Course Curriculam

Course Code: AIML202

Course Level UG

Course Title Machine Learning: Theory and Applications

Course Description :

Credit Units

L	T	P/S	SW	AS/DS	FW	No. of PSDA	Total Credit Unit
3	0	2	0	0	0	0	4

Course Objectives :

SN.	Objectives
1	The goal of the course is to understand the basic theory underlying machine learning, be able to formulate machine learning problems corresponding to different applications, to understand a range of machine learning algorithms along with their strengths and weaknesses and be able to apply machine learning algorithms to solve problems of moderate complexity. To apply the algorithms to a real-world problem, optimize the models learned and report on the expected accuracy that can be achieved by app

Pre-Requisites : General

Course Contents / Syllabus :

SN.	Module	Descriptors / Topics	Weightage
1	Module I Introduction to Machine Learning	• What is machine Learning? • Machine learning applications. • Types of machine learning: Supervised learning, unsupervised learning, reinforcement learning • Feature Representation, data cleaning, feature scaling, feature engineering, • Supervised and unsupervised learning Tasks: Regression, Classification and Clustering	10.00
2	Module II Supervised Learning: Regression	• Types of Regression models: Linear regression, Multiple linear regression, Polynomial regression • Model Complexity: overfitting and underfitting • Evaluation Measures: R square, Mean Square Error (MSE), Mean Absolute Error • Regularization techniques: Ridge, Lasso, Elastic Net • Real-world applications	20.00
3	Module III Supervised Learning: Classification	• Types of classification algorithms, Inductive Bias of machine learning classifiers, Curse of Dimensionality • Logistic regression • Support vector machines (SVMs): Hard and Soft-margin, kernel trick • Naive Bayes: Bayes theorem, Naive Bayes assumption, Types of Naive Bayes classifiers, Applications • K-Nearest Neighbour: Introduction to k-NN, Distance metrics, Choosing k • Cross-Validation strategies • Evaluation measures- Confusion Matrix, Precision, Recall, Specificity, AUC	25.00
4	Module IV Supervised Learning: Decision Trees and Ensemble Learning	• Decision Trees: Introduction, Attribute selection measures (Information Gain, Gini Index), Tree construction algorithm • Introduction to ensemble methods: Bagging and boosting, Random forests • Gradient boosting • Combining machine learning models: voting, averaging and stacking • Applications of ensemble methods	25.00
5	Module V Unsupervised Learning	• Why unsupervised learning? • Clustering methods: K-Means Clustering, Hierarchical Clustering • Dimensionality Reduction: Principal Component Analysis • Evaluation metrics for unsupervised learning (Silhouette Score, Calinski Harabaz Index) • Applications of unsupervised learning (anomaly detection, image segmentation, recommender systems)	20.00

Course Learning Outcomes :

SN.	Course Learning Outcomes

1	After completing this course, the student will be able to • Highlight the importance of visualization in the data analytics solution.
2	After completing this course, the student will be able to Implementing structured thinking to unstructured problems.
3	After completing this course, the student will be able to Summarizing a very broad collection of machine learning algorithms and problems.
4	After completing this course, the student will be able toAssessing algorithmic topics of machine learning and mathematically deep enough to introduce the required theory.
5	After completing this course, the student will be able to Articulate an appreciation for what is involved in learning from data.
6	After completing this course, the student will be able to Solving real life problems using machine learning algorithms.

Pedagogy for Course Delivery :

SN.	Pedagogy Methods
1	The course pedagogy will include lectures, e-contents, e-tutorials, numerical practice and experimental investigations

Theory /VAC / Architecture Assessment (L,T & Self Work): 75.00 Max : 100

Attendance+CE+EE : 5+35+60

SN.	Type	Component Name	Marks
1	Attendance		5.00
2	End Term Examination (OMR)		60.00
3	Internal	MID TERM EXAM	15.00
4	Internal	CLASS QUIZ	10.00
5	Internal	HOME ASSIGNMENT	10.00

Lab/ Practical/ Studio/Arch. Studio/ Field Work Assessment : 25.00 Max : 100

Attendance+CE+EE : 5+35+60

SN.	Type	Component Name	Marks
1	Attendance		5.00
2	External	PRACTICAL	30.00
3	External	Viva	30.00
4	Internal	PERFORMANCE	15.00
5	Internal	Lab record including Attendance,Ethics,Precautions,Interactions	10.00
6	Internal	Viva	10.00

Lab/ Practical details, if applicable :

SN.	Lab / Practical Details
1	Importing Data and Data Visualization Task: Construct various type of plots/charts like histogram, bar chart, pie-chart and scatter plot by importing data from a CSV format file. Further label different axes and data in a plot.
2	Data Cleaning and Pre-processing Task: Fill the missing values, removing/inserting columns, labelling the output column, feature scaling, converting the categorical values to numerical values etc.
3	Dimension Reduction using PCA: Use Principal Component Analysis to reduce the number of dimensions and thereby the size of the raw dataset.
4	Linear Regression: Implement an ordinary least squares linear regression model based on sample data and test the accuracy of the model

5	5. Multiple Linear Regression: Predict the output of the dependent variable based on multiple features of the sample dataset.
6	6. Logistic Regression: Implement a classification/ logistic regression problem on a sample dataset to separate the data into their labelled categories.
7	7. Write a program to train the Naïve Bayes classifier for a sample training data set stored as a .csv file. Compute the accuracy of the classifier, considering the test data set.
8	8. Write a program to implement K-Nearest Neighbour algorithm to classify the sample data set into different classes. Print both correct and wrong predictions.
9	9. Write a program to train a Decision Tree Classifier to classify the sample dataset into different categories.
10	10. Clustering: Implement K-means algorithm to form clusters in a sample dataset.For example, use wine dataset from UCI Machine Learning repository. It contains 178 observations of wine grown in the same region in Italy. Each observation is from one of three cultivars (the 'Class' feature), with 13 constituent features that are the result of a chemical analysis. Their annual spending on diverse product categories, like milk, grocery, region, etc.

List of Professional skill development activities :

No.of PSDA : 0

N/A

Text & References :

SN.	Type	Title/Name	Description	ISBN/ URL
1	Book	"Machine Learning: A Probabilistic Perspective" by Kevin Murphy, published by MIT Press, 2012		
2	Book	"Pattern Recognition and Machine Learning" by Christopher M. Bishop, published by Springer, 2006		
3	Book	"Python Machine Learning" by Sebastian Raschka and Vahid Mirjalili, published by Packt Publishing, 2		
4	Book	"Machine Learning Yearning" by Andrew Ng, published by Goodfellow Publishers, 2018		
5	Book	"Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron, published by		
6	Book	"Applied Predictive Modeling" by Max Kuhn and Kjell Johnson, published by Springer, 2013.		
7	Book	"Reinforcement Learning: An Introduction" by Richard S. Sutton and Andrew G. Barto, published by MIT		
8	Url	https://nptel.ac.in/courses/106106139		