Bharatiya Vidya Bhavan's

SARDAR PATEL COLLEGE OF ENGINEERING

Government Aided Autonomous Institute under Mumbai University

Andheri (W), Mumbai - 400058

COURSE CONTENTS

Minor in AI-ML and Data Science

Regulation 23

Programming and Data Science

Course Code	**Course Name**
MI-BT051 Sem IV	Programming and Data Science
Course pre-requisites	

Course Objectives

1. Learn Python programming from basics.
2. Use Python tools to work with simple datasets.
3. Understand basic statistics used in data analysis.

Course Outcomes

Upon successful completion of the course, students should be able to

1. Write a code for simple programming tasks with a data set.
2. Use python Libraries like Numpy, Pandas for Exploratory Data Analysis.
3. Appreciate role of statistics & probability in Data Science.
4. Visualize the data using tools such as Matplotlib and Seaborn.

Course Content

Module No	***Details***	***Hrs***
1	**Python Programming Basics** • Variables, data types (integers, strings, lists) • Operators, expressions, and basic input/output • Conditional statements (if-else), loops (for, while)	**04**
2	**Functions and File Handling** • Defining and calling functions • Parameters, return values, and scope • Reading from and writing to text and CSV files	**04**
3	**Introduction to packages like NumPy, SciPy** • Creating arrays, indexing, and slicing • Basic array operations and mathematical functions • Array broadcasting and reshaping • EDA (Exploratory Data Analysis),	**04**
4	**Data Handling with Pandas** • Creating and using Series and DataFrames • Reading data from CSV, Excel, and JSON files	**04**

	• Filtering, sorting, merging, and grouping data • Understand the dataset quality and integrity using EDA.	
5	**Basic Statistics and Probability** • Measures of central tendency (mean, median, mode) • Measures of dispersion (range, variance, standard deviation) • Basics of probability and normal distribution	**04**
6	**Data Cleaning and Preprocessing** • Identifying and handling missing data • Encoding categorical variables • Scaling and normalizing features • Principle Component Analysis for reducing data dimensions	**04**
7	**Data Visualization** • Introduction to Matplotlib and Seaborn • Creating line plots, bar charts, histograms, and scatter plots • Customizing plots with labels, legends, and titles	**04**

Text Books
• *Python for Data Analysis* by Wes McKinney • *Think Python* by Allen B. Downey • *Head First Statistics* by Dawn Griffiths
References
• NPTEL: "Python for Data Science" by Prof. Ragunathan Rengasamy, IIT Madras • Google Colab tutorials, Kaggle beginner Python exercises

Statistical Learning

Course Code	Course Name
MI-BT052 Sem V	Statistical Learning
Course pre-requisites	

Course Objectives

1. Build the mathematical, statistical, and optimization foundation for machine learning.
2. Demonstrate feature engineering, data preprocessing, and cross-validation techniques.
3. Develop supervised and unsupervised models using standard Python libraries.
4. Design end-to-end predictive pipelines for real-world case studies.

Course Outcomes

Upon successful completion of the course, students should be able to

1. Understand and apply linear algebra and probability concepts to solve data problems.
2. Process raw data and select features to improve model performance.
3. Evaluate classification and clustering models using standard performance metrics.
4. Deploy time-series forecasting models using structured machine learning pipelines.

Course Content

Module No	*Details*	*Hrs*
1	**Mathematical background** • Basics of Linear Algebra (orthogonality, independence etc.) • Basic terminology in Optimization (cost function etc) • Understanding probability and statistics. (various distributions, inferential statistics)	**08**
2	**Feature Engineering & Cross Validation** • Selecting and extracting features • Handling missing and categorical data • Normalizing and scaling features • Data Labeling & Bias identification • *k*-fold cross-validation • Bias-Variance Trade-Off for *k*-fold	**08**
3	**Regression & Classification Models I** • Simple and multiple linear regression • Logistic regression for classification • Decision trees for classification and regression • Using scikit-learn to build models • Model accuracy evaluation, P-value, R^2 Confusion matrix etc	**08**

4	**Regression & Classification Models II** • Bagging, Random Forest, Boosting • Basics of Support Vector Machines (SVM) • Naive Bayes, K-Nearest Neighbors • Model accuracy evaluation Accuracy, precision, recall, F1-score, Confusion matrix, ROC curve, AUC	**06**
5	**Unsupervised learning** • K-Means and Hierarchical Clustering • DBSCAN for density-based clustering • Principle Component Analysis for reducing data dimensions • Model accuracy evaluation	**06**
6	**Time-Series Forecasting** • Understanding trends and seasonal data • ARIMA and Facebook Prophet for forecasting Example: weather, load, or demand prediction in infrastructure	**03**
7	**Case Study** using open datasets **(e.g. Kaggle)** • Steps of an ML pipeline (preprocessing to prediction) • Case studies using open datasets • Python project with train/test split and evaluation	**03**

Text Books
• An Introduction to Statistical Learning by Gareth James et al. • A First Course in Machine Learning, S. Rogers and M. Girolami, CRC Press, 2nd Ed, (2015) • Linear Algebra & its Applications by Gilbert Strang • https://www.youtube.com/playlist?list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab • **https://youtu.be/LvySJGj-88U?si=zA6V6b8vY6QLmC0H**
References
• NPTEL: "Machine Learning" by Prof. P. Abinash (IIT Kharagpur) • scikit-learn official documentation • Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow by Aurélien Géron

AI-ML & Deep Learning

Course Code	Course Name
MI-BT053 Sem VI	AI-ML & Deep Learning
Course pre-requisites	

Course Objectives

1. Introduce to Artificial Intelligence & neural network.
2. Develop a foundational understanding of deep learning methodologies.
3. Discuss applications of AI & Deep Learning.

Course Outcomes

Upon successful completion of the course, students should be able to

1. Explain search algorithms in AI and choose appropriate one for given problem.
2. Build and train a basic neural network as well as models like CNN, RNN, Deep learning for various applications.
3. Design AI models for practical mini-projects.

Course Content

Module No	*Details*	*Hrs*
1	**Introduction to Artificial Intelligence (AI)** What is AI? Agents & Environment, Problem solving Agent, Uninformed Search Methods: Breadth First Search (BFS), Depth First Search (DFS), Informed Search Methods: Greedy best first Search A* Search, Memory bounded heuristic Search.	**06**
2	**Local Search Algorithms and Optimization Problems**: Hill- climbing search, Simulated annealing, Local beam search, Genetic algorithms. Adversarial Search: Games, Optimal strategies, The minimax algorithm, Alpha-Beta Pruning.	**08**
3	**Basics of Neural Network** • simple concepts: neurons, weights, layers • Need of neural network and data requirements for training the models. • Single layer Neural Networks, Multilayer Neural Networks, • Training methods - Back propagation	**08**
4	**Convolution Neural Networks & Recurrent Neural networks**	**08**

	• Convolution layers, Architecture of CNN, Data Augmentation, Application of CNN for Image data • RNN for sequence data, Application of RNN in Time-series forecasting	
5	**Deep Learning** • Basic terminology such as Encoders, transformers etc. • Basics of Long term Short term Memory (LSTM)	**06**
6	• **Tools and Practice** • Using TensorFlow, Keras or PyTorch easily with examples • Mini-projects with real data	**06**

Text Books
• *Deep Learning* by Ian Goodfellow, Yoshua Bengio, and Aaron Courville • *Artificial Intelligence: A Modern Approach* by Stuart Russell & Peter Norvig • https://youtu.be/II4giR4vOOo?si=l2-EgJp0fLJFhgzl
References
• *Neural Networks and Deep Learning* by Michael Nielsen • PyTorch & TensorFlow documentation • NPTEL: "Deep Learning" by Prof. Balakrishnan (IIT Madras)

Advanced Data Science and Applications

Course Code	**Course Name**
MI-BT054 SemVII	Advanced Data Science and Applications
Course pre-requisites	

Course Objectives

1. Understand Big Data ecosystems and practical handling challenges.
2. Learn the principles of recommender systems and basic natural language processing workflows for text analysis.
3. Explore model deployment techniques using web frameworks and containerization tools for practical usability.
4. Study machine learning ethics, explainability, and industry-driven concepts like Physics-Informed AI.

Course Outcomes

Upon successful completion of the course, students should be able to

1. Apply Hadoop, Spark, and PySpark utilities to process large-scale files and real-time datasets.
2. Implement collaborative filtering algorithms and NLP preprocessing pipelines for classification tasks.
3. Deploy trained machine learning models into web applications using Flask, Streamlit, and Docker.
4. Design an integrated project incorporating fairness principles, open datasets, and industry-relevant AI designs.

Course Content

Module No	***Details***	***Hrs***
1	**Big Data Overview** • What is big data? Why is it important? • Basics of Hadoop and Spark • Using PySpark to handle simple large files (CSV/logs) • Real time data Applications -parallel data processing, speed, latency, • Practical challenges in handling Big Data	08
2	**Recommender Systems** • What are recommendation systems? • Collaborative vs. content-based filtering • Simple example: tool or material selection recommender	08
3	**Natural Language Processing (NLP)** • What is text data? Simple preprocessing (cleaning, tokenizing) • Text classification: feedback analysis or issue sorting	08

	• Tools: NLTK, basic Hugging Face models	
4	**Model Deployment** **(to be discussed with an example such as Anomaly Detection)** • Making ML models usable by others • Flask/Streamlit for web-based apps • Packaging with GitHub and Docker (simple intro)	08
5	**Project, Ethics and Explainability** • Small project integrating learning from all modules • Concepts of fairness, bias, and transparency in ML	04
6	**Case Studies using open Data set (e.g. Kaggle)** Introduction to ML and NN designs that are influenced by science and have applications in the industrial world like PI-AI (Physics Informed AI)	06

Text Books
• *Data Science from Scratch* by Joel Grus • *Practical Statistics for Data Scientists* by Peter Bruce
References
• *Forecasting: Principles and Practice* by Rob J. Hyndman (free online) • NPTEL: "Big Data Computing" – IIT Patna • Kaggle and UCI ML Repositories (for projects) • https://www.sciencedirect.com/science/article/abs/pii/S0021999118307125

Suggested Tools and Platforms

- **Languages**: Python (primary), R (optional for stats)
- **IDEs**: Jupyter Notebook, VS Code, Google Colab
- **Libraries**: Pandas, Scikit-Learn, TensorFlow, Keras, PyTorch, Matplotlib, Seaborn, NLTK, SpaCy, XGBoost
- **Datasets**: UCI, Kaggle, OpenML, real-world case data (CSV, JSON)