Course Code	Advanced Data Analytics	Course Type: LTP
NAS2001	Job Role: SSC/Q2101	**Credits :4**
		v.1.0

Course 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

Expected Course Outcome:

1. Use a tool to implement typical clustering algorithms for different types of applications
2. Identify applications suitable for different types of machine learning with suitable
3. justification
4. Apply machine learning techniques to text classification and clustering which is used for efficient Information Retrieval
5. Implement statistical analysis techniques for solving practical problems.
6. Ability to apply and implement learned algorithm design techniques and models to solve problems.

Student Learning Outcomes (SLO)	1,2,4, 12, 14

[1] Having an ability to apply mathematics and science in engineering applications
[2] Having a clear understanding of the subject related concepts and of contemporary issues
[4] Having Sense-Making Skills of creating unique insights in what is being seen or observed
[12] Having adaptive thinking and adaptability
[14]Having an ability to design and conduct experiments, as well as to analyze and interpret data

Module:1	**Correlation**	**6 hours**

Basic statistics: mean, median, standard deviation, variance, correlation, covariance

Module:2	**Regression Analysis**	**6 hours**

Linear regression: simple linear regression - Regression Modelling - Correlation, ANOVA, Forecasting, Autocorrelation

Module:3	**Classification**	**6 hours**

Logistic Regression, Decision Trees, Naïve Bayes-conditional probability

Module:4	**Clustering**	**4 hours**

K-means, K-medoids, Hierarchical clustering

Module:5	Managing Health and Safety	4 hours

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.

Module:6	Data and Information Management	4 hours

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

Module:7	Data and Information Management	3 hours

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

	Total Lecture hours:	30 hours

Text Book(s)

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

1. Glenn J. Myatt, "Making Sense of Data : A Practical Guide to Exploratory Data Analysis and Data Mining", John Wiley & Sons, Second Edition, 2014.
2. G. K. Gupta, ―Introduction to Data Mining with Case Studies", Easter Economy Edition, Prentice Hall of India, 2006.
3. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.
4. Colleen Mccue, "Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis", Elsevier, 2007.
5. R N Prasad, Seema Acharya, "Fundamentals of Business Analytics", Wiley; Second edition, 2016.
6. https://www.sscnasscom.com/qualification-pack/SSC/Q2101/

List of Experiments (Indicative)		**SLO: 1,2,4, 12, 14**
1.	Visualizing data using R with different type of graphs and charts	
2.	Linear regression analysis	
3.	Forecasting - weather dataset using R	
4.	R- Hadoop Integration for Analytics	
5.	Text Analytics – Sentiment Analysis using R, Word cloud analysis using R	
6.	Time Series Components(Trend, Seasonality, Cyclicity and Level)	
7.	Banking Sector: Understand customer spend & repayment behavior, along with evaluating areas of bankruptcy, fraud, and collections. Also, respond to customer requests for help with proactive offers and service.	
8.	Retail Case Study: A retail store requires analyzing the day-to-day transactions and keeping a track of its customers spread across various locations and their purchases/returns across various categories. The objective of the case study is to understand customer behavior in-terms of purchase and returns through various Data Manipulation steps in R.	
9	Movie Recommendation System: To understand the functioning of how a recommendation system works. Develop an Item Based Collaborative Filter using Netflix dataset	
	Total Laboratory Hours	**30 hours**
Recommended by Board of Studies		
Approved by Academic Council		Date	