SUMMARY

Degree and Academic Programs

March 19, 2023

Actions

Attached: Materials prepared for approval by the State Council of Higher Education for Virginia (SCHEV)

College of Science

Approve New MS Degree in Applied Data Science

Board of Visitors
March 19-20, 2023

Summary Information – New Degree Program

Action: New Master of Science (M.S.) in Applied Data Science degree program in the College of Science

Virginia Tech requests approval to establish a new degree program – Master of Science (M.S.) in Data Science. The degree program will be housed in the College of Science. The proposed effective date is Spring, 2024.

The purpose of the proposed degree program is to prepare students to be data scientists who can analyze complex, large-scale data sources to confirm hypotheses, make predictions, and inform decision making based on the interdisciplinary perspectives of applied statistics, computer science, mathematics, computational modeling and data engineering and mining. Graduates will be prepared to relate scientific, business and policy questions, problems or goals to the data needed to answer such questions; to define and acquire data in order to create new knowledge; to develop, compare and evaluate different methods for analyzing data in relationship to the questions or problems being addressed; to use high-performance computing techniques effectively and efficiently; and to effectively communicate the results of analyses to decision makers and other stakeholders.

Table of Contents

Description of the Proposed Program

Program Background

Virginia Polytechnic Institute and State University seeks approval for a new Master of Science (M.S.) in Applied Data Science (CIP Code 30.7001), to be implemented in the Fall 2024 semester. The proposed M.S. degree will be offered by the College of Science.

The purpose of the proposed degree program is to prepare data scientists to analyze complex, large-scale data sources to confirm hypotheses, make predictions, or inform decision making based on the interdisciplinary perspectives of applied statistics, computer science, mathematics, computational modeling and data engineering and mining. The proposed degree will benefit students by providing them the knowledge and skills that are required and in high demand for jobs from a variety of employers.

Scholars have debated the definition and origins of data science.[1] However, there is a consensus that it is an interdisciplinary science that expands data analytics beyond the classical boundaries of theoretical statistics to include algorithmic and computational modeling.[2] Consistent with an interdisciplinary definition of data science, the proposed degree incorporates statistics, computer science and mathematics as well as electives from application domains. The curriculum is designed to provide the analytic, communication and problem-solving skills required in the workforce, while also considering the ethical issues that can arise in the application of data science methods.

Through a combination of coursework and project-based experiential learning, students will learn to relate scientific, business and policy questions, problems or goals to the data needed to answer such questions; to define and acquire the necessary data (including acquiring and integrating large-scale data from multiple sources) in order to create new knowledge; to develop, compare and evaluate different methods for analyzing data in relationship to the questions or problems being addressed; to use high-performance computing techniques effectively and efficiently; and to effectively communicate the results of analyses to decision makers and other stakeholders. In addition, students will learn to work in interdisciplinary teams that include specialists from business, scientific or policy areas; information technology and computer engineering specialists; and other data scientists.

The program will provide training in data analysis methods; machine learning; predictive modeling; computational and mathematical modeling; high performance computing; programming; algorithm development; experimental design; data extraction, transformation, and preparation; visual analytics; teamwork as part of an interdisciplinary team; the ethics and potential biases of modeling and "big data"; and the skills to communicate the results of complex data analysis and modeling.

[1] Donoho, D. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, 26(4), 745-766. https://www.tandfonline.com/doi/full/10.1080/10618600.2017.1384734

[2] Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, 16(3), 199–231. https://projecteuclid.org/journals/statistical-science/volume-16/issue-3/Statistical-Modeling--The-Two-Cultures-with-comments-and-a/10.1214/ss/1009213726.full

The degree is supported by the departments of Statistics and Mathematics, in the College of Science, and the department of Computer Science, in the College of Engineering. Its core curriculum, undergraduate prerequisites and electives include content from all three departments. In addition, in reference to application domains, the degree will offer electives from a variety of other departments including Economics, Geoscience, Agricultural and Applied Economics, Forest Resources and Environmental Conservation, and Industrial and Systems Engineering. The math, statistics and computer science electives will provide students the opportunity to delve deeper into the foundational disciplines of data science and pursue job opportunities with employers who require individuals with the most advanced applied data science skills, or to pursue PhD education in data science. The electives from other departments will provide students the opportunity to pursue career paths with employers who are looking for individuals with a combination of advanced data science skills and domain knowledge.

Institutional Mission

The Virginia Tech Mission Statement states "Inspired by our land-grant identity and guided by our motto, *Ut Prosim* (That I May Serve), Virginia Tech is an inclusive community of knowledge, discovery, and creativity dedicated to improving the quality of life and the human condition within the Commonwealth of Virginia and throughout the world."

The M.S. in Applied Data Science aligns with the university's mission by training students in the quantitatively based techniques of knowledge discovery and will do this with a constant eye to the real problems of science and society. This aligns with Virginia Tech's mission: the discovery and dissemination of new knowledge, through its focus on teaching and learning, research and discovery, and outreach and engagement. Knowledge discovery is intimately connected with the university's commitment to create, convey, and apply knowledge to expand personal growth and opportunity, advance social and community development, foster economic competitiveness, and improve the quality of life and the human condition within the Commonwealth of Virginia and throughout the world.

Admission Criteria

Admission to the proposed M.S. in Applied Data Science degree program will require additional requirements beyond those of the admissions policies of Virginia Tech. Applicants will be required to also meet the following requirements. In order to be admitted to the proposed Applied Data Science degree program, students must:

- Have three recommendation letters from faculty or employers with at least one letter from a faculty member.
- Have a grade equivalent to C or better in undergraduate calculus, introduction to linear algebra, introduction to statistical methods and theory, and introduction to programming in Python or R.

Curriculum

The proposed Master of Science (M.S.) in Applied Data Analytics is a 33-credit hour degree program. The degree program is a non-thesis program only and does not offer a thesis option.

The curriculum will provide a solid educational foundation in the tools and methods used to gather, analyze, and gain knowledge from data. Students will learn to integrate algorithms and methods from computer science, mathematics, and statistics to use an end-to-end, process-based approach to data science projects. Coursework will focus on the advanced understanding and skills needed to import, prepare, and manipulate large data sets for analysis such as data engineering and mining, computer programming and scripting, and algorithms and data analysis techniques used in a multitude of scenarios. Courses focus on topics that help students understand how and when to select and apply the appropriate tools, methods, and systems of analyses based on the needs of a data science project (e.g., algorithms for clustering and classification, algorithms for supervised, unsupervised, and visual learning). The capstone project courses and courses in data science foundations and team-based communication teach students to iteratively evaluate the methodology, tools, and outcomes of the data science process and the ability to translate results in a meaningful, transparent way that supports ethical considerations in real-world contexts.

Students will choose a concentration to gain a focused depth of knowledge, skills, and abilities that are specific to an applied disciplinary area that aligns with their individual academic and career goals. We propose initially to offer concentrations in Modeling and Algorithms, Economics, Geosciences, Agricultural and Applied Economics, Forest Resources and Environmental Conservation, Industrial and Systems Engineering, and an Individual concentration.

Coursework in the Modeling and Algorithms Concentration focuses on the advanced skills in algorithms and modeling techniques needed for data science projects involving high performance, scalable computing and the analysis of “big data”. Students can learn about advanced modeling methods and the optimization and efficiency of algorithms. The Economics Concentration is designed to teach students about using economic data and applying data science methods in the field of economics. Coursework in this concentration focuses on sources of economic data and how to combine that data with data science principles to better understand the economy and economic trends, inform business decisions, and support socioeconomic problem-solving. Coursework in the Geosciences Concentration focuses on understanding the data and data science methods commonly used in the geosciences. Students will learn about applying data science fundamentals in the context of earth science data (e.g., fault zones depth, dependent surface deformation) and how data science can be used in the geosciences to solve problems and make predictions (e.g., write computer programs that use complex data sets to make predictions about earthquakes and potential impact on human life). The Agricultural and Applied Economics Concentration focuses on econometric forecasting and applied economic management. The Forest Resources and Environmental Conservation concentration focuses on forecasting and data science methods relevant to environmental management of forestry and water resources. The Industrial and Systems Engineering Concentration focuses on operations research methods for statistical problems and application of data science theory and tools in manufacturing. Students will learn machine learning problem formulation, theory of linear and nonlinear programming, data generation and analysis from discrete-event simulation, and data science methods and tools for sensing, monitoring, prediction, and variation reduction in manufacturing processes. The Individual Concentration allows students the opportunity to select coursework that supports an academic pathway, skillset, and/or career trajectory that meets their individual interests and aspirations.

Through the capstone courses, students are required to complete a collaborative capstone project. Students are required to work as a team to complete a real-world, data science project supplied by program partners from Virginia Tech research institutes and centers and external business and industry partners secured by the program and college leadership. Students will work with the partners to complete a data science project from inception to deliverable including: 1) understand and define the research question/problem; 2) translate the research question/problem into a data science question; 3) determine and conduct the appropriate tools and techniques to answer the question; and 4) communicate the findings using visual, written, and verbal presentations skills.

Program Requirements

Five new courses were developed for the core of the proposed program. All new courses are denoted with an asterisk (*).

Core Courses: 18 credit hours

*ADS 5064: Foundations of Data Science (3 credits)
*ADS 5224: Communication in Team-Based Data Science (3 credits)
CS 5045: Computation for the Data Sciences (3 credits)
*CS 5054: Programming Models for Big Data (3 credits)
STAT/CS 5525: Data Analytics (3 credits)
STAT/CS 5526: Data Analytics (3 credits)

Other Required Courses: 3 credit hours

*ADS 5804: Capstone Experience I: Definition and Data (1 credit)
*ADS 5814: Capstone Experience II: Implementation (2 credits)

Concentration Areas: 12 credit hours

Students must choose a designated area of concentration or an individually developed concentration. Students will work with faculty advisors to select a designated concentration and appropriate courses within each concentration to meet their individual professional goals.

Agricultural and Applied Economics Concentration

The purpose of the Agricultural and Applied Economics concentration is to allow students to build proficiency in modern econometric, forecasting, and data science tools needed to succeed in modern applied economic management and related fields. The emphasis is on foundational methods spanning economic analysis, data analysis, predictive analytics, data visualization, and relevant business and agribusiness applications.

AAEC 5014: Applied Economic Analytics (3 credits)
AAEC/STAT 5484: Applied Economic Forecasting (3 credits)
AAEC 5804G: Fundamentals of Econometrics (3 credits)
AAEC 5824: Advanced Applied Economic Analytics (3 credits)

Economics Concentration

The purpose of the Economics Concentration is to teach students how to apply data science methods in the field of economics. Coursework focuses on different types of economic data, how that data relates to economic activities, and how to use data science methods to better understand the economy and economic trends, inform business decisions, and support socioeconomic problem-solving.

ECON 4084: Industry Structure (3 credits)
*ECON 4514: Applied Analysis of Banking and Financial Markets (3 credits)
ECON 5134: Managerial Economics* (3 credits)
*ECON 5314G: Advanced Big Data Economics (3 credits)

Forest Resources and Environmental Conservation Concentration

The purpose of the Forest Resources and Environmental Conservation concentration is to allow students to build proficiency in modern environmental management, forecasting, and data science tools needed to succeed in modern natural resources and related fields. The emphasis is on foundational methods spanning geospatial data analysis, environmental time-series data analysis, predictive modeling of forest and other ecosystems, and data visualization.

BIOL/FREC 5034: Ecosystem Dynamics (4 credits)
FREC 5114G: Advanced Information Technology for Natural Resource Management (3 credits)
FREC/GEOG 5154: Hyperspectral Remote Sensing (3 credits)
FREC 5224: Forest Biometry (3 credits)
FREC 5244G: Advanced Hydroinformatics (3 credits)
FREC 5254: Remote Sensing of Natural Resources (3 credits)
FREC/AAEC/GEOG 5544: Remote Sensing in the Social Sciences (3 credits)
FREC/GEOG 6214: Forestry Lidar Applications (3 credits)

Geosciences Concentration

The Geosciences Concentration provides students with the knowledge and skills to mine and analyze Earth's observational data such as atmospheric and climate data, Earth positioning and distancing data, oceans and coastal data, and natural resource data. Students in this concentration can learn more about different types of geoscience data and modeling techniques to investigate and create knowledge from that data.

GEOS/MATH 5144: Inverse Theory and Geoscience Applications (3 credits)
*GEOS 5184: Advanced Geodesy in the Earth Sciences (3 credits)
GEOS 5314: Advanced Coastal Hazards (3 credits)
GEOS 5814: Numerical Modeling of Groundwater (3 credits)
GEOS 6104: Advanced Topics in Geosciences (3 credits)

Industrial and Systems Engineering Concentration

The purpose of the Industrial and Systems Engineering concentration is for students to develop enhanced knowledge of operations research methods used in data science and in-depth understanding and problem-solving skills in manufacturing data science problems. Students in this concentration will be provided enhanced knowledge and skills in optimization, discrete event simulation, statistics for industrial engineers, and manufacturing and Industrial Internet data generation, problem formulation, and analytics.

ISE 5034: Mathematical Probability and Statistics for Industrial Engineers (3 credits)
ISE 5104: Operations Research (3 credits)
ISE 5204: Manufacturing Systems Engineering (3 credits)
ISE 5405: Optimization: Linear and Nonlinear Programming (3 credits)
ISE 5424: Simulation I (3 credits)

<u>Individual Concentration</u>
The purpose of the Individual Concentration is to offer students the opportunity to select coursework that supports an academic pathway, skillset, and/or career trajectory that meets the need of the individual student. For example, a student may work at a health insurance company and want to understand what data is available, how to use that data, and what data science methods would best support business decisions related to social media marketing campaigns. In this context, the student and faculty advisor could choose courses to meet this need and align with this career path.

<u>Modeling and Algorithms Concentration</u>
In collaboration with a faculty advisor, students select courses from a restricted list of courses from the computer science, computational modeling and data analysis, mathematics, and statistics areas.

The purpose of the Modeling and Algorithms concentration is for students to develop a more in-depth understanding of the algorithms and modeling techniques used in advanced data science. Students with an interest in obtaining an enhanced understanding and skillset in areas such as the interpretability of models, quantification of uncertainty, optimization and efficiency of algorithms, high performance and scalable computing can take courses offered in this concentration.

Computer Science Courses
CS 5644: Machine Learning with Big Data (3 credits)
CS 5664: Social Media Analytics (3 credits)
CS 5764: Information Visualization (3 credits)
CS 5834: Introduction to Urban Computing (3 credits)

Computational Modeling and Data Analysis
CMDA 4634: Scalable Computing for Computational Modeling and Data Analytics (3 credits)

Mathematics Courses
MATH 5424: Numerical Linear Algebra (3 credits)
MATH 5544: Mathematical Optimization for Machine Learning (3 credits)
MATH 5564: Model Reduction: System-Theoretic Methods (3 credits)

Statistics Courses
STAT 5054: Introduction to Statistical Computing (3 credits)

STAT 5154: Statistical Computing for Data Analytics (3 credits)
*STAT 5234: Experimental Design for Data Science (3 credits)
STAT 6554: Advanced Statistical Computing (3 credits)

See Appendix A for a sample plan of study.
See Appendix B for course descriptions.

Faculty Resources

The six core courses, and the required capstone experience sequence in the proposed degree will be taught by faculty from the Computer Science, Mathematics, and Statistics Departments. Collectively, the three departments have a total of 192 full-time faculty: Mathematics 81, Statistics 37, and Computer Science 74. There are a group of 30 faculty members from these departments available to teach the courses in the proposed program, six faculty members from Computer Science, nine from Mathematics, and fifteen from Statistics. Of these, 11 faculty are available for teaching the core requirements of the proposed degree: 4 from Computer Science, 7 from Statistics and 4 from Mathematics.

All these faculty members have doctoral degrees or the equivalent in their respective fields of Computer Science, Mathematics, Statistics, or closely related areas. Twenty-one of the 30 faculty are of full or associate professor rank. Collectively they have published dozens of scholarly articles in peer-reviewed professional journals and served as manuscript reviewers for scholarly journals. For many their work is recognized through external funding agencies, and they serve as leaders of national organizations. These faculty members have been involved in the development of the proposed curriculum.

The electives that comprise the other concentrations will be taught by 42 full-time faculty from the respective departments: 7 faculty members from Agricultural and Applied Economics will teach in courses in the Agricultural and Applied Economics concentration; 6 faculty members from Economics will teach in courses in the Economics concentration; 5 faculty members from Forest Resources and Environmental Conservation will teach in courses in the Forest Resources and Environmental Conservation concentration; 6 faculty members from Geosciences will teach in courses in the Geosciences concentration; 7 faculty members from Industrial and Systems Engineering will teach in courses in the Industrial and Systems Engineering concentration and; 11 faculty members who teach core and required courses will also teach courses in the Modeling and Algorithms concentration (4 faculty members from Computer Science, 7 faculty members from Statistics, and 4 faculty members from Mathematics). These faculty members have been involved in the selection and development of the courses comprising their respective concentrations. Collectively they have published dozens of scholarly articles in peer-reviewed professional journals and have served as manuscript reviewers for scholarly journals.

See Appendix C for faculty curriculum vitae (abbreviated).

Student Learning Assessment

The learning outcomes for the Master of Science in Applied Data Science degree program are specific to graduate-level knowledge, skills, and abilities that students will acquire in the proposed degree program, such as analyzing large-scale data sources and validating predictive models. Student learning will be assessed throughout the program through a variety of formative and summative measures. Assessment measures will include, but are not limited to, tests, written reports, oral presentations, case studies, and individual and team-based project work. The final work products, analyses, reports, and presentations from the capstone experiences will be used to assess students' ability to effectively explain complex data science results and concepts to a broad audience.

Faculty members in the College will review assessment feedback annually in August. The feedback will be used to make appropriate changes to specific course instruction for the coming academic year to enhance learning and strengthen outcomes.

Learning Outcomes

Students will be able to:

- Plan and oversee the lifecycle of a data science project from identification of problem, through collection and assessment of data and data analysis to communication of actionable insight to stakeholders.
- Organize, prepare, and manage large data sets for analysis, including structured and unstructured data.
- Formulate data analysis strategies utilizing sophisticated analytical methods, contemporary programming languages, and high-performance computing techniques.
- Evaluate statistical, machine learning, AI and other data analytical methods to ensure valid and accurate conclusions and recommendations.
- Prepare written and oral reports, including data visualizations, to communicate the results and limitations of data analyses, including potential bias and ethical considerations, and recommended actions to stakeholders.

Each concentration area has additional learning outcomes or competencies that students who select the concentration area should acquire.

Agricultural and Applied Economics Concentration

Students in this concentration will be able to:

- Develop and validate econometric models for analyzing problems in agricultural and applied economics.
- Program and implement time-series models for analyzing economic data.
- Adapt, validate, and implement appropriate data science methods for forecasting economic trends with emphasis on agricultural economics.

Economics Concentration

Students in this concentration will be able to:

- Apply economics concepts, theories, and reasoning to formulate relevant research questions in real-world contexts.
- Construct applicable macroeconomic or financial forecast models needed to address a research question.
- Adapt, scale, and implement appropriate data science methods to leverage high-dimensional economic datasets and extrapolate economic insight to the relevant application.

Forest Resources and Environmental Conservation Concentration
Students in this concentration will learn be able to:

- Apply domain knowledge and data-driven approaches to synthesize diverse observations of natural and managed ecosystems to address societal relevant environmental issues.
- Develop and validate models for analyzing geospatial data, including classification and change detection.
- Program and implement time-series models for analyzing and predicting environmental dynamics.

Geosciences Concentration
Students in this concentration will be able to:

- Apply Geosciences domain knowledge and data-driven approaches to formulate and explain hypotheses about modern Earth science problems that aren't fully understood, and that include highly uncertain natural phenomena.
- Derive physics-driven models of the climatic, isostatic, and tectonic processes related to phenomena such as sea-level change and explain the associated characteristics of such models such as the uncertainty associated with the governing equations and their parameters.
- Use appropriate mathematical inverse methods to identify model parameters from geodetic data sets and comment on challenges such as non-uniqueness of the inverse problem.

Industrial and Systems Engineering Concentration
Students in this concentration will be able to:

- Perform probability and statistical modeling and analysis in Industrial Engineering problems.
- Apply data science methods for manufacturing functional activities such as manufacturing engineering and quality control.
- Apply theoretical concepts and numerical algorithms to solve linear and nonlinear programming problems in data science.
- Use stochastic simulation methodology to design and analyze simulation experiments.

Modeling and Algorithms Concentration
Students in this concentration will be able to:

- Compare the accuracy, stability, and performance trade-offs of different optimization methods available to find optimal solutions.
- Derive and adapt reduced models that approximate high-dimensional physics-based models by using data collected as the computation proceeds.
- Apply optimization strategies to improve the performance of scalable applications and use computational science strategies to exploit massively parallel processors

Curriculum map for M.S. in Applied Data Science:

Learning Outcomes	Core and Required Courses	Assessment Measures
Plan and oversee the lifecycle of a data science	ADS 5064: Foundations of Data Science	Written assignments, tests, class assignments, case

project from identification of problem, through collection and assessment of data and data analysis to communication of actionable insight to stakeholders.		studies (for example, studying use of public health data related to COVID-19 virus), presentations.
Organize, prepare, and manage large data sets for analysis, including structured and unstructured data.	ADS 5064: Foundations of Data Science CS 5045: Computation for the Data Sciences	Projects (for example, cleaning and preparing a raw dataset for analysis or), tests, class assignments (for example, creating visualizations of raw data to identify problems/trends/correlations), case studies, presentations, collaborative written/oral assignments.
Formulate data analysis strategies utilizing sophisticated analytical methods, contemporary programming languages, and high-performance computing techniques.	STAT/CS 5525: Data Analytics STAT/CS 5526: Data Analytics CS 5045: Computation for the Data Sciences CS 5054: Programming Models for Big Data	Projects, tests, class assignments, case studies (for example, creating models for various datasets and modalities), presentations, collaborative written/oral assignments.
Evaluate statistical, machine learning, AI, and other data analytical methods to ensure valid and accurate conclusions and recommendations.	STAT/CS 5525: Data Analytics STAT/CS 5526: Data Analytics CS 5054: Programming Models for Big Data	Projects (for example, identifying and track improvements to programming code with progressive optimizations using big data software technologies), tests, class assignments, case studies, presentations, collaborative written/oral assignments.
Prepare written and oral reports, including data visualizations, to communicate the results and limitations of data analyses, including potential bias and ethical considerations, and recommended actions to stakeholders.	ADS 5064: Foundations of Data Science ADS 5224: Communication in Team-Based Data Science ADS 5804: Capstone Experience I: Definition and Data	Class assignments, reports (for example, preparing written reports on result of analyzing problem datasets), creating models and performing validation of predictive power, creating visualizations of data

	ADS 5814: Capstone Experience II: Implementation	

Employment Skills

The proposed M.S. in Applied Data Science degree program will offer graduates knowledge skills and abilities for employment opportunities in the fields of data science, data analytics, applied statistics, machine learning and "big data".

Graduates of the proposed M.S in Applied Data Science degree program will be able to:

- Identify and collect the data that are needed to answer specific questions and problems
- Identify and lead design and execution of experiments or surveys
- Ensure quality and accuracy of data
- Identify and lead development of novel data assets
- Identify and lead development of statistical analyses
- Identify and lead development and deployment of statistical/machine learning models to analyze data
- Write, test and document computer programs for data analysis using various languages and libraries
- Organize, manage, clean and manipulate very large data sets using data science software
- Analyze data to identify trends or relationships among variables
- Analyze data to inform operational decisions or activities
- Test, validate and reformulate models to ensure accurate predictions of outcomes of interest
- Prepare data analysis reports including data visualizations
- Explain the limitations and uncertainties associated with predictive and machine learning models and other types of data analysis
- Document programming scripts, processes and deliverables
- Interpret and communicate results of data analyses to decision makers, researchers and executives
- Recommend data-driven solutions to key stakeholders

Graduates in each concentration will demonstrate the following additional workplace competencies.

Agricultural and Applied Economics Concentration

- Develop useful new computational and analytical tools by combining agricultural economics knowledge with computing platforms such as "cloud computing".
- Identify, acquire and integrate data from multiple sources, including official government data sources and other public data sources.
- Prepare written reports and oral presentations to communicate the results and limitations of complex models to business or policy decision makers.

Economics Concentration

- Identify official and unofficial data sources and utilize characteristics of big economic data.

- Develop useful new computational tools by combining economic knowledge with a computing platform such as Python.
- Identify formal assumptions made by data analytics tools and reconcile these assumptions with big data econometrics applications.

Forest Resources and Environmental Conservation Concentration

- Develop useful new computational and analytical tools by combining forest resources and environmental conservation knowledge with widely used software languages (e.g. R and Python) and on diverse computation platforms such as "cloud computing".
- Identify, acquire, and integrate data from multiple sources, including remote sensing, field observations, and environmental sensors.
- Prepare written reports and oral presentations to communicate the results and limitations of complex models to natural resource managers.

Geosciences Concentration

- Explain and identify techniques of geodetic methods of measurement including Global Navigation Satellite Systems (GNSS), Interferometric Synthetic Aperture Radar (InSAR), and Terrestrial Laser Scanning (TLS).
- Design computer programs to analyze geodetic data and to perform reference frame conversions.
- Evaluate the approaches and challenges of communicating coastal hazards and related topics in non-expert language for coastal policymakers and the public.

Industrial and Systems Engineering Concentration

- Identify and implement advanced optimization and simulation methods in data science problems.
- Formulate and implement data science tools to improve operational excellence and advance domain knowledge discovery in manufacturing and the Industrial Internet.
- Differentiate advantages and limitations of different solutions, convey interpretable and actionable insights and results, and inform risks in Industrial Engineering problems with effective written and oral formats.

Modeling and Algorithms Concentration

- Solve problems that require an optimal algorithmic solution that can be used to make a data-driven decision that provides a trusted prescriptive course of action that businesses can confidently operationalize.
- Explain the limitations in the applicability of different approaches used to derive reduced dimensional models such as the parameter range, and the associated computational challenges.
- Identify appropriate algorithms and data structures for scalable computing environments such as servers in "the cloud" or at high-performance computing facilities.

Relation to Existing Programs

Virginia Tech offers four degree programs that are related to the proposed degree program: an M.S. in Statistics, an M.S. in Mathematics and an M.A. in Data Analysis and Applied Statistics offered by the College of Science; and an M.S./M.Eng. in Computer Science and Applications offered by the College of Engineering.

All four programs and the proposed program share related interests in data analytics or data science. Additionally, both degree programs in the College of Engineering and the proposed program share interests in computer programming. However, there are distinct differences in the purpose and focus of the programs. There are also distinct differences in the knowledge and skills of the graduates. The proposed degree is not an expansion of any certificate or program. It has a different focus, scope, and intent from other degree programs and will attract students specifically interested in applying data science in industry. No degree programs will be negatively impacted because of the initiation and operation of the proposed degree program.

Master of Science (M.S.) in Mathematics
The M.S. in Mathematics degree program focuses on the theoretical foundations of mathematics and its applications to scientific and engineering problems. The program is intended to provide a basic foundation of graduate level work from which students can pursue doctoral level studies in mathematics or a technically oriented job in the private sector or government service. Depending on their program of study, students in this program acquire the skills to apply mathematical theories and techniques to develop new mathematical methods or to apply mathematical techniques to the solution of engineering or scientific problems. Students take courses from one of several sets of course sequences in the areas of abstract algebra, calculus of variations, differential equations, and computational math. This is supplemented by courses in specific topics such as mathematics education, mathematical biology, optimization, and control and systems theory. This degree and the proposed degree do not share any core courses. The M.S. in Applied Data Science is an interdisciplinary degree program that prepares students to use a combination of mathematical, statistical, and computational techniques to acquire, prepare and interpret data to extract insights from it, make predictions and support informed decision making.

Master of Science (M.S.) in Statistics
The M.S. in Statistics degree program focuses on the theoretical foundations of statistics and the application of statistical methods to the analysis of experimental or engineering data. The program is intended to provide a basic foundation of graduate level work from which students can pursue doctoral level studies in statistics or a technically oriented job in the private sector or government service. Depending on their program of study, students in this program acquire the skills to apply statistical theories and techniques to develop new statistical methods or to apply statistical techniques to the analysis of experimental or engineering data. The core curriculum of the M.S. in Statistics degree focuses on fundamentals of inference, regression and analysis of variance, probability and distribution theory, linear models theory, and experimental design. There are two core courses in the proposed program, STAT/CS 5525 and STAT/CS 5526, that are electives in the M.S. in Statistics degree program. The M.S. in Applied Data Science focuses on the combination of mathematical, statistical, and computational techniques needed to extract insights from and make decisions based on data.

Master of Arts (M.A.) in Data Analysis and Applied Statistics
The M.A. in Data Analysis and Applied Statistics focuses on the application of statistical methods. It is intended to provide a basic foundation of graduate level work in the applications of

statistical methods either for students who are pursuing a PhD in another field who need to be able to analyze the data associated with their program of study, or for individuals who are employed full-time and are currently involved in statistical data analysis on behalf of their employer, or who are seeking to take advantage of the career opportunities for individuals with foundational applied statistics skills. The degree comprises 30 credits. The core curriculum of 18 credits emphasizes applied statistics, regression methods, design of experiments, communications, and basic theoretical statistics. Depending on the student's program of study, the remaining 12 credits are electives that may be chosen from a variety of departments, including two courses, STAT/CS 5525 and STAT/CS 5526, that are part of the core curriculum of the proposed degree. Unlike the proposed degree, there are no specific mathematics, statistics, or computer science prerequisites for this degree, nor does this degree include any computer science courses in its core curriculum. The proposed degree has an emphasis on statistical and mathematical modeling with large datasets, high-performance computing, programming, and properties of algorithms and models that this degree does not provide. Also, the proposed degree offers electives from statistics and mathematics that will typically be unavailable to students of this degree for lack of the required mathematical prerequisites.

<u>Master of Science/Master of Engineering (M.S./M.Eng.) in Computer Science and Applications</u>
The M.S./M.Eng. degree program in Computer Science and Application is designed to teach students about the advanced theory and application of computer hardware and software. Students will learn how to write programs, build systems, analyze algorithms, and evaluate the performance of computer hardware and software. The M.S. program has a research focus and the M.Eng. program has an applied focus. Students take courses in algorithms and theory, computer systems, programming languages, numerical and scientific computing, computer architecture and networking, data and information, software engineering, human-computer interaction, intelligent systems, and computational biology and bioinformatics. Students develop skills in modifying tool chains, building new tool chains, and evaluating newly proposed tool chains. Students will learn about research methods and ethical frameworks and how to apply those to computer science work. Graduates of the M.S. degree will be prepared to conduct research in computer science as well as apply for doctoral programs in computer science and related fields. Graduates of the M.Eng. degree will be prepared to work in fields such as software development, artificial intelligence, or computer security designing, implementing, improving, and monitoring computer applications and software systems. The proposed program has a required two-course sequence in its core, STAT/CS 5525 and STAT/CS 5526, that also serve as electives in M.S./M.Eng. in Computer Science and Application degree program. Information Visualization and Introduction to Urban Computing (each 3 credits) are electives in both programs. The M.S. in Applied Data Science is an interdisciplinary degree designed to teach students how to formulate data analysis strategies and perform exploratory data analysis using pre-built computational tools and systems.

<u>Justification for the Proposed Program</u>

Response to Current Needs (Specific Demand)

Businesses and Government agencies are increasingly enabled by data as a strategic asset and the ability to analyze it and extract information to inform decision making and spur innovation. Organizations big and small are seeking individuals with the skills required to "mine" their data to make "data-driven" decisions, to innovate, to better meet customer needs, and to increase profits. The current demand for individuals with such skills in Virginia and nationally include: 1) industry demand for professionals with knowledge and skills in advanced mathematical and statistical computing to handle large volumes of data; 2) industry demand for professionals with knowledge and skills in advanced computational analysis to handle various forms of data including images, video, and natural language; and 3) industry demand for professionals with knowledge and skills to work with business leaders to develop data science projects to innovate and make data driven decisions.

There are several technical factors fueling the demand for data science skills:

- The *volume* of data that is accumulating and the computational power available to analyze it.
- The speed, or *velocity*, with which the data continue to accumulate.
- The *variety* of data formats including images, video, and natural language in addition to numeric data.
- The need for companies to remain competitive by being able to make data-driven decisions and rapidly innovate in response to economic trends and consumer preferences.

Volume and Velocity

Every interaction on a phone, computer, or smart device uses and creates data. When someone posts, likes, or shares a picture on social media, data is collected. When an article is read online or a song, TV show, or movie is streamed, data is collected. When something is purchased online, data is collected. Information about where and when the product or service was purchased, its cost, and information about the individual who purchased the product can be collected and stored. As a result of interactions such as these, the total amount of data created, captured, copied, and consumed worldwide in 2020 is estimated to be 64.2 billion terabytes.[3] Illustrating the speed with which data is accumulating, the volume is expected to reach 180 zettabytes by 2025,[4] almost tripling the amount of data created and consumed. To put this into perspective, if that much data were stored on DVD's and placed in a stack, the stack would circle the earth 222 times. Described another way, at current average internet download speeds it would take 1.8 billion years to download that amount of data.[5]

This volume of information, in the form of numbers, natural language, pictures, and videos, can be used to predict users' behaviors and improve their experiences. For example, Amazon collects data to assess purchase history to 'learn' and predict future purchases. By tracking how often a user ordered laundry detergent last month, it can send a notification that it may be time to purchase more or even set up an automatic delivery so the user no longer has to 'think' about it.[6]

Extracting meaningful and relevant information from large, rapidly accumulating stores of data requires more than basic data management and statistical analysis skills. It requires individuals with the knowledge and skills to create sophisticated analytical models utilizing artificial

[3] Statista Research Department. (2022, May 23). *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025.* Statista. https://www.statista.com/statistics/871513/worldwide-data-created/.

[4] Ibid.

[5] S Marr, B. (2021). *How Much Data is There in the World?* Bernard Marr & Co. https://bernardmarr.com/how-much-data-is-there-in-the-world/. Accessed 29 Jun. 2022.

[6] Matt, B. (2021, July 23). Amazon: Using Big Data to Understand Customers. Bernard Marr & Co. https://bernardmarr.com/amazon-using-big-data-to-understand-customers/. Accessed 29 Jun. 2022.

intelligence and machine learning to gather and build new data sets and extract insights from them.[7] The proposed degree will address this need by training graduates who will help companies transform from reactive environments with static and aged data, to automated ones that continuously learn in real time.[8] Students will learn how to manage and process large data sets for analysis, including structured and unstructured data; formulate data analysis strategies utilizing sophisticated analytical methods, contemporary programming languages, and high-performance computing techniques; create statistical models; and use machine learning and artificial intelligence to discover insights from data in ways that will help their client or organization make decisions.

<u>Variety of Data Formats</u>
Data can be structured or unstructured. Structured data typically refers to data in a tabular format consisting of columns of numerical or categorical data (for example, gender or blood type) stored in an Excel spreadsheet or a data warehouse. Structured data consists of about 20% of the current data that exists today. Unstructured data, on the other hand, has no predetermined structure. It encompasses a variety of formats including video, images, audio, and natural spoken language. Approximately 80% of current data is unstructured data.

To analyze unstructured data requires advanced knowledge and skills in programming languages to preprocess data before it is ready to be analyzed. Among other tasks preprocessing includes handling missing values, removing redundant or irrelevant data, identifying key features or variables, partitioning data, and joining datasets. Analyzing the data requires advanced skills in algorithm development, machine learning, predictive analytics, visualization, and model validation. Additionally, individuals must have a strong understanding of the domain in which they are working (e.g., geosciences, economics) in order to identify 1) what information will be useful to answer specific questions, and 2) how to collate the data from its various formats and integrate it in a way to extract usable information from it. For example, predictive analytics can be used to evaluate historical weather patterns along with other climate change variables including temperature, humidity, surface pressure, erosion patters, and greenhouse gas emissions to predict how often an area will experience heavy rainfall and the impact of those events on things like coastal erosion.

The proposed MS degree program will respond to the need for these skills by training students who can identify and develop methodologically sound and reproducible approaches for analyzing data sets that are large and messy, create models and software that predict what is going to happen or prescribe what should happen, and create the data sets and analytical tools necessary to solve industry problems and spur innovation.

---

[7] Business Higher Education Forum. (2017, April). *Investing in America's data science and analytics talent. The case for action.* PwC. Accessed Jun. 29, 2022.
[8] Kowalczyk, A. (2018, March 22). Why do we need more data scientists and why should you become one? Deepsense.ai. https://deepsense.ai/why-do-we-need-more-data-scientists-and-why-should-you-become-one/. Accessed Jun. 29, 2022.

Data-Driven Innovation

Successful companies are using data science techniques to innovate, create new products and enhance user experiences.[9] One real-world example is evidenced by a data scientist at LinkedIn[10] who improved their members' ability to network with other members. This innovation generated millions of new page views on LinkedIn's site and shifted their growth trajectory significantly upward.[11]

LinkedIn uses biographical data on its members (hometown, education, career path, etc.) to suggest possible connections among members based on such common interests. This was designed and developed by a data scientist who first identified that members were not seeking out connections with other people on the site at the rate executives had expected to fuel LinkedIn's growth. He found patterns in the data that allowed him to predict which members a given member would be interested in connecting with thereby building their network. He tested what would happen if you presented members with names of people they hadn't yet connected with but would likely be interested in meeting. He did this by creating a custom pop-up ad that displayed the three best new matches for each member based on their LinkedIn profile. These "People You May Know" ads achieved a click-through rate 30% higher than other prompts to visit more pages on the site and spurred LinkedIn's growth. This innovative thinking required a data scientist, someone with strong knowledge and skills in statistics and computer science who also had a strong understanding of LinkedIn's mission and vision.

This example illustrates that businesses and organizations need individuals with strong data science skills that also have content knowledge within the field in which they are working.[12] Similarly, individuals analyzing geoscience to understand the effects of climate change on coastal areas are more effective if they understand the drivers of coastal change and hazards fluid flow and transport in ground water flow systems. Likewise, individuals trying to predict the risk for commercial loans need strong knowledge and skills with tools like python and cross validation techniques but also need knowledge about factors that affect risk such as business position, capital and earnings, risk position, and funding and liquidity.[13] Through its concentration areas, the proposed program will provide students the skills necessary to plan and oversee the lifecycle of a data science project from identification of problem, through collection and assessment of data and data analysis, to communication of actionable insight to stakeholders. Students will learn how to prepare written and oral reports, including data visualizations, to

---

[9] Palak, J. (2019, May 21). *How Data Science Has Helped Amazon to be the World's Prime?* DataFlair. https://medium.com/dataflair/how-data-science-has-taken-amazon-on-top-360043e146ec. Accessed Jun. 29, 2022.

[10] LinkedIn Engineering. (2022). *Data Science.* LinkedIn. https://engineering.linkedin.com/teams/data/data-science. Accessed Jun. 29, 2022.

[11] Davenport, T.H., and Patil, D.J. (2012, October). *Data Scientist: The Sexiest Job of the 21st Century.* Harvard Business Review. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century. Accessed Jun. 29, 2022.

[12] Business Higher Education Forum. (2017, April). *Investing in America's data science and analytics talent. The case for action.* PwC. Accessed Jun. 29, 2022.

[13] Farragher, M. (2022, April 20). Why Economists Should Embrace Data Science. Cambridge Spark. https://www.cambridgespark.com/info/why-economists-should-embrace-data-science#:~:text=A%20brief%20comparison%20of%20economics,data%20scientist%20focuses%20on%20prediction. Accessed Jun. 29, 2022.

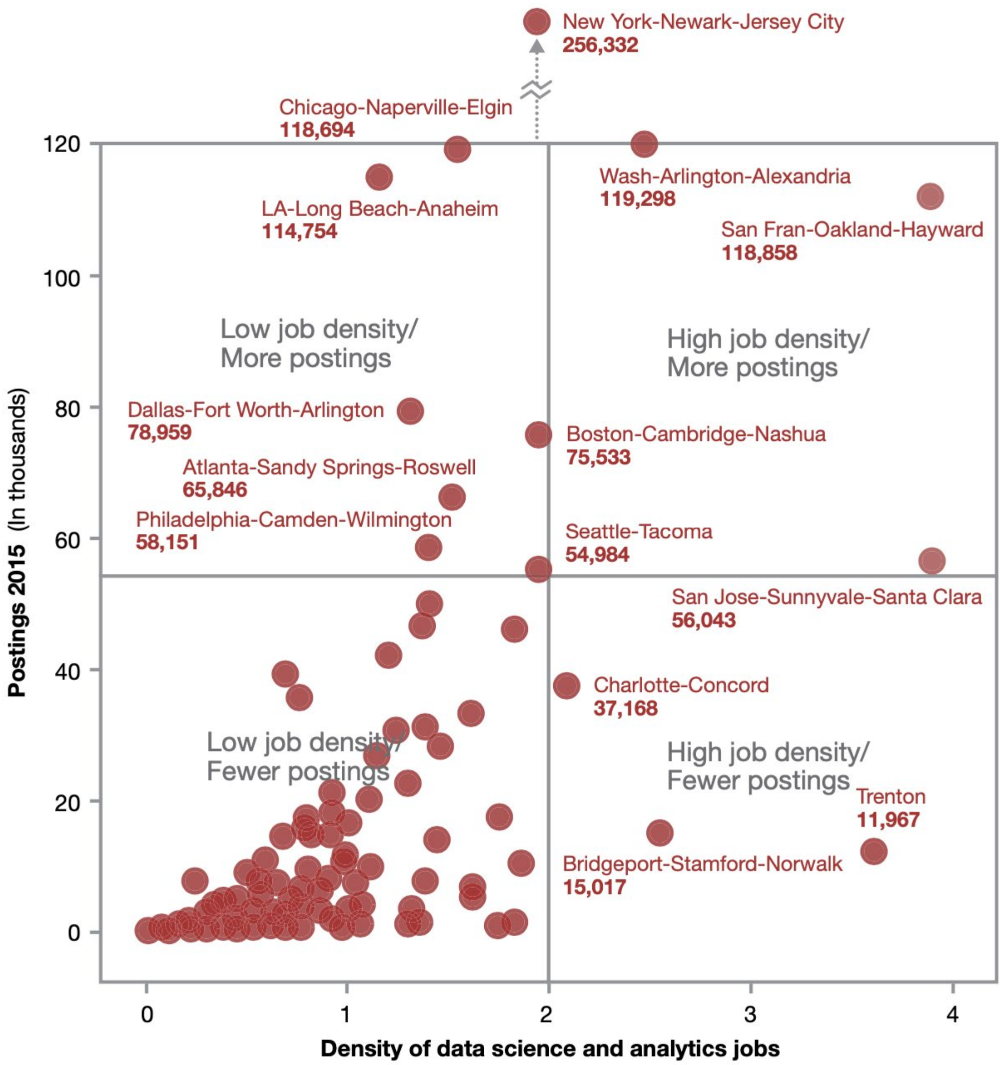
communicate the results and limitations of data analyses, including potential bias and ethical considerations, and to recommend actions to stakeholders.

Employment Demand

Various organizations have noted the demand for data scientists. For example, the Business-Higher Education Forum (BHEF),[14] projects continued growth and demand for graduates with Data Science skills. Prepared in partnership with PwC,[15] a well-known business and government consultancy, the report[16] notes that in 2015 data science-related job postings exceeded the total number of postings combined for registered nurses and truck drivers, two of the largest hiring occupations in the US.

The report includes the following graphic[17] illustrating where the demand for data science and analytics talent is greatest nationally as of 2015. It shows where the demand is concentrated (referred to as density) as well as the level of demand measured as job postings. We note that the northern Virginia area, indicated as "Wash-Arlington-Alexandria" in the upper right area of the plot, is one of a handful of locations that has both high concentration and high volume of postings, an indicator of high demand. According to the report, the Northern Virginia area had the second highest number of data science and analytics postings at 119,298, second only to the New York City metropolitan area. In a related report,[18] BHEF "… identified four foundational areas for 21st century skills". Among these are what they refer to as "digital skills" which include data science, analytics, and big data, the skills that our proposed degree will deliver to students.[19]

[14] Business Higher Education Forum. https://www.bhef.com
[15] PwC. (2017). https://www.pwc.com/us/en
[16] Business Higher Education Forum. (2017, April). Investing in America's Data Science and Analytics Talent. https://www.bhef.com/sites/default/files/bhef_2017_investing_in_dsa.pdf
[17] Ibid.
[18] "Creating the Future Workforce Today", Business Higher Education Forum, 2021
[19] Ibid https://www.bhef.com/sites/default/files/bhef_2017_investing_in_dsa.pdf



Consistent with the findings of the BEHF report, a recent report[20] from LinkedIn, the leading professional networking site that mines its membership data to identify jobs in high demand, identified 15 job categories, including Data Scientist,[21] that are in demand and hiring now. It asserts that businesses are leaning more heavily on Data Scientists as they navigate an increasingly uncertain world and that hiring for data scientist roles has grown nearly 46% since 2019. It lists the Washington DC area, including northern Virginia, among the top locations hiring Data Scientists and quotes a salary range of $100,000 to $130,000. The report includes a link to job postings that can be filtered to the Washington, DC area. As of July 2021, a sample of major companies posting jobs in the DC area included Microsoft, Deloitte, Facebook, MedStar Health, a healthcare provider, and Uber.

[20] LinkedIn. (2021, January 12). LinkedIn Jobs on the Rise: 15 opportunities that are in demand and hiring now. https://apastyle.apa.org/style-grammar-guidelines/references/examples/webpage-website-references#4

[21] Ibid. https://apastyle.apa.org/style-grammar-guidelines/references/examples/webpage-website-references#4

In 2019 a unit of the Wharton School of the University of Pennsylvania[22] and Working Nation[23] hosted a "town hall" titled "The Future is Now: Closing the Data Analytics Skills Gap."[24] The town hall's panel of a dozen leaders from business, academia, government, and the non-profit sectors all cited the need for employees with data science and analytics skills. As an example, one panel discussion[25] touches on the participants struggle to find data analytics talent with the requisite degree of "quantitative rigor" that the proposed degree will provide to our students.

The Demand in Virginia

The demand for data science talent is as great in the Commonwealth. In May 2021 the Virginia Economic Development Partnership[26] (VEDP) commissioned McKinsey & Company[27] to conduct a survey of 14 corporations in the National Capital Region (NCR) concerning data science talent.

See appendix D for the full slide presentation.

About the Companies

- 50% of the companies had revenue greater than $5 billion (slide 4)
- 70% have 500 or more employees in the NCR (slide 4)
- 86% are headquartered in the NCR with no other headquarters elsewhere (slide5)
- 85% report that the largest portion of their employee base is in Virginia or DC (slide 5)

Demand for Data Scientists

- 28% of the companies report that competition is "fierce" to fill Data Science positions with well-qualified candidates (slide 9)
- 26% report that competition is "fierce" to fill Data Science positions with recent graduates (slide 10)
- 32% report that a larger pool of BS and MS graduates in Data Science programs would be helpful in meeting their organizations Data Science needs (sidebar slide 12)
- On a priority scale from 1 to 10 the average score was 7 as the priority for hiring candidates for Data Science roles (slide 12)

The demand for data science talent extends to some of Virginia's key industries. As cited by the VEDP, Virginia is home to one of the largest life science hubs on the East Coast including proximity to federal agencies such as NIH, DARPA, and FDA, all of which are heavily reliant on a workforce with Data Science skills. The same holds true for Virginia's robust aerospace sector that includes Boeing, Northrup Grumman, and Raytheon, among other companies headquartered

---

[22] Wharton University of Pennsylvania. (2022). Wharton Customer Analytics. https://wca.wharton.upenn.edu/about-wca/
[23] WorkingNation. (2021). https://workingnation.com/
[24] WorkingNation and The Wharton Customer Analytics Initiative (WCAI). (2019, February 26). The Future Is Now: Closing the Data Analytics Skills Gap Town Hall. [YouTube]. https://workingnation.com/wharton/
[25] WorkingNation and The Wharton Customer Analytics Initiative (WCAI). (2019, April 17). The Future Is Now: The struggle to find data analytics talent Town Hall. [YouTube]. https://www.youtube.com/watch?v=a1qLClgKCZc
[26] The Virginia Economic Development Partnership. (2022). https://www.vedp.org/
[27] McKinsey & Company. (2022). https://www.mckinsey.com/

here. The proposed degree will help meet the demand for individuals with Data Science skills in these and other key sectors of Virginia's economy.

The Commonwealth's Expectations (2019 – 2020 Biennium)

In the "2019 Virginia Acts of Assembly, Chapter 854",[28] the Commonwealth set its expectations for increased production of degrees awarded in Data Science and Technology on the part of VA Tech and other state-supported universities. Out of the total appropriation for VA Tech, $5,215,880 was "designated to address increased degree production in Data Science & Technology, Science and Engineering, Healthcare and Education."[29] Out of this amount "Virginia Tech is expected to increase Data Science and Technology awards by 60 in the second year."[30] We submit that this reflects the importance of Data Science education in the Commonwealth. Given the demand for individuals with Data Science degrees as presented here and below, we anticipate the Assembly's emphasis on continued production of Data Science degrees, including at the master's level, will continue. The proposed degree is in response to this emphasis.

Bureau of Labor Statistics (BLS) Information

Graduates of the proposed MS in Data Science degree will be qualified to work as data scientists employed by industry, academic institutions, and government. In response to a data analysis assignment they will be able to determine which data are available and needed to perform the analysis; collect, organize and analyze the data; select and utilize algorithms and models to analyze the data; validate the results; and communicate the results of their analysis including visualization techniques as well as and make recommendations based on their analysis.

In its latest Occupational Outlook Handbook, the BLS projects that employment for Data Scientists will grow "36 percent from 2021to 2031, much faster than the average for all occupations."[31] About 13,500 openings for data scientists are projected each year, on average, over the decade. "The median annual wage for data scientists was $100,910 in May 2021" compared to a "median annual wage of $45,760 across all occupations."[32]

According to the BLS, while data scientists "typically need at least a bachelor's degree in mathematics, statistics, computer science, or a related field to enter the occupation" some employers "require or prefer candidates that have a masters or doctoral degree."[33] Among important qualities data scientists should possess, they include "communications skills" and "problem solving skills." [34] In addition, "at the college level, courses in computer science are important in addition to math and statistics" and that students "must learn data-oriented programming languages" as well as other software.[35] The proposed degree program has been designed so graduates gain knowledge and skills in programming, modeling, data analytics, as

---

[28] Commonwealth of Virginia. (2019, May 1). 2019 Virginia Acts of Assembly, Chapter 854 https://budget.lis.virginia.gov/get/budget/3929/

[29] Ibid. Page 255. https://budget.lis.virginia.gov/get/budget/3929/

[30] Ibid. Page 255. https://budget.lis.virginia.gov/get/budget/3929/

[31] BLS Occupational Outlook Handbook. https://www.bls.gov/ooh/math/data-scientists.htm#tab-6

[32] Ibid. https://www.bls.gov/ooh/math/data-scientists.htm#tab-5

[33] Ibid. https://www.bls.gov/ooh/math/data-scientists.htm#tab-4

[34] Ibid. https://www.bls.gov/ooh/math/data-scientists.htm#tab-4

[35] Ibid. https://www.bls.gov/ooh/math/data-scientists.htm#tab-4

well as oral, written, and visual communication skills and working on teams to solve real-world problems.

Additionally, the BLS says that "some data scientists choose to focus on a particular area of work. For example, data scientists who have a strong coding or engineering background may develop or recommend systems, build machine learning algorithms, and devise ways to enhance web-browsing functions. Others conduct research for reports or academic journals. Still others focus on improving business strategy for activities such as marketing, sales, and user engagement." [36] The degree program's concentrations provide students the opportunity to focus their talents on a wide variety of high-value areas of interest.

Virginia Employment Commission Information

The Virginia Employment Commission (VEC) does not have a category for data scientists, therefore does not make occupational projections for such. Nonetheless, other relevant employment categories in the VEC's "Long-Term Occupational Projections" are appropriate for graduates from this degree program and show substantial positive projections over the next ten years. [37]

Virginia Employment Commission, Labor Market Information 2020-2030 (10-Yr)

Occupation Title	Base Year Employment	Projected Employment	Total Projected Difference	Total Percent Change	Annual Change	Education
Statisticians	1,276	1,744	468	36.7%	47	Master's
Computer and Mathematical Occupations	245,652	287,041	41,389	16.8%	4,139	Not applicable
Mathematical Science Occupations	12,356	15,658	3,302	26.7%	330	Not applicable

See Appendix E for employment announcements.
See Appendix F for letters of support from prospective employers.

Duplication

Three (3) public institutions in Virginia that offer a total of four M.S. degrees similar or related to our proposed M.S. in Applied Data Science. University of Virginia and Old Dominion University each offer one degree, both with CIP Code 30.7001. George Mason University offers two degrees, neither of which uses the referenced CIP Code.[38]

[36] Ibid. https://www.bls.gov/ooh/math/data-scientists.htm#tab-4
[37] Virginia Employment Commission, Labor Market Information. https://virginiaworks.com/Occupational-Projections
[38] As of January 31, 2023 the only MS degrees with CIP Code 30.7001 in the Commonwealth are those of UVA and ODU.

University of Virginia (UVA)
M.S. in Data Science (CIP Code 30.7001)[39]
UVA offers a Master of Science degree in Data Science. The UVA program requires students to complete 32 credit hours comprising nine core courses (23 credits), elective coursework (6 credits), and a capstone project sequence totaling 3 credits.

Similarities
Both programs require core coursework in statistics and computer science topics as discussed below. Both programs require capstone project work that emphasize teamwork and experiential learning. The UVA program requires a course in Programming for Data Science (3 credits) and one in Foundations of Computer Science (3 credits). The proposed program requires two core courses, Computation for Data Sciences (3 credits) and Programming Models for Big Data (3 credits) that cover similar topics.

The UVA program requires a core course in Linear Models for Data Science (3 credits) and a core course in Statistical Learning (3 credits). The proposed program requires a core two-course sequence in Data Analytics (each 3 credits) that covers similar topics such as linear models and statistical learning, in addition to other data science topics.

The UVA program requires a core course in the Practice and Application of Data Science (3 credits) and a core course in Ethics of Big Data (2 credits). The proposed program requires a core course in Foundations of Data Science (3 credits) that discusses the practice of Data Science, and a core course in Communication in Team-Based Data Science (3 credits), both of which include coverage of ethics and bias in data science.

Differences
The UVA program allows students to select 2 courses (6 credits), in consultation with a program advisor, from a list of electives from across the university. The proposed program offers students 12 credits of electives organized into concentrations, including an individually tailored concentration, as described above. The UVA program does not provide such concentrations.

An additional difference is that the UVA program requires core courses in Bayesian Machine Learning (3 credits) and Deep Learning (3 credits). The proposed program does not require similar core courses but offers elective courses in Machine Learning with Big Data (3 credits) and Scalable Computing for Computational Modeling and Data Analytics (3 credits).

Old Dominion University (ODU)
M.S. in Data Science and Analytics (CIP Code 30.7001)[40]
ODU offers a Master of Science degree in Data Science and Analytics. The ODU program requires students to complete 30 credit hours including a capstone course (3 credits). The program requires students to complete 5 core courses (15 credits), four courses (12 credits)

---

[39] University of Virginia. (2022). Master of Science in Data Science. https://datascience.virginia.edu/degrees/info/academics
[40] Old Dominion University. (2022). The Graduate Catalog. https://catalog.odu.edu/graduate/graduateschool/

included as part of a concentration area, chosen from one of four areas, and a capstone project (3 credits).

Similarities
The ODU and the proposed program require core coursework in statistics and computer science topics. Both programs have a capstone project requirement. The ODU program requires a course in Statistical/Probability Modeling for Data Science and Analytics (3 credits) and Statistical Tools for Data Science and Analytics (3 credits) in its core curriculum. The proposed program includes a required 2 course sequence in Data Analytics (each 3 credits) in its core covering similar topics.

The ODU program requires Introduction to Data Science and Analytics (3 credits) in its core. The proposed program requires Foundations of Data Science (3 credits) in its core.

The ODU program allows students to pursue a concentration in one of four areas, each totaling 12 credit hours. The proposed program allows students to pursue one of several defined concentrations or to choose electives from among approved electives for a total of 12 credit hours.

Both programs have capstone project requirements totaling 3 credits.

Differences
The proposed program is significantly different in its core curriculum totaling 21 credit hours compared to ODU's program of 18 credit hours, each including 3 credits of capstone project work. The proposed program has two courses, Foundations of Data Science (3 credits) and Communication in Team-Based Data Science (3 credits) that have no counterpart in the ODU program in terms of the range of topics covered. For example, our course in Team-Based Communication in Data Science includes content related to ethics and bias in data science. The ODU program does not appear to have a similar course in its core curriculum.

The ODU program requires Data Visualization (3 credits) in its core curriculum. The proposed program requires Communication in Team-Based Data Science (3 credits) that covers data visualization in addition to other communications topics.

The ODU program has a core computer science course in Data Analytics and Big Data (3 credits). The proposed program has two required computer science courses in its core curriculum, Computation for Data Sciences (3credits) and Programming Models for Big Data (3credits) totaling 6 credits.

George Mason University (GMU)
M.S. in Computational Science (CIP Code 30.3001)[41]
GMU offers a Master of Science degree in Computational Science. The GMU program requires students to complete 30 credit hours. The program requires students to select 2 courses (6 credits) from a list of four core courses and five courses (15 credits) from its computational

[41] George Mason University (2022). Master of Science in Computational Science. https://catalog.gmu.edu/colleges-schools/science/computational-data-sciences/computational-sciences-ms/#text

extended core whose courses come from three extended core areas[42]: Computational and Data Sciences, Computational Science and Informatics, and Computational Social Science. Students also select three courses (9 credit hours) from approved electives.

Similarities
The proposed program requires core courses in Computation for the Data Sciences and Programming Models for Big Data (each 3 credits). The GMU program offers three similar courses, Computational Science Tools, Introduction to Scientific Programming, and Quantitative Foundations for Computational Sciences (each 3 credits), as electives only.

Differences
The GMU required courses comprise a “core” of two courses (6 credits) selected from four courses and a “computational extended core”, consisting of five courses (15 credits) selected from a large number of courses from three different areas. Three of the core courses (each 3 credits) are Numerical Methods, Scientific Databases, and Scientific and Statistical Visualization, which do not correspond to any courses in the proposed program. Several of the courses in the computational extended core, such as Introduction to Computational Social Science and Origins of Social Complexity, are oriented to the social sciences. There are no such courses in the proposed program.

Contrast this with the proposed program that has a defined core curriculum that includes two statistics courses (6 credits), two computer science courses (6 credits), and two courses covering the fundamentals of data science and communications for data science (6 credits). Although GMU has a statistics department there are no statistics courses among either its core or computational extended core courses.

Notable also is that the proposed program has a capstone project requirement and offers several well-defined concentrations for its electives. The GMU program has neither.

George Mason University (GMU)
M.S. in Data Analytics Engineering (CIP Code 11.0802)[43]
GMU offers a Master of Science degree in Data Analytics Engineering. The program requires students to complete 30 credit hours. The GMU program requires students to complete 4 courses of 3 credits each plus a data analytics project of 3 credit hours for a total of 15 credit hours. Students may choose electives from several concentrations totaling 15 credit hours to complete the degree.

Similarities
The GMU program requires students to take a core course in Analytics: Big Data to Information (3 credits). This is similar to the proposed program’s required core course in Foundations of Data Science (3 credits).

[42] George Mason University (2022). Master of Science in Computational Science. https://catalog.gmu.edu/colleges-schools/science/computational-data-sciences/computational-sciences-ms/#requirementstext
[43] George Mason University (2022). Master of Science in Data Analytics Engineering. https://catalog.gmu.edu/colleges-schools/engineering-computing/data-analytics-engineering-ms/

The GMU program requires students to complete a Data Analytics Project (3 credits). This is similar to the proposed program's 2 course capstone project sequence (3 credits total).

The GMU program allows students to pursue an elective concentration totaling 15 credit hours or to choose 15 credits of electives from among a list of elective courses. This is similar to the proposed program's 12 credit hours of concentrations or electives.

Differences

The proposed program requires two core courses: Computation for the Data Sciences (3 credits) and Programming Models for Big Data (3 credits) totaling 6 credits. The GMU program requires students to select either Principles of Data Management and Mining (3 credits) or Theory and Applications of Data Mining (3 credits). The GMU courses have content similar to the proposed program's required core courses, but students take only one of the two totaling 3 credits.

The proposed program requires a core two-course sequence in Data Analytics (each 3 credits) for a total of 6 credits. The GMU program requires students to choose either Applied Statistics I (3 credits) or Applied Statistics and Visualization for Analytics (3 credits) for a total of 3 credits.

The GMU program requires students to select from two operations research courses, either Analytics and Decision Analysis (3 credits) or Operations Research: Deterministic Models (3 credits). The proposed program does not require operations research courses.

The proposed program requires core courses in Communication in Team-Based Data Science (3 credits) and Foundations of Data Science (3 credits). The GMU program does not have similar courses.

Enrollment and Degrees Awarded at Comparable Programs in Virginia

Enrollments[44]	**Fall 2018**	**Fall 2019**	**Fall 2020**	**Fall 2021**	**Fall 2022**
George Mason University (Data Engineering Analytics)	383	476	527	602	727
George Mason University (Computational Science)	-	-	-	-	0
Old Dominion University	0	0	0	0	0
University of Virginia	69	99	231	-	-
Degrees Awarded[45]	**2018**	**2019**	**2020**	**2021**	**2022**
George Mason University	119	152	198	222	153
George Mason University	-	-	-	-	0
Old Dominion University	0	0	-	-	-
University of Virginia	32	57	54	101	-

[44] State Council of Higher Education for Virginia (SCHEV). *Enrollment Summary by Program Level*. https://research.schev.edu//enrollment/E01_Report.asp. (Accessed: November 13, 2022).
[45] State Council of Higher Education for Virginia (SCHEV). Completions, Program Detail. https://research.schev.edu//Completions/C1Level2_Report.asp. (Accessed: November 13, 2022).

Student Demand

Evidence of student demand comes from 1 source: A student survey

Virginia Tech evaluated student demand for the proposed M.S. in Applied Data Science degree program via a student survey.

<u>Student Survey</u>
In February 2022, instructors in courses required for undergraduate Computer Science, Computational Modeling and Data Analytics, Mathematics, and Statistics degree programs administered an electronic survey via Canvas (VA Tech's learning management system) to all students enrolled. Accompanying the survey was a video that introduced the students to the program's admission requirements and core courses as well as example programs of study. The online survey remained open for 1.5 weeks. A total of 129 responses were received, 46 of which were from seniors and 48 from juniors (94 total). The prompt most relevant to student interest in the proposed degree is listed below, followed by the percent of survey respondents who answered: "very likely" or "likely" on a 5-point scale[46].

- If Virginia Tech offers a Master's degree in Applied Data Science, how likely are you to enroll? 44 (47%) of the 94 seniors and juniors responded they were very likely or likely to enroll. 22 said they were very likely to enroll and 22 said they were likely to enroll.

See Appendix G for the original survey.

State Council of Higher Education for Virginia
Summary of Projected Enrollments in Proposed Program

Year 1		Year 2		Year 3		Year 4 **Target Year (2-year institutions)**			Year 5 **Target Year (4-year institutions)**		
2023 - 2024		2024 - 2025		2025 - 2026		2026 - 2027			2027 - 2028		
HDCT 10	FTES 3	HDCT 35	FTES 24	HDCT 45	FTES 34	HDCT 50	FTES 35	GRAD 30	HDCT 60	FTES 45	GRAD 40

Assumptions:
Retention percentage: 95%
Full-time students: 100% Part-time students: 0%
Full-time students credit hours per semester: 11-13
Full-time students graduate in 1 year
Summer semester (6 credits) is required

[46] A 5-point Likert scale was used: 5) very likely, 4) likely, 3) possibly, 2) unlikely, 1) very unlikely

Projected Resources for the Proposed Program

Resource Needs

Virginia Tech and the Departments of Statistics, Mathematics and Computer Science have all of the faculty, classified support, equipment, space, library, and other resources necessary to initiate the proposed M.S. in Applied Data Science degree program. Assessments of the need for full-time and part-time faculty are based on the ratio of 1.00 FTE of faculty effort for every 11.00 FTE of student enrollment. The proposed program will utilize a total of 0.375 FTE faculty instructional effort in the initial year of 2023-2024; a total of 2.25 FTE faculty instructional effort in year two; a total of 3.125 FTE faculty instructional effort in year three; a total of 3.25 FTE faculty instructional effort in year four; and q total of 4.125 FTE faculty instructional effort by the fifth year of 2027-2028.

Full-time Faculty

Year 1: 2023-2024

In year one, no full-time faculty FTE of instructional effort will be dedicated to the proposed program.

Year 2: 2024-2025

In year two, two new full-time faculty members will be hired to support the proposed program. Each new faculty member will support the proposed program with 0.5 FTE instructional effort.

New Hire 1

The Department of Statistics will hire one new faculty member who will dedicate 0.5 FTE instructional effort to the proposed program. The faculty member will be hired at an assistant or associate professor faculty rank with expertise in statistics with an approximate salary of $110,000 and fringe benefits of $42,000 for a total of $152,000. The portion of salary devoted to the proposed program will be $55,000 with fringe benefits of $21,000 for a total of $76,000.

New Hire 2

In year two, 2024-2025, the Department of Statistics will hire a second new faculty member who will dedicate 0.5 FTE instructional effort to the proposed program. The faculty member will be hired at an assistant or associate professor faculty rank with expertise in statistics with an approximate salary of $110,000 and fringe benefits of $42,000 for a total of $152,000. The portion of salary devoted to the proposed program will be $55,000 with fringe benefits of $21,000 for a total of $76,000.

A total of 1.0 FTE of full-time faculty instructional effort will be dedicated to the proposed program in year two, 2024-2025.

Year 3: 2025-2026

In year three, two existing faculty members from the Department of Statistics will continue to dedicate 0.5 FTE instructional effort each for a total of 1.0 FTE instructional effort dedicated to the proposed program.

A total of 1.0 FTE of full-time faculty instructional effort will continue to be dedicated to the proposed program in year three, 2025-2026.

Year 4: 2026-2027
In year four, two existing faculty members from the Department of Statistics will continue to dedicate 0.5 FTE instructional effort each for a total of 1.0 FTE instructional effort dedicated to the proposed program.

A total of 1.0 FTE of full-time faculty instructional effort will continue to be dedicated to the proposed program in year four, 2026-2027.

Year 5: 2027-2028
In year five, two existing faculty members from the Department of Statistics will continue to dedicate 0.5 FTE instructional effort each for a total of 1.0 FTE instructional effort dedicated to the proposed program.

In year five, one additional existing faculty member from the Department of Statistics will be reallocated to dedicate 0.5 FTE instructional effort to the proposed program.

In year five, one existing faculty member from the Department of Mathematics will be reallocated to dedicate 0.5 FTE instructional effort to the proposed program.

A total of 2.0 FTE of full-time faculty instructional effort will be dedicated to the proposed program in year five, 2027-2028.

Part-time Faculty
Year 1: 2023-2024
In year one, one existing faculty member from the College of Science, Academy of Data Science, will be reallocated to dedicate 0.125 FTE instructional effort to the proposed program.

In year one, one existing faculty member from the Department of Statistics will be reallocated to dedicate 0.125 FTE instructional effort to the proposed program.

In year one, one existing faculty member from the Department of Computer Science will be reallocated to dedicate 0.125 FTE instructional effort to the proposed program.

A total of 0.375 FTE of part-time faculty instructional effort will be dedicated to the proposed program in year one, 2023-2024.

Year 2: 2024-2025
In year two, one existing faculty member from the College of Science, Academy of Data Science, who had previously dedicated 0.125 FTE, will rise to 0.25 FTE instructional effort dedicated to the proposed program.

In year two, one existing faculty member from the Department of Statistics who had previously dedicated 0.125 FTE will rise to 0.25 FTE instructional effort dedicated to the proposed program.

In year two, one existing faculty member from the Department of Mathematics will be reallocated to dedicate 0.25 FTE instructional effort to the proposed program.

In year two, one existing faculty member from the Department of Computer Science will continue to dedicate 0.125 FTE instructional effort to the proposed program.

In year two, three existing faculty members from the Department of Computer Science will be reallocated to dedicate 0.125 FTE instructional effort each for a total of 0.375 FTE instructional effort dedicated to the proposed program.

A total of 1.25 FTE of part-time faculty instructional effort will be dedicated to the proposed program in year two, 2024-2025.

Year 3: 2025-2026
In year three, one existing faculty member from the College of Science, Academy of Data Science, who had previously dedicated 0.25 FTE, will rise to 0.375 FTE instructional effort dedicated to the proposed program.

In year three, one existing faculty member from the Department of Statistics will continue to dedicate 0.25 FTE instructional effort dedicated to the proposed program.

In year three, one existing faculty member from the Department of Mathematics will continue to dedicate 0.25 FTE instructional effort to the proposed program.

In year three, four existing faculty members from the Department of Computer Science will continue to dedicate 0.125 FTE instructional effort each for a total of 0.5 FTE instructional effort dedicated to the proposed program.

In year three, three new full-time faculty members will be hired to support the proposed program. Each new faculty member will support the proposed program with 0.25 FTE instructional effort for a total of 0.75 FTE instructional effort dedicated to the proposed program.

New Hire 3
In year three, the Department of Statistics will hire one new faculty member who will dedicate 0.25 FTE instructional effort to the proposed program. The faculty member will be hired at an assistant or associate professor faculty rank with expertise in statistics with an approximate salary of $120,000 and fringe benefits of $45,000 for a total of $165,000. The portion of salary devoted to the proposed program will be $30,000 with fringe benefits of $11,250 for a total of $41,250.

New Hire 4
In year three, the Department of Mathematics will hire one new faculty member who will dedicate 0.25 FTE instructional effort to the proposed program. The faculty member will be hired at an assistant or associate professor faculty rank with expertise in mathematics with an approximate salary of $120,000 and fringe benefits of $45,000 for a total of $155,000. The portion of salary devoted to the proposed program will be $30,000 with fringe benefits of $11,250 for a total of $41,250.

New Hire 5
In year three, the Department of Computer Science will hire one new faculty member who will dedicate 0.25 FTE instructional effort to the proposed program. The faculty member will be hired at an assistant or associate professor faculty rank with expertise in computer science with an approximate salary of $150,000 and fringe benefits of $51,900 for a total of $201,900. The portion of salary devoted to the proposed program will be $37,500 with fringe benefits of $12,975 for a total of $50,475.

A total of 2.125 FTE of part-time faculty instructional effort will be dedicated to the proposed program in year three, 2025-2026.

<u>Year 4: 2026-2027</u>
In year four, one existing faculty member from the College of Science, Academy of Data Science, will continue to dedicate 0.375 FTE instructional effort dedicated to the proposed program.

In year four, one existing faculty member from the Department of Statistics will continue to dedicate 0.25 FTE instructional effort dedicated to the proposed program.

In year four, one existing faculty member from the Department of Statistics who had previously dedicated 0.25 FTE, will rise to 0.375 FTE instructional effort dedicated to the proposed program.

In year four, two existing faculty members from the Department of Mathematics will continue to dedicate 0.25 FTE instructional effort each for a total of 0.5 FTE instructional effort dedicated to the proposed program.

In year four, four existing faculty members from the Department of Computer Science will continue to dedicate 0.125 FTE instructional effort each for a total of 0.5 FTE instructional effort dedicated to the proposed program.

In year four, an additional existing faculty member from the Department of Computer Science will continue to dedicate 0.25 FTE instructional effort to the proposed program.

A total of 2.25 FTE of part-time faculty instructional effort will be dedicated to the proposed program in year four, 2026-2027.

Year 5: 2027-2028

In year five, one existing faculty member from the College of Science, Academy of Data Science, will continue to dedicate 0.375 FTE instructional effort dedicated to the proposed program.

In year five, one existing faculty member from the Department of Statistics will continue to dedicate 0.375 FTE instructional effort to the proposed program.

In year five, one existing faculty member from the Department of Mathematics who had previously dedicated 0.25 FTE, will rise to 0.375 FTE instructional effort dedicated to the proposed program.

In year five, one existing faculty member from the Department of Computer Science will continue to dedicate 0.25 FTE instructional effort to the proposed program.

In year five, three existing faculty members from the Department of Computer Science will continue to dedicate 0.125 FTE instructional effort each for a total of 0.375 FTE instructional effort dedicated to the proposed program.

In year five, one existing faculty member from the Department of Computer Science who had previously dedicated 0.125 FTE, will rise to 0.375 FTE instructional effort dedicated to the proposed program.

A total of 2.125 FTE of part-time faculty instructional effort will be dedicated to the proposed program in year five, 2027-2028.

Adjunct Faculty

Adjunct faculty will not be required to launch or sustain the proposed program.

Graduate Assistants

No new graduate assistants will be required to initiate or sustain the proposed program.

Classified Positions

No new classified positions will be required to launch or sustain the proposed program.

Targeted Financial Aid

No targeted financial aid will be available or is needed to launch or sustain the proposed program.

Equipment

Equipment will be purchased for all new faculty members. Equipment costs of $4,000 for each new faculty hire is provided for a total of $20,000 for equipment, including computers. Existing office furniture will be used for the new faculty hires.

Library
No new library resources are needed to launch or sustain the proposed program. The library has an adequate collection to support the proposed degree program. Resources include journals and publications for computer science, data science, mathematics, and statistics. As a member of the Virtual Library of Virginia (VIVA), on-line access to journals is also available.

Telecommunications
The proposed program requires no new telecommunications to launch or sustain the proposed program.

Space
Sufficient faculty office space and instructional space currently exists or is under construction to support and sustain the program. No additional space is required.

Cost and Funding Sources to Initiate and Operate the Program			
Informational Category		**Program Initiation Year 2023 - 2024**	**Program Full Enrollment Year[47] 2025 - 2027**
1.	Projected Enrollment (Headcount)	10	45
2.	Projected Enrollment (FTE)	3	30
3.	Projected Enrollment Headcount of In-State Students	10	15
4.	Projected Enrollment Headcount of Out-of-State Students	0	30
5.	Estimated Annual Tuition and E&G Fees for In-state Students in the Proposed Program	$82,610	$123,915
6.	Estimated Annual Tuition and E&G Fees for Out-of-State Students in the Proposed Program	$0	$471,645
7.	Projected Total Revenue from Tuition and E&G Fees Due to the Proposed Program	$82,610	$595,560
8.	Other Funding Sources Dedicated to the Proposed Program (e.g., grant, business entity, private sources)	$0	$0

[47] For the "Full Enrollment Year" use: for associate degrees, initiation year plus 1; for baccalaureate degrees, initiation plus 3; for masters degrees, initiation plus 2; for doctoral degrees, initiation plus 3.

Part V: Certification Statements

1. A request of any kind will be submitted to the General Assembly for funds to initiate and/or maintain the proposed degree program.

   Yes ☐
   No ☒

   If "Yes" is checked, include narrative text to describe: when the request will be made, how much will be requested, what the funds will be used for, and what will be done if the request is not fulfilled.

2. The proposed degree program is included in the institution's most recent six-year plan.

   Yes ☒
   No ☐

   If "No" is checked, include narrative text to explain why the program is being advanced at the present time despite not being included in the six-year plan.

3. The institution's governing board has been provided information regarding duplication (if applicable) and labor market projections as part of its approval action.

   Yes ☒
   No ☐

   If "No" is checked, include narrative text to explain why the governing board has not been provided the information.

The institution's Chief Academic Officer attests to the accuracy of the above statements

____
Name

____
Signature   Date

Appendices

Appendix A
Sample Plans of Study

Full-time Students

Year One	
Summer	**Credits**
ADS 5064: Foundations of Data Science	3
STAT/CS 5525: Data Analytics	3
CS 5045: Computation for the Data Sciences	3
Total	**9**
Fall	**Credits**
ADS 5224: Communication in Team-Based Data Science	3
CS 5054: Programming Models for Big Data	3
ADS 5804: Capstone Experience I: Data and Definition	1
Concentration Course	3
Concentration Course	3
Total	**13**
Spring	**Credits**
STAT/CS 5526: Data Analytics	3
ADS 5814: Capstone Experience II: Implementation	2
Concentration Course	3
Concentration Course	3
Total	**11**

Total: 33 credit hours

Part-time Students

Year One	
Fall	**Credits**
ADS 5064: Foundations of Data Science	3
CS 5045: Computation for the Data Sciences	3
Total	**6**
Spring	**Credits**
STAT/CS 5525: Data Analytics	3
CS 5054: Programming Models for Big Data	3
Total	6
Summer	**Credits**
STAT/CS 5526: Data Analytics	3
ADS 5224: Communication in Team-Based Data Science	3
Total	**6**
Year Two	
Fall	**Credits**
ADS 5804: Capstone Experience I: Data and Definition	1
Concentration Course	3
Concentration Course	3
Total	**7**
Spring	**Credits**
ADS 5814: Capstone Experience II: Implementation	2
Concentration Course	3
Concentration Course	3
Total	**8**

Total: 33 credit hours

Appendix B
Course Descriptions

New courses are denoted with an asterisk (*).

Core Courses
ADS 5064: Foundations of Data Science (3 credits) *
History and evolution of data science. End to end process-based approach for data science. Application of tools, methods, and phases of the data science process. Evaluation of the data science process and tools in data science projects. Integrating data science solutions in the context of data structures, big data, network architecture, and software development. Project management as applied to data science projects. Applied ethical decision making in the context of the data science project. Prerequisite: Graduate Standing.

ADS 5224: Communication in Team-Based Data Science (3 credits) *
Oral, written, and visual communication skills required to successfully execute data science projects in a team environment. Interpersonal and formal communication in the context of data science projects. Aspects of cross-cultural communication. Principles of effective data visualization. Explaining complex concepts to non-data scientists. Facilitation of reproducible and replicable data science. Prerequisite: Graduate Standing.

CS 5045: Computation for the Data Sciences (3 credits)
Covers fundamentals of computer science and background in data sciences needed by graduate students without a computer science background. Programming language syntax and semantics; abstraction and object-oriented programming; data structures; databases; visualization; ethics and data manipulation. Prerequisite: Graduate Standing.

CS 5054: Programming Models for Big Data (3 credits) *
Survey of computer science concepts and tools that enable efficient computational science and data analytics with big data. Data structure design and implementation. Analysis of data structure and algorithm performance. Introduction to high-performance computer architectures and parallel computation. Basic operating systems concepts that influence the performance of large-scale computational modeling and data analytics. Software tools for computational modeling. Prerequisite: CS 5045.

STAT/CS 5525: Data Analytics (3 credits)
Basic techniques in data analytics including the preparation and manipulation of data for analysis and the creation of data files from multiple and dissimilar sources. The data mining and knowledge discovery process. Overview of data mining algorithms in classification, clustering, association analysis, probabilistic modeling, and matrix decompositions. Detailed study of classification methods including tree-based methods, Bayesian methods, logistic regression, ensemble, bagging and boosting methods, neural network methods, use of support vectors and Bayesian networks. Detailed study of clustering methods including k-means, hierarchical and self-organizing map methods. Prerequisite: Graduate Standing.

STAT/CS 5526: Data Analytics (3 credits)
Techniques in unsupervised and visualized learning in high dimension spaces. Theoretical, probabilistic, and applied aspects of data analytics. Methods include generalized linear models in high dimensional spaces, regularization, lasso and related methods, principal component regression (pca), tree methods, and random forests. Clustering methods including k-means, hierarchical clustering, biclustering, and model-based clustering will be thoroughly examined. Distance-based learning methods include multi-dimensional scaling, the self-organizing map, graphical/ network models, and isomap. Supervised learning will consist of discriminant analyses, supervised pca, support vector machines, and kernel methods. Prerequisite: CS 5535 or STAT 5525.

Additional Required Courses
ADS 5804: Capstone Experience I: Definitions and Data (1 credit) *
Application of underlying principles and initial phases of the data science process to a specific project. Problem definition, writing questions to be addressed through data analysis, identification of available data sources, project feasibility, evaluation of adequacy and quality of data sources for addressing the project questions, data preparation and cleaning, preliminary exploratory data analysis, reproducible analysis, ethical aspects including fairness/bias, communication with non-data scientists. Prerequisites: ADS 5224, CS 5045; Corequisite: STAT/CS 5526.

ADS 5814: Capstone Experience II: Implementation (2 credits) *
Application of latter phases of the data science process to a specific project, plan and execute modeling and analysis needed to address the research, business, or policy questions proposed in Capstone Experience I: Definition & Data, solution quality assessment and implementation, assessment and mitigation of risks from ethical dilemmas and fairness/bias issues, communication of results to non-data scientist project stakeholders, comprehensive documentation, production of reproducible code for data scientist colleagues, critique of data science project for improvement. Prerequisites: ADS 5804, STAT/CS 5526.

Concentration Area Courses

<u>Agricultural and Applied Economics Courses</u>
AAEC 5014: Applied Economic Analytics (3 credits)
Mathematical and statistical methods used in applied economic decision making. Applied mathematical optimization, statistical simulation, data visualization, probability theory and linear econometric models to economic, agricultural, and environmental data and problems. Extensive application of quantitative models and modern programming platforms used in applied economic analysis. Prerequisite: Graduate Standing.

AAEC/STAT 5484: Applied Economic Forecasting (3 credits)
Forecasting economic, agricultural, and environmental data using basic linear and non-linear time series models. Emphasis on programming and computational implementation of time series model-selection techniques and practical applications. Prerequisite: Graduate Standing.

AAEC 5804G: Fundamentals of Econometrics (3 credits)
Introduction to economic applications of mathematical and statistical techniques: regression, estimators, hypothesis testing, lagged variables, discrete variables, violations of assumptions, simultaneous equations, instrumental variables, panel data methods. Prerequisite: Graduate Standing.

AAEC 5824: Advanced Applied Economic Analytics (3 credits)
Advanced econometric analysis of problems in agricultural and applied economics. Modern techniques in econometrics and data analytics including multiple regression, classification, instrumental variables, clustering, and regression trees. Prerequisites: Graduate Standing.

<u>Economics Concentration Courses</u>
ECON 4084: Industry Structure (3 credits)
The structure and performance of American industry. Dimensions and measures of market structure. Factors affecting market structure. The relationship between structure and performance. Purpose and effects of antitrust policy, regulation, and other public policies toward industry. Prerequisite: ECON 3104 or ECON 4924.

ECON 4514: Applied Analysis of Banking and Financial Markets (3 credits) *
Analysis of economic data with focus on understanding of decision-making in financial markets. Behavior of and optimal outcomes for individuals (consumption, savings, and investment), financial institutions (lending, borrowing, and risk management), regulators, and policy makers. Statistical tools and inference using recent data sets. Prerequisite: ECON 3254 or ECON 4304.

ECON 5134: Managerial Economics (3 credits) *
Applications of price theory to applied problems at the graduate level. Utility maximization and demand, price indices, and short- and long-run demand. Discrete choice and product quality, location choice, and human capital. Profit maximization, cost minimization, factor demand, and industry model. Durable goods, capital accumulation, and value of a statistical life. Empirical analysis: formulating research questions, finding data from multiple sources, econometric analysis, and presenting findings to non- specialists. Students are assumed to have an understanding of principles-level microeconomics. Prerequisite: Graduate Standing.

ECON 5314: Advanced Big Data Economics (3 credits) *
Applied econometrics dealing with big data. Theoretical, computational, and statistical underpinnings of big data analysis. The use of econometric models and deep machine learning algorithms to analyze the high-dimensional data sets. Implications in research focusing on economic questions that arise from rapid changes in data availability and computational technology. Materials are hands-on tutorials that come with Python codes and real-world data sets. Prerequisite: Graduate Standing.

<u>Forest Resources and Environmental Conservation Courses</u>
BIOL/FREC 5034: Ecosystems Dynamics (4 credits)

Application of the systems perspective to functional characteristics and dynamics of ecosystems: energy flow, biogeochemical cycling, and stability/resilience in the response to environmental change. Computer simulation of ecosystem response to change. Prerequisites: MATH 2014, CS 1014, or MATH 2014.

FREC 5114G: Advanced Information Technology for Natural Resources Management (3 credits)
Course will introduce students to the theory and applications of database management systems (DBMS) and geographic information systems (GIS). Uses, challenges, and limitations of these technologies in natural resource management application will be discussed. Prerequisite: FREC 2214 or GEOG 2314.

FREC/GEOG 5154: Hyperspectral Remote Sensing for Natural Resources (3 credits)
Theory of spectroscopy and spectrometry from portable spectroradiometers to airborne and spaceborne hyperspectral sensors as relevant to natural resource applications, including vegetation species identification and vegetative health, soil and peat properties, mineral and geothermal characteristics, and water applications. Practical investigation of research tools and techniques used to analyze hyperspectral data. Prerequisite: Graduate Standing.

FREC 5224: Forest Biometry (3 credits)
Theory and practice involved in the measurement and modeling of the growth and yield of forest trees and stands. Prerequisites: FREC 3215, FREC 3216, STAT 5606, STAT 5616.

FREC 5244G: Advanced Hydroinformatics (3 credits)
Analysis and examination of hydrologic data using basic statistics and computer programming. Calculation and interpretation of flow frequency and duration, hydrologic analysis of geospatial digital terrain data, and implementation and analysis of simple hydrologic models. Advanced methods of temporal and spatial hydrologic data visualization using computer programming. Prerequisite: Graduate Standing.

FREC 5254: Remote Sensing of Natural Resources (3 credits)
Philosophy and rationale of remote sensing as a part of the resource management process; comparisons of analogic and digital sensors; sensor selection and proper use; accuracy assessment; signature development; and identification of factors which affect the quality of remotely sensed information. Prerequisite: Graduate Standing.

FREC/AAEC/GEOG 5544: Remote Sensing in the Social Sciences (3 credits)
Principles on the use of remotely sensed (satellite) data in social science research, with key applications in environment, agriculture, and economic development. Basic scripting techniques to extract, visualize, and analyze satellite remote sensing data across the electromagnetic spectrum with cloud-based computing platforms. Development of social science research proposals using remotely sensed data and based on review of relevant seminal and current research articles. Prerequisite: Graduate Standing.

FREC/GEOG 6214: Forestry Lidar Applications (3 credits)
Theoretical underpinning of established and emerging research using light detection and ranging (lidar) technology for forestry applications including detailed terrain mapping and digital

elevation models, canopy height modeling, prediction of forest biophysical parameters, forest physiology and the canopy light regime, watershed mapping and stream modeling, ecological modeling, landscape classifications, and wildlife habitat. Advanced research tools and techniques used to analyze lidar data for different applications. Prerequisites: FREC/FOR 5254, FREC/FOR 5264, GEOG 5034.

Geosciences Concentration Courses

GEOS/MATH 5144: Inverse Theory and Geoscience Applications (3 credits)
Overview of inverse theory, utilizing geophysical examples to illustrate the concepts of model construction, parameter estimation, resolution, and non-uniqueness. Emphasis is on the linear problem, concluding with an overview of nonlinear inversion. Prerequisite: Graduate Standing.

GEO 5184: Advanced Geodesy in the Earth Sciences (3 credits)
Study of measurement of Earth's geometric shape, orientation in space, the gravity field, and how these properties change over time. Geodetic methods of measurement (i.e., GNSS, InSAR, TLS, gravity). Reference frames, geodetic applications, and geodetic advances. Quantitative analysis of geodetic observations with applications to Earth science. Prerequisites: Graduate Standing.

GEOS 5314: Advanced Coastal Hazards (3 credits)
Study of past, current, and future drivers of coastal change and hazards. Integration of concepts and skills from: climatic, isostatic, and tectonic processes that drive sea-level change; geologic (e.g., coastal stratigraphy, microfossils) and instrumental (e.g., tide gauges, satellite altimetry) coastal change reconstructions, models, measurements, and projections. Coastal earthquake, tsunami, hurricane, and storm-surge hazards. Approaches and challenges of communicating coastal hazards to the public. Coastal hazards and public policy. Quantitative analysis of coastal change using observational data. Prerequisite: Graduate Standing.

GEOS 5814: Numerical Modeling of Groundwater (3 credits)
Theory and practice of numerical techniques are developed and applied to fluid flow and transport in ground-water flow systems. Governing equations are formulated using FD and FE techniques with appropriate BCs and ICs. Additional topics include: model conceptualization and grid design in multidimensional systems; practical applications of numerical models including calibration, validation, and prediction; concepts and techniques of advective transport using particle tracking and dispersive transport. Introduction to MODFLOW, MODPATH, MT3D, and others. Prerequisite: Graduate Standing.

GEOS 6104: Advanced Topics in Geophysics (3 credits)
Advanced analysis of one or more topics of geophysics using the most recent techniques, interpretations, and data. Prerequisite: Graduate Standing.

Industrial and Systems Engineering Courses

ISE 5034: Mathematical Probability & Statistics for Industrial Engineers (3 credits)
One-semester mathematical foundations for graduate study of engineering oriented probability and statistics. Re-introduces probability in a rigorous mathematical fashion and re-introduces mathematical statistics as an application of mathematical probability. Establishes a solid mathematical foundation for the type of probability (and statistics) modeling and analysis that is

characteristic of graduate industrial engineering curricula and practice. Prerequisite: Graduate Standing.

ISE 5104: Operations Research (3 credits)
Basic techniques and methods of operations research, including the operations research approach to decision making, model formulations, and analytical methods for their solution. Introduction to implementing these models and methods. Fundamental aspects of optimization techniques (e.g., linear, integer, dynamic and nonlinear programming), network analysis, and Markov processes. Not for credit for students in the Operations Research or Manufacturing Systems Engineering tracks or by students who have taken 5405. Prerequisite: Graduate standing.

ISE 5204: Manufacturing Systems Engineering (3 credits)
Conceptual models of manufacturing, process, and service organizations for various operational levels. Functional activities and interrelationships for each type of manufacturing model. Typical objectives, operating constraints, and informatics for functional activities: production planning/control, material management, facility design/material handling, manufacturing engineering, and quality control. Prerequisite: Graduate Standing.

ISE 5405: Optimization: Linear and Nonlinear Programming (3 credits)
Introduction to theory of linear and nonlinear programming. A mix of theoretical concepts and numerical algorithms to solve the linear and nonlinear programming problems. Modeling for real world problems using linear programming and integer linear programming. Geometric foundations for linear programs – characterization for polyhedral sets and convex analysis. Numerical algorithms for linear programs: simplex method (its geometry and algebra), primal-dual simplex algorithm, revised simplex, two-phase and big-M methods. Farkas' Lemma and Optimality Karush-Kuhn-Tucker (KKT) conditions for linear programs. Duality theory, sensitivity analysis, state-of-the-art modeling language and solvers. Prerequisite: Graduate standing.

ISE 5424: Simulation I (3 credits)
Introduction to stochastic discrete-event simulation. Theoretical foundations for stochastic simulation methodology and design, and analysis of simulation experiments. Simulation modeling and programming in general-purpose languages and using some specialized simulation software packages. Applications are drawn from various settings, such as manufacturing, financial, logistics, and service systems. Course projects on solving a real-world decision-making problem under uncertainty, each involving building a simulation model based on a case description, conducting design of simulation experiments, and performing simulation-based analysis. Prerequisite: ISE 5034 or STAT 5104.

Modeling and Algorithms Concentration Courses
CMDA 4634: Scalable Computing for Computational Modeling and Data Analytics (3 credits)
A focused study of concepts and tools that accelerate computational and data science at scale. Design, analysis, optimization, and modeling of application-driven algorithms suitable for state-of-the-art scalable computing platforms. Software development and engineering for scalable computational science. Prerequisites: CMDA/CS 3634 or CS 4234, CMDA/CS/STAT 3654, CMDA 3605 or CS 3414 or MATH 3414 or MATH 4445.

CS 5644: Machine Learning with Big Data (3 credits)
Basic principles and techniques for big data analytics, including methods for storing, searching, retrieving, and processing large datasets; introduction to basic machine learning libraries for analyzing large datasets; data visualization; case studies with real-world datasets. Prerequisite: CS 5044.

CS 5664: Social Media Analytics (3 credits)
Social media platforms, media feeds, and data formats; machine learning and graph theory foundations of social media analytics; Forms of social media analytics - text analytics, network analytics, and action analytics; Forecasting models and applications, including in marketing, event tracking, surveying, and A/B testing. Prerequisite: CS 5644.

CS 5764: Information Visualization (3 credits)
Examine computer-based strategies for interactive visual presentation of information that enable people to explore, discover, and learn from vast quantities of data. Learn to analyze, design, develop, and evaluate new visualizations and tools. Discuss design principles, interaction strategies, information types, and experimental results. Research-oriented course surveys current literature, and group projects contribute to the state of the art. Prerequisite: Graduate Standing.

CS 5834: Introduction to Urban Computing (3 credits)
Computational approaches to address urban challenges; sensor network testbeds; algorithms for storing, processing, and mining data from urban settings; communicating patterns to decision makers; special focus on epidemiology, sustainability, transportation, social science, urban economics; case studies with applications. Prerequisite: Graduate Standing.

MATH 5424: Numerical Linear Algebra (3 credits)
Matrix factorizations and iterative algorithms. Topics in numerical linear algebra include stability analysis, accuracy, and problem conditioning. Applications include solving large systems of linear equations, linear least-squares, and eigenvalue problems. Prerequisite: Graduate Standing.

MATH 5544: Mathematical Optimization and Machine Learning (3 credits)
Formulation and analyses of mathematical problems to minimize or maximize functions. Convex, nonlinear, constrained, large-scale, and stochastic optimization problems. Gradient based and higher order optimization methods. Convergence analysis and implementation of optimization strategies. Accuracy and performance trade-offs. Introduction to neural networks, backpropagation, supervised learning, and machine and deep learning strategies, including stochastic approximation and sample average approximation. Regression and classification applications. Prerequisite: Graduate Standing.

MATH 5564: Model Reduction: System-Theoretic Methods (3 credits)
Transfer functions, reachability and observability, Gramians. Balanced truncation. Interpolatory model reduction and optimal interpolation. Data-driven modeling. Parametric Systems. Balanced truncation for nonlinear systems. Interpolatory model reduction for nonlinear systems. Prerequisite: Graduate Standing.

STAT 5054: Introduction to Statistical Computing (3 credits)
Introduction to modern programming packages for data analysis. Basics of coding, language syntax, and statistical functionality to read in raw data files and data sets, subset data, create variables, and recode data. Summaries in the form of tables and graphs. Data analysis using standard statistical methods and data management and analysis of large data sets. Parallel computing. Applied data analysis is emphasized rather than statistical theory. Prerequisite: Graduate Standing.

STAT 5154: Statistical Computing for Data Analytics (3 credits)
Computational techniques for advanced applied statistical analyses and machine learning methods. Project management for larger data projects including computational constraints, pitfalls, and techniques related to different data types. Advanced report generation across different media, efficient R programming, advanced statistical function writing, parallel statistical computing with R, handling missing data, numerical optimization methods, the EM algorithm, and Monte Carlo methods. Prerequisite: STAT 5054.

STAT 5234: Experimental Design for Data Science (3 credits) *
Understanding data, data collection, and proper data analysis for knowledge discovery and decision-making. Randomization, replication, blocking, data quality evaluations (e.g., representativeness of training data), analysis quality assessment (e.g., robustness of the machine learning algorithm to representativeness of training data). Strengths and weaknesses of experimental designs for data science. Modern qualitative and quantitative techniques for constructing experimental designs and analyzing experimental data. Interpretation and reporting of results. Prerequisites: STAT 5615 or STAT 5616 or STAT/CS 5525.

STAT 6554: Advanced Statistical Computing (3 credits)
A second course on statistical and scientific computing. Hands-on, statistical implementation leveraging modern desktop computing (multiple cores), cluster computing (multiple nodes) and distributed computing (hadoop/Amazon EC2) and the coming wave of exascale computing (GPU/TPU/Xeon Phi). Fundamentals of the Unix shell, manipulating data therein, compiling libraries with make, version control (e.g., Git), good habits/best practice with code development and data management. Using advanced R skills to design statistical applications and bind together other languages (e.g., C, C++, Fortran, awk, sed, Cuda, etc.), databases, computing architectures and interfaces to address statistical problems. Prerequisite: STAT 5054.

Appendix C
Faculty Curriculum Vitae (abbreviated)

Statistics
Datta, Jyotishka, Ph.D. in Statistics, 2014, Purdue University, Assistant Professor, Specialization Area: Bayesian methodology and theory, Sparse signal recovery, Global-local shrinkage priors, Default Bayes, Discrete data, High-dimensional data, Epidemiology, Geo-spatial prediction, Bioinformatics, Compositional data, Applied probability, and Bayesian nonparametrics.

Deng, Xinwei, Ph.D. in Industrial and Systems Engineering major concentration Statistics, 2009, Georgia Institute of Technology, Associate Professor, Specialization Area: Interface between experimental design and machine learning, model and analysis of high-dimensional data, covariance matrix estimation and its applications, design and analysis of computer experiments, statistical methods for Nano and emerging technology.

Robertson Evia, Jane, Ph.D. in Education, 2012, University of North Carolina, Collegiate Associate Professor, Specialization Area: Statistics education, Collaborative learning, Student self-efficacy, Program evaluation, STEM Communication.

Feirreira, Marco A. R., Ph.D. in Statistics, Duke University, 2002, Associate Professor, Specialization Area: Highly structured models: multi-scale models, spatial-temporal models, Markov random fields, state-space models, Bayesian Statistics, Computational statistics: Markov chain Monte Carlo, simulated annealing and genetic algorithms, Inverse problems: solution and uncertainty characterization, Applications to environment, medicine, epidemiology.

Gramacy, Robert B., Ph.D. in Applied Mathematics and Statistics, University of California, Santa Cruz, 2005, Professor, Specialization Area: Bayesian modeling methodology, statistical computing, machine learning, Monte Carlo inference, nonparametric regression, sequential design, and optimization under uncertainty. Application areas include spatial data, computer experiments, ecology, epidemiology, finance and public policy.

Guo, Feng, Ph.D. in Statistics and Ph.D. in Transportation Engineering, University of Connecticut, Storrs, 2007, Professor, Specialization Area: quantitative transportation modeling, analyses of naturalistic driving studies, transportation infrastructure safety evaluation, advanced vehicle proactive safety device evaluation, and automated driving systems.

Lucero, Christian L., Ph.D. in Mathematical and Computer Sciences, 2013, Colorado School of Mines, Collegiate Assistant Professor, Specialization Area: Inverse Problems, Uncertainty Quantification, Machine Learning, Computational and Statistical Methodology, Statistical Applications in Scientific Fields.

Higdon, David, Ph.D. in Statistics, 1994, University of Washington, Professor, Specialization Area: space-time modeling, inverse problems in hydrology and imaging, statistical modeling in ecology, environmental science, and biology, multiscale models, parallel processing in posterior exploration, statistical computing, Monte Carlo and simulation-based methods.

Hong, Yili, Ph.D. Statistics, 2009, Iowa State University, Professor, Speciality area: Machine Learning and Engineering Applications, Reliability Analysis, Safety of Artificial Intelligence Systems, Survival Analysis, Longitudinal Data Analysis, Biostatistics, Spatial Data Analysis, Epidemiology, Statistical Computing.

Mahmoud, Hamdy F.F., Ph.D. in Statistics, 2014, Virginia Tech, Collegiate Assistant Professor, Specialization Area: Semi/nonparametric regression models, spatial and spatio-temporal analysis, linear mixed models, and change points detection.

McCarty, Frances, Ph.D. in Research, Management and Statistics, 2001, Georgia State University, Collegiate Assistant Professor, Specialization Area: Survey Data Methodology, Measurement Theory

Van Mullekom, Jennifer, Ph.D. in Statistics, 1998, Virginia Tech, Associate Professor of Practice, Specialization Area: statistical engineering, statistical process control, statistical education, effective consulting practices.

Woteki, Thomas, Ph.D. in Statistics, 1974, Virginia Polytechnic Institute and State University, Professor of Practice, Statistics and Director Academy of Data Science, Specialization Area: statistical computing, data-driven modeling, exploratory data analysis, data analysis application engineering.

Xing, Xin, Ph.D. in Statistics, 2020, Harvard University, Assistant Professor, Specialization Area: minimax nonparametric testing, smoothing spline, dimension reduction, controlled variable selection, neural network, domain adaptation, transfer learning, Computational biology: metagenomics, single cell, epigenomics, neuroimaging.

Zhu, Hingxiao, Ph.D. in Statistics, 2009, Rice University, Associate Professor, Specialization Area: Bayesian Methods, Functional Data Analysis, Statistical Machine Learning, Applications in Medicine, Neuroscience, Engineering, Bioinformatics, and Genetics.

Mathematics

Borggaard, Jeff, Ph.D. in Mathematics, 1995, Virginia Tech, Professor, Specialization Area: Applied and Computational Mathematics.

Childs, Lauren, Ph.D. in Applied Mathematics, 2010, Cornell University, Assistant Professor, Specialization Area: Computational and Mathematical Biology, differential equations, dynamical systems and stochastic analysis.

De Sturler Eric, Ph.D. Applied Mathematics, 1994, Delft University of Technology, Professor, Specialization Area: Fast solvers for large scale linear and nonlinear systems, inverse problems and parameter estimation, optimization, and design, high performance computing.

Embree, Mark, D. Phil. Numerical Analysis, 2000, University of Oxford, Professor, Specialization Area: Numerical analysis, numerical linear algebra, and spectral theory.

Gugercin, Serkan, Ph.D. Electrical Engineering, 2003, Rice University, Professor, Specialization Area: Dynamical systems, numerical analysis and scientific computing.

Iliescu, Traian, Ph.D. Mathematics, 2000, University of Pittsburgh, Professor, Specialization Area: mathematics of turbulent flows.

Liu, Honghu, Ph.D. in Mathematics, 2013, Indiana University Bloomington, Assistant Professor, Specialization Area: Stochastic PDEs, parameterizations, manifolds and approximations, data driven modeling, Geophysical fluid dynamics.

Warburton, Timothy, Ph.D. in Mathematics, 1999, Brown University, Professor, Specialization Area: Finite Element Methods, High Performance Computing, Complete radiation boundary conditions, applications.

Zietsman, Lizette, Ph.D. in Applied Mathematics, 2000, University of Pretoria, Associate Professor, Specialization Area: Numerical analysis, partial differential equations, control and estimation of distributed parameter systems, optimization, applications.

Computer Science

Naren Ramakrishnan, Ph.D. in Computer Science, 1997, Purdue University. Specialization area: Applied machine learning, forecasting, and urban analytics.

Chris North, Ph.D. in Computer Science, 2000, University of Maryland, College Park. Specialization area: Visual analytics, information visualization, and human computer interaction.

Lenwood Heath, Ph.D. in Computer Science, 1985, University of North Carolina, Chapel Hill. Specialization area: Algorithms, Graph Theory, Computational Biology, and Bioinformatics.

Liqing Zhang, Ph.D. in Molecular Evolution/Population Genetics, 2002, University of California, Irvine. Specialization area: Computational Genomics, Bioinformatics.

Anuj Karpatne, Ph.D. in Computer Science, 2017, University of Minnesota, Twin Cities. Specialization area: Theory-Guided Data Science and Machine Learning.

Yalong Yang, Ph.D. in Computer Science, 2019, Monash University. Specialization area: Visualization, Human Computer Interaction, and Immersive Analytics.

Appendix D
McKinsey Survey

In May 2021 the Virginia Economic Development Authority[48] (VEDP) commissioned McKinsey & Company[49] to perform a survey of 14 corporations in the national capital region concerning attracting data science talent. The following are the results of that survey in its entirety as provided to us by the VEDP and presented with their permission.

[48] https://www.vedp.org
[49] https://www.mckinsey.com

Appendix E
Employment Announcements

We are working on employment announcements.

Appendix F
Student Demand Survey

A student demand survey concerning interest in the proposed program was distributed to instructors of both undergraduate service courses and advanced courses required for degree completion in the Departments of Mathematics, Statistics, and Computer Science, and the Academy of Data Science Computational Modeling and Data Analytics program. Instructors were asked to post the survey on the Canvas instructional system management page for their class and/or allow class time to complete the survey. Instructors were encouraged to send a follow-up email to improve response rate for the survey. A video which introduced the students to the program admission requirements and core courses as well as example programs of study accompanied the survey. The survey was opened on February 16, 2022 and closed on February 28, 2022.

A total of 129 responses were received from undergraduate students in these courses. Tables and graphical summaries of the survey questions follow. The full text of the survey is included at the end of the appendix.