Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati

Department of Artificial Intelligence and Data Science

S.Y. B. Tech Syllabus 2025-26 (As per NEP 2020)

Syllabus: Double Minor w. e. f. AY: 2025-2026
SEMESTER-III

Double Minor in Artificial Intelligence and Data Science

SEM	Course Code	Courses Name	Teaching Scheme			Examination Scheme and Marks								Credits		
			TH	PR	TUT	Activity	ISE	ESE	TW	PR	OR	Total	TH	PR	TUT	Total
III	AI24261	Artificial Intelligence and Data Science	2	2	-	10	20	50	20	20		120	2	1		3

Dept. Autonomy Coordinator
Mrs. R. S. Naik

Dept. Academic Coordinator
Mr. P. N. Shendage

HOD, AI&DS
Dr. C. S. Kulkarni

Dean Autonomy
Dr. C. B. Nayak

Dean Academics
Dr. S. M. Bhosle

Principal
Dr. S. B. Lande

VPKBIET BARAMATI
Affiliated to S. P. P. U. Pune
Id No./PU/PN/Engg./152/2008
PLACE (MS.)

Principal
Vidya Pratishthan's
Kamalnayan Bajaj Institute of
Engineering & Technology, Baramati
Vidyanagari, Baramati-413102

BUCKET OF DOUBLE MINOR DEGREE

DOUBLE MINOR DEGREE (only for students having CGPA >= 7.5)
AI24261: Artificial Intelligence and Data Science

Vidya Pratishthan's
Kamalnayan Bajaj Institute of Engineering and Technology, Baramati
(Autonomous Institute)

AI24261 - Artificial Intelligence and Data Science

Teaching Scheme: **Theory: - 2 Hours/Week** **Practical: 2 Hour/Week**	**Credits** **03**	**Examination Scheme:** **Activity: - 10 Marks** **In-Sem:- 20 Marks** **End-Sem:- 50 Marks** **Term Work: 20 Marks** **Practical: 20 Marks**

Prerequisites: Basic python programming

Course Objectives:

- To provide a comprehensive introduction to the concepts and applications of Artificial Intelligence (AI) and Data Science.
- To equip students with the foundational knowledge of machine learning algorithms and data handling techniques.
- To develop practical skills in using AI and Data Science tools and libraries.
- To foster the ability to apply AI and Data Science methods to solve real-world engineering problems.

Course Outcomes (COs): The students will be able to:

CO1: Understand the fundamental principles and applications of AI and Data Science.
CO2: Collect, preprocess, and visualize data effectively.
CO3: Implement basic machine learning algorithms for regression, classification, and clustering tasks.
CO4: Apply neural networks for simple AI tasks and understand their structure and function.

Course Contents

Unit I: Introduction to Artificial Intelligence and Data Science (06 Hours)

Basics of AI and Data Science: Definition and history of AI, Key concepts and applications of AI, Introduction to Data Science and its importance.

Fundamentals of Machine Learning: Types of machine learning: supervised, unsupervised, and reinforcement learning, Basic concepts: features, labels, training, and testing data, Overview of common algorithms: linear regression, classification, clustering.

Tools and Technologies: Introduction to Python for AI and Data Science, Overview of popular libraries: NumPy, pandas, matplotlib, scikit-learn, Setting up the development environment.

Unit II: Data Handling and Pre-processing (06 Hours)

Data Collection and Cleaning: Sources of data: databases, web scraping, APIs, Data cleaning techniques: handling missing values, outliers, duplicates.

Data Manipulation: Data types and structures, Basic operations with pandas: filtering, grouping, merging.

Data Visualization: Importance of data visualization, Visualization tools and techniques: matplotlib, seaborn, Creating basic plots: bar charts, histograms, scatter plots.

Unit III: Core Machine Learning Concepts (06 Hours)

Supervised Learning: Regression algorithms: linear regression, polynomial regression, Classification algorithms: logistic regression, decision trees, k-nearest neighbors, Model evaluation metrics: accuracy, precision, recall, F1-score.

Unsupervised Learning: Clustering algorithms: k-means, hierarchical clustering, Dimensionality reduction: PCA, t-SNE, Applications and use cases.

Model Training and Evaluation: Splitting data: training, validation, and test sets, Cross-validation techniques, Avoiding overfitting and underfitting.

Unit IV: Advanced Topics and Applications(06 Hours)

Introduction to Neural Networks: Basics of neural networks and deep learning, Structure of neural

networks: neurons, layers, activation functions, Overview of frameworks: TensorFlow, Keras.
AI in Real-World Applications: AI in healthcare, finance, and manufacturing, Ethical considerations and challenges in AI, Case studies of successful AI implementations.

Text Books:

1. Stuart Russell and Peter Norvig, "Artificial Intelligence – A Modern Approach", Fourth Edition, Pearson Education, 2021.
2. Jake VanderPlas, "Python Data Science Handbook", O'Reilly Media, Inc.,2016.
3. Aurélien Géron, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow", Third Edition, O'Reilly Media, Inc.,2022.

Reference Books:

1. Dan W. Patterson, "Introduction to AI and ES", Pearson Education,2007
2. Kevin Night, Elaine Rich, and Nair B., "Artificial Intelligence", McGraw Hill, 2008
3. Patrick H. Winston, "Artificial Intelligence", Third Edition, Pearson Education, 2006
4. Deepak Khemani, "Artificial Intelligence", Tata McGraw Hill Education, 2013.

E-Resources:

1. https://nptel.ac.in/courses/106/102/106102220/
2. https://nptel.ac.in/courses/106/105/106105077/
3. https://nptel.ac.in/courses/106/105/106105078/
4. https://nptel.ac.in/courses/106/105/106105079/

List of Assignments

1. Data Collection and Cleaning:
   a. Collect a dataset from an online source (e.g., Kaggle, UCI Machine Learning Repository).
   b. Identify and handle missing values in the dataset.
   c. Remove duplicates and handle outliers.
2. Data Exploration and Visualization:
   a. Load the cleaned dataset using pandas.
   b. Perform basic statistical analysis (mean, median, mode).
   c. Create visualizations using matplotlib and seaborn (e.g., histograms, bar plots, scatter plots).
3. Linear Regression:
   a. Implement linear regression from scratch using numpy.
   b. Use scikit-learn to fit a linear regression model to a given dataset.
   c. Visualize the regression line and interpret the results.
4. Classification with Logistic Regression:
   a. Implement logistic regression using scikit-learn.
   b. Evaluate the model using accuracy, precision, recall, and F1-score.
   c. Visualize the confusion matrix and ROC curve.
5. Clustering with K-Means:
   a. Implement K-Means clustering using scikit-learn.
   b. Determine the optimal number of clusters using the elbow method.

c. Visualize the clustered data and analyze the results.

6. Dimensionality Reduction with PCA:
    a. Implement PCA using scikit-learn.
    b. Visualize the explained variance ratio.
    c. Plot the data in the new reduced dimension and interpret the results.
    d. Evaluate the model's performance and interpret the results.
7. Neural Networks with Keras:
    a. Create a simple neural network for a classification problem using Keras.
    b. Train and evaluate the model.
    c. Visualize the training history (loss and accuracy plots).

Syllabus: Double Minor w. e. f. AY: 2025-2026
SEMESTER-IV

Double Minor in Artificial Intelligence and Data Science

SEM	Course Code	Courses Name	Teaching Scheme			Examination Scheme and Marks							Credits		
			TH	PR	TUT	Activity	ISE	ESE	TW	PR	OR	Total	TH	PR	TUT	Total
IV	AI24271	Data Operation and Interpretation	2	2	-	10	20	50	20	20		120	2	1		3

Dept. Autonomy Coordinator
Mrs. R. S. Naik

Dept. Academic Coordinator
Mr. P. N. Shendage

HOD, AI&DS
Dr. C. S. Kulkarni

Dean Autonomy
Dr. C. B. Nayak

Dean Academics
Dr. S. M. Bhosle

Principal
Dr. S. B. Lande

VPKBIET BARAMATI
Affiliated to S.P.P.U. Pune Id No. / PU / PN / Engg. / 152 / 2008
PUNE (MS)

Principal
Vidya Pratishthan's
Kamalnayan Bajaj Institute of
Engineering & Technology, Baramati
Vidyanagari, Baramati-413133

BUCKET OF DOUBLE MINOR DEGREE

DOUBLE MINOR DEGREE (only for students having CGPA >= 7.5)
AI24271: Data operation and Interpretation

Vidya Pratishthan's
Kamalnayan Bajaj Institute of Engineering and Technology, Baramati
(Autonomous Institute)

AI24271- Data operation and Interpretation

Teaching Scheme: **Theory: 2 Hours/Week** **Practical: 2 Hour/Week**	**Credits** **03**	**Examination Scheme:PR:20** **Activity:10 Marks** **ISE: 20 Marks** **ESE: 50 Marks** **Term Work: 20 Marks**

Prerequisites: Python Programming

Course Objectives:.

- To introduce the foundational concepts of data science.
- To learn data manipulation and preprocessing techniques.
- To develop proficiency in data visualization.
- To explore predictive data models and their evaluation.

Course Outcomes (COs): The students will be able to learn:

CO1: Perform data manipulation using Pandas.
CO2: Understand and apply data preprocessing techniques.
CO3: Create insightful visualizations using Matplotlib and Seaborn.
CO4: Implement Predictive data models and evaluate their performance.

Course Contents

Unit I: Foundations of Data Science and Pandas Basics (06 Hours)
Introduction to Data Science : Data, Definition and significance of Data Science. Overview of the Data Science lifecycle. **Pandas library:** Data Structures: Series and DataFrames. Basic Operations: Data Exporting, Data Loading and Inspection , Data Selection and Filtering,

Unit II: Data Preparation and Statistical Foundations (06 Hour)
Statistical Measures: Mean, median, mode, variance, and standard deviation. Correlation and its significance in data relationships, Skewness, Kurtosis and Outliers.
Data Preprocessing: Handling missing values, scaling, encoding and feature extraction, Data merging, reshaping, and transformation.

Unit III: Visualizing Data Trends and Relationships with Python (06 Hour)
Importance of Visualization in data science, Plot types: Line, bar, histogram, scatter.
Libraries: Matplotlib and Seaborn, Visualizing relationships and trends in data. Advanced plots using Seaborn: Heatmaps, pair plots, violin plots. correlation matrics.

Unit IV: Predictive Data Analytics (6 Hours)
Analytics Types: Predictive, Descriptive and Prescriptive ,Types: Supervised (classification, regression) and unsupervised learning. Algorithms: Simple Linear Regression, Logistic regression, Multiple linear Regression. K-Nearest neighbor. Metrics: Accuracy, precision, recall, F1-score.

Text Books:

1. **"Introduction to Machine Learning with Python: A Guide for Data Scientists"** by Andreas C. Müller and Sarah Guido (1st Edition, 2016, O'Reilly Media, ISBN-13: 978-1449369415).
2. Wes McKinney, "Python for Data Analysis",O'REILLY, ISBN:978-1-449-31979-3, 1st edition, October 2012.

3. Rachel Schutt & O'neil, "Doing Data Science", O'REILLY, ISBN:978-1-449-35865-5, 1st edition, October 2013.

4. **"Machine Learning: A Probabilistic Perspective"** by Kevin P. Murphy (1st Edition, 2012, The MIT Press, ISBN-13: 978-0262018029

Reference Books:

1. Wes McKinney - *Python for Data Analysis* Publisher: O'Reilly MediaISBN: 978-1491957660
2. Joel Grus - *Data Science from Scratch: First Principles with Python* Publisher: O'Reilly Media ISBN: 978-1492041139
3. **David Spiegelhalter** - *The Art of Statistics: How to Learn from Data* ,Publisher: Basic Books ISBN: 978-1541618510
4. **Andy Kirk** - *Data Visualization: A Handbook for Data-Driven Design* ,Publisher: SAGE ISBN: 978-1473960543
5. **Ben Jones** - *Communicating Data with Tableau* , Publisher: O'Reilly MediaISBN: 978-1449372026
6. **irag Shah** - *A Hands-On Introduction to Data Science* ,Publisher: Cambridge University Press ISBN: 978-1108472449

E-Resources:

1. **https://onlinecourses.nptel.ac.in/noc21_cs69/preview**
2. **https://www.coursera.org/learn/machine-learning**
3. **https://www.datacamp.com/courses/data-visualization-with-python**

List of Assignments

1. Find an open-source dataset (e.g., from Kaggle), and provide its description and source URL. Load the dataset into a pandas DataFrame, then check for missing values, display basic statistics using `describe()`, and provide variable descriptions and data types. verifying dimensions and ensure proper data preprocessing.
2. Analyze the Iris dataset by computing the mean and standard deviation for sepal length, sepal width, petal length, and petal width. Based on these statistics, determine the best measurement for identifying the Iris species.
3. Create various plots (line, bar, scatter, histogram) using **Matplotlib** and **Seaborn**. Perform visualize distributions, relationships, and correlations.
4. Implement a simple machine learning model (e.g., K-Nearest Neighbour) and evaluate its performance using different Metrics.
5. Demonstrate the Simple Linear Regression model and evaluate the performance on any dataset.