Gyanmanjari Innovative University

Course Syllabus
Faculty of Engineering &Technology
Semester-1 (M.Tech)

Subject: Foundation of Data Science - METAI11502

Type of course: Major (Core)

Prerequisite: Basic knowledge of programming (preferably Python), mathematics, statistics, and database concepts (SQL).

Rationale:

Data Science integrates domain expertise from programming, mathematics, and statistics to extract meaningful insights from data and support informed decision-making. By transforming raw data into actionable knowledge, organizations can minimize uncertainty and gain strategic advantages.

This course provides a foundational understanding of data science concepts, tools, and techniques used in various applications such as statistical analysis, predictive modeling, and time-series analytics. It emphasizes the importance of data preprocessing and visualization as essential steps before applying statistical or machine learning models. Through this course, students will develop the analytical and technical skills required to handle data-driven challenges effectively.

Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks					Total Marks
				Theory Marks		Practical Marks		CA	
CI	T	P	C	ESE	MSE	V	P	ALA	
4	0	2	5	60	30	10	20	30	150

Legends: CI-ClassRoom Instructions; T – Tutorial; P - Practical; C – Credit; ESE - End Semester Examination; MSE- Mid Semester Examination; V – Viva; CA - Continuous Assessment; ALA- Active Learning Activities.

Course Content:

Sr. No	Course Content	Hrs.	% Weightage
1	**Fundamentals of Data Science** Introduction to Data Science and its Applications, Data Analytics Life Cycle, Types of Data Analysis: Descriptive, Diagnostic, Predictive, Prescriptive, Roles and Responsibilities in Data Science (Data Scientist, Data Analyst, Data Engineer, ML Engineer), Overview of Data Science Tools and Environments, Fundamental Areas of Study in Data Science (Statistics, Machine Learning, Data Engineering, Visualization, Big Data Concepts), Role of Databases and SQL in Data Science, Introduction to Data Storage and Management (SQL vs NoSQL Systems, Data Warehousing Concepts), Advantages, Challenges, and Limitations of Data Science	12	20
2	**Data Preprocessing, Feature Engineering, and Visualization** Data types and forms, Possible data error types, Various data preprocessing operations,Common Data Quality Issues: Missing Values, Outliers, Duplicates, Inconsistencies, **Feature Engineering:** Handling Categorical and Numerical Data, Encoding Techniques, Feature Selection and Extraction, Introduction to data visualization, Visual encoding, Data visualization libraries, Basic data visualization tools.	15	25
3	**Statistical Data Analysis and Machine Learning** Role of statistics in data science, Descriptive statistics, Inferential statistics, Overview of machine learning, Supervised machine learning, Regression methods, Classification methods, Unsupervised machine learning, Clustering methods, Association analysis.	15	25
4	**Time-series Analysis** Overview of time-series analysis, Components of time-series, Time-Series Forecasting Models: AR, MA, ARIMA, SARIMA, Exponential Smoothing, Model Evaluation Metrics for Forecasting, Basics of Advanced Approaches: LSTM (Overview only)	09	15
5	**Ethics, Privacy, and Governance in Data Science** The Five Cs: Consent, Clarity, Consistency and Trust, Control and Transparency, Consequences, Implementing Five Cs, Data's Day of Reckoning, Ethics and Security Training, Developing Guiding Principles, Building Ethics into Data-drive Culture, Regulation, Data Governance and Regulation (GDPR, India's DPDP Act, Data Protection Standards), Case Studies on Ethical and Privacy Challenges in Data Science	09	15

Continuous Assessment:

Sr. No	Active Learning Activities	Marks
1	**Exploratory Data Analysis (EDA) and Visualization** In this activity, students will perform exploratory data analysis using Python. They will clean and preprocess a real-world dataset, apply feature engineering, and create effective visualizations using Matplotlib or Seaborn. The report with findings and plots will be submitted on the GMIU Web Portal.	10
2	**Machine Learning Model Implementation** In this activity, students will select a dataset, apply appropriate statistical analysis, and implement a basic machine learning model (e.g., regression or classification). They will evaluate performance using model metrics such as accuracy, precision, and recall. The final summary and code will be submitted on the GMIU Web Portal.	10
3	**Ethics and Data Governance Case Study** In this case study, students will analyze real-world ethical or privacy challenges in data science (e.g., bias in AI models or misuse of personal data). They will discuss the Five Cs of Data Ethics and propose guidelines for responsible data use. A short report will be submitted on the GMIU Web Portal.	10
	Total	30

Suggested Specification table with Marks (Theory):60

Distribution of Theory Marks (Revised Bloom's Taxonomy)						
Level	Remembrance (R)	Understanding (U)	Application (A)	Analyze (N)	Evaluate (E)	Create (C)
Weightage %	20%	25%	25%	20%	10%	-

Course Outcome:

After learning the course the students should be able to:	
CO1	Explain the fundamental concepts, life cycle, tools, and roles in data science and recognize the relevance of SQL and data management systems in analytics.
CO2	Apply various data preprocessing, cleaning, and feature engineering techniques, and create effective visual representations using modern data visualization tools.

CO3	Analyze data using statistical methods and implement basic machine learning algorithms for regression, classification, clustering, and model evaluation.
CO4	Develop and interpret time-series models for forecasting and understand their applications in real-world scenarios.
CO5	Evaluate ethical, legal, and governance issues in data science, and integrate responsible practices in data-driven decision-making.

Instructional Method:

The course delivery will depend on the content requirements and student needs. Along with conventional classroom teaching using the blackboard and presentations, the instructor may employ various active learning methods such as **demonstrations, quizzes, brainstorming sessions, case studies,** and **group discussions**.

Use of **MOOCs**, **NPTEL videos**, and **online e-learning resources** is encouraged to supplement classroom learning. Students may also engage with **virtual labs** and **open datasets** for self-learning and practice.

The **internal evaluation** will be carried out through **continuous assessment** based on assignments, tutorials, class participation, and mini-projects to ensure conceptual understanding and analytical skill development.

Reference Books:

[1] Data Science Fundamentals and Practical Approaches by Gypsi Nandi and Rupam Kumar Sharma, First Edition, BPB Publications India.

[2] Ethics and Data Science by Mike Loukides, Hilary Mason, & D J Patil, First Edition, O'Reilly.

[3] Introducing Data Science by Davy Cielen, Arno D. B. Meysman, Mohamed Ali, Latest Edition, Manning.

[4] Data Science and Analytics with Python by Jesus Rogel-Salazar, Latest Edition, CRC Press Taylor & Francis Group.