MSDS 6120: Capstone A
Course Syllabus

Welcome to the Capstone A Course

Welcome to MSDS 6120, Capstone A. The objective of this course is to provide students the opportunity to solve a complex, ill-defined and open-ended problem in the realm of data science. The objective of this course syllabus document is to acquaint you with the administration, procedures, and policies of this course. Refer to this document throughout this course and save this document for future reference.

Course Overview

The Master of Science in Data Science (MSDS) program provides a rigorous education in the foundation of data science for both data analysts and data engineers. The Capstone Courses (MSDS 6120 and MSDS 6130) are a two-sequence set of courses designed to have student groups solve a significant problem. Students may define a problem statement of their choosing or may attempt to work on a sponsored project. Students will find one or more advisors for their project to guide them along their course of problem solution. Students report out their work through a combination of lightning and poster presentations at the conference held during their second (MSDS 6130) term and a journal paper published in the SMU Data Science Review Journal.

In this course (Capstone A), students begin the definition and solution of a significant ill-defined and open-ended problem, similar to problems a data scientist may be asked to solve. The primary course activity requires the students to form teams, identify an advisor, define the problem with specificity, and to begin the solution of the problem. Students present their problems in Café Talks to give and obtain feedback and peer guidance on their problem and solution approach. There may also be helpful presentations which will aid in students with their deliverables. Presentations may be in PowerPoint or video form, and they will be spread throughout the course at the professor's discretion.

Course Student Learning Outcomes

Learning outcomes, or learning goals, are what you are able to do as a result of the activities that you participate in and complete during this course. The primary learning outcomes of this course are:

1. The learner will be able to define with specificity a concrete problem statement, given a question or problem domain.
2. The learner will be able to solve ill-defined and open-ended problems.
3. The learner will be able to understand the basics of ethical theory.
4. The learner will be able to discuss ethical and privacy issues in the use and collection of large data sets.
5. The learner will be able to discuss the impact of data analysis on society.
6. The learner will be able to write technical papers that present their work in a clear, comprehensible manner.

This course supports, through its various synchronous, asynchronous, and other activities, broad general learning outcomes that are supported by the Master of Science in Data Science program, including:

1. An ability to design and conduct experiments that yield relevant and reproducible data.
2. An ability to manage and clean data sets.
3. An ability to apply knowledge of data analytics to explore and identify relevant information contained within a data set.
4. An ability to design a system or algorithm to meet desired needs within realistic constraints.
5. An ability to function on teams using data science tools and technologies.
6. An ability to identify, formulate, and solve data science problems based on a fundamental understanding of

concepts of data science.
7. An ability to communicate effectively in both oral and written form.
8. Knowledge of the broad foundational data science education necessary to understand the impact of data science solutions in a global, economic, environmental, and social context.
9. Knowledge of contemporary issues in data science.
10. An ability to use the techniques, skills, and modern data science tools necessary for data science practice.

Course Instruction Using Synchronous and Asynchronous Sessions

The course uses a combination of asynchronous videos, materials, and synchronous meetings to teach students the course material and guide them through the learning process.

Course Prerequisites

A student taking MSDS 6120 must be enrolled in the Master of Science in Data Science program at SMU, and they must have passed both MSDS 6110 and MSDS 7331 with a grade of C- or better.

Course Textbook and Other Course Material

There is no required textbook for this course. Assigned readings are made available on the Learning Management System used for this course.

Technology Requirements

MSDS 6120 is a course taught online with Synchronous and Asynchronous portions requiring the transfer of video. Students are expected to have access to a computer with reliable, high-speed Internet access. Students are expected to have access to a computer with a web camera with the computer capable of running the required software to access the Learning Management System, to read online documents, to watch course videos, and to participate in the Synchronous classes (including being on camera). Students are also expected to have access to a reliable phone connection in order to participate in the Synchronous classes.

All students enrolled in SMU have an SMU email account. Notifications from the Learning Management System and from the course instructor utilize your SMU email account. Students are encouraged to check this email regularly.

Course Access

This course is accessible to registered students in the SMU MSDS program only. Course asynchronous material, course information, and course communications occur through the 2DS Learning Management System. Access to the 2DS Learning Management System is available at https://2ds.datascience.smu.edu/.

Students who experience technical issues with the Learning Management System or the AdobeConnect classroom should contact technical support as described below.

Students will have access to only those courses and course sections in which they are currently enrolled or have been enrolled in previous terms. Access to other sections is at the discretion of the section instructor. Access to recordings of synchronous sections where the student did not participate or was not an enrolled student are prohibited to protect the privacy of the students that do attend and participate.

Communication and Technical Support

Direct communications with the instructor should be made in the manner indicated by the instructor. General questions and questions that are relevant to multiple students, that is, questions that are not specific to an individual and involve that individual's private information, should be posted on the course wall.

Technical support for the Learning Management System and the online classroom may be reached 24 hours a day, seven days a week via:

- *Chat Support*: Click on 'Live Support' in the lower right-hand corner of the 2DS screen after logging into the system to chat with a technical support representative. Chat support generally responds and engages in five minutes or less.
- *Phone*: Students should call 1-844-768-5637 (Toll Free) to speak with a Technical Support Representative.
- *Email*: studentsupport@datascience.smu.edu to initiate a support request with a Technical Support Representative.

For other questions or concerns, please contact the appropriate SMU department for your questions or concerns or send email to datascience@smu.edu.

It is the student's responsibility to ensure that all communications are received or acted upon.

Course Procedures and Policies

This course has a number of policies and procedures that students should understand and follow if appropriate. The following sections present the general course policies and procedures that students must follow. Additional policies and procedures may be given by the instructor. Please discuss as early in the term as possible with the instructor any questions or concerns that you may have regarding the course procedures and policies as defined herein or any additions made by the instructor to the course procedures and policies. All policies are subject to change by the professor.

Course Grading Policy

This course consists of a number of asynchronous videos, synchronous meetings, and deliverables throughout the term. It is expected that all students will put forth the effort required to earn an 'A' letter grade for this course. Assignment grades will be determined using evaluation rubrics. You are responsible for reviewing the rubrics and raising questions or concerns related to the assignments, their rubrics, and their grading prior to the submission of each assignment. Questions regarding the grading of any assignments should be directed to the course instructor as soon as possible and in accordance with any regrading policy instituted by the instructor.

The final grade for the course will be calculated on the bases of the earned cumulative percentage and the grade received for each of the components of the cumulative percentage. This course is NOT graded on a curve. The required cumulative percentage needed to earn each letter grade is given in Table 2.

Table 2: Cumulative Percentage Required to Reach Each Letter Grade

Cumulative Percentage	Earned Grade
100 – 92	A
92 – 90	A-
90 – 88	B+
88 – 82	B
82 – 80	B-
80 – 78	C+
78 – 72	C
72 – 70	C-
70 – 60	D
< 60	F

The cumulative percentage for the course is determined by the course assignment components with their corresponding percentages defined in Table 3.

Table 3: Grade Components and Weightings of the Cumulative Percentage

Percentage of Cumulative Percentage	Component
10%	Group Formation and Initial Draft Zero

10%	Presentation of Idea
50%	First Paper Draft
15%	Meeting with Advisor and Instructor
15%	Other Course Participation

You will receive a grade for each component. If you earn less than 60% (grade of D) in any one of these components, you will receive a final grade of F for this course.

A course grade of *Incomplete* (I) will be given only in the case of extraordinary circumstances that prevent the student from finishing the semester. Students must have completed at least 50% of the course with a passing grade to be eligible for an *Incomplete* grade.

Grade Grievance Policy

Students are responsible for saving all graded materials as evidence in case of a discrepancy with the assigned grades. Students are responsible for ensuring that all grades are correctly reflected on the grade store. Any identified discrepancies should be brought to the attention of the instructor as soon as the discrepancy is found.

Refer to the university catalogue for the university policy and process for grade grievances.

Assignment and Collaboration Policy

Data Science is an inherently collaborative subject, and learning often occurs best when subjects are taught both to and from peers. Collaboration is expected to occur both in learning the course material and in performing the course work. However, each student must hand in their own work performed by themselves unless explicitly allowed by written directions given by the instructor. Collaboration means helping one another learn the material. Collaboration does not mean copying answers from one another.

Assignment submissions that contain substantially the same answers shall receive a grade of zero on the first instance and a course grade of F upon a second instance. In order to mitigate potential issues and questions of similarity, peers with whom a student collaborates should be clearly identified by that student in their submissions.

Scholarly Expectations

Work submitted at the graduate level is expected to demonstrate critical and creative thinking skills and be of significantly higher quality than work produced at the undergraduate level. To achieve this expectation, all students are responsible for giving and receiving peer feedback of their work. Students are also expected to resolve technical issues, be active problem solvers, and to embrace challenges as positive learning opportunities. Data Science professionals must be able to teach themselves and teach others to fill in any gaps in their knowledge or to find a way of learning new material that is most conducive to their learning style. Data Science professionals must also be able to work cooperatively and collaboratively with others – skills that students are expected to practice in this course. Students are expected to ask questions and ask for help when they need it and to offer help when others are in need.

Absent questions or requests for assistance, instructors must assume that students understand the material being covered and are able to complete the assignments. It is primarily through your questions that the instructor learns where the students are struggling to understand and on which topics more time needs to be spent for the students' benefit.

Timeliness

Because a 15-week term goes by quickly, assignments must be submitted by the designated due dates. Full credit cannot be earned by late or incomplete assignments. Assignments may lose up to 10% of their possible value each day late if submitted after the posted due date/time. (e.g. Assignments can lose all of their value at 10 days past due.) When a project incorporates peer review, it is imperative that all projects be available at the beginning of the review period and that reviews are completed by the end of the review period so that others may incorporate feedback into project revisions. You will have plenty of notification and time to complete course assignments. If you know you are

going to be out of town, involved in a special event/project, or unable to access a computer, please plan ahead. Also ensure that you have a backup plan ready in the event you might lose power, Internet access, or your available technology.

Time Commitment

As a technical graduate level course, and a Capstone course, it is expected that students will spend a significant amount of time on this course. The time needed to define and solve an ill-defined and open-ended problem can be considerable. Additionally, all students will produce a final paper published in a peer-reviewed journal publication. Therefore, it is expected that students will spend between 5 and 15 hours per week on this course.

Attendance Policy

Attendance and on-camera participation during any presentations is highly encouraged. Scheduled meetings that are noted to be mandatory are, in fact, mandatory.

Drop Policy

Refer to the university drop policy for a complete description of the drop and withdrawal policies for this course.

Campus Concealed Carry

Concealed handguns are prohibited on the Southern Methodist University campus. Pursuant to section 30.06, Penal Code (Trespass by License Holder with a Concealed Handgun), a person licensed under subchapter H, Chapter 411, Government Code (Handgun License Law), may not enter SMU property with a concealed handgun. Report violations to the Southern Methodist University Police Department by dialing 9-1-1 or 214-768-3388 (non-emergency) or 214-768-3333 (emergency).

Americans With Disabilities Act

Disability Accommodations: Students needing academic accommodations for a disability must first be registered with Disability Accommodations & Success Strategies (DASS) to verify the disability and to establish eligibility for accommodations. Students may call 214- 768-1470 or visit http://www.smu.edu/alec/dass to begin the process. Once registered, students should then schedule an appointment with the professor to make appropriate arrangements. (See University Policy No. 2.4.)

Religious Observance

Religiously observant students wishing to be absent on holidays that require missing class should notify their professors in writing at the BEGINNING of the semester and should discuss with the professor, in advance, acceptable ways of making up any work missed because of the absence. (See University Policy No. 1.9.) Failure to notify your professor prior to your absence will result in an unexcused absence and possibly a grade of zero for any assignments.

Excused Absences for University Extracurricular Activities

Students participating in an officially sanctioned, scheduled University extracurricular activity should be given the opportunity to make up class assignments or other graded assignments missed as a result of their participation. It is the responsibility of the student to make arrangements with the instructor prior to any missed scheduled examination or other missed assignment for making up the work.

Academic Integrity

It is the philosophy of Southern Methodist University that academic dishonesty is a completely unacceptable mode of conduct and will not be tolerated in any form. All persons involved in academic dishonesty will be disciplined in accordance with University regulations and procedures. Discipline may include suspension or expulsion from the University.

Scholastic dishonesty includes but is not limited to cheating, plagiarism, collusion, the submission for credit of any work or materials that are attributable in whole or in part to another person, taking an examination for another person, any act designed to give unfair advantage to a student or the attempt to commit such acts. This includes reworking one's own projects for this class. This class should be an entirely new project for all involved parties.

Students caught being academically dishonest shall receive a grade of F for this course.

University Honor Code

When you signed your letter of intent to enroll in the MSDS program, you initialed the following statement:

> "I have read and agree to abide by the SMU Honor Code available online at:
> https://www.smu.edu/StudentAffairs/StudentLife/StudentHandbook/HonorCode"

The Honor Code is taken seriously at all levels within the university. Students that are found to have violated the honor code will be disciplined which often includes expulsion from the university.

Plagiarism

Plagiarism is the "practice of taking someone else's work or ideas and passing them off as one's own" (this definition is from Google Dictionary). An example of plagiarism is as follows:

A regression is a statistical analysis assessing the association between two variables. It is used to find the relationship between two variables.

The following is NOT plagiarism:

"A regression is a statistical analysis assessing the association between two variables.
It is used to find the relationship between two variables." (https://www.easycalculation.com/statistics/learn-regression.php).

The difference is in the punctuation and the attribution. Note that one can self-plagiarize. If you are using something that you wrote (e.g. a blog or a previously published article), please reference yourself.

DO NOT PLAGIARIZE. If you have any question as to what is and what is not plagiarism, ask your instructor. As a general rule, always use your own words and cite your source.

The consequence for being caught plagiarizing is to earn at least a zero on the identified assignment and may include earning a course grade of F and a referral to the SMU Honor Council for your Honor Code violation.

Best Practices for Success in the Course

Attendance. Take responsibility for your commitment. Attendance means not only being there for synchronous sessions but also participating in asynchronous work.

Citizenship. You need to be actively engaged to succeed in this class. Talking on cell phones, texting, "facebooking," tweeting, or leisure web browsing are prohibited in class. I consider these to be a disruption (not to mention rude).

Integrity. A lot of the graded work occurs outside of class, so I expect honesty and integrity in what you submit for evaluation. Evidence of academic dishonesty will minimally result in zeros for all involved parties, and perhaps University-level disciplinary action. Don't risk your career.

Humility. Don't get lost! Ask questions in class. If something isn't clear to you, it probably isn't clear to others either. Questions may arise because I haven't made a connection clear or have inadvertently left out an important point. Your question gives me a chance to explain more clearly. Don't be proud or shy.

Organization. Don't procrastinate! This is a technology-driven course. Count on your computer failing or your wireless connection breaking the night before a due date. Start early and give yourself a chance to succeed.

Deadlines. You will generally have a week to complete an assignment. Due dates and times will be clearly indicated. Late submissions will be penalized, but it is much better to turn in work late than not at all (or to turn in incomplete/sloppy work). Work turned in after solutions have been posted to the course website will receive no credit.

Getting help. If questions arise while doing assignments/exams, do your best to resolve these questions before the assignment is due, first by taking time to seek answers yourself, next by asking questions on the wall, and finally via email to your instructor or other students. I encourage you and expect you to seek help. For questions during exams, please email the live session instructor directly.

Collaboration. I encourage the formation of study groups and collaboration with your fellow students in tackling the assignments. Working together in groups on homework is permitted, even encouraged. However, every student should write up and complete his or her homework independently. Talking about problems with other people does help in learning, but just copying the solutions from one another doesn't help!

Looks do matter! All assignments must be NEATLY executed and organized. You risk a zero on any assignment submitted in a sloppy manner. See submission guidelines for more detail.

Have Fun! Learning is meant to be a fun activity. While it can be difficult, time consuming, frustrating, and sometimes disappointing, always seek to find the fun in what you are doing and learning. The gratification from learning complex concepts and applying them to solve hard problems is what we are all striving to achieve. Having fun while we are learning and teaching others just makes the learning easier and friendships better.

MSDS 6120 Fall 2018 Term

Section Instructor Contact Information

Instructor: Sadie Thomas, PhD
Email: sadiet@smu.edu
Office Hours: Upon Request

Assignment Expectations and Grading Rubrics Summary

The primary work deliverables for this course are the paper drafts. During the term of this course, Paper Draft Zero and Paper Draft One are submitted and graded. Very early (typically either the first or second week of the term) in the following course in this sequence (MSDS 6130 Capstone B), Paper Draft Two is due. More details associated with all drafts and the Capstone Projects are provided in a separate document (and/or announcements) and by the instructor. All questions regarding the projects and drafts should be directed to the course instructor.

All paper drafts and the final submitted Capstone paper use the same template. The template is provided in both Word and LaTeX format by the instructor. You are expected to use, and follow without modification, the provided templates. The templates follow the Lecture Notes in Computer Science formatting.

Paper Draft Zero – Draft Zero is meant to provide an early checkpoint to force teams to be formed, to force the teams to identify a problem or problem domain, to force the teams to find one or more advisors, and to force the teams to begin the process of defining and understanding the problem they are about to solve. Writing is the primary method used to both document and refine the problem statement and the work performed. Draft Zero begins that documentation.

Note that Draft Zero is exactly that, the very first beginnings of the paper that will become your Capstone paper published in the SMU Data Science Review Journal. This draft, and all drafts, should be written as a draft for your final paper. Include only the material that will be included in your final paper. Do not include information on timelines, milestones, project plans, etc. None of that information will be in your final paper, so do not include it in any of your drafts. If you do wish to include notes, timelines, or any other information that either may not or will not be in your final paper, create one or more Appendices and place the material in there. Do not reference Appendices from the main body of your paper.

Draft Zero is expected to be at least one page in length, including the draft title, author list (include all advisors as authors in the list), abstract draft, and, within the Introduction section, one paragraph motivation and one paragraph problem statement. Teams are expected to have performed a related work and tutorial topic search. This information should be included in Draft Zero in some form. A preliminary set of sections with their titles should be created to indicate the general and expected flow of the paper. This is particularly important for you to identify the tutorial sections required to ensure that the general audience reader will be able to understand your work.

There is no maximum page limit on your final paper; however, all papers are expected to be at least fifteen (15) pages in length, excluding any appendices, in their final form. All papers are expected to have at least twenty (20) peer reviewed references in their final form. These may sound like huge numbers now, but they are actually very small numbers in practice. Very few teams have difficulties reaching these minimum requirements, and those that do have difficulty have that difficulty because they choose to fight the requirements instead of embracing them. As you move forward in your documentation, keep these final lengths and reference requirements in mind. You will have no difficulty in reaching them if you are diligent in your documentation and your problem solution.

Table 6 contains the grading rubric used for Draft Zero.

Table 6: Draft Zero Grading Rubric

Component	Percent of Draft Zero Grade
Using the Correct Template and Using it Correctly	40%
Advisor in Author List	20%
Problem Statement	20%
One Plus Page and Sections Drafted	20%

Paper Draft One – Draft One is meant to provide a mid-project checkpoint on the progress of teams in further defining and refining the problem statement and in solving the stated problem. In Draft One, you will continue to use the given template, and you will have made significant progress in documenting the tutorial material, related work, solution approach, and may even have some very preliminary data.

Draft One is expected to be at least eight (8) pages in length. A preliminary Abstract draft should be written with placeholder sentences for the main result and main conclusion. The Introduction section should be written with placeholder paragraphs for the results and conclusions. The Introduction should be 2 to 4 pages in length in the given format. Remember that both the Abstract and the Introduction are executive summaries of your work. It is strongly recommended that you have someone who is not on your team read and comment on both your Abstract and Introduction section prior to your Draft One submission.

Draft One should include all of the tutorial sections in well-written draft form. Your paper is targeted at a general technical audience (think – students who are just beginning the MSDS program but have not yet taken any classes in the program, except the Stats Bridge and the Programming Bridge). Therefore, background sections on your problem domain are necessary for all readers to be able to understand your work properly. Draft One should include early draft sections on the algorithms or techniques being used or developed, the data being used and its collection methodology, the solution approach and evaluation methodology, and related results from prior publications.

Draft One is expected to be a draft with imperfect grammar and all that normally comes with a draft. However, it is expected to be a draft document and not simply a collection of notes. All citations should be correctly used (citations are always contained within the same sentence that they are citing, i.e., the citations come before the period ending the sentence – preferably right next to the words needing to be cited). Plagiarism in Draft One will result in a grade of 'F' for the course and a referral to the SMU Honors Council for an Honor Code violation.

Table 7 contains the grading rubric used for Draft One.

Table 7: Draft One Grading Rubric

Component	Percent of Draft One Grade
Using the Correct Template and Using it Correctly	40%
Clear Problem Statement	10%
Introduction Well Written	10%
Tutorial Sections Drafted and Complete	20%
Eight Plus Pages and All Sections Drafted to a Reasonable Level of Detail	20%

Paper Draft Two – Draft Two is due at the beginning of the following term, i.e., the beginning of the MSDS 6130 Capstone B term. However, the work necessary to complete Draft Two by the beginning of the next term must be performed during this term. Do not delay in continuing your work on your Capstone project. Do not wait for comments on Draft One. Draft Two is expected to be a significantly more polished draft than Draft One, and it is meant to be approaching (but not quite) completed work, at least in many of its sections.

Draft Two is meant to provide a second mid-project checkpoint on the progress of teams in further defining and refining the problem statement and in solving the stated problem. In Draft Two, you will continue to use the given template, and you will have made significant progress in documenting, in well-polished prose and figures and tables, the tutorial material, related work, solution approach, data, early results, analysis of those results, ethics, and early conclusions.

Draft Two is expected to be at least twelve (12) pages in length. A preliminary Abstract draft should be well written with placeholder sentences for the main result and main conclusion if early results and conclusions have not been obtained already. The Introduction section should be written with draft paragraphs for the results and conclusions. The Introduction should be 2 to 4 pages in length in the given format. Remember that both the Abstract and the Introduction are executive summaries of your work. It is strongly recommended that you have someone who is not on your team read and comment on both your Abstract and Introduction section prior to your Draft Two submission.

Draft Two should include all of the tutorial sections in well-written, nearly polished form. Your paper is targeted at a general technical audience (think – students who are just beginning the MSDS program but have not yet taken any classes in the program, except the Stats Bridge and the Programming Bridge). Therefore, background sections on your problem domain are necessary for all readers to be able to properly understand your work. Draft Two should include (polished) draft sections on the algorithms or techniques being used or developed, the data being used and its collection methodology, the solution approach and evaluation methodology, related results from prior publications, your results, your analysis, ethics, and early conclusions.

It is expected that additional work will be performed after the submission of Draft Two. This additional work will result in additional results, possibly additional algorithm development or modification, and other expansions of the work involved in solving the problem and providing evidence of solution goodness. For this reason, it is expected that some sections will be incomplete, may not exist (if their existence is determined to be needed after Draft Two submission), may disappear or be merged, and the entire paper may be completely reorganized at any point in time. While Draft Two is expected to begin approaching final form, significant changes may occur after Draft Two submission. Comments from the instructor may engender such large additions or modifications.

Draft Two is expected to be a draft with good grammar and all the issues that normally come with a second draft. However, it is expected to be a draft document approaching final form. All citations should be correctly used (citations are always contained within the same sentence that they are citing, i.e., the citations come before the period ending the sentence – preferably right next to the words needing to be cited). Plagiarism in Draft Two will result in a grade of 'F' for the course and a referral to the SMU Honors Council for an Honor Code violation.

Table 8 contains the grading rubric used for Draft Two.

Table 8: Draft Two Grading Rubric

Component	Percent of Draft Two Grade
Using the Correct Template and Using it Correctly	**40%**
Clear Problem Statement	**10%**
Abstract and Introduction Well Written	**10%**
Tutorial Sections Well Written	**20%**
Twelve Plus Pages and All Sections Drafted to a Reasonable Level of Detail	**20%**