New Academic Program - M.S. in Applied Data Science (PENDING)

Data Science is the study of techniques for collecting, processing, and drawing inferences from all manner of data. A multi-disciplinary area, it uses statistics in conjunction with a large repertory of programming tools to extract and organize knowledge from large volumes of structured and unstructured data.

The growing importance of Data Science in technology, industry, health services, administration, and other arenas calls for a well-trained Data Science workforce. The proposed Master’s Degree in Applied Data Science (M.S.A.D.S.) will help to alleviate this need by promoting data literacy across many of ETSU’s disciplines, while enhancing Data Science expertise by providing students with comprehensive and in-depth training. The new degree will enhance students’ mathematical and computational proficiency while providing real-world experiences through internships with local industries, healthcare or administrative units. At same time, an extensive elective track will guide their efforts to identify suitable concentrations within the rich spectrum of data-related disciplines.
Proposed M.S.A.D.S. Curriculum

2.1 Prerequisites
- CSCI 1250, 1260  Introduction to Computer Science 1 and 2. Key competencies: basics of contemporary programming languages, including state change, selection, and iteration; coding style; code modularization (functions, classes); and object-oriented programming (inheritance, polymorphism)
- CSCI 2020  Intro. to Databases. Key competencies: creating, maintaining, and querying relational databases
- MATH 2050  Calculus-Based Probability and statistics. Key competencies: basic probability, mathematical expectation, discrete and continuous probability distributions, sampling distributions, one and two-sample estimation; hypothesis testing; linear regression and correlation.
- Familiarity with Python and R, possibly obtainable through a "boot camp"

ETSU’s Department of Computing and Department of Mathematics & Statistics will create online “boot camp” courses that students who lack these prerequisites can take prior to starting the program, potentially in the summer, or within their first semester of enrollment.

2.2 Program Proper: 39 credits

2.2.1 Core curriculum - 30 credits (Thesis option: 33 credits)
- MATH 5830 - Analytics and Predictive Modeling  3 credits
- STAT 5047 - Mathematical Statistics 1  3 credits
- STAT 5710 - Statistical Methods 1: Linear Models 3 credits
- STAT 5720 - Statistical Methods 2: Generalized Linear Models (prerequisite: STAT 5710) 3 credits
- STAT 5730 - Applied Multivariate Statistical Analysis  3 credits
- CSCI 5260 - Artificial Intelligence  3 credits
- CSCI 5270 - Machine Learning  3 credits
- CSCI 5620 Analysis of Algorithms  3 credits

Culminating Experience: Industrial-Based Project with Industrial partner
- Option I – Internship Experience in Data Science I and II  STAT 5910 and 5920  6 credits
- Option II – Thesis MATH 5960 and Internship Experience in Data Science I STAT 5910 – 9 credits

These could be team projects, with students serving on 2 different teams with different companies that stem across the year.

2.2.2 Focus Areas - 9 credits (Thesis Option: 6 credits)
Possible concentrations for the program would center on choices of electives. The following are six candidates:
- Theory - e.g., three courses from MATH 5257, MATH 5850, STAT 5057, STAT 5217
- Computation-e.g., three courses from CSCI 5260, CSCI 5300, CSCI 5047, CSCI 5000, CSCI 5050
- Sport science-e.g., three courses from PEXS 5270, PEXS 5520, PEXS 5670, SALM 5670
- Health sciences - e.g., three courses from EPID 5100, EPID 5480, BSTA 5380, and MATH 5880
- Business – e.g., three courses from BADM 5140, MSDM 5010, MSDM 5050, and MTKG 5717
- General – three complementary courses from areas of a students’ choosing – most likely, courses like CJCR that are not included in the remaining five focus areas

2.2.3 List of Electives
- AMBA 5140 - Data Analysis and Modeling
- ALHE 5500 - Methods of Research in Allied Health
- BADM 5140 - Data Analysis Modules for Business
- BIOL 5367 Modeling Biological Systems
- BIOL 5500 - Biometry
- BSTA 5310 - Biostatistics I
- BSTA 5350 - Intermediate Biostatistics
- BSTA 5370 - Categorical Data Analysis
- BSTA 5380 - SAS Programming with Statistical Application
- BSTA 5385 - Applied Longitudinal Data Analysis
- CDIS 5400 - Research Methods in Communicative Disorders
• CJCR 5950 - Quantitative Methods in Criminology
• COBH 5250 - Community-Based Methods in Public Health
• COBH 5210 - Adv. Theoretical Models and Survey Development
• COMM 5950 - Quantitative Research Methods in Communication
• CSCI 5260 - Artificial Intelligence
• CSCI 5270 – Machine Learning
• CSCI 5300 - Software Design
• CSCI 5047 - Data Analysis
• CSCI 5000 – Data Management
• CSCI 5050 – Decision Support Systems
• CSCI 5620 – Data Management
• CSCI 5000 – Decision Support Systems
• COMM 5950 - Quantitative Research Methods in Communication
• CSCI 5047 - Data Analysis
• CSCI 5000 – Data Management
• CSCI 5050 – Decision Support Systems
• CSCI 5620 – Analysis of Algorithms
• ECON 5010 - Essentials of Statistics
• EDFN 5950 - Methods of Research
• ELPA 5300 - Professional Needs of Individuals and Groups
• ENTC 5037 - Quality Assurance I
• EPID 5100 - Analytic Methods in Public Health
• EPID 5405 - Intermediate Epidemiology
• EPID 5430 - Epidemiology of Infectious Disease
• EPID 5460 - Environmental Epidemiology
• EPID 5480 - Genetic Epidemiology
• EPID 6410 - Advanced Multivariate Analysis
• EPID 6420 - Applied Epidemiological Analysis
• EPID 6470 - Risk Behavior
• GEOG 5000 - Quantitative Techniques
• GEOG 5217 - Geographic Information Systems
• GEOG 5317 - Advanced Geographic Information Systems
• GEOG 5327 – Advanced Remote Sensing
• GEOG 5807 - Advanced Field Methods in Geography
• GEOS 5010 - Geospatial Analysis
• GEOS 5300 - Topics in Geospatial Analysis
• GEOS 5320 - Geographic Information Systems Projects
• GEOS 5350 - Statistics for Geosciences
• HDAL 5817 - Introduction to Psychological Testing
• HSMP 5040 - Data-Informed Decision-Making in Health Service Organizations
• HSMP 5300 - Quality Improvement in Health Organizations
• HSMP 6310 - Population Health Management
• HSMP 6320 - Applied Health Services Research Methods
• MATH 5257 - Numerical Analysis
• MATH 5267 - Numerical Linear Algebra
• MATH 5810 - Operations Research IMATH 5820 - Operations Research II
• MATH 5850 - Modeling of Infectious Diseases and Social Networks
• MATH 5890 - Stochastic Modeling
• MATH 58010 - Biometry and Biomedical Computing I
• MATH 58020 - Biometry and Biomedical Computing II
• MSDF 5010 – Digital Marketing Research
• MSDF 5050 - Web Analytics
• MSDF 5060 - Business Analytics, Data Visualization and Online Metrics
• PEXS 5270 - Sport Biomechanics
• PEXS 5520 - Instrumentation in Exercise and Sport Science
• PEXS 5670 - Research Design and Analysis
• PHYS 5007 - Computational Physics
• PMGT 5180 - Quantitative Inquiry and Policy Analysis for Public Managers
• PSYC 5210 - Statistical Methods
• PSYC 5410 - Correlation and Multiple Regression
• PSYC 6410 - Structural equation modeling
• PSYC 6220 - Meta-analytic Research Methods
• PSYC 6510 - Topical Seminar
• PUBR 5325 - Brand Insight & Analytics
• SALM 5670 - Sport Management Research
• SOCI 5444 - Applied Data Analysis
• SOCI 5820 - Skills in Applied Sociology and Anthropology
• SOCI 5210 - Sociological Research
• SOCI 5320 - Program Evaluation
• SOCI 5444 - Data Analysis
• STAT 5057 - Mathematical Statistics 2
• STAT 5217 - Statistical Machine Learning
• STAT 5307 - Sampling and Survey Techniques
• STAT 5287 - Applications of Statistics
• STAT BAYxx0 - Bayesian Probability