NAZARBAYEV UNIVERSITY
SCHOOL OF ENGINEERING AND DIGITAL SCIENCES

Department of Computer Science

Program Handbook

Master of Science in Data Science

Academic year 2020-21 and onwards

Welcome Note

Dear Students!

It is my great pleasure to welcome you to Nazarbayev University and the Master of Science in Data Science program. I wish you all the best and thank you for making NU your choice of postgraduate studies.

There are several reasons why students choose to pursue their higher education. Many do it for career advancement, while others do it for self-satisfaction. Whatever your reason, our diverse and experienced staff at Nazarbayev University will provide you with first-rate education and research knowledge. Just as you are proud of the high quality of professional work that we know you can do, we are proud of the success and reputation of our graduate programs. As such, we have aligned the courses that you are about to undertake with the recommendation of higher education accreditation bodies and local and international information technology companies to respond research expectations.

The Data Science program will run over 4 semesters with a duration of 2 years where you will need to complete 120 ECTS credits. A master's thesis begins in the 3rd semester (1st semester of the 2nd year) and ends in the last semester of the master's course where you will be asked to demonstrate your knowledge of current literature in the field; defend your thesis proposal; develop a solution to a real problem by developing a model and/or an algorithm and/or a software system, analyze and evaluate the results of the solution you offer; support your conclusions in a scholarly manner according to disciplinary standards and finally, defend your thesis work. With this in mind, it is recommended to start planning and interacting with the supervisors of your choice at the start of the program so that you can discuss your research topic, organize supervised meetings and receive advice on how to complete your thesis on time.

I wish you success in your Master of Science program!

Adnan YAZICI, Prof. Dr.

Chair of Department of Computer Science,

School of Engineering and Digital Sciences, Nazarbayev University,

Nur-Sultan / Kazakhstan

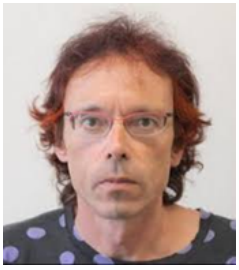
Contents

Data Science Faculty

The Data Science program is interdisciplinary by design. While based primarily in the Department of Computer Science, the largest department in the university, the program was designed by senior leadership of the related disciplines, as noted below:

Dr. Adnan Yazici, Chair of the Department of Computer Science.

Dr. Atakan Varol, Chair of the Department of Robotics and Mechatronics



Dr. Christian Schoenbach, Chair of the Department of Biology

Prof. Michael Lewis, Director of the Data Science Program

The core team is comprised of the faculty of the Department of Computer Science (shown below), with topical expertise and research interests across the full spectrum of Data Science topics, and all stages of the Data Science project life-cycle.

The Data Science is further augmented by specialists throughout the university, from the fields of Biology, Robotics, Mathematics, Physics, and beyond. The full roster of faculty with backgrounds, areas of expertise, and current research interests can be found on the web page.

Program Overview

Digitalization is a phenomenon which now permeates nearly all human endeavor and most aspects of daily life, encompassing the spheres of education, finance, health, industry, and culture.

It is now possible to collect data on the scale of Exabytes (1 billion gigabytes), from an incredibly wide range of digital sources, such as cameras, sensors (thermal, light, motion), smart phones, and medical devices. In addition to the data sources, we have enormous volumes of the routine transactional data of digitalized daily life, data that was previously uncollected or, if collected, ignored.

It is not just the data but the meta-data about the data, including the details of how, when, and where it was collected, that facilitates the identification of patterns, and, conversely, anomalies, in ways that are increasingly useful. This phenomenon of high-volume data collection is motivating the creation of advanced analytic techniques that can achieve results across disciplinary boundaries.

What began as "Big Data" has become "Data Science, enabling significant advances in applications ranging from health care (improved diagnostics and outcomes) to energy consumption (smart homes, factories, and cities). The ubiquity of digital devices is greatly expanding both the quality and quantity of data generated, with an effect amplified by the rapid emergence and utilization of the IPv6 protocols and the resulting plethora of "internet-of-things" applications.

The exponential growth in data sources and volume has presented non-trivial challenges to every step of data management: collection, storage, validation, preservation, transmission, access, analytics, and also raises new challenges to the related issues of anonymity, privacy, and security.

Data Science is the scientific discipline that covers the full range of the data life cycle. It includes both theoretical and practical methods for organizing, processing, and analyzing the data and transforming data into information and, increasingly, information into actionable "intelligence".

In this regard, "intelligence" is feasible through major innovations in the application areas of Artificial Intelligence, Machine Learning, and Deep Learning, which are benefitting from new models of cognition and learning, and substantial improvements in computing resources and methods to manage computational complexity.

The result is that Data Science is achieving results that were previously conceived only speculatively or in the literature of science fiction. As but a single example, a machine recently taught itself chess over a weekend, and proceeded to defeat the top human players in the world.

Similar demonstrations are emerging in more practical application areas, such as medical imaging diagnostics, where AI analytics are achieving more accurate recognition rates than human experts.

The field of Data Science is a rapidly growing interdisciplinary specialty that is directly relevant to the national development priorities for Kazakhstan.

The Master of Science in Data Science at Nazarbayev University will provide in-depth education in Data Science, incorporating key concepts from the fields of Computer Science, Mathematics, Robotics, and related engineering disciplines.

The program includes the following subject areas: Databases, Data Mining, Big Data, Business Analytics, Artificial Intelligence, Information Retrieval, Machine Learning, Deep Learning, Image and Video Processing, Bioinformatics, Cybersecurity, Data Analysis and Visualization, Mathematical and Statistical Modeling, Data Storage and Processing Infrastructures, and Cloud-based Solutions.

Aims and Objectives

The Master of Science in Data Science is a two-year full-time program (120 ECTS credits) at the School of Engineering and Digital Sciences (SEDS) of Nazarbayev University.

The first term of study is designed to provide a foundation in the fundamentals of the field, and to provide a baseline for all incoming admissions streams from several related but distinct fields of undergraduate study.

The program includes discipline core courses, and a set of topical electives that provide the continuity of study across the two years. The discipline courses are augmented by an arc of courses that provide milestones for program advancement and completion of the Master Thesis Project. This arc consists of a course in Research Methods, Research Seminar, Thesis Proposal, and then a final term to conclude and defend the work.

The program is aligned with professional society guidance from the ACM and IEEE disciplinary societies, and incorporates specific topical coverage as requested from Kazakh government and industry partners. The program provides a framework for collaboration with partners from education, government, and industry, so as to align opportunities for targeted research and collaboration on projects related to national development priorities as articulated in a series of national strategy documents such as Digital Kazakhstan, Kazakhstan 2030, the 100 Steps, and Kazakhstan 2050.

The Data Science program is unique in Kazakhstan, due to its interdisciplinary focus and a pedagogical approach which integrates research and emphasizes innovation. Starting from the first semester students interact closely with faculty. Graduates of the program are trained to become data science professionals prepared to enter careers in industry, government, or education.

Graduate Attributes

The Data Science MSc program immerses the student within a scientific and technological ecosystem to favor their professional growth according to NU Graduate Attributes, which are very well aligned with the vision and mission of NU. These attributes are listed as:

1) Possess an in-depth and sophisticated understanding of their domain of study;
2) Be intellectually agile, curious, creative, and open-minded;
3) Be thoughtful decision makers who know how to involve others;

4) Be entrepreneurial, self-propelling and able to create new opportunities;
5) Be fluent and nuanced communicators across languages and cultures;
6) Be cultured and tolerant citizens of the world;
7) Demonstrate high personal integrity;
8) Be prepared to take a leading role in the development of their country.

The MSc program delivers these attributes by providing the students opportunities to be involved in: (a) working on individual and group assignments; (b) team-building exercises for developing decision-making skills; (c) designing tasks for developing creativity; (d) delivering and attending project-research presentations to polish their communication skills; and (e) engaging on group discussions among students and faculty in order to develop personal integrity and cultural tolerance. The same attributes are also addressed by the program learning outcomes that follow.

Program Learning Outcomes

On successful completion of the program, graduates will be able to:

1) design and execute research methodologies and communicate strategies and outcomes to broader audiences of both technical and non-technical nature (LO1);
2) conduct complex ICT projects in their area of expertise in accordance with ethical and professional standards, prepare and present professional posters, lectures, and publications of the research findings in conference and journal settings (LO2);
3) demonstrate comprehensive knowledge of advanced data science and ICT technology concepts by successfully completing advanced courses and seminars (LO3);
4) identify authoritative sources and conduct relevant search and discovery of topical literature sources and tools to be used in the design and development of solutions to complex problems in their domain of discourse (LO4);
5) identify and apply appropriate methodologies in successful execution of complex software projects and implement analytical, statistical and/or numerical solutions of data-driven or theoretical question related to IT-related problems (LO5);
6) demonstrate expertise with the phases and stages of the research process through successful presentation of a research seminar and/or proposal writing (LO6);
7) evaluate the relation between computer and data science concepts and state of the art IT technologies and how this drives innovation; communicate competently with expert audiences (LO7);
8) explain scientific solutions for complex problems, concepts and research findings, using various modalities of communication, with particular emphasis on tertiary education instruction (LO8).

Tabulated Program Learning Outcomes against NU Graduate Attributes:

		Program Learning Outcomes							
		1	2	3	4	5	6	7	8
NU Graduate Attributes	1	●	●	●	●			●	
	2	●	●		●			●	
	3		●			●	●		●
	4	●	●			●	●	●	
	5			●	●			●	●
	6			●	●	●			●
	7		●			●	●		
	8	●	●					●	●

Program Duration

The nominal MSc program duration is **two years**, while the maximum allowable duration can be extended up to **two and half years** (excluding leave of absence and deferment of admission; see "*ACADEMIC POLICIES AND PROCEDURES FOR GRADUATE PROGRAMS OF THE AUTONOMOUS ORGANIZATION OF EDUCATION 'NAZARBAYEV UNIVERSITY'*" for further details).

Assessment

The Master of Science in Data Science program will use the full range of outcome measures within comprehensive multimodal assessment methods (AM) of student achievement of the competencies:

Program Assessment Methods	Name of Assessment Method
AM1	Written or oral examination
AM2	Home assignment
AM3	Presentation of research papers
AM4	Research proposal writing
AM5	Thesis writing and presentation
AM6	Team project assignment
AM7	Thesis research
AM8	Supervision of Master Thesis research

The specific assessment methods include:

1. Subject examination
2. Performance-based assessment of student skills and abilities
3. Performance in small group sessions
4. Master thesis proposal defense in the graduate seminar
5. Performance on the research
6. Oral Master thesis defense at the end of Year 2

Taken together, the combination of knowledge assessments, performance-based assessments, faculty observation, peer assessments, and the research will provide a comprehensive and cumulative portfolio of information on the progress of students and their achievement of the competencies. The multimodal assessment information will be used in a formative and ultimately a summative fashion towards the performance of students.

		Program Assessment Method							
		AM1	AM2	AM3	AM4	AM5	AM6	AM7	AM8
Program Learning Outcomes	LO1	●	●	●	●	●			
	LO2			●	●	●			
	LO3		●				●		
	LO4			●			●		
	LO5		●		●		●	●	
	LO6			●	●	●	●	●	●
	LO7	●		●	●	●		●	●
	LO8	●		●	●	●		●	●

Assessment is aligned with the learning outcomes of the program and of those of each course. Course assessment tasks are performed during and at the end of each course. Types of assessment vary form successful completion of integrated coursework, assignments, and project work to evaluation of performance of case studies, interviews, and deliverance of presentations.

The following table summarizes assessment and evaluation points for all stages of the MSc program:

Stage of Program	***Significance***	***Possible Results***	***Evaluation Point***
ADMISSION TO PROGRAM	*Initial Evaluation*	*Admission* *Admission with Conditional Status, Subject to Satisfactory Completion of Conditions* *Rejection*	***Key Evaluation Point*** *Admission is handled on a case-by-case basis by evaluating the student's undergraduate curriculum, English proficiency and letters of recommendation among other documents and interview (only for shortlisted candidates)*
COURSEWORK	*Determination of Student Competence in Fundamentals of Discipline*	*Continue in Program* *Continue on Probation* *Dismissed from Program*	***Continuous Evaluation*** *The coursework component for the Master of Science is assessed by the module instructor. It is enforced that all faculty provide a module descriptor to students at the start of the course outlining the weight of each assessment.*
DEGREE CANDIDACY	*Demonstration of Student's Mastery of Content Knowledge and Skills in the Discipline*	*Pass and Continue in Program* *Required to Re-Take Some Courses* *Dismissed from Program*	***Key Evaluation Point***

COMPLETION OF THESIS PROJECT	*Demonstration of Student's Mastery of Content Knowledge and Skills Needed to Graduate*	*Pass* *Recommend Changes with or without re-defense* *Fail and dismissal from Program*	***Key Evaluation Point***

Coursework Assessment methods by course & correspondence to Program Learning outcomes are summarized in the following table:

Program Learning Outcome	***Where addressed*** *(course)*	***How addressed*** *(L&T Methods)*
1.	*Research Methods, Research Seminar, Thesis Proposal, Thesis*	1. *Lectures* 2. *Workshops* 3. *Individual and Group Presentations* 4. *Individual and Group Projects* 5. *Paper writing and presentation.* 6. *Exams* 7. *Flipping/Blending Learning in which on-line and in-class classes are combined resulting in an effective technique to improve the process of learning.*
2.	*Research Methods, Research Seminar, Thesis Proposal, Thesis*	
3.	*Fundamentals of Data Science, Statistical Learning, Probability and Statistics for Data Science, Database Management Systems, Process and Project Management, Data Mining and Decision Support, Big Data Analytics*	
4.	*Research Methods, Research Seminar, Thesis Proposal, Thesis*	
5.	*Fundamentals of Data Science, Statistical Learning, Probability and Statistics for Data Science, Data Mining and Decision Support, Big Data Analytics*	
6.	*Research Methods, Research Seminar, Thesis Proposal, Thesis*	
7.	*Fundamentals of Data Science, Database Management Systems, Process and Project Management, Data Mining and Decision Support, Big Data Analytics, Innovation and Entrepreneurship, Data-Driven Innovation*	
8.	*Research Methods, Research Seminar, Thesis Proposal, Thesis, Innovation and Entrepreneurship*	

MASTER OF SCIENCE - PROGRAM CALENDAR YEAR-1

Course-type key

Program Core courses
Program Elective courses

SEMESTER 1	FALL	August – December
TYPE	**COURSE CODE & TITLE**	**ECTS**
Program core	DS 501 FUNDAMENTALS OF DATA SCIENCE	6
Program core	DS 502 PROBABILITY AND STATISTICS FOR DATA SCIENCE	6
Program core	DS 507 DATABASE MANAGEMENT SYSTEMS	6
Innovation	DS 551 PROCESS AND PROJECT MANAGEMENT	6
Research	SEDS 591 RESEARCH METHODS	6

SEMESTER 2	SPRING	January – May
TYPE	**COURSE CODE & TITLE**	**ECTS**
Program core	DS 504 DATA MINING AND DECISION SUPPORT	6
Program core	CSCI 545 BIG DATA ANALYTICS	6
Research	SEDS 592 RESEARCH SEMINAR	6
Elective	ELECTIVE 1 - SEDS 504 INNOVATION AND ENTREPRENEURSHIP (RECOMMENDED)	6
Elective	ELECTIVE 2	6

MASTER OF SCIENCE - PROGRAM CALENDAR YEAR-2

SEMESTER 3	FALL	August-December
TYPE	**COURSE CODE & TITLE**	**ECTS**
Research	DS 693 THESIS PROPOSAL	6
Innovation	DS 552 DATA DRIVEN INNOVATION	6
Elective	ELECTIVE 3 - MATH 540 STATISTICAL LEARNING (RECOMMENDED)	6
Elective	ELECTIVE 4	6
Elective	ELECTIVE 5	6

SEMESTER 4	SPRING	January-May
TYPE	**COURSE CODE & TITLE**	**ECTS**
Research	DS 694 THESIS	30

Academic Policies and Procedures

All academic policies and procedures that are not explicitly covered in this handbook are conformant with the corresponding items described in "*SCHOOL OF ENGINEERING AND DIGITAL SCIENCES MASTERS STUDENT HANDBOOK",* which covers School of Engineering and Digital Sciences Master Programs, and the "*ACADEMIC POLICIES AND PROCEDURES FOR GRADUATE PROGRAMS OF THE AUTONOMOUS ORGANIZATION OF EDUCATION "NAZARBAYEV UNIVERSITY" (APP-Graduate Programs-NU)"*, which covers all graduate programs in Nazarbayev University. These policies and procedures include, among others, the following:

1. Admissions
2. Registration
3. Credits (Requirements, awarding & transfers)
4. Grading issues such as: administrative grades, grade appeals
5. Course re-takes
6. Degree withdrawals
7. Academic code of behavior
8. Leaves of absence, including medical reasons, immediate family member issues and others
9. Dismissal & voluntary withdrawal.

Every student participating in the Data Science program is expected to have read and understood all the policies, rules, procedures, and guidelines described in this program specific handbook, school's MSc handbook and the general APP for graduate programs in NU.

Grading System

Graded courses

Letter Grade	Grade Points	Percentage
A	4.00	95-100%
A-	3.67	90-94.9%
B+	3.33	85-89.9%
B	3.00	80-84.9%
B-	2.67	75-79.9%
C+	2.33	70-74.9%
C	2.00	65-69.9%
C-	1.67	60-64.9%
D+	1.33	55-59.9%
D	1.00	50-54.9%
F	0.00	0-49.9%

Non-graded (PASS/FAIL) courses

In the case of a non-graded course, the following assessment percentages apply

Description	Percentage
Pass	59% or Above
Fail	Below 59%

Program Completion Requirements

Satisfactory completion of the MSc program requires that the student progress through a number of distinct stages, each of which is characterized by a key evaluation point (See Appendix) The necessary stages are:

1) Satisfactory application to the program;
2) Completing all required coursework in the program (90 ECTS);
3) Satisfactory completion of the master thesis (30 ECTS);
4) Satisfactory achievement of minimum GPA for continuation through semesters and graduation (Candidacy).

Continuation / normal progress

To continue in the Data Science graduate program at SEDS, NU, a student must maintain a minimum CGPA of no less than a **B- (2.67 on a 4-point scale)** after each grading period and conform to all program rules and policies to maintain normal progress toward degree. A student who fails to satisfy the continuation requirement for the program is subject to dismissal.

Appealing against grades

If a student believes that she or he has received an unfair or erroneous grade, the student may appeal. The following are cases for appeal:

1) In the case of an examination. The student must first consult with the instructor within 5 working days of her or his receipt of the contested grade (this time may be extended in the event that the instructor can be shown to have been unavailable during the period following the student's receipt of the grade in question). The Instructor must respond within the next 5 working days. In the event that the student is still dissatisfied, she or he may appeal to the Dean of the School (or the Dean's designee) within 5 working days. The Dean (or her or his designee) shall consult with the Instructor before making any decision. The decision of the Dean (or of her or his designee) shall be final;

2) In the case of a Final Course Grade. The student must first consult with the instructor within 5 working days of her or his receipt of the contested grade (this time may be extended in the event that the instructor can be shown to have been unavailable during the period following the student's receipt of the grade in question). The date to be used for appeals of Final Course Grades is the date published in the Academic Calendar. The Instructor must respond within the next 5 working days (that time may be extended in the event the instructor is shown to have been unavailable during the period following the student's receipt of their final grade). In the event that the student still believes that the grade is incorrect, or the Instructor has not replied within 15 days, the student may appeal to the Dean of the School (or the Dean's designee) within 5 days. The Dean (or her or his designee) shall consult with the Instructor before making any decision. The decision of the Dean (or her or his designee) shall be final.

Plagiarism

In any coursework or thesis assessment, unacknowledged copying or plagiarism is not acceptable. Plagiarism can result in extremely serious academic actions including cancellation of any or all results, suspension from the program, or even expulsion. Plagiarism means using the work of others in preparing an assignment and presenting it as your own without explicitly acknowledging – or referencing – where it came from. Plagiarism can also mean not acknowledging the full extent of indebtedness to a

source. Work can be plagiarized from many sources including books, articles, the internet, and other media. Plagiarism can also occur unconsciously or inadvertently. Direct copying is definitely plagiarism. Paraphrasing of another's work without acknowledgment is also plagiarism. Submitting someone else's work or ideas without attribution is not evidence of your own grasp of the material and cannot earn you marks.

Nazarbayev University's policy on plagiarism sets out student responsibilities in regard to copying. Students are responsible for ensuring that:

- They are familiar with the expected conventions of authorship and the appropriate use and acknowledgement of all forms of intellectual material relevant to their discipline.
- The work submitted for assessment is their own.
- They take all reasonable steps to ensure their work cannot be accessed by others who might seek to submit it, in whole or in part, as their own.

Whenever you refer to another person's research or ideas -either by directly quoting or by paraphrasing them-, you must acknowledge your source by proper referencing. Turnitin is a useful web-based originality checking service that can help in assessing the originality of one's submitted work. More information on Turnitin can be found in **Error! Reference source not found.** and service's web site (http://turnitin.com/).

Description of Courses

Course-type key

Core courses
Elective courses

***Note: In future, minor changes in courses and/or program, subject to approval by SEDS Teaching and Learning Committee, may not be reflected in this document, however, these would be reflected in the corresponding Course Specification Forms.**

Program Core Courses

DS 501, Fundamentals of Data Science

This course is designed to introduction object-based programming and programming concepts with Python, Algorithm analysis and design, basic data structures, and Scientific computation and visualization with Python. Moreover, the course will include the following sub topics; Introduction to Python and its basic data types; Object-oriented programming with Python; Recursion; Algorithm analysis; Searching and sorting; Abstract data types such as linked lists, stacks, queues, priority queues, trees, hashs and their implementation in Python; Vectors and matrices in Python; Plotting and visualization with Python; Solving systems of linear equations with Python.

CLOs

By the end of the course students will be able to:
1) design basic and advanced data structures according to the requirements of a problem.

2) implement data structures in Python.
3) analyze and compare algorithms based on their space and time complexities.
4) solve basic numerical problems using Python.

DS 502, Probability and Statistics for Data Science

The fundamental concepts of probability are essential to solve the data science problem. This course will introduce a broad range of probability theory and mathematical statistics that emphasizes the probabilistic foundations required to understand probability models and statistical. The topics covered include fundamentals of probability theory, statistical inference and graph data structure, including probabilistic models, random variables, useful distributions, expectations, the law of large numbers, the central limit theorem, point and confidence interval estimation, hypothesis tests, basic data structure and algorithms for graphs. After taking this course, students will be able to use calculators and tables to perform simple statistical analyses for small samples and use popular statistics packages (e.g., R Studio) to perform simple and sophisticated analyses for large samples.

CLOs

By the end of the course students will be able to:

1) Demonstrate the ability to apply fundamental concepts in exploratory data analysis, including random variable, probability of an event, additive rules, and conditional probability.
2) Able to understand several well-known distributions, including Binomial, Geometrical, Negative Binomial, Normal, and Exponential Distribution.
3) Able to derive the probability density function of transformations of random variables and use these techniques to generate data from various distributions.
4) Able to analyze hypotheses tests of means, proportions, and variances using both one-and two-sample data sets.
5) Able to understand fundamental properties of graphs and related discrete structures, and be able to model problems in Data Science.

DS 504, Data Mining and Decision Support

Introduction to data mining techniques, including data preprocessing, data mining primitives, association rules, decision trees, cluster analysis, classification and machine learning, data visualization, and data warehousing. Applications from a wide variety of domains will be analyzed.

CLOs

By the end of the course students will be able to:

1) After taking and successfully passing this course, students will:
2) Establish a foundational understanding of basic data mining concepts and techniques
3) Understand a variety of supervised and unsupervised learning algorithms and what types of problems each is appropriate for
4) Be able to assess the efficacy of data mining techniques with quantitative metrics
5) Gain practical experience with small programming projects

CSCI 545, Big Data Analytics

Recent advances in technology have led to orders-of-magnitude increases in the amount of data that is being produced and archived. This phenomenon has been generically referred to as Big Data. Examples of such data include internet traffic and logs, social media, large scale scientific projects, and health records. The process of deriving actionable insights from these large volume and heterogeneous data sets is referred to as Big Data Analytics, and poses new challenges in all aspects of computing. This course will review essential theory as well as statistical and machine learning techniques for data processing and examine their scalability. Students will gain an understanding of current best practices in Big Data processing and complete assignments and projects with a variety of relevant software tools.

CLOs

By the end of the course students will be able to:

1) Understand key tools and techniques of data science
2) Gain practical experience with technologies that enable analysis of large data sets
3) Appreciate the current position of Big Data and analytics in the value chain

DS 507, Database Management Systems

This course is a one-semester course intended for graduate students in Data Science and Computer Science graduate programs. This course focuses on the design, analysis and implementation of database management systems. Most of the topics will be covered as case studies using commercial and open-source database components such that the students can learn a wide-range of tools and techniques and also assess the trade-offs. Student will implement their own database management system as the integrated course project. Therefore, strong programming skills are required. The course will cover topics such as relational, document, key/value data models, storage models and architectures, query languages (e.g. SQL), various types of indexing, transaction management and recovery, query processing, distribute/parallel database architectures, and big data and NoSQL database models and systems.

CLOs

By the end of the course students will be able to:

1) use relational algebra and relational calculus, to express database queries;
2) use query languages for database management systems;
3) implement database tables, using functional dependencies and normal forms;
4) use table heaps and indexes to implement a database;
5) understand the tradeoffs between different concurrency control methods;
6) compare different database systems and choose the suitable one for the given specifications;
7) implement database recovery algorithms and verify their correctness.

SEDS 591, Research Methods

This course introduces students to representative research methodologies such as surveys, interviews, experimentation, and case studies. Students then review the fundamental approaches to research design, data collection, analysis, and presentation.

CLOs

By the end of this course, students will be able to:

1) Understand the need for and role of scientific research;
2) Undertake research activities in a structured and formalized manner;
3) Recognize and observe proper ethical standards used in the conduct of scientific research;
4) Identify potential misconduct or ethical breaches;
5) Identify authoritative sources;
6) Recognize and select relevant methodologies;
7) Understand and implement proper techniques of data acquisition, storage, access, and management;
8) Recognize the role of intellectual property rights (know-how, trade secrets, patents, etc.).
9) Communicate effectively in written and oral formats.

DS 551, Process and Project Management

This course is a one-semester course intended for graduate students in Data Science graduate program. In this course, students will learn the concepts of process and project management, professional practices, software lifecycle, process models, requirements engineering, and design. Students will be taught theoretical concepts, applications, tools, processes, and practical insights to help them understand how to successfully manage projects in a wide variety of industries including software development and information technology. Students will be exposed to study aids, real-time projects, and tools to provide plenty of hands-on practice.

CLOs

By the end of the course the student will be expected to be able:

1) Learn about process and project management, the organizational structure, and key personnel positions associated with planning and managing projects.
2) To explain the software lifecycle and compare major process models and methodologies.
3) To demonstrate effective process and project management practices through participation in a medium-sized team project.
4) Work successfully with a group of your peers on a common problem.
5) To understand the use case model and to produce a software requirements specification.
6) Estimate project size, effort, and schedule for new program proposals or enhancements to existing programs. Staff the project and orienting new employees.
7) Plan and document projects. Manage cost and schedule.
8) Assess and control project risks.
9) Continuously improving processes and the role of the project manager.
10) Evaluate performance.

DS 552, Data-Driven Innovation

With the rapid advance of information technologies, data is emerging as the world's most important resource. Data-driven innovation offers a huge potential for economic, social and technological progress. In this course, students will learn and discuss concepts such as data-driven research and development, transforming traditional goods and services to data-intensive products, optimizing processes, logistics and delivery using data-driven techniques, and personalized marketing using data analytics, social networking and cloud computing. The course will also discuss how organizational evolution and optimization can be facilitated using data-driven techniques as an accelerator. The students will engage in a semester-long team-based project to create a data-driven innovation project proposal for a governmental or private organization.

CLOs

By the end of the course students will be able to:

1) explain the principles of data driven concepts an data management and engineering, and apply these principles when engineering data.
2) analyze larger amounts of data using a variety of data engineering tools, and
3) interpret correctly the results derived from these data-driven tools

SEDS 592, Research Seminar

The course requires students to select a topic, identify authoritative sources, generate an annotated bibliography, and compose a survey paper representing the current state-of-the-art in the selected topic area, and prepare a presentation based on the survey paper. The course includes a series of research presentations, presented by domain experts, to expose students to current research programs, and facilitate the thesis topic selection by students.

CLOs

By the end of this course, students will be able to:

1) Conduct a literature search;
2) Prepare a detailed bibliography;
3) Analyze current research and critically review for strengths and weaknesses;
4) Recognize and extract relevant information from scientific papers, seminars, and presentations;
5) Prepare a survey paper representing the state-of-the art in the topic area;
6) Present the results to both technical and non-technical audiences.

DS 693, Thesis Proposal

Students are led through the processes required to plan and initiate an independent research project that will eventually lead to a thesis. Early on, students are required to finalize their topic and secure the participation of a thesis advisor. They must then develop a research proposal based on a literature analysis of the chosen topic, defend the proposal, and then begin the execution of the project.

CLOs

By the end of this course, students will be able to:

1. Identify relevant authoritative sources, conduct a literature search, and produce a detailed bibliography;

2. Define the scope of a research project to replicate published work and/or extend current knowledge in terms of achievable aims;
3. Develop a methodology to conduct and execute the project, including outlining a timeline with milestones, cognizant of resource availability;
4. Prepare and defend a thesis proposal.

DS 694, Thesis

The student will conduct independent work under the direction of a supervisor on a research project in the student's designated area of inquiry. This work will culminate in a thesis document that the student must successfully defend. The project itself should demonstrate student mastery of the some or all of the stages of the Data Science life cycle, such as data acquisition, storage, preparation / cleaning, analytics & data mining, modeling & machine learning, and visualization.

CLOs

By the end of this course, students will be able to:

1. Conduct independent work on the topic as designated in the Thesis Proposal;
2. Manage task completion in accordance with the established project milestones, and revise the project schedule as needed using critical-path management techniques;
3. Write, using an iterative process, the thesis document in consultation with the thesis advisor;
4. Present the results of the independent work to a technical audience.

Program Elective Courses

- MATH 540 Statistical Learning (Recommended Elective)
- SEDS 504 Innovation and Entrepreneurship (Recommended Elective)
- CSCI 512 Information Theory
- CSCI 515 Modeling and Simulation for Computer Science
- CSCI 547 Algorithmic Trading
- CSCI 581 Acquisition and Analysis of Biomedical Data
- CSCI 594 Deep Learning

Elective Courses Descriptions:

MATH 540, Statistical Learning (Recommended Elective)

The course starts with a review of basic probability and statistics and proceeds to detailed discussion of linear regression, quality measures, diagnostic tests and plots, and remedial methods. It continues with an extension to multiple linear regression, predictor subset selection, interactions, variable transformations, use of categorical predictors. Other topics include introduction to nonlinear regression and neural networks, logistic regression, Poisson regression and generalized review of basic probability and statistics (probability, random variables, special distributions, point and interval estimation, hypothesis testing, one-sample and two-sample t-tests, chi-square tests).

CLOs

By the end of the course students will be able to:

1) Use labeled quantitative and categorical data to draw conclusions about real world phenomena using advanced regression methods,
2) Build regression models and validate their quality,
3) Use modern statistical software packages for building statistical models.

SEDS 504, Innovation and Entrepreneurship (Recommended Elective)

This course, which requires no background in business studies, exposes the students to fundamental ideas regarding innovation and entrepreneurship. Topics covered will include: how to identify business opportunities and to acquire customers, how to develop a business model, understand investments, and manage risks.

CLOs

By the end of this course, students will be able to:

1) Conduct a market assessment and niche analysis;
2) Prepare a business plan which describes the opportunity and presents the innovation concept;
3) Provide a technical characterization of the solution strategy;
4) Design and/or prototype the solution;
5) Prepare a marketing plan for the proposal.

CSCI 512, Information Theory

This course covers the mathematical and technological foundations of communication and data storage. We begin by introducing the idea of entropy as a way to quantify information and proceed to major results in the field such as the source and channel coding theorems. The second part of the course examines topics such as detection and estimation theory (but, in the framework of Information Theory). We also cover other concepts and meanings of *information*.

CLOs

By the end of the course students will be able to:

1) Understand the fundamental limits of communication in the presence of noise
2) Quantify the information requirements and overhead of a given problem or system
3) Understand and be able to apply different types of coding

CSCI 515, Modeling and Simulation for Computer Science

The roots of computer science can be traced to modeling things such as trajectories of artillery shells and cryptographic protocols, both of which pushed the development of early computing systems in the early and mid-1940's. A principal approach to modeling and simulation is abstraction, so that real world systems can be effectively simulated on a machine. In this course, we will examine modeling techniques such as Monte Carlo methods, stochastic processes, queuing theory, and Markov chains. Assessment and evaluation of approaches will also be covered. Important application areas that may be investigated include computer systems and networks, diagnostics, economics and finance, and urban planning.

CLOs

By the end of the course the student will be expected to be able to:

1) To acquire skills in handling situations involving more than one random variable and functions of random variables.
2) To apply basic probability techniques and models to analyze the performance of computer systems, and, in particular, of networks and queues.
3) To have a well – founded knowledge of standard distributions which can describe real life phenomena.
4) To understand and characterize phenomena which evolve with respect to time in a probabilistic manner.
5) To expose the basic characteristic features of a queuing system and acquire skills in analyzing queuing models.
6) To use discrete time Markov chains to model computer systems.
7) To learn how to analyze a network of queues with Poisson external arrivals, exponential service requirements and independent routing

CSCI 547, Algorithmic Trading

Financial practices have substantially benefited from the domains of computer science (CS) and artificial intelligence (AI), to the point that most trading in major financial markets are now governed by algorithms. This course aims to provide the CS students with the necessary background and guidance in order to have them practically apply their CS and AI knowledge in financial domains. The course consists of 4 parts: In the first part of the course, all necessary financial background will be provided. In the second part, tangible mechanics of the domain (like the Financial Information Exchange, FIX, protocol, and order book dynamics) will be covered. In the third part, algorithmic side of the financial transactions will be elaborated on. And in the last part, quantitative strategies (like portfolio construction, high-frequency trading, news handling, etc.) which make heavy use of statistics, AI, data mining, time series analysis, and performance evaluations will be covered.

CLOs

By the end of the course the student will be expected to be able:

1) to become comfortable and fluent with the terminology and practices of the financial applications of computer science and artificial intelligence
2) to design and develop both client and server sides of trading applications using native Financial Information Exchange protocol
3) to understand the strengths and weaknesses of major trading algorithms, and to assess and compare their performances
4) to develop quantitative financial solutions which make use of statistics, artificial intelligence and computer science practices
5) to be ready and well-prepared to participate, work in and/or scientifically contribute to financial organizations and industries

CSCI 581, Acquisition and Analysis of Biomedical Data

Students will learn a number of important topics in the field of computational biomedicine. They will study various topics within the biomedical domain, including primers of anatomy, physiology as well as the basis of various biomedical imaging techniques. Signal processing tools, such as filter theory, artifact rejection, as well as PCA and CSP will be covered. Additionally, we will cover how uni- as well as multi-variate features can be employed for decoding. Practical computing sessions will be carried out and students will also perform a project, where all research-related steps will be covered.

CLOs

By the end of the course the student will be expected to be able:

1) to know some basic anatomical and physiological concepts
2) to understand the basis of various biomedical imaging techniques, such as EEG, NIRS, fMRI, ECG, EMG, among others
3) to understand which different imaging correlates can be measured and analyzed by various imaging modalities
4) to understand and implement a range of data analytical techniques, that are common to biomedical related data analysis in particular, but also for computational approached in biomedicine in general
5) to design and conduct a small biomedical study, i.e. experimental design, implementation, conduction, data analysis, report writing

CSCI 594, Deep Learning

This course is a one-semester course intended for graduate students in Computer Science and Data Science graduate programs. It introduces the students to the concepts and existing models and techniques of deep learning. The main themes of the course are benefits, properties and challenges of Deep Learning; Introduction to Machine Learning and Optimization; Challenges and Common Approaches; Regularization; Convolutional Neural Networks; Deep Recurrent Networks and Sequence Learning; Generative Adversarial Networks; Autoencoders; Advanced Deep Learning Topics; Applications. The students will learn how to design a deep learning architecture. The course presents several common deep learning techniques and exposes the students in learning how to use deep learning in a very efficient manner. Students will gain experience through assignments and projects, with an emphasis on proper deep learning practices with appropriate applications.

CLOs

By the end of the course the student will be expected to be able:

1) to develop solutions using deep learning with Python programming language and environments to solve problems and perform specified tasks, using proper hyper parameters and optimization techniques;
2) to understand and apply concepts related to deep learning, such as logistics regression; gradient descent; regularization; Convolutional Neural Networks; Deep Recurrent Networks; Generative Adversarial Networks; Autoencoders;
3) to critically determine how to select a deep learning approach for a domain and task;
4) to gain knowledge about state-of-the-art solutions using deep learning
5) to understand the strengths and weaknesses of various deep learning approaches;

Master Thesis Guidelines

The guidelines presented here form a manual designed to provide you with a quick reference for planning, preparation, and compilation of your thesis project.

Aims and Objectives

The Master's thesis constitutes a piece of applied research and in this context, your primary goal is to analyze, solve and present your research findings for a problem relevant to your field of study. This process should be based on existing scientific and engineering knowledge and follow the principles of responsible research conduct.

The topic of your thesis should be related to your degree program and should be decided in agreement with your thesis supervisor and approved by the MSc program coordinator.

The primary focus of your research project is usually expressed in terms of **aims** and **objectives**. Your aims should comprise aspirations and/or intentions defined in broad terms which essentially describe what you are hoping to achieve. These aims set out what you targeting to deliver at the end of the project. Objectives, on the other hand, are specific statements that define measurable outcomes and comprise specific goals and steps that must be followed for achieving your aims. Your objectives should be:

Specific; provide precise descriptions of what you are going to do.
Measurable; be able to provide concrete evidence when reaching a goal.
Achievable; avoid setting infeasible goals.
Realistic; plan your steps and goals based on the available resources (time, equipment, skills, etc.)
Timely delivered; create a timetable, know when each stage needs to be completed, allow extra time for unexpected delays

Thesis Content

You should consider the following when conducting research and compiling your thesis manuscript:

- Always include a detailed literature review. The literature review should describe the existing theory and research in your thesis area and provide a context for your work. Reference all sources mentioned in the review and give full citation in thesis's Reference List.
- Explain the methods used in researching and developing your work. It is important to explain what research methods you used to acquire data.
- Discuss with your thesis supervisor the extent and level of detail required; different levels of research depth will obviously require different levels of detail.
- Clearly present your findings and describe what have been discovered. Include tables, graphs, illustrations etc., so that it is easier for the reader to understand your results.
- Always, include a discussion of your findings. Use a discursive and evaluative writing approach and fully present your interpretations and judgements of your results. Contextualize your ideas in relation to other theories and with similar research, particularly in reference to the works mentioned in your literature review.

Stages and Procedures

Students should follow the roadmap described in five stages to accomplish the MSc thesis:

STAGE 1: Identify Thesis Supervisor (Thesis Committee)

Students are expected to choose their MSc thesis supervisor before the end of the second semester of the first year, and spend the summer conducting preliminary work on their thesis project. Constructive supervision is a significant component aiming in the success of your thesis work and requires the vivid interaction between you and your supervisors. Your Supervisor is responsible for ensuring that the Master's thesis meets the goals and requirements set by SEDS.

Furthermore, the thesis committee must have at least three members. One of these can be assigned as the Co-Supervisor of the student. The thesis committee should include an external committee member who is external to the department and can be an academic from another NU department, university or, alternatively, an expert from the industry holding an appropriate academic degree and specializing in your thesis's scientific field.

STAGE 2: Thesis Topic Selection

Supervisors are in position to suggest appropriate MSc thesis's topics. These may stem from research work being conducted at the school/department or may arise from material covered during your coursework. Furthermore, thesis topics may be also related to work carried out in the context of research projects involving industrial partners. The topic of the thesis is decided in discussions between you and your chosen supervisors; however, the final choice is always made by you. In this regard, it is customary for faculty members to announce topics in helping you choose your thesis topic.

STAGE 3: Submission of Thesis Proposal

The MSc proposal should be submitted by the end of the third semester and approved by the thesis supervisor. Your thesis proposal should clearly address the following items:

- Outline of the problem/area of application
    - Explain why you think it is worth investigating
    - Set your ideas into a theoretical/academic context
- Aims and Objectives
    - Describe what you are aiming to achieve
    - Present the steps and approaches you will employ for reaching your goals
- Methodology
    - Explain what methods you intend to use when researching and developing your work
    - Use a descriptive writing approach corresponding to the detail required for the panel's comprehension of your approach.
- Scope and constraints
    - Clearly set your scope and anticipated constraints:
        - Your selected topic may be vast with numerous applications and thus, you might want to limit your work in an area of application
        - You may not be able to conduct some research due to constraints on time, cost, or availability of resources
- Discuss requirements on resources
    - Do you need any special lab equipment?
    - Is literature review possible with library's resources?
    - Are any materials and/or consumables required in your research?
- Propose a draft timetable for your thesis

STAGE 4: Carrying out Research and Thesis Manuscript Preparation

Once your proposal has been officially approved, the actual work may begin. It is crucial that you are always well-prepared in meetings with your supervisor. In this context, it is a good practice to always keep minutes of your meetings and circulate agendas with clearly outlined discussion points and expected results prior to your meetings. This makes it easier for the supervisor to focus on significant issues, leading to a better response for you. If you feel that you may have misunderstood a concept, or you are not certain of the steps required for performing a task, ask your supervisor for clarifications or further guidance. The supervisors should always guide you with advices on the topics and tasks you should put emphasis on and at the same time turn you away of tasks that may waste your time.

Finally, you must keep in mind that writing a thesis takes significant time and effort. You should keep track of your work, make notes, write intermediate reports so that when your work has approached a certain maturity, you will be able to compile a successful thesis's manuscript.

STAGE 5: Thesis Submission and Defense

Before submitting your manuscript, your supervisor will check it thoroughly and give you feedback on corrections and changes that need to be made.

When you have prepared the revised document, you submit to your supervisory committee and external examiner for evaluation. Keep also in mind that you should aim at meticulously following your supervisor's comment and corrections so that a series of multiple revisions can be avoided. When your supervisory committee & the external examiner approve the final document, your Lead supervisor will get permission to submit your final thesis report for evaluation.

Thesis Grading

The MSc. Thesis must be compiled in a report (manuscript) according to the specification provided by the course instructor of the DS 694 Thesis course and defended in front of the MSc. Thesis committee, which comprises of the supervisor and committee members (including the external committee member). The MSc. Thesis manuscript and MSc Defense Oral presentation will be evaluated according to the following assessment criteria.

MSc Thesis Assessment

- ***Presentation of the research problem and thesis's objectives***
  - Is the research problem clearly specified and contextualized?
  - Are the research questions and hypotheses clearly formulated?
  - Does the thesis capture the relevance, rationale, and objectives of the proposed research?
- ***Literature review***
  - Does the thesis include a comprehensive review and critical discussion of the relevant literature and/or technological developments?
  - Is there a description on how the conducted research positions itself within the generic context of works which have been published in the area?
  - Is the relevant background theory covered? Are the presentation, discussion and explanation provided, adequate? Has the theory been contextualized appropriately within the framework of the research problem being investigated?
  - Have the latest theoretical developments in the area been presented and described?
  - Does the student demonstrate a systematic understanding of the relevant background material and knowledge?
- ***Methodology, design and implementation***
  - Are the adopted methodologies and/or design approaches clearly justified and described?

- Is the implementation well explained?
- Is there a clear identification of any limitations, assumptions and constraints which affect the application of the employed methodology, design approach and implementation?

- ***Testing, results, analysis, evaluation concluding remarks & future work***
  - Are the test procedures sound and objective?
  - Do the proposed tests address the research problem being investigated?
  - Are the test conditions, assumptions, constraints, and limitations clearly identified?
  - Are the results clearly presented, analyzed objectively and critically evaluated?
  - Do the concluding remarks summarize the work done? Are there suggestions for any future development and/or enhancements?
- ***Structure and presentation of thesis***
  - Are the thesis contents well structured, focused, and easy to follow?
  - Are the student's contributions and assumptions clearly communicated to the reader?
  - Is it in compliance with the given guidelines?
  - Is it clearly presented and organized? Is the grammar and usage of English of an appropriate level?

Oral Presentation Assessment (MSc Thesis Defense)

PRESENTATION:

- ***Speech & Style***
  - Clear and easily understood. Correct use of terms.
  - Easy-to-understand sequence. Professional appearance. Use of good English.
- ***Structure of the Presentation***
  - Logical sequence, good flow. Supporting body of literature mentioned.
  - Development of topics described clearly. Smooth progression from topic to topic.
  - Key points & challenges sufficiently highlighted.
- ***Layout of Visual Aids***
  - Clear power point slides, uncluttered. Concise & precise slides.
  - Use of good English. Good use of charts, tables, diagrams, etc.
- ***Questions & Answers***
  - Clear understanding of the questions.
  - Concise answer responding to the point of the question.

TECHNICAL CONTENT:

- ***Introduction***
  - Problem statement & project objectives. Coverage of all main points of the project.
  - Literature review and conclusions. Relevance to the need of industry, society etc.
- ***Technical Competency***
  - Viability of the design concept. Justification of the approach
  - Design methodology. Practical Implications.
  - Quality of the concept presentation. Interpretation of the achieved results.
  - Use of relevant tools/equipment/software.
- ***Conclusions, Future Work & Professional ethics***
  - Conclusions: advantages and disadvantages.
  - Level of the project objectives achievement.
  - Future work and possible improvements.
  - Consideration in design and solution. Applicability to real-life situations.
  - Compliance with good practices and standards.