**APPLIED MACHINE LEARNING**

**Course Objectives**

1. Understand the core concepts and techniques of machine learning and artificial intelligence.
2. Develop machine learning models using popular libraries and frameworks.
3. Evaluate the performance of machine learning models using appropriate metrics.
4. Apply machine learning to various real-world problems and domains.

**Course Outcomes**

On completion of this course, the students will be able to

1. Recall and define key machine learning concepts, terminologies, and algorithms.
2. Describe the differences between supervised, unsupervised, and reinforcement learning.
3. Apply data preprocessing techniques to clean, transform, and prepare datasets for machine learning.
4. Apply, compare, and contrast the strengths and weaknesses of different machine learning algorithms.

**Catalog Description**

This course introduces students to the foundational principles, algorithms, and applications of machine learning. Through a combination of lectures, hands-on exercises, and projects, students will gain a solid understanding of supervised and unsupervised learning techniques. Topics include data preprocessing, model selection, evaluation, and ethical considerations in machine learning. The course equips students with the skills necessary to apply machine learning to real-world problems in various domains.

**Course Content**

**Unit 1: Introduction:** **8 hours**

Overview of machine learning and its applications, Types of machine learning: supervised, unsupervised, reinforcement, GenAI; Loss functions:Mean Squared Error (MSE), Mean Absolute Error (MAE), Huber Loss, Binary Cross-Entropy Loss (Log Loss), Categorical Cross-Entropy Loss, Sparse Categorical Cross-Entropy Loss, Hinge Loss (SVM Loss), Triplet Loss; Optimizer function:Stochastic gradient descent, Mini-Batch Gradient Descent, Momentum, Adaptive gradient algorithm (Adagrad), Adam (Adaptive Moment Estimation), RMSprop (Root Mean Square Propagation), Adadelta

**Unit 2: Data Preprocessing: 10 hours**

Python and libraries for machine learning (e.g., NumPy, Pandas, Seaborn, scikit-learn), Data Cleaning: handling Missing Data, Handling Outlier, Normalization, Statistical Interpolation of Data, Data Transformation: Feature Scaling, Feature Encoding, Feature Engineering, Data Reduction: Dimensionality reduction technique, feature selection, LDA, PCA, Data Splitting: Cross validation techniques, Handling imbalanced data: Oversampling techniques, under sampling techniques.

**Unit 3: Regression: 8 hours**

Introduction to Regression, Regression examples, Regression models, Steps in regression analysis, Linear regression, Simple linear regression, Mathematical proof of linear regression, Least squares estimation, Least squares regression-Line of best fit, Illustration, Direct regression method, Maximum likelihood estimation, Coefficient of determination (R-squared), Checking model adequacy, Over-fitting, Detecting over-fit models: Cross validation, Logistic regression, Mathematical proof of logistic regression, multiple linear regression, Multiple linear regression model building, Mathematical proof of Multiple linear regression model, Interpretation of multiple linear regression coefficients-Partial regression coefficients, Standardized regression coefficients, Missing data, Validation of multiple regression model, regularization, ridge and lasso regularization.

**Unit 4: Classification: 12 hours**

Introduction, ML classifier, Classification and general approach, Classification algorithms: K-Nearest Neighbour, Naïve Bayes Classifier, SVM, Instance based learning, Decision trees, Attribute selection measure: Information gain, ID3 algorithm, Converting a tree to rules, Ensemble, Ensemble of classifiers, Bagging, Boosting, Random forests; Neural networks: Activation functions, Feedforward neural network, Multi-layer perceptron, Back propagation algorithm, Recurrent or feedback architecture, Perceptron rule, Multilayer networks and back propagation algorithm, Support vector machine, Classification model evaluation and selection, ROC curves, AUC curves.

**Unit 5: Clustering Techniques: 10 hours**

Introduction to Clustering, Clustering algorithms, Statistics associated with cluster analysis, General applications of clustering, Clustering as a pre-processing tool, Similarity and dissimilarity between objects, Type of data in clustering analysis, Binary variables, Nominal variables, Ordinal variables, Cluster centroid and distances, Hierarchical clustering, Hierarchical Agglomerative Clustering (HAC), Hierarchical Agglomerative Clustering: Linkage method, Hierarchical Agglomerative Clustering: Variance and Centroid method, Cluster distance measures, agglomerative clustering, Distance between two clusters, Hierarchical clustering: Time and Space requirements, K - means clustering, The K-medoids clustering method, CLARA (Clustering Large Applications), Density based clustering methods, DBSCAN, Cluster validity indices.

**Textbook**

* "Introduction to Machine Learning with Python" by Andreas C. Müller and Sarah Guido
* "Pattern Recognition and Machine Learning" by Christopher M. Bishop

**Reference book**

* "Machine Learning: A Probabilistic Perspective" by Kevin P. Murphy
* "Python Machine Learning" by Sebastian Raschka and Vahid Mirjalili

**Modes of Evaluation: Quiz/Assignment/ presentation/ extempore/ Written Examination**

**Examination Scheme:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Components** | **Quiz/Assignment/Class Test etc.** | **MSE** | **ESE** |
| **Weightage (%)** | **50%** | **20%** | **30%** |

**Relationship between the Course Outcomes (COs), Program Outcomes (POs) and Program Specific Outcomes (PSOs):**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PO/CO | PO  1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO  9 | PO  10 | PO  11 | PO  12 | PSO  1 | PSO  2 | PSO  3 |
| CO1 | 1 | 1 | 1 | - | - | - | - | - | - | - | - | - | 1 |  | 3 |
| CO2 | 1 | 1 | 1 | - | - | - | - | - | - | - | - | - | 1 |  | 3 |
| CO3 | 1 | 1 | 1 | - | - | - | - | - | - | - | - | - | 1 |  | 3 |
| CO4 | 1 | 1 | 1 | - | - | - | - | - | - | - | - | - | 2 |  | 3 |
| Average | 1 | 1 | 1 | - | - | - | - | - | - | - | - | - | 1.4 |  | 3 |

1=weak 2= moderate 3=strong