MDS611 - Introduction to Data Science

Businesses, governments, and individuals create massive collections of data as a by-product of their activity. Increasingly, decision-makers and systems rely on intelligent technology to analyze data systematically in order to improve decision-making. In many cases automating analytical and decision-making processes is necessary because of the volume of data and the speed with which new data are generated. We will examine how data analysis technologies can be used to improve decision-making. We will study the fundamental principles and techniques of data science, and we will examine real-world examples and cases to place data science techniques in context, to develop data-analytic thinking, and to illustrate that proper application is as much an art as it is a science. In addition, we will work hands-on with the Python programming language and its associated data analysis libraries.

MDS612 - Statistical Computing for Data Science

This course provides foundation understanding of classical statistics within the broader context of data science. Topics include exploratory analysis and descriptive statistics, probability theory and foundation of statistical modeling, estimators, hypothesis testing, classical linear regression, inferential statistics, statistical decision theory. Students will become familiar with current trends in the application of computational statistics in a number of application disciplines.

MDS613 - Data Mining

This course will introduce data mining and statistical methods for extracting knowledge from data. The principles and theories of data mining methods will be discussed and will be related to the issues in applying data mining to real world problems. Data Mining studies algorithms and computational paradigms that allow computers to find patterns and regularities in databases, perform prediction and forecasting, and generally improve their performance through interaction with data. It is currently regarded as the key element of a more general process called Knowledge Discovery that deals with extracting useful knowledge from raw data. The knowledge discovery process includes data selection, cleaning, coding, using different statistical and machine learning techniques, and visualization of the generated structures. Important related technologies, as data warehousing and on-line analytical processing (OLAP) will be also discussed. This course will utilize data mining techniques on real time data, and best practices as it examines the topics of data preprocessing, data modeling, and discovering knowledge from the data. In this course, Weka and Rapid Miner tools will be used to mine the data.

MDS614 - Programming for Data Science

This course introduces the programming techniques required for data science to enable the students to implement data science concepts in Python and R Language. Students are introduced to various structures of R such as Objects, Symbols and Environments. Students are exposed to probability distributions, statistical tests and regression models. Students are trained on how to

analyze the data in R for making predictions. Students are introduced to Python environment and how to make use of Python packages for making predictions and for visualizing data.

MDS621 - Research Methodology

This course will provide an overview of the important concepts of research design, data collection, statistical and interpretative analysis. Topics include:

- Engineering research and Approach's
- Literature Review
- Quantitative theory
- Developing a hypothesis, a research problem and related questions
- Framing the problem with the correct research methodology
- Collecting data that accurately addresses the research problem
- Quantitative, Qualitative and Mixed Methods & evaluating feasibility of research proposals.

MDS623 - Data Visualization

This course aims to develop skills on how to design and create data visualizations based on data available and tasks to be achieved. Students will learn the processes of data modeling, data processing, mapping data attributes to graphical attributes, and strategic visual encoding based on known properties of visual perception as well as the task(s) at hand. In addition, students will also learn to evaluate the effectiveness of visualization designs, and think critically about each design decision, such as choice of color and choice of visual encoding.

MDS624 - Advanced Data Analytics

This course aims to develop strong data analytic skills using both theoretical and case based approach to apply data mining and advanced statistical techniques to real world problems facing the society. The students will learn about the use of various multivariate methods, how to design the study to collect data amenable for such analysis, and the issues involved in acquiring, storing, accessing, analyzing and visualizing large, heterogeneous and real-time data associated with diverse real-world domains

MDS632 - Big Data Applications and Analytics

In today's world, enormous amount of data is generated from a variety of sources which include web server logs, Internet click-stream data, social media content, text from customer emails and survey responses, mobile phone call detail records and machine data captured by sensors . It has been predicted by experts that, it may result in a huge tidal wave of data which is referred to as Big Data. The big data has become so ubiquitous that it has become necessary to have

technologies and tools to store, process and analyze it in an efficient way. Big data finds applications in a number of domains including scientific research, government, and industry. Big data analytics is a broad term for the wide range of technologies which help companies in making informed business decisions and analyze large volumes of transactional data, as well as other forms of data that may be untapped by more conventional Business Intelligence(BI) programs. This course provides an exposure to the real-time use cases of big data, introduces to the imperative big data technology, Hadoop. It provides a comparison of Hadoop data processing with other conventional systems. Then we dive deep into the working mechanism of Hadoop focusing upon the Hadoop Distributed File System (HDFS) and Map Reduce.

MDS633 - Information Retrieval and Web Search Engines

This course intends to introduce special topics of current trends in Information Retrieval and Web Search. The course is designed to enable students to study a variety of topics that are new and emerging in the discipline and are not covered in any core or elective courses of the department. Each semester, the contents of the course could cover a new topic or a mix of many topics and could be taught by one or more faculty members

MDS634 - Advanced Data Mining

This course will cover a number of advanced topics in data mining. The students will be familiarized with recent methods and tools that have proven effective in dealing with data sets which often involve uncertain description or other complexities that cause difficulty for the conventional approaches of logistic regression, neural network models and decision trees. This course will discuss various data mining methods with simple examples, review applications, and evaluate relative advantages of several contemporary methods.

MDS635 - Advanced Database Systems

The course gives an overview of motivation and background of the new developments related to databases, and is intended as an introduction to the most important advances in database systems with respect to the classical relational database systems.

MDS641 - Ethics and Data Protection

This course will provide an overview of data security, privacy concerns, and ethics. Data security seeks to protect government, corporate, and individual information. Data security will be addressed as a continuous issue in data management, and the challenge it poses as the uses for data expand. The areas where information technology and privacy meet is also covered. Data ethics is included as a critical topic in terms of privacy, data manipulation, data sharing and ownership, conflict of interest, and communications. This course will utilize case studies, trends,

techniques, and best practices as it examines the topics of data quality, data security, and ethical questions associated with dealing in data.