



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

Board of Studies: Artificial Intelligence and Data Science

**Syllabus: Double Minor (Artificial Intelligence and Data Science) 2024 Pattern
w.e.f 2025-2026
(As per NEP 2020)**

Vidya Pratishthan's

Kamalnayan Bajaj Institute of Engineering and Technology, Baramati

Board of Studies: Artificial Intelligence and Data Science

Syllabus: Double Minor (Artificial Intelligence and Data Science) 2024 Pattern w.e.f 2025-2026

Sem	Course Code	Course Name	Teaching Scheme			Examination Scheme and Marks							Credits			
			TH	PR	TUT	CAA	ISE	ESE	TW	PR	OR	Total	T H	PR	TUT	Total
III	AI24261TH	Artificial Intelligence and Data Science	2	-	-	10	-	60		-	-	70	2	-	-	3
	AI24261PR	Artificial Intelligence and Data Science	-	2	-	-	-	-	30	-	-	30	-	1	-	
IV	AI24271TH	Data Operation and Interpretation	2	-	-	10	-	60		-	-	70	2		-	3
	AI24271PR	Data Operation and Interpretation	-	2	-	-	-	-	30	-	-	30		1		
V	AI24361TH	Machine Learning: Concepts, Techniques, and Applications	3	-	-	10	30	60	-	-	-	100	3	-	-	4
	AI24361PR	Machine Learning: Concepts, Techniques, and Applications	-	2	-	-	-	-	30	-	-	30	-	1	-	
VI	AI24371TH	Deep Learning: Foundations, Architectures, and Application	3	-	-	10	30	60	-	-	-	100	3	-	-	4
	AI24371PR	Deep Learning: Foundations, Architectures, and Application	-	2	-	-	-	-	30	-	-	30	-	1	-	
VII	AI24461TH	Artificial Intelligence and Data Science Advanced Applications and Trends	3	-	-	10	30	60	-	-	-	100	3	-	-	4
	AI24461PR	Artificial Intelligence and Data Science Advanced Applications and Trends	-	2	-	-	-	-	30	-	-	30	-	1	-	
Total			13	10	-	50	90	300	150	-	-	590	13	05	-	18

Mr.P.N. Shendage
Academic Coordinator

Dr.C.S. Kulkarni
Head of Department

Dr.S.M. Bhosale
Dean Academics

Dr.A.M. Kolekar
Controller of Examination



Dr.S.B. Lande
Principal



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 Hours/Week

Credits

03

Examination Scheme:

CAA: - 10 Marks

ESE: - 60 Marks

TW: 30 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. Python Basics:
 - a. Write a Python program to perform basic arithmetic operations.
 - b. Create a program to manipulate strings and lists.
 - c. Implement a simple calculator using functions.
2. 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.
3. 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).
4. 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.
5. 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).



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 Hours/Week

Credits

03

Examination Scheme:

CAA:10 Marks

ESE: 60 Marks

TW: 30 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 matrices.

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 Media ISBN: 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 Media ISBN: 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.



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

AI24361- Machine Learning: Concepts, Techniques, and Applications

Teaching Scheme:
Theory: 3 Hours/Week
Practical: 2 Hours/Week

Credits
04

Examination Scheme:
CAA:10 Marks
ISE: 30 Marks
ESE: 60 Marks
TW: 30 Marks

Prerequisites: Python Programming

Course Objectives:

- To understand and analyze various machine learning models, their principles, and real-world applications
- To design and implement machine learning models that predict outcomes using linear, non-linear, and logistic regression techniques.
- To develop and evaluate classification models for solving binary and multiclass classification problems.
- To apply and evaluate clustering algorithms for effective segmentation and analysis of unlabeled data.
- To explore ensemble learning techniques for enhancing the performance of machine learning models.
- To understand the fundamentals of reinforcement learning and apply its algorithms to real-world applications.

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

CO1: Explain and differentiate various machine learning models based on their characteristics and applications

CO2: Develop and implement predictive models using linear, non-linear, and logistic regression techniques.

CO3: Apply and evaluate classification techniques for solving binary and multiclass classification problems.

CO4: Implement and analyze clustering techniques for handling and segmenting unlabeled datasets.

CO5: Utilize ensemble learning methods to enhance model performance through the combination of multiple machine learning algorithms.

CO6: Apply reinforcement learning techniques to solve real-world decision-making problems.

Course Contents

Unit I Introduction to Machine Learning(06 Hour)

Introduction: Machine Learning, Definitions and Real-life applications, Comparison of Machine Learning with Traditional Programming, ML vs AI vs Data Science. **Learning Paradigms:** Learning Tasks- Descriptive and Predictive Tasks, Supervised, Unsupervised, Semi-supervised and Reinforcement Learnings.

Unit II: Regression (06 Hour)

Introduction: Regression, Need of Regression, Compare Regression and Correlation, Bias-Variance tradeoff, Overfitting and Under fitting. **Regression Techniques :** Support Vector Regression, Ridge Regression, Lasso Regression, Elastic Net Regression

Unit III: Classification (06 Hour)

Introduction: Need of Classification, Types of Classification (Binary and Multiclass), Binary-vs-Multiclass Classification, Balanced and Imbalanced Classification Problems , **Algorithms:** Logistic Regression for Multiclass , Decision Tree Classifier, Linear Support Vector Machines (SVM) – Introduction, Soft Margin SVM, **Kernel functions:** Radial Basis Kernel, Gaussian, Polynomial, Sigmoid

Unit IV: Evaluation Metrics (6 Hours)

Evaluation Metrics for Regression: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared ,Adjusted R-squared, **Classification Performance Evaluation:** Confusion Matrix, Accuracy, Precision, Recall

Unit V: Unsupervised Machine learning algorithms (6 Hours)

Clustering, Need of Clustering, Types of Clustering, Hierarchical clustering algorithm, K-Means clustering algorithm, Advantages and disadvantages of K-Means clustering algorithm, Elbow method.

Unit VI: Reinforcement Learning (6 Hours)

Reinforcement Learning, Need for Reinforcement Learning, Supervised vs Unsupervised vs Reinforcement Learning, Types of Reinforcement, Elements of Reinforcement Learning, Real time applications of Reinforcement learning.

Text Books:

1. **Goodfellow, I., Bengio, Y., & Courville, A.** (2016). Deep Learning. MIT Press. ISBN: 978-0262035613
2. **Raschka, S., & Mirjalili, V.** (2020). Python Machine Learning. Packt Publishing. ISBN: 978-1789955750
3. **Zhang, C., & Ma, Y.** (2021). Ensemble Machine Learning. Springer. ISBN: 978-3030985064

Reference Books:

1. Ian H Witten, Eibe Frank, Mark A Hall, “Data Mining, Practical Machine Learning Tools and Techniques”, Elsevier, 3rd Edition
2. Jiawei Han, Micheline Kamber, and Jian Pie, “Data Mining: Concepts and Techniques”, Elsevier Publishers Third Edition, ISBN: 9780123814791, 9780123814807
3. Shalev-Shwartz, Shai, and Shai Ben-David, “Understanding machine learning: From theory to algorithms”, Cambridge university press, 2014
4. McKinney, “Python for Data Analysis O' Reilly media, ISBN : 978-1-449- 31979-3
5. Hastie, T., Tibshirani, R., & Friedman, J. (2017). The Elements of Statistical Learning. Springer. ISBN: 978-0387848570
6. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press. ISBN: 978-0521865715

E-Resources:

4. <https://www.coursera.org/learn/machine-learning>
5. <https://www.edx.org/learn/machine-learning>
6. <https://www.kaggle.com/learn/intro-to-machine-learning>

List of Assignments

1. Use the diabetes data set from UCI and Pima Indians Diabetes data set for performing the following:

- a. Univariate analysis: Frequency, Mean, Median, Mode, Variance, Standard Deviation, Skewness and Kurtosis
- b. Bivariate analysis: Linear and logistic regression modeling
Dataset link: <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>
2. Predict House Prices using Regression
 - a. Apply Linear, Ridge, and Lasso regression to predict house prices.
Dataset: Boston Housing Dataset
3. Predict the price of the Uber ride from a given pickup point to the agreed drop-off location. Perform following tasks:
 - a. Pre-process the dataset.
 - b. Identify outliers.
 - c. Check the correlation.
 - d. Implement linear regression and ridge, Lasso regression models.
 - e. Evaluate the models and compare their respective scores like R2, RMSE.
Dataset link: <https://www.kaggle.com/datasets/yasserh/uber-fares-dataset>
4. Handwritten Digit Recognition
 - a. Train a Logistic Regression model on MNIST for digit classification.
Dataset: MNIST Handwritten Digits
5. Implementing K-Means Clustering on the Iris Dataset and Determining the Optimal Number of Clusters using the Elbow Method.
6. Use the Wine Quality Dataset to predict wine quality based on physicochemical properties.
<https://archive.ics.uci.edu/dataset/186/wine+quality>
7. Stock Price Prediction using Regression
 - a. Use Support Vector Regression (SVR) to predict stock market trends.
Dataset: Yahoo Finance Stock Data
8. Use the Student Performance Dataset to predict student performance based on their study habits, family background, etc.
<https://archive.ics.uci.edu/dataset/320/student+performance>
9. Use the **Air Quality Dataset** to predict air quality based on several environmental features.
<https://archive.ics.uci.edu/dataset/360/air+quality>
10. Segmenting Customers using K-Means
 - a. Apply K-Means Clustering to divide customers into groups.
Dataset: Mall Customers Segmentation



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

AI24371- Deep Learning: Foundations, Architectures, and Application

Teaching Scheme:

Theory: 3 Hours/Week

Practical: 2 Hours/Week

Credits
04

Examination Scheme:

CAA:10 Marks

ISE: 30 Marks

ESE: 60 Marks

TW: 30 Marks

Prerequisites: Python Programming

Course Objectives:

- To build a strong foundation in neural networks, including their architectures, working principles, and training mechanisms.
- To analyze the evolution of deep learning, its fundamental concepts, and real-world applications.
- To explore hyperparameter tuning techniques and optimization strategies for improving deep learning model performance.
- To design and implement convolutional neural networks (CNNs) for computer vision applications.
- To develop expertise in Recurrent Neural Networks (RNNs) for sequence modeling tasks such as language processing and time series forecasting.
- To investigate advanced recurrent architectures, including LSTMs and Bi-LSTMs, for handling long-term dependencies in sequential data.

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

CO1: Explain the fundamentals of neural networks, including perceptrons, activation functions, and training mechanisms...

CO2: Compare and apply deep learning techniques for various real-world applications, understanding their strengths and challenges.

CO3: Optimize neural network performance through hyperparameter tuning, gradient-based learning, and advanced optimization techniques.

CO4: Implement and train convolutional neural networks (CNNs) for solving image classification and object detection problems.

CO5: Develop and apply recurrent neural networks (RNNs) and their variants (LSTMs, Bi-LSTMs) for sequence-based data processing.

CO6: Solve complex sequence modeling problems using advanced RNN architectures and backpropagation through time (BPTT).

Course Contents

Unit I Fundamentals of Neural Networks (06 Hours)

Biological Neuron, The Perceptron(AND,OR,NOT,XOR), Deep Forward Network, Multilayer Feed-Forward Networks, **Training Neural Networks:** Backpropagation and Forward Propagation, **Activation Functions** :Linear, Sigmoid, Tanh, Hard Tanh, Softmax, Rectified Linear, **Loss Functions:**Loss Function Notation, Loss Functions For Regression, Loss Functions For Classification.

Unit II Introduction To Deep Learning Foundations (06 Hours)

Machine Learning And Deep Learning, History Of Deep Learning, Advantage And Challenges Of Deep Learning. Learning Representations From Data , Understanding How Deep Learning Works In Three Figures(Input, Hidden Layer, Output), Common Architectural Principles Of Deep Network, Architecture Design, Applications Of Deep Learning

Unit III Hyperparameter Tuning of Neural Network (06 Hours)

Hyperparameters : Learning Rate, Regularization, Cost Functions, Error Back Propagation, Gradient-Based Learning, Implementing Gradient Descent, Vanishing And Exploding Gradient Descent, Optimization Algorithm(SGD)

Unit IV Convolutional Neural Network (CNN) (06 Hours)

CNN Architecture Overview, The Basic Structure Of A Convolutional Network- Padding, Strides, Typical Settings, The Relu Layer, Pooling, Fully Connected Layers, The Interleaving Between Layers, Training a Convolutional Network.

Unit V Recurrent Neural Networks (6 Hours)

Recurrent Neural Networks Overview, Need for RNNs, RNNs And Traditional Feedforward Networks, Basic RNN Architecture, Working of an RNN, Applications of RNNs.

Unit VI Advanced Recurrent Neural Networks and Sequence Learning (06 Hours)

Vanishing Gradient And Exploding Gradient Problem, RNN Differs From Feedforward Neural Networks, Backpropagation Through Time (BPTT), Types Of Recurrent Neuralnetworks, Bidirectional Rnns, Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM)

Text Books:

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. ISBN: 978-0262035613.
2. Patterson, J., & Gibson, A. (2017). Deep Learning. O'Reilly Media. ISBN: 978-1491925614.
3. Aggarwal, C. (2018). Neural Networks and Deep Learning. Springer. ISBN: 978-3319944623.
4. Buduma, N. (2017). Fundamentals of Deep Learning. SPD. ISBN: 978-9386551551.
5. Chollet, F. (2017). Deep Learning with Python. Manning Publications. ISBN: 978-1617294433.

Reference Books:

1. Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT Press. ISBN: 978-0262039246.
2. Weidman, S. (2019). Deep Learning from Scratch: Building with Python from First Principles. O'Reilly Media. ISBN: 978-1492047489.
3. Duval, F. (2020). Deep Learning for Beginners: Practical Guide with Python and TensorFlow. Independently Published. ISBN: 979-8618122373.
4. **Sutton, R. & Barto, A.** (2018). Reinforcement Learning: An Introduction (2nd Ed.). MIT Press. ISBN: 978-0262039246.

E-Resources:

7. <http://csis.pace.edu/ctappert/cs855-18fall/DeepLearningPractitionersApproach.pdf>
8. https://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-1col-dkrieselcom.pdf
9. **Deep Learning Specialization – Andrew Ng (Coursera)**
<https://www.coursera.org/specializations/deep-learning>
10. **Practical Deep Learning for Coders (FastAI)**
<https://course.fast.ai/>
11. **CS231n: Convolutional Neural Networks for Visual Recognition (Stanford University)**
<http://cs231n.stanford.edu/>
12. **MIT Deep Learning Course**
<https://deeplearning.mit.edu/>

List of Assignments

1. **Perceptron Model Implementation:** Develop a simple perceptron-based neural network to classify logical operations (AND, OR, XOR).
2. **Train a Multi-Layer Perceptron (MLP) for Digit Recognition.** Use an MLP model to classify handwritten digits. Dataset: MNIST Dataset.
3. **Visualizing Activation Functions** Implement different activation functions (Sigmoid, ReLU, Tanh) and plot their outputs. Dataset: Synthetic data.
4. **Exploring Deep Learning Architectures** Implement a simple feedforward network for classification tasks. Dataset: Fashion MNIST.
5. **Compare Performance of Shallow vs. Deep Networks** Train 3-layer vs. 10-layer neural networks and compare accuracy. Dataset: CIFAR-10.
6. **Deep Learning for Image Classification** Build a deep fully connected network for image classification. Dataset: Tiny ImageNet.
7. **Implementing Regularization Techniques** Compare L1 (Lasso) and L2 (Ridge) regularization on a neural network. Dataset: Diabetes Dataset.
8. **Impact of Different Optimizers** train a model using SGD, Adam, RMSProp and compare results.
9. **Build a Basic CNN for Image Classification** implement a CNN model for digit classification.
10. **Sentiment Analysis using LSTM** train a LSTM for IMDB movie review sentiment classification. Dataset: IMDB Dataset



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

AI24461- Artificial Intelligence and Data Science Advanced Applications and Trends

Teaching Scheme:

Theory: 3 Hours/Week
Practical: 2 Hours/Week

Credits
04

Examination Scheme:

CAA:10 Marks
ISE: 30 Marks
ESE: 60 Marks
TW: 30Marks

Prerequisites: Machine Learning: Concepts, Techniques, and Applications, Deep Learning: Foundations, Architectures, and Application, Python programming, Statistics and probability, Linear algebra.

Course Objectives:

1. provide a strong foundation in advanced concepts of Artificial Intelligence and Data Science
2. develop the ability to apply AI and DS techniques to real-world problem
3. expose learners to modern trends such as deep learning, generative AI, explainable AI, edge AI, and responsible AI
4. build practical skills in data analysis, model development, evaluation, and deployment
5. encourage students to explore domain-specific AI applications in healthcare, finance, smart systems, education, agriculture, and industry

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

- CO1: Understand advanced methods in AI and Data Science and their practical relevance
CO2: Analyze structured and unstructured data using suitable computational technique
CO3: Develop machine learning and deep learning models for predictive and intelligent applications
CO4: Apply natural language processing, computer vision, and generative AI methods in selected domains
CO5: Evaluate AI systems using performance, fairness, explainability, and ethical considerations
CO6: Design data-driven solutions for real-life applications and emerging industrial trends

Course Contents

Unit I: Foundations of Advanced Artificial Intelligence and Data Science

Overview of AI and Data Science ecosystems, AI life cycle and data science process, Supervised, unsupervised, semi-supervised, and reinforcement learning overview, Data collection, data preprocessing, feature engineering, and data pipelines, Model evaluation metrics and validation strategies, Introduction to scalable AI systems and modern AI tools

Unit II: Advanced Machine Learning and Predictive Analytics

Regression and classification revisited, Ensemble learning: bagging, boosting, random forests, XGBoost basics, Support Vector Machines and kernel methods, Clustering and dimensionality reduction, Time series analysis and forecasting, Anomaly detection and recommendation systems

Unit III: Deep Learning and Neural Network Applications

Artificial neural networks and deep neural networks, Backpropagation and optimization methods, Convolutional Neural Networks, Recurrent Neural Networks, LSTM, GRU, Transfer learning and fine-tuning, Deep learning frameworks and practical model building

Unit IV: Natural Language Processing, Generative AI, and Intelligent Interaction

Fundamentals of NLP, Text preprocessing, embeddings, and vectorization, Sentiment analysis, text classification, and named entity recognition, Transformers and attention mechanism, Large Language Models and Generative AI overview, Prompt engineering basics and AI-assisted applications, Chatbots, question answering systems, and text generation

Unit V: Computer Vision, Edge AI, and Domain-Specific Applications

Image processing fundamentals, Object detection and image segmentation, Video analytics and activity recognition, AI at the edge and IoT-enabled intelligent systems, AI applications in healthcare,

agriculture, education, finance, cybersecurity, and smart cities, Case studies of industrial AI systems

Unit VI: Explainable AI, Responsible AI, Deployment, and Future Trends

Model interpretability and explainable AI concepts, Bias, fairness, accountability, and transparency, Privacy-preserving AI and secure data analytics, MLOps basics and model deployment pipelines, Federated learning and AutoML overview, Emerging trends: multimodal AI, AI agents, human-AI collaboration, sustainable AI

Text Books:

1. **Artificial Intelligence: A Modern Approach**, Stuart Russell and Peter Norvig, 4th Edition, 2021.
2. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow**, Aurélien Géron, 3rd Edition, 2022.
3. **Deep Learning**, Ian Goodfellow, Yoshua Bengio, and Aaron Courville, 1st Edition, 2016.
4. **Pattern Recognition and Machine Learning**, Christopher M. Bishop, 1st Edition, 2006.
5. **Natural Language Processing with Transformers**, Lewis Tunstall, Leandro von Werra, and Thomas Wolf, 1st Edition, 2022

Reference Books:

1. **Machine Learning: A Probabilistic Perspective**, Kevin P. Murphy, 1st Edition, 2012.
2. **Data Science for Business**, Foster Provost and Tom Fawcett, 1st Edition, 2013.
3. **Interpretable Machine Learning**, Christoph Molnar, 2nd Edition, 2022.
4. **Deep Learning for Coders with fastai and PyTorch**, Jeremy Howard and Sylvain Gugger, 1st Edition, 2020.
5. **Speech and Language Processing**, Daniel Jurafsky and James H. Martin, 3rd Edition, draft/online resource widely used.
6. **Introduction to Machine Learning with Python**, Andreas C. Müller and Sarah Guido, 1st Edition, 2016.

1. NPTEL – Artificial Intelligence domain

https://nptel.ac.in/domains/discipline/106?course=106_0

2. NPTEL – An Introduction to Artificial Intelligence

https://onlinecourses.nptel.ac.in/noc22_cs56/preview

3. TensorFlow – Tutorials <https://www.tensorflow.org/tutorials>

4. PyTorch – Official Website

<https://pytorch.org/>

List of Assignments

1. Data Preprocessing and Exploratory Data Analysis

Collect a real dataset, clean it, handle missing values, perform visualization, and summarize insights.

2. Predictive Modeling using Regression or Classification

Build a prediction model on a structured dataset and compare at least two algorithms.

3. Ensemble Learning Performance Study

Implement random forest and boosting techniques on a dataset and evaluate performance differences.

4. Clustering and Dimensionality Reduction

Apply clustering and PCA/t-SNE on a dataset and interpret discovered patterns.

5. Deep Learning for Image Classification

Develop a CNN model for image classification using a benchmark or custom dataset.

6. Sequence Modeling using RNN/LSTM

Perform time series forecasting or text sequence modeling using RNN/LSTM.

7. Natural Language Processing Mini Project

Create an NLP system for sentiment analysis, spam detection, or topic classification.

8. Generative AI / Prompt Engineering Task

Design prompts for a specific application and analyze the quality, reliability, and limitations of generated outputs.

9. Computer Vision or Domain Application Case Study

Implement a simple computer vision task or prepare a case-based AI solution for healthcare, agriculture, finance, or education.

10. Responsible AI and Explainability Report

Select a model, apply explainability methods, and prepare a report discussing fairness, trust, risk, and ethical use.