Data Science & Analytics (DSAN)

DSAN 5000 Data Science and Analytics 0 or 3 Credits
Introduction of data science concepts (data collection, cleaning, filtering, pre-processing, modeling, knowledge extraction, actionable recommendations). The data science process and its connections to statistical techniques. Elements of database use and of SQL. Algorithms. Data exploration and elements of visualization. Ethical issues and ways to implement them. Examples applications include fraud detection, social networks, and spam filters, among others. Prerequisites: Intermediate coding experience in Python3, and knowledge of introductory statistics, 3 credits.
Level: Graduate, Undergraduate
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment limited to students in the MS-ANLY or MS-DSAN programs.

DSAN 5100 Prob Modeling/Stat Computing 0 or 3 Credits
Fundamentals of probability theory: sampling, random variables, independence and conditioning, probability distributions, expected value, law of large numbers, Sampling distributions and central limit theorem. Elements of Monte Carlo methods. Descriptive statistics and graphics, tests, regression, ANOVA, nonparametric methods, linear models, logistic regression. Algorithms used in high dimensional distributions, such as inference using maximum likelihood estimation and Bayesian approaches (expectation maximization, Markov Chain Monte Carlo). If time permits: bootstrapping, density estimation. Prerequisites: Introductory statistics, some coding experience (e.g. R), 3 credits.
Level: Graduate, Undergraduate
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to students with a program in Analytics, Analytics, Analytics or Analytics.

Enrollment is limited to Graduate level students.

DSAN 5200 Advanced Data Visualization 3 Credits
Presenting quantitative information in visual form is an essential communication skill for data professionals. This course introduces representation methods and visualization techniques for complex data, drawing on insights from cognitive science and graphic design. Students will obtain an overview the human visual system, learn to use models for data and for images, and acquire good design practices, using the “grammar of graphics”. Students will use common statistical design tools such as Tableau and the R package ggplot2. Prerequisites: Introduction to Data Analytics, coding experience, knowledge of R. Three credits.
Level: Graduate, Undergraduate
Prerequisites: (ANLY 501 or DSAN 5000)
Grading: Main Campus (UGrad, Grad)

DSAN 5300 Statistical Learning for ANLY 3 Credits
Basic concepts: Model accuracy, prediction accuracy, interpretability, supervised and unsupervised learning. Linear regression. Classification, logistic regression, linear discriminant analysis. Resampling methods, cross validation. Model selection, dimension reduction, and other high-dimensional considerations. Support vector machines. Unsupervised methods such as PCA and Clustering. If time permits: Splines, general additive models, tree-based methods. Prerequisites: Probabilistic Modeling and Statistical Computing (ANLY-511) or equivalent. Good knowledge of R or Python.
Level: Graduate, Undergraduate
Grading: Main Campus (UGrad, Grad)

DSAN 5400 Computational Ling Adv Python 3 Credits
This course presents topics in Natural Language Processing (NLP) and Python programming. The goal of this class is to explore techniques in NLP, with a strong emphasis on hands-on instruction that progressively matures basic Python users into expert Python developers. We will examine topics such as text classification, model evaluation, machine translation, and distributed representations. Throughout the semester, students will select and read a book on AI ethics to motivate discussions on the social impact of modern NLP technologies. Applications include authorship identification, retrieval, and textual similarity, to name a few.

About half of the total class time is devoted to addressing an essential but often neglected piece in software development education: moving from typical data science programming workflows (such as writing basic scripts) to developing sophisticated Python projects. In other words, students will learn to design professional-grade software that they and others will be proud to contribute to together. Programming topics are explored in great depth, including Python best practices, object-oriented design, project structuring, and more. This class will give students the skills they need to contribute to the professional software repositories they work with already and even develop their own.
Level: Graduate, Undergraduate
Grading: Main Campus (UGrad, Grad)

DSAN 5450 Data Ethics and Policy 3 Credits
This graduate-level course will train students to navigate the landscape of ethical issues which inevitably arise, across a variety of fields and industries, in each step of the data science process. Students will explore and critically evaluate a range of data-related issues in contemporary society, such as responsible data collection, algorithmic bias, privacy, transparency, accountability, democratic participation in data usage and data-driven decisions, and the ethical implications of emerging technologies like artificial intelligence and machine learning (self-driving cars, ChatGPT, crowd-sourced training data, etc.). Through a combination of theoretical discussions and real-world case studies, students will examine the profound public policy and social issues associated with data ethics. This course will empower students by introducing a set of general ethical frameworks (consequentialism, deontological ethics, and virtue ethics) and discussing their relative strengths and weaknesses in terms of their ability to address modern ethical dilemmas and guide ethical decision-making processes in business, healthcare, government, and academia. These theoretical frameworks will then be discussed in light of more practical regulatory and policy considerations, so that students will have the tools they need to draw conclusions (in the form of their final projects) about best practices for data handling within a particular field or topic of interest to them. The course will thus equip students with a robust ethical "toolbox" for conscientiously gathering, interpreting, and extracting meaning from data while respecting privacy, fairness, transparency, democratic accountability, and other social concerns.
Level: Graduate, Undergraduate
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to Graduate level students.

DSAN 5500 Data Strctrs, Objcts, Algrthms 3 Credits
The Data Structures, Objects, and Algorithms in Python course will look at built-in data structures, such as dictionaries, lists, tuples, sets, strings, and frozen sets. The course will also cover objects and classes in Python, as well as building new structures and objects. The class will cover algorithms including runtime, recurrence, and development. Applications will include data science problems. Prerequisite: A working or intermediate knowledge of Python.
Level: Graduate, Undergraduate
Grading: Main Campus (UGrad, Grad)

DSAN 5550 Data Science Climate Change 3 Credits
Data Science, as a key component of Artificial Intelligence, is helping shape profound changes to society. An equally forceful phenomenon affecting our world is climate change. This course will investigate the myriad ways Data Science can be used to address climate change. This will include aspects of climate change which Data Science is already beginning to tackle, such as mitigating emissions from the five most carbon-intensive societal activities – energy, manufacturing, agriculture / land use, transportation, and buildings / infrastructure. We will also look at Data Science's emerging role in areas such as climate modeling, biodiversity conservation, carbon capture, climate mitigation finance, geoengineering, climate ethics, and reducing the carbon footprint of Data Science itself. We will see how the following Data Science and Machine Learning topics can be used to address climate change: regression, gradient boosting, causal inference, interpretability, optimization, image processing, natural language processing, reinforcement learning, time-series analysis, and several neural network architectures. Using a variety of existing data sources, students will undertake a final project of their choosing to apply a Data Science technique to an aspect of climate change.
Level: Graduate, Undergraduate
Prerequisites: DSAN 5000
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment limited to students in the MS-DSAN program.

DSAN 5600 Applied Time Series for DS 3 Credits
Time series analytics focuses on trends that occur in data over time. The Time Series class will use R to perform time series analysis for a variety of applications, including financial, econometrics, policy, health, engineering, forecasting, etc. Time series data can be used to better understand temporal forces and to generate predicative models. Often, underlying or latent effects can lead to observable trends. Analytics techniques will include model fitting, statistical methods, visualization, and storytelling. This course will include structured programming with the R language, statistical computing, the use of models to make forecasts, data formatting, cleaning and manipulation of data, solving statistical and time series equations, building predictive models, utilizing graphical applications, and applying applicable machine learning methods and models. It is recommended that students know multivariate calculus, linear algebra, and probability and statistics at the undergraduate level.
Level: Graduate, Undergraduate
Prerequisites: DSAN 5000 and DSAN 5100
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to students with a program in Analytics, Analytics, Analytics or Analytics.

DSAN 5650 Causal Methods Comp Social Sci 3 Credits
Causal Inference for Computational Social Science

This course provides students with the opportunity to take the analytical skills, machine learning algorithms, and statistical methods learned throughout their first year in the program and explore how they can be employed towards carrying out rigorous, original research in the behavioral and social sciences. With a particular emphasis on tackling the additional challenges which arise when moving from associational to causal inference, particularly when only observational (as opposed to experimental) data is available, students will become proficient in cutting-edge causal Machine Learning techniques such as propensity score matching, synthetic controls, causal program evaluation, inverse social welfare function estimation from panel data, and Double-Debiased Machine Learning.

In-class examples will cover continuous, discrete-choice, and textual data from a wide swath of social and behavioral sciences: economics, political science, sociology, anthropology, quantitative history, and digital humanities. After gaining experience through in-class labs and homework assignments focused on reproducing key findings from recent journal articles in each of these disciplines, students will spend the final weeks of the course on a final project demonstrating their ability to develop, evaluate, and test the robustness of a causal hypothesis.
Level: Graduate
Prerequisites: DSAN 5000 and DSAN 5100
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment limited to students in the MS-DSAN program.

DSAN 5700 Blockchain Technologies in DS 3 Credits
This course is designed to provide hands-on experience in building public and private blockchains. Students will gain the critical insight, practical knowledge, and technical skills required to design and integrate successful blockchain technologies into a business domain. The course covers the basics of Blockchain technologies including decentralized ledgers, consensus mechanisms, Public and Private key cryptography, smart contracts, etc. The class also examines the public policy and social issues addressed by and arose from adopting blockchain technology in finance, supply chain, and healthcare. The course utilizes Blockchain services (such as Azure) to teach blockchain development on platforms such as Ethereum (or other) and provides the necessary path for learning to build Blockchain networks at scale. Blockchain technologies (distributed ledger, smart contracts, etc.) are a new paradigm in data management and sharing technologies. Gartner Technology hype cycle predicts blockchain to be the industrial ready phase in the next five years. Students will take a deep dive into organizations with integrated blockchain technologies as part of their business strategy. Blockchain's mainstream use by firms like Walmart, Merck, BlackRock will mean that our Data Analytics program graduates will be interacting with the technology often and will require expertise in handling, extracting, converting data to and from the blockchain. This course will provide an early introduction to the graduates, making them Blockchain-ready.
Level: Graduate, Undergraduate
Grading: Main Campus (UGrad, Grad)

DSAN 5800 Advanced NLP 3 Credits
This course provides a formalism for understanding the statistical machine learning methods that have come to dominate natural language processing. Divided into three core modules, the course explores (i) how language understanding is framed as a tractable statistical inference problem, (ii) a formal yet practical treatment of the DNN architectures and learning algorithms used in NLP, and (iii) how these components are leveraged in modern AI systems such as information retrieval, recommender systems, and conversational agents. In exploring these topics, the course exposes students to the foundational math, practical applications, current research directions, and software design that is critical to gaining proficiency as an NLP/ML practitioner. The course culminates in a capstone project, conducted over its final six weeks, in which students apply NLP to an interesting problem of their choosing. In past semesters students have built chatbots, code completion tools, stock trading algorithms, just to name a few. This course assumes a basic understanding of linear algebra, probability theory, first order optimization methods, and proficiency in Python. This is an advanced course. Suggested prerequisites are DSAN 5000, DSAN 5100 and DSAN 5400. However, first-year students with the necessary math, statistics, and deep learning background will be considered.
Level: Graduate, Undergraduate
Prerequisites: ANLY 501 or DSAN 5000
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to students with a program in Analytics, Analytics, Analytics or Analytics.

DSAN 5900 Digital Storytelling 3 Credits
To be successful, a data scientist needs to have many skills beyond coding. This course will teach you how to communicate your findings and data to users, clients, stakeholders to make the biggest impact on your organization and your career. The course consists of two types of activities: lectures and exercises. Lectures provide a theoretical foundation of storytelling; the exercises are designed to help you learn practical skills that work best for different audiences. You will learn what fits your personal style and practice the power of storytelling that can inform and influence decision-makers.

Writing topics will include technical writing, writing for action, and writing for communication in data science with respect to non-technical readers. Storytelling and visualization topics will include methods for presenting results as conclusions, presenting results as actionable items, and creating visual narratives. Interactive visualizations will focus not only on clearly illustrating information within data for use in decision-making, but also for use in discovery and exploration, as well as question generation. Topics in expression, color use, and style will be included. You will learn how to make a point with your charts, and how to make your charts clear to the audience. Further topics in the area of information presentation will include speaking with clarity to a team, group, or large audience.

All parts of the course will contain applications that focus on the utilization of the results of data science and analytics to promote public good, to encourage social concerns and equality, and to support change in areas such as business, public health, and public policy.
Level: Graduate, Undergraduate
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to students with a program in Analytics, Analytics, Analytics or Analytics.

DSAN 5925 Internship 0.25 Credits
An internship provides the student with practical, hands-on work experience. Students must obtain approval from the Program Director to register for this course. To be approved, an internship must be aligned with the DSAN program goals and provide a significant learning experience for the student. At the end of the internship, the student must submit a 10 page paper to the course instructor, detailing why and how the internship was a critical component of their graduate program, as well as the specific skills, methods, and models they were able to apply from their coursework. Internships should be at least 5 hours per week for a minimum of 6 weeks (30 hours total), and may be in-person, remote, or hybrid.

F-1 international students who take this course must request Curricular Practical Training (CPT) authorization via the Office of Global Services (OGS). The program does not require students to complete an internship.
Level: Graduate, Undergraduate
Grading: Main Campus (UGrad, Grad)

DSAN 6000 Big Data and Cloud Computing 0 or 3 Credits
In this course, students will learn the technology, business, science, and social implications of "big data" processing. In recent years there has been an explosion of tools, techniques, and technologies for working with massive data sets. Students will build real word systems, using stand-alone Hadoop/Spark environments running in VirtualBox on personal systems, and scalable clusters on Amazon Web Services. Topics: Big Data terminology, scaling from one computer to thousands, data storage and data privacy, Spark, data formats and data wrangling, text processing and web mining, streaming data, graph processing. Students will be provided Amazon Web accounts with allowances that are sufficient to cover the course work. Prerequisites: ANLY 501 or equivalent, working knowledge of Python and the Unix command line. Students need to own a laptop computer with at least 100GB of free disk space and 8GB of RAM. Computer Science students may register for a section of this course with ANLY department approval.
Level: Graduate
Prerequisites: (ANLY 501 or DSAN 5000)
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to students with a program in Analytics, Analytics, Analytics, Analytics, Computer Science, Computer Science, Computer Science, Computer Science, Mathematics and Statistics, Mathematics and Statistics, Mathematics and Statistics or Mathematics and Statistics.

Enrollment is limited to Graduate level students.

DSAN 6150 BiologicalBiomed Data Science 3 Credits
Biological and Biomedical Data Science

We are bombarded everyday with multiple claims of health risks (doing this will ruin your health), new treatments and cures (just take this for 30 days for a new you), and better lifestyle choices. How are these claims made, evaluated and validated using data science? Data drives our knowledge of biology, disease and effective treatments. This data is diverse, complex, large, and in many respects unique. This data drives our understanding of whether risk factors or treatments causally change our health outcomes, whether our genes or our environment affects our health, and decisions about drugs, protocols and public health that affect all of us everyday. In this class we explore this rich, diverse data landscape and the specialized methods needed to make sense of it, leveraging the instructor's decades-long experience in collaborative epidemiological and biomedical research across academia, government and industry. We will explore designing good experiments to extract causal relationships, and how we might still make valid decisions even in non-ideal settings. We will explore high-dimensional multivariate data and evaluate the validity of finding a "needle in a haystack" biomarker that can be targeted for treatment. We will see how statistical modeling (survival analysis in particular), machine learning, AI, and explainable AI have made an impact in helping us understand this world within. We will see how data-driven decision making using Bayesian analysis works. This journey will take us through real-life applications in bioinformatics (understanding how genes, proteins and other molecular markers affect disease), epidemiology (how do diseases spread and how interventions can prevent it) and clinical research (clinical trials, observational studies, case-control studies).

DSAN 5100 is a prerequisite. However, first-year DSAN students with the necessary statistics background will be considered.
Level: Graduate
Prerequisites: DSAN 5100
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment limited to students in the MS-DSAN program.

Enrollment is limited to Graduate level students.

DSAN 6300 Database Systems and SQL 3 Credits
ANLY 640 / PPOL 740: Relational and Semi-Structured Databases and SQL Programming This course will explore several aspects of modern database management systems, database programming, relational databases, semi-structured databases, and SQL. The course will begin with an introduction to relational models, normal forms and schema design, relational algebra, and SQL Programming. The course will focus on application development using relational databases and will introduce Big Data concepts and discuss Big Data Processing. Both structured and semi-structured data will be considered, such as XML, JSON, and record-style. Query processing methods will be applied and evaluated. Topics will also include recursion in SQL, constraints and triggers, indices and transactions, data storage including column-oriented and distributed storage, noSQL, and different types of databases, such as non-relational, scientific, parallel, and streaming. The course will discuss types of database-system architectures, including cloud-based services. Applications will coincide with data science and analytics, as well as public policy, intelligence generation, and narratives. Tools may also include cloud-based DBMS.
Level: Graduate, Undergraduate
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to Graduate level students.

DSAN 6400 Network Analytics 3 Credits
The design and analysis of networks to represent interactions between and within data is a quickly emerging discipline of significant importance. Data Analytics combines graph theory, optimization, data science, data visualization, community and cluster analysis, and more. Topics in this course will help answer intriguing questions such as, “How can we make sense of large, highly-associated data sets, ranging from social networks to the smart power grid?” or “Which models are more accurate for predicting popularity on Twitter?” or “How can we estimate the spread of a contagion or of information?” The course will begin with a discussion of applications, specifically to data science and analytics. From there, a formal framework for analysis of graphs and trees will be introduced. This will include graph theory and representation, optimization, and graph-based algorithms. Next packages in Python and/or R will be investigated for the purposes of exploring and visualizing data that contain relationships. These packages will then be used to model and analyze complex data sets for the purposes of community detection, path analysis, influencer assessment, logistics analytics, contagion or information spread (such as rumor spreading), web page ranking, and more. Examples of data science applications are provided with real-world data sets including social network data, web-based data, attributed data, flow data, biological data, and more.
Level: Graduate
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to Graduate level students.

DSAN 6500 CompVision GenImage Modeling 3 Credits
Computer Vision & Generative Image Modeling - Computer Vision Analytics & Generative Image Modeling offers a comprehensive introduction to image mining and computer vision. The course covers image acquisition, representation, and processing, including convolution, Fourier transforms, filters, and feature generation. Advanced topics include classification, segmentation, spatial relations, deep-fake detection, object tracking, and image sentiment analysis. Students will explore cutting-edge techniques such as diffusion models, Vision Transformers, advanced GANs (StyleGAN, CycleGAN), Neural Radiance Fields (NeRFs), zero-shot and few-shot learning, image-to-image translation, self-supervised learning, 3D generative models, cross-modal generation, and adversarial robustness. Practical applications include facial recognition, OpenCV in Python, deep fake detection, gesture analysis, and object/scene categorization. This course equips students with the skills to develop sophisticated computer vision and generative image solutions.
Level: Graduate
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to Graduate level students.

DSAN 6550 Adaptive Measurement with AI 3 Credits
This course provides an opportunity for students to engage in learning new algorithms and data science methods in measurement that is applied across all research fields. Unlike traditional one-size-fits-all assessments, how to make an adaptive test, survey, scale, or game for individuals that are tailored by their ability, interests, behavior, health status, and learning requirements is the major theme to be explored in this course. Topics will include but not limited to fundamental psychometric modeling, item response theory, item bank, item information, equating and differential item functioning, cognitive diagnostic modeling, adaptive testing, game-based assessment, sequence mining on process data, generalized models in large-scale assessment, automated item scoring with NLP, personalized assessment design, and generative AI for automated item generation. In addition, this course will invite 1-2 experts from industry to share the fresh ideas and latest products in measurement to the class. The knowledge taught in this course is emergent and arousing incredibly increasing attention in recent years, especially in high demand at organizations or government that manage large-scale assessments (e.g., World Bank, USAID, OECD, AIR, NCES), high-tech learning, education, game companies (e.g., Pearson, ETS, Duolingo, Roblox), e-commercial high tech platforms imbedded with behavioral science (e.g., Amazon, Meta, Google), public health (e.g., NIH), medical recovery services (e.g., hospital, mental health center), just name a few, and extended to general needs in interdisciplinary research and survey designs.
Level: Graduate
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment limited to students in the MS-DSAN program.

Enrollment is limited to Graduate level students.

DSAN 6600 NeuralNets Adv Deep Learning 3 Credits
Neural Networks and Advanced Deep Learning - Neural Networks and Advanced Deep Learning explores both foundational and cutting-edge deep learning techniques. Students start with a short review of core concepts such as feed-forward networks, activation functions, backpropagation, and optimization using TensorFlow and Keras. The course reviews convolutional and recurrent neural networks, auto-encoders, and methods to prevent overfitting. Advanced topics include Transformers and attention mechanisms, Graph Neural Networks, self-supervised and contrastive learning, neural architecture search, adversarial robustness, energy-based models, and neural ODEs. Additionally, the curriculum delves into advanced optimization, neuro-symbolic AI, and bio-inspired deep learning. Various practical applications are covered to equip students with the skills to address complex deep learning challenges across diverse domains.
Level: Graduate
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to students with a program in Analytics, Analytics, Analytics or Analytics.

Enrollment is limited to Graduate level students.

DSAN 6650 Reinforcement Learning 3 Credits
The field of machine learning is typically divided into three fundamental sub-paradigms. These include supervised learning, unsupervised learning, and reinforcement learning (RL). The discipline of reinforcement learning focuses on how intelligent agents learn to perform actions, inside a specified environment, to maximize a cumulative reward function. Over the past several decades, there has been a push to incorporate concepts from the field of deep-learning into the agents used in RL algorithms. This has spawned the field of deep reinforcement learning. To date, the field of deep RL has yielded stunning results in a wide range of technological applications. These include, but are not limited to, self-driving cars, autonomous game play, robotics, trading and finance, and Natural Language Processing. This course will begin with an introduction to the fundamentals of traditional, i.e. non-deep, reinforcement learning. After reviewing fundamental deep learning topics the course will transition to deep RL by incorporating artificial neural networks into the models. The course includes a coding emphasis to showcase applied implementations of RL within the python ecosystem. Topics include multi-armed bandits, contextual bandits, function approximation via neural networks, Markov decision processes, Monte Carlo methods, temporal difference learning, deep Q-learning, actor-critic methods, and policy gradient methods. Students must have intermediate coding experience in Python as well as knowledge of introductory statistics and multivariable calculus.
Level: Graduate
Prerequisites: DSAN 5300
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to Graduate level students.

DSAN 6700 Machine Learning AppDeployment 3 Credits
Machine learning application deployment bridges the gap between training a machine learning model and running it reliably in production. Many data scientists can build a model that functions well in a notebook. But far more interesting in industry is the question that immediately follows: how do I make this available to users, at scale, without it silently breaking over time? This class answers that question concretely, using a single semester-long project as the vehicle. By the end of the course, students will have built, deployed, and monitored a full ML-powered pipeline.

The course is organized around a deliberate progression. The first part of the class establishes the engineering foundations that every subsequent week builds on: modern Python packaging, automated quality gates, probabilistic data structures, system architecture, and containerization. In the second part, we build the running service: a REST API, experiment tracking, distributed caching, and cloud deployment. The final part of the course focuses on making the service production-grade: reliability patterns, security, orchestration, and observability. Each week adds a concrete component to the project.

The culminating activity of the course is the final project, where students will build and deploy their own sophisticated application that draws both from the skills we learn from the running example we study throughout the semester. The strategies discussed here are often components of state-of-the-art production systems. This class will help students build and deploy robust production-grade systems. Incrementally, students will construct production-grade, cloud-deployed AI systems, drawing from cutting-edge tools used in today's AI ecosystems. Students are expected to be able to easily program in Python at an intermediate level and have strong prior exposure to deep learning or machine learning.
Level: Graduate, Undergraduate
Prerequisites: (ANLY 501 or DSAN 5000)
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to Graduate level students.

DSAN 6725 ApplGenerativeAI forDevelopers 3 Credits
Applied Generative AI for AI Developers - This course is designed for AI developers aiming to build cutting-edge Generative AI (GenAI) applications. Focusing on the applied side of AI, students will explore key techniques such as in-context learning (ICL), retrieval-augmented generation (RAG), AI agents, and responsible AI principles. The course covers advanced tools and methods, including embedding models, inference optimizations (e.g., quantization, multi-adapter swapping), fine-tuning of pre-trained models and benchmarking LLMs. Students will gain hands-on experience with open-source tools like LangChain, LlamaIndex, and platforms such as AWS, applying their skills in practical GenAI applications. The course culminates in a capstone project, preparing participants to deploy scalable, optimized AI systems in real-world scenarios. This course bridges the gap between data science knowledge and applied AI development, empowering students to solve industry-level challenges.
Level: Graduate
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment limited to students in the MS-DSAN program.

DSAN 6750 Geographic Information Systems 3 Credits
Geographic Information Systems (GIS) and Applications Geographic Information Systems (GIS) and Applications Geographic Information Systems (GIS) are used as tools for describing, analyzing, managing, and presenting information about the relationships between geographical and spatial locations, sizes, and shapes. This is known as attribute data. GIS uses techniques that can represent social and environmental data as a map, with a significant number of applications including those in engineering, architecture, public health, environmental science, and business. GIS data will be created through a variety of methods including those offered by global positioning system (GPS) technologies. This course will assume knowledge of R and Python.
Level: Graduate
Prerequisites: (ANLY 501 or DSAN 5000)
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment limited to students in the MS-ANLY program.

Enrollment is limited to Graduate level students.

DSAN 6800 Principles of Cybersecurity 3 Credits
This course explores several aspects of modern security systems, risk management, security policies, and covers an overview of digital forensics. The course begins with a definition of what information security is, the need for security policies and controls, then moves into security management, risk management, incident response and planning, discussion of ethical and legal issues surrounding security, and into technical aspects of security to include authentication, authorization, security appliances, cryptography, ending with an overview of digital forensics along with security maintenance and auditing.
Level: Graduate, Undergraduate
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to Graduate level students.

DSAN 6825 Vector Search+LLM Engineering 3 Credits
This course focuses on how modern language model-based applications combine retrieval systems and large language models (LLMs) to build groundbreaking technologies. As LLMs become increasingly central to decision-making pipelines, their effectiveness depends critically on the retrieval and integration of external knowledge. This course teaches the tools and techniques behind these systems, from inverted indices and vector databases to end-to-end retrieval-augmented generation (RAG) pipelines. Students will build scalable retrieval systems using industry-standard tools including Elasticsearch and Qdrant, while learning how dense and sparse neural retrievers, late-interaction architectures, and approximate nearest neighbor search enable fast and effective search over large datasets. Classical information retrieval concepts are introduced first, with the main focus on modern neural and vector-based approaches. The course then turns to RAG and LLM system design. Students will construct end-to-end pipelines that are evaluated and monitored for observability and failure analysis. Students will serve models and explore efficient inference techniques, while considering tradeoffs between model quality, latency, and cost. Applications will involve speech, text, image, and video modalities. By the end of the course, students will be able to design and deploy production-grade RAG systems that can be used over complex, real-world data. The course culminates in a final project in which students build a complete RAG system tailored to their chosen applications. Students are expected to be proficient in Python and have prior exposure to machine learning or deep learning.
Level: Graduate
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment limited to students in the MS-DSAN program.

DSAN 6850 NLP with Large Language Models 3 Credits
In recent times, Large Language Models (LLMs) have earned the attention of the world. OpenAI's infamous generative LLM, ChatGPT, became the fastest-growing consumer application in history in only two months–and the feverish interest around LLMs continues to grow. This course is concerned with applying LLMs to natural language processing (NLP) problems in real-life settings. This is a seminar-based course, so students will spend the majority of time outside of class reading, with in-class time dedicated to presenting and discussing recent research developments in NLP. The course will begin with a review of the transformer architecture that underlies LLMs and describe its prominent role in modern NLP. Then, we will discuss modern issues using transformers, including: the training and scaling of transformer-based models, variations on the classic transformer, transfer learning in low-resource settings, model deployment, distributed systems, and more. Meta-learning, multimodal learning, and societal impact will also be covered. Students will work on applications such as cross-language information retrieval, machine translation, prompt engineering, and select tasks outside of NLP. By the end of the course, students will have mastered transformer-based models and will be poised to use them at the cutting edge of NLP practice today.
Level: Graduate
Prerequisites: DSAN 6600
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment limited to students in the MS-DSAN program.

Enrollment is limited to Graduate level students.

DSAN 7000 Capstone Project 3 Credits
The Capstone Project course is designed to equip students with advanced research methodologies for developing impactful data science projects, alongside the principles of effective scientific writing. Students will engage in real-world, collaborative projects under the joint supervision of internal faculty members and external industry mentors. Through an immerse curriculum, students will learn the principles of research questions development, research methods selection, effective results interpretation, scientific writing, journal paper submission, peer review, authorship, and communicating scientific ideas to academic and non-academic audiences. In addition, this course will provide guidance on paper publishing, including how to identify an appropriate journal, navigating the selection process, editing, and measuring impact. Ideal outcomes include improved self-editing, development of effective strategies for offering and receiving concise editorial recommendations among peers and finalizing a research paper to be submitted to academic publications (e.g., journal, conference, research reports, etc.). Capstone projects are designed to bridge theory and practice and foster meaningful collaboration with academic researchers and industry mentors. These projects are expected to generate substantive contributions to students' professional development and future career trajectories. This course is particularly well-suited for students interested in academic research practices and/or pursuing doctoral studies. While publication cannot be guaranteed, students will fully engage in the scholarly experience of preparing a high-quality research paper for academic review. Prerequisites: DSAN 5000, DSAN 5100, DSAN 5300; Knowledge about data science, introductory statistics, statistical learning, and coding experience in R and/or Python. 3 credits. Offered in the Fall semester.
Level: Graduate
Prerequisites: (ANLY 501 or DSAN 5000) and (ANLY 511 or DSAN 5100) and (ANLY 512 or DSAN 5300)
Grading: Main Campus (UGrad, Grad)
Course registration restrictions:
Enrollment is limited to Graduate level students.

DSAN 7949 Tutorial: Data Analytic 0-3 Credits
Level: Graduate, Undergraduate
Grading: Main Campus (UGrad, Grad)