اسم التخصص	الذكاء الاصطناعي وتحليل البيانات Artificial Intelligence and Data Analytics (AIDA)
الدرجة العلمية للتخصص	ماجستير .M.Sc
مدة الدراسة	سنتان
الكلية	الدراسات العليا
مسار الشامل	

A. Core Requirements (21 CHs)

COURSE CODE	COURSE NAME	CHS	PRE-REQUISITE	ILO Number
700101	**STATISTICAL COMPUTING**	**3**		1, 2, 3, 7, 8
700103	**DATA PREPROCESSING**	**3**		1, 4, 6
700105	**INTRODUCTION TO MACHINE LEARNING**	**3**		1, 4
700201	**BIG DATA ANALYTICS**	**3**		1, 3, 6, 8
700203	**ADVANCED MACHINE LEARNING**	**3**		1, 3, 5, 6, 8
700205	**FUNDAMENTALS OF DEEP LEARNING**	**3**		1, 3, 4, 7, 8
700107	**ARTIFICIAL INTELLIGENCE**	**3**		1, 2

A. Comprehensive Exam Requirements (15 CH)

Course Code	Course Name	CHs	Pre-requisite	ILO Number
700299	project	3		1-8
Elective Courses		12		
700199	Comprehensive Exam	-	Completing 36 CH with an Accumulative Average ≥ 75%	All Core and Taken Elective Courses

B. Elective Requirements (12 CH)

Course Code	Course Name	CHs	Pre-requisite	ILO Number
700207	Special Topics in Data Analytics	3		1-8
700123	Advanced Algorithms	3		1, 8
700109	Python Programming for Analytics	3		1, 2, 3, 5
700111	Regression Analysis and R Programming	3		1, 3, 8
700209	Neural Nets and Regression Predictive Analytics	3		1, 3, 8
700211	Dimension Reduction, Clustering and Association Rules Predictive Analytics	3		1, 3, 8
700212	Business Intelligence	3		1, 3, 5, 8
700214	SAS Programming	3		1, 8
700113	Natural Language Processing	3		1, 2, 8
700115	Cloud Computing	3		1, 8
700213	Web and Social Networks Analysis	3		1, 3, 6, 8
700215	Internet of Things	3		1-8

700217	Spatial Statistics with Geographic Information Systems	3		1, 2, 5, 8
700219	Data Exploration and visualization	3		
700221	Mining Massive Datasets	3		
700121	Selected Topics in Data Science	3		1-8

C. Courses Vs. ILOs: The following table shows the various courses against the ILOs:

Courses	Intended Learning Outcomes							
	1	2	3	4	5	6	7	8
Statistical Computing	✓	✓	✓				✓	✓
Data Preprocessing	✓			✓		✓		
Introduction to Machine Learning	✓			✓				
Big Data Analytics	✓		✓					✓
Advanced Machine Learning					✓			
Fundamentals of Deep learning	✓			✓			✓	✓
Artificial Intelligence	✓	✓						
Special Topics in Data Analytics	✓	✓	✓	✓	✓	✓	✓	✓
Python Programming for Analytics	✓	✓	✓			✓		
Regression Analysis and R Programming	✓		✓					✓
Neural Nets and Regression Predictive Analytics	✓		✓					✓
Dimension Reduction, Clustering and Association Rules Predictive Analytics	✓		✓					✓
SAS Programming	✓							✓
Natural Language Processing	✓	✓						✓
Cloud Computing	✓							✓
Web and Social Networks Analysis	✓		✓			✓		✓
Internet of Things	✓	✓	✓	✓	✓	✓	✓	✓
Spatial Statistics with Geographic Information Systems	✓	✓			✓			✓
Selected Topics in Data Science	✓	✓	✓	✓	✓	✓	✓	✓
Advanced Algorithms	✓							✓
Project	✓	✓	✓	✓	✓	✓	✓	✓
Thesis	✓	✓	✓	✓	✓	✓	✓	✓

Study plan (Comprehensive Exam track)

First Semester (9 CHs)			
Course Code	**Course Name**	**CHs**	**Pre-requisites**
700101	Statistical Computing	3	
700103	Data Preprocessing	3	
700203	Introduction to Machine Learning	3	

Second Semester (9 CHs)			
Course Code	**Course Name**	**CHs**	**Pre-requisites**
700201	Big Data Analytics	3	
700203	Advanced Machine Learning	3	
700107	Artificial Intelligent	3	

Third Semester (9 CHs)			
Course Code	**Course Name**	**CHs**	**Pre-requisites**
700207	Special Topics in Data Analytics	3	700201
	Elective course	3	
700205	Fundamentals of Deep Learning	3	700203 & 700107

Comprehensive Exam Fourth Semester (9 CHs)			
Course Code	**Course Name**	**CHs**	**Pre-requisites**
700299	Project	3	
700199	Comprehensive Exam	-	
	Any Two Electives	6	

Courses Description:

Data Preprocessing: The course will present the Needed Data pre-processing, Attributes and Data types, Statistical descriptions of Data, Handling missing Data, Data sampling, Data cleaning, Data Integration and transformation, Data reduction, Discretization and generating concept hierarchies. Python will be used as a medium of coding.

Statistical Computing: This course will introduce students to Data and Statistics, Descriptive Statistics: Tabular and Graphical Presentations, Descriptive Statistics: Numerical Measures, Probability, Discrete Probability Distributions, Continuous Probability Distribution, Sampling and Sampling Distributions, Interval Estimation, Fundamentals of Hypothesis Testing, Two-Sample Tests, Inferences about Population Variances, Tests of Goodness of Fit and Independence, Experimental Design and ANOVA, and Simple Linear Regression.

Introduction to Machine Learning: The course aims at introducing the basic concepts and techniques of Machine Learning, developing the skills in using recent machine learning software for solving practical problems as well as being familiar with a set of well-known supervised, semi-supervised and unsupervised learning algorithms. Thus, the course covers topics including: Supervised, Unsupervised and Reinforcement Learning; visualization of algebraic concepts Linear Regression; Gradient descent algorithm, cost function to find 'beta' values and concept Gradient Descent and how to represent matrix of problem, how to use Gradient descent for multiple features and scaling techniques, in addition to the types of feature scaling, finding coefficients analytically. Furthermore, Logistic Regression will be covered such as the Sigmoid function model and its graphical representation. The Receiver-operating characteristic (RoC) curve and its use; Optimization and Classifications including the Optimization objective from logistic regression, the large margin classifier such as SVM.

Big Data Analytics: This course will introduce students to big data. This introduction includes, distributed file system, Big Data and its importance, Drivers, Big data analytics, Big data applications. Algorithms, Matrix-Vector, Multiplication by Map Reduce. In addition, the course will introduce students to HADOOP: Big Data, Apache Hadoop & Hadoop Ecosystem, MapReduce, Data Serialization as well as HADOOP Architecture: Architecture, Storage, Task trackers, Hadoop Configuration. Furthermore, the course will present the HADOOP ecosystem and yarn: Hadoop ecosystem components, Hadoop 2.0 New Features NameNode High Availability, HDFS Federation, MRv2, YARN, Running MRv1 in YARN**.**

Advanced Machine Learning: The course covers three major approaches. In particular, the Decision trees and random forests: Concept, diagrammatic representation, random forest as a voting committee of decision trees, parameter meaning and explanation; the Naive Bayes: Venn diagrams, Naive Bayes algorithm, application and problems, Naive Bayes learning, Bayesian inference, Retail basket analysis; Concept of boosting and bagging; the Unsupervised learning methods/Clustering: K-means algorithm, optimization objective, graphical representation, random initialization, choosing number of clusters; as well as the Association Rules: Association rule mining, K-nearest neighbours' algorithm.

Artificial Intelligent: This course introduces students to artificial intelligence and its trends and applications, exhausted research methods, diligent research methods, first-class logic for representing knowledge, other methods for representing knowledge such as semantic networks, frameworks, conclusion rules, principles of expert systems, knowledge extraction, planning and scheduling, techniques Machine learning, decision trees, neural networks, similar learning, simple biz learning, biz networks, learning theory.

Fundamentals of Deep learning: The course introduces students to deep learning and its trends and applications. The course will present various topics including: Neural Networks, cost functions, hypotheses and tasks; training data; maximum likelihood based cost, cross entropy, MSE cost; feed-

forward networks; MLP, sigmoid units; neuroscience inspiration; Learning in neural networks output vs hidden layers; linear vs nonlinear networks; In addition, to the backpropagation learning via gradient descent; recursive chain rule (backpropagation). Furthermore, the course will present the Deep learning strategies like: GPU training, regularization, RLUs, dropout, etc, CNNs I Convolutional neural networks, Deep Belief Nets **(**probabilistic methods RNNs I Recurrent neural networks Other DNN variants), Neural Turing Machines, Unsupervised deep learning like autoencoders, deep generative models., and Deep reinforcement learning Vision and NLP applications.

Advanced Algorithms: The course reviews the latest methods for analysing and designing algorithms and covers topics such as: Advanced issues in dynamic programming, advanced issues in graphical algorithms, network flow algorithms, text matching algorithms, random algorithms, Fully Difficult Problems: Classification of Problems by Degree of Complexity, Proofs of Proof of Difficulty, Addressing Fully Difficult Problems (Approximate Solutions, Full-Branching Solutions, Linear Programming). Furthermore, Selected Topics will be addressed like number theory, computational engineering, and parallel algorithms.

Cloud Computing: This course introduces the Cloud and its Services, types of services, and service Providers: Google, Amazon, Microsoft Azure, Ibm, Sales Force; and Collaborating Using Cloud Services: Email Communication Over The Cloud - CRM Management - Project Management-Event, Management - Task Management – Calendar - Schedules - Word Processing – Presentation, Spreadsheet - Databases – Desktop - Social Networks And Groupware; Virtualization For Cloud: Need for Virtualization, Pros and Cons of Virtualization, types of Virtualization System, VM, Process VM, Virtual Machine Monitor, Virtual Machine Properties, Interpretation and Binary Translation, HLL VM Hypervisors – Xen, KNM, VMWare, Virtual Box, Hyper-V. The course will also preset Security standards and ts applications including: cloud security challenges, software as a service Security, common standards, the Open Cloud Consortium, the distributed management task force, standards for application developers, standards for messaging, standards for security, end user access to cloud computing, mobile internet devices and the cloud.

Regression Analysis and R Programming: This course reviews Regression and addresses estimates relationships between independent (predictor or explanatory) variables and a dependent (response or outcome) variable. Regression models can be used to help understand and explain relationships among variables; they can also be used to predict actual outcomes. In this course, students will learn how multiple linear regression models are derived, use software to implement them, learn what assumptions underlie the models, learn how to test whether available data meet those assumptions and what can be done when those assumptions are not met, and develop strategies for building and understanding useful models. The various covered Topics include: Calculate a simple linear regression model; Assess the model with standard error, R-squared, and slope; Review and check model assumptions; Extend the model to multiple linear regression; Assess parameter estimates globally, in subsets, and individually; Test model assumptions; Transform predictors and response variables to improve model fit; Deal with qualitative predictors; Handle interactions among predictors; Identify influential points; Deal with autocorrelation, multi-collinearity, and missing data; and Exercise appropriate caution with respect to extrapolation. In addition, this course introduces basic concepts of R Programming: File formats and R syntax, text editors to write code, reading in files, use symbols and assignments, and iterate simple loops, discussion of data structures and sub setting and apply functions.

Python Programming for Analytics: This course introduces the Python language for data analysis. The course starts with basic Python skills and data structures, how to load data from different sources, rearrange and aggregate it, and finally how to analyse and visualize it to create high-quality products. Students will be able to: Construct conditional statements and loops; Work with strings, lists and dictionaries (in addition to variables); Read and write data; Use Pandas for data analysis; Group,

aggregate, merge and join; Handle time series and data frames; Use matplotlib for visualization and Create format, and output figures.

Neural Nets and Regression Predictive Analytics: The course will continue work from 700105, and introduce additional techniques in predictive analytics (modelling) as other forms of data mining. Following the review of data mining topics, the course introduces four topics. The Linear and Logistic Regression by addressing for descriptive modelling including Odds and logit, Fitting the model and Interpreting output; and using Logistic regression for classification. The next topic will be the Discriminant Analysis for classification defining and using statistical (Mahalanobis) distance, linear classification functions, and Generating classifications. Basics of the Neural Network structure (Input, Hidden, and Output layers), the Supervised learning (Back propagation and iterative learning) will be covered. The fourth topic is the Text Mining that include: tabulation of text; Term-document matrix; Bag of words; Preprocessing of text (Tokenization, Text reduction, and Term Frequency - Inverse Document Frequency (TF-IDF)); as well as Fitting a predictive model. Furthermore, Additional Topics will be addressed like: Multiclass classification, Network analytics, and Text analytics. Hands-on work will be included on the software like XLMiner (a Data Mining add in for Excel, and the Matlab.

Dimension Reduction, Clustering and Association Rules Predictive Analytics: The course covers more predictive analytics techniques including, Dimension Reduction, and unsupervised learning approaches like Clustering, and Association Rules. Furthermore, an integration of supervised and unsupervised learning techniques will be covered. The Dimension Reduction part includes: Detecting information overlap using domain knowledge and data summaries and charts; Removing or combining redundant variables and categories; Dealing with multi-category variables; Automated dimension reduction techniques (Principal Components Analysis (PCA. The Cluster Analysis part includes: Clustering approaches and potential applications; Hierarchical Clustering; and K-Means Clustering. The Association Rules part includes: Discovering association rules in transaction databases; and Item-based, and Person-based Collaborative filtering. Furthermore, Integrating Supervised and Unsupervised Methods will be covered and addressed in the context of network and text analytics. Hands-on work will be included on the software like XLMiner, and the Matlab.

Selected Topics in Data Science: This course explores the latest trends and emerging technologies based on global forecasting consultancies' Firms and focuses on the hottest and emerging predicted topic(s) so as to prepare the university, community and enrolled students for the upcoming Technology and trends of the future in the field of data science and analytics.

SAS Programming: This course covers the programming in Statistical Analysis System (SAS) and using SAS tools to manage data, analyze data, and create data reports. The course will focus on basic data manipulation, including reading raw data into SAS, using formats and in formats, functions, conditional processing, and sub-setting and joining data sets. In addition, the course will cover the basic reporting procedures, including PROC Print, PROC Freq, PROC Means, PROC Tabulate, and PROC Report, as well as the Output Delivery System (ODS). Furthermore, the course will present advanced topics for intermediate SAS users, such as SQL, SAS/GRAPH, Macros, and Arrays.

Internet of Things: This course introduces the emerging topic of the Internet of Things (IoT) and covers the opportunities of IoT and shows the transformation of the Internet to the IoT and describing the 4 components (People, Data, Process ad Things) of IoT: Connecting People in more relevant valuable ways; Leveraging data into more useful information for decision making; Delivering the right information to the right person (or machine) at the right time; and the physical devices and objects connected to the Internet and each other for the intelligent decision making. Following a brief introduction of the IoT, the course describes the technologies used with particular emphasis of what makes a product or system a Thing in the IoT world. Then, the course covers the IoT Platform and IoT Application Technologies with particular emphasis of what makes a product or system a Thing in the

IoT world. Furthermore, the course describes the four components that make up IoT from a higher level discussing functional components as well as major suppliers in each area as well as some key challenges for creating IoT solutions.

Spatial Statistics with Geographic Information Systems: The course addresses problems in which spatial location is the most important explanatory variable. The course introduces essential background in the geospatial concepts. The course introduces basics of geographical data, statistics and describes spatial data using maps. Also, the course presents the analysis of Patterns in Point Data by covering the introductory methods for detecting non-randomness in dot/pin map distributions. Furthermore, the course covers the Analysis of Patterns in Area Data by detecting and measuring spatial autocorrelation in lattice data and then moves forward to introduce the Analysis of Continuous Field Data by creating contour-type maps using inverse distance weighting and geo-statistical methods. Examples and hands-on work of the analysis that can be conducted in a geographic information system such as ArcGIS or Mapinfo.

Special Topics: This course is taken individually for every student where students are expected to choose a faculty member and select a Artificial Intelligence & Big Data Analytics - related topic of interest. The major outcome of this course is a thesis proposal. Students will explore their topic of interest, conduct a literature survey, define a problem, and propose a solution in the Artificial Intelligence & Big Data Analytics context. Students will be assessed by two presentations in front of a jury of faculty members.

Business Intelligence: This course focuses on business intelligence – an information technology approach to data collection and data analysis to support a wide variety of management tasks, from performance evaluation to trend spotting and policy making. Students learn analytical components and technologies used to create dashboards and scorecards, data/text/Web mining methods for trend and sentiment analysis, and artificial intelligence techniques used to develop intelligent systems for decision support. Students will actively participate in this course through class discussions, project preparation and presentation, and visual tool utilization.

Natural Language Processing: The course intends to present an introduction to Natural Language Processing (NLP), the study of computing systems that can process, understand, or communicate in human language. The primary focus of the course will be on understanding various NLP algorithms for effectively solving problems, and methods for evaluating their performance. There will be a focus on statistical and neural-network learning algorithms that train on (annotated) text corpora to automatically acquire the knowledge needed to perform the task.

Web and Social Networks Analysis: This course covers both quantitative and qualitative methods for describing, measuring and analyzing social data networks. The course also will explore how to identify influential individuals, track the spread of information through networks, and see how to use these techniques on real problems. The various presented topics include: Network Analysis Basics: Basic Terminology, Metrics, and Visualization; the Social Network: Tie strength and Trust - User attributes and behavior; Analytics: Modeling, Sampling, Content Analysis; and Propagation Applications: Location, Filtering and recommender systems, and Business use. Furthermore, the course will see how AI, network analysis, and statistical methods can be used to study these topics.

Data Exploration and Visualization: This course is all about data visualization, the art and science of turning data into readable graphics. The course will also explore how to design and create data visualizations based on data available and tasks to be achieved. This process includes data modeling, data processing (such as aggregation and filtering), 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. 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. Students will create their own data visualizations, and learn to use Open Source data visualization tools, especially D3.js. Students will also read papers from the current and past visualization literature and create video presentations of their findings.

Mining Massive Datasets: This course introduces the students to modern distributed file systems and MapReduce, including what distinguishes good MapReduce algorithms from good algorithms in general. The rest of the course is devoted to algorithms for extracting models and information from large datasets. Students will learn how Google's PageRank algorithm models importance of Web pages and some of the many extensions that have been used for a variety of purposes. The course will also cover locality-sensitive hashing, a bit of magic that allows you to find similar items in a set of items so large you cannot possibly compare each pair. When data is stored as a very large, sparse matrix, dimensionality reduction is often a good way to model the