School of Engineering
B. Tech. (common to all disciplines) I Year
B. Tech. (CSE) with Specialization Artificial Intelligence and Machine Learning
(In Association with IBM)

Semester – I

BCO118A	**Data Science Methodology, Data Science 101**	**1:0:0(1 Credit)**

Course Objectives:

1. The major steps involved in tackling a data science problem.
2. The major steps involved in practicing data science, from forming a concrete business or research problem, to collecting and analyzing data, to building a model, and understanding the feedback after model deployment.
3. How data scientists think!
4. Find out the truth about what Data Science is.
5. Hear from real practitioners telling real stories about what it means to work in data science.

UNIT 1: From Problem to Approach
Business Understanding, Analytic Approach, What is data science?, There are many paths to data science, Any advice for a new data scientist?, What is the cloud?.

UNIT 2: From Requirements to Collection
Data Requirements, Data Collection, A day in the life of a data science person, R versus Python?, Data science tools and technology, "Regression".

UNIT 3: From Understanding to Preparation
Data Understanding, Data Preparation, How should companies get started in data science?, Tips for recruiting data science people, "The Final Deliverable"

UNIT 4: From Modeling to Evaluation
Modeling, Evaluation, Applications for data science, "The Report Structure"

UNIT 5: From Deployment to Feedback
Deployment, Feedback, Things data science people say, "What Makes Someone a Data Scientist?"

Course outcome:
CO1: Understand basic concepts of data science and cloud technologies.
CO2: Identify data requirements and tools used in data science.
CO3: Apply techniques for data understanding and preparation.
CO4: Explain data modeling, evaluation, and reporting methods.
CO5: Understand deployment, feedback, and professional practices in data science.

CO-PO Mapping:

Course Outcome	Program Outcome												Program Specific Outcome		
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	H	M			*H*							*M*	H	M	
CO2	H	H			H							*M*	*H*	H	
CO3	H	H	M	M	M								H	H	*M*
CO4	H	H	M	H	H					*M*			H	H	
CO5	M	H	H	M	H	*M*		*H*	*M*	*H*	*M*	*M*	M	M	H

BCO117A	**Data Visualization**	**2:0:0(2 Credit)**

Course objectives:

- Learn basic concepts such as the mean, median etc.
- Learn practical implementation statistical concepts using R, IBM Watson Studio and python.
- The Data Visualization Training Module will give the reader a thorough introduction to Data Science, Statistics, R, IBM Watson Studio and python using real life examples.

UNIT 1: From Problem to Approach, Business Understanding - Concepts & Case Study, Analytic Approach - Concepts & Case Study, Introduction to Statistics: Difference between inferential, statistics and descriptive statistics, Inferential Statistics, Drawing Inferences from Data, Random Variables, Normal Probability Distribution, Sampling, Sample Statistics and Sampling Distributions

UNIT 2: R overview and Installation: Overview and About R, Data visualization with R

UNIT 3:Introduction to Python: Python and Anaconda Installation, Introduction to Jupyter Notebook, Python scripting basics, Numpy and Pandas, Numpy overview - Creating and Accessing Numpy Arrays, Introduction to pandas, Pandas read and write csv, Descriptive statistics using pandas, Pandas working with text data and datetime columns, Pandas Indexing and selecting data, Pandas – groupby,Merge / Join datasets

UNIT 4: Introduction to Data Visualization Tools in Python, Introduction to Matplotlib, Read a CSV and Generate a line plot with matplotlib, Basic plots using matplotib, Area Plots, Bar Charts, Histograms, Specialized Visualization Tools using Matplotlib, Pie Charts, Box Plot, Scatter Plots, Bubble Plots, Advanced Visualization Tools using Matplotlib: Waffle Charts, Word Clouds

UNIT 5: Introduction to Seaborn: Seaborn functionalities and usage with Hands-on, Spatial Visualizations and Analysis in Python with Folium, Introduction to Folium, Case Study (Analyze New York City Taxi Trip Ride Data Set to Identify best locations for taxi stops)

Course outcome:

CO1: Understand business problems and apply basic statistical concepts for data analysis.
CO2: Use R for data handling and visualization.
CO3: Apply Python with NumPy and Pandas for data processing.
CO4: Create visualizations using Matplotlib to explore data insights.
CO5: Use Seaborn and Folium for advanced and spatial data visualization through case studies.

CO PO mapping:

Course Outcome	Program Outcome												Program Specific Outcome		
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	H	H		M									H	M	
CO2		M			H									H	
CO3	H	H			H								H	H	
CO4		H		H	H								M	H	
CO5	H	H	H	H	H		M						H	M	H

Semester – II

BCO119B	Introduction to Python	2-0-0(2 Credits)

Course Objectives:

1. Learn basic concepts of python variables, loops, dictionaries, tuples, operators and operands.
2. Learn about python libraries, error handling and concepts Machine Learning Algorithms.
3. This course will introduce basic python programming concepts such as variables, loops, dictionaries, tuples etc , then advance concepts on libraries.
4. This course also throw light on Machine Learning Algorithms.

UNIT 1: Introduction to Python: Introduction, Installation, variables, Operators and Strings
UNIT 2: Deep Dive into Python: Input Output functions, Loops, List, dictionaries, tuples, File Handler
UNIT 3: Python Libraries: Pandas, Series and Data Frames, Grouping, aggregating, and applying
UNIT 4: Error Handling: Dealing with syntax errors, Exceptions, Handling exceptions with try/except
UNIT 5: Advance: Regression, Correlation Matrix, Linear Regression, Machine Learning Algorithms.

Course Outcome:
CO1: Understand basic Python concepts including variables, operators, and strings.
CO2: Apply control structures, data structures, and file handling in Python programs.
CO3: Use Python libraries like Pandas for data manipulation and analysis.
CO4: Implement error handling techniques using exceptions in Python.
CO5: Apply regression and basic machine learning algorithms using Python.

CO-PO Mapping:

Course Outcome	*Program Outcome*												*Program Specific Outcome*		
	PO1	*PO2*	*PO3*	*PO4*	*PO5*	*PO6*	*PO7*	*PO8*	*PO9*	*PO10*	*PO11*	*PO12*	*PSO1*	*PSO2*	*PSO3*
CO1	*H*	*M*			*M*							*M*	*H*	*M*	
CO2	*H*	*H*			*H*							*M*	*H*	*H*	
CO3	*H*	*H*		*M*	*H*							*M*	*H*	*H*	
CO4	*H*	*M*		*M*	*M*							*M*	*M*	*M*	
CO5	*H*	*H*	*M*	*H*	*H*					*M*	*M*	*M*	*H*	*H*	*H*

DCO120A	ML with Sound Lab	0-0-2(1 Credits)

Course Objectives:

1. Prepare data so that it can be consumed by machine learning models
2. Build a binary classification model that can predict which animal (dogs and cats) is making a specific sound
3. Build a multiclass classification model to detect whether a birdsong is from a bird from a specific order and view the confidence level of that prediction
4. Make predictions on audio files by using a Node-RED application built as a web page
5. Create an application in Node-RED that integrates the Watson Visual Recognition service with your machine learning model to recognize images of cats and dogs

Module 1: IBM Watson studio, Gather and prepare the data.
Module 2: Build a machine learning model.
Module 3: Create predictions in a node-RED application.
Module 4: Create multiclass classification model.
Module 5: Create UIs and integrate visual recognition.

Course outcomes:
CO1: Preprocess and transform raw audio data into suitable formats for effective machine learning model training and evaluation.
CO2: Design, train, and evaluate binary classification models to accurately predict animal sounds (dogs vs. cats).
CO3: Develop and deploy multiclass classification models to identify bird species from specific orders and interpret prediction confidence scores.
CO4: Implement a Node-RED based web application for making real-time predictions on audio files using trained machine learning models.
CO5: Integrate Watson Visual Recognition service with a Node-RED application to recognize and classify images of cats and dogs.

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1:	H	M	L	M	H	–	–	–	–	M	–	M	H	M	M
CO2:	H	H	H	M	H	–	–	–	–	M	–	M	H	H	H
CO3:	H	H	H	H	H	–	L	–	–	M	–	M	H	H	H
CO4:	H	M	H	M	H	–	–	–	M	H	M	M	H	H	H
CO5:	H	M	H	M	H	–	–	–	M	H	M	M	H	H	H

Semester – III

BCO 121A	Rapid Development for AI (AI Services)	3-0-0 [3]

Objective:

1. Introduction to Watson.
2. Natural Language Understanding
3. Conversational AI
4. Build Your Own AI using Watson Tools
5. Understand Ai services

UNIT 1: *Introduction to Watson*
- Overview of IBM Watson and its capabilities
- Core Watson services and their applications
- Accessing Watson through IBM Cloud

UNIT 2: *Natural Language Understanding (NLU)*
- Fundamentals of NLU
- Text analysis, entity recognition, sentiment analysis
- Using Watson NLU for extracting structured information from unstructured text

UNIT 3: *Conversational AI*
- Basics of conversational AI and chatbots
- Designing conversational flows
- Implementing chatbots using Watson Assistant

UNIT 4: *Build Your Own AI Using Watson Tools*
- Overview of Watson Studio and related tools
- Creating, training, and testing custom AI models
- Deploying AI models in applications

UNIT 5: *Understanding AI Services*
- Overview of key AI services (Vision, Speech, Language, Data)
- Integration of multiple AI services
- Use cases and real-world applications of Watson AI services

Course Otcomes:
CO1: Explain the features, architecture, and applications of IBM Watson AI platform.
CO2: Apply Watson Natural Language Understanding (NLU) tools to analyze and interpret unstructured text data.
CO3: Design and implement conversational AI solutions using Watson Assistant.
CO4: Develop and deploy custom AI models using Watson Studio and related tools.
CO5: Integrate multiple Watson AI services to create intelligent, real-world applications.

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1:	H	M	L	L	M	–	–	–	–	M	–	M	H	M	M
CO2:	H	H	M	M	H	–	–	–	–	M	–	M	H	H	H
CO3:	H	H	H	M	H	–	–	–	M	H	M	M	H	H	H
CO4:	H	H	H	H	H	–	–	–	M	H	M	M	H	H	H
CO5:	H	H	H	M	H	L	L	–	H	H	H	H	H	H	H

BCO122A	Text Analytics 101	1-0-0 [1]

Objective:

1. **Introduction to Information Extraction (IE):** Understanding the process of distilling structured information from unstructured text, common IE tasks, and evaluation techniques.
2. **Classical Grammar-Based IE Systems:** Overview of traditional grammar-based approaches, their expressivity limitations, and runtime performance challenges.
3. **Declarative Information Extraction Paradigm:** Exploration of the emerging declarative approach for building high-performance IE systems.
4. **SystemT and AQL:** Introduction to SystemT as a scalable, accurate, and user-friendly IE system, and its declarative language AQL for creating basic and advanced extractors.
5. **Performance Optimization in AQL:** Best practices for writing efficient extractors and addressing common performance issues in AQL-based systems.

Module 1 - Getting to Know Information Extraction

Module 2 - Limitations in Information Extraction

Module 3 - Getting to Know SystemT

Module 4 - Information Extraction with AQL

Module 5 - AQL Basics

Module 6 - Advanced AQL

Module 7 - Declarative Information Extraction and the SystemT Optimizer

Module 8 - Best Practices

Course Outcomes (COs)

- **CO1:** Explain the concepts, features, limitations, and applications of Information Extraction systems.
- **CO2:** Demonstrate the architecture and functionalities of IBM SystemT for Information Extraction.
- **CO3:** Apply AQL (Annotation Query Language) to design and implement information extraction tasks.
- **CO4:** Develop optimized declarative information extraction solutions using advanced AQL and SystemT Optimizer.
- **CO5:** Apply best practices to build scalable, efficient, and real-world information extraction applications.

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1:	H	M	L	L	M	–	–	–	–	M	–	M	H	M	M
CO2:	H	M	M	L	H	–	–	–	–	M	–	M	H	H	H
CO3:	H	H	H	M	H	–	–	–	M	H	M	M	H	H	H
CO4:	H	H	H	H	H	–	–	–	M	H	M	M	H	H	H
CO5:	H	H	H	M	H	L	L	–	H	H	H	H	H	H	H

BCO 123A	IBM Watson Studio	2-0-0 [2]

OBJECTIVE:

1. To provide an overview of data projects and their role in modern organizations.
2. To develop skills for preparing, cleaning, and organizing data for analysis.
3. To introduce core concepts of data science and their applications.
4. To familiarize students with machine learning and AI models for real-world problem solving.
5. To understand the importance of data catalogs and governance in ensuring secure, reliable, and ethical use of data.

Unit 1: Introduction and Overview

- Introduction to Data-Driven Ecosystems
- Importance of Data in Decision-Making
- Data Lifecycle: Collection, Storage, Processing, Analysis
- Types of Data: Structured, Semi-Structured, and Unstructured
- Overview of Data Projects and their Business Impact

Unit 2: Data Projects and Preparation

- Defining and Scoping a Data Project
- Phases of a Data Project: Requirement, Design, Implementation, Evaluation
- Data Sources: Databases, APIs, Sensors, and Web Data
- Data Collection Techniques and Challenges
- Data Preparation: Cleaning, Normalization, Integration, and Transformation
- Handling Missing and Noisy Data

Unit 3: Data Science Foundations

- Introduction to Data Science Process and Methodologies
- Exploratory Data Analysis (EDA)
- Data Visualization and Interpretation
- Statistical Foundations for Data Science
- Tools and Platforms for Data Science (Python, R, Jupyter, etc.)
- Case Studies: Applications of Data Science in Various Domains

Unit 4: Machine Learning and AI Models

- Fundamentals of Machine Learning (ML)
- Types of ML: Supervised, Unsupervised, Reinforcement Learning
- Model Building and Training Process
- Evaluation Metrics for ML Models (Accuracy, Precision, Recall, F1-score, etc.)
- Introduction to AI Models and Deep Learning Basics

Unit 5: Data Catalogs and Governance

- Importance of Data Catalogs for Data Management
- Metadata Management and Data Discovery
- Data Quality, Lineage, and Traceability
- Principles of Data Governance and Compliance

Course Outcomes:

CO1: Explain the fundamental concepts of data-driven projects, including data preparation, catalogs, and governance.
CO2: Demonstrate the ability to collect, clean, and organize datasets for effective data science workflows.
CO3: Apply data science methods and techniques to solve real-world problems.
CO4: Build and evaluate machine learning and AI models to extract insights and support decision-making.
CO5: Integrate governance and catalog management principles to ensure secure, reliable, and ethical data practices.

CO-PO Mapping:

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1:	H	M	M	–	M	M	–	M	–	–	–	M	M	M	–
CO2:	H	H	M	–	M	–	–	–	–	–	–	M	H	H	–
CO3:	M	H	H	M	H	–	M	–	–	M	–	M	H	H	M
CO4:	H	H	H	H	H	–	M	–	–	M	–	M	H	H	H
CO5:	M	M	M	–	M	L	M	H	–	–	–	M	M	M	–

Semester – IV

BCO 124A	COGNITIVE DECISION SYSTEM FOR MANAGERS	3-0-0 [3]

OBJECTIVE:

1. **To introduce** the fundamental concepts, scope, and applications of modern data-driven technologies.
2. **To develop understanding** of analytics methods, tools, and techniques for deriving insights from structured and unstructured data.
3. **To provide exposure** to the principles and applications of Artificial Intelligence (AI) for solving real-world problems.
4. **To familiarize students** with data platforms, architectures, and technologies that enable scalable storage, processing, and management of data.
5. **To enable learners** to integrate systems, solutions, and governance frameworks for effective, secure, and ethical use of emerging technologies.

Module 1: Introduction

- Overview of Data-Driven Technologies and Digital Transformation
- Evolution of Analytics and Artificial Intelligence
- Data Ecosystem: Sources, Types, and Characteristics (Structured, Semi-Structured, Unstructured)
- Introduction to Data Value Chain: Collection → Processing → Insights → Decisions
- Real-world Applications of Data in Business, Healthcare, Education, and Governance

Module 2: Analytics

- Fundamentals of Descriptive, Diagnostic, Predictive, and Prescriptive Analytics
- Data Preparation: Cleaning, Normalization, and Integration
- Exploratory Data Analysis (EDA) and Visualization
- Statistical Foundations: Correlation, Regression, Hypothesis Testing
- Tools and Platforms for Analytics (Excel, Python, R, Power BI, Tableau)
- Case Studies: Applications of Analytics in Business and Society

Module 3: Artificial Intelligence

- Introduction to AI: Definitions, History, and Applications
- Machine Learning (ML): Supervised, Unsupervised, and Reinforcement Learning
- AI Models: Decision Trees, Regression, Clustering, Neural Networks
- Natural Language Processing (NLP) and Computer Vision Basics
- Model Training, Testing, and Evaluation Metrics
- Ethical AI: Bias, Transparency, and Responsible Use of AI

Module 4: Data Platforms

- Data Management Systems: Databases, Data Warehouses, and Data Lakes
- Big Data Ecosystem: Hadoop, Spark, and Cloud Platforms (AWS, Azure, GCP)
- Data Pipelines: Ingestion, Transformation, and Orchestration
- Metadata Management and Data Catalogs
- Data Security, Privacy, and Compliance Regulations (GDPR, HIPAA)
- Scalability, Performance, and Reliability of Data Platforms

Module 5: Systems and Solutions

- Designing End-to-End Data Solutions
- Integration of Analytics, AI, and Data Platforms into Systems
- Business Intelligence (BI) Dashboards and Decision-Support Systems
- Case Studies: AI-Driven Business Solutions, Smart Cities, Healthcare Systems
- Governance and Ethical Considerations in Deploying Data Systems
- Future Trends: Cloud-Native AI, Edge Analytics, and Autonomous Systems

Course Outcome:

CO1: Explain the fundamental concepts, scope, and applications of data-driven technologies.
CO2: Apply analytics methods, tools, and visualization techniques to derive insights from data.
CO3: Build and evaluate Artificial Intelligence and Machine Learning models for real-world applications.
CO4: Demonstrate understanding of data platforms, architectures, and tools for scalable data storage, processing, and management.
CO5: Integrate systems and solutions with governance and ethical considerations to address industry and societal challenges.

CO-PO mapping:

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	H	M	M	–	M	M	M	M	–	–	–	M
CO2	H	H	M	M	H	–	M	–	–	M	–	M
CO3	H	H	H	H	H	–	M	–	–	M	–	M
CO4	H	M	M	M	H	–	H	–	–	–	–	M
CO5	M	M	M	–	M	H	H	H	–	M	M	M

BCO 125A	Deep Learning with TensorFlow	1-0-0 [1 credit]

OBJECTIVE:

- Understand the core concepts of TensorFlow and its role in deep learning.
- Install and configure TensorFlow in a development environment.
- Explain the computational graph, tensors, and sessions in TensorFlow.
- Develop simple neural network models using TensorFlow.
- Debug and optimize TensorFlow models for efficiency.

Module 1: Introduction to TensorFlow

- Overview of Deep Learning frameworks
- TensorFlow architecture and computational graph
- Tensors, operations, and sessions
- Building and training simple neural networks
- Model evaluation and optimization

Module 2: Convolutional Neural Networks (CNNs)

- Introduction to CNNs and applications
- Convolution, pooling, and fully connected layers
- Popular architectures (LeNet, AlexNet, VGG, ResNet)
- Image classification and object detection
- Implementation using TensorFlow/Keras
- Model evaluation and optimization

Module 3: Recurrent Neural Networks (RNNs)

- Introduction to RNNs and sequence modeling
- Long Short-Term Memory (LSTM) networks
- Gated Recurrent Units (GRUs)
- Text generation, sentiment analysis, and time series prediction
- Training challenges: vanishing and exploding gradients
- Implementation and evaluation using TensorFlow

Module 4: Restricted Boltzmann Machines (RBMs)

- Introduction to energy-based models
- Structure and working of RBMs
- Training RBMs using contrastive divergence
- Feature learning and dimensionality reduction
- Applications: collaborative filtering, unsupervised learning
- Limitations and advantages

Module 5: Autoencoders

- Basic autoencoders and reconstruction tasks
- Denoising autoencoders
- Variational autoencoders (VAEs)
- Applications: anomaly detection, data compression, dimensionality reduction
- Implementation in TensorFlow
- Model evaluation and optimization

Course Outcomes (COs)

CO	Course Outcome
CO1	Explain the fundamentals of TensorFlow and implement simple neural networks.
CO2	Design and implement convolutional neural networks for image-based tasks.
CO3	Develop recurrent neural network models for sequence data and time series.
CO4	Understand and implement Restricted Boltzmann Machines for unsupervised learning.
CO5	Apply autoencoders for dimensionality reduction, anomaly detection, and data reconstruction.

CO / PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	H	M	M	L	L	-	-	-	-	-
CO2	H	H	M	M	L	-	-	-	-	-
CO3	H	M	H	M	L	-	-	-	-	-
CO4	M	M	M	L	L	-	-	-	-	-
CO5	H	M	M	M	M	-	-	-	-	-

BCO 126A	Statistics 101, Spark Mllib, Machine Learning- Dimensionality Reduction	2-0-0 [2]

OBJECTIVES:

1. Understand the fundamentals of dimensionality reduction and its importance in simplifying high-dimensional datasets while retaining essential information.
2. Differentiate between various dimensionality reduction techniques, including Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA), and identify suitable applications for each.
3. Develop the ability to implement PCA and EFA using R for real-world datasets, including survey data, to extract meaningful patterns and relationships.
4. Analyze and interpret the results of dimensionality reduction techniques, such as factor loadings, explained variance, and principal components, to make informed data-driven decisions.
5. Apply dimensionality reduction to practical problems, including feature selection, data visualization, and grouping of correlated variables, to enhance subsequent machine learning or statistical modeling.

Unit 1: Introduction to Dimensionality Reduction

- Overview of dimensionality reduction and its significance in machine learning
- Curse of dimensionality and challenges with high-dimensional data
- Types of dimensionality reduction: feature selection vs. feature extraction
- Applications in data analysis, visualization, and preprocessing

Unit 2: Principal Component Analysis (PCA) – Theory

- Concept of variance, covariance, and correlation
- Eigenvalues and eigenvectors in PCA
- Principal components and explained variance
- Choosing the number of components

Unit 3: Principal Component Analysis (PCA) – Implementation in R

- Hands-on coding of PCA using R
- Data preprocessing for PCA (standardization, normalization)
- Interpreting PCA results: loadings, scores, and scree plots
- Visualization of high-dimensional data in reduced dimensions

Unit 4: Exploratory Factor Analysis (EFA)

- Introduction to EFA and its objectives
- Differences between PCA and EFA
- Factor extraction methods (Principal Axis Factoring, Maximum Likelihood)
- Determining the number of factors (Kaiser criterion, scree plot, parallel analysis)

Unit 5: Applications and Case Studies

- Grouping and summarizing variables in survey datasets
- Dimensionality reduction for feature selection in predictive modeling
- Hands-on practice with PCA and EFA on survey and real-world datasets
- Interpretation of results and reporting insights

Course Outcomes

- **CO1:** Explain the fundamentals and significance of dimensionality reduction in machine learning.
- **CO2:** Differentiate between feature selection and feature extraction techniques, including PCA and EFA.
- **CO3:** Apply Principal Component Analysis (PCA) on datasets using R to extract meaningful patterns.
- **CO4:** Implement Exploratory Factor Analysis (EFA) in R and interpret factor loadings and structures.
- **CO5:** Analyze and visualize high-dimensional data, and apply dimensionality reduction techniques to real-world problems.

CO-PO mapping

CO / PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	H	M	M	L	L	-	-	-	-	-
CO2	H	H	M	M	L	-	-	-	-	-
CO3	H	M	H	M	L	-	-	-	-	-
CO4	M	M	M	L	L	-	-	-	-	-
CO5	H	M	M	M	M	-	-	-	-	-

Semester – V

BCO 127A	R for Data Science	2-0-0 [2]

Module 1 - R basics

- Math, Variables, and Strings
- Vectors and Factors
- Vector operations

Module 2 - Data structures in R

- Arrays & Matrices
- Lists
- Dataframes

Module 3 - R programming fundamentals

- Conditions and loops
- Functions in R
- Objects and Classes
- Debugging

Module 4 - Working with data in R

- Reading CSV and Excel Files
- Reading text files
- Writing and saving data objects to file in R

Module 5 - Strings and Dates in R

- String operations in R
- Regular Expressions
- Dates in R

Course Outcome:

 CO1: Understand and apply the fundamentals of R, including math operations, variables, strings, vectors, and factors.

 CO2: Develop proficiency in handling R data structures such as arrays, matrices, lists, and dataframes for efficient data management.

 CO3: Write and debug R programs using conditions, loops, functions, and object-oriented programming concepts.

 CO4: Import, export, and manipulate data in R, including CSV, Excel, and text files.

 CO5: Perform advanced string operations and work with dates, including regular expressions for text processing.

CO-PO mapping:

CO / Outcome	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
CO1	H	M	L	L	M	-	-	-	-	-	-	M	H	M	L
CO2	H	H	M	M	H	-	-	-	-	-	-	M	H	H	M
CO3	H	H	H	M	H	-	-	-	-	-	-	M	H	H	M
CO4	H	M	M	M	H	-	-	-	-	-	-	M	H	H	M
CO5	H	M	M	L	H	-	-	-	-	-	-	M	H	H	M

BCO 128A	DIGITAL ANALYTICS AND REGRESSION	1-0-0 [1]

Module 1: Case Study 'CEO vs. CMO'

Learning Objectives, Case Study Assignment: 'CEO vs. CMO', Formulating the Business Objective, A Data Driven Approach, CEO vs. CMO Case Study Summary (5:41)

Module 2: Importing Google Trends data in R

Learning Objectives, What is R, Lab - Exercise - Import Google Trends Data

Module 3: Plotting & Correlation

Learning Objectives, Lab - Exercise - Basic Plotting in R, Lab Exercise - Correlation in R

Module 4: Simple Linear Regression in R

Learning Objectives, Box Plots & Histograms in R, Scatter Plots & Lines of Best Fit in R, Simple Linear Regression in R, Lab Exercise - Linear Regression in R

Module 5: Presenting Data Findings in Business

Learning Objectives, Using data to answer a business question, Summarizing the Data Analytics Process, Presenting Data Insights, Presenting Data Insights (5:48)

Course Outcome:

 CO1: Understand the role of data analytics in business decision-making and interpret case studies like 'CEO vs. CMO'.

 CO2: Import, clean, and manage real-world datasets, such as Google Trends data, using R.

 CO3: Perform data visualization and correlation analysis in R to extract meaningful insights.

 CO4: Apply simple linear regression techniques in R to model relationships between variables and make predictions.

 CO5: Summarize and present data findings effectively to support business decisions.

CO / Outcome	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	H	H	M	M	M	M	L	L	M	H	-	M	H	M	L
CO2	H	H	M	M	H	L	L	L	M	M	-	M	H	H	M
CO3	H	H	H	M	H	L	L	L	M	H	-	M	H	H	M
CO4	H	H	H	H	H	L	L	L	M	H	-	M	H	H	M
CO5	H	M	H	M	M	M	L	L	H	H	-	M	H	H	M

BCO 129A	DATA SCIENCE TOOLS, DATA SCIENCE FOR SCALA, MACHINE LEARNING WITH APACHE SYSTEMML	3-0-0 [0]

Module 1 - Data Scientist's Toolkit

Languages of Data Science, Introduction to Python, Introduction to R Language, Introduction to SQL, Other Languages, Categories of Data Science Tools, Open Source Tools for Data Science - Part 1, Open Source Tools for Data Science - Part 2, Commercial Tools for Data Science, Cloud Based Tools for Data Science, Practice Quiz – Tools, Libraries for Data Science, Application Programming Interfaces (API), Data Sets - Powering Data Science, Sharing Enterprise Data – Data Asset eXchange, Machine Learning Models, The Model Asset Exchange, Practice Quiz - Packages, APIs, Data Sets, Models, Lab: Explore Data Sets and Models

Module 2 - Open Source Tools

Overview of Git/GitHub, GitHub - Part 1, GitHub - Part 2 (Optional), GitHub - Part 3 (Optional), Lab: Getting Started with GitHub, Practice Quiz – GitHub, Getting Started with Jupyter Notebooks, Getting Started with JupyterLab, Jupyter Architecture,

Lab: Jupyter Notebooks - The Basics

Lab: Jupyter Basics - on Cloud

Lab: Jupyter Notebooks - Advanced Features 1

Lab: Jupyter Notebooks - Advanced Features 2

What is RStudio IDE?, Installing Packages and Loading Libraries in RStudio IDE, Plotting within RStudio IDE,

Lab: RStudio – The Basics

Lab: RStudio Basics on Cloud

Lab: Creating an Interactive Map in R

Module 3 - IBM Tools for Data Science

What is IBM Watson Studio?, Watson Studio Introduction, Creating an Account on IBM Watson Studio, Jupyter Notebooks in Watson Studio - Part 1, Jupyter Notebooks in Watson Studio - Part 2, Lab: Creating a Watson Studio Project with Jupyter Notebooks, Linking GitHub to Watson Studio, Practice Quiz - Watson Studio, IBM Watson Knowledge Catalog, Data Refinery, SPSS Modeler Flows in Watson Studio, Lab: Modeler Flows in Watson Studio (1hr),

IBM SPSS Modeler, IBM SPSS Statistics, Model Deployment with Watson Machine Learning, Auto AI in Watson Studio, IBM Watson OpenScale

Data Science with Scala

Module 1: Basic Statistics and Data Types

Vectors and Labelled Points (7:06), Local and Distributed Matrices (8:47), Summary Statistics, Correlations, and Random Data (5:30), Sampling (6:49), Hypothesis Testing

Module 2: Preparing Data

Statistics, Random data and Sampling on Data Frames (12:16), Handling Missing Data and Imputing Values (11:07), Transformers and Estimators (8:39), Data Normalization (5:00), Identifying Outliers

Module 3: Feature Engineering

Feature Vectors (3:28), Categorical Features (6:53), Using Explode, User Defined Functions, and Pivot (4:18), Principal Component Analysis (PCA) in Feature Engineering, RFormulas

Module 4: Fitting a Model

Decision Trees (8:45), Random Forests (7:15), Gradient-Boosting Trees (7:10), Linear Methods (6:16), Evaluation (11:04)

Module 5: Pipeline and Grid Search

Predicting Grant Applications: Introduction (4:27), Predicting Grant Applications: Creating Features (4:52), Predicting Grant Applications: Building a Pipeline (4:16), Prediciting Grant Applications: Cross Validation and Model Tuning (3:47), Predicting Grant Applications: Wrapping up

Machine learning with Apache SystemML

COURSE SYLLABUS

- **Module 1 - What is SystemML?**

1. Explain the purpose and the origin of SystemML
2. List the alternatives to SystemML
3. Compare performances of SystemML with the alternatives

- **Module 2 - SystemML and the Spark MLContext**

1. Use MLContext to interact with SystemML (in Scala)

- **Module 3 - Working with BigSheets**

1. Describe and use a number of SystemML algorithms

- **Module 4 - Working with BigSheets**

1. Explain the purpose of DML
2. Describe the DML language
3. List some of the built-in functions

- **Module 5 - Working with BigSheets**

1. Describing the optimizer stack
2. Explaining how SystemML know it's better to run on one machine
3. Explaining why SystemML is so much faster than single-node R

Course Outcome:

 CO1: Understand the data science toolkit, including programming languages (Python, R, SQL), open-source and commercial tools, and cloud platforms for data analysis.

 CO2: Apply open-source tools such as Jupyter Notebooks, RStudio, and GitHub to perform data manipulation, visualization, and collaborative projects.

 CO3: Utilize IBM Watson Studio tools, including AutoAI, SPSS Modeler, and Watson OpenScale, for building, deploying, and managing machine learning models.

 CO4: Perform statistical analysis, data preparation, feature engineering, and model fitting using Scala, including decision trees, random forests, gradient boosting, and linear methods.

 CO5: Understand and apply Apache SystemML for large-scale machine learning, including MLContext, DML language, built-in algorithms, and performance optimization.

CO-PO mapping:

CO / Outcome	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	H	M	M	L	H	L	L	L	M	M	-	M	H	H	M
CO2	H	H	H	M	H	L	L	L	H	H	-	M	H	H	M
CO3	H	H	H	H	H	L	L	L	M	H	-	M	H	H	M
CO4	H	H	H	H	H	L	L	L	M	H	-	M	H	H	M
CO5	H	H	H	H	H	L	L	L	M	H	-	M	H	H	M

BCO 130A	DATA VISUALIZATION WITH R ,DATA ANALYSIS WITH PYTHON	3-0-0 [0]

Data Visualization with R

COURSE SYLLABUS

Module 1 - Basic Visualization Tools

- Bar Charts
- Histograms
- Pie Charts

Module 2 - Basic Visualization Tools Continued

- Scatter Plots
- Line Plots and Regression

Module 3 - Specialized Visualization Tools

- Word Clouds
- Radar Charts
- Waffle Charts
- Box Plots

Module 4 - How to create Maps

- Creating Maps in R

Module 5 - How to build interactive web pages

- Introduction to Shiny
- Creating and Customizing Shiny Apps
- Additional Shiny Features

Data Analysis with Python

Module 1 - Introduction Learning Objectives Understanding the Domain Understanding the Dataset Python package for data science Importing and Exporting Data in Python Basic Insights from Datasets

Module 2 - Data Wrangling Identify and Handle Missing Values Data Formatting Data NormalizationSets Binning Indicator variables

Module 3 - Exploratory Data Analysis Descriptive Statistics Basic of Grouping ANOVA Correlation Correlation 2

Module 4 - Model Development Simple and Multiple Linear Regression Model Evaluation using Visualization Polynomial Regression and Pipelines R-squared and MSE for In-Sample Evaluation Prediction and Decision Making Module 5 - Working with Data in Python Model Evaluation Over Fitting, Under fitting and Model Selection Ridge Regression Grid Search Model Refinement

Course Outcome:

 CO1: Understand and apply basic data visualization techniques in R, including bar charts, histograms, pie charts, scatter plots, and line plots.

 CO2: Utilize specialized visualization tools in R such as word clouds, radar charts, waffle charts, and box plots to extract meaningful insights.

 CO3: Develop interactive data visualizations and dashboards using Shiny in R for effective communication of data findings.

 CO4: Apply Python for data analysis, including data wrangling, exploratory data analysis, and handling missing or inconsistent data.

 CO5: Build and evaluate predictive models in Python, including linear and polynomial regression, model selection, and evaluation metrics.

CO-PO mapping:

CO / Outcome	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO11	PO12	PSO 1	PSO 2	PSO 3
CO1	H	M	M	L	H	L	L	L	M	H	-	M	H	H	M
CO2	H	H	H	M	H	L	L	L	M	H	-	M	H	H	M
CO3	H	M	H	M	H	L	L	L	H	H	-	M	H	H	M
CO4	H	H	H	H	H	L	L	L	M	H	-	M	H	H	M
CO5	H	H	H	H	H	L	L	L	M	H	-	M	H	H	M

Semester – VI

BCO131A	**Machine Learning with R**	**1-0-0 [1]**

Module 1 - Machine Learning vs Statistical Modeling

Learning Objectives, Introduction to Machine Learning (1:46), Learning vs Statistical Modeling (4:05), Supervised VS Unsupervised Learning (5:02), Supervised Learning - Classification (2:37), Unsupervised Learning (1:41), Lab

Module 2 - Supervised Learning I

Learning Objectives, K-Nearest Neighbors (7:17), Decision Trees (5:18), Random Forests (3:46), Reliability of Random Forests (3:34), Advantages & Disadvantages of Decision Trees (1:25)

Module 3 - Supervised Learning II

Learning Objectives, Regression Algorithms (3:48), Model Evaluation (5:10), Model Evaluation Overfitting & Underfitting (2:19), Understanding Different Evaluation Models (2:38)

Module 4 - Unsupervised Learning

Learning Objectives, K-Means Clustering plus Advantages & Disadvantages (5:06), Hierarchical Clustering plus Advantages & Disadvantages (5:59), Measuring the Distances Between Clusters - Single Linkage Clustering (2:13), Measuring the Distances Between Clusters - Algorithms for Hierarchical Clustering (4:16), Density Based Clustering (3:44)

Module 5 - Dimensionality Reduction & Collaborative Filtering

Learning Objectives, Dimensionality Reduction - Feature Extraction & Selection (5:30), Collaborative Filtering & Its Challenges (5:01), Lab

Course outcomes:

 CO1: Understand the differences between machine learning and statistical modeling, and distinguish between supervised and unsupervised learning approaches.

 CO2: Apply supervised learning algorithms, including K-Nearest Neighbors, Decision Trees, and Random Forests, for classification and regression tasks.

 CO3: Evaluate machine learning models using performance metrics, and understand issues like overfitting and underfitting.

 CO4: Implement unsupervised learning techniques such as K-Means, Hierarchical, and Density-Based clustering for data segmentation.

 CO5: Apply dimensionality reduction techniques and collaborative filtering methods to improve model efficiency and handle high-dimensional data.

CO-PO mapping:

CO / Outcome	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
CO1	H	H	M	L	M	L	L	L	M	H	-	M	H	M	L
CO2	H	H	H	M	H	L	L	L	M	H	-	M	H	H	M
CO3	H	H	H	H	H	L	L	L	M	H	-	M	H	H	M
CO4	H	H	H	H	H	L	L	L	M	H	-	M	H	H	M
CO5	H	H	H	H	H	L	L	L	M	H	-	M	H	H	M

BCO 132A	DATA VISUALIZATION WITH PYTHON	1:0:0 [1]

Data Visualization with Python

COURSE SYLLABUS

Module 1 - Introduction to Visualization Tools

- Introduction to Data Visualization
- Introduction to Matplotlib
- Basic Plotting with Matplotlib
- Dataset on Immigration to Canada
- Line Plots

Module 2 - Basic Visualization Tools

- Area Plots
- Histograms
- Bar Charts

Module 3 - Specialized Visualization Tools

- Pie Charts
- Box Plots
- Scatter Plots
- Bubble Plots

Module 4 - Advanced Visualization Tools

- Waffle Charts
- Word Clouds
- Seaborn and Regression Plots

Module 5 - Creating Maps and Visualizing Geospatial Data

- Introduction to Folium
- Maps with Markers
- Choropleth Maps

Course outcome:

 CO1: Understand the fundamentals of data visualization and apply basic plotting techniques using Matplotlib.

 CO2: Create standard visualizations such as area plots, histograms, and bar charts to analyze datasets effectively.

 CO3: Apply specialized visualization techniques, including pie charts, box plots, scatter plots, and bubble plots, for exploratory data analysis.

 CO4: Utilize advanced visualization tools such as Waffle Charts, Word Clouds, and Seaborn regression plots to derive insights.

 CO5: Create maps and visualize geospatial data using Folium, including markers and choropleth maps

CO-PO mapping

CO / Outcome	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO11	PO12	PSO 1	PSO 2	PSO 3
CO1	H	M	M	L	H	L	L	L	M	H	-	M	H	H	M
CO2	H	H	H	M	H	L	L	L	M	H	-	M	H	H	M
CO3	H	H	H	M	H	L	L	L	M	H	-	M	H	H	M
CO4	H	H	H	M	H	L	L	L	M	H	-	M	H	H	M
CO5	H	H	H	M	H	L	L	L	M	H	-	M	H	H	M

BCO 133A	MACHINE LEARNING WITH PYTHON	1:0:0 [1]

Machine Learning with Python

COURSE SYLLABUS

Module 1 - Intro to Machine Learning

- Applications of Machine Learning
- Supervised vs Unsupervised Learning
- Python libraries suitable for Machine Learning

Module 2 - Regression

- Linear Regression
- Non-linear Regression
- Model evaluation methods

Module 3 - Classification

- K-Nearest Neighbour
- Decision Trees
- Logistic Regression
- Support Vector Machines
- Model Evaluation

Module 4 - Unsupervised Learning

- K-Means Clustering
- Hierarchical Clustering
- Density-Based Clustering

Module 5 - Recommender Systems

- Content-based recommender systems
- Collaborative Filtering

Course Outcomes:

 CO1: Understand the fundamentals of machine learning, including its applications, types (supervised vs. unsupervised), and relevant Python libraries.

 CO2: Apply regression techniques, both linear and non-linear, to model and predict continuous outcomes, and evaluate model performance.

 CO3: Implement classification algorithms such as K-Nearest Neighbors, Decision Trees, Logistic Regression, and Support Vector Machines, and assess model accuracy.

 CO4: Apply unsupervised learning methods including K-Means, Hierarchical, and Density-Based clustering to explore and segment datasets.

 CO5: Develop recommender systems using content-based and collaborative filtering approaches to provide personalized recommendations.

CO-PO mapping

CO / Outcome	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
CO1	H	M	M	L	H	L	L	L	M	H	-	M	H	M	L
CO2	H	H	H	M	H	L	L	L	M	H	-	M	H	H	M
CO3	H	H	H	M	H	L	L	L	M	H	-	M	H	H	M
CO4	H	H	H	H	H	L	L	L	M	H	-	M	H	H	M
CO5	H	H	H	H	H	L	L	L	M	H	-	M	H	H	M