University of Mumbai

No. AAMS_UGS/ICC/2024-25/123

CIRCULAR:-

All the Principals of the Affiliated Colleges and Directors of the Recognized Institutions and the Head, University Departments in Faculty of Science & Technology are hereby informed that the recommendations made by the Board of Deans at its meeting held on 24th May, 2024 **vide** item No. 6.2 have been accepted by the Academic Council at its meeting held on 24th May, 2024 **vide** item No. 6.7 (N) and that in accordance therewith **M.S. (Data Analytics)** programme is introduced with exit option after one year **PG Diploma in Data Analytics** as per NEP 2020 in collaboration with St. Louis University, USA along with the Ordinance, Regulations and syllabus of **Sem I & II** as per appendix with effect from the academic year 2024-25.

(The said circular is available on the University's website www.mu.ac.in).

(Dr. Prasad Karande)
REGISTRAR

MUMBAI – 400 032
02nd September, 2024

To,
The Principals of the Affiliated Colleges, Directors of the recognized Institutions and the Head, University Departments.

A.C/6.7(N)/24/05/2024

Copy forwarded with Compliments for information to:-

1) The Chairman, Board of Deans,
2) The Dean, Faculty of Science & Technology,
3) The Chairman, Board of Studies,
4) The Director, Board of Examinations and Evaluation,
5) The Director, Department of Students Development,
6) The Director, Department of Information & Communication Technology,
7) The Director, Institute of Distance and Open Learning (IDOL Admin), Vidyanagari.
8) The Deputy Registrar, Admissions, Enrolment, Eligibility & Migration Department (AEM).

Copy forwarded for information and necessary action to :-	
1	The Deputy Registrar, (Admissions, Enrolment, Eligibility and Migration Dept)(AEM), dr@eligi.mu.ac.in
2	The Deputy Registrar, Result unit, Vidyanagari drresults@exam.mu.ac.in
3	The Deputy Registrar, Marks and Certificate Unit,. Vidyanagari dr.verification@mu.ac.in
4	The Deputy Registrar, Appointment Unit, Vidyanagari dr.appointment@exam.mu.ac.in
5	The Deputy Registrar, CAP Unit, Vidyanagari cap.exam@mu.ac.in
6	The Deputy Registrar, College Affiliations & Development Department (CAD), deputyregistrar.uni@gmail.com
7	The Deputy Registrar, PRO, Fort, (Publication Section), Pro@mu.ac.in
8	The Deputy Registrar, Executive Authorities Section (EA) eau120@fort.mu.ac.in He is requested to treat this as action taken report on the concerned resolution adopted by the Academic Council referred to the above circular.
9	The Deputy Registrar, Research Administration & Promotion Cell (RAPC), rapc@mu.ac.in
10	The Deputy Registrar, Academic Appointments & Quality Assurance (AAQA) dy.registrar.tau.fort.mu.ac.in ar.tau@fort.mu.ac.in
11	The Deputy Registrar, College Teachers Approval Unit (CTA), concolsection@gmail.com
12	The Deputy Registrars, Finance & Accounts Section, fort draccounts@fort.mu.ac.in
13	The Deputy Registrar, Election Section, Fort drelection@election.mu.ac.in
14	The Assistant Registrar, Administrative Sub-Campus Thane, thanesubcampus@mu.ac.in
15	The Assistant Registrar, School of Engg. & Applied Sciences, Kalyan, ar.seask@mu.ac.in
16	The Assistant Registrar, Ratnagiri Sub-centre, Ratnagiri, ratnagirisubcentre@gmail.com
17	The Director, Centre for Distance and Online Education (CDOE), Vidyanagari, director@idol.mu.ac.in
18	Director, Innovation, Incubation and Linkages, Dr. Sachin Laddha pinkumanno@gmail.com
19	Director, Department of Lifelong Learning and Extension (DLLE), Dlleuniversityofmumbai@gmail.com

Copy for information :-	
1	P.A to Hon'ble Vice-Chancellor, vice-chancellor@mu.ac.in
2	P.A to Pro-Vice-Chancellor pvc@fort.mu.ac.in
3	P.A to Registrar, registrar@fort.mu.ac.in
4	P.A to all Deans of all Faculties
5	P.A to Finance & Account Officers, (F & A.O), camu@accounts.mu.ac.in

To,

1	The Chairman, Board of Deans pvc@fort.mu.ac.in
2	**Faculty of Humanities,** **Dean** 1. Prof.Anil Singh Dranilsingh129@gmail.com **Associate Dean** 2. Dr.Suchitra Naik Naiksuchitra27@gmail.com 3.Prof.Manisha Karne mkarne@economics.mu.ac.in
	Faculty of Commerce & Management, **Dean** 1. Dr.Kavita Laghate kavitalaghate@jbims.mu.ac.in **Associate Dean** 2. Dr.Ravikant Balkrishna Sangurde Ravikant.s.@somaiya.edu 3. Prin.Kishori Bhagat kishoribhagat@rediffmail.com

	Faculty of Science & Technology **Dean** 1. Prof. Shivram Garje ssgarje@chem.mu.ac.in **Associate Dean** 2. Dr. Madhav R. Rajwade Madhavr64@gmail.com 3. Prin. Deven Shah sir.deven@gmail.com
	Faculty of Inter-Disciplinary Studies, **Dean** 1.Dr. Anil K. Singh aksingh@trcl.org.in **Associate Dean** 2.Prin.Chadrashekhar Ashok Chakradeo cachakradeo@gmail.com
3	Chairman, Board of Studies,
4	The Director, Board of Examinations and Evaluation, dboee@exam.mu.ac.in
5	The Director, Board of Students Development, dsd@mu.ac.in@gmail.com DSW direcotr@dsw.mu.ac.in
6	The Director, Department of Information & Communication Technology, director.dict@mu.ac.in

AC- 24/05/2024
Item No. 6.7 (N)

As Per NEP 2020

University of Mumbai

Title of the program

A- P.G. Diploma in Data Analytics _ **2024-25**
B- M.S. (Data Analytics) (Two Year)
C- M.S. (Data Analytics) (One Year) - **2027-28**

Syllabus for

Semester – Sem I & II

Ref: GR dated 16th May, 2023 for Credit Structure of PG

University of Mumbai

(As per NEP 2020)

Sr. No.	Heading	Particulars	
1	**Title of program** **O: _____A**	**A**	**P.G. Diploma in Data Analytics**
	O: _____B	**B**	**M.S. (Data Analytics) (Two Year)**
	O: _____C	**C**	**M.S. (Data Analytics) (One Year)**
2	**Eligibility** **O: _____A**	**A**	Any Undergraduate from the recognized institutions of Apex bodies both Nationally and Internationally.
	O: _____B	**B**	St. Louis University, USA (SLU) shall admit Program Students enrolled at UM's M.Sc.(Data Analytics) to SLU's M.S(Data Analytics) in their second year when they have: a. completed UM in Program Year One courses , i.e **Program A** b. earned a SLU-established minimum score on a SLU-approved English language proficiency examination (or equivalent approval per SLU policy); and c. earned the UM equivalent of a 3.0 cumulative grade point average at SLU in their UM courses in the Program. d. Completed 24 credits at the 1st year of M.Sc(Data Analytics)
	O: _____C	**C**	Graduate with 4 year U.G. Degree (Honours / Honours with Research) with Specialization in concerned subject or equivalent academic level 6.0 **OR** Graduate with four years UG Degree program with maximum credits required for award of Minor degree is allowed to take up the Post graduate program in Minor subject provided the student has acquired the required number of credits as prescribed by the concerned Board of Studies.

3	**Duration of program** **R: _____**	**A**	1 Year
		B	2 Year
		C	1 Year
4	**R: _____ Intake Capacity**	40 (A collaborative program at Department of Information technology, University of Mumbai with St. Louis University, USA))	
5	**R: _____Scheme of Examination**	NEP 50% Internal 50% External, Semester End Examination Individual Passing in Internal and External Examination	
6	**Standards of Passing** **R: _____**	40% in each component	
7	**Credit Structure** **R: _____**	Attached herewith	
8	**Semesters**	**A**	Sem. I & II
		B	Sem. I, II, III& IV
		C	Sem. I & II
9	**Program Academic Level**	**A**	6.0
		B	6.5
		C	6.5
10	**Pattern**	Semester	
11	**Status**	New	
12	**To be implemented from Academic Year Progressively**	A	2024-25
		B	
		C	2027-28

Sign of the
BOS Chairman
Dr. Mrs. R. Srivaramangai
Board of Studies in Data Science

Sign of the
Offg. Associate Dean
Dr. Madhav R. Rajwade
Faculty of Science & Technology

Sign of the
Offg. Dean
Prof. Shivram S. Garje
Faculty of Science & Technology

PREAMBLE

1) Introduction

A joint degree postgraduate program in Data Analytics is a specialized academic program that combines curriculum from two different disciplines, typically in collaboration between two departments or schools within a university. This program is designed to provide students with a comprehensive understanding of both data science and analytics, preparing them for successful careers in a data-driven world.

Students enrolled in a joint degree program in Data Analytics will typically study a combination of courses in statistics, computer science, data mining, machine learning, and business analytics. This interdisciplinary approach equips students with a diverse skill set that enables them to effectively collect, analyze, and interpret data to inform decision-making and drive organizational success.

By combining expertise from multiple fields, joint degree programs in Data Analytics offer students a more holistic and well-rounded education, preparing them to tackle complex challenges and make meaningful contributions in a variety of industries such as business, healthcare, finance, and technology. Graduates of these programs are well-equipped to pursue careers as data scientists, business analysts, data engineers, and more, making them highly sought-after professionals in today's data-driven economy.

2) Aims and Objectives

Aims: The primary aim of M.S(Data Analytics) is to provide a joint degree program with University of Mumbai, Mumbai, India and St. Louis University, USA, where the students will be enrolling for the program at University of Mumbai in the 1st year and complete their PG Diploma in Data Analytics and pursue their 2nd year at St. Louis University, USA to complete the program of M.S(Data Analytics). Those who complete from both the universities will be awarded with a joint degree convocation.

Objectives:

1. To impart knowledge of statistical methods, data mining techniques, and machine learning algorithms used in data analytics.
2. To enable students to apply data analytics tools and software to analyze real-world datasets.
3. To cultivate problem-solving skills through hands-on projects and case studies in data analytics.
4. To encourage students to critically evaluate the ethical and legal implications of data analytics.
5. To facilitate networking opportunities with industry professionals and organizations in the field of data analytics.
6. To support students in developing a research project or thesis that contributes to the advancement of knowledge in the field of data analytics.

3) Learning Outcomes

Upon completion of a postgraduate program in Data Analytics, learners can expect to achieve several learning outcomes, including:

1. Develop a strong foundation in statistical techniques and data analysis methods.
2. Gain hands-on experience with various data analytics tools and software.
3. Acquire advanced knowledge in data management, cleaning, and visualization.
4. Learn how to interpret and communicate data insights effectively to non-technical stakeholders.
5. Understand the ethical considerations and implications of working with data.
6. Gain expertise in machine learning algorithms and predictive analytics.
7. Develop critical thinking and problem-solving skills in the context of data analysis.
8. Enhance their ability to work collaboratively in a team environment on real-world data projects.
9. Prepare to pursue careers in data science, business intelligence, or other related fields that require strong analytical skills.
10. Continuously adapt to emerging technologies and industry trends in the field of data analytics.

4) Joint degree program description

5) For the international students entering the US on a student (F1) visa, to be considered eligible for Curricular Practical Training (CPT) and Optional Practical Training (OPT), 51% or higher of the overall curricular requirements in their program of study must be completed at the US institution (SLU). This means that the M.Sc. DA students need to take 18 (or more) credits of the required 33 credits at SLU, in the M.S. A program. Interdisciplinary knowledge: AI intersects with various disciplines, such as computer science, mathematics, cognitive science, and ethics. Expanding your knowledge beyond AI-specific topics by studying related disciplines will enable you to approach AI challenges from a broader perspective.

Credit Structure of the Program (Sem I, II, III, IV)

Credit Distribution Structure for Two Years/ One Year PG
M.S (Data Analytics)

<table>
<tr><th rowspan="2">Year</th><th rowspan="2">Level</th><th rowspan="2">Sem</th><th colspan="3">Major</th><th rowspan="2">RM</th><th rowspan="2">OJT/FP</th><th rowspan="2">RP</th><th rowspan="2">Cum. Cr.</th><th rowspan="2">Degree</th></tr>
<tr><th colspan="2">Mandatory</th><th>Electives</th></tr>
<tr><td rowspan="8">1</td><td rowspan="8">6.0</td><td rowspan="5">Sem I</td><td colspan="2">3*4+2*2</td><td>4</td><td>4</td><td rowspan="8">-</td><td rowspan="8">-</td><td rowspan="8">24</td><td rowspan="13">PG Diploma (after 3 Years Degree)</td></tr>
<tr><td>Fundamentals of Data Analysis and Technology (501)</td><td>TH 2</td><td rowspan="5">SPARK Technologies (506a)
2 TH+2PR
(OR)
SnowFlake(506b)
2 TH + 2 PR
(OR)
Hadoop (506c)
2 TH +2PR</td><td rowspan="5">Research Methodology (510)</td></tr>
<tr><td>Data Visualization Tools and Techniques((502)</td><td>PR 4</td></tr>
<tr><td>Statistical Methods for Data Analytics (503)</td><td>TH 4</td></tr>
<tr><td>Statistical Methods for Data Analytics using R(504)</td><td>PR 2</td></tr>
<tr><td>Big Data Analytics (505)</td><td>PR 4</td></tr>
<tr><td rowspan="6">Sem II</td><td colspan="2">3*4+2*2</td><td>4</td><td>-</td><td>(517)4</td><td>-</td><td>24</td></tr>
<tr><td>Regression Modelling (511)</td><td>TH 4</td><td rowspan="6">Geospatial Data Analytics (PR) (516a)
(OR)
Social Network Analysis (PR) (516b)
(OR)
Healthcare Analytics (PR) (516c)</td></tr>
<tr><td>Regression Modelling Practical(512)</td><td>PR 2</td></tr>
<tr><td>Datamining (513)</td><td>PR 4</td></tr>
<tr><td>Database Management</td><td>PR 4</td></tr>
<tr><td>Data Ethics and Privacy (515)</td><td>TH 2</td></tr>
<tr><td colspan="3">Cum. Cr. For PG Diploma</td><td colspan="2">32</td><td>8</td><td>4</td><td>4</td><td></td><td>48</td></tr>
<tr><td colspan="11">Exit Option: PG Diploma (48 credits) after Three Year UG Degree</td></tr>
</table>

Year	Level	Sem (2yr)	Major				RM	OJT/FP	RP	Cum. Cr.	Degree
2	6.5	Sem III	2*4+2*2 + 2			4	-	-	(607)4	24	PG Degree after 3-yr UG or PG Degree after 4-yr UG
			Time Series Analysis (601)	TH	4	Image and Video Analysis (606a) **(OR)** Fraud Detection and Risk Analytics (PR) (606b) **(OR)** Sports Data Analytics (PR) (606c)					
			Time Series Analysis Practical (602)	PR	2						
			Machine Learning For Data Analytics (603)	TH	4						
			Machine Learning For Data Analytics Practical (604)	PR	2						
			Optimization Techniques for Data Analytics (605)	TH	4						
		Sem IV	2*4+2*2			4	-	-	(616)6 Capstone	26	
			Deep learning (611)	TH	4	Financial Analysis(TH) (615a) **(OR)** Market Analysis and Customer Segmentation (615b) **(OR)** Business Intelligence (615c)					
			Deep Learning Practical (612)	PR	4						
			Natural Language Processing (613)	TH	4						
			Natural Language Processing Practical (614)	PR	4						
Cum. Cr. For 1 Yr PG Degree			32			8			10	50	
Cum. Cr. For 2 Yr PG Degree			64			16	4	4	10	98	

Syllabus

M.S (Data Analytics)

(Sem. I & II)

Sem-I

Programme Name: M.S (Data Analytics)

Course Code:501 [Mandatory] **Total Credits:** 04 (60 Lecture Hrs) **University assessment:** 50 marks	**Course Name:** Fundamentals of Data Analysis and Technology (Theory) **Total Marks:** 100 marks **College/Department assessment:** 50 marks

Prerequisite:

Basic understanding of Statistics and Excel

Course Objectives (CO)

CO1 To develop a comprehensive understanding of data, and its related terms

CO2 To understand the difference between various types of data, clean, and transform data.

CO3 To explore various Data Analysis Tools and evaluate

MODULE I: Introduction to Data Analysis	**(2 CREDITS)**
Unit I **What Is The Problem?** Questions a Data Head Should Ask, Understanding Why Data Project Fail, Working on Problems that Matter. **(Reference book 1 – Becoming a Data Head Wiley 2021)** **What is Data?** Data vs Information **(Reference book 1)** **Data, measurements, and data preprocessing** Data types, Statistics of data, Similarity and distance measures, Data quality**. (Reference book 2 Data Mining: Concepts and Techniques Morgan Kaufmann Publishers 2023)**	10 Hrs [OC1]
Unit II **Data Analytics in a Data-Driven World** What is Data Analytics? Types of Data Analytics, The Data Analytics Process **(Reference book 3 Fundamentals of Data Analytics by Russell Dawson 2nd Edition December 2023)** **Data Management Techniques for Data Collection and Storage** Data Management, Where Can Data Be Sourced and Collected? Types of Data Sources, Data Collection Tools and Software, Storing Huge Amounts of Data, Finding the Correct Source **(Reference book 3 Fundamentals of Data Analytics by Russell Dawson 2nd Edition December 2023)** Data cleaning, and data integration, Data transformation, and Dimensionality reduction **(Reference book 2 Data Mining: Concepts and Techniques Morgan Kaufmann Publishers 2023)**	30 Hrs [OC2]
MODULE II: Exploratory Data Analysis and Data Analysis Tools	

Unit III **Exploratory Data Analysis** Elements of Structure Data, Rectangular Data, Estimates of Location, Estimates of Variability, Exploring the Data Distribution, Exploring Binary and Categorical Data, Correlation, Exploring Two or More Variables **(Reference book 4 Practical Statistics for Data Scientists O'Reilly 2020 2nd Edition**)	10 Hrs [OC3]
Unit IV **Data Analysis Tools, Techniques, and Best Practices** The Data Professional's Toolkit, Tools and Software for Efficient Data Analysis, Data Analysis Pro Tips. **(Reference book 3 Fundamentals of Data Analytics by Russell Dawson 2nd Edition December 2023)**	10 Hrs [OC3]

References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Becoming a Data Head	Alex J. Gutman and Jordan Goldmeier	Wiley	First	2021
2	Data Mining Concepts and Techniques	Han Kamber	Morgan Kaufmann Publishers	Fourth	2023
3	Fundamentals of Data Analytics	Russell Dawson	Jws Publishers	Second	2023
4	Practical Statistics for Data Scientists	Peter Bruce, Andrew Bruce & Peter Gedeck	O'Reilly	Second	2020
5	Data Analytics 3 in 1- Beginners Guide	Benjamin Smith	Independe -ntly Published	First	2021

Course Outcomes (OCs)

Upon completion of this course, students will be able to:

OC1	Identify different types of data, sources, preprocess and transform data
OC2	Apply basic statistical concepts to summarize data into meaningful insights.
OC3	Conduct and Interpret results of exploratory data analysis.

Course Code:502 [Mandatory] **Total Credits:** 04 (120 Lecture Hrs) **University assessment: 50** marks	**Course Name:** Data Visualization Tools and Techniques Practical **Total Marks: 100** marks **College/Department assessment: 50** marks

Prerequisite:

Basic understanding of Visualization and Graphs, Logical Analysis of Data

Course Objectives(Cos):

OC1 To equip learners with a strong foundation in data visualization

OC2 To enable learners to create impactful visuals that communicate data stories effectively.

OC3 To explore tools and software for data visualization

Sr.No.	Practical List	4 Credits
	Module I Unit I	
1.	**Introduction to Data Visualization – Basics**	30 Hrs
a.	Introduction to Data Visualization	[OC1]
b.	Interpretation of Visual Information	
c.	Data Types and Use Cases	
d.	Chart Types and Use Cases	
e.	Analyze poorly designed charts and identify flaws	
f.	Redesign poorly designed charts and remove flaws identified	
g.	Create and explore visualization charts as per data using Excel or Google Sheets	
h	Experiment with formatting options to enhance the visual clarity and effectiveness of the charts	
i	Analyze real-world examples and scenarios for good and bad visualization	
	Module I- Unit II	
2.	**Python Visualization Libraries**	30 Hrs.
	Use datasets available in repositories like Kaggle or UCI Machine Learning Repository	[OC2,OC3]
a	Introduction to Python Visualization libraries and uses	
b.	Load a dataset and use plots to create a visualization distribution and relationships between variables	
c.	Use styling options to make the visualization informative.	
d.	Create multiple plots in a single frame to analyze different aspects of the data. Use grid layouts and shared axes.	
e.	Use seaborn to generate box plots, pair plots, heatmaps, etc. to analyze and explore the distribution of data, correlations, and categorical relationships.	
f.	Fit probability distribution namely normal and exponential to visualize the data and result using histogram and density plots.	
	Module II- Unit III	
3.	**Advanced Visualization Tools and Libraries**	30 Hrs.
a.	Create an interactive scatter plot.	[OC2,OC3]
b.	Visualize a time-series data with an animated line chart that shows trends and changes over time	
c.	Use Plotly to create geographic visualizations	
d.	Visualize three-dimensional data using a 3-D scatter plot.	
e.	Explore the interactive dashboard with visual elements	
f.	Implement Network visualization	
g.	Implement Hierarchical visualization	
h.	Implement Multivariate analysis and dimension reduction	
	Module II- Unit IV	
4.	**Case Studies**	30 Hrs.

a.	Sales Performance Analysis	[OC1, OC2,OC3]
b.	Customer Churn Production	
c	Website Traffic Analysis	
d.	Sentiment Analysis	
e.	Survival Patterns on the Titanic dataset	

References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Hands-on Data Visualization	Jack Dougherty and Ilya Ilyankou	O'Reilly	First	2021
2	Better Data Visualizations	Jonathan Schwabish	Columbia University Press	Second	2021
3	Interactive Data Visualization with Python	Luca Canali	Packt	Second	2021
4.	Fundamentals of Data Visualization	Claus O. Wilke	O'Reilly	First	2019

Course Outcomes (OCs)

Upon completion of this course, students will be able to:

1. Understand and Identify basic and different visualization charts for analysis.
2. Apply visual design principles to create effective visualization.
3. Demonstrate proficiency in using variety of data visualization tools and libraries.

Course Code: 503 [Mandatory] **Total Credits:** 04 (60 Lecture Hrs) **University assessment:** 50 marks	**Course Name:** Statistical Methods for Data Analytics **Total Marks:** 100 marks **College/Department assessment:** 50 marks

Pre-requisite: Basic Knowledge on Stattistics.

Course Objectives (COs):
To enable the students to:

CO1: Understand concepts of dynamic Routing protocols such as Routing Information Protocol (RIP), Open Shortest Path First (OSPF) and Border Gateway Protocol (BGP)

CO2: Acquire knowledge of MPLS, VPN Technology and benefits of NAT.

CO3: Understand the various standards for wireless networks.

CO4: Understand the functionality of each wireless component in cellular networks.

CO5: Understand the functionality of LTE and 5G

CO6: Understand the conceptual knowledge of Software Defined Networks (SDN) and Network Functions Virtualization (NFV)

MODULE I:	**(2 CREDITS)**
Unit I a) Introduction to Statistics and Data Analytics : Understanding the role of statistics in data analytics, Overview of data analytics and its applications, Exploring statistical concepts and terminology; b) Descriptive Statistics : Measures of central tendency: mean, median, mode, Measures of dispersion: range, variance, standard deviation, Data visualization techniques: histograms, box plots, scatter plots.	15 Hrs OC1, OC2, OC3
Unit II a) Probability Theory : Basic concepts of probability, Probability distributions: normal, binomial, and Poisson distributions, Random variables and their properties; Inferential Statistics : Estimation: point estimation and interval estimation, Hypothesis testing: significance levels, p-values, and confidence intervals, Understanding sampling methods and their implications. **b)** Correlation and Regression Analysis : Pearson correlation coefficient, Simple linear regression: model, estimation, and interpretation, Multiple linear regression: model building and diagnostics; Statistical Modeling : Introduction to statistical modeling, Model selection techniques: AIC, BIC, cross-validation, Model diagnostics and validation	15 Hrs OC4, OC5
MODULE II:	**(2 CREDITS)**
a) Experimental Design and Analysis of Variance (ANOVA) : Principles of experimental design, One-way and two-way ANOVA, Post-hoc tests and multiple comparisons ; b) Time Series Analysis : Time series data and its characteristics, Forecasting methods: moving averages, exponential smoothing, ARIMA models, Seasonal decomposition and trend analysis.	15 Hrs OC6, OC8
Unit IV a) Non-parametric Statistics : Wilcoxon rank-sum test, Kruskal-Wallis test, Other non-parametric methods and their applications; Statistical Software and Tools for Data Analytics : Introduction to statistical software packages (e.g., R, Python, SAS), Best practices for data analysis and presentation	15 Hrs OC7, OC9, OC10, OC11

b) Ethical and Legal Aspects of Data Analytics: Privacy, security, and ethical considerations in data analytics, Regulatory compliance and legal requirements Data governance and responsible use of data.	

References:

1. "Introduction to the Practice of Statistics" by David S. Moore, George P. McCabe, and Bruce A. Craig.
2. "Statistical Methods for Data Analysis" by John Fox
3. "Practical Statistics for Data Scientists: 50 Essential Concepts" by Peter Bruce and Andrew Bruce
4. "Statistics in Plain English" by Timothy C. Urdan

Course Outcomes (OCs):
Upon completing this course, the student will be able to:

OC1	Understand Fundamental Statistical Concepts
OC2	Apply Statistical Techniques
OC3	Implement Data Collection and Organization
OC4	Perform Inferential Statistics
OC5	Interpret Correlation and Regression Analyses
OC6	Conduct Analysis of Variance (ANOVA)
OC7	Apply Non-parametric Methods
OC8	Analyze Time Series Data
OC9	Conduct Real-world Data Analysis Projects
OC10	Utilize Statistical Software
OC11	Adhere to Ethical Standards

Course Code: 504 [Mandatory] **Total Credits:** 02 (60 Lecture Hrs) **University assessment:** 25 marks	**Course Name**: Statistical Methods for Data Analytics using R(Practical) **Total Marks:** 50 **College/Department assessment:** 25 marks

Pre-requisite:

1. Basic knowledge of statistics and basic knowledge in programming preferred.

Course Objectives (COs):

CO1. Introduction to R and RStudio
CO2. Data Import and Manipulation
CO3. Implementing Descriptive Statistics
CO4. Implementing Probability Distributions and Sampling:
CO5. Implementing Hypothesis Testing and Confidence Intervals:
CO6. Problem solving using Correlation and Regression Analysis in R
CO7. Analysis of Variance (ANOVA) in R
CO8. Non-parametric Methods in R
CO9. Time Series Analysis in R
CO10. Multivariate Statistical Analysis in R
CO11. Real-world Applications and Projects:
CO12. Ethical Use of Data and Reporting using R.

MODULE I:	**(2 CREDITS)**
Unit I a) Introduction to R and RStudio Setting up R and RStudio Basic operations in R: data types, variables, and functions Working with RStudio: scripts, console, and environment b) Data Import and Manipulation Importing data from various sources (CSV, Excel, databases) Data cleaning, transformation, and manipulation using dplyr and tidyr packages Handling missing data and outliers c) Exploratory Data Analysis (EDA) using R Descriptive statistics and visualization: summary statistics, histograms, boxplots, and scatter plots Exploring relationships between variables using R's ggplot2 package d) Probability Distributions and Sampling Simulating probability distributions (e.g., normal, binomial) using R Generating random samples and understanding sampling distributions e) Hypothesis Testing and Confidence Intervals Performing t-tests, z-tests, and chi-square tests for hypothesis testing Constructing confidence intervals for population parameters f) Correlation and Regression Analysis in R Calculating correlation coefficients and conducting simple linear regression Visualizing regression models using R's ggplot2 package	30 Hrs OC1 – OC5
MODULE II:	**(2 CREDITS)**
Unit II a) Analysis of Variance (ANOVA) in R Applying one-way and two-way ANOVA using built-in functions and ANOVA packages Post hoc tests and interpretation of ANOVA results b) Non-parametric Methods in R Implementing non-parametric tests such as Wilcoxon rank sum test and Kruskal-Wallis test Comparing non-parametric methods with their parametric counterparts c) Time Series Analysis in R	30 Hrs OC6 – OC10

Time series visualization, trend analysis, and seasonality detection Forecasting using time series models and evaluating forecast accuracy d) Multivariate Statistical Analysis in R Conducting principal component analysis (PCA) and factor analysis Applying clustering techniques using R's cluster package e) Real-world Applications and Projects Working on real-world data analytics projects using R and presenting the findings Utilizing statistical methods and R programming to solve practical data analysis challenges **f)** Ethical Use of Data and Reporting	

References:

1. An Introduction to Statistical Learning: with Applications in R" by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani
2. "Practical Statistics for Data Scientists: 50 Essential Concepts" by Peter Bruce and Andrew Bruce

Upon completing this course, the student will be able to:

Course Otucomes(OCs):

OC1	Utilize R Programming for Data Analytics
OC2	Apply Descriptive Statistics and Visualization Techniques in R
OC3	Perform Probability Distributions and Sampling Simulations
OC4	Conduct Hypothesis Testing and Construct Confidence Intervals in R
OC5	Utilize Correlation and Regression Analysis Techniques in R
OC6	Apply Analysis of Variance (ANOVA) and Non-parametric Methods in R
OC7	Analyze Time Series Data and Conduct Forecasting in R
OC8	Perform Multivariate Statistical Analysis using R
OC9	Participate in Real-world Data Analytics Projects
OC10	Evaluate and Interpret Ethical Considerations in Data Analytics

Course Code: 505 [Mandatory] **Total Credits:** 04 (120 Lecture Hrs) **University assessment:** 50 marks	**Course Name**: Big Data Analytics(Practical) **Total Marks:** 100 marks **College/Department assessment:** 50 marks

Pre requisite:

a) Programming Skills, Database knowledge
b) Statistics and Mathematics
c) Datamining, Machine learning and distributed systems

Course Objectives (COs):

a) Understand the key issues in existing technologies leading to the evolution of big data analytics and its associated applications in business analytics.
b) Understand the architecture of Hadoop and its components
c) Explore map reduce framework and optimize its jobs.
d) Explore popular Hadoop tools like Hive, Pig, Hbase, Spark

MODULE I:	**(2 CREDITS)**
Unit 1: Introduction to Big Data a) **Introduction to Big data:** Evolution of Big Data, Definition of Big Data, Challenges with Big Data, Traditional Business Intelligence (BI) versus Big Data. b) **Big data analytics:** Classification of Analytics, Importance and challenges facing big data, Terminologies Used in Big Data Environments, The Big Data Technology Landscape, NoSQL c) **Introduction to Hadoop:** Introducing Hadoop, RDBMS versus Hadoop, Distributed Computing Challenges, History and overview of Hadoop, Use Case of Hadoop, Hadoop Distributors, Processing Data with Hadoop, Interacting with Hadoop Ecosystem	305 Hrs [OC1, OC2]
Unit 2: Hadoop Ecosystem a) **Avro:** Data Types and Schema, Serialization, Datafiles, Schema Resolution b) **Flume:** Transactions and Reliability, Fan Out, Distribution c) **Sqoop:** Sqoop Connectors, Imports, Working with Imported Data, Importing Large Objects d) **Pig**: Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators e) **Hive**: Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying, Data and User Defined Functions. f) **Spark:** Spark Applications, Jobs, Stages, and Tasks, Resilient Distributed Datasets, Anatomy of a Spark Job Run g) **Hbase:** HBasics, Concepts, Clients, Example, Hbase Versus RDBMS h) **ZooKeeper:** Group Membership, ZooKeeper Service i) Oozie, Solr, Storm	30 Hrs [OC3, OC4]
MODULE II :	**(2 CREDITS)**
Unit 3 : Hadoop Tools Part 1 a) **Installation** of PIG. b) **Pig commands:** Write pig latin scripts sort, group, join, project, and filter your data. c) **Pig latin modes, programs:** i. Run the Pig Latin Scripts to find Word Count	30 Hrs [OC4, OC6]

ii. Run the Pig Latin Scripts to find a max temp for each and every year. d) **Installation** of HIVE. e) **Hive operations:** Use Hive to create, alter, and drop databases, tables, views, functions, and indexes.	
Unit 4 : Hadoop Tools Part 2 a) **MongoDB:** Install MongoDB and manipulate it using Python b) **Apache Spark:** Install, configure and run Apache Spark. Create & transform RDDs c) **Flume:** Install, configure and run Apache Flume. Configure Source, Sink & Flume Agent	30 Hrs [OC4, OC6]

References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Big Data and Analytics	Seema Acharya, Subhashini Chellappan	Wiley	2nd edition	2019
2.	Hadoop: The Definitive Guide	Tom White	O'reily	4th Edition	2012
3.	Big Data Science and Analytics – A hands on approach	Arshdeep Bagha, Vijay Madisetti		1st edition	2019
4.	Data Analytics with Hadoop- An Introduction for Data Scientists	Benjamin Bengfort and Jenny Kim	O'reily	1st edition	2016

Course Outcomes(OCs)

OC 1. Familiarize & understand what is Big Data, its evolution, describe the elements of big data- volume, variety, velocity and veracity & understand the importance and challenges of big data.

OC 2. Understand the importance and challenges of big data & define big data analytics advantages and its applications.

OC 3. Understand architectural components involved in hadoop ecosystem, describe in detail about Distributed file system, Understand the concept of Hadoop cluster architecture.

OC 4. Understand & explain the architecture frameworks like Pig, Hive, Hbase, Sqoop, Avro, Flume, Spark, Zookeeper

OC 5. Familiarize with frameworks like Oozie, Solr, Storm

OC 6 Implement the concepts of Pig, Hive, MongoDb, Spark, Flume

Course Code: 506a [Elective] **Total Credits:** 04 (120 Lecture Hrs) **University assessment:** 50 marks	**Course Name**: Spark Technologies(Practical) **Total Marks:** 100 marks **College/Department assessment:** 50 marks

Pre-requisite:

a. Fundamentals of Programming, Databases and mathematical background.

Course Objectives (COs):

To enable the students to:

CO1. Develop Proficiency in Apache Spark
CO2. Real-time Data Processing and Analytics
CO3. Machine Learning with Spark MLlib
CO4. Graph Processing and Analysis
CO5. Integration with NoSQL Databases
CO6. Performance Optimization and Scalability
CO7. Statistical Methods for Data Analysis
CO8. Practical Data Analysis and Visualization

MODULE I:	**(2 CREDITS)**
Unit 1: a) Setting Up Spark Environment Installation of Spark on a local machine or cluster Configuring Spark clusters for performance b) Working with RDDs Creating and manipulating Resilient Distributed Datasets (RDDs) Implementing transformations and actions on RDDs	20 Hrs [OC1, OC2]
Unit 2: a) Data Processing with DataFrames Loading data into Spark DataFrames Performing data transformations and analysis using DataFrames b) Real-time Data Processing with Spark Streaming Setting up Spark Streaming job Processing streaming data from sources like Kafka	20 Hrs [OC2]
MODULE II :	**(2 CREDITS)**
Unit 3: a) Machine Learning Modeling Implementing machine learning algorithms with Spark MLlib Evaluating models and tuning hyperparameters b) Graph Analytics with GraphX Creating and analyzing graphs using GraphX Implementing graph algorithms for various use cases c) Integration with NoSQL Databases Connecting Spark with NoSQL databases like MongoDB or CassandraReading and writing data between Spark and NoSQL databases	40 Hrs [OC3, OC4, OC5]
Unit 4: a) Performance Optimization Monitoring and optimizing Spark application performance b) Understanding memory management and caching in Spark One Mini Task of Implementing a Big Data solution using Apache Spark	40 Hrs [OC6]

References:

1. "Big Data Analytics with Spark: A Practitioner's Guide to Using Spark for Large Scale Data Analysis
2. Advanced Analytics with Spark: Patterns for Learning from Data at Scale" by Sandy Ryza, Uri Laserson, Sean Owen, and Josh Wills

Course Outcome (OCs)

Upon completing this course, the student will be able to:

OC1	Proficiency in Apache Spark and Related Technologies
OC2	Real-time Data Processing and Analytics Skills
OC3	Machine Learning Proficiency with Spark MLlib
OC4	Graph Processing and Analysis Competence
OC5	Integration with NoSQL Databases and Performance Optimization
OC6	Practical Data Analysis and Visualization Proficiency

Course Code: 506b [Elective] **Total Credits:** 04 (120 Lecture Hrs) **University assessment:** 50 marks	**Course Name**: SnowFlake Practical **Total Marks:** 100 marks **College/Department assessment:** 50 marks

Pre-requisite: basics of networking and Cloud Computing.

Course Objectives (COs)

To enable the students to learn:

CO1. Introduction to Snowflake Cloud Platform
CO2. Navigating the Snowflake Interface
CO3. Data Loading and Management
CO4. Querying and Analyzing Data
CO5. Data Security and Governance
CO6. Performance Optimization
CO7. Integration and ETL Processes
CO8. Monitoring and Troubleshooting
CO9. Scalability and Resource Management
CO10. Real-World Use Cases and Best Practices

MODULE I:	**(2 CREDITS)**
Unit I a) Introduction to Snowflake Overview of Snowflake's architecture and key features Understanding Snowflake's unique capabilities in data warehousing and analytics Setting up a Snowflake account and navigating the user interface b) Data Ingestion and Integration Methods for ingesting data into Snowflake: Snowpipe, batch ingestion, and streaming Integrating data from various sources such as S3, Azure Blob Storage, and Google Cloud Storage Hands-on exercise: Data ingestion and integration using Snowflake	30 Hrs OC1
Unit II a) Data Warehousing and Modeling Designing and implementing scalable data warehousing solutions in Snowflake Data modeling best practices for optimizing querying and analytics Hands-on exercise: Creating and optimizing data models in Snowflake	30 Hrs OC2, OC3

b) Querying and Performance Optimization Advanced SQL querying in Snowflake, including window functions and common table expressions Techniques for optimizing query performance and resource utilization Hands-on exercise: Performance optimization of complex queries in Snowflake	
MODULE II:	**(2 CREDITS)**
Unit III a) Security and Governance Implementing robust security measures in Snowflake, including user access management and data encryption Ensuring data governance and compliance with industry regulations Hands-on exercise: Configuring security and governance settings in Snowflake b) Data Transformation and Analytics Applying functions and procedures for data transformation within Snowflake Performing advanced analytics using Snowflake's built-in analytical functions and tools Hands-on exercise: Data transformation and advanced analytics in Snowflake	30 Hrs OC3, OC4
Unit IV a) Snowflake Architecture and Infrastructure Understanding the architecture of Snowflake, including virtual warehouses and storage layers Configuring and optimizing Snowflake instances for specific use cases and performance requirements Hands-on exercise: Configuring virtual warehouses and storage in Snowflake b) Data Pipelines and Automation Designing and implementing data pipelines for automating ETL/ELT processes in Snowflake Scheduling and orchestrating data workflows efficiently using Snowflake's features and third-party tools Hands-on exercise: Building data pipelines and automating workflows in Snowflake	30 Hrs OC5, OC6

References:

1) Snowflake: The Definitive Guide by Joyce Kay Avila
2) "Snowflake DBA: Comprehensive Guide to Master Snowflake Cloud Data Platform, Author: Martin Muller and Justin Somers

Course Outcomes (OCs)

Upon completing this course, the student will be able to:

OC1	Data Ingestion and Integration
OC2	Data Warehousing and Modeling
OC3	Querying and Performance Optimization
OC4	Security and Governance
OC5	Data Transformation and Analytics
OC6	Snowflake Architecture and Infrastructure
OC7	Data Pipelines and Automation

Course Code: 506c [Elective] **Total Credits:** 04 (120 Lecture Hrs) **University assessment:** 50 marks	**Course Name**: Hadoop Practical **Total Marks:** 100 marks **College/Department assessment:** 50 marks

Prerequisite:

1. Sound knowledge of Python, Java
2. Sound knowledge of concepts in probability, statistics & mathematics
3. Sound knowledge in using Linux OS.

Course Objectives (COs):

To enable the students to:

1. Optimize business decisions and create competitive advantage with Big data analytics
2. Practice java concepts required for developing map reduce programs.
3. Impart the architectural concepts of Hadoop and introducing map reduce paradigm.
4. Implement best practices for Hadoop development.

Prac No	Practical Description	4 CREDITS (120 hrs)
1	Installation of VMWare to setup the Hadoop environment and its ecosystems	20 [OC1]
2	**Hadoop Modes** a. Perform setting up and Installing Hadoop in its three operating modes.	20 [OC2]

	i. Standalone. ii. Pseudo distributed. iii. Fully distributed. b. Use web based tools to monitor your Hadoop setup	
3	Implementing the basic commands of LINUX Operating System – File/Directory creation, deletion, update operations.	20 [OC3]
4	**File Management In Hadoop** Implement the following file management tasks in Hadoop: i. Adding files and directories ii. Retrieving files iii. Deleting files	20 [OC4]
5	**Mapreduce Program 1** Hadoop Programming: Word Count Mapreduce Program Using Eclipse	10 [OC5]
6	**Mapreduce Program 2** Implementing Matrix Multiplication Using One Map-Reduce Step	15 [OC6]
7	**Mapreduce Program 3** Write a Map Reduce program that mines weather data.	15 [OC7]

References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
5.	Big Data and Analytics	Seema Acharya, Subhashini Chellappan	Wiley	2nd edition	2019
6.	Hadoop: The Definitive Guide	Tom White	O'reily	4th Edition	2012
7.	Big Data Science and Analytics – A hands on approach	Arshdeep Bagha, Vijay Madisetti		1st edition	2019
8.	Data Analytics with Hadoop- An Introduction for Data Scientists	Benjamin Bengfort and Jenny Kim	O'reily	1st edition	2016

Course Outcomes (OCs)

Upon completing this course, the student will be able to:

1. Understand & Implement the installation of VMWare
2. Understand and Installing Hadoop in its three operating modes.
3. Implementing the basic commands of LINUX Operating System
4. Implement the file management tasks in Hadoop.
5. Understand Map Reduce Paradigm.
6. Apply Map Reduce program that mines weather data.
7. Implement matrix multiplication with Hadoop MapReduce

Course Code: 510 **Total Credits:** 04 (60 Lecture Hrs) **University assessment:** 50 marks	**Course Name**: Research Methodology **Total Marks:** 100 marks **College/Department assessment:** 50marks

Pre requisite:

Basic programming skills, orientation towards research and conceptual understanding of IT subjects

Course Objectives (COs)

To enable the students to:

CO1. Know basics of how research problems are defined, research methods are adopted and/or developed,
research is undertaken
CO2. Make understand how research results are communicated to the peers.
CO3. Learn research methods, some of which are general in nature and the remaining specific to the field of
Information Technology and the specialization.

MODULE I:	**(2 CREDITS)**
Unit 1: **Research Methodology and Problem Identification and Formulation:** Meaning and objectives, motivation of research, types of research, research methods v/s methodology, research and scientific methods, research process and stages of research, defining and formulating the research problem, technique involved in defining a problem, importance of literature review in defining a problem, role of literature review, ways to perform literature review, methods to find open problem and research problems, critical literature review, identifying gap areas from literature study, hypothesis building	15 Hrs [OC1]
Unit 2: **Research Design and Data Collection and Analysis:** Need of research design, concepts related to research design, different research designs, research plan, basic principles of experimental design and setup, collection of primary data, observation methods, interview methods, collection of data through questionnaire and schedules, collection of secondary data, selection of appropriate method for data collection, case study method, guidelines for developing questionnaire, successful interview, survey v/s experiment, processing and data analysis, use of statistical packages, measure of asymmetries and other measures. Fieldwork-The Nature of Field Work, Selection and Training of Investigators, Sampling Frame and Sample Selection, Field Operation, Field Administration.	15 Hrs [OC2, OC3]
MODULE II:	**(2 CREDITS)**
Unit 3: **Probability Distribution and Hypothesis Testing:** Sampling and probability distribution, definitions and basic concepts of hypothesis testing, procedures of hypothesis testing, flow diagram for hypothesis testing, test of hypothesis, important parametric test, hypothesis testing of mean, proportion, tests for equality of mean and variances of two population, confidence interval, z-test, and X2 test for goodness to fit, limitation of test of hypothesis. **Analysis of Variance and Covariance:** Basic principle of Analysis of Variance, ANOVA Technique, Setting up Analysis of Variance Table, short-cut method for one- way ANOVA, Coding method, Two-way ANOVA, ANOVA in Latin-square design, analysis of co-variance (ANCOVA), assumptions in ANCOVA.	20 Hrs [OC4]

Unit 4: **Academic Ethics:** Plagiarism, exposure on anti-plagiarism tools. **Technical Writing and IPR:** Academic writing, sources of information, assessment of quality of journals and articles, writing scientific report, structure and component of research report, types of report – technical reports and thesis, SCOPUS Index, citations, search engines beyond google, impact factor, H-Index. **IPR:** What is IPR?, importance of patents, types of IPR, process of patent.	10 Hrs [OC5, OC6, OC7]

References:

1) Dawson, Catherine, 2002, Practical Research Methods, New Delhi, UBS Publishers' Distributors.
2) Kothari, C.R.,1985, Research Methodology-Methods and Techniques, New Delhi, Wiley Eastern Limited.
3) Kumar, Ranjit, 2005, Research Methodology-A Step-by-Step Guide for Beginners, (2nd.ed), Singapore, Pearson Education.
4) Neeraj Pandey, Intellectual Property Rights ,1st Edition, PHI
5) Shrivastava, Shenoy& Sharma, Quantitative Techniques for Managerial Decisions, Wiley
6) Goode W J &Hatt P K, Methods in social research, McGraw Hill
7) Basic Computer Science and Communication Engineering – R. Rajaram (SCITECH)

Course Outcomes (OCs)

Upon completing this course, the student will be able to:

OC1: Basic understanding of research and how to formulate a research problem

OC2: Understand and develop methodological design for the research problem

OC3: Identify the required data and use data collection methods for acquiring data

OC4: Set hypothesis for the given research problem and apply testing methods

OC5: Follow the research ethics

OC6: Write research proposals, documentations related with research

OC7: Understand and apply IPR and patent filing

Sem-II

Course Code: 511 [Mandatory] **Total Credits:** 04 (60 Lecture Hrs) **University assessment**: 50 marks	**Course Name:** Regression Modelling **Total Marks:** 100 marks **College/Department assessment**: 50 marks

Pre requisite:

1. Knowledge of standard statistical methodology, such as ordinary least squares linear regression methods, including t-tests
2. Knowledge of constructions of standard confidence intervals, and standard distributions such as the univariate and multivariate normal distributions.

Course Objectives (COs)

To enable the students to:

CO1. Fundamental concepts and significance of regression analysis
CO2. Principles and application of simple linear regression
CO3. Construct and interpret multiple linear regression models
CO4. Gain knowledge of generalized least squares estimation methods
CO5. Knowledge and application on all regression models

MODULE I:	**(2 CREDITS)**
Unit 1: a) Introduction, Simple Linear Regression Analysis b) Multiple Linear Regression Model, Model Adequacy Checking	15 Hrs [OC1]
Unit 2: a) Transformation and Weighting to Correct Model Inadequacies, Diagnostic for Leverage and Influence b) Generalized and Weighted Least Squares Estimation, Indicator Variables	15 Hrs [OC2, OC4]
MODULE II:	**(2 CREDITS)**
Unit 3: a) Multicollinearity, Heteroskedasticity, b) Autocorrelation. Polynomial Regression Models	20 Hrs [OC3]
Unit 4: a) Variable Selection and Model Building, Logistic Regression Models b) Poisson Regression Models, Generalized Linear Models	10 Hrs [OC5, OC6]

References:

1. Draper, N. R., and Smith, H. (1998), Applied Regression Analysis (3rd ed.), New York: Wiley.
2. Montgomery, D. C., Peck, E. A., and Vining, G. (2001), Introduction to Linear Regression Analysis (3rd ed.), Hoboken, NJ: Wiley.

Course Outcomes (OCs)

Gain knowledge of generalized least squares estimation methods

Upon completing this course, the student will be able to:

OC1	Gain knowledge of generalized least squares estimation methods
OC2	Gain competence in conducting and interpreting simple linear regression analyses.
OC3	Assess the significance of multiple linear regression models.
OC4	Incorporating indicator variables
OC5	Apply all regression models
OC6	Understanding the flexibility of Generalized linear models

Course Code:512 [Mandatory]	**Course Name:** Regression Modelling (Practical)
Total Credits: 02 (60 Lecture Hrs)	**Total Marks:** 50 marks
University assessment: 25 marks	**College/Department assessment:** 25 marks

Pre requisite:

1. Knowledge of standard statistical methodology, such as ordinary least squares linear regression methods, including t-tests
2. Knowledge of constructions of standard confidence intervals, and standard distributions such as the univariate and multivariate normal distributions.
3. Programming skills

Course Objectives (COs)

To enable the students to perform and apply:

CO1. Fundamental concepts and significance of regression analysis
CO2. Principles and application of simple linear regression
CO3. Construct and interpret multiple linear regression models
CO4. Gain knowledge of generalized least squares estimation methods
CO5. Knowledge and application on all regression models

MODULE I:	**(2 CREDITS)**
Unit 1: a) Problem solving based on i. Simple Linear Regression Analysis ii. Multiple Linear Regression Model, iii. Model Adequacy Checking	15 Hrs [OC1]
Unit 2: a) Problem solving based on i. Transformation and Weighting to Correct Model Inadequacies, ii. Diagnostic for Leverage and Influence iii. Generalized and Weighted Least Squares Estimation, iv. Indicator Variables	15 Hrs [OC2, OC4]
MODULE II:	**(2 CREDITS)**
Unit 3: a) Problem solving based on i. Multicollinearity, ii. Heteroskedasticity, iii. Autocorrelation. iv. Polynomial Regression Models	20 Hrs [OC3]
Unit 4: a) Problem solving based on i. Variable Selection and Model Building, ii. Logistic Regression Models iii. Poisson Regression Models, iv. Generalized Linear Models * **All implementations can be done using R/Python or any other tool**.	10 Hrs [OC5, OC6]

References:

1. Draper, N. R., and Smith, H. (1998), Applied Regression Analysis (3rd ed.), New York: Wiley.
2. Montgomery, D. C., Peck, E. A., and Vining, G. (2001), Introduction to Linear Regression Analysis (3rd ed.), Hoboken, NJ: Wiley.

Course Outcomes (OCs)

Gain knowledge of generalized least squares estimation methods

Upon completing this course, the student will be able to solve problems for application of :

OC1	Generalized least squares estimation methods
OC2	Conducting and interpreting simple linear regression analyses.
OC3	Assessing the significance of multiple linear regression models.
OC4	Incorporating indicator variables
OC5	Apply all regression models
OC6	Understanding the flexibility of generalized linear models

Course Code: 513 [Mandatory]	**Course Name:** Data Mining Practical
Total Credits: 04 (120 Lecture Hrs)	**Total Marks:** 100 marks
University assessment: 50 marks	**College/Department assessment:** 50 marks

Pre requisite:

1. Basic knowledge of Python programming
2. Familiarity with data manipulation and analysis

Course Objectives (COs)

To enable the students to perform and apply:

CO1 Understand the fundamentals of data mining
CO2 Explore different data mining techniques
CO3 Learn how to preprocess and clean data for analysis
CO4 Implement various machine learning algorithms for data mining
CO5 Gain hands-on experience in data visualization and interpretation

MODULE I:	**(2 CREDITS)**
Unit 1: a) Introduction to Data Mining i. Overview of data mining concepts ii. Introduction to Python libraries for data mining (NumPy, Pandas) iii. Data preprocessing techniques	25 Hrs [OC1]
Unit 2: a) Data Exploration i. Exploratory data analysis (EDA) ii. Data visualization with Matplotlib and Seaborn iii. Understanding data distributions and patterns	25 Hrs [OC2, OC3]
MODULE II:	**(2 CREDITS)**
Unit 3: a) Supervised Learning Algorithms i. Introduction to supervised learning ii. Linear regression and logistic regression iii. Decision trees and random forests	30 Hrs [OC4]
Unit 4: a) Unsupervised Learning Algorithms i. Introduction to unsupervised learning ii. Clustering algorithms: K-means, hierarchical clustering iii. Dimensionality reduction techniques b) Advanced Topics i. Association rule mining ii. Text mining and sentiment analysis	40 Hrs [OC5, OC6]

iii. Model evaluation and tuning * **All implementations can be done using R/Python or any other tool**.	

References:

1. Modern Data Mining with Python: A risk-managed approach to developing and deploying explainable and efficient algorithms using ModelOps by Dushyant Singh Sengar , Vikash Chandra, BPB Publications
2. Online tutorials and forums for additional support

Course Outcomes (OCs)

Upon completing this course, the student will be able to:

OC1	Understanding of Data Mining Fundamentals
OC2	Proficiency in Data Preprocessing and Cleaning
OC3	Exploratory Data Analysis Skills
OC4	Application of Supervised Learning Algorithms
OC5	Understanding of Unsupervised Learning Techniques
OC6	Knowledge of Advanced Data Mining Topics

Course Code:514 [Mandatory] **Total Credits:** 02 (120 Lecture Hrs) **University assessment:** 50 marks	**Course Name:** Database Management Systems (Practical) **Total Marks:** 100 marks **College/Department assessment:** 50 marks

Pre requisite:

1. Basic Understanding of Database Concepts
2. Knowledge of SQL (Structured Query Language)
3. Basic Programming Skills

Course Objectives (COs)

To enable the students to perform and apply:
Understand the Fundamentals of Database Management:

CO1 Master SQL Fundamentals
CO2 Learn Database Design and Implementation
CO3 Explore Relational Database Management Systems (RDBMS)
CO4 Hands-On Experience with Database Operations
CO5 Database Security and Integrit
CO6 Data Analysis and Interpretation
CO7 Optimize Database Performance
CO8 Troubleshooting and Problem-Solving:
CO9 Prepare for Real-World Database Management

MODULE I:	**(2 CREDITS)**
Unit 1: a) Introduction to Database Concepts i. Overview of database management systems ii. Relational database fundamentals iii. Data modeling and normalization	30 Hrs [OC1]
Unit 2: a) SQL Fundamentals i. Introduction to SQL (Structured Query Language) ii. Data manipulation language (DML) operations: SELECT, INSERT, UPDATE, DELETE iii. Data definition language (DDL) operations: CREATE, ALTER, DROP	30 Hrs [OC2]
MODULE II:	**(2 CREDITS)**
Unit 3: a) Database Design and Implementation i. Entity-Relationship (ER) modeling ii. Normalization techniques iii. Designing efficient and normalized databases	40 Hrs [OC3, OC4]

a) Relational Database Management Systems (RDBMS) i. Understanding RDBMS architecture ii. Hands-on experience with MySQL, PostgreSQL, or Oracle iii. Managing database objects: tables, views, indexes	
Unit 4: a) Data Security and Integrity i. Enforcing data integrity constraints ii. Access control and user privileges iii. Backup and recovery strategies b) Creation of a small DBMS for any real-world problem	20 Hrs [OC5, OC6]

References:

1. **"Database System Concepts" by Abraham Silberschatz, Henry F. Korth, and S. Sudarshan**
2. **"Learning SQL: Generate, Manipulate, and Retrieve Data" by Alan Beaulieu**
3. **Any reference of RDBMS like MYSQL, ORACLE, MS SQL SERVER etc.**

Course Outcomes (OCs)

Upon completing this course, the student will be able to:

OC1	Comprehensive Understanding of Database Management Systems (DBMS)
OC2	Proficiency in SQL
OC3	Effective Database Design Skills
OC4	Hands-On Experience with RDBMS Platforms
OC5	Data Security and Integrity
OC6	Real-World Database Management Preparation

Course Code: 515 [Mandatory] Total Credits: 02 (30 Lecture Hrs) University assessment: 25 marks	Course Name: Data Ethics and Privacy Total Marks: 50 marks College/Department assessment: 25 marks

Pre-requisite:

1. Basic understanding of data privacy regulations
2. Familiarity with data management and ethical considerations in technology

Course Objectives (COs):

To enable the students to:

CO1	Understand the ethical implications of data collection and usage
CO2	Explore privacy regulations and their impact on data handling
CO3	Gain insights into the ethical considerations of data-driven decision-making
CO4	Develop strategies for ethical data management and protection of privacy rights

MODULE I:	**(2 CREDITS)**
Unit I a) Introduction to Data Ethics and Privacy i. Overview of ethical considerations in data management ii. Historical perspective on data ethics and privacy regulations b) Data Collection and Consent i. Ethical considerations in data collection practices ii. Consent management and informed consent principles c) Data Usage and Privacy Rights i. Ethical implications of data usage and privacy rights ii. Understanding the rights of data subjects under various privacy regulations d) Data-driven Decision-making and Accountability i. Ethics in data-driven decision-making processes ii. Accountability and transparency in algorithmic decision-making e) Privacy Regulations and Compliance i. Overview of privacy regulations (e.g., GDPR, CCPA, etc.) ii. Compliance requirements and impact on data handling practices	15 Hrs OC1
MODULE II:	**(2 CREDITS)**
Unit II a) Privacy by Design and Default i. Principles of privacy by design and default ii. Integrating privacy considerations into data management processes b) Ethical Data Handling in AI and Machine Learning i. Ethical challenges in AI and machine learning algorithms ii. Fairness, accountability, and transparency in AI ethics c) Data Security and Privacy Protection i. Strategies for data security and privacy protection ii. Best practices for protecting sensitive data and preventing privacy breaches d) Ethical Data Governance and Responsible AI	15 Hrs OC3-OC5

i. Ethical considerations in data governance and responsible AI ii. Establishing ethical frameworks for data governance and AI development e) Ethical Data Impact Assessment and Future Trends i. Conducting ethical data impact assessments ii. Future trends in data ethics, privacy, and responsible data management	

References:

1. Practical Strategies for Implementing Ethical Information Management and Governance, Katherine O'Keefe, Daragh O Brien, Kogan Page Publishers
2. Online resources on National and International data and privacy policies

Course Outcomes (OCs)

Upon completing this course, the student will be able to

OC1	Ethical Data Handling Proficiency
OC2	Compliance with Privacy Regulations
OC3	Critical Analysis of Data Ethics Dilemmas
OC4	Implementation of Privacy by Design Principles
OC5	Strategic Data Security and Privacy Protection
OC6	Ethical Data Governance and Accountability
OC7	Application of Ethical Data Impact Assessments
OC8	Foresight into Future Trends in Data Ethics and Privacy

Course Code: 516a [Elective] **Total Credits:** 04 (120 Lecture Hrs) **University assessment:** 50 marks	**Course Name:** Geospatial Data Analytics Practical **Total Marks:** 100 marks **College/Department assessment**: 50 marks

Pre-requisite:

1) Basic understanding of GIS concepts
2) Familiarity with spatial data formats (e.g., shapefiles, GeoJSON)
3) Proficiency in data analysis tools (e.g., Python, R)

Course Objectives (COs):

To enable the students to:

CO1	Understand the principles of geospatial data analytics
CO2	Learn geospatial data processing and manipulation techniques
CO3	Apply spatial analysis methods to identify spatial patterns and relationships
CO4	Visualize geographic data using mapping tools and software
CO5	Explore applications of geospatial data analytics in different domains

MODULE I:	**(2 CREDITS)**
Unit I a) Introduction to Geospatial Data Analytics i. Overview of geospatial data concepts and applications ii. Introduction to GIS software and tools for geospatial analysis	20 Hrs OC1
Unit II a) Geospatial Data Processing i. Data acquisition and preprocessing techniques for geospatial data ii. Hands-on exercises in cleaning and formatting spatial datasets	20 Hrs OC2
MODULE II:	**(2 CREDITS)**
Unit III a) Spatial Analysis Techniques i. Spatial interpolation methods for raster data analysis ii. Spatial querying and geoprocessing operations in GIS b) Geostatistical Analysis i. Introduction to geostatistics and spatial data modeling ii. Exploratory spatial data analysis (ESDA) and spatial autocorrelation	40 Hrs OC3-OC5
Unit IV a) Spatial Data Visualization i. Mapping techniques for visualizing geographic data ii. Interactive mapping tools and software for spatial visualization b) Applications of Geospatial Data Analytics	

i. Case studies and projects applying geospatial data analytics techniques ii. Future trends in geospatial data analytics and emerging technologies	40 Hrs OC2, OC4, OC6, OC7

References:

1. Mastering Geospatial Analysis with Python" by Silas Toms
2. Python Geospatial Analysis Cookbook" by Michael Diener, Benjamin Buckley, and Joel Lawhead

Course Outcomes (OCs):

Upon completing this course, the student will be able to:

OC1	Proficiency in Geospatial Data Processing
OC2	Application of Spatial Analysis Techniques
OC3	Advanced Geostatistical Analysis Skills
OC4	Effective Spatial Data Visualization Capabilities
OC5	Real-world Applications of Geospatial Data Analytics
OC6	Hands-on Experience and Project Implementation
OC7	Critical Thinking and Problem-solving in Spatial Contexts
OC8	Proficiency in GIS Software and Tools
OC9	Strategic Application of Geospatial Data Analysis
OC10	Understanding Future Trends in Geospatial Data Analytics

Course Code: 516b [Elective]	**Course Name:** Social Network Analysis **Practical**
Total Credits: 04 (120 Lecture Hrs)	**Total Marks:** 100
University assessment: 50 marks	**College/Department assessment:** 50 marks

Pre-requisite:

1. Basic understanding of networks and graphs
2. Familiarity with data analysis and visualization tools
3. Proficiency in a programming language (Python or R)

Course Objectives (COs):

To enable the students to:

CO1	Understand the basic concepts and terminology of social network analysis
CO2	Learn to collect, preprocess, and analyze social network data
CO3	Apply network analysis algorithms to identify key network properties
CO4	Visualize social networks and interpret the results
CO5	Explore real-world applications of social network analysis in various fields

MODULE I:	**(2 CREDITS)**
Unit I b) Introduction to Social Network Analysis i. Overview of social networks and key concepts in network analysis ii. Introduction to network representation (graphs) and basic terminology	20 Hrs OC1
Unit II c) Data Collection and Preprocessing i. Methods for collecting social network data ii. Data preprocessing techniques to clean and format social network datasets	30 Hrs OC2
MODULE II:	**(2 CREDITS)**
Unit III d) Network Analysis Algorithms i. Centrality measures: Degree centrality, betweenness centrality, and closeness centrality ii. Community detection algorithms: Louvain method, Girvan-Newman algorithm	30 Hrs OC3-OC5
Unit IV e) Visualization and Interpretation i. Network visualization tools and techniques ii. Interpretation of network metrics and visual patterns in social networks f) Real-world Applications	

	i. Social network analysis in sociology, marketing, health, and other domains ii. Case studies and hands-on projects applying social network analysis techniques	40 Hrs OC2, OC4, OC6, OC7

References:

3. Python or R programming resources for social network analysis
4. Network analysis libraries (e.g., NetworkX, igraph)
5. Visualization tools (e.g., Gephi, Cytoscape)

Course Outcomes (OCs):

Upon completing this course, the student will be able to:

OC1	Proficiency in Social Network Analysis Tools and Techniques
OC2	Data Collection and Preprocessing Skills
OC3	Understanding of Network Analysis Algorithms
OC4	Visualization and Interpretation Capabilities
OC5	Application of Social Network Analysis in Real-world Contexts
OC6	Hands-on Experience and Project Implementation
OC7	Critical Thinking and Problem-solving Skills
OC8	Effective Communication of Analytical Findings
OC9	Ability to Apply Social Network Analysis in Research and Business Settings
OC10	Confidence in Using Social Network Analysis Tools

Course Code: 516c [Elective]	Course Name: Healthcare Analytics Practical
Total Credits: 04 (120 Lecture Hrs)	Total Marks: 100 marks
University assessment: 50 marks	College/Department assessment: 50 marks

Pre-requisite:

1. Fundamental knowledge of healthcare
2. Database systems, Fundamental concepts of Data Analysis
3. Knowledge of any Data Analysis Tool

Course Objectives (COs):

To enable the students to:

CO1.	Understanding Healthcare Data
CO2.	Analytics Foundations
CO3.	Data Processing Skills
CO4.	Predictive Modeling
CO5.	Clinical Decision Support
CO6.	Healthcare Performance Metrics
CO7.	Compliance and Fraud Detection
CO8.	Data Visualization and Interpretation
CO9.	Ethical and Regulatory Considerations
CO10.	Advanced Trends and Applications

MODULE I:	**(2 CREDITS)**
Unit I a) Introduction to Healthcare Analytics Overview of Healthcare Data and Information Systems Role of Analytics in Healthcare Decision-Making Ethical and Regulatory Considerations in Healthcare Analytics b) Data Sources in Healthcare Electronic Health Records (EHR) and Health Information Exchanges (HIE) Claims and Billing Data Medical Imaging and Diagnostic Data	20 Hrs OC1
Unit II a) Data Preprocessing and Cleaning in Healthcare Data Cleaning Techniques for Healthcare Data De-identification and Anonymization of Patient Data Handling Missing and Inconsistent Data in Healthcare Datasets b) Descriptive Analytics in Healthcare Exploratory Data Analysis for Healthcare Datasets Patient Demographics Analysis Disease Prevalence and Distribution Analysis c) Predictive Analytics and Machine Learning in Healthcare Predictive Modeling for Disease Diagnosis and Prognosis Risk Prediction and Stratification Models Clinical Decision Support Systems	30 Hrs OC2
MODULE II:	**(2 CREDITS)**
Unit III a) Prescriptive Analytics in Healthcare Treatment Optimization and Personalized Medicine Resource Allocation and Capacity Planning Intervention and Preventive Care Recommendations b) Healthcare Data Visualization Visual Representation of Healthcare Data Dashboard Design for Healthcare Analytics Interpretation of Visual Analytics for Healthcare Professionals c) Healthcare Performance Metrics and KPIs Key Performance Indicators for Healthcare Organizations Quality Measures and Patient Outcomes Analysis Value-Based Care Analytics	30 Hrs OC3-OC5
Unit IV a) Healthcare Fraud Detection and Compliance Analytics Fraud Detection Techniques in Healthcare Claims Compliance Monitoring and Risk Assessment Anomaly Detection and Investigation b) Emerging Trends in Healthcare Analytics Real-Time Analytics and Decision-Making in Healthcare Telemedicine and Remote Patient Monitoring Blockchain and Data Security in Healthcare c) Case Studies and Hands-On Projects	40 Hrs OC2, OC4, OC6, OC7

Analyzing and Predicting Disease Outbreaks Predictive Modeling for Patient Readmission Rates Designing Clinical Decision Support Systems	

References:

Healthcare Analytics: From Data to Knowledge to Healthcare Improvement" by Laura A. Tafe and Amanda K. LaRoche

Healthcare Analytics for Quality and Performance Improvement" by Trevor L. Strome

Course Outcomes (OCs):

Upon completing this course, the student will be able to:

- **OC1** Apply Healthcare Data Analysis Techniques
- **OC2** Develop Predictive Models for Healthcare Applications
- **OC3** Analyze Healthcare Performance Metrics
- **OC4** Conduct Ethical Healthcare Analytics Practices
- **OC5** Visualize Healthcare Data Effectively
- **OC6** Explore Emerging Trends in Healthcare Analytics
- **OC7** Engage in Hands-On Healthcare Analytics Projects

Evaluation Scheme

Theory courses of 4 credits: Total marks 100. Out of the total, 50 % each for internal and external evaluation.

A. Internal Evaluation (30m + 10m + 10m = 50 Marks)

The internal assessment marks shall be awarded as follows:

1. **30 marks (Any one of the following):**
   a. Written Test of 30 Marks
   b. SWAYAM (Advanced Course) of minimum 20 hours and certification exam completed or
   c. NPTEL (Advanced Course) of minimum 20 hours and certification exam completed or
   d. Valid International Certifications (Prometric, Pearson, Certiport, Coursera, Udemy and the like)
   e. Certification marks of one completed exam shall be awarded to one course only. For four courses, the students will have to complete four certifications.
   (Note: Only those certification/courses suggested by the department shall be deemed valid, Student cannot do any certification on their own)
2. **10 marks**
   10 marks from every course (Two 4 credits mandatory courses, one 2 credits mandatory course, one 4 credits elective course) coming to a total of 40 marks, shall be awarded on publishing of research paper in UGC approved / Other Journal with plagiarism less than 15%. The marks can be awarded as per the impact factor of the journal, quality of the paper, importance of the contents published, social value.
3. **10 marks**
   Open Book examination based on problem solving related to the respective subject.

i. Suggested format of Question paper of 30 marks for the written test.

Q1.	Attempt *any two* of the following:	16 marks
a.		
b.		
c.		
d.		
Q2.	Attempt *any two* of the following:	14 marks
a.		
b.		
c.		
d.		

B. External Examination: (50 marks) Duration : 2 hrs

	All questions are compulsory	
Q1	(Based on all units) Attempt *any two* of the following:	10 marks
a.	Unit 1	
b.	Unit 2	
c.	Unit 3	
d.	Unit 4	
Q2	(Based on Unit 1) Attempt *any two* of the following:	10 marks
Q3	(Based on Unit 2) Attempt *any two* of the following:	10 marks
Q4	(Based on Unit 3) Attempt *any two* of the following:	10 marks
Q5	(Based on Unit 4) Attempt *any two* of the following:	10 marks

Theory courses of 2 credits: Total marks 50. Out of the total, 50 % each for internal and external evaluation.

A. Internal Evaluation (25 Marks)

The internal assessment marks shall be awarded as follows:

1. 10 marks from every course (Two 4 credits mandatory courses, One 2 credits mandatory course, One 4 credits elective course) coming to a total of 40 marks, shall be awarded on publishing of research paper in UGC approved / Other Journal with plagiarism less than 15%. The marks can be awarded as per the impact factor of the journal, quality of the paper, importance of the contents published, social value.
2. 10 marks - Open Book examination based on problem solving related to the respective subject.
3. 5 marks - Assignment/Group discussion.

B. External Examination: (25 marks) Duration : 1 hr

	All questions are compulsory	
Q1	(Based on Unit 1) Attempt *any two* of the following:	13 marks
Q2	(Based on Unit 2) Attempt *any two* of the following:	12 marks

Practical courses of 2 credits: Total marks 50. Out of the total, 50 % each for internal and external evaluation.

A. Practical Evaluation Internal (25 marks)

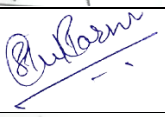
1.	Performance during all practical sessions	10
2.	Problem solving with the acquired programming skills	10
3.	Viva Voce	5

B. Practical Evaluation External (25 marks)

A Certified copy of hard-bound journal is essential to appear for the practical examination.

1.	Practical Question	15
2.	Journal	5
3.	Viva Voce	5

Signatures of Team Members

Sr. No.	Name	Signature
1	Dr. Srivaramangai R	
2	Mr. Nikhil Pawanikar	
3	Ms. Maria Muthukumar	
4	Mr. Suvee Kulkarni(Industry Expert)	
5	Mr. Nishant Vibhute (Industry Expert)	

Appendix B

Justification for M.S. (Data Analytics)

1.	Necessity for starting the program:	The exponential growth of data in businesses and organizations has created a significant demand for skilled professionals who can analyze and interpret this data to derive valuable insights. By offering the M.S. in Data Analytics program, institutions can meet the pressing need for skilled professionals in this field, contributing to the growth and success of industries and organizations while also fostering innovation and research in data analytics.
2.	Whether the UGC has recommended the program:	Yes
3.	Whether all the programs have commenced from the academic year 2024-25	The program is commencing from 2024-2025
4.	The programs started by the University are self-financed, whether adequate number of eligible permanent faculties are available?:	Yes. Some experts are called as visiting faculties
5.	To give details regarding the duration of the program and is it possible to compress the program?:	2 years. Not possible to compress the program
6.	The intake capacity of each program and no. of admissions given in the current academic year:	40 seats. 2024-2025 admission is yet to start
7.	Opportunities of Employability / Employment available after undertaking these courses:	The program prepares individuals to work as business intelligence analysts, specializing in transforming data into actionable intelligence to assist in strategic decision-making. Graduates can pursue careers as data scientists, leveraging their expertise to analyze complex datasets, develop predictive models, and derive actionable insights to drive business decisions. The M.S. in Data Analytics opens up diverse and rewarding career opportunities across industries, with the potential for high growth, impact, and contribution to the evolving field of data analytics.

Sign of the
BOS Chairman
Dr. Mrs. R. Srivaramangai
Board of Studies in Data Science

Sign of the
Offg. Associate Dean
Dr. Madhav R. Rajwade
Faculty of Science & Technology

Sign of the
Offg. Dean
Prof. Shivram S. Garje
Faculty of Science & Technology