Name of the Programme: M.Sc. in Artificial Intelligence
Course Code: CSI-529
Title of the Course: Essentials of Data Analytics
Number of Credits: 4(2L+2T+0P)
Effective from AY: 2023-24

Prerequisites for the course	Probability and Statistics	
Objectives	1. To understand the concepts of analytics using various machine learning models. 2. To appreciate supervised and unsupervised learning for predictive analysis. 3. To understand data analytics as the next wave for businesses looking for competitive advantage. 4. Carry out rule-based analysis of the data in line with the analysis plan. 5. Validate the results of their analysis according to statistical guidelines. 6. Validate and review data accurately and identify anomalies. 7. To learn aspects of computational learning theory. 8. Apply statistical models to perform Regression Analysis, Clustering and Classification.	
Theory:	**Module:1** Regression Analysis Linear regression: simple linear regression - Regression Modelling - Correlation, ANOVA, Forecasting, Autocorrelation	5 hours
	Module:2 Classification Logistic Regression, Decision Trees, Naïve Bayes-conditional probability - Random Forest - SVM Classifier	5 hours
	Module:3 Clustering K-means, K-medoids, Hierarchical clustering	5 hours
	Module:4 Optimization Gradient descent - Variants of gradient descent - Momentum - Adagrad - RMSprop - Adam - AMSGrad	5 hours
	Module:5 case study -Managing Health and Safety Comply with organization's current health, safety and security policies and procedures - Report any identified breaches in health, safety, and security policies and procedures to the designated person - Identify and correct any hazards that they can deal with safely, competently and within the limits of their authority - Report any hazards that they are not competent to deal with to the relevant person in line with organizational procedures and warn other people who may be affected.	5 hours
	Module:6- requirement analysis - Data and Information Management Establish and agree with appropriate people the data/information they need to provide, the formats in which they need to provide it, and when they need to provide it - Obtain the data/information from reliable sources - Check that the data/information is accurate, complete and up-to-date	3 hours
	Module:7 Learning and Self Development Obtain advice and guidance from appropriate people to develop their knowledge, skills and competence - Identify accurately the knowledge and skills they need for their job role - Identify accurately their current level of knowledge, skills and competence and any learning and development needs - Agree with appropriate people a plan of learning and development activities to address their learning needs	2 hours

Practicals to be discussed during Tutorial Slots:	1. Web Scraping:- a. While you'll find no shortage of excellent (and free) public data sets on the internet, you might want to show prospective employers that you're able to find and scrape your own data as well. Plus, knowing how to scrape web data means you can find and use data sets that match your interests, regardless of whether or not they've already been compiled. b. If you know some Python, you can use tools like Beautiful Soup or Scrapy to crawl the web for interesting data. If you don't know how to code, don't worry. You'll also find several tools that automate the process (many offer a free trial), like Octoparse or ParseHub. c. If you're unsure where to start, here are some websites with interesting data options to inspire your project: d. Reddit, Wikipedia, Job portals 2. Data Cleaning a. A significant part of your role as a data analyst is cleaning data to make it ready to analyze. Data cleaning (also called data scrubbing) is the process of removing incorrect and duplicate data, managing any holes in the data, and making sure the formatting of data is consistent. b. As you look for a data set to practice cleaning, look for one that includes multiple files gathered from multiple sources without much curation. Some sites where you can find "dirty" data sets to work with include: c. CDC Wonder, Data.gov, World Bank, Data.world/r/datasets 3. Exploratory data analysis (EDA) a. Data analysis is all about answering questions with data. Exploratory data analysis, or EDA for short, helps you explore what questions to ask. This could be done separate from or in conjunction with data cleaning. Either way, you'll want to accomplish the following during these early investigations. b. Ask lots of questions about the data. c. Discover the underlying structure of the data. d. Look for trends, patterns, and anomalies in the data. e. Test hypotheses and validate assumptions about the data. f. Think about what problems you could potentially solve with the data. 4. Sentiment analysis a. Sentiment analysis, typically performed on textual data, is a technique in natural language processing (NLP) for determining whether data is neutral, positive, or negative. It may also be used to detect a particular emotion based on a list of words and their corresponding emotions (known as a lexicon). b. This type of analysis works well with public review sites and social media platforms, where people are likely to offer public opinions on various subjects. c. To get started exploring what people feel about a certain topic, you can start with sites like: Amazon (product reviews), Rotten Tomato (movie reviews), Facebook witter, News sites 5. Data visualization a. Humans are visual creatures. This makes data visualization a powerful tool for transforming data into a compelling story to	5 * 6 = 30 hours

	encourage action. Great visualizations are not only fun to create, they also have the power to make your portfolio look beautiful.	
Pedagogy	Lectures/Assignments/Seminar Presentations/Mini-Project	
References/ Readings	1.Cathy O'Neil and Rachel Schutt. "Doing Data Science, Straight talk from the Frontline",O'Reilly. 2014. 2.Dan Toomey, "R for Data Science", Packt Publishing, 2014. 3.Trevor Hastie, Robert Tibshirani and Jerome Friedman. "Elements of Statistical Learning",Springer , Second Edition. 2009. 4.Kevin P. Murphy. "Machine Learning: A Probabilistic Perspective", MIT Press; 1st Edition, 2012. **Reference Books** Glenn J. Myatt, "Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining", John Wiley & Sons, Second Edition, 2014. G. K. Gupta, —Introduction to Data Mining with Case Studies", Easter Economy Edition, Prentice Hall of India, 2006. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007. Colleen Mccue, "Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis", Elsevier, 2007. R N Prasad, Seema Acharya, "Fundamentals of Business Analytics", Wiley; Second edition, 2016. https://www.sscnasscom.com/qualification-pack/SSC/Q2101/ Mode of Evaluation: ISA/Assignment / Quiz / Project / Seminar	
Course Outcomes	1. Identify and apply the appropriate supervised learning techniques to solve real world problems with labeled data. 2. Choose and implement typical unsupervised algorithms for different types of applications with unlabelled data. 3. Implement statistical analysis techniques for solving practical problems. 4. Understand different techniques to optimize the learning algorithms.	

(Back to top)