Semester- III

Data Mining and Visualization			
Course Code	**MMCA311A**	CIE Marks	50
Teaching Hours/Week (L:P:SDA)	3:0:0	SEE Marks	50
Total Hours of Pedagogy	40	Total Marks	100
Credits	03	Exam Hours	3

Course Learning objectives:

- Understand foundational concepts of data mining, including data preprocessing, pattern discovery, and classification techniques.
- Apply data mining algorithms to extract useful patterns, trends, and insights from large datasets.
- Analyse and interpret mined data using appropriate visualization techniques and tools.
- Develop skills to evaluate the performance of various data mining models and choose suitable techniques based on problem context.
- Use data visualization tools and libraries to present complex data and mining results in an intuitive and meaningful way.

Module-1

Foundations of Data Mining and Data Preprocessing

Introduction to Data Mining & Preprocessing Techniques: Introduction to data mining: Motivation, architecture, KDD process. Types of data: Structured, semi-structured, unstructured.
Data preprocessing: Cleaning, integration, reduction, transformation, Missing Values and Noisy Data. Data summarization and visualization techniques for preprocessing analysis. Implementation using Python: Pandas, NumPy for basic preprocessing.

Module-2

Data Mining Techniques and Algorithms

Mining Techniques: Classification, Clustering & Association: Classification: Decision Trees, k-NN, Naive Bayes – concepts and implementation. Clustering: k-Means, Hierarchical clustering.
Association rule mining: Market basket analysis, Apriori algorithm, FP-Growth. Evaluation methods: Confusion matrix, precision, recall, ROC.

Module-3

Data Visualization Techniques

Static and Interactive Data Visualization with Python: Principles of effective data visualization. Visualization tools and libraries: Matplotlib, Seaborn, Plotly, Bokeh. Histograms, bar charts, scatter plots, heatmaps, and pair plots.
Dashboard creation using Jupyter notebooks and interactive widgets. Case studies and real-world examples using multi-dimensional data.

Module-4

Visualizing Streaming and Real-Time Data

Real-Time Analytics and Streaming Data Visualization Overview of streaming data: Sources, characteristics, and tools. Real-time processing with Apache Kafka, PySpark Streaming (introductory overview).
Visualization strategies for streaming data. Tools: Dash by Plotly, Streamlit, Grafana.
Case studies: Sensor data, web server logs.

Module-5

Advanced Data Mining Applications and Trends

Emerging Trends and Applications in Data Mining social media and text data. Sentiment analysis and NLP basics using Python. Time series analysis and visualization. Anomaly detection and predictive analytics. Ethical issues and future trends in data mining.

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 50% of the maximum marks. Minimum passing marks in SEE is 40% of the maximum marks of SEE. A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures not less than 50% (50 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

1. Two Unit Tests each of **25 Marks**
2. Two assignments each of **25 Marks** or **one Skill Development Activity of 50 marks** to attain the COs and POs

The sum of two tests, two assignments/skill Development Activities, will be **scaled down to 50 marks**

CIE methods /question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

1. The SEE question paper will be set for 100 marks and the marks scored will be proportionately reduced to 50.
2. The question paper will have ten full questions carrying equal marks.
3. Each full question is for 20 marks. There will be two full questions (with a maximum of four sub-questions) from each module.
4. Each full question will have a sub-question covering all the topics under a module.
5. The students will have to answer five full questions, selecting one full question from each module

Suggested Learning Resources:

Books

1. **Jiawei Han, Micheline Kamber, Jian Pei**, *Data Mining Concepts and Techniques*, Morgan Kaufmann.
2. **Wes McKinney**, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Python*, O'Reilly Media.
3. **Tim Grobmann, Mario Dobler**, *Data Visualization with Python*, O'Reilly Media.

Web links and Video Lectures (e-Resources):

- **NPTEL – Data Mining by IIT Kharagpur** (Prof. Pabitra Mitra)
  https://nptel.ac.in/courses/106105174
- **Data Mining Full Course by Great Learning** (YouTube)
  https://www.youtube.com/watch?v=RlD5q_pIWkM
- **Data Visualization using Python (Edureka)**
  https://www.youtube.com/watch?v=UB3DE5Bgfx4
- **Harvard Data Science: Visualization (edX)**
  https://cs50.harvard.edu/
- **Tableau for Data Visualization (Simplilearn)**
  https://www.youtube.com/watch?v=IFM03Nis2dg

Skill Development Activities Suggested

- **Hands-on Data Mining Projects**
- Work on **real-world datasets** (e.g., Kaggle, UCI ML Repository).
- Implement **data preprocessing, cleaning, and feature engineering**.
- Apply **classification, clustering, and association rule mining** techniques.
- **Learning and Using Data Visualization Tools**
- Get hands-on with **Tableau, Power BI, and Matplotlib/Seaborn in Python**.
- Use **SQL for data extraction and processing**.

Course outcome (Course Skill Set)

At the end of the course the student will be able to:

Sl. No.	Description	Blooms Level
CO1	Understand foundational concepts of data mining and apply preprocessing techniques using Python.	L2
CO2	Implement key data mining techniques such as classification, clustering, and association rule mining.	L3
CO3	Design and develop effective static and interactive data visualizations using Python libraries.	L3
CO4	Apply real-time visualization strategies for streaming data using tools like Dash, Streamlit, and Grafana.	L3
CO5	Analyse advanced data mining applications including sentiment analysis, time series, and anomaly detection.	L2

Mapping of COS and POs

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	2	1					
CO2	3	3	2	2				
CO3	2	2	3	2	1			
CO4	2	3	3	3	2			
CO5	3	2	3	3	2	1		

Semester- III

Big Data Analytics			
Course Code	**MMCA311B**	CIE Marks	50
Teaching Hours/Week (L:P:SDA)	3:0:0	SEE Marks	50
Total Hours of Pedagogy	40	Total Marks	100
Credits	03	Exam Hours	3

Course Learning objectives:

- Understand Big Data Concepts – Gain a comprehensive understanding of Big Data, its characteristics, and its significance in modern computing.
- Explore Big Data Technologies – Learn about various Big Data tools and frameworks such as Hadoop, Spark, and NoSQL databases.
- Perform Data Processing & Analysis – Develop skills in processing, storing, and analysing large-scale data using distributed computing techniques.

Module-1

Big Data Fundamentals and Ecosystem Overview

Introduction to Big Data: Concepts and Ecosystem: Definition and Evolution of Big Data. Characteristics of Big Data (Volume, Velocity, Variety, Veracity, Value). Traditional vs Big Data Systems,
Introduction to Hadoop Ecosystem: HDFS, YARN, MapReduce. Architecture and components of Hadoop. Limitations of Hadoop and the shift to Spark

Module-2

Hadoop Architecture and MapReduce Programming

Distributed Data Processing using Hadoop: Hadoop Distributed File System (HDFS): Design and operations. Hadoop MapReduce: Programming model, job execution flow. Writing MapReduce programs (Word Count, Sorting, Joins). Advanced Hadoop: Combiners, Partitioners, Counters. Hadoop Streaming and integration with Python.

Module-3

Apache Spark for Big Data Analytics

In-Memory Big Data Processing with Spark: Spark architecture and components: RDDs, DAG, Executors. Transformations and Actions on RDDs. Introduction to Data Frames and Spark SQL.
Introduction to Spark MLlib for machine learning. PySpark: Setting up and running Spark jobs using Python.

Module-4

NoSQL and Big Data Storage Systems

Scalable Data Storage with NoSQL Databases: Need for NoSQL: Limitations of RDBMS in Big Data. Types of NoSQL Databases: Key-Value, Document, Column, Graph. Introduction to HBase: Architecture and CRUD operations. Working with Cassandra and MongoDB. Data modelling for scalability and performance.

Module-5

Big Data Tools and Industry Applications

Real-World Big Data Applications and Tools: Overview of Big Data Tools: Hive, Pig, Sqoop, Flume, Oozie. Data ingestion with Flume and Sqoop. Use cases in healthcare, finance, e-commerce, IoT, social media. Real-time analytics introduction using Kafka and Spark Streaming. Ethics and challenges in Big Data (privacy, bias, governance).

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 50% of the maximum marks. Minimum passing marks in SEE is 40% of the maximum marks of SEE. A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures not less than 50% (50 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

1. Two Unit Tests each of **25 Marks**
2. Two assignments each of **25 Marks**or **oneSkill Development Activity of 50 marks** to attain the COs and POs

The sum of two tests, two assignments/skill Development Activities, will be **scaled down to 50 marks**

CIE methods /question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

1. The SEE question paper will be set for 100 marks and the marks scored will be proportionately reduced to 50.
2. The question paper will have ten full questions carrying equal marks.
3. Each full question is for 20 marks. There will be two full questions (with a maximum of four sub-questions) from each module.
4. Each full question will have a sub-question covering all the topics under a module.
5. The students will have to answer five full questions, selecting one full question from each module

Suggested Learning Resources:

Books

1. . **Tom White** – *Hadoop: The Definitive Guide*, O'Reilly Media.
2. **Vignesh Prajapati** – *Big Data Analytics with R and Hadoop*, Packt Publishing.
3. **Jure Leskovec, Anand Rajaraman, Jeff Ullman** – *Mining of Massive Datasets*, Cambridge University Press.
4. **Venkat Ankam** – *Big Data Analytics with Spark*, Packt Publishing.

Weblinks and Video Lectures (e-Resources):

- **NPTEL Big Data Analytics Course** – https://nptel.ac.in/courses/106/104/106104189/
- **Simplilearn Big Data Tutorial (YouTube)** – https://www.youtube.com/watch?v=-FrXAKGthF8
- Detailed explanation of Big Data concepts and tools.
- **Big Data Analytics Using Python (YouTube - Great Learning)** – https://www.youtube.com/watch?v=ZkZclIFmgVY

Skill Development Activities Suggested

- **Hands-on with Hadoop and Spark** – Work on real-time data processing using Hadoop (HDFS, MapReduce) and Apache Spark. Set up a small Hadoop cluster and practice writing Spark applications.
- **Data Processing and SQL** – Master SQL-based tools like Hive, Impala, and Presto. Work with large datasets to optimize queries and improve performance.
- **Machine Learning with Big Data** – Implement machine learning algorithms using libraries like MLlib (Spark) and TensorFlow with large datasets.

Course outcome (Course Skill Set)

At the end of the course the student will be able to :

Sl. No.	Description	Blooms Level
CO1	Understand the fundamental concepts, evolution, and architecture of Big Data, including the Hadoop ecosystem.	L1
CO2	Develop and execute distributed data processing tasks using HDFS and MapReduce programming techniques.	L2
CO3	Analyze and implement in-memory data processing using Apache Spark and perform machine learning tasks with Spark MLlib.	L3
CO4	Compare and evaluate NoSQL data models (Key-Value, Document, Column, Graph) and perform operations on HBase, MongoDB, and Cassandra.	L4
CO5	Apply big data tools (Hive, Pig, Sqoop, Flume, Kafka) in real-world domains and understand ethical issues related to Big Data analytics.	L5

Mapping of COS and POs

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	1	2						
CO2	1	2	3					
CO3		2	3		3			3
CO4		2						3
CO5		2	3	3			3	

Semester- III

Business Data Analytics			
Course Code	**MMCA311C**	CIE Marks	50
Teaching Hours/Week (L:P:SDA)	3:0:0	SEE Marks	50
Total Hours of Pedagogy	40	Total Marks	100
Credits	03	Exam Hours	3

Course Learning objectives:

- Understand the Fundamentals of Business Analytics – Learn the core concepts, tools, and techniques used in data-driven decision-making for businesses.
- Data Collection and Processing – Gain knowledge of data acquisition, cleaning, transformation, and management techniques for business applications.
- Exploratory Data Analysis & Visualization – Develop skills to analyze and visualize business data using statistical and graphical methods.
- Predictive Analytics & Machine Learning – Apply statistical models and machine learning techniques to predict business trends and customer behavior.

Module-1

Foundations of Business Data Analytics

Introduction to Business Analytics and Data-Driven Decision Making: Introduction to Business Analytics: Scope, Types (Descriptive, Predictive, Prescriptive). Data in Business: Structured vs Unstructured, Sources of Data. Business Intelligence vs Business Analytics.

Analytics Life Cycle: CRISP-DM methodology. Role of Business Analyst: Tools, Skills, and Case Examples

Module-2

Data Preprocessing and Exploratory Analysis

Data Wrangling, Cleaning, and Exploration for Business Insights:

Data Preparation: Cleaning, Integration, Transformation. Handling Missing Data, Outliers, and Noise. Descriptive Statistics: Mean, Median, Mode, Variance, Skewness. Correlation and Covariance.

Visualization for Exploration: Histograms, Boxplots, Heatmaps. Business Case: Customer Segmentation and Sales Data Analysis.

Module-3

Predictive Analytics in Business

Forecasting and Predictive Modelling for Business Decision Making:

Introduction to Regression: Simple & Multiple Linear Regression. Logistic Regression: Applications in classification. Time Series Analysis and Forecasting Techniques. Model Evaluation: RMSE, MAE, Accuracy, Precision, Recall. Business Applications: Sales Forecasting, Customer Churn Prediction.

Module-4

Prescriptive Analytics and Optimization

Optimization and Decision-Making Techniques:

Introduction to Prescriptive Analytics. Linear Programming and Solver in Excel. Optimization Models: Objective functions, Constraints. Sensitivity and Scenario Analysis. Decision Trees and Business Rules. Case Study: Resource Allocation, Pricing Models, Supply Chain Optimization.

Module-5

Data Visualization and Business Intelligence Tools

Storytelling and Visualization for Business Insights

Principles of Data Visualization and Dashboards. BI Tools: Introduction to Power BI, Tableau. Designing Interactive Dashboards. KPI Definition and Visualization. Business Reporting and Data-Driven Story telling. Final Capstone: Complete Business Analytics Solution

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 50% of the maximum marks. Minimum passing marks in SEE is 40% of the maximum marks of SEE. A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures not less than 50% (50 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

1. Two Unit Tests each of **25 Marks**
2. Two assignments each of **25 Marks** or **one Skill Development Activity of 50 marks** to attain the COs and POs

The sum of two tests, two assignments/skill Development Activities, will be **scaled down to 50 marks**

CIE methods /question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

1. The SEE question paper will be set for 100 marks and the marks scored will be proportionately reduced to 50.
2. The question paper will have ten full questions carrying equal marks.
3. Each full question is for 20 marks. There will be two full questions (with a maximum of four sub-questions) from each module.
4. Each full question will have a sub-question covering all the topics under a module.
5. The students will have to answer five full questions, selecting one full question from each module

Suggested Learning Resources:

Books

1. **Abdulhamit Subasi**, *Practical Machine Learning for Data Analysis Using Python*, Academic Press.
2. **U. Dinesh Kumar**, *Business Analytics: The Science of Data-Driven Decision Making*, Wiley.
3. **Wes McKinney**, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*, O'Reilly.
4. **Wayne Winston**, *Microsoft Excel Data Analysis and Business Modeling*, Microsoft Press.

Weblinks and Video Lectures (e-Resources):

- **Introduction to Business Analytics** – NPTEL Course
- **Harvard Data Science and Business Analytics Lectures** – YouTube Playlist
- **Coursera: Business Analytics by Wharton** – Coursera

Skill Development Activities Suggested

- **Hands-on Experience with Data Tools**
- Practice using **Excel, Power BI, and Tableau** for data visualization.
- Work with **SQL and NoSQL databases** (MySQL, MongoDB).
- Learn **Python and R** for data analysis.
- **Real-world Data Projects**
- Analyse publicly available datasets (Kaggle, UCI Machine Learning Repository).
- Work on **case studies in business analytics** (sales forecasting, customer segmentation).
- Implement **predictive analytics models** using **machine learning**.

Course outcome (Course Skill Set)

At the end of the course the student will be able to :

Sl. No.	Description	Blooms Level
CO1	Understand the foundational concepts of business analytics, types of analytics, and the data analytics lifecycle.	L1
CO2	Apply data preprocessing techniques and perform exploratory data analysis to extract meaningful business insights.	L2
CO3	Develop and evaluate predictive models using regression and time series forecasting for business decision making.	L3
CO4	Implement prescriptive analytics using optimization techniques to support data-driven decisions in a business context.	L4
CO4	Design interactive dashboards using BI tools and present data-driven stories for effective communication of business insights.	L5

Mapping of COS and POs

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	1	2						
CO2		2		2	2			
CO3		2	3					3
CO4			3	3				3
CO5								

Semester- III

Enterprise Resource Planning			
Course Code	**MMCA311D**	CIE Marks	50
Teaching Hours/Week (L:P:SDA)	3:0:0	SEE Marks	50
Total Hours of Pedagogy	40	Total Marks	100
Credits	03	Exam Hours	3

Course Learning objectives:

- Understand ERP Concepts – Explain the fundamentals of ERP, its evolution, and its significance in modern businesses.
- Analyze ERP Modules – Explore core ERP modules like Finance, HR, Supply Chain, and Customer Relationship Management (CRM).
- ERP Implementation Strategies – Understand the phases of ERP implementation, challenges, and best practices.
- ERP Technologies & Trends – Examine emerging trends in ERP, such as cloud-based solutions, AI integration, and analytics.
- Real-world Applications – Analyse case studies of ERP implementations in various industries to understand its impact on business efficiency.

Module-1

ERP Systems and Business Process Integration

Fundamentals of ERP and Business Process Mapping Evolution of ERP – MRP, MRP II to ERP. Business Functions and Business Processes. Need for Integration and ERP as an Integrator, Benefits, Risks, and Misconceptions of ERP. Overview of Functional Modules: Finance, HR, Production, Sales. Case Example: Business Process before and after ERP

Module-2

ERP Architecture and Technologies

ERP System Architecture and Technological Infrastructure Client/Server Architecture, Service-Oriented Architecture (SOA), Cloud-based ERP vs On-Premise ERP, ERP Platforms and Databases, ERP and Web Integration, Security, Customization, and Interoperability in ERP Systems.
Overview of leading ERP systems: SAP, Oracle, Microsoft Dynamics, Odoo

Module-3

ERP Modules and Functional Features

Core ERP Modules and Organizational Applications Finance and Accounting Module, Manufacturing and Production Planning, Sales and Distribution, Human Resource Management (HRM).
Supply Chain Management (SCM), CRM and Business Intelligence Features, Industry Examples: ERP use in Retail, Healthcare, Logistics.

Module-4

ERP Implementation and Project Management

ERP Life Cycle and Implementation Strategies Phases of ERP Implementation Life Cycle, Business Process Reengineering (BPR) and Change Management, Implementation Methodologies (ASAP, AIM), ERP Project Planning, Testing, Training, Go-Live & Support, Cost-Benefit Analysis, Vendor Selection, Risk Management, Failure Cases and Lessons Learned.

Module-5

Emerging Trends in ERP and Industry Practices

Future Directions and ERP in the Digital Era ERP with AI, ML, IoT, and Blockchain, ERP and Digital Transformation, Mobile ERP and UX Design, ERP in SMEs and Cloud ERPs.
ERP Data Analytics and Reporting, Future Trends: Low-Code ERP, Industry 4.0 Integration, Capstone: Evaluation of ERP for a case enterprise.

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 50% of the maximum marks. Minimum passing marks in SEE is 40% of the maximum marks of SEE. A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures not less than 50% (50 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

1. Two Unit Tests each of **25 Marks**
2. Two assignments each of **25 Marks** or **one Skill Development Activity of 50 marks** to attain the COs and POs

The sum of two tests, two assignments/skill Development Activities, will be **scaled down to 50 marks**

CIE methods /question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

1. The SEE question paper will be set for 100 marks and the marks scored will be proportionately reduced to 50.
2. The question paper will have ten full questions carrying equal marks.
3. Each full question is for 20 marks. There will be two full questions (with a maximum of four sub-questions) from each module.
4. Each full question will have a sub-question covering all the topics under a module.
5. The students will have to answer five full questions, selecting one full question from each module

Suggested Learning Resources:

Books

1. **Alexis Leon**, *Enterprise Resource Planning*, McGraw Hill Education.
2. **Mary Sumner**, *Enterprise Resource Planning*, Pearson Education.
3. **Mahadeo Jaiswal & Ganesh Vanapalli**, *Enterprise Resource Planning*, Macmillan India.
4. **Ellen Monk, Bret Wagner**, *Concepts in Enterprise Resource Planning*, Cengage Learning.

Weblinks and Video Lectures (e-Resources):

- https://www.youtube.com/watch?v=qgHlU_ll6mk
- https://www.youtube.com/watch?v=pSttK5Op1rI&utm_source=chatgpt.com
- https://www.youtube.com/watch?v=JnSrp4k1gJw&utm_source=chatgpt.com
- https://www.youtube.com/watch?v=ppfBvofxCM0&utm_source=chatgpt.com
- https://www.youtube.com/watch?v=cblNqNEThcE&utm_source=chatgpt.com

Skill Development Activities Suggested

- Map business processes of a small business (e.g., order-to-cash or procure-to-pay) using a flowchart or BPMN tool (e.g., Draw.io or Lucidchart).
- Group discussion/debate on ERP benefits vs. risks in real-world businesses.
- Create a basic ERP system architecture diagram using PowerPoint or any diagramming tool.
- Create a module-wise feature matrix comparing SAP, Oracle, and Odoo ERP.

Course outcome (Course Skill Set)

At the end of the course the student will be able to:

Sl. No.	Description	Blooms Level
CO1	Understand the fundamentals of ERP systems, their evolution, business processes, and the need for integration.	L1
CO2	Analyze ERP architectures, technologies, and distinguish between various deployment models and ERP solutions.	L2
CO3	Examine ERP functional modules and their applications across different business domains.	L3
CO4	Apply ERP implementation strategies, project management techniques, and evaluate risk and success factors.	L4
CO5	Explore and assess emerging trends in ERP such as AI, IoT, Blockchain, and digital transformation practices.	L5

Mapping of COS and POs

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	1	2						
CO2	1	2			3			
CO3		2	3		3			
CO4			3			3		
CO5					3		3	3

Semester- III

Exploratory Data Analysis			
Course Code	**MMCA311E**	CIE Marks	50
Teaching Hours/Week (L:P:SDA)	3:0:0	SEE Marks	50
Total Hours of Pedagogy	40	Total Marks	100
Credits	03	Exam Hours	03

Course Learning objectives:

- To introduce the fundamental concepts and principles of exploratory data analysis.
- To equip students with skills to summarize and visualize both univariate and multivariate data effectively.
- To develop the ability to clean, pre-process, and transform raw data for analysis.
- To expose learners to current trends and tools used in the field of EDA.

Module-1

Introduction to Exploratory Data Analysis: Historical background and role of EDA in data science, Philosophy and goals of EDA, Comparison with classical statistical methods, Types of data and scales of measurement, Importance of visual summaries before formal modelling.

Module-2

Univariate Data Exploration: Distribution shape: symmetry, skewness, kurtosis Summary statistics: mean, median, mode, variance, standard deviation, range, IQR Graphical techniques: histograms, dot plots, stem-and-leaf plots, boxplots, Identifying outliers and anomalies.

Module-3

Bivariate and Multivariate Data Exploration: Scatter plots, trend analysis, Correlation vs. causation,Crosstabs and pivot tables, Pair plots and heatmaps, Data smoothing (moving averages, LOESS).

Module-4

Data Transformation and Cleaning: Motivation for data transformation (e.g., to achieve normality or reduce skew) Log, square root, and other transformations, Handling missing values and duplicates, Introduction to resistant statistics (median, trimmed mean).

Module-5

Emerging Trends and case studies: EDA as a storytelling tool, AI-powered visualizations, integration with big data platforms.Case studies: EDA on real-world datasets (Titanic, Iris, planet), pitfalls in EDA.

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 50% of the maximum marks. Minimum passing marks in SEE is 40% of the maximum marks of SEE. A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures not less than 50% (50 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

1. Two Unit Tests each of **25 Marks**
2. Two assignments each of **25 Marks**or **oneSkill Development Activity of 50 marks**
   to attain the COs and POs

The sum of two tests, two assignments/skill Development Activities, will be **scaled down to 50 marks**

CIE methods /question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

1. The SEE question paper will be set for 100 marks and the marks scored will be proportionately reduced to 50.
2. The question paper will have ten full questions carrying equal marks.
3. Each full question is for 20 marks. There will be two full questions (with a maximum of four sub-questions) from each module.
4. Each full question will have a sub-question covering all the topics under a module.
5. The students will have to answer five full questions, selecting one full question from each module

Suggested Learning Resources:

Books

.1. *Exploratory Data Analysis* by John W. Tukey.

2. *An Introduction to Statistical Learning* by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani

3. *Think Stats: Exploratory Data Analysis in Python* by Allen B. Downey

Reference Books:

1. *Practical Statistics for Data Scientists* by Peter Bruce, Andrew Bruce, and Peter Gedeck

Web links and Video Lectures (e-Resources):

- https://youtu.be/fHFOANOHwh8?si=MFGfiOEvPQSF-g2H
- https://youtu.be/w2QVZHcJapU?si=xfacUu80VK8J4fzc
- https://youtu.be/clblk_NwEU8?si=e4O8qLB6TnuaejdQ

Skill Development Activities Suggested

- Hands-on labs using Python for real-world datasets (Titanic, Iris, Sales, etc.)
- Participation in Kaggle or similar online EDA competitions
- Group project: Collaborative analysis and presentation of EDA findings

Course outcome (Course Skill Set)

At the end of the course the student will be able to :

Sl. No.	Description	Blooms Level
CO1	Understand the role and importance of Exploratory Data Analysis in the data science pipeline.	L2
CO2	Analyse univariate and bivariate datasets using appropriate summary statistics and visualization techniques.	L2
CO3	Apply data analysis techniques to explore relationships between multiple variables and derive insights using Python.	L2
CO4	Apply data transformation and cleaning methods to prepare raw data for further analysis.	L2
CO5	Interpret insights from real-world datasets and communicate findings through visual storytelling and reporting.	L2

Mapping of COS and POs

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	1							
CO2		2						
CO3	1	2			3			
CO4	1			3				3
CO5	1	2						

Semester- III

Social Media Analytics			
Course Code	**MMCA311F**	CIE Marks	50
Teaching Hours/Week (L:P:SDA)	3:0:0	SEE Marks	50
Total Hours of Pedagogy	40	Total Marks	100
Credits	03	Exam Hours	3

Course Learning objectives:

- Understand the fundamentals and evolution of social media platforms.
- Explore key concepts and techniques in social media data collection and analysis.
- Apply analytics tools to extract insights from social media data.
- Develop skills in sentiment analysis, trend prediction, and influence measurement.
- Design data-driven strategies for business and marketing using social media insights.

Module-1

Introduction to Social Media Analytics

Foundations of Social Media Data Analysis Introduction to Social Media Analytics: Definition, Applications, and Importance, Overview of Popular Social Media Platforms: Facebook, Twitter, Instagram, LinkedIn, YouTube

Social Media Data Types: Structured vs. Unstructured Dat, Social Media Metrics and KPIs: Engagement, Reach, Impressions, Sentiment Score, Data Collection Techniques: Web Scraping, APIs (Twitter, Facebook, YouTube), Streaming Data

Module-2

Sentiment Analysis and Text Mining

Natural Language Processing for Social Media Data Fundamentals of Sentiment Analysis: Positive, Negative, Neutral Sentiments**,** Text Preprocessing: Tokenization, Stopword Removal, Stemming, Lemmatization

Machine Learning Approaches for Sentiment Classification : Naïve Bayes, SVM, LSTM, Word Embeddings and Sentiment Scoring: TF-IDF, Word2Vec, BERT.

Module-3

Social Network Analysis and Trend Detection

Graph-based Social Media Analytics Basics of Social Network Analysis (SNA) Key Metrics: Centrality, Clustering Coefficients, Community Detection, Influencer Identification and User Engagement Analytic

Hashtag Analysis and Topic Modelling using LDA, Trend Detection on Social Media: Time Series Analysis, Virality Prediction.

Module-4

Visualizing and Interpreting Social Media Insights

Data Visualization and Interpretation for Social Media Analytics Importance of Data Visualization in Social Media Analytic.

Visualization Techniques: Word Clouds, Heatmaps, Network Graphs Sentiment Heatmaps and Hashtag Trends Visualization, Dashboard Creation using Tableau and Power BI.

Module-5

Applications of Social Media Analytics in Business and Research

Business and Industry Applications of Social Media Analytics

Social Media Analytics in Digital Marketing: Ad Performance and Customer Engagement, Social Media in Business Intelligence: Brand Monitoring and Crisis Management, Ethical Considerations in Social Media Analytics: Privacy, Bias, and Data Protection.

Future Trends in Social Media Analytics: AI-Driven Social Insights, Capstone Project: Analyzing Real-World Social Media Data for Business Insights.

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 50% of the maximum marks. Minimum passing marks in SEE is 40% of the maximum marks of SEE. A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures not less than 50% (50 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

1. Two Unit Tests each of **25 Marks**
2. Two assignments each of **25 Marks** or **one Skill Development Activity of 50 marks** to attain the COs and POs

The sum of two tests, two assignments/skill Development Activities, will be **scaled down to 50 marks**

CIE methods /question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

1. The SEE question paper will be set for 100 marks and the marks scored will be proportionately reduced to 50.
2. The question paper will have ten full questions carrying equal marks.
3. Each full question is for 20 marks. There will be two full questions (with a maximum of four sub-questions) from each module.
4. Each full question will have a sub-question covering all the topics under a module.
5. The students will have to answer five full questions, selecting one full question from each module

Suggested Learning Resources:

Books

1. Matthew A. Russell, Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More, O'Reilly Media.
2. Wasim Ahmed, Social Media Analytics: Techniques and Insights for Extracting Business Value Out of Social Media, Wiley
3. Reza Zafarani, Mohammad Ali Abbasi, Huan Liu, Social Media Mining: An Introduction, Cambridge University Press.
4. Piyushimita Thakuriah, Nebiyou Tilahun, Moira Zellner, Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics, Springer.

Web links and Video Lectures (e-Resources):

- NPTEL Course on Social Media Analytics – nptel.ac.in
- IBM Social Media Analytics Tutorials – ibm.com
- YouTube Channel: Analytics Vidhya – youtube.com/analyticsvidhya

Skill Development Activities Suggested

- Hands-on training with social media analytics tools (e.g., Hootsuite, Google Analytics).
- Data collection from various social media platforms.
- Analyzing engagement metrics and sentiment.
- Creating visual dashboards and reports.
- Case studies on brand performance and campaign impact.

Course outcome (Course Skill Set)

At the end of the course the student will be able to :

Sl. No.	Description	Blooms Level
CO1	Understand the fundamentals of social media analytics, data types, key metrics, and data collection techniques.	L2
CO2	Apply sentiment analysis and text mining techniques to analyze social media data using NLP and ML models.	L2
CO3	Analyze social network structures and trends using graph-based analytics, hashtag modeling, and virality detection.	L3
CO4	Visualize and interpret social media insights using tools like Tableau and Power BI to support decision-making.	L4
CO5	Evaluate real-world applications of social media analytics in business, digital marketing, bran monitoring, and ethical considerations.	L5

Mapping of COS and POs

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	**1**	**2**						
CO2		**2**	**2**		**3**			
CO3		**2**		**3**	**3**			
CO4			**2**		**3**			**3**
CO5					**3**	**3**		