SREENIVASAINSTITUTEOFTECHNOLOGYANDMANAGEMENTSTUDIES
(Autonomous)

Minor - DATA ANALYTICS - Offered by CSD Department

S.No	Course Code	CourseTitle	Scheme of Instructions HoursperWeek				Scheme ofExaminationMaximumMarks		
			L	T	P	C	I	E	Total
1	23MRDAS1	Introduction to Data Analytics	3	-	-	3	30	70	100
2	23MRDAS2	Data Engineering	3	-	-	3	30	70	100
3	23MRDAS3	Predictive Analytics	3	-	-	3	30	70	100
4	23MRDAS4	Big Data Analytics (Hadoop,Spark)	3	-	-	3	30	70	100
5	23MRDAS5	Data Analytics with Power I/Tableau/Matplotlib	3	-	-	3	30	70	100
6	23MRDAS6	Data Analytics Tools Lab	-	-	3	1.5	30	70	100
7	23MRDAS7	Big Data &NoSQL Lab	-	-	3	1.5	30	70	100

SREENIVASAINSTITUTEOFTECHNOLOGYANDMANAGEMENTSTUDIES
(Autonomous)

CSD Department

DATA ANALYTICS

23MRDAS1	INTRODUCTION TO DATA ANALYTICS	L	T	P	C
		3	0	0	3

PRE-REQUISITES:

COURSE EDUCATIONAL OBJECTIVES:

- To introduce the fundamental concepts of data analytics.
- To understand the process of collecting, cleaning, analyzing, and interpreting data.
- To gain insights into exploratory data analysis (EDA) and visualization techniques.
- To apply statistical and machine learning techniques to draw inferences from data.
- To use analytical tools (like Python/R) to solve real-world data problems.

UNIT –1: Introduction to Data Analytics (9)

Data, Information, and Knowledge,Types of data: Structured, Unstructured, Semi-Structured, Phases of the Data Analytics Lifecycle, Introduction to Business Analytics, Tools for data analytics: Excel, Python, R, SQL.

UNIT –2: Data Collection and Pre processing (9)

Sources of data: Databases, APIs, Web Scraping, Handling missing data and outliers, Data cleaning, transformation, normalization, Feature engineering and selection, Data partitioning and sampling techniques.

UNIT –3: Descriptive and Inferential Statistics (9)

Central tendency and dispersion: Mean, Median, Mode, Standard Deviation, Correlation and covariance, Probability distributions: Normal, Binomial, Poisson, Hypothesis testing, t-test, Chi-square test, Confidence intervals and p-values.

UNIT –4: Exploratory Data Analysis (EDA) and Visualization (9)

Univariate, Bivariate, Multivariate Analysis, Histograms, Boxplots, Scatter plots, Pair plots, Heatmaps and correlation matrices, Data dashboards and storytelling using visualizations, Tools: Matplotlib, Seaborn, Plotly, Tableau basics.

UNIT –5: Data-Driven Decision Making (9)

Predictive analytics vs descriptive analytics, Introduction to regression and classification, Clustering and segmentation basics, Model evaluation metrics: Accuracy, Precision, Recall, F1-score, Case studies in marketing, healthcare, finance, operations

Total Hours: 45

COURSE OUTCOMES:

On successful completion of the course- students will be able to		**Bloom's Level**
CO1	Understand the core concepts and applications of data analytics.	Understand (L2)
CO2	Prepare and preprocess data for analysis using appropriate techniques.	Apply (L3)
CO3	Analyze data using descriptive and inferential statistical methods.	Analyze (L4)
CO4	Visualize data insights through various graphical representations.	Evaluate (L5)
CO5	Apply data analytics tools for solving domain-specific problems.	Create (L6)

TEXT BOOKS:

1. **V. Uday Shankar** – *Data Analytics*, Cengage Learning
2. **P. N. Murthy** – *Introduction to Data Analytics*, Himalaya Publishing
3. **Anil Maheshwari** – *Data Analytics Made Accessible*, Amazon Independent

REFERENCE BOOKS:

1. **Foster Provost & Tom Fawcett** – *Data Science for Business*, O'Reilly
2. **Joel Grus** – *Data Science from Scratch*, O'Reilly
3. **Wes McKinney** – *Python for Data Analysis*, O'Reilly
4. **Allen Downey** – *Think Stats*, Green Tea Press

REFERENCE WEBSITE:

Platform	Course Title	Link
Coursera	Introduction to Data Analytics – IBM	

SREENIVASAINSTITUTEOFTECHNOLOGYANDMANAGEMENTSTUDIES
(Autonomous)

23MRDAS2	DATA ENGINEERING	L	T	P	C
		3	0	0	3

PRE-REQUISITES:

COURSE EDUCATIONAL OBJECTIVES:

- To introduce the architecture, components, and workflow of modern data engineering systems.
- To provide knowledge on ingestion, storage, and processing of large-scale data.
- To apply tools and techniques like ETL, data pipelines, and distributed processing frameworks.
- To explore data modeling, warehousing, and real-time data processing.
- To build scalable and maintainable data systems with cloud and open-source technologies.

UNIT –1: Introduction to Data Engineering (9)

What is Data Engineering?,Role of Data Engineer in the data ecosystem, Types of data: Structured, Semi-structured, Unstructured, Data Lifecycle: Collection to Consumption, Introduction to ETL and ELT processes.

UNIT –2: Data Ingestion and Storage (9)

Batch vs Streaming data, Data ingestion tools: Sqoop, Flume, Kafka, Logstash, Structured storage: Relational DBs, Data Lakes, File formats: CSV, JSON, Avro, Parquet, ORC, NoSQL overview: HBase, MongoDB, Cassandra.

UNIT –3: Distributed Data Processing (9)

Hadoop ecosystem: HDFS, MapReduce, YARN, Apache Spark: RDD, DataFrame, and SQL APIs, Data cleaning and transformation techniques, Resource allocation and job scheduling in Spark, Introduction to cloud-based data processing (AWS/GCP).

UNIT –4: Data Warehousing & Modeling (9)

Concepts of OLTP vs OLAP, Star and Snowflake Schemas, Fact and Dimension tables, Data Warehousing tools: Amazon Redshift, Google BigQuery, Snowflake, Data modeling tools: dbt (data build tool), ER diagrams.

UNIT –5: Real-Time Data Engineering and Pipelines (9)

Streaming with Apache Kafka and Spark Streaming, Message Queues and Pub/Sub systems, Lambda and Kappa architectures, Monitoring and logging tools: Prometheus, Grafana, Case study: End-to-end data pipeline implementation.

Total Hours: 45

COURSE OUTCOMES:

On successful completion of the course- students will be able to		**Bloom's Level**
CO1	Understand core concepts of data engineering and data pipelines.	Understand (L2)
CO2	Apply data ingestion, transformation, and storage techniques.	Apply (L3)
CO3	Analyze big data processing frameworks and streaming systems.	Analyze (L4)
CO4	Evaluate data warehousing models and real-time processing strategies.	Evaluate (L5)
CO5	Design and implement scalable data engineering workflows.	Create (L6)

SREENIVASAINSTITUTEOFTECHNOLOGYANDMANAGEMENTSTUDIES
(Autonomous)

TEXT BOOKS:

1. **Andreas François Vermeulen** – *Data Engineering with Python*, Packt Publishing
2. **V. Uday Shankar** – *Big Data Analytics & Data Engineering*, Cengage Learning
3. **Sam Newman** – *Building Microservices*, O'Reilly (for data infrastructure components)

REFERENCE BOOKS:

1. **Bill Inmon** – *Building the Data Warehouse*, Wiley
2. **Tom White** – *Hadoop: The Definitive Guide*, O'Reilly
3. **Jules S. Damji et al.** – *Learning Spark: Lightning-Fast Big Data Analysis*, O'Reilly
4. **Noel Markham** – *Learning Apache Kafka*, Packt Publishing

REFERENCE WEBSITE:

Platform	Course Title	Link
Coursera	Data Engineering on Google Cloud – Google	

SREENIVASAINSTITUTEOFTECHNOLOGYANDMANAGEMENTSTUDIES
(Autonomous)

23MRDAS3	PREDICTIVE ANALYTICS	L	T	P	C
		3	0	0	3

PRE-REQUISITES:

COURSE EDUCATIONAL OBJECTIVES:

- To understand the fundamentals of predictive analytics and modeling.
- To learn statistical, machine learning, and probabilistic techniques used for prediction.
- To build and evaluate predictive models for various types of data.
- To apply data pre processing, feature engineering, and model tuning techniques.
- To explore real-world use cases of predictive analytics in business, health, and industry.

UNIT –1: Introduction to Predictive Analytics (9)

Basics of predictive analytics, Types of predictive models: Classification vs Regression, Predictive modeling workflow, Applications in finance, healthcare, marketing, etc., Introduction to supervised learning.

UNIT –2: Data Preparation and Feature Engineering (9)

Data cleaning and preprocessing, Handling missing values, outliers, Categorical encoding, normalization, scaling, Feature selection and dimensionality reduction (PCA, LDA), Feature importance.

UNIT –3: Regression and Classification Models (9)

Linear Regression, Ridge & Lasso Regression, Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, Random Forests, Model diagnostics: R^2, MAE, RMSE, Confusion Matrix, AUC

UNIT –4: Advanced Predictive Modeling (9)

Support Vector Machines (SVM), Ensemble Models: Bagging, Boosting, XGBoost, Model selection and hyperparameter tuning, Cross-validation, Grid Search, Avoiding overfitting/underfitting.

UNIT –5: Real-Time Predictive Analytics and Deployment (9)

Case studies: Predictive maintenance, customer churn, disease prediction, Tools: Python (scikit-learn, pandas), R, AutoML, Model interpretability: SHAP, LIME, Model deployment using Flask, FastAPI, or cloud (AWS/GCP), Ethics in predictive analytics.

Total Hours: 45

COURSE OUTCOMES:

On successful completion of the course- students will be able to		Bloom's Level
CO1	Understand key concepts, tools, and algorithms used in predictive analytics.	Understand (L2)
CO2	Apply classification and regression techniques for prediction problems.	Apply (L3)
CO3	Analyze model performance using error metrics and validation techniques.	Analyze (L4)
CO4	Evaluate predictive models in real-world applications and domains.	Evaluate (L5)
CO5	Design complete predictive analytics pipelines with preprocessing to prediction.	Create (L6)

TEXT BOOKS:

1. **Dean Abbott** – *Applied Predictive Analytics: Principles and Techniques*, Wiley
2. **GalitShmueli et al.** – *Data Mining for Business Analytics: Concepts, Techniques, and Applications*, Wiley
3. **John D. Kelleher et al.** – *Fundamentals of Machine Learning for Predictive Data Analytics*, MIT Press

REFERENCE BOOKS:

1. **Trevor Hastie, Robert Tibshirani, Jerome Friedman** – *The Elements of Statistical Learning*
2. **Ian H. Witten, Eibe Frank, Mark A. Hall** – *Data Mining: Practical Machine Learning Tools and Techniques*
3. **Tom Fawcett** – *Data Science for Business*, O'Reilly

REFERENCE WEBSITE:

Platform	Course Title & Link
Coursera	Advanced Predictive Modeling – University of Washington

SREENIVASAINSTITUTEOFTECHNOLOGYANDMANAGEMENTSTUDIES
(Autonomous)

23MRDAS4	BIG DATA ANALYTICS (Hadoop, Spark)	L	T	P	C
		3	0	0	3

PRE-REQUISITES:

COURSE EDUCATIONAL OBJECTIVES:

- To understand the need and concepts of Big Data and its ecosystem.
- To study architectures and tools such as Hadoop and Spark for large-scale data processing.
- To explore Hadoop Distributed File System (HDFS) and MapReduce programming.
- To learn in-memory computation using Apache Spark.
- To implement big data solutions using real-time tools for analytics and decision-making.

UNIT –1: Introduction to Big Data (9)

Characteristics of Big Data: Volume, Velocity, Variety, Veracity, and Value, Challenges of Big Data, Traditional vs. Big Data systems, Big Data architecture: Storage, processing, analytics, Introduction to Hadoop Ecosystem

UNIT –2: Hadoop Distributed File System (HDFS) and MapReduce (9)

HDFS Architecture, Blocks, NameNode&DataNode, File Read/Write operations, MapReduce Architecture, Developing MapReduce programs (Word Count Example), Combiners, Partitioner, and Counters

UNIT –3: Apache Spark and RDDs (9)

Need for Apache Spark over Hadoop, Spark architecture and components, Spark RDDs and transformations/actions, DataFrames and Spark SQL, Working with Spark MLlib for basic ML tasks

UNIT –4: Big Data Tools and Streaming (9)

YARN architecture, Sqoop: Import/export between RDBMS and Hadoop, Flume: Ingesting streaming data, Kafka for distributed messaging, Spark Streaming: DStreams, windowing, real-time analytics

UNIT –5: Big Data Applications and Case Studies (9)

Social media analytics, IoT data pipelines, Retail analytics (recommendation engines, customer sentiment), Fraud detection, End-to-End Mini Project: Data ingestion, processing, and visualization

Total Hours: 45

COURSE OUTCOMES:

On successful completion of the course- students will be able to		Bloom's Level
CO1	Explain the concepts, characteristics, and challenges of Big Data.	Understand (L2)
CO2	Apply HDFS and MapReduce programming for large-scale data processing.	Apply (L3)
CO3	Analyze and optimize data pipelines using Spark RDDs and DataFrames.	Analyze (L4)
CO4	Evaluate Big Data tools and frameworks for distributed processing.	Evaluate (L5)
CO5	Design end-to-end Big Data workflows for analytics use cases.	Create (L6)

TEXT BOOKS:

1. **Tom White**, *Hadoop: The Definitive Guide*, O'Reilly Media
2. **Jules S. Damji et al.**, *Learning Spark: Lightning-Fast Big Data Analysis*, O'Reilly
3. **VenkatAnkam**, *Big Data Analytics Using Spark*, Wiley

REFERENCE BOOKS:

1. **Chuck Lam**, *Hadoop in Action*, Manning
2. **Alex Holmes**, *Hadoop in Practice*, Manning Publications
3. **P. J. Sadalage& M. Fowler**, *NoSQL Distilled*, Addison-Wesley
4. **Sridhar Alla**, *Big Data Analytics with Hadoop 3*, Packt Publishing

REFERENCE WEBSITE:

Platform	Course Title & Link
Coursera	Big Data Specialization – UC San Diego

SREENIVASAINSTITUTEOFTECHNOLOGYANDMANAGEMENTSTUDIES
(Autonomous)

23MRAIA5	DATA ANALYTICS WITH POWER BI / TABLEAU / MATPLOTLIB	L	T	P	C
		3	0	0	3

PRE-REQUISITES:

COURSE EDUCATIONAL OBJECTIVES:

- To understand the fundamentals of data visualization and storytelling using analytics tools.
- To learn practical implementation of visual analytics using Power BI, Tableau, and Matplotlib.
- To develop dashboards, custom reports, and interactive charts.
- To explore real-world datasets and generate insights.
- To design end-to-end data pipelines and reporting workflows.

UNIT –1: Introduction to Data Visualization & BI (9)

What is data analytics and visualization?,Importance of business intelligence, Types of charts and graphs, Visualization best practices and perception, Introduction to tools: Power BI, Tableau, and Python/Matplotlib

UNIT –2: Power BI for Data Analytics (9)

Power BI interface and data sources, Data transformations using Power Query, DAX (Data Analysis Expressions) basics, Visualizations: Bar, Line, Area, Pie, Tree Maps, Report publishing and dashboard sharing

UNIT –3: Tableau for Data Analytics (9)

Tableau workspace and connecting to datasets, Dimensions vs. Measures, Filters, Parameters, and Calculated Fields, Visualizations: Heatmaps, Maps, Scatter Plots, Building Interactive Dashboards

UNIT –4: Matplotlib & Seaborn (Python Visualization) (9)

Introduction to Matplotlib architecture, Basic plots: line, bar, pie, histogram, Subplots, customization (color, label, grid), Seaborn: heatmaps, boxplots, pairplots, Exporting and saving visuals

UNIT –5: Real-time Case Studies and BI Projects (9)

Designing end-to-end dashboards for:Sales analytics, Healthcare data, IoT or Sensor data dashboards, Social media analytics, Performance monitoring and KPI analysis, BI deployment best practices

Total Hours: 45

COURSE OUTCOMES:

On successful completion of the course- students will be able to		Bloom's Level
CO1	Understand data analytics lifecycle and fundamentals of visualization tools.	Understand (L2)
CO2	Apply Power BI, Tableau, and Matplotlib to visualize structured data.	Apply (L3)
CO3	Analyze visual trends and compare different types of charting techniques.	Analyze (L4)
CO4	Evaluate dashboards and performance metrics for decision-making.	Evaluate (L5)
CO5	Design interactive data dashboards and business visual reports.	Create (L6)

TEXT BOOKS:

1. **Alberto Ferrari & Marco Russo**, *Introducing Microsoft Power BI*, Microsoft Press.
2. **Joshua N. Milligan**, *Learning Tableau*, Packt Publishing.
3. **AdrienChauveau**, *Data Visualization with Python and Matplotlib*, Packt Publishing.

REFERENCE BOOKS:

1. **Ben Jones**, *Storytelling with Data in Tableau*, O'Reilly.
2. **Jake VanderPlas**, *Python Data Science Handbook*, O'Reilly.
3. **Cole NussbaumerKnaflic**, *Storytelling with Data*, Wiley.

REFERENCE WEBSITE:

Platform	Course Title & Link
Coursera	Data Visualization with Tableau – UC Davis

SREENIVASAINSTITUTEOFTECHNOLOGYANDMANAGEMENTSTUDIES
(Autonomous)

23MRDAS6	DATA ANALYTICS TOOLS LAB	L	T	P	C
		-	-	3	1.5

PRE-REQUISITES: Nil.

COURSE EDUCATIONAL OBJECTIVES:

- To visualize and interpret data using modern tools.
- To create dashboards and analytical views.

Experiments:

1. Introduction to Power BI / Tableau interface
2. Loading data and transforming in BI tools
3. Creating bar, pie, and line charts
4. Cross filters and slicers
5. Using DAX in Power BI
6. Map visualization
7. Dashboards creation
8. Connecting to real-time data
9. Storytelling with data
10. Introduction to Matplotlib for comparison
11. Using Seaborn and Plotly

Course Outcomes:

- Use BI tools for analytical visualizations.
- Interpret datasets using graphical methods.

SREENIVASAINSTITUTEOFTECHNOLOGYANDMANAGEMENTSTUDIES
(Autonomous)

23MRDAS7	BIG DATA & NOSQL LAB	L	T	P	C
		-	-	3	1.5

PRE-REQUISITES: Nil.

COURSE EDUCATIONAL OBJECTIVES:

- To process big datasets using distributed systems.
- To implement NoSQL solutions.

Experiments:

1. Introduction to HDFS and Hadoop
2. MapReduce word count example
3. Spark RDD operations
4. DataFrames and Spark SQL
5. PySpark ML basics
6. Connecting MongoDB to Python
7. CRUD operations in MongoDB
8. Indexing and Aggregation
9. Replication and Sharding demo
10. Case study: Big Data pipeline
11. Real-time analytics using Spark Streaming

Course Outcomes:

- Handle large datasets using Hadoop/Spark.
- Perform CRUD operations in NoSQL.