Carnegie Mellon
Heinz College

<table>
<tr><td>Course Information</td><td>Course Title: 95829 Software Design for Data Scientists
Instructor: Neelam Dwivedi (ndwivedi@andrew.cmu.edu).
Office hours: Please check the course home page.</td></tr>
<tr><td>Prerequisites</td><td>None</td></tr>
<tr><td>Description</td><td>Less than half of data science projects across the industry get deployed in production and a key factor contributing to this problem is the lack of process centric approach. How can the principles and practices of software design and engineering help alleviate this problem? KDD, CRISP-DM, ASUM, SEMMA, TDSP from the world of data mining, and Agile, Lean, and DevOps from the world of software engineering are now coming together in DataOps to address this challenge. Get an appreciation for these methodologies that can help you increase the success rate of your data science projects. Expand your learnings through case studies on how organizations are managing their data analytics pipeline and productionizing their data science experiments. Apply these learnings in a small project with your peers . This course will help you learn how a short-term data science initiative can evolve into a long-term organizational capability.</td></tr>
<tr><td>Course Materials</td><td>Reference materials will be provided during the course via Canvas.</td></tr>
<tr><td>Evaluation Method</td><td>The final grade will be out of 100 points. The grading breakdown is listed below. A detailed description of each of these activities is given on the next page.
<table>
<tr><th>Activity</th><th>Distribution</th><th>Points</th></tr>
<tr><td>Self-assessments (SA)</td><td>3 points per week for 6 weeks</td><td>18</td></tr>
<tr><td>Feedback surveys (FS)</td><td>~0.67 point per week for 6 weeks</td><td>4</td></tr>
<tr><td>Discussion posts (DP)</td><td>1.5 point per week for 6 weeks</td><td>9</td></tr>
<tr><td>Project artifacts</td><td>3 deliverables of 5 points each</td><td>15</td></tr>
<tr><td>Project presentation</td><td>3 presentations of 5 points each</td><td>15</td></tr>
<tr><td>Peer rating</td><td>3 ratings of 3 points each</td><td>9</td></tr>
<tr><td>Case Studies</td><td>3 case studies for 10 points each</td><td>30</td></tr>
<tr><td>Total</td><td></td><td>100</td></tr>
</table></td></tr>
<tr><td>Learning / Course Objectives</td><td>• Understand the unique nature of Data Science projects and how they compare against Software Engineering projects
• Develop an appreciation for how software development methodologies have evolved. Analyze their suitability or lack thereof for data science projects
• Compare and analyze various data science methodologies and their application in real world projects
• Trace the development of a data-science product starting with an organizational initiative to its deployment.</td></tr>
<tr><td>Grading Scale</td><td>A+ 97-100% B+ 87 - 89% C+ 77 - 79%
A 93 – 96% B 83 - 86% C 73 - 76%
A- 90 - 92% B- 80 - 82% C- 70 - 72%</td></tr>
<tr><td>Course Policies & Expectations</td><td>Graded activities

1. Weekly self-assessments (SA), Feedback Survey (FS), and Discussion Posts (DP): There is a significant part of course-content provided in the form of readings and videos that you are expected to peruse</td></tr>
</table>

	each week. You are expected to perform the following graded activities after perusing these materials before Sunday midnight US-Eastern Time. a. Self-assessment: Complete weekly self-assessment quizzes for which you will get two attempts. The higher of the two scores will be considered for grading. b. Feedback Survey: Take the weekly feedback survey. This is a channel for me to get your feedback on your learning experience each week, and for you to ask any questions. c. Discussion Posts: Respond to the question posted in the Weekly discussion forum. 2. **Project**: In this activity, you need to work with a team to develop a small data science product while tracing the stages of its development lifecycle. You will be graded based on the following elements: a. Project artifact: Your team will submit three artifacts as per the instructions provided to you. b. Project presentation: Each of the three artifacts needs to be presented in a video recording of up to 10 minutes by each team. c. Peer rating: Part of your grade will be based on peer-rating by your team members in a survey administered after each artifact and presentation is submitted. 3. **Case Studies**: There will be three case studies assigned to you throughout the course. You will be assigned some activity related to the case study that needs to be completed before the due date. The specifics will vary with each case study **Grades**: Grade disputes, if any, must be reported to the TA or the instructor within one week from the day of grade-distribution. Copying from any source without citation, sharing your work with other students, or copying from other students will be considered as cheating and plagiarism and will be addressed according to the university policies http://www.cmu.edu/academic-integrity/. You are responsible for being familiar with the university standard for academic honesty and plagiarism. Please see the CMU Student Handbook for information. To deter and detect plagiarism, online tools and other resources are used in this class.

Course Schedule / Topical Outline: (Subject to change as needed)

Wk#	Starting on	Topics	Course activities *	
1	13-May-23	• Why methodologies • Data science vs Software Engineering	• SA1, FS1, DP1	• Project survey
2	20-May-23	• Software Engineering methodologies in Data Science	• SA2, FS2, DP2 • CQ1	• Project kickoff (team)
3	27-May-23	• CRISP-DM • Project: Where to begin? Development architecture	• SA3, FS3, DP3 • Case study 1	
4	3-Jun-23	• Data science project trajectories • Project: Deploy and consume the model	• SA4, FS4, DP4 • CQ2	• Project Report 1 • Checkpoint review 1
5	10-Jun-23	• Best practices • Project: From Dev to Staging	• SA5, FS5, DP5 • Case study 2	• Project Report 2 • Checkpoint
6	17-Jun-23	• From Dev Ops to Data Ops, MLOps, and XOps	• SA6, FS6, DP6 • Case study 3	
7	24-Jun-23	Wrap up!	CQ3	• Project Report 3 • Checkpoint review 3

*Please see Canvas for exact due dates and times.

Students with Disabilities:
Our community values diversity and seeks to promote meaningful access to educational opportunities for all students. CMU and your instructors are committed to your success and to supporting Section 504 of the Rehabilitation Act of 1973 as amended and the Americans with Disabilities Act (1990). This means that in general no individual who is otherwise qualified shall be excluded from participation in, be denied benefits of, or be subjected to discrimination under any program or activity, solely by reason of having a disability. If you believe that you need accommodations for a disability, please contact us ASAP, and we will work together to ensure that you have the correct access to resources on campus to assist you through your coursework and time at CMU.

Academic Integrity:
Carnegie Mellon University sets high standards for academic integrity. Those standards are supported and enforced by students, including those who serve as academic integrity hearing panel members and hearing officers. The presumptive sanction for a first offense is course failure, accompanied by the transcript notation "Violation of the Academic Integrity Policy." The standard sanction for a first offense by graduate students is suspension or expulsion. Please see http://www.cmu.edu/academic-integrity/ for any questions. You are responsible for being familiar with the university standard for academic honesty and plagiarism.

Specifics for this course:
All assignments in this course are meant to be your individual work, unless assigned as a group activity. Copying from any source without citation, sharing your work with other students, or copying from other students will be considered as cheating and will be addressed according to the university policies. In order to deter and detect plagiarism, online tools and other resources are used in this class.

- **Acceptable**
  - You may discuss the requirements of the assignment, but not specifics, such as code
  - You may refer to code samples from the textbook, lectures and class handouts
- **Not Acceptable and Considered Cheating**
  - You may not discuss specific code in labs and homework
  - You may not look at or copy other's assignment code, in whole or in part
  - You may not have someone else write code for you
  - You may not copy code you find on the web
  - You may not submit another's work as your own
  - You may not have in your possession other students' assignments or exams from the current or past semesters
  - You may not share your assignment code with others
  - You may not use an alternate, stand-in, or proxy during an exam
  - You may not receive help from someone else during an exam

If students are found to be sharing code, both the student who shared their code and the student who used the code will be found in violation of the academic integrity policy. All students involved will be penalized equally.

Policy Regarding Students Using English as a Foreign Language:
Assignments in this course are graded with reference to evidence of the acquisition of concepts, presentation format, and accuracy of information. Having done business in countries that use languages other than English, we understand that the use of an unfamiliar language can result in unusual word choices or grammatical errors that are not critical to the overall understanding of the information. Therefore, we will take into account your need to function in a language that may be unfamiliar to you. We will provide feedback as appropriate if we feel that language or grammar you have used in assignments would be best if it were configured in a different way.

Use of Canvas System for this course: In this course, we will use the Canvas system generally to post lecture notes and related documents and to receive assignments electronically from students. To access Canvas, go to https://cmu.instructure.com