PUCIT

MACHINE LEARNING

Subject Code: STML401

Total Hours: 40 Credits: 3

Course Learning Objectives (CLO)

This program provides a comprehensive understanding of machine learning and its applications. Participants will learn to apply supervised and unsupervised learning algorithms, evaluate model performance, and interpret results. The program also covers graph analysis. Through practical exercises and projects, learners gain the skills needed for roles in data science, AI, and machine learning. The goal of this course is to help Students/working professionals upskill and equip themselves with skills required to build and analyse Machine Learning models in production.

UNIT 1: Introduction to different types of ML [6 hours]

- Types of machine learning: supervised, unsupervised, and reinforcement learning, Real-world examples of machine learning applications.
- Maths Behind ML: Linear Algebra, Probability, Statistics, Calculus, Optimization

UNIT 2: Supervised ML [12 hours]

- Linear regression and polynomial regression
- Logistic regression
- Naive Bayes,
- Decision trees and random forests
- Support vector machines (SVM)
- k-Nearest Neighbors (kNN)
- Regularization Methods,
- Gradient boosting
- Linear Discriminant Analysis (LDA)

UNIT 3: Unsupervised ML [8 hours]

- Introduction to unsupervised learning

- Clustering algorithms: k-means, hierarchical clustering, and DBSCAN
- Principal Component Analysis (PCA)

UNIT 4: Latent Variable Models: [8 hours]

- Introduction to latent variable models
- Factor analysis
- matrix factorization
- singular value decomposition
- Gaussian mixture models
- Hidden Markov models (HMM)

Unit 5: Graph Analysis [6 hours]

- Introduction to graph theory
- Graph representations and properties
- Network analysis and visualization
- Applications of graph analysis in machine learning

SKILL BASED EXERCISE (SBE):

Note: - These Projects/activities are only indicative; the faculty member can innovate

Assignments/Mini Projects:

- Implement linear regression from scratch to predict the housing prices based on the number of bedrooms, bathrooms, and square footage. Use the dataset from Kaggle or any other available source.
- Predicting the Price of a Football Player
- Road Accident Severity Prediction:

  Develop a machine learning model that can predict the severity of road accidents based on the attributes provided during reporting. The goal is to build a system that can assist emergency response departments in prioritizing and allocating appropriate resources based on the predicted severity of accidents.

- Spam Classifier: Train a logistic regression model to classify emails as spam or not spam based on their content.
- Implement k-means clustering to group the customers based on their purchase behavior. Use the retail dataset from UCI Machine Learning Repository or any other available source.
- Iris Flower Classification: Use a decision tree or random forest model to classify different species of iris flowers based on their petal and sepal sizes.
- Implement logistic regression to predict whether a bank customer will default or not. Use the dataset from Kaggle or any other available source.
- Implement a random forest classifier to classify the handwritten digits. Use the MNIST dataset or any other available source.
- Customer Segmentation: Use clustering algorithms like k-means or DBSCAN to segment customers based on their purchasing habits.
- Image Compression: Use PCA to compress an image and reconstruct it from the compressed form.
- Implement principal component analysis (PCA) to reduce the dimensions of the iris dataset. Visualize the data using scatter plots.
- Anomaly Detection: Use clustering algorithms or autoencoders to detect anomalies in network traffic data.
- Implement hierarchical clustering to group the movies based on their similarity. Use the MovieLens dataset or any other available source.
- Evaluate the performance of different classification models (Logistic Regression, Random Forest, and SVM) on the bank customer default dataset. Use metrics such as confusion matrix, precision, recall, and F1 score.
- Handwriting Recognition: Use a Hidden Markov Model (HMM) to recognize handwritten digits.
- Topic Modeling: Use a Gaussian mixture model to perform topic modeling on a set of documents.
- Factor Analysis: Use factor analysis to extract underlying factors from a dataset of survey responses.
- Social Network Analysis: Analyze a social network dataset using graph analysis techniques like centrality measures and community detection.

Course Outcomes: Upon completion of this course, students will be able to:

- Understand the basic concepts and applications of machine learning
- Identify real-world examples of machine learning applications
- Apply regression, classification, and ensemble learning algorithms to solve supervised learning problems
- Understand the trade-offs and limitations of different supervised learning and unsupervised learning algorithms
- Apply clustering, dimensionality reduction, and autoencoder algorithms to unsupervised learning problems
- Understand the concept of latent variables and their application in machine learning
- Develop and apply factor analysis, Gaussian mixture models, and HMMs to various datasets
- Understand the basic concepts and applications of graph theory and network analysis
- Apply graph analysis techniques like centrality measures and community detection to various datasets
- Understand the basic concepts and applications of artificial neural networks
- Develop and train feedforward neural networks, convolutional neural networks, and recurrent neural networks on various datasets
- Understand the basic concepts and applications of advanced deep learning techniques like GANs, reinforcement learning, transfer learning, and NLP architectures

Benefits of our Program

- Learn from renowned Professors & Industry Experts.
- A program designed by Subject-Matter Experts
- Recognized Certification from PU
- LIVE classes + Lifetime recorded videos
- Job Placement Assistance
- A digital portfolio through "Github"
- Hands-on Workshops & Hackathons

- Technical Support Available in English, தமிழ் and हिंदी

Textbook:

1. "Pattern Recognition and Machine Learning" by Christopher Bishop
2. "Machine Learning: A Probabilistic Perspective" by Kevin P. Murphy
3. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron
4. "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
5. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman

References and Online Resources:

1. Andrew Ng's Machine Learning course on Coursera: Offers a comprehensive introduction to machine learning concepts and algorithms.
2. Scikit-learn Documentation: Provides detailed documentation, examples, and tutorials for implementing machine learning algorithms using the scikit-learn library.
3. TensorFlow Documentation: Offers comprehensive documentation and examples for deep learning using the TensorFlow library.
4. Deep Learning Book: Provides a comprehensive and in-depth introduction to deep learning concepts and architectures.
5. Kaggle: A platform for machine learning competitions and a community with a vast collection of datasets, code, and notebooks for learning and practice.