III) <u>**Course Description:**</u>

11740: Data Engineering - 3 credits

The course starts by examining the modern data ecosystem and how it relates to running a smart and efficient data hub. Then, it shows the student how to perform the principle tasks involved in managing extracting, transforming and loading (ETL) data. This course will explain the data life cycle in a Data science project. In addition, it will cover types of data, such as structured, semi-structured and unstructured and the different formats of data and techniques used in the ETL process. The course also covers the elementary visualization aspects needed to understand the data. It also takes the student through staging, profiling, cleansing, and migrating data.

11771: Computational Statistics - 3 credits

The objectives of this course are to develop an understanding of modern computationally intensive methods for statistical inference, exploratory data analysis. Advanced computational methods for statistics will be introduced, including univariate, multivariate and combinatorial optimization methods and simulation methods. In addition, the course will demonstrate how to apply the above techniques effectively for use on large data sets in practice. Finally, this course will show how to make inferences about populations of interest in data mining problems. In addition to that, other topics that will be covered include: theory of sampling distributions; principles of data reduction; interval and point estimation, sufficient statistics, order statistics, hypothesis testing, correlation and regression.

11745: Data Mining - 3 credits

This course provides a practical and technical introduction to knowledge discovery and data mining. The topics that will be covered include problems of data analysis in databases, discovering patterns in the data, and knowledge interpretation, extraction and visualization. The topics include all data mining and machine learning techniques used for descriptive and predictive analysis. Such as clustering association rules mining, classification, prediction. This course is an absolute necessity for those interested in joining the data science workforce, and for those who need to obtain more experience in data mining.

11741: Big Data - 3 credits

This course provides the data science students with understanding of the Big Data and its role in data analysis. It provides the terminology and the core concepts behind big data problems, applications, and systems. It provides an introduction to one of the most common frameworks, Hadoop and Spark that have made big data analysis easier and more accessible. Also, it will provide you with the necessary skill in manipulating big data distributed over a cluster using functional concepts and in-memory distributed collections framework written in Scala or Spark. We'll cover Spark's programming model in detail, being careful to understand how and when it differs from familiar programming models, like shared-memory parallel collections or sequential collections. Through hands-on examples in Spark and Scala, student learns when important issues related to distribution like latency and network communication should be considered and how they can be addressed effectively for improved performance.

14791: Research Methodology – 1 Credit hours

This one credit hour course will review the major considerations needed in conducting scientific research, particularly in the fields of Computer Science and data science. The topics covered include: Definitions and characteristics of research; Types of research; Topic Selection; Research methodology; Evaluation and validation of research results; writing, publishing, presenting research work; intellectual property and ethics.

11774: Business Modeling for Data Science - 3 Credits hours

This course aims at providing students with essential skills needed to design and develop innovative and consistent business models to increase profits, decrease expenses, minimize risks, or comply with laws and regulations in smart and proactive ways in organizations. The course introduces students to the dimensions of business models along with their elements and relationships. It also enables students in the area of understanding and analyzing business requirements of data-driven projects. The course introduces students to data analytics lifecycle and allowing them to understand the methodological process that business modelers go through while developing big data applications. Finally, the course will provide students with many practical use cases that are related to various sectors.

11711: Algorithms - 3 Credits hours

Review of algorithm design and analysis techniques: asymptotic notation and design techniques. Advanced problems in dynamic programming (edit distance, matrix-chain multiplication and the partition problem).Advanced topics in graph algorithms: all-pairs shortest paths, graph connectivity. Network flow and bipartite matching. String matching and suffix trees. Randomized algorithms. NP-Completeness: complexity classes (P, NP, NP-complete, NP-hard), NPcompleteness reductions, dealing with NP-complete problems (approximation algorithms,

branchand-bound, integer linear programming).Selected advanced topics, such as number theoretic algorithms, computational geometry and parallel algorithms.

11743: Data Exploration and Visualization – 3 Credit hours

This course covers the essential exploratory techniques for summarizing data. These techniques are typically applied before formal modeling commences and can help to inform the development of more complex statistical models. Exploratory techniques are also important for eliminating or sharpening potential hypotheses about the world that can be addressed by the data. We will cover in detail the plotting systems using various tools as well as some of the basic principles of constructing data graphics. We will also cover some of the common multivariate statistical techniques used to visualize high-dimensional data. The course also aims to facilitate the data analytics process through Information Visualization. The challenge for Visual Analytics is to design and implement "effective Visualization methods that produce pictorial representation of complex data so that data analysts from various fields (bioinformatics, social network, software visualization and network) can visually inspect complex data and carry out critical decision making.

11745: Mining Massive Datasets – 3 Credit hours

Pivotal issues pertaining to mining massive data sets will range from how to deal with huge document databases and infinite streams of data to mining large social networks and web graphs. This course will discuss data mining and machine learning algorithms for analyzing very large amounts of data. The emphasis will be on Map Reduce/Spark as a tool to implement parallel algorithms such as PageRank, EdgeRank and graph centrality that can process very large amounts of data. Hands on experience will be obtained through case studies that demonstrate how big data problems and their solutions allow organizations to succeed in the market.

11725: Cloud Computing – 3 Credit hours

This course provides a hands-on experience and study of Cloud concepts and capabilities across the various Cloud service models including Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), and Business Process as a Service (BPaaS). Mainstream Cloud infrastructure services and related vendor solutions are also covered in detail. PaaS topics cover a broad range of Cloud vendor platforms including AWS, Google App Engine, Microsoft Azure The SaaS and PaaS topics covered in the course will familiarize students with the use of vendor maintained applications and processes available on the Cloud on a metered on-demand basis in multi-tenant environments. Through hands-on assignments and projects, students will learn how to configure and program IaaS services. They will also learn how to develop Cloud-based software applications on top of various Cloud platforms, how to integrate application-level services built on heterogeneous Cloud platforms, and how to leverage SaaS and BPaaS solutions to build comprehensive end-to-end business solutions on the Cloud.

11747: Business Intelligence - 3 Credits hours

This course is intended to provide an integrative foundation in the field of Business Intelligence (BI) at the operational, tactical, and strategic levels. Topics such as value chain, customer service

management, business process analysis and design, transaction processing systems, management information systems, and executive information systems will be covered, along with other topics relevant to the field of business intelligence. Students are exposed to the latest applications and theories that add value to any organization through data, information, knowledge, processes, and communications technologies. In this course, students will be familiar with the basic and current technologies together with advanced concepts, applications, and competitive strategies in the context of enterprise business intelligence supported by practical examples. The course will explain what business intelligence can offer to organizations and demonstrate how business intelligence is used in the real world; and finally provide an action plan for identifying and acting on the BI opportunities that exist in an organization.

11757: Natural Language Processing - 3 Credits hours

This course covers the fundamental concepts and ideas of natural language processing (NLP). It develops an in-depth understanding of both the algorithms available for the processing of linguistic information and the underlying computational properties of natural languages with focus on Arabic language. Word level, syntactic, and semantic processing from both a linguistic and an algorithmic perspective are considered. The focus is on modern quantitative techniques in NLP: using large corpora, statistical models for acquisition, disambiguation, and parsing. The main NLP applications will be presented: Information Extraction, Question Answering, Summarization, Dialogue and Conversational Agents, and Machine Translation.

11789: IT Projects Management for Data Analysis - 3 Credits hours

The main goal of this course is to gain a clear understanding of the five IT Project Management Process Groups (Initiating, Planning, Executing, Monitoring and controlling, and Closing) and learn how these processes interact with each other to successfully achieve project objectives. Discover how to integrate the ten Knowledge Area processes, tools and templates in the work place. Concepts include stakeholders, scope, quality, time, cost, human resources, communication, risk, procurement and project integration management. Students will also apply techniques such as stakeholder analysis, work breakdown structure, scheduling, estimating, risk assessments, contracts and change control. Students will have opportunity to apply project management principles to real-world situations.

11753: Artificial Intelligence - 3 Credits hours

The course is divided into four parts, the first covering knowledge representation, the second introducing heuristic search and constraint satisfaction and the third is dedicated to advanced topics such as rule-based Expert Systems, case-based reasoning, model-based reasoning. The fourth part is dedicated to machine learning techniques and theory. The following topics will be discussed in the course: introduction to AI and applications; exhaustive search methods; heuristic search methods; First order logic for knowledge representation; other knowledge representation schemes such as semantic networks, frames; production rule systems; principles of expert systems; Knowledge acquisition, planning and scheduling, machine learning techniques: decision trees, neural networks, Instance-Based-learning, Naïve Bayesian learning, Bayesian networks and learning theory.

11742: Thesis - 9 Credits hours

11739 Selected Topics in Data Science - 3 Credits hours

Topics are selected from different areas in Data Science that are not covered in the description of the courses listed in the curriculum. This course will cover recent trends and issues in the field of data science and will be chosen at the discretion of the instructor. A sampling of just a few of the subjects covered in this course are: NoSQL, Deep Learning. Students are assigned individual projects in specific fields. Project reports and seminars will be required in order for the students to demonstrate their ability in research and oral presentations. Projects are discussed in groups in order involve the whole class in these subjects.

11748 Web and Social Network Analysis - 3 Credits hours

The course covers concepts and techniques for retrieving, exploring, visualizing, and analyzing social network and social media data, website usage, and clickstream data. Students learn to use key metrics to assess goals and return on investment, perform social network analysis to identify important social actors, subgroups, and network properties in social media.

11747 Capstone Project - 3 Credits hours

The capstone project will focus on tackling a data science problem sourced from science, government or industry. Capstone projects should be development oriented.