-centered pedagogy emphasizes the active role of learners and regards teachers as facilitators in constructing an environment in which learners collaboratively experience. In this approach,

students are encouraged to explore their interests and goals, to make connections between new and prior knowledge, to solve problems and challenges, and to reflect on their own learning process [Moate and Cox, 2015]. Subject-centered pedagogy concentrates on the content and structure of the discipline and the development of the expertise. In this approach, students are exposed to the original and complex form of the subject matter, to the methods and tools of inquiry, and to the disciplinary standards and criteria [Mascolo, 2009].

While subjects of classes are important, I believe that human agency, including both teachers and students, is crucial for successful education and has not been addressed adequately. I first discuss the backward design flow which puts the learning experience at the center of the curriculum, then I present the practices I adopted to advocate humanity in a graduate-level data science curriculum, including the design logic, implementation, feedback, and modifications. Details of the syllabus, assignments, project design, etc. are included in the appendix. I aim to contribute to the community of STEM education with these teaching tips.

2 Center of the Curriculum

My practices prioritize the learning experience as the center of the curriculum. I believe all entities – teachers, students, and subjects – are essential components of the learning experience. Teachers act as facilitators to foster an inclusive and stimulating learning environment, students engage and collaborate to advance the learning process, and the subjects provide the foundation of the learning experience. To achieve this goal, I began with the question "what should students be upon graduation?" The answer is straightforward: a competent team member with a desirable *personality* and a *skill set* that aligns with their domain-specific position and life-long *learning ability*.

2.1 Personality

Students are expected to develop the capacity for effective interaction with others, encompassing communication, social, teamwork, and conflict management skills crucial for their professional journeys. These competencies extend beyond colleague interactions to encompass dealings with clients and customers. Additionally, students should adeptly present themselves in public and collaborate seamlessly with colleagues to achieve team objectives. Proficiency in handling emergencies or conflicts among interested parties is also emphasized. The learning outcomes in this category include:

- Effective Communication and Collaboration: Students will excel in communicating with diverse individuals, working collaboratively in team settings, and contributing positively to group discussions and activities. They will demonstrate the ability to fulfill assigned roles and responsibilities.

- Leadership and Decision-Making: Students will showcase leadership skills when needed, take initiative in planning and organizing team activities, and make well-informed and timely decisions. They will adeptly engage in constructive negotiation during conflicts and effectively assign tasks based on individual strengths and project requirements.
- Adaptability, Resilience, and Self-Reflection: Students will demonstrate the capacity to adapt to changing circumstances and unforeseen challenges. They will exhibit resilience in the face of setbacks or difficulties and engage in thoughtful self-reflection regarding their contributions to the team, actively seeking continuous improvement.

2.2 Learning Ability

Students are expected to excel in learning as data science is a field that evolves rapidly and has a highly dynamic environment. While I aim to teach them all the necessary knowledge, it's impossible to teach them new concepts that currently do not exist. Therefore, they must be adaptable and able to learn independently. Improving their learning ability is crucial to their success. This includes not only the acquisition of knowledge but also the development of effective learning strategies. The learning outcomes in this category include:

- Adaptability and Continuous Learning: Students will demonstrate adaptability in the rapidly evolving field of data science, showcasing the ability to navigate and embrace new concepts as they emerge. They will independently seek out and acquire new knowledge beyond the curriculum, demonstrating a proactive approach to staying updated in the dynamic environment of data science.
- Independent Learning and Resourcefulness: Students will develop the capability to learn independently, recognizing the importance of self-directed learning in a field with constantly evolving technologies and methodologies. They will exhibit resourcefulness in finding and utilizing learning materials beyond the classroom, showcasing a proactive attitude toward their ongoing educational journey.
- Effective Learning Strategies: Students will not only acquire domain-specific knowledge but will also refine and apply effective learning strategies. They will develop the skill to analyze, synthesize, and apply information efficiently, fostering a deep understanding of data science concepts and methodologies.

2.3 Course Material

The course material is an essential component of the curriculum. However, I view it as a means to train students to be good people and good learners, rather than an end in itself. The course content covers both theoretical and practical aspects

of data science. I aim to expose students to a wide range of topics, delve deeply into some of them, and leave others for them to explore on their own. The learning outcomes in this category include:

- Values and Ethical Decision-Making: Students will develop a strong sense of ethical responsibility and integrity, understanding that the course material serves as a foundation for fostering not only technical skills but also moral character. They will demonstrate the ability to make ethically sound decisions in the context of data science, recognizing the impact of their work on individuals and society.
- Critical Thinking and Exploration: Students will cultivate critical thinking skills by engaging with both theoretical and practical aspects of data science. They will analyze, evaluate, and synthesize information to make informed decisions. They will be encouraged to explore topics beyond the course material, fostering curiosity and a lifelong learning mindset.
- Holistic Learning and Personal Development: Students will recognize the importance of the course material as a tool for personal and professional development rather than a mere academic requirement. They will actively seek opportunities for holistic learning, combining theoretical knowledge with practical applications, and view the course as a platform for continuous growth as both individuals and learners.

The execution of the design flow is summarized in Table 1.

Outcome	Assessment	Grade %	Activity
Course material	Individual	9	Theory
Application	Individual	15	Application
Learning	Class	6	Learning Competition
Application, Learning	Individual	10	Mini Projects
Personality, Learning	Team	10	Team Discussion
All	Team	5	Project Debut
All	Team	5	Project Checkpoint
All	Team	10	Final Presentation
All	Team	10	Final Report
Learning	Individual	0	Case Studies
Course Material, Learning	Individual	20	Exams
All	Team	(3)	Investment Game

Table 1: Backward Design

3 Humanizing Data Science Curriculum

Data science is a multidisciplinary field aimed at extracting insights from data to inform decision-making processes. It encompasses three fundamental components:

mathematics, programming, and domain knowledge. Mathematics elucidates the significance of data for human understanding, programming translates human commands into computer operations, and domain knowledge facilitates comprehension and communication of data insights among individuals.

From the initiation of degree programs to classroom instruction, and from admissions to graduations, mathematics and programming have played dominating roles in every stage. Undoubtedly, these components are crucial, as they enable us to use programs for data analysis based on statistics and algorithms. However, it's essential to recognize that data science, being a distinctive field, demands learners who can analyze, synthesize, evaluate, and create knowledge across diverse sources and domains. Additionally, effective communication and collaboration skills are imperative, given that data scientists interact not only with data and programs but, most importantly, with fellow humans.

While mathematics and programming emphasize objectivity, the nature of data science involves subjective skills. Students need to gather data from clients, collaborate with team members, and present findings to varied audiences. This necessitates skills like communication, leadership, project management, self-learning, and effective self-management. Adopting a constructive pedagogy is crucial as it provides learners with opportunities to explore their interests and goals, establish connections between new and existing knowledge, tackle problems and challenges, and reflect on their own learning journey.

3.1 Whole Person Learning

Students as learners play a vital role in the learning experience. They are humans, but often we expect them to follow every instruction, purchase required books and devices, complete every assignment, solve every question, and pass every exam without emotions. The whole-person learning approach acknowledges students as real people, who may face illness, time constraints, financial difficulties, stress, and diverse needs and backgrounds [Yorks and Kasl, 2002]. The following practices aim at providing support to students as individuals, thus, we can ensure that each and every learner is actively and collaboratively engaged in the learning experience.

Open Education Resource To remove the barrier of the cost of textbooks and promote inclusion, I developed an Open Education Resource, titled Data Mining with Python, free and accessible to students. This book focuses on the hands-on approach and is designed to give students an understanding of data mining concepts in an applicable way. The tutorials in this book will help students to gain practical skills to implement data mining techniques in real-world scenarios. The book serves as complementary to a theoretical course and can be accessed at `https://scholar.colorado.edu/concern/defaults/9880vs25m`.

Reflective Writing Reflective writing is an effective approach to help students to better understand the curriculum, better understand the learning experience, and better understand themselves and their surroundings.

For lectures, assignments, and mini projects, reflective writing can help students look through their learning journey and ask some specific questions:

- How did you understand this lecture?
- If you had difficulties in learning this concept, what are they? How did you overcome them?
- What was your experience working with your teammates? Did you have hard moments coordinating with each other? How did you reach agreements and get the work done?
- What is your takeaway from today's lecture? Why is it? If you were the instructor, how would you make a question and test it?

Reflective writing can help students slow down, take a pause, and think about some big questions:

- Why do you study data science? Why do you want a job in data science?
- What do you want after successfully graduating and taking the job?
- How to stay current? How will you apply the learning techniques for new techniques in the future?

Reflective techniques we can have for this task are Loop Writing and Collaborative Learning Questions for individual thinking, Dialectical Notebooks and Writing in the Zone in a group for team thinking, and Process Writing for deeper and complete thinking overall. Readers can find more details of the techniques and examples at `https://www.lehman.edu/institute-literacy-studies/wac/critical-reading.php`.

The use of reflective writing has been successful in achieving its intended purpose. Various activities in the course, including team discussions, theoretical portions of assignments, and summary and reflection components of exams, involve reflective writing. At the outset, students often struggle with this type of thinking and writing. However, with regular and consistent practice, I believe they have come to appreciate the value of reflective writing. I hope that they will continue to develop this habit even after graduating.

Some example questions I used for team discussion are:

- What are the *three* things you learned in today's class? What are the *two* things DIFFERENT than your understanding before today's class? What is the *one* topic you'd like to have in Exam 2?
- What are the *three* similarities between classification and regression? What are the *two* differences between them? Explain *one* similarity or difference in detail with examples.

- From the four categories of clustering methods we learned, choose *three* for comparison with examples. How do the *two* partitioning methods – K-Means and K-Medoids – differ? Explain *one* real-life scenario in which we can apply DBSCAN.

- What is your *one* sentence summary of this presentation (by a team)? Write *one* aspect that you like most about their presentation. Provide *one* suggestion that you think the team should improve.

Meditation Transitioning students from their busy schedules into the classroom requires consideration and a gradual approach. Providing them with time and establishing a routine is essential for creating an effective learning environment. One way to facilitate this transition is through meditation or a ceremony that embodies the unique atmosphere of our class.

In the past, I used to play piano music before lectures, but I recognized that not all students share the same taste, and some even dislike it. Consequently, I have shifted my approach. Instead of piano music, I now play recent news snippets and connect them to the course material. For instance, I discuss the 2021 trend of the "metaverse," the 2022 renaming of "Facebook," the 2022 successes and failures of Zillow's real estate model, and topics like the Fed's fight against inflation and the rise of interest rates. This strategy encourages students to ponder the relevance of current events to the upcoming lecture, fostering engagement throughout the class.

Flexible Assignments We often have to meet the expectations of a very diverse classroom. Some experienced students may want to get more advanced material, some students simply love to take challenging questions and learn under stress, and some students simply just want to learn the basics and cannot tolerate frustration at all. To create a comfortable learning experience for everyone, I designed flexible assignments. The flexibility is achieved through multiple formats:

- There are mandatory assignments that are required for all students. However, some questions are optional. Students can accept the challenge and advance their learning by taking the optional questions, they can also ignore the challenge and just finish the required questions. A sample assignment is shown in Figure 1 in the appendix.

- There are optional assignments and students need to finish a certain number of them, but not all. For example, we had 20 mini-projects which are all real-life case studies. Students need to finish 10 of them to get full credit. While some students finished all 20, some students just finished 10 and skipped the rest. The average completion number is 15. A sample mini project is shown in Figure 2 in the appendix.

- There are optional assignments and students are not obliged to finish them at all. These assignments may be additional readings, self-exploratory ad-

vanced topics, or learning new Python packages. A sample optional assignment is shown in Figure 3 in the appendix.

The flexible assignments in addition provide students with a lot of opportunities to make choices and to know themselves better, which are the essential skills for their success in their careers.

3.2 Whole Person Teaching

Teachers play an important role in the learning experience as well. We are humans, but often we are expected to be perfect. Whole-person teaching approach recognizes our imperfections as real people, who may be in a sad mood, who do not know every method, who may not know current memes, and who may have a hard time picking up new technology. The following practices aim at providing support to teachers as individuals, thus, we can ensure that learners have a clear and acceptable expectation of their educators in the learning experience.

Teaching Philosophy To establish the expectations of the course, I developed a 30-minute presentation addressing the teaching philosophy in the first class. With this presentation, students will know exactly what they are expected to do so they can plan accordingly. The teaching philosophy is a higher-level syllabus, which often works as the contract for each class. It provides students with a detailed understanding of my approach to teaching, as well as insights into my personal interests, such as my fondness for World of Warcraft. During the presentation, I emphasized my definition of learning and my expectations for students to be diligent learners.

Meal Interviews To keep current, I want to create and maintain a routine that is a stable, open, and concentrated channel for getting inputs. Dr. Michael Wesch in his TEDxMHK talk introduced his experience in lunch interviews with students (readers can access the video at `https://youtu.be/SP7dbl0rJS0`). I borrowed that idea and conducted weekly/biweekly meetings with students, industry practitioners, and academic colleagues. During the meetings, we didn't do small talk but directly got to the business. To encourage open and frank communication, I kept all interviews anonymous. I found conducting interviews to be an extremely valuable experience. Without sitting down with my students and spending an hour talking with them, I wouldn't have been able to gain such a deep understanding of their true feelings. Furthermore, the interviews, even though they took place in a restaurant, set up an expectation that we would be discussing serious topics. Colleagues, particularly practitioners in the industry, went above and beyond my expectations. Some takeaways from the interviews are:

- I know many students are beginners, and you are trying to make everyone comfortable. However, the practices and assignments required are too easy and too boring for me. I have industry experience and I am way more

advanced than others, but I still have to do the assignments. I felt like I was wasting my time.

- I like how you designed the course. It is unique and I love to learn by doing. I wish we could do in-class tutorials for every topic we learn.
- You should be more assertive with students. We are from a culture where professors won't smile at you.
- You did a wonderful job in explaining the theoretical background of the topics, but I want to see more real-life examples and intensive coding.
- The optional mini-projects are great, but I just don't have time to do them. Please make them mandatory.
- For the lectures, I really like yours. But, you should look at [another Professor]'s assignments. They are very challenging, but I feel very good after I get them done.
- I understand team project is good for learning. I just don't like teamwork, and I don't want to push team members to do the work. Either I don't care, or I have to do everything. I prefer to do the project on my own.
- Your slides, videos, and other course materials are so complete (so sometimes I feel I will still be fine if I skip some classes).
- We hire talent, not a tool. We do not expect the student to know everything. However, we want to hire someone who can learn quickly (in a team).
- Our company is in a bad position, and it will remain falling next year (2023). Students should be prepared to compete with laid-off employees.
- In a lot of cases, what you have done is not that important, but what you can deliver is.
- My department [English and Communication] actually cares about data and data analysis. We learned and used Machine Learning and AI for research and teaching too.
- Everyone is busy. If you have a problem, ask around, but don't just wait for answers. Do some research quickly by yourself.

Based on these (and many other) comments, I made adjustments to my courses:

- I modified assignments, and made some questions optional, but more challenging. By modifying the assignments, you have given students more choices and made the class more accessible to students with different skill levels and backgrounds. The optional challenging questions provide an opportunity for students who want to push themselves, while the mandatory questions ensure that students receive the necessary knowledge.

- I added many applicable, case-study style, hands-on tutorials for the content this semester. It is a great way to help students apply what they've learned and make the content more engaging. Students recognized the value of these tutorials and found they are useful as complementary to the introduction of theoretical background in lectures.
- For 20 mini projects (almost two mini projects per week), I set up 10 as the threshold to get the full credit. Students don't have to do all, but they cannot drop all. This, again, provides flexibility. By setting a threshold, I encourage students to complete a reasonable amount of work while still allowing them to choose which projects they want to focus on. The other thing I learned, is that students want to have all projects unlocked at the same time, so they can plan, even if these projects are due in different weeks.
- I delivered the comments from industry practitioners. It helps prepare students for real-world situations and gives them a better understanding of what is expected of them in their future careers.
- I added some in-person-only activities, which are also a good way to engage students who might otherwise prefer to attend classes online.
- I weighted a lot in their presentations, and this emphasis on communication skills and presentation ability is a valuable lesson for students. They need to understand that being able to communicate their ideas effectively is just as crucial as being able to code.

Survey in Exams I included a mandatory survey in exams and the survey is worth 5% of the grade. In this way, I can get feedback from students, and make some changes in the course so students know their opinions matter. The survey has become a regular part of exams. I made the survey anonymous to encourage students to provide more candid feedback. Similar to the meal interviews, the survey has been incredibly beneficial in helping our understanding of students' perspectives. Most importantly, when students see that I make immediate adjustments to the course design based on their feedback, they feel acknowledged and are more willing to collaborate with us to enhance the course. Additionally, each semester brings a different cohort of students, and while some activities may be well-received by the current cohort, I must remain open to making changes in the future based on feedback from subsequent cohorts. The questions I asked in the survey are:

- How do you feel about this exam? How shall we improve it for the next exam?
- How do you feel about our homework so far? What should be improved for the rest of the homework?
- How do you feel about our class so far? What should be improved for the next half of the course?

- How do you feel about our mini projects so far? What should be improved for the next semester? For the 17 projects, shall we make 10 rather than 7 as required?
- How do you feel about our class so far? What should be improved for the next semester?
- What is your favorite activity in this class? What is the least?

3.3 Humanizing Subjects

Flipped Classroom To foster a subject-centered community, I implemented several student-led lectures in the form of a flipped classroom activity. We successfully implemented a student-led lecture activity in the Data Structure course, which proved to be very effective. In each class session, students were given 7 to 10 minutes to introduce a topic of their choice, after which we continued the lecture based on their presentation. This activity not only engaged the students but also demonstrated that the classroom was a collaborative space. It also encouraged them to learn how to find resources, deliver knowledge, and allocate their time wisely. However, to ensure the success of this activity, it is crucial to provide clear and detailed instructions to the students; otherwise, they may feel confused and lack direction. A sample flipped classroom activity is shown in Figure 5 in the appendix.

Cheat-sheet Competition Summarizing a subject is an important aspect of the learning experience and serves as a valid assessment for achieving learning outcomes. As such, we require students to prepare cheat sheets for their exams and hold a competition for the best cheat sheet. The winner will be selected through a vote among both the students and the instructor. This competition motivates students to strive for excellence and improve their understanding of the course material. Moreover, the peer-review process enables students to learn from each other's strengths and weaknesses, which enhances their knowledge in all three subjects. A sample cheat-sheet competition activity is shown in Figure 6 in the appendix.

Learning in Cohort Because the cheat-sheet competition proved to be highly successful during the fall 2022 semester, I decided to break it down into several smaller activities for the spring 2023 semester. For weekly assignments, there will be a section called "Learning in Cohort," which requires students to find and share the best resources available for a given topic in the discussion board. They must also provide reasons for choosing their selected resources. The most popular resource, as determined by the number of likes, will earn its author a small reward, such as being exempt from the next assignment. A sample learning in cohort assignment is shown in Figure 4 in the appendix.

Course Project Course projects play a crucial role in determining the final grade and facilitating students in achieving the outlined course objectives. This semester, we've provided comprehensive instructions for the course projects, with no restrictions on their scope. To help students stay on track, we've established four milestones, each aligning with a significant phase of the course project.

In week five, students will accomplish milestone 1, the project debut. Following this, in week 10, students are required to present their progress in milestone 2, known as the project checkpoint. Week 15 marks the completion of the final presentation, designated as milestone 3. Finally, in week 16, students are expected to submit their final report, concluding the series of milestones. This structured approach ensures a step-by-step progression through the course project, aligning with the semester timeline.

To facilitate student learning, we have started to candidly point out their errors and imperfections, even though we know this may be frustrating for them. It's important to recognize that happiness does not equate to wellness. We hope that students can learn from their mistakes and imperfections and use that knowledge to improve their performance in future projects, ultimately preparing them for success in the real world.

To foster critical thinking, we introduced the Investment Game. In the initial three milestones, which involve public presentations, students take on the dual roles of presenters and investors. As they assess the presentations and projects of other teams, their task is to identify the most promising team and allocate their tokens accordingly. Recognizing the teams with the highest token counts, we have the top three team awards. The leading team rewards its investors with a remarkable 1000% return, the second team with 500%, and the third team with 300%. Investors face token losses if they invest in teams that do not make the top three. Determined by the number of tokens each investor gathers, we have the top three investor awards. This engaging activity stimulates the cohort to attentively listen, observe, and refine their skills in discerning valuable lessons while honing their evaluative abilities.

For these project milestones and investment game, please find the details in Figure 7, 8, 9, 10, and 11 in the appendix.

4 Conclusion and Future Work

Many of these pedagogical practices can be applied to STEM in general, as they can help students develop critical thinking, creativity, communication, and collaboration skills that are essential for STEM fields. Moreover, they can help students appreciate the ethical and social implications of STEM applications, and foster a culture of inclusion and diversity in STEM education and careers. However, some disciplines in STEM require absolute objective and independent thinking, which puts different challenges to education. Applying data science pedagogy to STEM in general may require some adaptation or customization to fit the specific needs and characteristics of each discipline or context. It may also require collabora-

tion and coordination among different stakeholders in STEM education, such as educators, researchers, practitioners, policymakers, etc.

The Supreme Court recently ruled that Affirmative Action violates the Constitution and ended its use in higher education. This decision has significant implications for college admissions, as it eliminates the consideration of race as one of the factors to increase the representation of historically marginalized groups. Minority and underrepresented students may face more disadvantages due to social class, income, and other factors that affect their access and opportunities to quality education. However, it is not the end of diversion, equity, and inclusion in pedagogy. We will and must be adaptive to this change, and continue to construct a DEI learning environment for all humans.

Acknowledgment

The practices were inspired and supported by the ASSETT 2022-2023 Program at University of Colorado Boulder. The development of the OER book was supported by Open CU Boulder 2022-2023. I would like to thank the reviewers for their valuable comments that improved the paper.

DR. DI WU is currently an Assistant Professor of Finance, Information Systems, and Economics department of Business School, Lehman College, CUNY. Dr. Di Wu's research interests are Temporal extensions to RDF and semantic web, Applied Data Science, and Experiential Learning and Pedagogy in business education.

References

[Brophy, 1999] Brophy, J. (1999). *Perspectives of Classroom Management: Yesterday, Today and Tomorrow*. Allyn & Bacon, Boston.

[Cuban, 1993] Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms, 1890-1990*. Teachers College Press.

[Dollard et al., 1996] Dollard, N., Christensen, L., Colucci, K., and Epanchin, B. (1996). Constructive classroom management. *Focus on Exceptional Children*, 25:X–12.

[Krahenbuhl, 2016] Krahenbuhl, K. S. (2016). Student-centered education and constructivism: Challenges, concerns, and clarity for teachers. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 89(3):97–105.

[Mascolo, 2009] Mascolo, M. F. (2009). Beyond student-centered and teacher-centered pedagogy: Teaching and learning as guided participation. *Pedagogy and the human sciences*, 1(1):3–27.

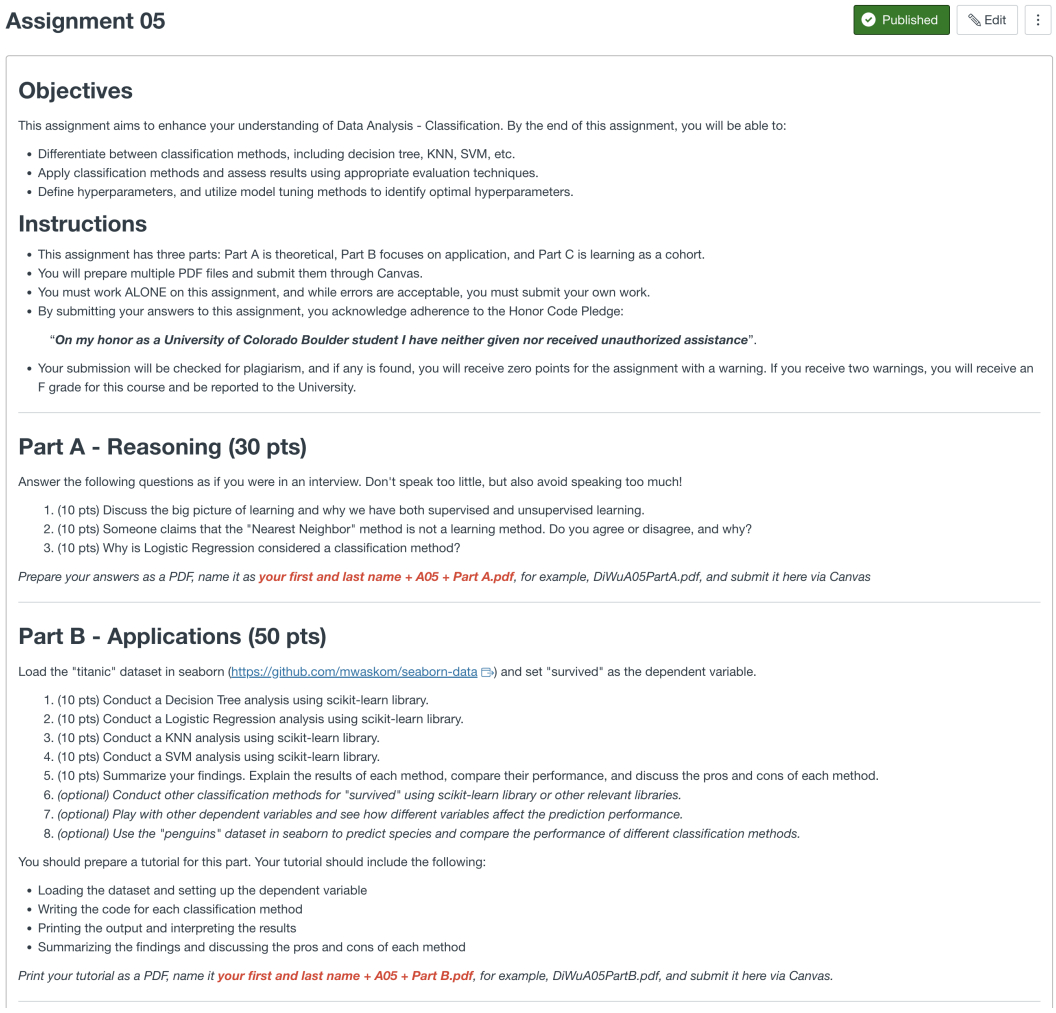
[Matthews, 1998] Matthews, M. R. (1998). *Constructivism in science education: A philosophical examination*. Springer Science & Business Media.

[Moate and Cox, 2015] Moate, R. M. and Cox, J. A. (2015). Learner-centered pedagogy: Considerations for application in a didactic course. *Professional Counselor*, 5(3):379–389.

[Steffe and Gale, 1995] Steffe, L. P. and Gale, J. E. (1995). *Constructivism in education*. Psychology Press.

[Yorks and Kasl, 2002] Yorks, L. and Kasl, E. (2002). Toward a theory and practice for whole-person learning: Reconceptualizing experience and the role of affect. *Adult Education Quarterly*, 52(3):176–192.

A Sample Assignments



Assignment 05

Published Edit

Objectives

This assignment aims to enhance your understanding of Data Analysis - Classification. By the end of this assignment, you will be able to:

- Differentiate between classification methods, including decision tree, KNN, SVM, etc.
- Apply classification methods and assess results using appropriate evaluation techniques.
- Define hyperparameters, and utilize model tuning methods to identify optimal hyperparameters.

Instructions

- This assignment has three parts: Part A is theoretical, Part B focuses on application, and Part C is learning as a cohort.
- You will prepare multiple PDF files and submit them through Canvas.
- You must work ALONE on this assignment, and while errors are acceptable, you must submit your own work.
- By submitting your answers to this assignment, you acknowledge adherence to the Honor Code Pledge:

  "***On my honor as a University of Colorado Boulder student I have neither given nor received unauthorized assistance***".

- Your submission will be checked for plagiarism, and if any is found, you will receive zero points for the assignment with a warning. If you receive two warnings, you will receive an F grade for this course and be reported to the University.

Part A - Reasoning (30 pts)

Answer the following questions as if you were in an interview. Don't speak too little, but also avoid speaking too much!

1. (10 pts) Discuss the big picture of learning and why we have both supervised and unsupervised learning.
2. (10 pts) Someone claims that the "Nearest Neighbor" method is not a learning method. Do you agree or disagree, and why?
3. (10 pts) Why is Logistic Regression considered a classification method?

*Prepare your answers as a PDF, name it as **your first and last name + A05 + Part A.pdf**, for example, DiWuA05PartA.pdf, and submit it here via Canvas*

Part B - Applications (50 pts)

Load the "titanic" dataset in seaborn (https://github.com/mwaskom/seaborn-data) and set "survived" as the dependent variable.

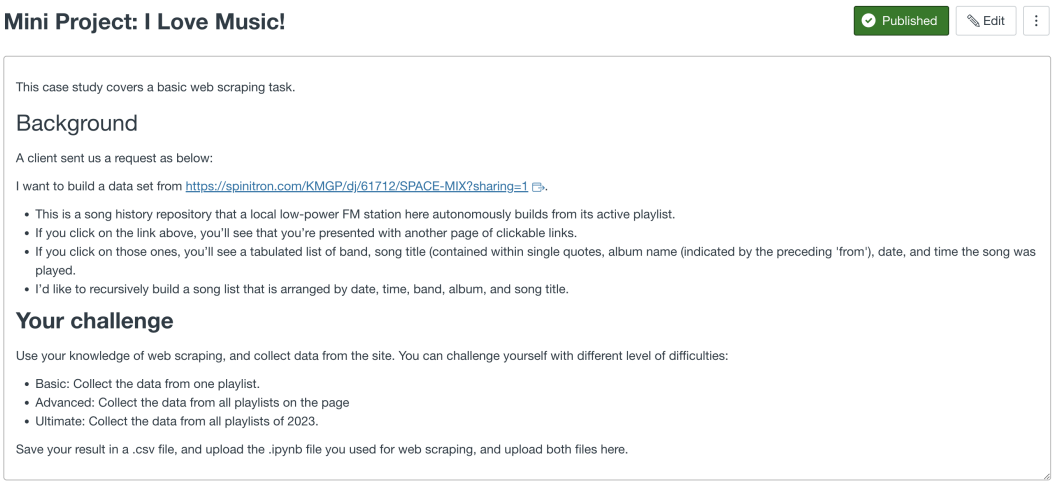
1. (10 pts) Conduct a Decision Tree analysis using scikit-learn library.
2. (10 pts) Conduct a Logistic Regression analysis using scikit-learn library.
3. (10 pts) Conduct a KNN analysis using scikit-learn library.
4. (10 pts) Conduct a SVM analysis using scikit-learn library.
5. (10 pts) Summarize your findings. Explain the results of each method, compare their performance, and discuss the pros and cons of each method.
6. *(optional) Conduct other classification methods for "survived" using scikit-learn library or other relevant libraries.*
7. *(optional) Play with other dependent variables and see how different variables affect the prediction performance.*
8. *(optional) Use the "penguins" dataset in seaborn to predict species and compare the performance of different classification methods.*

You should prepare a tutorial for this part. Your tutorial should include the following:

- Loading the dataset and setting up the dependent variable
- Writing the code for each classification method
- Printing the output and interpreting the results
- Summarizing the findings and discussing the pros and cons of each method

*Print your tutorial as a PDF, name it **your first and last name + A05 + Part B.pdf**, for example, DiWuA05PartB.pdf, and submit it here via Canvas.*

Figure 1: Assignment 5



Mini Project: I Love Music!

Published Edit

This case study covers a basic web scraping task.

Background

A client sent us a request as below:

I want to build a data set from https://spinitron.com/KMGP/dj/61712/SPACE-MIX?sharing=1.

- This is a song history repository that a local low-power FM station here autonomously builds from its active playlist.
- If you click on the link above, you'll see that you're presented with another page of clickable links.
- If you click on those ones, you'll see a tabulated list of band, song title (contained within single quotes, album name (indicated by the preceding 'from'), date, and time the song was played.
- I'd like to recursively build a song list that is arranged by date, time, band, album, and song title.

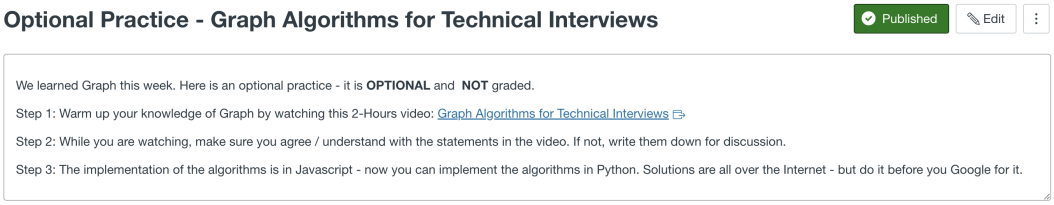
Your challenge

Use your knowledge of web scraping, and collect data from the site. You can challenge yourself with different level of difficulties:

- Basic: Collect the data from one playlist.
- Advanced: Collect the data from all playlists on the page
- Ultimate: Collect the data from all playlists of 2023.

Save your result in a .csv file, and upload the .ipynb file you used for web scraping, and upload both files here.

Figure 2: Mini Project: I Love Music!



Optional Practice - Graph Algorithms for Technical Interviews

Published Edit

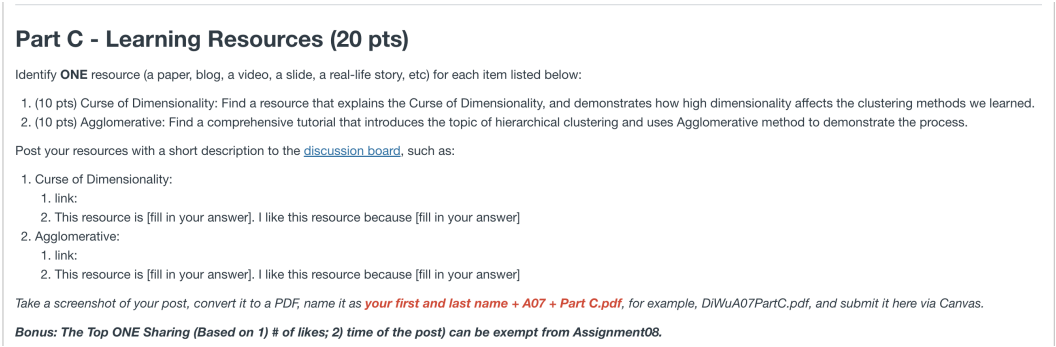
We learned Graph this week. Here is an optional practice - it is **OPTIONAL** and **NOT** graded.

Step 1: Warm up your knowledge of Graph by watching this 2-Hours video: Graph Algorithms for Technical Interviews

Step 2: While you are watching, make sure you agree / understand with the statements in the video. If not, write them down for discussion.

Step 3: The implementation of the algorithms is in Javascript - now you can implement the algorithms in Python. Solutions are all over the Internet - but do it before you Google for it.

Figure 3: Optional Assignment: Graph Algorithms



Part C - Learning Resources (20 pts)

Identify **ONE** resource (a paper, blog, a video, a slide, a real-life story, etc) for each item listed below:

1. (10 pts) Curse of Dimensionality: Find a resource that explains the Curse of Dimensionality, and demonstrates how high dimensionality affects the clustering methods we learned.
2. (10 pts) Agglomerative: Find a comprehensive tutorial that introduces the topic of hierarchical clustering and uses Agglomerative method to demonstrate the process.

Post your resources with a short description to the discussion board, such as:

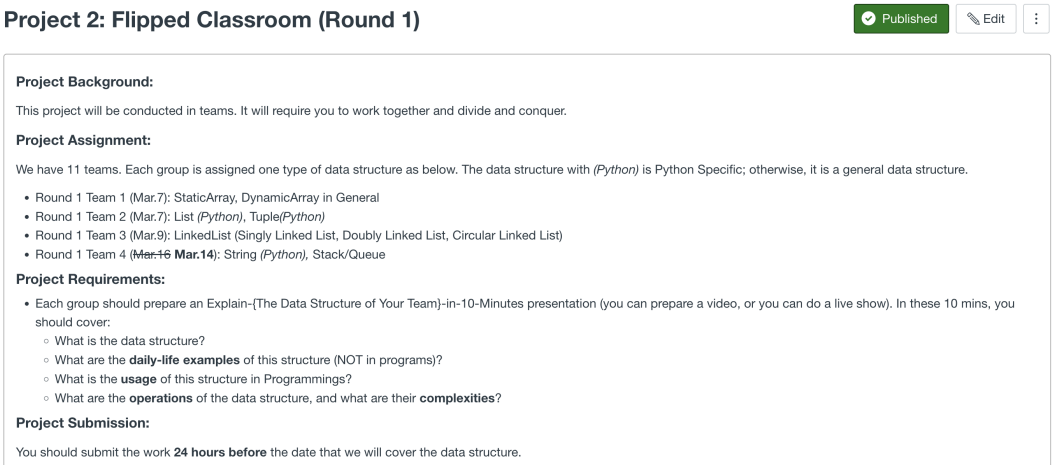
1. Curse of Dimensionality:
    1. link:
    2. This resource is [fill in your answer]. I like this resource because [fill in your answer]
2. Agglomerative:
    1. link:
    2. This resource is [fill in your answer]. I like this resource because [fill in your answer]

*Take a screenshot of your post, convert it to a PDF, name it as **your first and last name + A07 + Part C.pdf**, for example, DiWuA07PartC.pdf, and submit it here via Canvas.*

Bonus: The Top ONE Sharing (Based on 1) # of likes; 2) time of the post) can be exempt from Assignment08.

Figure 4: Learning in Cohort

B Flipped Classroom



Project 2: Flipped Classroom (Round 1)

Published Edit

Project Background:

This project will be conducted in teams. It will require you to work together and divide and conquer.

Project Assignment:

We have 11 teams. Each group is assigned one type of data structure as below. The data structure with *(Python)* is Python Specific; otherwise, it is a general data structure.

- Round 1 Team 1 (Mar.7): StaticArray, DynamicArray in General
- Round 1 Team 2 (Mar.7): List *(Python)*, Tuple*(Python)*
- Round 1 Team 3 (Mar.9): LinkedList (Singly Linked List, Doubly Linked List, Circular Linked List)
- Round 1 Team 4 (~~Mar.16~~ **Mar.14**): String *(Python),* Stack/Queue

Project Requirements:

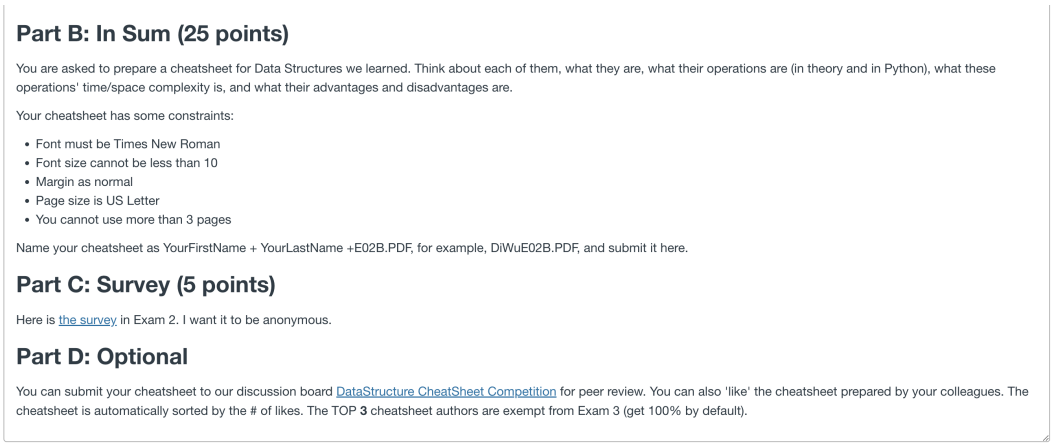
- Each group should prepare an Explain-{The Data Structure of Your Team}-in-10-Minutes presentation (you can prepare a video, or you can do a live show). In these 10 mins, you should cover:
  - What is the data structure?
  - What are the **daily-life examples** of this structure (NOT in programs)?
  - What is the **usage** of this structure in Programmings?
  - What are the **operations** of the data structure, and what are their **complexities**?

Project Submission:

You should submit the work **24 hours before** the date that we will cover the data structure.

Figure 5: Flipped Classroom Exercise

C Cheat-sheet Competition



Part B: In Sum (25 points)

You are asked to prepare a cheatsheet for Data Structures we learned. Think about each of them, what they are, what their operations are (in theory and in Python), what these operations' time/space complexity is, and what their advantages and disadvantages are.

Your cheatsheet has some constraints:

- Font must be Times New Roman
- Font size cannot be less than 10
- Margin as normal
- Page size is US Letter
- You cannot use more than 3 pages

Name your cheatsheet as YourFirstName + YourLastName +E02B.PDF, for example, DiWuE02B.PDF, and submit it here.

Part C: Survey (5 points)

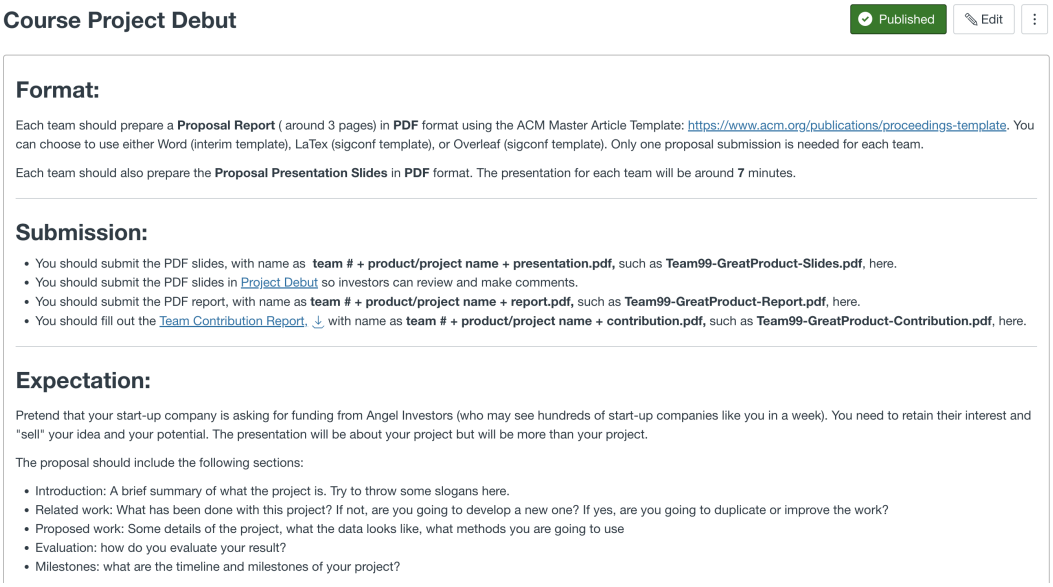
Here is the survey in Exam 2. I want it to be anonymous.

Part D: Optional

You can submit your cheatsheet to our discussion board DataStructure CheatSheet Competition for peer review. You can also 'like' the cheatsheet prepared by your colleagues. The cheatsheet is automatically sorted by the # of likes. The TOP **3** cheatsheet authors are exempt from Exam 3 (get 100% by default).

Figure 6: Cheat-sheet Competition

D Course Project



Course Project Debut

Published Edit

Format:

Each team should prepare a **Proposal Report** (around 3 pages) in **PDF** format using the ACM Master Article Template: https://www.acm.org/publications/proceedings-template. You can choose to use either Word (interim template), LaTex (sigconf template), or Overleaf (sigconf template). Only one proposal submission is needed for each team.

Each team should also prepare the **Proposal Presentation Slides** in **PDF** format. The presentation for each team will be around **7** minutes.

Submission:

- You should submit the PDF slides, with name as **team # + product/project name + presentation.pdf,** such as **Team99-GreatProduct-Slides.pdf**, here.
- You should submit the PDF slides in Project Debut so investors can review and make comments.
- You should submit the PDF report, with name as **team # + product/project name + report.pdf,** such as **Team99-GreatProduct-Report.pdf**, here.
- You should fill out the Team Contribution Report, with name as **team # + product/project name + contribution.pdf,** such as **Team99-GreatProduct-Contribution.pdf**, here.

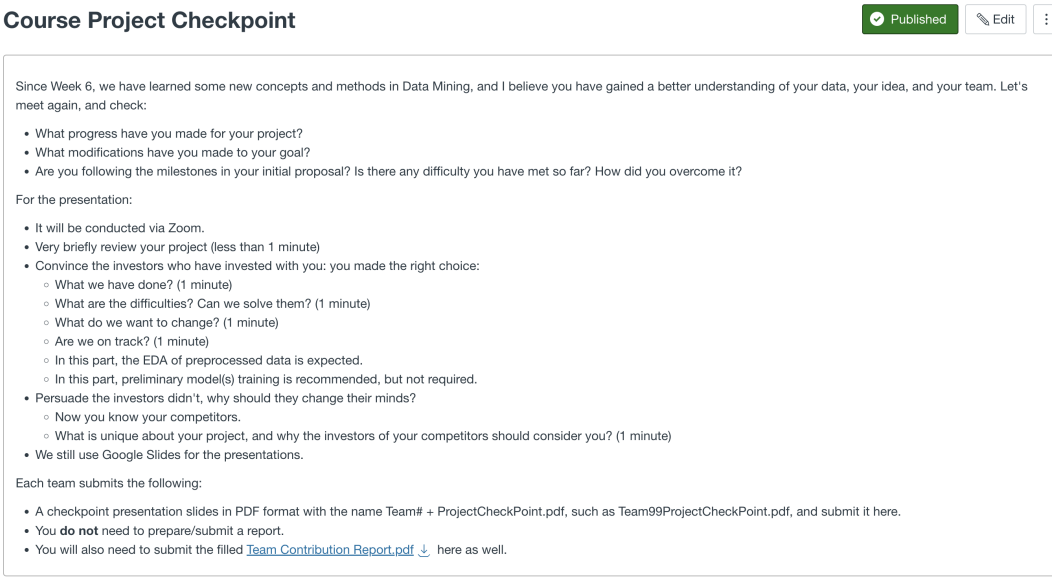
Expectation:

Pretend that your start-up company is asking for funding from Angel Investors (who may see hundreds of start-up companies like you in a week). You need to retain their interest and "sell" your idea and your potential. The presentation will be about your project but will be more than your project.

The proposal should include the following sections:

- Introduction: A brief summary of what the project is. Try to throw some slogans here.
- Related work: What has been done with this project? If not, are you going to develop a new one? If yes, are you going to duplicate or improve the work?
- Proposed work: Some details of the project, what the data looks like, what methods you are going to use
- Evaluation: how do you evaluate your result?
- Milestones: what are the timeline and milestones of your project?

Figure 7: Course Project – Milestone 1 – Debut



Course Project Checkpoint

Published Edit

Since Week 6, we have learned some new concepts and methods in Data Mining, and I believe you have gained a better understanding of your data, your idea, and your team. Let's meet again, and check:

- What progress have you made for your project?
- What modifications have you made to your goal?
- Are you following the milestones in your initial proposal? Is there any difficulty you have met so far? How did you overcome it?

For the presentation:

- It will be conducted via Zoom.
- Very briefly review your project (less than 1 minute)
- Convince the investors who have invested with you: you made the right choice:
  - What we have done? (1 minute)
  - What are the difficulties? Can we solve them? (1 minute)
  - What do we want to change? (1 minute)
  - Are we on track? (1 minute)
  - In this part, the EDA of preprocessed data is expected.
  - In this part, preliminary model(s) training is recommended, but not required.
- Persuade the investors didn't, why should they change their minds?
  - Now you know your competitors.
  - What is unique about your project, and why the investors of your competitors should consider you? (1 minute)
- We still use Google Slides for the presentations.

Each team submits the following:

- A checkpoint presentation slides in PDF format with the name Team# + ProjectCheckPoint.pdf, such as Team99ProjectCheckPoint.pdf, and submit it here.
- You **do not** need to prepare/submit a report.
- You will also need to submit the filled Team Contribution Report.pdf here as well.

Figure 8: Course Project – Milestone 2 – Checkpoint

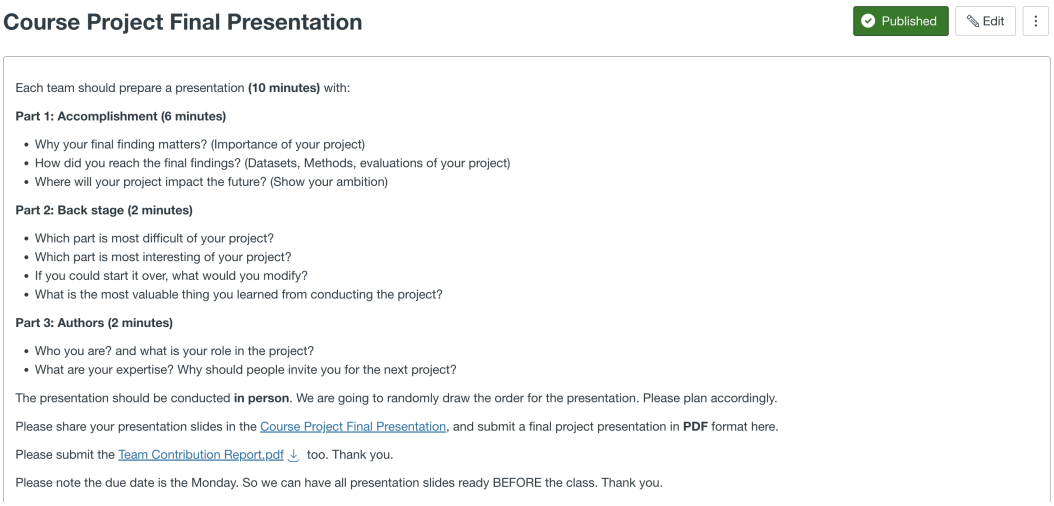
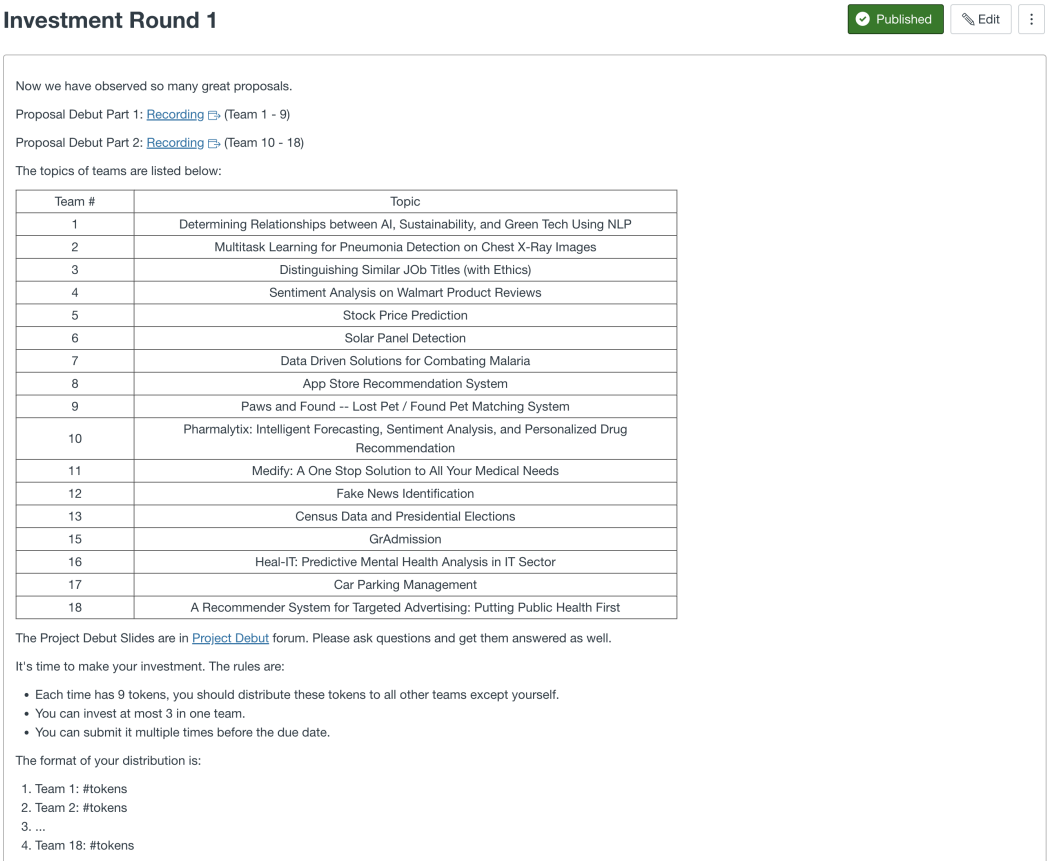


Figure 9: Course Project – Milestone 3 – Final Presentation

Figure 10: Course Project – Milestone 4 – Final Report



Investment Round 1

Published Edit

Now we have observed so many great proposals.

Proposal Debut Part 1: Recording (Team 1 - 9)

Proposal Debut Part 2: Recording (Team 10 - 18)

The topics of teams are listed below:

Team #	Topic
1	Determining Relationships between AI, Sustainability, and Green Tech Using NLP
2	Multitask Learning for Pneumonia Detection on Chest X-Ray Images
3	Distinguishing Similar JOb Titles (with Ethics)
4	Sentiment Analysis on Walmart Product Reviews
5	Stock Price Prediction
6	Solar Panel Detection
7	Data Driven Solutions for Combating Malaria
8	App Store Recommendation System
9	Paws and Found -- Lost Pet / Found Pet Matching System
10	Pharmalytix: Intelligent Forecasting, Sentiment Analysis, and Personalized Drug Recommendation
11	Medify: A One Stop Solution to All Your Medical Needs
12	Fake News Identification
13	Census Data and Presidential Elections
15	GrAdmission
16	Heal-IT: Predictive Mental Health Analysis in IT Sector
17	Car Parking Management
18	A Recommender System for Targeted Advertising: Putting Public Health First

The Project Debut Slides are in Project Debut forum. Please ask questions and get them answered as well.

It's time to make your investment. The rules are:

- Each time has 9 tokens, you should distribute these tokens to all other teams except yourself.
- You can invest at most 3 in one team.
- You can submit it multiple times before the due date.

The format of your distribution is:

1. Team 1: #tokens
2. Team 2: #tokens
3. ...
4. Team 18: #tokens

Figure 11: Investment Game Round 1

