Mahatma Education Society's

Pillai College of Arts, Commerce & Science (Autonomous)

Affiliated to University of Mumbai

'NAAC Accredited 'A' grade (3 cycles)'
'Best College Award' by University of Mumbai
ISO 9001:2015 Certified

SYLLABUS

Program: Master of Science (M. Sc.) in Data Analytics

M.Sc.- Part I Data Analytics

PCACS/PMSDA/SYL/2024-25/PI

As per National Education Policy
Choice Based Credit & Grading System

Academic Year 2024-25

Mahatma Education Society's

Pillai College of Arts, Commerce & Science
(Autonomous)
Affiliated to University of Mumbai
NAAC Accredited 'A' grade (3 cycles)
Best College Award by University of Mumbai
ISO 9001:2015 Certified

Board of Studies in the Department of Computer Science

Sr. No.	Name of the	Details	Sign
1	Prof. Deepika Sharma	Chairperson (Head of Department of Information Technology & Computer Science), Vice Principal	*signature*
2	Dr. Gajanan Wader	Principal	*signature*
3	Mrs.Munawira Kotyad Founder and CEO Wonderwheel Enterprises, Pillai, Director Pillai Center for Innovation & Research	Faculty Specialization	Absent
4	Dr. Amiya Kumar Tripathy Director Center for GeoAI & ML, Professor, Computer Engineering, Don Bosco Institute of Technology, Mumbai	Subject Expert From Outside Parent University	*signature*
5	Dr.Mrs. Anjali Kulkarni CKT College, New Panvel	Vice Chancellor Nominee, University of Mumbai	*signature*
6	Mr. Tito Idicula, Director, Programming Hub	Alumni representative	*signature*
7	Mr. Anant Baddi, Security Solution Architect, cloud Google Google	Industry Representative (Industry/Corporate/Allied Sector)	Absent
8	Mr. Bhupendra Kesariya Professor,N. M. .College, Vile Parle	Subject Expert in Mathematics From Outside Parent University	*signature*
9	Mrs. Anju Somani	Faculty Specialization	*signature*
10	Mrs. Shubhangi Pawar	Faculty Specialization	*signature*

11	Dr. Kumudini Das	Faculty Specialization	Absent
12	Mrs. Soly Zachariah	Faculty Specialization	
13	Mrs. Ramya S. Kumar	Faculty Specialization	
14	Mrs. Sujata Shahabade	Faculty Specialization	
15	Mrs. Sreevidya T.V.	Faculty Specialization	
16	Mr. Omkar Sherkhane	Faculty Specialization	
17	Mr. Abhijeet Salvi	Faculty Specialization	

Introduction to M.Sc. Data Analytics Program

M.Sc. Data Analytics is a two years post graduate programme that concentrates on creating links between theory and practice. Data Analytics is the process of analyzing sets of data to guide business decisions, plus its business benefits and different types of analytics.

It covers a wide variety of software and hardware technologies and their applications. Students will also gain practical problem solving and program design skills; Students learn how to think more carefully and how to solve real world problems more effectively.

They will not only develop a diverse set of skills to prepare for their curriculum and for employment, but will also be encouraged to launch their own startups or venture into new types of careers using their interdisciplinary training.

Our curriculum exposes students to modern advancements. As we all know M.Sc. degrees in Data Analytics lead to rewarding and lucrative careers, excellent placement and incubation assistance is provided.

Program Outcomes

Sr No	PO Title	POs in brief
PO1	Advanced Knowledge and Expertise	Demonstrate a systematic, extensive and coherent knowledge and understanding of their academic discipline as a whole and its applications, and links to related disciplinary areas/subjects of study; demonstrate a critical understanding of the latest developments in the subject, and an ability to use established techniques of analysis and enquiry within the subject domain with a global perspective.
PO2	Research and Innovation	Acquire comprehensive knowledge about current research and innovation, and acquire techniques and skills required for identifying problems and issues to produce a well-researched written work that engages with various sources employing a range of disciplinary techniques and scientific methods applicable.
PO3	Interdisciplinary Perspective	Commitment to intellectual openness and developing understanding beyond subject domains; answering questions, solving problems and addressing contemporary social issues by synthesizing knowledge from multiple disciplines.
PO4	Leadership Abilities & Entrepreneurial Mindset	Inculcate Leadership skills, including the ability to lead teams, manage projects, and make strategic decisions. innovation and entrepreneurship, business development, technology commercialization, and startup creation
PO5	Communication Competence	Demonstrate effective oral and written communication skills to covey disciplinary knowledge and to communicate the results of studies undertaken in an academic field accurately in a range of different contexts using the main concepts, constructs and techniques of the subject(s) of study
PO6	Ethical Conduct and Research Integrity	Understanding and adherence to ethical standards and research integrity by developing commitment towards professional ethics and responsibilities as a social endeavor.
PO7	Career development	Demonstrate subject-related knowledge and skills that are relevant to academic, professional, soft skills and employability required for higher education and placements.
PO8	Commitment to the society and to the Nation	Recognise the importance of social, environmental, human and other critical issues faced by humanity at the local, national and international level; appreciate the pluralistic national culture and the importance of national integration.

Program Specific Outcomes

Sr. No	PSOs in brief
PSO1	Competence in employing principles, techniques and tools of data analytics for business analytics.
PSO2	Curiosity and readiness to deal with small and big data and ability to engage in exploratory research.
PSO3	Capability to become a successful trainer in data science, data analytics, productive decision maker and therefore well-respected personality at work and in life.
PSO4	Students completing this programme will become full-stack data scientists, well-equipped for a variety of job roles depending on their choices and temperament including Data Scientists, Data Analysts as well as Industrial and Societal Entrepreneurs.

Course Structure

Semester I						
Course Code	**Course Type**	**Course Title**	**Theory/ Practical**	**Marks**	**Credits**	**Lectures/ Week**
PMSDA101	**Major**	Foundations of Data Science	Theory	100	4	4
PMSDA102	**Major**	Next Generation Databases	Theory	100	4	4
PMSDA103	**Major**	Predictive Analytics	Theory	100	4	4
PMSDA104	**Major Elective**	a. Visual Perception and Recognition a. Computer Forensics b. Malware Analysis	Theory	50	2	4
PMSDA105P	**Major Practical**	Practicals (PMSDA101 + PMSDA102)	Practical	100	2	4
PMSDA106P	**Major + Elective Practical**	Practicals (PMSDA103 +PMSDA104)	Practical	100	2	4
PMSDA107	**RM**	Research Methodology	Theory	50	2	2
PMSDA108P	**RM Practical**	Practicals (PMSDA107)	Practical	50	2	2
Total				650	22	28
All Subjects having Field Project as part of Continuous Assessment-2						

Abbreviations:

RM: Research Methodology

	Semester II					
Course Code	Course Type	Course Title	Theory/ Practical	Marks	Credits	Lectures/ Week
PMSDA201	**Major**	Advanced Statistical Methods & Testing of Hypothesis	Theory	100	4	4
PMSDA202	**Major**	Big Data Analytics	Theory	100	4	4
PMSDA203	**Major**	Cloud Computing & Virtualization	Theory	100	4	4
PMSDA204	**Major Elective**	a. Modern Networking b. Information Security Auditing c. Network Security & Cryptography	Theory	50	2	4
PMSDA205P	**Major Practical**	Practicals (PMSDA201 + PMSDA202)	Practical	100	2	4
PMSDA206P	**Major + Elective Practical**	Practicals (PMSDA203 + PMSDA204)	Practical	100	2	4
PMSDA207	**OJT/FP/RP**	Internship / Apprenticeship	-	100	4	(Min. 90 Hrs)
Total				650	22	24
All Subjects having Field Project as part of Continuous Assessment-2						

Abbreviations:

OJT : On Job Training: Internship/ Apprenticeship

FP : Field Projects

RP : Research Project

Evaluation Pattern

Marking Code	Marking Scheme
A	60 Marks Final Exam, 20 Marks Internal Exam, 15 Marks Field Project, 5 Marks Attendance
B	50 marks distributed within report /case study/ project/ presentation etc.
C	50 Marks Practical Examination.
D	100 Marks Practical Examination. Subject 1 Practical (50 Marks) + Subject 2 Practical (50 Marks) =100
E	100 Marks distributed within report / internship period / certificate etc.
F	50 Marks Theory Exam.

SEMESTER - I				
Course Code	**Course Type**	**Course Title**	**Evaluation Pattern**	**Marks**
PMSDA101	**Major**	Foundations of Data Science	A	100
PMSDA102	**Major**	Next Generation Databases	A	100
PMSDA103	**Major**	Predictive Analytics	A	100
PMSDA104	**Major Elective**	a. Visual Perception and Recognition b. Computer Forensics c. Malware Analysis	F	50
PMSDA105P	**Major Practical**	Practicals (PMSDA101 + PMSDA102)	D	100
PMSDA106P	**Major + Elective Practical**	Practicals (PMSDA103 +PMSDA104)	D	100
PMSDA107	**RM**	Research Methodology	B	50
PMSDA108P	**RM Practical**	Practical (PMSDA107)	C	50
Total				650

SEMESTER – II				
Course Code	Course Type	Course Title	Evaluation Pattern	Marks
PMSDA201	**Major**	Advanced Statistical Methods & Testing of Hypothesis	A	100
PMSDA202	**Major**	Big Data Analytics	A	100
PMSDA203	**Major**	Cloud Computing & Virtualization	A	100
PMSDA204	**Major Elective**	a. Modern Networking b. Information Security Auditing c. Network Security & Cryptography	F	50
PMSDA205P	**Major Practical**	Practicals (PMSDA201 + PMSDA202)	D	100
PMSDA206P	**Major + Elective Practical**	Practicals (PMSDA203 + PMSDA204)	D	100
PMSDA207	**OJT/FP/RP**	Internship / Apprenticeship	E	100
Total				650

SEMESTER - I

BOS	COMPUTER SCIENCE
Class	M.Sc. Data Analytics Part - I
Semester	I
Course Name	Foundations of Data Science
Course Code	PMSDA101
Type of course	Major
Level of the Subject	Advanced
Credit points	04 Theory+01 Practical

Course Objectives:

Having successfully completed this module, you will be able to demonstrate knowledge and understanding of

1. Key concepts in data science, including tools, approaches, and application scenarios.
2. State-of-the-art tools to build data-science applications for different types of data, including text and CSV data.

Unit No.	Name of Unit	Topic No.	Content	Hours
1	Introduction to Data Science	1.1	Introduction : What is Data Science, Defining data science project roles: Project roles, Understanding the stages of a data science project : Defining the goal,Data collection and management,Modeling ,Model evaluation and critique,	**15**
		1.2	data science project : Presentation and documentation,Model deployment and maintenance, Setting expectations for a new data science project: Determining lower and upper bounds on model performance	
2	Basic of R & Loading data into R	2.1	Getting Started with R : Installing R , Using R , Creating and Using Vectors ,Creating Data frames-Exploring data frames-Accessing columns in a Data frame- Reading a CSV text file-Removing rows and columns-Renaming rows and columns-sorting data frames, Understanding R's data frame structure,	**15**
		2.2	**Working with data from files:**Working with well-structured data from files or URLs,Using R on less-structured data, **Working with relational databases:** A production-size example,Loading data from a database into R, Working with the PUMS data.	
3	Working with R-Studio, Exploring data &	3.1	Onward with RStudio-Creating R scripts-Creating Functions using R-Testing Functions-Use of Statistics-Sampling a population **Exploring data : Using summary statistics to spot problems**: Typical problems revealed by data summaries.	**15**

	Managing data		**Spotting problems using graphics and visualization** :Visually checking distributions for a single variable, Visually checking relationships between two variables	
		3.2	**Managing data** **Cleaning data** :Treating missing values (NAs), Data transformations, **Sampling for modeling and validation:** Test and training splits, Creating a sample group column, Record grouping, Data provenance,	
4	Modeling methods	4.1	**Choosing and evaluating models** **Mapping business problems to machine learning tasks**: Solving classification problems, Solving scoring problems, Working without known targets, Problem-to-method mapping. **Evaluating models :** Evaluating classification models, Evaluating scoring models.	**15**
		4.2	**Evaluating models :**Evaluating probability models,Evaluating ranking models ,Evaluating clustering models **Validating models :** Identifying common model problems, Quantifying model soundness, Ensuring model quality,	
Total No. Of Lectures				**60**

Course Outcomes:
1. Understand the basics of the R programming language.
2. Demonstrate how to install and configure RStudio.
3. Analyze data and generate reports based on the data by using the ggplot2 package.
4. Develop necessary skills for data processing, statistical and computational analysis, visualization and document preparation.
5. Improve skills in solving scientific problems of interdisciplinary nature in academia and industry.
6. Apply various data mining algorithms to solve real world problems.

References:
1. "Data Science from Scratch: First Principles with Python" by Joel Grus
2. "R for Data Science: Import, Tidy, Transform, Visualize, and Model Data" by Garrett Grolemund and Hadley Wickham
3. "Doing Data Science: Straight Talk from the Frontline" by Cathy O'Neil and Rachel Schutt
4. "Practical Statistics for Data Science" by Peter Bruce and Andrew Bruce
5. "Statistics for Data Science" by James D. Miller:

CASE STUDY:

1	You are working as a data scientist for a telecommunications company, and you are tasked with reducing customer churn (i.e., the rate at which customers switch to a competitor). You have a dataset containing customer information and historical churn data.
2	You are a data scientist tasked with building a predictive model to determine the likelihood of passengers surviving the Titanic disaster. You have a dataset containing passenger information, including age, gender, class, and other factors.

PRACTICALS

Practical No.	Details
1	Basic of R (variable, Decision making, loops, function etc.)
2	R Statistical Examples (Descriptive Statistics & Advanced Descriptive Statistics)
3	Working with CSV files & excel file dataset in R (Import and plot)
4	R and SQL Database Connecting R with SQL Database and performing inserting, retrieving, updating and deleting.
5	Data visualization with R and ggplot2
6	Advanced Data visualization with R and ggplot2
7	Cleaning data ,modeling and validation.
8	Evaluating classification models & Evaluating scoring models.
9	Evaluating probability models & Evaluating clustering models
10	Create a simple web application using R

BOS	Computer Science
Class	M.Sc. Data Analytics Part - I
Semester	I
Course Name	Next Generation Databases
Course Code	PMSDA102
Type of course	Major
Level of the Subject	Advanced
Credit points	04 Theory+01 Practical

Course Objectives:
1. Explain differences between Relational and NoSQL databases.
2. Use NoSQL databases to Store the big data for useful business applications.

Unit No.	Name of Unit	Topic No.	Content	Hours
1	Database Revolutions	1.1	**Introduction :** Different Databases for Different Requirements: Relational Database Design, Early Database Management Systems, The Relational Database Revolution, Motivations for Not Just/NoSQL (NoSQL) Databases.	15
		1.2	**Working with Relational Database:** Relational Database: Database Design, Data Storage, Transaction Management, Working with SQL command, CRUD Operation, group by clause, having clause , joins etc.	
2	Big Data Revolution:	2.1	**Big Data Revolution:** CAP Theorem, Birth of NoSQL. Types of NoSQL databases. **Document Databases :** Document Data Model, Relational Vs Document Database , Working of Document Data Model, Advantages, Disadvantages.	15
		2.2	**Working with MongoDB:** MongoDB Installation ,Create Database, Create Collection, Insert, Find, Update, Delete,Query Operators, Update Operators, Aggregations,Indexing/Search, Replication, sharding. **Working with CouchDB:** Features of CouchDB, CouchDB Installation. CouchDB Database : Create Database, Delete Database. CouchDB Document : Create, Update, Delete, Create View	
3	Distributed Database Patterns	3.1	**Column Databases:** Data Warehousing Schemes, Columnar Alternative, Sybase IQ, C Store and Vertica, Column Database Architectures,	

			The End of Disk: SSD and In-Memory Databases, Berkeley Analytics Data Stack and Spark.		**15**
		3.2	**Working with Cassandra :** Introduction, Architecture, Data Model, Installation, Referenced Api, Cqlsh, Shell Commands Cassandra Keyspace Operations: Create Keyspace , Alter Keyspace, Drop Keyspace Cassandra Table Operations: Create Table, Alter Table, Drop Table, Truncate Table, Create Index,Drop Index, Batch Cassandra CRUD Operations: Create Data, Update Data, Read Data, Delete Data Cassandra CQL Types: CQL Datatypes, CQL Collections , CQL User Defined Datatypes		
4	Data Models and Storage	4.1	Graph Oriented Databases : What is a Graph Database, Why Graph Databases? The Property Graph Model, What is Neo4j? **Working with Neo4j :** Overview, , Data Model, Environment Setup, Building Blocks , Neo4j CQL, Neo4j CQL - Introduction, Neo4j CQL - Creating Nodes, Neo4j CQL - Creating a Relationship, Neo4j CQL Write Clauses, Merge Command, Set Clause , Delete Clause, Remove Clause, Foreach Clause, Neo4j CQL Read Clause, Match Clause, Optional Match Clause, Where Clause, Count Function Neo4j CQL General Clauses: Return Clause , Order By Clause, Limit Clause, Skip Clause, With Clause, Unwind Clause Neo4j CQL Functions : String Functions, Aggregation Function**.**		**15**
		4.2	**Key-Value Databases :** How do key-value databases work?, When to use a key-value database, Features, Advantages, Disadvantages, examples. **Working with Redis :** Redis TutorialRedis vs RDBMS , Redis vs Other Key-value, Installation, Redis Configuration, Redis Data types, Redis All Commands.		
			Total No. of Lectures		**60**

Course Outcomes :
1. Understand the concept of different types of NoSQL database.
2. Describes the database revolution from file based to SQL to NOSQL.
3. Helps to understand the different distributed database patterns .
4. Students will work with mongodb and enhance their knowledge of Big Data.
5. Differentiate between relational databases and other databases.
6. Enhance knowledge on data retrieval with data mining and data warehouse.

Books and References:
1. NoSQL for Mere Mortals Dan Sullivan Addison-Wesley NoSQL for Dummies Adam Fowler John Wiley & Sons
2. Abraham Silberschatz, Henry F. Korth, S. Sudarshan, "Database System Concepts", Sixth Edition, McGrawHill.

3. Guy Harrison, "Next Generation Databases", Apress, 2015.
4. Eric Redmond, Jim R Wilson, "Seven Databases in Seven Weeks", LLC. 2012.
5. Dan Sullivan, "NoSQL for Mere Mortals", Addison-Wesley, 2015.
6. Adam Fowler, "NoSQL for Dummies ", John Wiley.5 & Sons, 2015

CASE STUDY	
1	Hospital Management System: A DBMS Case Study Problem Statement A hospital needs an efficient system to manage its operations. The system should handle the following aspects: Departments: Keep track of different hospital departments (e.g., Cardiology, Pediatrics, Orthopedics, etc.). Each department has a unique ID, name, and description. Rooms: Maintain information about hospital rooms. Each room has a room number, capacity (number of beds), and department ID (indicating which department the room belongs to). Doctors: Store details of doctors working in the hospital. Each doctor has a unique ID, name, specialization, contact information, and department ID (specifying their area of expertise).

2	Case Study: Online Bookstore Database
	Background: An online bookstore named "BookWorm" maintains its inventory using a MongoDB database. The database contains two collections: books and users.
	Collection: books
	Each document in the books collection represents a book. Here's an example document:
	JSON
	
	`{`
	`  "_id": ObjectId("5f8a1b9e4c3a2d0017e6a1a1"),`
	`  "title": "The Great Gatsby",`
	`  "author": "F. Scott Fitzgerald",`
	`  "genre": "Classic",`
	`  "price": 15.99,`
	`  "stock": 50`
	`}`
	Collection: users
	Each document in the users collection represents a registered user. Example document:
	JSON
	
	`{`
	`  "_id": ObjectId("5f8a1c0e4c3a2d0017e6a1a2"),`
	`  "username": "booklover123",`
	`  "email": "booklover@example.com",`
	`  "purchased_books": [`
	`    {`
	`      "book_id": ObjectId("5f8a1b9e4c3a2d0017e6a1a1"),`
	`      "quantity": 2`
	`    },`
	`    {`
	`      "book_id": ObjectId("5f8a1c0e4c3a2d0017e6a1a3"),`
	`      "quantity": 1`
	`    }`
	`  ]`
	`}`

PRACTICALS

Practical No.	Details
1	MongoDB Basics & Implementing Aggregation 1) Write a MongoDB query to create and drop databases. 2) Write a MongoDB query to create, display and drop collection. 3) Write a MongoDB query to insert, query, update and delete a document. 4) Write a MongoDB query to use sum, avg, min and max expressions. 5) Write a MongoDB query to use push and addToSet expressions. 6) Write a MongoDB query to use the first and last expression.
2	Replication, Backup and Restore 1) Write a MongoDB query to create Replica of existing database. 2) Write a MongoDB query to create a backup of an existing database. 3) Write a MongoDB query to restore the database from the backup.
3	Java/Python/R and MongoDB 1) Connecting Java/Python/R with MongoDB and performing inserting, retrieving, updating and deleting.
4	Create Data Model using Cassandra 1) Cassandra Keyspace Operations. 2) Cassandra Table Operations. 3) Cassandra CRUD Operations. 4) Cassandra CQL Types
5	Java/Python/R and Cassandra 1) Connecting Java with Cassandra and performing inserting, retrieving, updating and deleting.
6	PostgreSQL 1) Table Operations (CRUD) 2) Aggregation Functions 3) Join & Sub query Operations 4) String Functions
7	Java/Python/R and PostgreSQL Connect Java with PostgreSQL and perform inserting, retrieving, updating and deleting.
8	CouchDB 1) Working with Database 2) Working with Document 3) Attaching file
9	Java/Python/R and CoachDB Connect Java with CoachDB and perform inserting, retrieving, updating and deleting.
10	Neo4j 1) Neo4j CQL 2) Neo4j CQL Write Clause , Read Clause & General Clause 3) Neo4j String & Aggregate functions 4) CQL Admin (Backup,Restore,index)

BOS	COMPUTER SCIENCE
Class	M.Sc. Data Analytics Part - I
Semester	I
Course Name	Predictive Analytics
Course Code	PMSDA103
Type of course	Major
Level of the Subject	Advanced
Credit points	04 Theory+01 Practical

Course Objectives:
1. Analyze the data and its obstacles along with different learning methods.
2. Apply knowledge of predictive models using statistical and data mining techniques.

Unit No.	Name of Unit	Topic No.	Content	Hours
1	Introduction to Predictive Analytics	1.1	**Overview of Predictive Analytics:** Predictive Analytics, Supervised vs. Unsupervised Learning, Parametric vs. Non-Parametric Models, Business Intelligence, Predictive Analytics vs Business Intelligence, Statistics and Analytics, Predictive Analytics vs. Data Mining	15
		1.2	**Challenges in Using Predictive Analytics**: Obstacles in Management, Obstacles with Data, Obstacles with Modeling, Obstacles in Deployment.	
2	Setting Up the Problem	2.1	**Predictive Analytics Processing Steps:** CRISP-DM, Business Understanding, The Three-Legged Stool, Business Objectives	15
		2.2	**Defining Data for Predictive Modeling:** Defining the Columns as Measures, Defining the Unit of Analysis, Defining the Target Variable, Temporal Considerations for Target Variable, Defining Measures of Success for Predictive Models, Success Criteria for Classification, Success Criteria for Estimation.	
3	Data Understanding	3.1	**Single Variable Summaries:** Mean, Standard Deviation, The Normal Distribution, Uniform Distribution, Applying Simple Statistics in Data Understanding, Skewness, Kurtosis, Rank-Ordered Statistics, Categorical Variable Assessment **Data Visualization in One Dimension** Histograms, Boxplot, stripplot, dot plot	15

		3.2	**Multiple Variable Summaries :**The Combinatorial Explosion of Interactions , Correlations Spurious Correlations Matrix, Covariance, Matrix ,Cross-Tabulation and Contingency Tables	
4	Data Preparation & Predictive Modeling	4.1	**Data Preparation:** Variable Cleaning, Incorrect Values, Consistency in Data Formats, Outliers, Multidimensional Outliers, Missing Values, Fixing Missing Data,	**15**
		4.2	**Feature Creation:** Simple Variable Transformations, Fixing Skew, Binning Continuous Variables, Numeric Variable Scaling, Nominal Variable Transformation, Ordinal Variable Transformations, Date and Time Variable Features, ZIP Code Features. **Predictive Modeling:** Decision Trees, Linear Regression , Logistic Regression, K-Nearest Neighbor, Assessing Predictive Models.	
			Total No. Of Lectures	60

Course Outcomes :

1. Understand the basics of Predictive analytics and the general challenges faced while working with Data.
2. Explain the CRISP-DM model and work with data by defining measures of success
3. Understand the data with help of Statistical concepts like Mean, Standard Deviation, Skewness, Kurtosis.
4. Apply exploratory data analysis using graphs to extract insights and patterns from Data.
5. Prepare and transform the data using techniques like fixing Missing values, Incorrect values, Fixing Skew etc to make it suitable for performing analysis.
6. Apply advanced Predictive Modeling techniques like Decision trees, K-Nearest Neighbor, Regression Models and test their efficiency.

Books and References:
1. Applied Predictive Analytics Principle and Techniques for the professional data analyst by Dean Abbot Wiley publication.
2. Data Mining and Predictive Analytics by Daniel T. Larose and Chantal D. Larose Wiley publication
3. "Applied Predictive Modeling" by Max Kuhn and Kjell Johnson
4. "Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die" by Eric Siegel
5. "Hands-On Predictive Analytics with Python" by Alvaro Fuentes
6. "Modeling Techniques in Predictive Analytics" by Thomas W. Miller

CASE STUDY	
1	The telecommunications industry is characterized by fierce competition, rapid technological advancements, and evolving customer preferences. With the proliferation of service providers offering similar products and services, customer loyalty has become increasingly elusive. High churn rates not only erode revenue but also incur substantial costs associated with customer acquisition and retention efforts. In an increasingly competitive telecommunications landscape, retaining customers is paramount for sustaining business growth and profitability. As customer churn continues to

	pose a significant challenge, telecommunications companies are turning to predictive analytics to proactively address this issue. how a telecommunications company can utilize predictive analytics to enhance customer retention and achieve a substantial 20% reduction in churn rate within the next year.
2	Our organization, like many others, regularly assesses invoices for payment of service rendered by vendors. Unfortunately, some of these invoices are submitted fraudulently leading to financial losses and operational disruptions. Due to resource constraints, ou organization can only afford to investigate a small number of invoices in detail. Therefore there is a pressing need to develop a predictive modeling solution that can efficiently identify the most suspicious invoices for thorough investigation Invoices play a critical role in the operation of businesses, serving as a means of payment fo services rendered. However, the submission of fraudulent invoices poses a significant risk t organizations, resulting in financial losses and reputational damage. To mitigate this risk, ou organization has embarked on a journey to leverage predictive modeling as a proactive approach to identify key invoices for investigation. This case study explores how th CRISP-DM (Cross-Industry Standard Process for Data Mining) model can be applied t address the challenge of invoice fraud detection.

PRACTICALS

Practical No.	Details
1.	Perform the Single Variable summary using Python.
2.	Perform the Multiple Variable non graphical summary using Python.
3.	Perform the Single Variable Graphical summary using Python.
4.	Perform the Multiple Variable summary using Python.
5	Perform Feature Transformation with all the types.
6	Perform the following Data Preparation task on any of the data • Missing Value Detection from all the columns • Feeding of Missing values. • Outlier Detection
7	Perform the following Data Preparation task on any of the data • Check the correlation between various columns • Check the skewness and kurtosis of data • Perform the transformation of data.
8	Perform the Data Transformation on date time and zip code feature.
9	Perform Logistics Regression on Diabetic dataset and Evaluate the model performance
10	Case Study : Amazon_cloths sells clothes online. Customers come into the store, have meetings with a personal stylist, then they can go home and order either on a mobile app or website for the clothes they want. The company is trying to decide whether to focus their efforts on their mobile app experience or their website. Following is predict is analysis for this company

BOS	Computer Science
Class	M.Sc. Data Analytics Part - I
Semester	I
Course Name	Visual Perception and Recognition
Course Code	PMSDA104
Type of course	Major Elective
Level of the Subject	Advanced
Credit points	02 Theory+01 Practical

Course Objectives:

1. The main objective of this course is making the students know about various fundamental principles of image and video analysis and have an idea of their application.
2. Understand how to apply filters in an image & transformations in images and videos.

Unit No.	Unit Name	Topic No.	Content	Hours
1	Digital image representation-	1.1	Digital image representation- Visual Perception- Sampling and Quantization- Basic Relations between Pixels	10
		1.2	Mathematical Tools Used in Digital Image Processing: Fundamental Operations –Vector and Matrix Operations- Image Transforms (DFT, DCT, DWT, Hadamard).	
2	Fundamentals of spatial filtering:	2.1	Fundamentals of spatial filtering: spatial correlation and convolution-smoothing blurring- sharpening edge detection -	10
		2.2	Basics of filtering in the frequency domain: smoothing-blurring-sharpening – Histograms and basic statistical models of image.	
3	Color models and Transformations	3.1	Color models and Transformations – Image and Video segmentation-Image and video demonizing Image and Video enhancement- Image and Video compression.	10
		3.2	Object detection and recognition in image and video-Texture models Image and Video classification models- Object tracking in Video.	10
4	Applications and Case studies	4.1	Applications and Case studies- Industrial- Retail Transportation & Travel	10
		4.2	Remote sensing-Video Analytics in WSN: IoT Video Analytics Architectures.	
Total No. of Lectures				**40**

Course Outcomes:

1. Understand the requirements of image processing.
2. Illustrate the principles and techniques of digital image in applications related to digital imaging systems.
3. Demonstrate image recognition and motion recognition.
4. Understand color models, transformations and the fundamentals of digital video processing.
5. Illustrate the object detection, recognition and tracking of objects.
6. Analysis of the video processing in applications.

Books and References:
1. "Digital Image Processing" by Rafael C. Gonzalez and Richard E. Woods
2. "Digital Image Processing: An Algorithmic Introduction Using Java" by Wilhelm Burger and Mark J. Burge
3. "Computer Vision: Algorithms and Applications" by Richard Szeliski
4. "Computer Vision: Models, Learning, and Inference" by Simon J.D. Prince
5. "Digital Image Processing and Analysis: Human and Computer Vision Applications with CVIPtools" by Scott E. Umbaugh
6. "Pattern Recognition and Machine Learning" by Christopher Bishop

CASE STUDY	

1	Case Study: Enhancing Traffic Management with Image and Video Analytics
	Background: A smart city initiative aims to improve traffic management using image and video analytics. The city has deployed cameras at key intersections and highways. These cameras capture real-time video feeds, which can be analyzed to enhance traffic flow, monitor violations, and optimize transportation.
	Collection: traffic_cameras
	Each document in the traffic_cameras collection represents a camera deployed at a specific location. Example document:
	JSON
	{
	  "_id": ObjectId("5f8a1b9e4c3a2d0017e6a1a1"),
	  "location": "Intersection A",
	  "camera_type": "Highway",
	  "status": "Active"
	}

2	Case Study: Detecting Anomalies in Security Camera Footage Background: You work for a security company that monitors surveillance cameras in public spaces. One of your clients, a shopping mall, has reported some unusual activity captured by their security cameras. They suspect that certain individuals are loitering in restricted areas after hours. Your task is to analyze the video footage and identify any anomalies. Scenario: Late one night, the security cameras at the mall captured the following events: Loitering in Restricted Area: A person wearing a hooded jacket was seen lingering near the mall's back entrance. The individual spent an unusually long time in the vicinity, looking around nervously. The person did not enter the mall but eventually left the area. Unattended Bag: Around the same time, another camera captured footage of an unattended bag left near a bench in the mall's central atrium. The bag appeared to be abandoned, and no one approached it for several minutes.

	Suspicious Interaction: A third camera recorded two people having a heated argument in the parking lot. Their gestures were aggressive, and they seemed to be exchanging something discreetly. After a few minutes, they separated and left in opposite directions.

PRACTICALS

Practical No	Details
1	Write a python program to perform fundamental operations on Image.
2	Write a python program to filter images by using spatial domain in Digital Image Processing.
3	Write a python program to filter image by using the frequency domain in digital image processing
4	Write a program to perform segmentation on Image and Video.
5	Write a program to apply enhancement on Image and video.
6	Write a program to perform compression Image and Video.
7	Write a program to build a Video classification model.
8	Develop an application to Object detection and recognition in image and video.
9	Develop an application to Object tracking in Video.
10	Develop an application to Video Analytics in WSN.

BOS	Computer Science
Class	M.Sc. Data Analytics Part - I
Semester	I
Course Name	Computer Forensics
Course Code	PMSDA104
Type of course	Major Elective
Level of the Subject	Advanced
Credit points	02 Theory+01 Practical

Course Objectives:

1. To correctly define and cite appropriate instances for the application of computer forensics Correctly collect and analyze computer forensic evidence.
2. Identify the essential and up-to-date concepts, algorithms, protocols, tools, and methodology of Computer Forensics

Unit No.	Unit Name	Topic No.	Content	Hours
1	Cyber Crime and computer crime	1.1	Introduction to Digital Forensics, Definition and types of cybercrimes, electronic evidence and handling, electronic media, collection, searching and storage of electronic media,	15
		1.2	introduction to internet crimes, hacking and cracking, credit card and ATM frauds, web technology, cryptography, emerging digital crimes and modules.	
2	Basics of Computer Computer organization, component	2.1	Computer organization, components of computer- input and output devices, CPU, Memory hierarchy, types of memory,	
		2.2	storage devices, system softwares, application softwares, and the basics of computer languages.	15
3	Computer Forensics	3.1	Definition and Cardinal Rules, Data Acquisition and Authentication Process, Windows Systems-FAT12, FAT16, FAT32 and NTFS,	
		3.2	UNIX file Systems, mac file systems, computer artifacts, Internet Artifacts, OS Artifacts and their forensic applications	15
4	Forensic Tools and Processing of Electronic Evidence	4.1	Introduction to Forensic Tools, Usage of Slack space, tools for Disk Imaging, Data Recovery, Vulnerability Assessment Tools, Encase and FTK tools, Anti Forensics and probable counters, retrieving information, process of computer forensics and digital investigations,	15

		4.2	Processing of digital evidence, digital images, damaged SIM and data recovery, multimedia evidence, retrieving deleted data: desktops, laptops and mobiles, retrieving data from slack space, renamed file, ghosting, compressed files.	
Total No. of Lectures				**60**

Course Outcomes:
1. To study the fundamentals of Computer Forensics
2. To learn, analyze and validate Forensics Data
3. To study the tools and tactics associated with Computer Forensics
4. To correctly define and cite appropriate instances for the application of computer forensics Correctly collect and analyze computer forensic evidence
5. Identify the essential and up-to-date concepts, algorithms, protocols, tools, and methodology of Computer Forensics
6. Understand the threats in networks and security concepts.

Books and References:
1. C. Altheide & H. Carvey Digital Forensics with Open Source Tools, Syngress, 2011. ISBN: 9781597495868.
2. "Modern Forensic Tools and Devices: Trends in Criminal Investigation": This book explores the latest technologies and techniques used in forensic investigations, including computer forensics, facial recognition, and more[2].
3. "Computer Organization" by Carl Hamacher, Zvonks Vranesic, and Safwat Zaky
4. "Computer Systems Architecture" by M. Morris Mano
5. "Computer Architecture: A Quantitative Approach" by John L. Hennessy and David A. Patterson:
6. "Modern Forensic Tools and Devices: Trends in Criminal Investigation"

CASE STUDY	
1	Case Study: Unauthorized Data Access at XYZ Corporation Background: XYZ Corporation, a medium-sized company, reported suspicious activity on their internal network.An employee noticed unauthorized access to sensitive files containing financial data and customer records.The company's IT team suspected an insider threat and decided to initiate a computer forensics investigation. Investigation Steps: Incident Identification: The IT team identified the compromised system and isolated it from the network. They collected relevant logs, system snapshots, and memory dumps. Evidence Preservation: The compromised system's hard drive was imaged using forensic tools (such as Encase or ProDiscover). The image was stored securely to prevent tampering. Forensic Analysis: Investigators analyzed the imaged data to identify signs of unauthorized access. They examined system logs, user accounts, and network traffic. Suspicious files and folders were flagged for further scrutiny. Timeline Reconstruction: Investigators reconstruct the timeline of events leading up to the breach. They correlated log entries, file modification timestamps, and user activity. User Interviews:

	The IT team interviewed employees who had access to the compromised system. They questioned the employee who reported the incident and any potential suspects. Digital Evidence Collection: Investigators collected evidence from various sources: Browser history: To trace the attacker's online activities. Registry entries: To identify installed software or modifications. Deleted files: To recover relevant data. Network traffic logs: To pinpoint communication with external servers. Analysis of Artifacts: Investigators analyzed artifacts such as: Browser cache: To find visited websites. Prefetch files: To determine executed programs. Recycle bin: To recover deleted files. Suspect Identification: Based on evidence, the prime suspect was identified: an employee with access to the compromised system. The suspect's computer was also imaged for further analysis. Legal Considerations: Investigators ensured compliance with relevant laws (e.g., Computer Misuse Act 1990 in the UK). They documented their findings meticulously for potential legal proceedings. Conclusion: The investigation revealed that the suspect had accessed sensitive files for personal gain. The company took appropriate action, including termination and legal measures.
2	Case Study: Suspicious Employee Activity Background A multinational corporation (XYZ Corp) suspects that one of its employees is involved in unauthorized data access and leaking sensitive information. The company's IT department has detected unusual network activity originating from the employee's workstation. Incident Details Employee: John Smith Position: Senior Systems Analyst Workstation: Dell OptiPlex 7010 Date and Time of Suspicious Activity: March 15, 2024, between 2:00 PM and 4:00 PM Observations Network Logs: The network logs show an unusually high volume of data transfers from John Smith's workstation to an external IP address. The external IP address is traced back to a competitor company (ABC Corp). File Access Logs: John accessed several confidential files during the suspicious time frame. The files include financial reports, product development plans, and customer databases. Email Activity: John sent an email with an attachment titled "XYZ Corp Financial Projections" to an unknown recipient. The email was sent from his corporate email account.

PRACTICALS

Practical No	Details
1	Live Case Studies
2	Linux commands
3	Open Source Forensic Tools
4	Disk Forensics and Data Recovery
5	Steganography
6	Keyloggers
7	Network monitors
8	Flowchart management
9	UML diagrams
10	eCommerce on websites

BOS	Computer Science
Class	M.Sc. Data Analytics Part - I
Semester	I
Course Name	Malware Analysis
Course Code	PMSDA104
Type of course	Major Elective
Level of the Subject	Advanced
Credit points	02 Theory+01 Practical

Course Objectives:
1. Understand , identify and analyze various malware types.
2. Recognize the basic signs of malware infection from a security analyst's perspective.

Unit No.	Unit Name	Topic No.	Content	Hours
1	Introduction to Malware Analysis & its techniques	1.1	**Malware Analysis:** Introduction, Techniques, Types of malware, General rules for Malware Analysis. **Basic Static Techniques**: Antivirus Scanning, Hashing, Finding Strings, Packed and Obfuscated Malware, Portable Executable Malware, Portable executable File Format, Linked Libraries and Functions, Static Analysis, The PE file headers and sections. **Malware Analysis in Virtual Machines:** Structure of VM, Creating and using Malware Analysis machine, Risks of using VMware for malware analysis, Record/Replay.	15
		1.2	**Basic Dynamic Analysis:** Sandboxes, Running Malware, Monitoring with process monitor, Viewing processes with process explorer, Comparing registry snapshots with regshot, Faking a network, Packet sniffing with Wireshark, Using INetSim, Basic Dynamic Tools. x86 Disassembly **IDA PRO:** Loading an executable, IDA Pro Interface, Using cross references, Analyzing functions, Using graphing options, Enhancing disassembly, Extending IDA with plug-ins.	
2	Malware Analyzing & Advanced Dynamic Analysis.	2.1	**Recognising C Code constructs in assembly:** Global v/s local variables, Disassembling arithmetic operations, recognizing if statements, recognizing loops, function call conventions, Analyzing switch statements, Disassembling arrays, Identifying structs, Analyzing linked list traversal. **Analyzing Malicious Windows Programs:** The windows API, The Windows Registry, Networking APIs Understanding running malware. Kernel v/s user mode, Native API.	15

		2.2	**Advanced Dynamic Analysis** – Debugging: Sourcelevel v/s Assembly-level debugging, kernel v/s user mode debugging, Using a debugger, Exceptions, Modifying execution with a debugger, modifying program execution **Advanced Dynamic Analysis – OLLYDBG:** Loading Malware, The Ollydbg Interface, Memory Map, Viewing threads and Stacks, Executing code, Breakpoints, Loading DLLs, Tracing, Exception handling, Patching, Analyzing shellcode, Assistance features, Plug-ins, Scriptable debugging.	
3	Working with malware activity.	3.1	**Kernel Debugging with WINDBG:** Drivers and kernel code, Using WinDbg, Microsoft Symbols, kernel debugging and using it, Rootkits, Loading drivers, kernel issues with windows. **Malware Functionality** – Malware Behavior: Downloaders and launchers, Backdoors, Credential stealers, Persistence mechanisms, Privilege escalation, covering the tracks. **Covert Malware Launching:** Launchers, Process injection, Process replacement, Hook injection, detours, APC injection.	15
		3.2	**Data Encoding:** Goal of Analyzing algorithms, Simple ciphers, Common cryptographic algorithms, Custom encoding, decoding. **Malware – focused network signatures:** Network countermeasures, Safely investigating attackers online, Content-Based Network Countermeasures, Combining Dynamic and Static Analysis Techniques, Understanding the Attacker's Perspective.	
4	Working with Anti-disassembly & Anti-Virtual machine techniques	4.1	**Anti-disassembly:** Concepts, Defeating disassembly algorithms, anti-disassembly techniques, Obscuring flow control, Thwarting stack-frame analysis. **Anti-debugging:** Windows debugger detection, debugger behavior, Interfering with debugger functionality, Debugger vulnerabilities **Anti-virtual machine techniques:** VMWare artifacts, Vulnerable functions, Tweaking settings, Escaping the virtual machine. **Packers and**	15
		4.2	**unpacking:** Packer anatomy, Identifying Packed Programs, Unpacking options, Automated Unpacking, Manual Unpacking, Common packers, Analyzing without unpacking, Packed DLLs, **Shellcode Analysis**: Loading shellcode for analysis, Position-independent Code, Identifying Execution Location, Manual Symbol Resolution, Shellcode encoding, NOP Sleds, Finding Shellcode. **64-bit Malware:** Why 64-bit malware? Differences in x64 architecture, Windows 32-bit on Windows 64-bit, 64- bit hints at malware functionality	
Total No. of Lectures				**60**

Course Outcomes:

After completion of the course, a student should be able to:
1. Understand various introductory techniques of malware analysis and creating the testing environment.
2. Perform advanced dynamic analysis and recognize constructs in assembly code.

3. Perform Reverse Engineering using OLLYDBG and WINDBG and study the behaviors and functions of malware
4. Understand data encoding, various techniques for anti-disassembly and anti-debugging.
5. Understand various anti virtual machine techniques and
6. Apply shellcode analysis of various languages along with x64 architecture.

Books and References:

1. "Learning Malware Analysis" by Monnappa K A
2. "Practical Malware Analysis: The Hands-On Guide to Dissecting Malicious Software" by Michael Sikorski and Andrew Honig
3. "Malware Analyst's Cookbook" by Michael Ligh, Steven Adair, and Blake Hartstein
4. "Malware Data Science: Attack Detection and Attribution" by Joshua Saxe and Hillary Sanders
5. "The Art of Memory Forensics: Detecting Malware and Threats in Windows, Linux, and Mac Memory" by Michael Hale Ligh, Andrew Case, Jamie Levy, and AAron Walters
6. "Antivirus Bypass Techniques: Learn Practical Techniques and Tactics to Combat, Bypass, and Evade Antivirus Software" by Miroslav Stampar

CASE STUDY	
1	Case Study: The Mysterious Malware Background You work as a cybersecurity analyst at a renowned security firm. Recently, your team received an anonymous tip about a suspicious executable file named "shadow.exe." The sender claims that this file is part of a sophisticated malware campaign targeting financial institutions. Your task is to analyze this malware and provide insights to protect potential victims. Scenario: Initial Inspection, You obtain a copy of "shadow.exe" and start your analysis. The file appears to be a Windows PE executable. Your antivirus scanner flags it as potentially malicious. Static Analysis Using IDA PRO, you load "shadow.exe." The disassembly reveals obfuscated code and multiple function calls. You notice references to the Windows Registry and network APIs. The malware seems to be packed or encrypted. Dynamic Analysis You run "shadow.exe" in a sandboxed environment. During execution, it creates registry keys, connects to an external IP address, and spawns additional processes. Wireshark captures network traffic, revealing communication with a command-and-control server. The malware exhibits anti-analysis techniques, such as detecting virtual machines. Debugging with OllyDbg You load "shadow.exe" into OllyDbg. The memory map shows suspicious sections. You set breakpoints to analyze specific code paths. Exception handling reveals evasion tactics. You identify shellcode execution.
2	Case Study: The Cryptic Ransomware Background You are part of an incident response team at a cybersecurity company. A client, a small

manufacturing firm, has fallen victim to a ransomware attack. Their critical systems are encrypted, and the attackers demand a hefty sum in cryptocurrency for the decryption key. Your mission is to analyze the ransomware, understand its behavior, and assist in recovery.
Scenario. Initial Assessment
The affected systems display a ransom note with instructions.
The ransomware executable is named "cryptolock.exe."
The client's backup servers are also compromised.
Static Analysis
You obtain a copy of "cryptolock.exe."
Using IDA PRO, you disassemble the binary.
The code reveals encryption routines, file manipulation, and network communication.
The malware appears to use a custom encryption algorithm.
Dynamic Analysis
In a controlled environment, you execute "cryptolock.exe."
It scans local drives, encrypts files, and modifies file extensions.
Network traffic analysis shows communication with a remote server.
The ransomware avoids virtual machines and sandboxes.
Decryption Attempts
You analyze the ransom note for clues.
The attackers demand payment in Bitcoin.
You explore the possibility of finding weaknesses in the encryption algorithm.
The client is desperate to recover their critical data.

PRACTICALS

Practical No	Details
1	Antivirus Scanning: ○ Assignment: Set up an antivirus scanner (such as ClamAV) and analyze a suspicious file using it. ○ Tools: ClamAV, VirusTotal.
2	Hashing and Finding Strings: ○ Assignment: Calculate hash values (MD5, SHA-256) for a set of files. Extract strings from a binary. ○ Tools: md5sum, sha256sum, strings.
3	Recognizing C Code Constructs in Assembly: ○ Assignment: Disassemble x86 code snippets. Identify global vs. local variables, loops, if statements, and function call conventions. ○ Tools: IDA Pro, objdump.
4	Analyzing Malicious Windows Programs: ○ Assignment: Investigate Windows API calls, registry interactions, and network-related activities in a malware sample. ○ Tools: Process Monitor, Process Explorer, Wireshark.
5	Dynamic Analysis in Virtual Machines: ○ Assignment: Set up a virtual machine for malware analysis. Understand risks and record/replay techniques. ○ Tools: VMware, VirtualBox.

6	Kernel Debugging with WinDbg:
	   ○ Assignment: Debug a kernel-mode driver. Analyze memory maps, threads, and exceptions.
	   ○ Tools: WinDbg, Driver Verifier.

7	Advanced Dynamic Analysis with OllyDbg:
	   ○ Assignment: Load malware in OllyDbg. Explore memory, breakpoints, loading DLLs, and analyzing shellcode.
	   ○ Tools: OllyDbg.

8	Malware Behavior Analysis:
	   ○ Assignment: Run malware samples and observe their behavior. Understand user vs. kernel mode execution.
	   ○ Tools: Procmon, API Monitor.

9	Anti-Disassembly Techniques:
	   ○ Assignment: Create a simple piece of code (in C/C++ or assembly) that employs anti-disassembly techniques. Use manual crafting to confuse disassembly tools (such as IDA Pro or radare2). For instance, you can insert junk instructions, reorder code, or use obfuscation.
	   ○ Tools: IDA Pro, radare2.

10	Anti-Virtual Machine Techniques:
	   ○ Assignment: Develop a shellcode that detects if it's running inside a virtual machine (e.g., VMware). Investigate VMWare artifacts (e.g., registry keys, device names) that can be used for detection. Experiment with tweaking VM settings (e.g., CPUID masking) to evade detection.
	   ○ Tools: VMware, VirtualBox, OllyDbg.

BOS	Computer Science
Class	M.Sc. Data Analytics Part - I
Semester	I
Course Name	Research Methodology
Course Code	PMSDA107
Type of course	Research Methodology
Level of the Subject	Advanced
Credit points	02 Theory+02 Practical

Course Objectives:
    1. To be able to conduct research with an understanding of all the latest theories.
    2. To develop the ability to explore research techniques used for solving any real world or innovative problem.

Unit No.	Name of Unit	Topic No.	Content	Hours
1	Introduction to Research	1.1	Meaning, Characteristics of Research, Objectives of research Types of research: Basic Research, Applied Research, Descriptive Research, Analytical Research, Empirical research	15
		1.2	Ethical issues and Problems in research, Meaning of research methodology, Stages in Scientific Research Process , Measurement concepts: Scales of measurement - nominal, ordinal, ratio and interval, Types of measurement scales:comparative scales-Paired Comparison Scale,Rank Order Scale,Q-Sort Scale and non comparative scales-Continuous Rating Scale, Itemized Rating Scale, Likert Scale, Stapel Scale, Semantic Differential Scale.	
2	Sampling Design and Questionnaire Design	2.1	Sampling design- meaning and significance, Essentials of good sampling, Stages in sampling design, Sampling errors. Types of Probability sampling- simple random sampling, stratified random sampling, systematic random sampling, and cluster random sampling.Types of non-probability sampling-Convenience sampling, Quota sampling, Self-selection (volunteer) sampling, Snowball sampling, Purposive (judgmental) sampling.	15

		2.2	Types of data- Primary data and secondary data meaning, significance and limitations, Collection of primary and secondary data Designing of a questionnaire- meaning, stages in questionnaire designing, Essentials of a good questionnaire.	
3	Data Analysis and Presentation:	3.1	Determination of sample size, Editing and Coding Basic Data Analysis Arithmetic Mean, Median, mode, Standard deviation, Correlation, Regression analysis	15
		3.2	Hypothesis testing for significance, Types of errors in Hypothesis testing Hypothesis testing for significance for mean/s and proportion for large samples, t test for single mean, paired and unpaired means, Chi-Square test. ANOVA-One way and Two way.	
	Total No. of Lectures			**45**

Course Outcomes :

After the deliverance of teaching module, it is desired that the student has:

1. Necessary skills to apply appropriate methods to the research in hand
2. Acquire aptitude for establishing rapport in the field
3. Basic knowledge about sampling methods
4. The ability to scan secondary sources for research
5. Acquire skills to analyze the quantitative and qualitative data
6. Acquire knowledge in using statistical tools and techniques
   .

Reference books:

1. "Research Methodology: Methods and Techniques" by C.R. Kothari
2. "Research Design: Qualitative, Quantitative, and Mixed Methods Approaches" by John W. Creswell and J. David Creswell
3. "Research Methods for Business Students" by Mark N.K. Saunders, Philip Lewis, and Adrian Thornhill
4. "Social Research Methods" by Alan Bryman
5. "Measurement and Scaling in Marketing Research" by Philip Kotler, Haider Ali, and Irfan Haider
6. "Sampling Techniques" by William G. Cochran

PRACTICALS

Practical No.	Details
1	Using R execute the basic commands

2	Import the data from Excel / .CSV find mean median mode,standard deviation variance
3	Perform R program for making Diagrams(Bar Diagram,Multiple Bar Diagram,Pie Chart)
4	Perform R program for making Graphs(Histogram,Frequency Polygon,Ogive)
5	Import the data from Excel / .CSV and perform the Chi-square Test,goodness of fit ,Independence of attributes
6	Perform an R program on z-test- one population mean, Two population means.One population proportion,two population proportion.
7	Perform an R program on t test- one sample, paired and unpaired
8	Perform an R program on Non ParametricTest -Sign test, wilcoxon signed rank test
9	Perform an R program on One way ANOVAand Two way ANOVA
10	Perform an R program on Friedman Test and Kruskal Wallis test

Semester-II

BOS	Computer Science
Class	M.Sc. Data Analytics Part - I
Semester	II
Course Name	Advanced Statistical Methods and Testing of Hypothesis
Course Code	PMSDA201
Type of course	Major
Level of the Subject	Advanced
Credit points	04 Theory+01 Practical

Course Objectives:
Students will get to know
1. Advance statistical concepts and some of their basic applications in real world,
2. The appropriateness of statistical analyses, results, and inferences , and Advanced data analysis in R / Python.

Unit No.	Name of Unit	Topic No.	Content	Hours
1	Estimation	1.1	Point Estimation, Method of moments, Likelihood function, Maximum likelihood equations. Mean square error, Unbiased estimator Minimum variance unbiased estimator, Uniformly minimum variance unbiased estimator.	15
		1.2	Consistent estimator, Efficient estimator, Sufficient estimator. Sampling distribution of mean and proportion	
2	Concept of a sampling distribution	2.1	Confidence Interval for single mean for large and small sample, Confidence Interval for difference of means for large and small sample Confidence Interval for single proportion and difference in proportions for large samples.	15
		2.2	Confidence interval for variance. Normal distribution, Students' t distribution, F distribution, chi square distribution	
3	Parametric test	3.1	Concept of Hypothesis, one tailed or two tailed test, level of significance, types of error. Hypothesis test for single mean , difference of means, for large and small sample, paired t test,Hypothesis test for single proportion, difference of proportions for large sample	15

		3.2	Chi square test for variance (single sample), *F*-Test for Equality of Two Variance Analysis of variance-One way ANOVA, Two way ANOVA, Post HOC ANOVA-Fishers' LSD	
4	Continuous Distributions:	4.1	Chi square test for goodness of fit, Chi square test for independence of attributes, Yates correction Kolmogrov Smirnov test for one sample and two sample, Fisher's exact test	**15**
		4.2	Wilcoxon test, Mann Whitney test, Kruskal Wallis test Comparison between parametric and nonparametric test, uses of non parametric test.	
Total No. of Lectures				**60**

Course Outcomes

1. Perform inference on parameter estimates, including constructing confidence intervals and testing.
2. Knowledge about formulating and testing a hypothesis, using critical values to draw conclusions and determining probability of making errors in hypothesis tests.
3. Knowledge about large sample tests and its applications and get an idea of order statistics with its applications.
4. Knowledge about small sample tests based on Chi-square, t and F distributions.
5. Differentiate between ANOVA and post HOC ANOVA test.
6. Knowledge about simple partial and Logistic regression.

Books and References:

1. "Statistical Inference" by George Casella and Roger L. Berger
2. "All of Statistics: A Concise Course in Statistical Inference" by Larry Wasserman
3. "Mathematical Statistics and Data Analysis" by John A. Rice
4. "Introduction to Mathematical Statistics" by Robert V. Hogg, Joseph W. McKean, and Allen T. Craig
5. "Statistical Methods" by Rudolf J. Freund, William J. Wilson, and Ping Sa
6. "Statistical Inference" by George E. P. Box, George C. Tiao, and Gwilym M. Jenkins

CASE STUDY	
1	In a pharmaceutical company, researchers are investigating the effectiveness of four different drug formulations in treating a particular condition. The formulations, labeled A, B, C, and D, have undergone preliminary testing, but now the researchers want to determine if there are any significant differences in their effectiveness. The primary objective of this study is to analyze the efficacy of the four drug formulations and determine if there are any statistically significant differences between them. The researchers conducted a randomized controlled trial involving 80 participants diagnosed with the condition. The participants were randomly assigned to one of the four treatment groups: A, B, C, or D. After a specified treatment period, the researchers measured the improvement in each participant condition using a standardized scale. Let assume the following hypothetical data represents the improvement scores of participants in each treatment group:

	Treatment A: 12, 15, 18, 14, 17 Treatment B: 13, 16, 19, 12, 15 Treatment C: 11, 14, 17, 13, 16 Treatment D: 10, 13, 16, 11, 14
2	A company is evaluating the effectiveness of two different advertising strategies, Strategy A and Strategy B, to promote a new product launch. The marketing team decides to conduct a study to compare the impact of these strategies on brand awareness among potential customers. They randomly select two groups of participants: Group A will be exposed to Strategy A, while Group B will be exposed to Strategy B. After the exposure, the participants will be asked to complete a survey to measure their brand awareness levels. Data: The following are the brand awareness scores (out of 100) obtained from 15 participants in each group: Group A (Strategy A): 78, 82, 75, 80, 79, 81, 77, 83, 76, 80, 84, 79, 82, 81, 78 Group B (Strategy B): 85, 88, 82, 86, 84, 87, 83, 89, 85, 87, 86, 83, 88, 84, 86

PRACTICALS

Practical No.	Details
1	Confidence intervals for single proportion and difference of two proportions
2	Testing of significance for single proportion and difference of two proportions
3	Confidence intervals for single mean and difference of two means
4	Testing of significance for single mean and difference of two means and paired tests.
5	Testing of independence of attributes using Chi square test.
6	F test
7	ANOVA (ONE way and TWO way)
8	Post HOC ANOVA test
9	Correlation analysis
10	Regression Analysis

BOS	Computer Science
Class	M.Sc. Data Analytics Part - I
Semester	II
Course Name	Big Data Analytics
Course Code	PMSDA202
Type of course	Major
Level of the Subject	Advanced
Credit points	04 Theory+01 Practical

Course Objectives:
1. To introduce the tools required to manage and analyze big data like Hadoop, NoSql MapReduce.
2. To teach the fundamental techniques and principles in achieving big data analytics with scalability and streaming capability.

Unit No.	Name of Unit	Toipc No.	Content	Hours
1	Introduction to Big Data & Big Data Analytics	1.1	Introduction to Big Data, Characteristics of Data, and Big Data Evolution of Big Data, Definition of Big Data, Challenges with big data, Why Big data? Data Warehouse environment, Traditional Business Intelligence versus Big Data. State of Practice in Analytics, Key roles for New Big Data Ecosystems, Examples of Big Data Analytics.	15
		1.2	Big Data Analytics, Introduction to big data analytics, Classification of Analytics, Challenges of Big Data, Importance of Big Data, Big Data Technologies, Data Science, Responsibilities, Soft state eventual consistency. Data Analytics Life Cycle	
2	Analytical Theory and Methods:	2.1	Clustering, Hierarchical Clustering, K-Means Clustering algorithm. Associated Algorithms, Association Rules, Apriori Algorithm, Candidate Rules, Applications of Association Rules, Validation and Testing, Diagnostics.	15
		2.2	Regression, Linear Regression, Logistic Regression, Additional Regression Models. Classification, Decision Trees, Naïve Bayes, Diagnostics of Classifiers, Additional Classification Methods.	
3	Analytical Theory	3.1	Time Series Analysis, Box Jenkins methodology, ARIMA Model, Additional methods Text Analysis, Steps, Text Analysis Example, Collecting Raw Text, Representing Text, Term Frequency-Inverse	15

	Methods & Hadoop Spark.		Document Frequency (TFIDF), Categorizing Documents by Topics, Determining Sentiments	
		3.2	Advanced Analytics-Technology and Tools: Hadoop Architecture, Working with Hadoop Distributed file system, Hadoop Environment, Data Science Pipeline and Hadoop Ecosystem, Hadoop - HDFS Operations, Hadoop - MapReduce, MapReduce example. Spark: What is Spark, Spark Architecture, Spark Components, Spark Installation, Spark RDD, What is RDD, RDD Operations, RDD Persistence, In-built Functions, Spark Examples :word count example,char count Example	
4	Advanced Analytics-Technology and Tools with various Databases.	4.1	HBase :What is HBase, HBase Model, HBase Read, HBase Write, HBase MemStore, HBase Installation, RDBMS vs HBase, HBase Commands, HBase Example. Hive : What is Hive, Hive Architecture, Hive Installation, Hive Data Types, Create Database, Drop Database, Create Table, Load Data, Drop Table, Alter Table, Static Partitioning, Dynamic Partitioning, Bucketing in Hive, HiveQL - Operators, HiveQL - Functions, HiveQL - Group By & Having, HiveQL - Order By & Sort BY Clause,HiveQL - Join	15
		4.2	Sqoop : What is Sqoop, Sqoop Installation, Starting Sqoop, Sqoop Import,Sqoop Where,Sqoop Export Pig : What is Pig,Pig Installation,Pig Run Modes,Pig Latin Concepts,Pig Data Types,Pig Example,Pig UDF, Relational Operators, Eval Functions, Math Function.	
Total No. of Lectures				60

Course Outcomes

1. Identify Big Data and its Business Implications.
2. List the types of analytical methods in big data.
3. Differentiate between Hbase and Hive databases.
4. Explain Job Execution in Hadoop Environment using Spark.
5. Analyze data by using pig language.
6. Develop Big Data Solutions using Hadoop EcoSystem.

Books and References:

1. Big Data and Analytics Subhashini Chellappan Seema Acharya Wiley
2. Data Analytics with Hadoop An Introduction for Data Scientists Benjamin Bengfort and Jenny Kim O'Reilly 2016
3. Big Data and Hadoop V.K Jain Khanna Publishing First 2018.
4. "Big Data: Principles and Best Practices of Scalable Realtime Data Systems" by Nathan Marz and James Warren
5. "Data Science for Business" by Foster Provost and Tom Fawcett
6. Apache Sqoop Cookbook – Unlocking Hadoop for Your Relational Database by Kathleen Ting & Jarek Jarec Cecho & Handbook on Pig Farming and Pork Processing

CASE STUDY	
1	Case Study: Hospital Readmission Prediction Using Association Rules Background: A large hospital system wants to reduce the rate of patient readmissions. They collect data on patients' demographics, medical history, and other relevant factors. The goal is to identify patterns that can help predict which patients are likely to be readmitted within 30 days after discharge. Data: The dataset contains the following features: Age: Patient's age (numeric). Gender: Patient's gender (categorical: Male/Female). Diagnosis: Primary diagnosis (categorical: Diabetes, Heart Disease, etc.). Medication: Medications prescribed (categorical: Insulin, Beta-blockers, etc.). Length of Stay: Number of days in the hospital (numeric). Readmission: Binary variable (Yes/No) indicating whether the patient was readmitted within 30 days. Approach: Data Preprocessing: Clean the data, handle missing values, and encode categorical variables. Create a binary target variable (1 for readmission, 0 otherwise). Association Rule Mining: Use the Apriori algorithm to discover frequent itemsets and association rules. Set appropriate support and confidence thresholds. Extract rules that indicate strong associations between patient characteristics and readmission. Interpretation: Analyze the mined rules to understand which factors contribute to readmission risk. Identify actionable insights for hospital management.
2	Case Study: Real-Time Data Processing with Hadoop Ecosystem Background: A retail company wants to analyze customer behavior in real-time to improve sales and marketing strategies. They collect data from various sources, including online transactions, social media, and customer feedback. The goal is to gain insights into customer preferences, identify trends, and personalize marketing campaigns. Data: The dataset includes the following features: Transaction Data: Purchase history, product details, and timestamps. Social Media Data: Customer interactions, likes, shares, and comments. Feedback Data: Ratings, reviews, and sentiment analysis. Approach: Data Ingestion: Use Sqoop to import transaction data from a MySQL database into HDFS. Social media data is collected using APIs and stored directly in HDFS. Data Processing: MapReduce processes transaction data to calculate metrics like total sales, popular products, and customer segments. Spark performs real-time analytics on social media data, identifying trending topics and sentiment patterns. Data Warehousing: Hive creates external tables over the transaction data in HDFS.

	Use HiveQL to query and join data for business intelligence reports. Machine Learning: Spark MLlib trains recommendation models based on customer behavior. Mahout provides collaborative filtering for personalized product recommendations. Workflow Management: Oozie schedules and orchestrates data processing workflows. Automate data pipeline execution.

PRACTICALS

Practical No	Details
1	Association Rules (Apriori Algorithm)
2	Regression Model a. Single Linear Regression b. Multiple Linear Regression c. Logistic Regression
3	Clustering Model a. Hierarchical Clustering b. K-Means Clustering
4	Classification Model a. Decision Tree Classification b. Naïve Bayes Classification
5	Working with Time Series Analysis & Text Analysis
6	Install, configure and run Hadoop and HDFS & Working with Spark.
7	Implement word count / frequency programs using MapReduce
8	Implement an application that stores big data in Hbase / MongoDB and manipulate it using R / Python.
9	Working with Pig & Sqoop..
10	Configure the Hive and implement the application in Hive.

BOS	Computer Science
Class	M.Sc. Data Analytics Part - I
Semester	II
Course Name	Cloud Computing and Virtualization
Course Code	PMSDA203
Type of course	Major
Level of the Subject	Advanced
Credit points	04 Theory+01 Practical

Course Objectives:

1. Computing concepts, technologies, architecture, implantations and applications.
2. To expose the learners to frontier areas of Cloud Computing, while providing sufficient foundations to enable further study and research.

Unit No.	Name of Unit	Topic No.	Content	Hours
1	Introduction to Cloud Computing & service models	1.1	Cloud Computing : The Cloud computing model, Cloud Deployment models – Private, Public and Hybrid models,	15
		1.2	Cloud service models- IaaS, PaaS, SaaS. Issues with performance, security, ROI, maintenance, QoS and ownership. Service Level Agreements, Cloud Data Centers	
2	Security	2.1	Security Issues in Cloud Computing: Common Types of Attacks and Policies, Cloud Security Alliance's Notorious nine issues.	15
		2.2	Security issues with Virtual Machines, IAM Systems, Authentication and Authorization, The OAuth protocol. Security at Cloud Data Centers, Data Security issues.	
3	Virtualization	3.1	Virtualization: Requirements of Virtualization, Hardware Virtualization and its types, Bare metal versus hosted Hypervisors, Storage Virtualization, Network Attached Storage versus Storage Area Networks. Virtualization versus Containers, Network Virtualization and Software Defined Networks.	15
		3.2	Load Balancing Based Live Migration: Introduction, Classification of Load Balancing Techniques, Policy engine, Load balancing Algorithm,Resource Load Balancing, Load Balancing in Virtual Infrastructure Management Software,	

4	Cloud computing in Practice	4.1	VMware Distributed Resource Scheduler. Cloud Computing in Practice: Globally reputed vendors and providers in categories of IaaS, PaaS, SaaS. Case studies of Amazon Web Services, Microsoft Azure, Google App Engine. SalesForce and Dropbox.	15
		4.2	Launching Virtual machines and remote login using SSH , using SSH clients like Putty, Deploying server side applications using services like Amazon BeanStalk etc., Using Relational and Flat storage services, deploying Load balancers on clouds.	
Total No. of Lectures				60

Course Outcomes:

1. Explain the core concepts of the cloud computing paradigm: how and why this paradigm shift came about, the characteristics, advantages and challenges brought about by the various models and services in cloud computing.
2. Identify security issues in cloud computing, Apply the fundamental concepts in data centers
3. Describe the importance of virtualization along with their technologies.
4. Explain load balancing techniques and policies
5. Use and Examine different cloud computing services analyze the components of open stack & Google Cloud platform and understand Mobile Cloud Computing. Describe the key components of Amazon web Service
6. Design & develop backup strategies for cloud data based on features.

Books and References:

1. Design and Use of Virtualization Technology in Cloud Computing by Das, Prashanta Kumar, Deka, Ganesh Chandra , IGI Global
2. Mastering Cloud Computing by Rajkumar Buyya and Christian Vecchiola , Tata McGraw-Hill Education
3. Virtualization: Requirements, Types, and Best Practices: "Mastering Virtualization" by Jon Kuhn:
4. Load Balancing and Cloud Infrastructure:"Load Balancing Servers, Firewalls, and Caches" by Chandra Kopparapu
5. Cloud Computing in Practice:"Amazon Web Services (AWS) Case Studies"
6. "Microsoft Azure Case Studies"

CASE STUDY	
1	**Case Study Example 1:** Japan's first digital bank, Minna Bank, as an organization to implement digital transformation in a single step without the constraints of the existing business. This bank was the first bank in the world to build a full cloud banking system, and the system was built in the midst of a pandemic, with overwhelming speed.Accenture is providing support in the adoption of Agile development and in multiple areas such as automation, strategy and talent development. Its Banking, Strategy & Consulting, Technology, and Interactive teams have come together from Fukuoka, Osaka, Tokyo, Aizu, Hokkaido and two overseas locations, transcending national and organizational boundaries to partner with Minna

	Bank.In addition to its own resources, Accenture has drawn on its vast ecosystem of technology partners—in this case, industry leaders such as Google, Microsoft, AWS, Salesforce and Oracle—to take advantage of their solutions and best practices.Specifically, in the "Zero Bank Core Solution" jointly developed by Minna Bank and Accenture, the core system will be implemented on Google Cloud using Accenture's Digital Experience and cloud-first approach, connected technology and cloud-native core solution. For contact center operations, Amazon Web Services (AWS), Amazon Connect and Salesforce's Service Cloud have been combined. Microsoft's Azure is being used for the virtual desktop infrastructure for employee and system operations, and Oracle Cloud is being used for the accounting system. Collaboration with these solution providers has allowed Minna Bank to build its foundation as a cloud-first business with the latest technology available worldwide. Minna Bank has three core business concepts: Give shape to everyone's voice—provide new financial services in line with changes in customer behavior. Deliver the best for everyone—become a comprehensive financial concierge based on an understanding of customers. Integrate into people's daily lives—realize the concept of a BaaS (Banking as a Service) business.
2	Pinterest on AWS Pinterest is a visual discovery engine that hosts billions of images for over 450 million users to explore, save, and share as "Pins" to personalized digital inspiration boards. Born in the Amazon Web Services (AWS) Cloud, Pinterest can scale processing, storage, and analysis of its rapidly increasing data, all while reducing infrastructure management and focusing on innovation. Using compute solutions from AWS, Pinterest migrated its iOS continuous integration and continuous delivery (CI/CD) pipeline from on-premises to reduce build failures by over 80 percent. Pinterest also uses machine learning (ML) to power its visual search tool Pinterest Lens that can recognize over 2.5 billion objects and match them to products. Today, Pinterest's exabyte data platform runs entirely on AWS, scaling log search and analytics to over 1.7TB while reducing operations costs by 30%.

PRACTICALS

Practical No	Details
1	Study and implementation of Infrastructure as a Service, Storage as a Service.
2	Study Cloud Security management.
3	Show the implementation of web services: Implementing "Big" Web Service.
4	Implementing Web Service that connects to MySQL database.
5	Implement xen virtualization and manage with xencenter
6	Implement virtualization using VMWare ESXi Server and managing with vCenter
7	Implement Windows Hyper V virtualization.
8	Develop application for Microsoft Azure.
9	Develop an application for Google App Engine.
10.	Deploying applications to Elastic Beanstalk environments.

BOS	Computer Science
Class	M.Sc. Data Analytics Part - I
Semester	II
Course Name	Modern Networking
Course Code	PMSDA204
Type of course	Major Elective
Level of the Subject	Advanced
Credit points	02Theory+01 Practical

Course Objectives:
1. To understand the state-of-the-art in network protocols, architectures and applications & Analyze existing network protocols and networks.
2. To investigate novel ideas in the area of Networking via term-long research projects.

Unit No.	Name of Unit	Topic No.	Content	Hours
1	Introduction to Modern Networking	1.1	Modern Networking Elements of Modern Networking The Networking Ecosystem ,Example Network Architectures,Global Network Architecture,A Typical Network Hierarchy Ethernet Applications of Ethernet Standards Ethernet Data Rates Wi-Fi Applications of Wi-Fi,Standards Wi-Fi Data Rates 4G/5G Cellular First Generation Second Generation, Third Generation Fourth Generation Fifth Generation, Cloud Computing Cloud Computing Concepts The Benefits of Cloud Computing Cloud Networking Cloud Storage, Internet of Things Things on the Internet of Things, Evolution Layers of the Internet of Things, Network Convergence Unified Communications, Requirements and Technology Types of Network and Internet Traffic,Elastic Traffic,Inelastic Traffic, Real-Time Traffic Characteristics Demand: Big Data, Cloud Computing, and Mobile	10
		1.2	TrafficBig Data Cloud Computing,,Mobile Traffic, Requirements: QoS and QoE,,Quality of Service,Quality of Experience, Routing Characteristics, Packet Forwarding, Congestion Control ,Effects of Congestion,Congestion Control Techniques, SDN and NFV Software- Defined Networking,Network Functions Virtualization Modern Networking Elements	

2	Software Defined Networks	2.1	Software-Defined Networks SDN: Background and Motivation, Evolving Network Requirements Demand Is Increasing,Supply Is IncreasingTraffic Patterns Are More ComplexTraditional Network Architectures are Inadequate,	10
		2.2	The SDN Approach Requirements SDN Architecture Characteristics of Software Defined Networking, SDN- and NFV-Related Standards Standards- Developing Organizations Industry Consortia Open Development Initiatives, SDN Data Plane and OpenFlow SDN Data Plane, Data Plane Functions Data Plane Protocols OpenFlow Logical Network Device Flow Table Structure Flow Table Pipeline, The Use of Multiple Tables Group Table OpenFlow Protocol, SDN Control PlaneSDN Control Plane Architecture Control Plane Functions, Southbound Interface Northbound InterfaceRouting, ITU-T Model, OpenDaylight OpenDaylight Architecture OpenDaylight Helium, REST REST Constraints Example REST API	
3	Virtualization & Network Functions Virtualization	3.1	Virtualization, Network Functions Virtualization: Concepts and Architecture, Background and Motivation for NFV, Virtual Machines The Virtual Machine Monitor, Architectural Approaches Container Virtualization, NFV Concepts Simple Example of the Use of NFV, NFV Principles High-Level NFV Framework, NFV Benefits and Requirements NFV Benefits, NFV Requirements, NFV Reference Architecture NFV Management and Orchestration, Reference Points Implementation, NFV Functionality, NFV Infrastructure,Container Interface,Deployment of NFVI Containers,Logical Structure of NFVI Domains,ComputeDomain, Hypervisor Domain,Infrastructure Network Domain, Virtualized Network Functions, VNF Interfaces,VNFC to VNFC Communication,VNF Scaling, NFV Management and Orchestration, Virtualized Infrastructure Manager,Virtual Network Function Manager,NFV Orchestrator, Repositories, Element Management, OSS/BSS, NFV Use Cases Architectural Use Cases, Service-Oriented Use Cases, SDN and NFV Network Virtualization, Virtual LANs ,The Use of Virtual LANs,Defining VLANs, Communicating VLAN Membership,IEEE	10
		3.2	802.1Q VLAN Standard, Nested VLANs, OpenFlow VLAN Support, Virtual Private Networks, IPsec VPNs,MPLS VPNs, Network Virtualization, Simplified Example, Network Virtualization Architecture, Benefits of Network Virtualization, OpenDaylight's Virtual Tenant Network, Software-Defined Infrastructure,Software-Defined	

				Storage, SDI Architecture	
4	Quality of Service	4.1		Defining and Supporting User Needs, Quality of Service, Background, QoS Architectural Framework, Data Plane, Control Plane, Management Plane, Integrated Services Architecture, ISA Approach ISA Components, ISA Services, Queuing Discipline, Differentiated Services, Services, DiffServ Field, DiffServ Configuration and Operation, Per-Hop Behavior, Default Forwarding PHB, Service Level Agreements, IP Performance Metrics, OpenFlow QoS Support, Queue Structures, Meters, QoE: User Quality of Experience, Why QoE?,Online Video Content Delivery, Service Failures Due to Inadequate QoE Considerations QoE-Related Standardization Projects, Definition of Quality of Experience, Definition of Quality, Definition of Experience Quality Formation Process, Definition of Quality of Experience, QoE Strategies in Practice, The QoE/QoS Layered Model Summarizing and Merging the ,QoE/QoS Layers, Factors Influencing QoE, Measurements of QoE, Subjective Assessment, Objective Assessment, End-User Device Analytics	10
		4.2		Modern Network Architecture: Clouds and Fog, Cloud Computing, Basic Concepts, Cloud Services, Software as a Service, Platform as a Service, Infrastructure as a Service, Other Cloud Services, XaaS, Cloud Deployment Models, Public Cloud Private Cloud Community Cloud, Hybrid Cloud, Cloud Architecture, NIST Cloud Computing Reference Architecture,ITU-T Cloud Computing Reference Architecture, SDN and NFV, Service Provider Perspective Private Cloud Perspective, ITU-T Cloud Computing Functional Reference Architecture, The Internet of Things: Components The IoT Era Begins, The Scope of the Internet of Things Components of IoT-Enabled Things, Sensors, Actuators, Microcontrollers, Transceivers, RFID, The Internet of Things: Architecture and Implementation, IoT Architecture,ITU-T IoT Reference Model, IoT World Forum Reference Model, IoT Implementation, IoTivity, Cisco IoT System, ioBridge, Security Security Requirements, SDN Security Threats to SDN, Software-Defined Security, NFV Security, Attack Surfaces, ETSI Security Perspective, Security Techniques, Cloud Security, Security Issues and	
Total No. of Lectures					40

Course Outcomes:
1. Understanding the elements of modern networking.
2. Demonstrate in-depth knowledge in the area of Computer Networking to work with software defined networks.
3. Differentiate between virtualization and software network virtualization.

4. Understand the architecture of modern networking.
5. Demonstrate scholarship of knowledge through performing in a group to identify, formulate and solve a problem related to Computer Networks
6. Prepare a technical document for the identified Networking System Conducting experiments to analyze the identified research work in building Computer Networks

Books and References:
1. Foundations of Modern Networking: SDN, NFV, QoE, IoT, and Cloud William Stallings AddisonWesley Professional October 2015
2. SDN and NFV Simplified A Visual Guide to Understanding Software Defined Networks and Network Function Virtualization Jim Doherty Pearson Education, Inc
3. Network Functions Virtualization (NFV) with a Touch of SDN Rajendra Chayapathi Syed Farrukh Hassan AddisonWesley
4. CCIE and CCDE Evolving Technologies Study Guide Brad dgeworth, Jason Gooley, Ramiro Garza Rios Pearson Education, Inc 2019
5. Virtualization & Network Functions Virtualization (NFV): "Network Function Virtualization" by O'Reilly Media[3]:
6. Quality of Service (QoS): "Service Quality" by Roland T. Rust and Richard L. Oliver.

CASE STUDY	
1	Modern Networks for Competitive Advantage A modern network seamlessly connects and enables various layers of the digital core. This core comprises: Cloud: Leveraging cloud services for scalability, agility, and cost-effectiveness. Data and AI: Harnessing data analytics and artificial intelligence to drive insights and decision-making. Applications and Platforms: Facilitating efficient application delivery and user experiences.Incorporating other new technologies as needed.The digital core acts as a catalyst for future innovation, fueling the creation of new products, services, and experiences that drive total enterprise reinvention. Rapid Enterprise Digitalization Outpaces Network Capacity and Spend: Over the past 15 years, digitalization (measured by the adoption of digital technologies, data growth, cloud adoption, and connected endpoints) has increased a staggering 162 times. However, network capacity has grown only 43 times, and innovation spend on networks has grown just two-fold. The gap between digitalization and network capacity is widening, necessitating increased investment in network modernization.The surge in new technologies (such as generative AI, edge computing, and 5G) and the volume of data generated further strain current networks. Network Limitations Impede Total Enterprise Reinvention: Businesses recognize that existing networks are becoming bottlenecks. In a global survey of business and IT executives, 87% of respondents acknowledged that the growing data demands of their new AI systems have outstripped the current capabilities of networks. To address this, enterprises must commit to continuous, dynamic reinvention, leveraging next-gen technologies like cloud, edge computing, 5G, data analytics, and AI.
2	Case Study: Next-Generation Wireless LAN Upgrade at Cisco Background: In the year 2000, Cisco IT designed and deployed a global Wireless LAN (WLAN) infrastructure to serve all Cisco offices. Initially conceived as a secondary network for intermittent data usage, the WLAN quickly gained popularity among Cisco's highly mobile workforce. Within just two years, nearly 25% of Cisco employees were using the WLAN as their primary network access method. By 2005, it became evident that an upgrade was necessary due to increased user adoption and changing business requirements.

Challenges:
User Adoption: The original WLAN, initially considered secondary, had become business-critical for the majority of Cisco employees.
End-of-Life Components: The existing infrastructure was aging, with many components no longer sold or supported.
Performance Limitations: The WLAN struggled to handle high levels of wireless voice and video traffic.
Goals:
Deploy an enterprise-class, on-demand wireless network suitable as a primary access medium.
Support at least 50% of users adopting wireless as their regular network access method.
Improve service availability, stability, security, and overall Service Level Agreement (SLA).
Solution:
Cisco embarked on the Next-Generation WLAN program in May 2006.
The program aimed to evolve the existing indoor wireless network infrastructure into a more available, stable, and secure network.
Key focus areas included native support for wireless voice and video and enhanced accessibility.
Results:
The upgraded WLAN architecture provided:
Higher bandwidth and broader coverage for users.
Substantial cost savings and ongoing operational efficiencies.
Continued improvements in employee productivity.
Cisco IT's real-world experience in this area helped support similar enterprise needs.

PRACTICALS

Practical No	Details
1.	Configure IP SLA Tracking and Path Control Topology
2.	Using the AS_PATH Attribute
3.	Configuring IBGP and EBGP Sessions, Local Preference, and MED
4.	Secure the Management Plane
5.	Configure and Verify Path Control Using PBR
6.	IP Service Level Agreements and Remote SPAN in a Campus Environment
7.	Inter-VLAN Routing
8.	Simulating MPLS environment
9.	Simulating VRF
10.	Simulating SDN with OpenDaylight SDN Controller with the Mininet Network Emulator . OFNet SDN network emulator

BOS	Computer Science
Class	M.Sc. Data Analytics Part - I
Semester	II
Course Name	Information Security Auditing
Course Code	PMSDA204
Type of course	Major Elective
Level of the Subject	Advanced
Credit points	02Theory+01 Practical

Course Objectives:

1. Assess an organization based on the needs and suggest the requisite information security policies to be deployed.
2. Audit the organization across relevant policies and assist the organization in implementing such policies along with suggesting improvements.

Unit No.	Unit Name	Topic No.	Content	Hours
1	Basic of Auditing	1.1	Secrets of a Successful Auditor, Understanding the Demand for IS Audits, Understanding Policies, Standards, Guidelines, and Procedures, Understanding Professional Ethics ,Understanding the Purpose of an Audit , Differentiating between Auditor and Auditee, Roles Implementing Audit Standards, Auditor Is an Executive Position, Understanding the Corporate Organizational	10
		1.2	Structure Governance, Strategy Planning for Organizational Control , Overview of Tactical Management Planning and Performance, Overview of Business Process Reengineering Operations Management Summary , Audit Process Understanding the Audit Program Establishing and Approving an Audit Charter, Preplanning Specific Audits Performing an Audit Risk Assessment ,	
2	Working with Auditing	2.1	Determining Whether an Audit Is Possible Performing the Audit, Gathering Audit Evidence , Conducting Audit Evidence Testing, Generating Audit Findings , Report Findings Conducting Follow-up (Closing Meeting)	10
		2.2	Information Systems Acquisition and Development , Project Governance and Management, Business Case and Feasibility , Analysis System Development Methodologies , Control Identification and Design Testing Methodologies, Configuration and Release Management System Migration, Infrastructure Deployment and Data Conversion, Post-implementation Review,	

3	End-User Computing Data Governance & Disaster Recovery	3.1	Information Systems Operations, Introduction, Common Technology Components, IT Asset Management, Job Scheduling and Production , Process Automation System Interfaces, End-user Computing Data Governance Systems Performance Management Problem and Incident Management Change, Configuration, Release and IT Service	10
		3.2	Level Management Database Management, Business Resilience Business Impact Analysis Data Backup, Storage and Restoration Business Continuity Plan Disaster Recovery Plans Information Systems Life Cycle ,Governance in Software Development ,Management of Software Quality, Overview of the Executive Steering Committee Change Management,	
4	Software Project Management	4.1	Management of the Software Project ,Overview of the System Development Life Cycle Overview of Data Architecture, Decision Support Systems Program Architecture Centralization vs. Decentralization Electronic Commerce System Implementation and Operations ,Understanding the Nature of IT Services Performing IT Operations Management Performing Capacity Management Using Administrative Protection Performing Problem Management,	10
		4.2	Monitoring the Status of Controls Implementing Physical Protection Protecting Information Assets Understanding the Threat Using Technical Protection Business Continuity and Disaster Recovery , Debunking the Myths Understanding the Five Conflicting Disciplines Called Business Continuity Defining Disaster Recovery Defining the Purpose of Business Continuity Uniting Other Plans with Business Continuity Understanding the Five Phases of a Business Continuity Program Understanding the Auditor Interests in BC/DR Plans	
Total No. of Lectures				40

Course Outcomes:

After completion of the course, a student should be able to:

1. Understand various information security policies and process flow, Ethics of an Information security Auditor.
2. Understand various information systems in an organization, their criticality and various governance and management policies associated with them.
3. Critically analyze various operational strategies like asset management, data governance etc.
4. Suggest requisite changes as per organizations requirements with improvements.
5. Understand the information flow across the organization and identify the weak spots, and also suggest improvements to strengthen them.
6. Come up with strong strategies to protect information assets and come up with an efficient business continuity plan, disaster recovery strategy etc.

Books and References:

1. CISA®: Certified Information Systems Auditor David Cannon SYBEX Fourth Edition 2016
2. "Auditing For Dummies": A beginner-friendly book that covers essential concepts in auditing[2].
3. "Business Continuity and Disaster Recovery for InfoSec Managers" by John Rittinghouse and James F.

4. The Phoenix Project: A Novel about IT, DevOps, and Helping Your Business Win" by Gene Kim, Kevin Behr, and George Spafford
5. "Lean Project Management: Eight Principles For Success" by Lawrence P. Leach
6. "Agile Project Management with Kanban" by Eric Brechner:

CASE STUDY	

1	**Case Study: Information Security Management Practices in Indian IT Organizations**
	 Background
	 In recent years, information security has become a critical concern for organizations across diverse industries. The increasing reliance on information handling, coupled with emerging IT/ICT mediums, has posed new challenges. This case study delves into the information security management (ISM) practices of two IT development and services organizations in India.
	 Methodology
	 1. Qualitative Research Approach:
	    ○ The study adopts a qualitative research route to understand the current ISM practices of the case organizations.
	    ○ Semi-structured interviews with key stakeholders provide valuable insights.
	 2. SAP-LAP Method of Inquiry:
	    ○ The Situation, Actor, Process—Learning, Action, Performance (SAP-LAP) framework is used to analyze the findings from the case studies.
	 Key Findings
	 1. Top Management Support:
	    ○ Consistent top management support is crucial for effective ISM.
	    ○ Organizations with strong leadership commitment tend to implement better security practices.
	 2. Organizational Information Security Culture:
	    ○ Cultivating a security-conscious culture is essential.
	    ○ Employees' awareness and adherence to security policies contribute significantly.
	 3. Monitoring Systems for ISM:
	    ○ Implement robust monitoring systems to detect and respond to security incidents.
	    ○ Regular audits and vulnerability assessments enhance security posture.
	 4. Incident Response Preparedness:
	    ○ Organizations should have well-defined incident response plans.
	    ○ Timely and effective responses mitigate the impact of security breaches.

2	**Case Study: Information Security Management Practices in Indian IT Organizations** **Background:** In recent years, information security has become a critical concern for organizations across diverse industries. The increasing reliance on information handling, coupled with emerging IT/ICT mediums, has posed new challenges. This case study examines the information security management (ISM) practices of two IT development and services organizations in India. **Facts:** 1. Organizations: The study focuses on two Indian IT companies. 2. Qualitative Research: The case study adopts a qualitative research approach. 3. Observations: Insights are derived from semi-structured interviews. 4. Methodology: The SAP-LAP (Situation, Actor, Process—Learning, Action,

Performance) method of inquiry is used for analysis.

Key Findings:
1. Top Management Support: Consistent support from top management is crucial for effective ISM.
2. Organizational Culture: A strong information security culture within the organization is essential.
3. Monitoring System: Implementing a robust monitoring system enhances ISM effectiveness.

PRACTICALS

Practical No	Details
1.	Pre-Audit Checklist Creation: ○ Develop a comprehensive pre-audit checklist covering aspects such as information gathering, risk assessment, and compliance requirements.
2.	Understanding Professional Ethics: ○ Research and analyze professional ethics codes relevant to information security auditing (e.g., ISACA's Code of Professional Ethics). ○ Create a presentation or report highlighting key ethical principles.
3.	Audit Program Development: ○ Design an audit program for a specific organizational process (e.g., financial transactions, access controls). ○ Specify audit objectives, scope, procedures, and resources required.
4.	Risk Assessment Simulation: ○ Simulate a risk assessment for an IT system or project. ○ Identify potential risks, assess their impact, and prioritize them. ○ Propose risk mitigation strategies.
5.	Audit Evidence Gathering: ○ Select a real-world organization or system. ○ Collect audit evidence (e.g., logs, policies, interviews). ○ Document findings and observations.
6.	Business Continuity and Disaster Recovery Planning Review: ○ Evaluate an organization's business continuity and disaster recovery plans. ○ Identify gaps or areas for improvement. ○ Provide recommendations.
7.	Security Controls Assessment: ○ Choose a specific security control (e.g., access control, encryption). ○ Assess its effectiveness within an organization. ○ Propose enhancements or adjustments..
8.	Post-implementation Review: ○ Analyze the success of an implemented system or project. ○ Assess whether it meets its intended goals. ○ Identify any deviations from the original plan.

9.	Change Management Audit:
	○ Investigate an organization's change management processes.
	○ Evaluate change request handling, approvals, and documentation.
	○ Suggest improvements.
10.	Database Management Security Assessment:
	○ Focus on database security (e.g., user access, encryption, backups).
	○ Assess the security posture of a database system.
	○ Generate a report with findings and recommendations..

BOS	Computer Science
Class	M.Sc. Data Analytics Part - I
Semester	II
Course Name	Network Security & Cryptography
Course Code	PMSDA204
Type of course	Major Elective
Level of the Subject	Advanced
Credit points	02Theory+01 Practical

Course Objectives:
1. Discuss the fundamental ideas of cryptographic algorithms, public-key cryptography, enhancements made to IPv4 by IPSec
2. Generate and distribute a PGP key pair and use the PGP package to send an encrypted email message.

Unit No.	Unit Name	Topic No.	Content	No. of Lect ures
1	Basic of Security.	1.1	**Security Concepts:** Introduction, The need for security, Principles of security, Types of Security attacks, A model for Network Security	10
		1.2	**Cryptography Concepts and Techniques:** Introduction, plain text and cipher text, substitution techniques, transposition techniques, encryption and decryption, symmetric and asymmetric key cryptography, steganography, possible types of attacks	
2	Symm etric & Asym metric Crypto graphy	2.1	**Symmetric key Ciphers:** Block Cipher principles, DES, AES, Blowfish, IDEA, Block cipher operation, Stream ciphers, RC4. **Asymmetric key Ciphers:** Principles of public key cryptosystems, RSA algorithm, Elgamal Cryptography, Diffie-Hellman Key Exchange, Knapsack Algorithm.	10
		2.2	**Cryptographic Hash Functions:** Message Authentication, Secure Hash Algorithm (SHA-512), Message authentication codes: Authentication requirements, HMAC, CMAC, Digital signatures	
3	Manag ement & Distrib ution of key ,	3.1	**Key Management and Distribution:** Symmetric Key Distribution Using Symmetric & Asymmetric Encryption, Distribution of Public Keys, Kerberos, X.509 Authentication Service, Public – Key Infrastructure	10

	Transport level security	3.2	**Transport-level Security:** Web security considerations, Secure Socket Layer and Transport Layer Security, HTTPS, Secure Shell (SSH)	
4	AI in Network Security and Cryptography	4.1	**AI in Network Security :** Applications of AI in NS , Challenge to work with AI	10
		4.2	**AI in Cryptography**: Applications , Challenges, Neural Cryptography , Neural key exchange protocol and its working. Case study on Key generation process with AI, Secure Inter-branch Payment Transactions	
Total No. of Lectures				40

Course Outcomes:
1. Understand basic cryptographic algorithms, message and web authentication and security issues.
2. Understand the basic categories of threats to computers and networks
3. Differentiate between Symmetric & Asymmetric cryptography algorithms.
4. Ability to identify information system requirements for both of them such as client and server.
5. Ability to understand the current legal issues towards information security.
6. Apply various security like email security, transport security etc.

Books and References:
1. Cryptography and Network Security - Principles and Practice: William Stallings, Pearson Education, 6th Edition
2. Cryptography and Network Security: Atul Kahate, Mc Graw Hill, 3rd Edition
3. Cryptography and Network Security: C K Shyamala, N Harini, Dr T R Padmanabhan, Wiley India, 1st Edition.
4. Cryptography and Network Security: Forouzan Mukhopadhyay, Mc Graw Hill, 3rd Edition
5. Information Security, Principles, and Practice: Mark Stamp, Wiley India.
6. Principles of Computer Security: WM. Arthur Conklin, Greg White, TMH

CASE STUDY	
1	Question : Consider Alice and Bob want to communicate securely. So they decide to generate the key with Diffie Hellman Key Exchange Algorithm by considering a large prime number p=17 and g = 3 Alice selects his private key as 7 and Bob selects 11.
2	Case Study: AI-Enhanced Key Generation for Secure Inter-branch Payment Transactions Background: In a rapidly digitizing world, secure payment transactions are paramount for financial institutions. Inter-branch transactions involve transferring funds between different branches of a bank or financial organization. Ensuring the confidentiality, integrity, and authenticity of these transactions is critical. Our case study focuses on an innovative approach to key generation using artificial intelligence (AI) and its impact on secure inter-branch payments. Scenario: ● Bank X operates multiple branches across a region. ● Inter-branch transactions occur daily, involving large sums of money. ● Traditional cryptographic key generation methods are time-consuming and resource-intensive.

Challenges:
1. Efficiency: Generating secure keys quickly without compromising security.
2. Robustness: Ensuring keys are resistant to attacks (both classical and quantum).
3. Scalability: Handling a growing number of transactions across branches.

Proposed Solution: AI-Enhanced Key Generation
● Neural Cryptography: Leveraging neural networks for key generation.
● Generative Adversarial Networks (GANs): Training GANs to create secure key pairs.
● Hybrid Approach: Combining AI-generated keys with traditional cryptographic methods.

Implementation Steps:
1. Data Collection:
    ○ Gather historical transaction data from inter-branch transfers.
    ○ Extract patterns, correlations, and anomalies.
2. Neural Network Training:
    ○ Train a neural network to learn from the data.
    ○ Optimize for key generation based on transaction characteristics.
3. GAN Training:
    ○ Create a GAN architecture with a generator and discriminator.
    ○ Train the generator to produce secure key pairs.
    ○ Evaluate against the discriminator's feedback.
4. Hybrid Key Generation:
    ○ Combine AI-generated keys with asymmetric encryption techniques.
    ○ Use traditional public-key infrastructure (PKI) for initial key exchange.

Benefits:
● Speed: AI-enhanced key generation significantly reduces the time required.
● Security: Neural networks adapt to emerging threats.
● Scalability: The system can handle a large volume of transactions.

Real-world Application: Secure Inter-branch Payments
● When two branches need to exchange funds, the AI system generates unique keys.
● These keys are used for secure communication during the transaction.
● The hybrid approach ensures robustness against attacks.

PRACTICALS

Practical No	Details
1.	Write programs to implement the following Substitution Cipher Techniques: - * Caesar Cipher - Encryption and Decryption
2.	Write programs to implement the following Substitution Cipher Techniques: * Playfair Cipher * Vernam Cipher
3.	Write programs to implement the following Transposition Cipher Techniques: - Rail Fence Cipher - Simple Columnar Technique
4.	Write program to encrypt and decrypt strings using - DES Algorithm - AES Algorithm
5.	Write program to encrypt and decrypt strings using

		- AES Algorithm
6.		Write a program to implement an IDEA algorithm to perform encryption / decryption of a given string.
7.		Write a program to implement the Diffie-Hellman Key Agreement algorithm to generate symmetric keys.
8.		Write a program to implement the MD5 algorithm to compute the message digest.
9.		Write a program to calculate HMAC-SHA1 Signature
10.		Write a Python program to illustrate ElGamal encryption

BOS	Computer Science
Class	M.Sc. Data Analytics Part - I
Semester	II
Course Name	On Job Training: Internship/ Apprenticeship
Course Code	PMSDA207
Type of course	OJT/FP
Level of the Subject	Advanced
Credit points	4

The syllabus proposes an internship for about 7 weeks to 8 weeks (min. 90Hrs) to be done by a student. It is expected that a student chooses a Data Analytics or IT-related industry and formally works as a full time/ part time intern during the period. The student should subject oneself with an internship evaluation with proper documentation of the attendance and the type of work he or she has done in the chosen organization. Proper certification (as per the guidelines given in Appendix 1 and 2) by the person, to whom the student was reporting, with the Organization's seal should be attached as part of the documentation.