AC: 02.06.2025 ITEM NO: 27.1

Deccan Education Society's

Kirti M. Doongursee College of Arts, Science and Commerce (AUTONOMOUS)

Affiliated to

UNIVERSITY OF MUMBAI

Syllabus for
Program: Masters of Science
Course: M.Sc Part-I
Subject: Data Science

Choice Based Credit System (CBCS)
with effect from
Academic Year 2025-2026

PROGRAM OUTCOMES

PO	Description
A student completing Master's Degree in Science Program will be able to	
PO1	Demonstrate an in-depth understanding of the relevant science discipline. Recall, explain, extrapolate, and organize conceptual scientific knowledge for execution and application and also to evaluate its relevance
PO2	Critically evaluate, analyze, and comprehend a scientific problem. Think creatively, experiment and generate a solution independently, check and validate it and modify if necessary.
PO3	Access, evaluate, understand, and compare digital information from various sources and apply it for scientific knowledge acquisition as well as scientific data analysis and presentation.
PO4	Articulate scientific ideas, put forth a hypothesis, design and execute testing tools and draw relevant inferences. Communicate the research work in appropriate scientific language.
PO5	Demonstrate initiative, competence, and tenacity at the workplace. Successfully plan and execute tasks independently as well as with team members. Effectively communicate and present complex information accurately and appropriately to different groups
PO6	Use an objective, unbiased and non-manipulative approach in the collection and interpretation of scientific data and avoid plagiarism and violation of Intellectual Property Rights. Appreciate and be sensitive to environmental and sustainability issues and understand its scientific significance and global relevance
PO7	Translate academic research into innovation and creatively design scientific solutions to problems. Exemplify project plans, use management skills, and lead a team in the planning and execution of a task
PO8	Understand the cross-disciplinary relevance of scientific developments and relearn and reskill to adapt to technological advancements.

PROGRAM SPECIFIC OUTCOMES

PSO	Description
On completion of M.Sc. Data Science program, students will be able to:	
PSO1	Identify, analyze and design solutions for data analytics problems using fundamental principles of mathematics, Statistics, computing sciences, and relevant domain disciplines
PSO2	Acquire the skills in handling data analytics programming tools towards problem solving and solution analysis for domain specific problems.
PSO3	Understand and commit to professional ethics, regulations, responsibilities, and norms of professional computing practices
PSO4	Understand the role of statistical approaches and apply the same to solve the real-life problems in the fields of data analytics.
PSO5	Apply the advanced concepts of Big Data to create a platform to gain analytical skills which impact business decisions and strategies
PSO6	Apply critical thinking to evaluate and optimize the computational resources in data analytics
PSO7	Understand, and compare different data sources and apply it for scientific knowledge acquisition as well as scientific data analysis and presentation.
PSO8	Use research-based knowledge to provide sustainable solutions in different domain using multidisciplinary data analytical approach
PSO9	Communicate effectively on the basis of presenting their research work and gain knowledge on documentation and reports writing in a professional way
PSO10	Create a new design for the complex computational problems which meets the specific needs for environmental and societal impact domains
PSO11	Demonstrate the principles of data science and apply these in the multidisciplinary environments to manage research project
PSO12	Gain hands-on experience with real datasets, work on team projects, and understand business-oriented data challenges in industry, academia, or government sectors

Deccan Education Society's

Kirti M. Doongursee College

(Autonomous) Proposed

Curriculum as per NEP-2020

Year of implementation- 2025-2026

Name of the Department-Data Science

Semester	Course Code	Course Title	Vertical	Credit
I	25DSMJ711	Essential Technologies for Data Science	Major	4
	25DSMJP71	Essential Technologies for Data Science – Practical	Major	2
	25DSMJ712	Data Analysis and Visualization	Major	4
	25DSMJP72	Data Analysis and Visualization-Practical	Major	2
	25DSMJ713	Statistical Methods for Data Science	Major	2
	25DSEL721 25DSELP71 25DSEL722 25DSELP72 25DSEL723 25DSELP73	Retail Marketing Analytics(2) Retail Marketing Analytics(2)-Practical [OR] SPARK Technologies(2) SPARK Technologies(2)-Practical (OR) Sports Data Analytics(2) Sports Data Analytics(2)-Practical	Elective	4
	25DSMJ714	Research Methodology	Research Methodology	4
	25DSMJ811	Artificial Intelligence	Major	4
	25DSMJP81	Artificial Intelligence – Practical	Major	2
	25DSMJ812	Time Series Analysis & Forecasting	Major	4
	25DSMJP82	Time Series Analysis & Forecasting – Practical	Major	2

II	**25DSMJ813**	**Ethical Issues in Data Science**	**Major**	**2**
	25DSEL821 **24DSELP81** **(OR)** **25DSEL822** **25DSELP82** **(OR)** **25DSEL823** **25DSELP83**	**Human Resource Analytics(2)** **Human Resource Analytics-Practical (2)** **Public Health care Analytics(2)** **Public Health care Analytics(2)-Practical** **Social Media Analytics(2)** **Social Media Analytics(2)-Practical**	**Elective**	**4**
OJT	**25DSOJT83**	**Internship**	**OJT**	**4**

Course Code	MAJOR SEM – I	Credits	Lectures /Week
25DSMJ711	PAPER I- Essential Technologies for Data Science	4	4

Course Outcomes:

After successful completion of this course, students would be able to

CO1: Learn different programming techniques and tools related to data science.

CO2: Know the core concepts of programming such as Python, R and Pandas.

CO3**:** Develop logic for Problem Solving and decision making.

CO4: Analyze Data using different programming techniques

Unit	Topics	No of Lectures
I	**Introduction to Data science and Python** a) Introduction to Data Science, data science life cycle, Applications, and advantages of Python over other programming languages b) What is Python? Why Should I learn Python? Installing Python How to execute Python program Writing your first program. c) Basic programming elements of Python-variables and constants, identifiers, Typecasting or Type Conversion in Python, indentation, comments, rules of writing identifiers, primitive data types, writing command line programs in python d) Operators in Python: Arithmetic operators, relational operators, Logical operators, Membership operators, Taking user input.	15
II	**Data structures and control flow** a) Collection data structures in Python- List, tuples, dictionary, sets and strings b) Control flow- Sequential, Branching or Conditional, Iteration or Repetition, Modular or Subroutines Conditional and iteration statements: if elif else statements, loops, for loop and while loops c) User defined functions in Python- No Value Pass and No Return, Value Pass and No Return, Value Pass and Return, Function with default arguments, Function with variable arguments, Higher order functions, list comprehension	15
III	**Statistics for Data Analysts** a) Permutations and combinations, probability, Descriptive statistics (mean, median, mode), point estimation, quartiles and boxplot, methods of dispersion, random variables and probability distribution b) Measures of shape- skewness, kurtosis, outlier detection, transformation (log, square root)	15

	c) Inferential statistics- Sampling techniques, Hypothesis testing, Z-score normalization, correlation, ANOVA d) Introduction to NumPy, creating NumPy arrays, indexing and slicing, vectorization, Boolean indexing, transformation, inferential statistics using NumPy	
IV	**Data wrangling using Pandas.** a) Introduction to data: NOIR (nominal, Ordinal, Interval and Ratio), continuous and discrete numeric data. Types of data analysis (descriptive, diagnostic, predictive and prescriptive analysis) b) Data wrangling using Pandas - Creating Series, Creating Data frame from dictionary, attributes, and method description of a data frame. Drop columns, add columns, add rows, iloc , loc, indexing and slicing data frames, selection with condition, group by summary operation, sorting operations c) Introduction to R IDE- components of R IDE, Basic data types in R, Data structures in R, data coercion, importing files, visualization using ggplot2. d) Basic visualization using matplotlib- Components of a chart, line chart, scatter chart, pie chart, sub plots.	**15**

References:

- Data Analysis with Pandas and Python by Boris Paskhaver, Manning Publications. Available at: https://www.perlego.com/book/2881120/pandas-in-action-pd

- Practical Statistics for Data Scientists: 50 Essential Concepts by Peter Bruce, Andrew Bruce, Peter Gedeck, O'Reilly Media, 2017 ISBN-10: 1491952962 ISBN-13: 978 1491952962,IInd Edition June 2020

- Foundations of Statistics for Data Scientists With R and Python By Alan Agresti, Maria Kateri, CRC Press Taylor and Francis group, 2022

Course Code	MAJOR SEM – I	Credits	Lectures/Week
25DSMJP71	**Paper I -Essential Technologies for Data Science – Practical**	**2**	**4**

Course Outcomes:
CO1: Gain the knowledge about Python and R programming language.
CO2: Comprehend Python data structures and able to use conditional and iterative control flow. Also, data wrangling and visualization concept in R and Pandas
CO3: Demonstrate descriptive, diagnostic, and inferential statistics using, Python, R or Excel (use Data analysis tool pack in Excel or Data analyzer tool in Microsoft office 365)
CO4: Interpret the results of Python, R and Pandas.

1	Write a Python program to accept inputs from users and perform arithmetic operations.
2	Write a program to demonstrate relational and logical operators in Python.
3	Write a Python program to demonstrate usage of loops. Use both for and while loops to distinguish between them. [e.g., Reversing the digits of a number without converting to String]
4	Demonstrate the use of data structures list, sets, dictionary.
5	Import a dataset and perform univariate analysis on the numeric columns to analyze the shape of the data. Write inference of the output. [Python or R or Excel]
6	Demonstrate Hypothesis testing, and ANOVA using a dataset [use Python, R or Excel]
7	Demonstrate correlation analysis. Use heatmap for visualization. Write inferences.
8	Import a csv or Excel dataset and demonstrate data wrangling, view shape, dimension, column names of the dataset, ways to select data using column number, column names, simple and compound conditional selection, update and modify dataset.
9	Demonstrate group by summary operations and sorting techniques.
10	Perform univariate, bivariate and multivariate analysis using visualization techniques in Python, R or Excel

Course Code	MAJOR SEM – I	Credits	Lectures/Week
25DSMJ712	Data Analysis and Visualization	4	4

Course Outcomes:

After successful completion of this course, students would be able to

CO1**:** Recall various data formats, sources and storage mechanisms.

CO2: Recognize the missing data and manage data wrangling and manipulation

CO3: Apply various tools for data visualization and making various report.

CO4: Study Data using different tools for creating interpretation and analyzing data visualizations.

Unit	Topics	No of Lectures
I	**Introduction to Data Analysis Data Analysis -** Exploratory Data Analysis and Data Science Process - Responsibilities of a Data Analyst - Data Analytics vs. Data Analysis - Types of Data - Understanding Different Types of File Formats - Sources of Data - Languages for Data Professionals - Overview of Data Repositories - Data Marts, Data Lakes, ETL, and Data Pipelines - Foundations of Big Data - Identifying Data for Analysis	15
II	**Data Wrangling Data Sources -** How to gather and Import Data - Data Loading, Storage and File Formats - Reading and Writing Data in Text Format, Web Scraping, Binary Data Formats, interacting with Web APIs, Interacting with Databases – Data Wrangling - Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting - Tools for Data Wrangling - Data Cleaning and Preparation - Handling Missing Data, Data Transformation, String Manipulation	15
III	**Data Visualization Intro to data visualization -** Introduction to Visualization and Dashboarding Software - Visualization Tools - Getting started with Tableau Desktop – Connecting to the dataset - Creating charts – Creating common visualizations (bar charts, line charts etc.) - Filtering and sorting data - Adding Titles, Labels, and descriptions - Publish your work to Tableau Cloud - Interactivity with text and visual tooltips - Interactivity with actions (filter, highlight, URL) – Assembling dashboards from multiple charts	15

IV	**Story Telling Introduction to Power BI - Understanding Desktop -** Understanding Power BI Report Designer - Report Canvas, Report Pages: Creation, Renames - Report Visuals, Fields and UI Options - Experimenting Visual Interactions, Advantages - Reports with Multiple Pages and Advantages - Pages with Multiple Visualizations - PUBLISH Options and Report Verification in Cloud - Adding Report Titles. Report Format Options - Introduction to data storytelling - Creating a data story	**15**

References:

- Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython by McKinney, W., 2nd edition. O"Reilly Media, 2017
- Doing Data Science:Straight Talk from the Frontline by O"Neil, C., & Schutt, R, O"Reilly Media, 2013
- The Big Book of Dashboards by Steve Wexler, Jeffrey Shaffer, Andy Cotgreave, John Wiley & Sons, 2017
- Practical Tableau by Ryan Sleeper, O"Reilly Media, 2018
- Power BI. Book-1, Business Intelligence Clinic: Create and Learn by Roger F Silva, 2018
- Introducing Microsoft Power BI by Alberto Ferrari and Marco Russo, Microsoft Press, Washington, 2016

Course Code	MAJOR SEM – I	Credits	Lectures/Week
25DSELP82	**Data Analysis and Visualization-Practical**	**2**	**4**

Course Outcomes:

After successful completion of this course, students would be able to

CO1: Recall the knowledge of working with data analysis tools.

CO2: Gain Knowledge of data wrangling for practical purposes.

CO3: Solve real-world data analysis problems with thorough, detailed examples,Creating a story using the dashboard to analyze data

CO4: Analyze the data to use Tableau to handle data from various sources and perform analysis of data, Handling and customizing Geospatial data using Tableau

Note: - The following set of practicals could be implemented in Python/ R/ Power BI/ Tableau or any other suitable software.	
1	Implement Data Loading, Storage and File Formats. Read data and store them in text format.
2	Implement the code to interact with Web APIs and to perform web scraping.
3	Demonstrate Data Cleaning and Preparation.
4	Implement Data wrangling on a data set.
5	Demonstrate the handling of missing data and string manipulation.
6	Create common charts with title, labels and descriptions using Tableau.
7	Perform sorting and filtering using tableau, create visualizations and publish it on Tableau Cloud.
8	Perform data visualization using Power BI.
9	Create reports using Power BI.
10	Create a data story in Tableau or power BI.

Course Code	MAJOR SEM – I	Credits	Lectures/Week
25DSMJ713	**Paper III – Statistical Methods in data Science**	**2**	**2**

Course Outcomes:

After successful completion of this course, students would be able to
CO1: Identify different types of data and their appropriate statistical measures.
CO2: Know the different statistical techniques necessary for the Data Science domain.
CO3: Execute the various statistical concepts in the relevant application areas.
CO4: Analyze problems and to make better decisions for the future in their fields, enhance their critical thinking skills and ability to solve problems using statistical methods.

Unit	Topics	No of Lectures
I	**Introduction to Applied Statistics**: The Nature of Statistics and Inference, What is "Big Data"?, Statistical Modelling, Statistical Significance Testing and Error Rates, Simple Example of Inference Using a Coin, Statistics Is for Messy Situations, Type I versus Type II Errors, Point Estimates and Confidence Intervals, Variable Types, Sample Size, Statistical Power, and Statistical Significance, The Verdict on Significance Testing, Training versus Test Data. **Computational Statistics**: Vectors and Matrices, The Inverse of a Matrix, Eigenvalues and Eigenvectors **Means, Correlations, Counts: Drawing Inferences**: Computing z and Related Scores, Statistical Tests, Plotting Normal Distributions, Correlation Coefficients, Evaluating Pearson's r for Statistical Significance, Spearman's Rho: A Nonparametric Alternative to Pearson, Tests of Mean Differences, t-Tests for One Sample, Two Sample t-Test, Paired-Samples t-Test, Categorical Data, Binomial Test, Categorical Data Having More Than Two Possibilities.	**15**
II	**Power Analysis and Sample Size Estimation:** Power for t-Tests, Power for One-Way ANOVA, Power for Correlations. Analysis of Variance: Fixed Effects, Random Effects, Mixed Models, Introducing the Analysis of Variance (ANOVA), Performing the ANOVA, Random Effects ANOVA and Mixed Models, One-Way Random Effects ANOVA, Simple and Multiple Linear Regression, Simple Linear Regression, Multiple Regression Analysis, Hierarchical Regression, How Forward Regression Works.	**15**

	Logistic Regression and the Generalized Linear Model: Logistic Regression, Logistic Regression, Predicting Probabilities, Multiple Logistic Regression, Training Error Rate Versus Test Error Rate. Multivariate Analysis of Variance (MANOVA) and Discriminant Analysis: Multivariate Tests of Significance, Example of MANOVA, Outliers, Homogeneity of Covariance Matrices, Linear Discriminant Function Analysis, Theory of Discriminant Analysis, Predicting Group Membership, Visualizing Separation	

References:

- Experimental Design in Data science with Least Resources N C Das Shroff Publishers 1st 2018
- Daniel W. W.: Applied Non-Parametric Statistics, First edition Boston-Houghton Mifflin Company.
- Univariate, Bivariate, and Multivariate Statistics Using R Daniel J. Denis Wiley 1st 2020

Course Code	MAJOR SEM – I	Credits	Lectures/Week
24DSMJ714	Paper IV- Research Methodology	4	4

Course Outcomes:

After successful completion of this course, students would be able to
CO1: List common research tools and instruments used for data collection, such as surveys, interviews, questionnaires, and observation.
CO2**:** Know the ethical considerations in research, such as informed consent, confidentiality, and the responsible use of data.
CO3**:** Implement appropriate research methods to design and conduct a research study, including selecting suitable data collection techniques and analyzing the data.
CO4: Compare the strengths and weaknesses of different research methodologies and select the most appropriate method for a given research problem.

Unit	Topics	No of Lectures
I	**Introduction:** Role of Business Research, Information Systems and Knowledge Management, Theory Building, Organization ethics and Issues Beginning Stages of Research Process: Problem definition, Qualitative research tools, Secondary data research	15
II	Research Methods and Data Collection: Survey research, communicating with respondents, Observation methods, Experimental research	15
III	Measurement Concepts, Sampling and Field work: Levels of Scale measurement, attitude measurement, questionnaire design, sampling designs and procedures, determination of sample size.	15
IV	Data Analysis and Presentation: Editing and Coding, Basic Data Analysis, Univariate Statistical Analysis and Bivariate Statistical analysis and differences between two variables. Multivariate Statistical Analysis.	15

References:

- Business Research Methods William, G.Zikmund, B.J, Babin, J.C. Carr, Atanu Adhikari, M.Griffin Cengage 8e 2016
- Business Analytics, Albright Winston, Cengage 5e 2015
- Research Methods for Business Students Fifth Edition, Mark Saunders 2011

Course Code	SEM I- ELECTIVE	Credits	Lectures/Week
25DSEL721	Retail Marketing Analytics	2	2

Course Outcomes:

After successful completion of this course, students would be able to

CO1**:Learn** how to use marketing analytics to perform predictive marketing

CO2**: Describe** the importance of marketing analytics for forward looking and systematic allocation of marketing resources

CO3**:** Execute the hypothesis to calculate power, and testing in R

CO4**: Interpret the** data and develop insights from it to address strategic marketing challenges

Unit	Topics	No of Lectures
I	**Introduction to Marketing Analytics and Exploratory Data analytics using R** a) Course Introduction ● why marketing analytics? ● course description and learning objectives b) Marketing Analytics Overview ● how analytics can assist marketing decision-making ● the framework of marketing optimization c) Tabulate and Summarize data ● what cleaned data looks like ● simple histogram plot ● use histogram and boxplot to inform data distribution d) Visualize data ● elements of data visualization ● histogram, scatter plot, line plot, bar charts, line fits with the ggplot() function	**15**
II	**a) Design and Conduct Experiments** ● design experiments, examples ● randomization/sample splitting ● conduct experiments **b) Assess Experiment Outcome Using Hypothesis Testing** ● why hypothesis testing for experiment outcomes ● terminologies for hypothesis testing ● how does hypothesis testing work ● power calculation ● conduct hypothesis testing in R	**15**

	c) Calculate and Predict CLV ● calculate CLV ● typical frameworks in predicting CLV ● using linear regression and logistic regression to predict CLV **d) CLV Analysis and Cohort Analysis Introduction to Experiment**	

References:

- Hands-on Data Science for Marketing by Yoon Hyup Hwang, Packt Publishing, 2019
- Retail Analytics: The Secret Weapon by Emmett Cox,1st edition , Weily , 2011
- Cutting Edge Marketing Analytics: Real World Cases and Data Sets for Hands on Learning by Venkatesan Rajkumar, Farris Paul and Ronald Wilcox, Pearson FT Press, 2014
- Marketing Analytics: A Practical Guide to Real Marketing Science by Grigsby Mike,Kogan Page, 2015

Course Code	**SEM I- ELECTIVE**	**Credits**	**Lectures/Week**
25DSELP71	**Retail Marketing Analytics-Practical**	**2**	**4**

Course Outcomes:

After successful completion of this course, students would be able to

CO1**:** Recall use of marketing analytics tool for prediction

CO2**:** Know how to use marketing analytics for forward looking and systematic allocation of marketing resources

CO3**:** Execute data analytics concepts on Marketing data

CO4**:** Interpret the data and develop insights to address strategic marketing challenges

1	**Learn how to tabulate and summarize marketing data using R.** ● Clean and preprocess the marketing data. ● Generate a simple histogram plot to visualize data distribution. ● Use tabulation and summary functions to gain insights from the data. ● Interpret the findings and discuss the implications for marketing analysis.
2	**Gain proficiency in visualizing marketing data using R.** ● Understand the key elements of data visualization. ● Create various visualizations such as histograms, scatter plots, line plots, and bar charts using the ggplot() function in R.

	• Apply appropriate visualization techniques to effectively communicate marketing insights.
3	**Design and conduct experiments for marketing campaigns.** • Learn about experimental design and its application in marketing. • Design experiments using examples from marketing scenarios. • Implement randomization and sample splitting techniques. • Conduct the experiments and collect relevant data for analysis.
4	**Understand the concept of hypothesis testing and its role in assessing experiment outcomes.** • Explore the purpose of hypothesis testing in analyzing experiment results. • Familiarize with key terminologies related to hypothesis testing. • Learn the process of hypothesis testing and power calculation. • Conduct hypothesis testing using R to evaluate experiment outcomes
5	**Calculate and predict Customer Lifetime Value (CLV).** • Calculate CLV using different approaches and frameworks. • Explore predictive modeling techniques such as linear regression and logistic regression for CLV prediction. • Assess the accuracy and reliability of CLV predictions.
6	**Apply CLV analysis and cohort analysis in marketing analytics.** • Analyze CLV data and identify patterns and trends. • Perform cohort analysis to segment customers based on their behavior or characteristics. • Interpret the results of CLV analysis and cohort analysis to derive actionable insights for marketing strategies.
7	**Extract data from social media platforms and perform analysis to gain insights into customer behavior and preferences.** • Utilize Python libraries like Beautiful Soup and requests to scrape data from social media platforms. • Clean and preprocess the scraped data. • Analyze the data to identify trends, sentiment analysis, or customer engagement metrics. • Visualize the findings using appropriate charts or graphs
8	**Analyze customer purchasing patterns and build a recommender system based on market basket analysis.** • Use transactional data to identify frequently occurring item sets using association rule mining algorithms. • Calculate support, confidence, and lift for the identified item sets. • Build a recommendation engine using collaborative filtering techniques. • Evaluate the performance of the recommender system and make recommendations based on customer preferences.
9	**Segment customers based on their recency, frequency, and monetary value (RFM) to better target marketing efforts.** • Analyze customer transaction data to calculate RFM scores. • Segment customers into different groups using clustering algorithms such as

	k-means or hierarchical clustering. ● Perform descriptive analysis on each customer segment to understand their characteristics. ● Develop targeted marketing strategies for each segment based on their RFM profiles.
10	**Conduct A/B testing to evaluate the impact of different marketing strategies and make data-driven decisions.** ● Design and implement A/B tests for marketing campaigns using randomized assignment. ● Collect relevant data and perform statistical analysis to compare the performance of different strategies. ● Calculate key metrics such as conversion rates, click- through rates, or revenue. ● Interpret the results and provide recommendations for optimizing marketing campaigns based on the findings.

Course Code	**SEM I- ELECTIVE**	**Credits**	**Lectures/Week**
25DSEL722	**SPARK Technologies**	**2**	**2**
Course Outcomes: After successful completion of this course, students would be able to CO1: Learn the concept of SPARK Technologies and its implementation CO2: Describe the different concepts of RDD CO3: Execute SPARK SQL, GraphX, Performance Tuning. CO4: Attribute large datasets across many CPUs.			
Unit	**Topics**		**No of Lectures**
I	**Introduction to SPARK Technologies** Components of the Spark unified stack, Features of Spark, Spark Web UI, an introduction to RDDs - Resilient Distributed Datasets, Launching and using Spark‟s Scala and Python shell, Spark Context, Spark Ecosystem, In-Memory data – Spark, Creating,Loading and Saving RDD, Transformations in RDD, Actions in RDD, Key-Value Pair RDD, Map Reduce and Pair RDD operations RDD Partitions		**15**
II	**implementation of SPARK Technologies**		**15**

	Spark Applications vs. Spark Shell, Creating Spark Context, Building a Spark Application, Spark and Hadoop Integration-HDFS, Handling Sequence File, Spark RDD-RDD Lineage, RDD Persistence Overview, Distributed Persistence. Spark Streaming, ML library for Spark, Working with Statistics, SPARK SQL, GraphX, Performance Tuning.	

References:

- Learning Spark: Lightning-Fast Data Analytics 2nd Edition, by Jules S. Damji, Brooke Wenig , Tathagata Das, Denny Lee, O"Reilly , 2020
- Apache Spark Machine Learning Blueprints 1st Edition, Kindle Edition by Alex Liu, Packt Publishing, 2016
- Apache Spark 2.x Cookbook: Cloud-ready recipes for analytics and data science 2nd

Course Code	SEM I- ELECTIVE	Credits	Lectures/Week
24DSELP72	**SPARK Technologies-Practical**	**2**	**4**

Course Outcomes:

After successful completion of this course, students would be able to

CO1: Recall the different SPARK Technologies and its use.

CO2: Know the implementation of Spark Technologies

CO3: Execute SPARK SQL, GraphX, Performance Tuning.

CO4: Organize Sparks Resilient Distributed Datasets to process and analyze large data sets across many CPUs

1	Installation of Apache Spark
2	Spark Basics and RDD interface
3	Filtering RDDs, and the Minimum Temperature by Location Example
4	Counting Word Occurrences using flatmap()
5	Executing SQL commands and SQL-style functions on a Data Frame
6	Implement Total Spent by Customer with DataFrames
7	Use Broadcast Variables to Display Movie Names Instead of ID Numbers
8	Create Similar Movies from One Million Rating
9	Using Spark ML to Produce Movie Recommendations
10	Use Windows with Structured Streaming to Track Most-Viewed URLs (Spark Streaming)

Course Code	SEM I- ELECTIVE	Credits	Lectures/Week
25DSEL723	**Sports Data Analytics**	2	2

Course Outcomes:

After successful completion of this course, students would be able to

CO1: Gain the knowledge in collecting, cleaning, and managing sports data

CO2: Know the role and importance of data analytics in sports

CO3: Execute statistical analysis techniques to analyze sports data

CO4**:** Compare data visualization methods to present sports data effectively

Unit	Topics	No of Lectures
I	**Fundamentals of Sports Data Analytics** A. Introduction to Sports Data Analytics ● Overview of sports data analytics ● Evolution of analytics in sports ● Importance and applications of sports data analytics B. Data Collection and Preprocessing ● Sources of sports data ● Data collection methods ● Data cleaning and preprocessing techniques ● Database management for sports analytics C. Data Visualization for Sports Analytics ● Principles of data visualization ● Visualization tools and libraries ● Creating effective visualizations for sports data ● Interactive dashboards for sports analytics D. Statistical Analysis in Sports ● Descriptive statistics for sports data ● Hypothesis testing in sports analytics ● Regression analysis in sports ● Analysis of variance (ANOVA) in sport	15
II	**Advanced Techniques in Sports Data Analytics** E. Predictive Modeling in Sports Analytics ● Introduction to predictive modeling ● Feature selection and engineering for sports data ● Linear regression models for sports prediction	15

	● Classification models for sports outcomes F. Machine Learning in Sports Analytics ● Overview of machine learning algorithms ● Decision trees and random forests in sports analytics ● Support vector machines for sports prediction ● Neural networks and deep learning in sports analytics G. Advanced Topics in Sports Data Analytics ● Sports performance analysis ● Player tracking and motion analytics ● Sports marketing and fan engagement analytics ● Sports injury prediction and prevention H. Sports Business Analytics ● Revenue generation and marketing in sports ● Fan engagement and customer analytics	

References:

1. Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers by Benjamin C. Alamar, Columbia university press, 2013
2. Sports Analytics and Data Science: Winning the Game with Methods and Models by Thomas Miller, 1st edition, Pearson FT Press, 2015
3. Sports Analytics: Analysis, Visualisation and Decision Making in Sports Performance By Daniel Memmert, Tim McGarry, and Tony Reilly, 2018
4. Cricket Analytics: Analytics and Data Science in Cricket by Tapan Bagchi and S. Raghunathan
5. Machine Learning using Python by Manaranjan Pradhan and U. Dinesh Kumar, Weily , 2020

Course Code	SEM I- ELECTIVE	Credits	Lectures/Week
25DSELP73	**Sports Data Analytics-Practical**	**2**	**4**

Course Outcomes:

After successful completion of this course, students would be able to

CO1: Recall sports data analytical concepts.

CO2: Describe the use of data analytics in sports

CO3: Execute data analytics techniques on sports data

CO4: Compare the data visualization methods to present sports data effectively

1	**Exploratory Data Analysis** ● Perform exploratory data analysis on a cricket dataset, analyzing variables such as number of matches, runs, not outs, wickets, etc. ● Visualize the distribution of player performance metrics using histograms, box plots, or scatter plots. Investigate the relationship between player age and performance metrics using correlation analysis.
2	**Batting Performance Analysis** ● Analyze batting performance in a Cricket dataset, calculating metrics such as batting average, strike rate, and runs scored by players. ● Identify top-performing batsmen based on performance metrics and compare their performance against different opponents or in specific conditions.
3	**Bowling Performance Analysis** ● Analyze bowling performance in the Cricket dataset, calculating metrics such as bowling average, economy rate, and wickets taken by players. ● Identify top-performing bowlers based on performance metrics and analyze their performance against different teams or in various match situations
4	**Performance Comparison** ● Compare the scoring averages of top-performing batsman in different seasons. ● Analyze the runs scoring (strike rate of Batting) of players from various teams in a specific league
5	**Player Position Analysis** ● Calculate position-specific performance metrics and compare players within each position.
6	**Injury Analysis** ● Investigate the relationship between player injuries and their subsequent performance using historical injury and performance data. ● Identify patterns and trends in the data to determine the impact of injuries on players performance and team success.

7	**Team Analysis** ● Analyze the impact of toss on a team's overall scoring and winning percentage. ● Study the relationship between batting averages of players and their team's win-loss record
8	**Sports Revenue Analysis** ● Analyze revenue generation in sports organizations by examining factors such as ticket sales, merchandise sales, and sponsorship deals. ● Identify key drivers of revenue and provide recommendations for maximizing financial performance
9	**Predictive Modeling** ● Build a regression model to predict the number of runs scored by players based on their historical performance data. ● Develop a classification model to predict the outcome of match based on team‟statistics.
10	**Visualization and Reporting: (Mini-Project)** ● Prepare a comprehensive report summarizing the findings of the analysis and providing actionable insights for sports teams or organizations.

SEMESTER II

Course Code	MAJOR SEM – II	Credits	Lectures/Week
25DSMJ811	Paper I- Artificial Intelligence	4	4

Course Outcomes:

After successful completion of this course, students would be able to

● Recall the foundational concepts of propositional logic, first-order logic, expert systems, probability theory, and basic machine learning techniques including supervised and unsupervised methods.

● Illustrate logical inference mechanisms, expert system architecture, probabilistic reasoning, and the working of machine learning models along with feature engineering techniques.

● Solve problems using different techniques such as logical inference, expert systems, probabilistic reasoning, and machine learning algorithms to address real-world challenges.

● Assess how inference methods, expert systems, and machine learning models help solve real-world problems.

Unit	Topics	No of Lectures
I	Logical Foundations and Inference Techniques Propositional Logic: Basics, Theorem Proving, Horn and Definite Clauses. Effective Propositional Model Checking First-Order Logic: Syntax, Semantics, and Knowledge Representation Inference in First-Order Logic: Unification, Lifting, Forward and Backward Chaining	15
II	Expert Systems and Probabilistic Reasoning Expert Systems: Development Phases, Architecture, and Comparison with Traditional Systems, Rule-Based Expert Systems. Probability Theory: Joint and Conditional Probability, Bayes's Theorem.	15
III	Introduction to Machine Learning and Supervised Learning Basics of Artificial Intelligence and Machine Learning. Types of Machine Learning. Steps in Machine Learning. Performance Metrics. Supervised Learning Algorithms: Decision Trees, Support Vector Machines (SVM). Bias-Variance Trade-off. Model Evaluation: Cross-validation, Confusion Matrix	15
IV	Unsupervised Learning and Feature Engineering Clustering Algorithms: K-Means, Hierarchical Clustering	15

	Dimensionality Reduction: Principal Component Analysis (PCA) Feature Engineering: Feature Scaling, Encoding Techniques, Handling Missing Data and Outliers Neural Networks: Introductions to Neural Networks, Architecture of Neural Networks	

References:

- Introduction to Machine Learning EthemAlpaydin PHI 2nd Edition 2013
- Elaine Rich, Kevin Knight (2017). Artificial Intelligence, Tata McGraw Hill.
- Rogers, A., & Kampffmeyer, M. (2020). The Data Science Workshop. Packt Publishing.
- Ross, S. M. (2014). A first course in probability (9th ed.). Pearson.
- Stair, R. M., & Reynolds, G. W. (2011). Expert systems: Principles and practice (2nd ed.). Course Technology.

Additional References:

- Luger, G. F. (2009). Artificial intelligence: Structures and strategies for solving complex problems (6th ed.). Pearson.
- Mendelson, E. (2015). Mathematical logic (6th ed.). Dover Publications.
- Mueller, J. P., & Massaron, L. (2016). Machine learning for dummies. John Wiley & Sons.
- Russell, S., & Norvig, P. (2020). Artificial intelligence: A modern approach (4th ed.).

Course Code	MAJOR SEM – II	Credits	Lectures/Week
25DSMJP81	Paper I-Artificial Intelligence – Practical	2 credits	4 lectures

Course Outcomes:

After successful completion of this course, students would be able to

CO1: Recall common algorithms used in AI and ML, including decision trees, support vector machines (SVM), k-means clustering, and Neural network architectures.

CO2: Understand the History of Artificial Intelligence and the future of AI

CO3: Apply Inference and machine learning algorithms

CO4: Interpret the results of inference, expert system and machine learning algorithms.

1	Propositional Logic Theorem Proving using Python: Implement basic propositional logic statements and perform simple theorem proving using Python logic libraries or custom code.
2	Forward and Backward Chaining for Inference: Create Python programs to simulate forward chaining and backward chaining inference techniques using a set of rules and facts.
3	Building a Simple Rule-Based Expert System: Develop a basic rule-based expert system for a domain like medical diagnosis or loan approval, applying if-then rules in Python.
4	Probabilistic Reasoning with Bayes' Theorem: Implement Bayesian inference for a simple problem such as spam email detection or disease diagnosis using conditional probabilities.
5	Decision Tree Classifier for Multi-class Problems: Implement a decision tree classifier on a dataset and visualize the tree structure using libraries like scikit-learn and matplotlib.
6	Support Vector Machine (SVM) Classification: Train and test an SVM classifier for a classification problem, experimenting with different kernel functions and parameter tuning
7	Cross-Validation and Model Evaluation: Implement k-fold cross-validation and compare the accuracy of different models (e.g., Decision Tree vs. SVM). Show confusion matrices and ROC curves.
8	K-Means Clustering for Customer Segmentation: Apply K-Means clustering on a customer dataset to find distinct customer groups and interpret cluster characteristics.
9	Feature Engineering and PCA: Perform feature scaling, one-hot encoding of categorical variables, and apply PCA for dimensionality reduction on a real-world dataset.
10	Neural Networks: Build and train a simple feedforward neural network to classify handwritten digits from the MNIST dataset.

Course Code	MAJOR SEM – II	Credits	Lectures/Week
25DSMJ812	Paper II – Time Series Forecasting	4	4

Course Outcomes:

After successful completion of this course, students would be able to

CO1: Forecast the trend pattern exhibited by the given data by using various methods

CO2: Run and interpret time series models and regression models for time series

CO3: Use the Box-Jenkins approach to model and forecast time series data empirically

CO4: Analyze and estimate the cyclic components using special processes.

Unit	Topics	No of Lectures
I	**Introduction to Trend** Introduction to times series data, application of time series from various fields, Components of a time series, Decomposition of time series. Trend: Estimation of trend by free hand curve method - method of semi averages - fitting a various mathematical curve and growth curves.	15
II	**Trend and Seasonal component** Method of moving averages – Detrending - Effect of elimination of trend on other components of the time series. Seasonal Component: Estimation of seasonal component by Method of simple averages, Ratio to Trend, Ratio to moving average and Link relatives	15
III	**Forecasting** Variate component method - Stationary Time series: Weak stationary, autocorrelation function and correlogram of moving average Forecasting: Exponential smoothing methods, short term forecasting methods: Brown‟s discounted regression, Box-Jenkins Method.	15
IV	**Cyclic Component** Deseasonalization - Cyclic Component: Harmonic Analysis. Some Special Processes: Moving-average (MA) process and Autoregressive (AR) process of orders one and two - Estimation of the parameters of AR (1) and AR (2) – Yule-Walker equations.	15

References:

- Kendall, M. (1976) Time Series. 2nd Edition, Charles Griffin and Co Ltd.,

London and High Wycombe.
- Chatfield C. (1980). The Analysis of Time Series –An Introduction, 6th Edition, Chapman & Hall.
- Mukhopadhyay P. (2011). Applied Statistics, 2nd ed. Revised reprint, Books and Allied
- Shumway, R. H., and Stoffer, D. S. (2006). Time Series Analysis and Its Applications With R Examples, 2 ed. Springer, New York, NY
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). Time Series Analysis: Forecasting and Control. Prentice - Hall, Inc., Upper Saddle River, NJ.
- Yaffee, R. and McGee, M. (2000). Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS. Academic Press, Inc., San Diego, CA

Course Code	**MAJOR SEM – II**	**Credits**	**Lectures/Week**
24DSMJP82	**Paper II – Time Series Analysis and Forecasting– Practical**	**2**	**4**

Course Outcome:
After successful completion of this course, students would be able to
CO1: Fit various growth curves, trend and to measure seasonal indices
CO2: Know the forecasting by different methods
CO3: Execute different forecasting methods and calculate variance of a random component
CO4: Analyze various time series analysis and forecasting techniques

Note: Software: Time Series Lab/any statistical software/any programming platform

1	Fitting and plotting of modified exponential curve.
2	Fitting and plotting of Gompertz curve.
3	Fitting and plotting of logistic curve.
4	Fitting of trend by Moving Average Method.
5	Measurement of Seasonal indices Ratio-to-Trend method.
6	Measurement of Seasonal indices Ratio-to-Moving Average method.
7	Measurement of seasonal indices Link Relative method.
8	Calculation of variance of random component by variate difference method.
9	Forecasting by exponential smoothing.
10	Forecasting by short term forecasting methods.

Course Code	MAJOR SEM – II	Credits	Lectures/Week
25DSMJ813	PAPER III-Ethical Issues in Data Science	2	2

Course Outcomes:

After successful completion of this course, students would be able to

CO1**:** Learn the fundamental ethical issues that arise in the field of data science

CO2**:** Illustrate the ethical implications of data collection, and decision-making processes

CO3**:** Implement the knowledge of Ethical issues in Data Science

CO4**:** Attribute different concerns related to data collection, storage, and sharing

Unit	Topics	No of Lectures
I	**Ethical foundation in Data Science** Introduction to ethical frameworks and principles, Ethics in data science: challenges and importance, Data collection methods, storage, sharing and its ethical considerations, Types of bias in data, Importance of transparency in data science, Ethical considerations in automated decision-making, Data governance frameworks and practices, Ensuring accountability in data science projects	**15**
II	**Emerging Ethical issues in Data Science** Ethical Issues in Data Visualization, Ethical Issues in Machine Learning, Ethical challenges in emerging technologies e.g., AI, IoT, biometrics, blockchain, Ethical challenges in data science research, Ethical considerations in collaborative data science environments, Ethical issues in using the internet, privacy and security, in the context of data science	

References:

- Data Science Ethics, David Martens ISBN: 9780192847263 Oxford University Press 2023
- Ethics of Big Data: Balancing Risk and Innovation" by Kord Davis and Doug Patterson O'Reilly 2012.
- Data Science Ethics Resources - Concepts, Techniques, and Cautionary Talesby David Martens-Oxford University Press-2022.

Course Code	ELECTIVE SEM – II	Credits	Lectures/Week
25DSEL821	**Human Resource Analytics**	**4**	**4**

Course Outcomes:

After successful completion of this course, students would be able to

CO1: Identify the application and uses of HR analytics in various HR sub-systems

CO2**:** Describe the problems and issues in HR and the relevance of HR analytics

CO3: Synthesize the tools, methods and techniques of HR analytics to understand real world corporate scenario.

CO4**:** Attribute the challenges and future trends in HR analytics

Unit	Topics	No of Lectures
I	**HR Measurement** Need for HR Measurement, Significance and concept of HR Analytics, HR Analytics and business linkages, Prerequisites of HR Analytics; Models and frameworks of HR Analytics; Measuring intellectual capital, need and rationale for HR Accounting & Audit, Approaches and methods of HR Accounting & Audit **HRIS for HR Analytics**: What is Human Resource Information System; Role of HRIS in analytics; HRIS development and Implementation, the development process- need analysis, systems design, structure and culture; HRIS Applications- making HRIS work.	**15**

II	**Analytics for HR sub-systems** HR Analytics for Staffing, Training & Development, Performance Management Systems, Career Planning Systems, Rewards and Compensation Management, Employee Relations Systems. **Analytics for HR system:** HR performance frameworks and measurement systems; Measuring HR Climate and People Management Capabilities; Competency Management Frameworks & Competency Mapping, Integration of competency-based HR System. Measuring HR Effectiveness, The HR Scorecard **Trends and Future Challenges:** Technology and changes in HR Analytics, Role of social media, Big Data and Predictive Analytics in HR, Assessing the effectiveness of HR Analytics, Post analysis steps, Review and monitoring, Issues in HR valuation and measurement; Emerging challenges: Global and Indian Experience	**15**

References:

- Ulrich, D. & Brockbank, W., The HR Value Proposition. Harvard Business School Press 2016
- How to measure HRM by Jac Fitz-enz 2002
- Predictive Analytics for Human Resources by Jac Fitz-enz, John Mattox II, Wiley 2014
- Making Human Capital Analytics Work: Measuring the ROI of Human Capital Processes and Outcomes. By by Jack Phillips,Patricia Pulliam Phillips- 2014

Course Code	ELECTIVE SEM – II	Credits	lectures/Week
25DSELP81	Human Resource Analytics-Practical	2	4

Course Outcomes:

After successful completion of this course, students would be able to

CO1: Recall the concepts of HR analytics and its use in various HR sub-systems

CO2: Summerize the issues in HR and its importance in HR analytics.

CO3: Execute the techniques of HR analytics to explore the scenario corporate scenario.

CO4: Interpret the real world corporate scenario using HR analytics.

Note: -Download R from http://cran.r-project.org/
Download R Studio fromhttp://www.rstudio.com/products/rstudio/download/

1	**Analyze employee turnover rates and identify factors contributing to attrition** ● Collect historical employee data, including tenure, performance ratings, salary, and job satisfaction. ● Calculate employee turnover rates for different departments and job roles. ● Conduct statistical analysis to identify correlations between turnover and variables such as salary, job satisfaction, and performance. ● Generate visualizations (e.g., charts, graphs) to present the findings and propose recommendations to reduce turnover.
2	**Develop a user-friendly HRIS dashboard for monitoring and analyzing HR metrics** ● Identify key HR metrics to be displayed on the dashboard (e.g., headcount, recruitment pipeline, training hours). ● Design the layout and interface of the HRIS dashboard using appropriate programming languages and tools. ● Integrate data from various HR systems and databases to populate the dashboard in real-time. ● Implement interactive features, such as drill-down capabilities and data filters, to facilitate data exploration and analysis
3	**Analyze training effectiveness and identify skill gaps in the organization** ● Collect training data, including participant demographics, training modules, pre/post-assessment scores, and performance metrics. ● Perform statistical analysis to evaluate the impact of training on employee performance. ● Identify areas of improvement and recommend targeted training programs based on identified skill gaps. ● Develop a visualization or report summarizing the training needs analysis results.
4	**Develop an HR scorecard to measure HR effectiveness and align HR strategies with organizational goals** ● Identify key HR performance indicators aligned with the organization's strategic objectives.

	● Collect relevant data for each HR indicator, such as employee satisfaction surveys, training investment data, and performance metrics. ● Calculate HR metrics and indicators, such as turnover rate, training ROI, and employee engagement index. ● Design a dashboard or report to present the HR scorecard and analyze trends over time.
5	**Use predictive analytics to forecast employee attrition and develop retention strategies** ● Gather historical HR data, including employee demographics, performance metrics, compensation, and employee exit data. ● Build a predictive model (e.g., logistic regression, decision tree) to predict employee attrition. ● Validate the model's accuracy and evaluate its performance using appropriate evaluation metrics. ● Generate actionable insights and recommendations to proactively address potential attrition risks.
6	**Use predictive analytics to forecast employee attrition and develop retention strategies** ● Gather historical HR data, including employee demographics, performance metrics, compensation, and employee exit data. ● Build a predictive model (e.g., logistic regression, decision tree) to predict employee attrition. ● Validate the model's accuracy and evaluate its performance using appropriate evaluation metrics. ● Generate actionable insights and recommendations to proactively address potential attrition risks.
7	**Measure and analyze employee engagement levels within the organization** ● Collect employee engagement survey data, including responses to survey questions related to job satisfaction, work environment, and organizational culture. ● Calculate engagement scores and identify key drivers of engagement. ● Conduct a sentiment analysis on employee feedback to understand areas of improvement. ● Present the findings and propose strategies to enhance employee engagement based on the analysis.
8	**Develop a program to automate repetitive HR processes, such as leave management or performance appraisal** ● Identify the HR process to be automated and define the required functionalities. ● Design and implement a web-based application or script to streamline the process using appropriate programming languages and frameworks. ● Integrate the application with relevant HR systems and databases to ensure data accuracy and consistency. ● Test and validate the automated process, considering different scenarios and user inputs.
9	**Analyze the effectiveness of the organization's performance management system and provide insights for improvement.** ● Collect performance evaluation data, including performance ratings, goal achievement metrics, and feedback. ● Analyze the distribution of performance ratings across different departments or job roles. ● Identify trends and patterns in performance data and assess the fairness

	and consistency of the evaluation process. ● Propose recommendations for enhancing the performance management system based on the analysis results.
10	**Analyze the organization's compensation structure and compare it to industry benchmarks.** ● Gather salary data for different job roles and levels within the organization. ● Perform a salary analysis, including measures like average salary, salary distribution, and salary competitiveness. ● Conduct benchmarking by comparing the organization's salary data with industry standards or competitor data.

Course Code	**ELECTIVE SEM – II**	**Credits**	**Lectures/Week**
25DSEL822	**Public Health Analytics**	**2**	**4**
Course Outcomes: After successful completion of this course, students would be able to CO1: Learn different concepts of healthcare data management. CO2: Acquire fundamental skills in using popular software tools for conducting data analyses CO3: Implement the data analysis techniques to create scenario specific perspectives. CO4: Examine the integrity of these data sets and their suitability for meaningful comparisons.			
Unit	**Topics**		**No of Lectures**
I	**Healthcare data management** ● What is Health Data Management? Benefits and challenges of health data management, how to store all that data ● Electronic Health Records– Components of EHR-Coding Systems- Benefits of EHR- Barrier to Adopting EHR Challenges ● Phenotyping Algorithms ● Statistical analysis of healthcare data- Measures of Central Tendency and Dispersion, Confidence Limits and Hypothesis Testing, Statistical Tests for Categorical Data, T-Tests for Related and Unrelated Data, Analysis of Variance ● Data Quality and Governance		**15**

I	**Healthcare data Analysis:** ● Biomedical Image and Signal Analysis, Genomic Data Analysis for Personalized Medicine ● Natural Language Processing and Data Mining for Clinical Text, Mining the Biomedical- Social Media Analytics for Healthcare. ● Predictive Models for Integrating Clinical and Genomic Data, Privacy- Preserving Data, Publishing Methods in Healthcare ● Mobile Imaging and Analytics for Biomedical Data, Data Analytics for Pharmaceutical Discoveries- Clinical Decision Support Systems	**15**
References: ● Healthcare Business Intelligence + Website - A Guide to Empowering Successful Data Reporting and Analytics Hardcover by LB Madsen (Author),2012 ● Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence (Marketing Science) 1st Edition by Dr. Steven Struhl,2016		

Course Code	ELECTIVE SEM-II	Credits	Practicals/Week
25DSELP82	**Public Health Analytics-Practical**	**2**	**4**

Course Outcomes:

After successful completion of this course, students would be able to

CO1: Recall the different concepts of Public Health Analytics.

CO2: Know different data analytics tools for exploring Health Sector data.

CO3: Develop data models that integrate patient information from multiple origins to create comprehensive and patient-centric perspectives.

CO4: Investigate the reliability, accuracy, and comparability of health and genomic data.

1	EHR data modeling, data mining, exploratory data analysis using tableau or power BI.
2	Medical expenditure prediction
3	A Twitter Healthcare data extraction, processing and sentiment analysis use keywords related to health (e.g., COVID-19 vaccine).
4	Using classification algorithms in prediction of diseases
5	Clinical Text Data Categorization and Feature Extraction
6	Medical Image classification (x-rays)
7	Image object detection using deep learning.
8	Health forecasting using time-series.
9	Heart sound classification using signal processing.
10	AI conversational chatbot for primary healthcare diagnosis.

Course Code	ELECTIVE SEM – II	Credits	Lectures/Week
25DSEL 823	**Social Media Analytics**	**2**	**4**

Course Outcomes:

After successful completion of this course, students would be able to

CO1: Familiarize with the concept of social media and social media analytics.

CO2: Describe the significance of Social media analytics.

CO3: Develop skills required for analyzing the effectiveness of social media.

CO4: Analyze different visualization techniques for social media decisions.

Unit	Topics	No of Lectures
I	**Social Media Analytics: An Overview** Core Characteristics of social media, Types of social media, social media landscape, Need for Social Media Analytics (SMA), Seven Layers of Social Media Analytics, Types of Social Media Analytics, Social Media Analytics Cycle, Location Analytics - Sources of Location Data, Categories of Location Analytics, Social Information Filtering: Social Sharing and filtering, Automated Recommendation systems, Traditional Vs social Recommendation Systems, understanding social media and Business Alignment, social media KPI, formulating a Social Media Strategy, Managing Social Media Risks **CaseStudy**: Online Behavior on Twitter	**15**
II	**Social Network Structure, Measures & Visualization** ● Basics of Social Network Structure - Nodes, Edges & Tie ● Describing the Networks Measures - Degree Distribution, Density, Connectivity, Centralization, Tie Strength & Trust. Network Visualization - Graph Layout, Visualizing Network features, Scale Issues. Capturing Correlations: Triangles, Clustering, and Assortativity. ● Social Media Network Analytics - Common Network Terms, Common Social Media Network Types, Types of Networks, Common Network Terminologies, Network Analytics Tools. **Case Study**: LinkedIn	**15**

References:

- Seven Layers of Social Media Analytics_ Mining Business Insights from Social Media Text, Actions, Networks, Hyperlinks, Apps, Search Engine, and Location Data, Gohar F. Khan, 2015
- Analyzing the Social Web 1st Edition by Jennifer Golbeck, 2013

Course Code	ELECTIVE-SEM-II	Credits	Lectures/Week
25DSELP83	**Social Media Analytics-Practical**	**2**	**4**

Course Outcomes:

Course Outcomes:

After successful completion of this course, students would be able to

CO1: Recall concepts of social media analytics.

CO2**:** Know the different analytical techniques use for Social media analytics.

CO3**:** Execute data analysis techniques for analyzing the effectiveness of social media.

CO4**:** Compare different visualization techniques for social media decisions.

Note: - The following set of practicals should be implemented in Scrape, python:
Link: -Python: https://www.python.org/downloads/

1	Study Various ● Social Media platforms (Facebook, twitter, YouTube etc) ● Social Media analytics tools (Facebook insights, google analytics netlytic etc) ● Social Media Analytics techniques and engagement metrics (page level, post level, member level) using Gephi Tool
2	Scrape an online Social Media Site for Data. Use python to scrape information from twitter. Exploratory Data Analysis and visualization of Social Media Data
3	Create sociograms for the persons-by-persons network and the community-by-community network for a given relevant problem. Create a one-mode network and two-node network for the same. Datasets: les-Misérables, Airlines, Internet Core Routers
4	Develop Content (text, emoticons, image, audio, video) based social media analytics model for business. (e.g., Content Based Analysis: Topic, Issue, Trend, sentiment/opinion analysis, audio, video, image analytics)
5	Develop Structure based social media analytics model for any business. (e.g., Structure Based Models -community detection, influence analysis)
6	Develop a dashboard and reporting tool based on real time social media data Using Power BI
7	Use Google Visualization Charts to analyze social media data
8	Analyze social media data Network Analysis with Orange Softwa
9	Use Graph Neural Networks on the datasets (Planetoid Cora Dataset)/ Jazz Musicians Network.
10	Analyze Twitter conversations to identify the most active and influential users using Machine Learning Algorithms with Gephi Tool.

Course Code	Course title	Credits
25DSOJT83	**On Job Training**	**4**

A. Introduction

- On Job training (OJT) is an integral component of the M.Sc. Data Science program that provides students with a unique opportunity to bridge the gap between theoretical knowledge gained in the classroom and practical application in a real-world environment. This training aims to equip students with both technical and non-technical skills that are essential for success in the industry.
- By participating in OJT, students are able to apply the concepts and theories learned during their coursework to real-world scenarios. They gain hands-on experience, problem-solving skills, and a deeper understanding of how the industry operates. This practical exposure enhances their competence and confidence, preparing them to tackle the challenges they may encounter in their professional careers.
- From an organizational perspective, hosting OJT programs allows companies to gain insights into the curriculum and content of the M.Sc. Data Science program. They can provide valuable feedback on the relevance of the coursework and industry requirements, enabling academic institutions to continually improve the program's alignment with industry needs. This collaboration between academia and industry fosters a mutually beneficial relationship, ensuring that graduates are well-prepared for the job market.
- Moreover, OJT benefits the faculty members involved in the program. They have the opportunity to gain firsthand exposure to the industry and observe the type of work being performed. This experience enables them to enhance their teaching methodologies and delivery techniques, ensuring that they remain up-to-date with the latest industry practices. The insights gained from OJT also enable faculty members to provide relevant guidance and mentorship to students, preparing them for successful careers in the field of data science.

B. Enhancing Practical Skills through OJT

● The On Job Training (OJT) program spans 4-6 weeks, requiring a minimum of 120 hours of physical presence at the organization.

● Students are expected to find their own OJT placements, although the institution provides support and guidance in securing positions with reputable organizations.

● OJT must be conducted outside the home institution to expose students to real-world work environments.

● OJT covers any subject within the syllabus, allowing students to align their experience with their academic interests.

● In recognition of changing dynamics, some OJT sessions can be conducted online to accommodate virtual work environments.

● OJT will offer students the opportunity to apply classroom learning in a real-world setting, fostering the development of technical and non-technical skills.

● Mutual Benefits: Organizations gain insights into the program's curriculum and industry requirements, enabling them to provide constructive feedback and enhance course relevance.

● OJT bridges the gap between theoretical knowledge and practical application, preparing students for successful careers in data science

C. Interning organization: Students have the flexibility to pursue their OJT in various types of organizations, including but not limited to:

● Software Development Firms: Gain practical experience in software development and programming.

● Hardware/Manufacturing Firms: Learn about hardware design, manufacturing processes, and quality assurance.

● Small-Scale Industries/Service Providers: Explore opportunities in diverse sectors such as banking, clinics, NGOs, and professional institutions like CA firms or law firms.

● Civic Departments: Engage with local civic departments such as ward offices, post offices, police stations, or panchayats to understand their functioning and contribute to their activities.

● Research Centre"s/University Departments/Colleges: Contribute as research assistants or in similar roles for research projects or initiatives, fostering collaboration between academia and industry.

Note: The listed options provide a range of possible OJT placements, offering students valuable exposure to different sectors and professional settings.

D. OJT mentors:

To enhance the learning experience and ensure the quality of the MSc program, each student participating in the OJT will be assigned two mentors: a faculty mentor from the institution and an industry mentor from the organization where the student is interning.

- **Industry Mentor Role:** The industry mentor plays a crucial role in guiding the student during the internship. They ensure that the internee fulfills the requirements of the organization and successfully meets the demands of the assigned project. Through their expertise and experience, industry mentors provide valuable insights into real-world practices and industry expectations.
- **Faculty Mentor Role:** The faculty mentor serves as the overall coordinator of the OJT program. They oversee the entire internship process and evaluate the quality of the OJT in a consistent manner across all students. The faculty mentor ensures that the OJT aligns with the program's objectives and provides valuable learning opportunities. They also facilitate communication between the institution, industry mentor, and student to ensure a fruitful OJT experience.

By having both an industry mentor and a faculty mentor, students benefit from a comprehensive guidance system that combines industry expertise and academic support. This dual mentoring approach ensures a well-rounded and rigorous OJT experience for every student in the program.

E. Submission of documentation for OJT

The student will make two documents as part of the OJT

- **Online diary:** This ensures that the student updates daily activity, which could be accessed by both the mentors. Daily entry can be of 3- 4 sentences giving a very brief account of the learning/activities/interaction taken place. The faculty mentor will be monitoring the entries in the diary regularly as shown in **Appendix-I**
- **OJT report:** A student is expected to make a report based on the OJT he or she has done in an organization. It should contain the following:

— Certificate: A certificate in the prescribed Performa (given in **Appendix II** and **Appendix III**) from the organization where the OJT was done.

— Title: A suitable title giving the idea about what work the student has performed during the OJT.

— Description of the organization: A small description of the organization where the student has interned

— Description of the activities done by the section where the intern has worked: A description of the section or cell of the organization where the intern worked. This should give an idea about the type of activity a new employee is expected to do in that section of the organization.

— Description of work allotted and done by the intern: A detailed description of the work allotted, and actual work performed by the intern during the OJT period. It shall be the condensed and structured version of the daily report mentioned in the online diary.

— Self-assessment: A self-assessment by the intern on what he or she has learned during the OJT period. It shall contain both technical as well as interpersonal skills learned in the process.

F. Interaction between mentors:

To ensure the smooth conduct of the OJT a meet-up involving the intern, industry mentor, and the faculty mentor will be scheduled as a mid-term review. The meeting can preferably be online to save time and resources. The meeting ensures the synergy between all stakeholders of the OJT. A typical meeting can be of around 15 minutes where at the initial stage the intern brief about the work and interaction goes for about 10 minutes. This can be followed by the interaction of the mentors in the absence of the intern. This ensures that issues between the intern and the organization, if any, are resolved amicably.

G. OJT workload for the faculty:

Every student is provided with a faculty member as a mentor. So, a faculty mentor will have a few students under him/her. A faculty mentor is the overall in charge of the OJT of the student. He/she constantly monitors the progress of the OJT by regularly overseeing the diary, interacting with the industry mentor, and guiding on the report writing etc. Considering the time and effort involved, a faculty mentor who is in-charge of 20 students shall be provided by a workload of 3 hours.

Evaluation Scheme for First Year (PG) under NEP (4 credits)

I. Internal Evaluation for Theory Courses – 40 Marks

1) Continuous Internal Assessment(CIA) Assignment - Tutorial/ Case Study/ Project / Presentations/ Group Discussion / Ind. Visit. – 20 marks

2) Continuous Internal Assessment(CIA) ONLINE Unit Test – 20 marks

II. External Examination for Theory Courses – 60 Marks

Duration: 2 Hours

Theory question paper pattern:

Question	Based on	Marks
Q.1	Unit I	15
Q.2	Unit II	15
Q.3	Unit III	15
Q.4	Unit IV	15

- All questions shall be compulsory with internal choice within the questions.
- Each Question may be subdivided into sub questions as a, b, c, d, etc. & the allocation of Marks depends on the weightage of the topic.

- **Paper Pattern of Theory Paper:(for 4 credits)**

DES's Kirti M. Doongursee College (AUTONOMOUS), Dadar (W), Mumbai-28			
Regular / Additional / ATKT Examination			
Duration: 2 Hour			**Max Marks: 60**
Date: **Time:** **Code:** ***(For office use)***			
N. B.	i)	*All Questions are compulsory.*	
	ii)	*Mixing of sub-questions is not allowed*	
	iii)	*Draw neat labeled diagrams wherever necessary.*	
Q. No.			**Marks**
Q.1 A			**07**
OR			
Q.1 B			**07**
Q.1 C			**08**
OR			

Q.1 D		**08**
Q.1 E		**07**
OR		
Q.1 F		**07**
Q.2 G		**08**
OR		
Q.2 H		**08**
Q.3 I		**07**
	OR	
Q.3 J		**07**
Q.3 K		**08**
	OR	
Q. 3 L		**08**
Q.4 M		**07**
	OR	
Q.4 N		**07**
Q.4 O		**08**
	OR	
Q.4 P		**08**

III. Practical Examination

- Each core subject carries 50 Marks
- Duration: 2 Hours for each practical course.
- Minimum 80% practical from each core subjects are required to be completed.
- Certified Journal is compulsory for appearing at the time of Practical Exam

NOTE: To pass the examination, attendance is compulsory in both Internal & External (Theory + Practical) Examinations.

Evaluation Scheme for First Year (PG) under NEP (2 credits)

I. Internal Evaluation for Theory Courses – 20 Marks

1) Continuous Internal Assessment(CIA) Assignment - Tutorial/ Case Study/ Project / Presentations/ Group Discussion / Ind. Visit. – 10 marks

2) Continuous Internal Assessment(CIA) ONLINE Unit Test – 10 marks

II. External Examination for Theory Courses – 30 Marks

Duration: 1 Hours

Theory question paper pattern:

Question	Based on	Marks
Q.1	Unit I	15
Q.2	Unit II	15

- All questions shall be compulsory with internal choice within the questions.
- Each Question may be subdivided into sub questions as a, b, c, d, etc. & the allocation of Marks depends on the weightage of the topic.
- **Paper Pattern of Theory Paper:(for 2 credits)**

DES's Kirti M. Doongursee College (AUTONOMOUS), Dadar (W), Mumbai-28			
Regular / Additional / ATKT Examination			
Duration: 1 Hour			**Max Marks: 30**
Date: **Time:** **Code:** ***(For office use)***			
N. B.	i)	*All Questions are compulsory.*	
	ii)	*Mixing of sub-questions is not allowed*	
	iii)	*Draw neat labeled diagrams wherever necessary.*	
Q. No.			**Marks**
Q.1 A			**07**
OR			
Q.1 B			**07**
Q.1 C			**08**

OR		
Q.1 D		**08**
Q.1 E		**07**
OR		
Q.1 F		**07**
Q.2 G		**08**
OR		
Q.2 H		**08**

III. Practical Examination

- Each core subject carries 50 Marks.
- Duration: 2 Hours for each practical course.
- Minimum 80% practical from each core subjects are required to be completed.
- Certified Journal is compulsory for appearing at the time of Practical Exam

NOTE: To pass the examination, attendance is compulsory in both Internal & External (Theory + Practical) Examinations.

Evaluation Scheme for First Year PG - On Job Training Course (4 Credit Course)

Internal Evaluation	
Online diary	25
Mid-term interaction	25
Total	50
External Evaluation	
OJT Documentation	25
Quality & Relevance	10
OJT Viva	15
Total	50

Appendix-I

Maintain the weekly online diary for each week in the following format.

	Day	**Date**	**Name of the topic/Module Completed**	**Remark**
1st Week	Monday			
	Tuesday			
	Wednesday			
	Thursday			
	Friday			
	Saturday			
Signature of the faculty member Seal of the College				

Appendix-II

This is to certify that Mr. /Ms. ..... of .....College/Institution worked as an intern as part of his/her M.Sc. course in Data Science of Kirti College, Mumbai.

The particulars of internship are given below:

Internship starting date: _____

Internship ending date: _____

Actual number of days worked: _____

Tentative number of hours worked: _____ Hours

Broad area of work: _____

A small description of work done by the intern during the period:

_____

_____

Signature: _____

Name:

Designation:

Contact details:

Email:

(Seal of the organization)

Appendix-III

(Proforma for the Evaluation of the intern by the industry mentor /to whom the intern was reporting in the organization)

Professional Evaluation of intern

Name of intern: _____

College/institution: _____

[Note: Give a score in the 1 to 5 scale by putting √ in the respective cells]

No	Particular	Excellent	Very Good	Good	Moderate	Satisfactory
1	Attendance & Punctuality					
2	Ability to work in a team					
3	Written and oral communication skills					
4	Problem solving skills					
5	Ability to grasp new concepts					
6	Technical skill in terms of technology, programming etc.					
7	Ability to complete the task					
8	Quality of overall work done					

Comments: _____

Signature: _____

Name :

Designation:

Contact details:

Email :

(Seal of the organization)