# Chandigarh University, Gharuan

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Subject Code CST-316** | | **Machine Learning** | | **L** | **T** | **P** | **C** |
| Total Contact Hours : 45Hours | | **3** | **0** | **0** | **3** |
| Common to all Specializations of CSE 3rdYear | |
| **Prerequisite:**   * Familiarity with the basic probability theory * Familiarity with the basic linear algebra | | | | | |
| **Marks-100** | | | | | | | |
| Internal-40 | | | External-60 | | | | |
| **Unit** | **Course Outcomes** | | | | | | |
| **1** | Apply the basic concept of Machine learning and statistics learning to deal with real-life Problems. | | | | | | |
| **2** | Understand different machine learning algorithms, as well as underlying theories the behind them. | | | | | | |
| **3** | Select and apply the appropriate machine learning algorithm to solve problems of moderate complexity | | | | | | |
| **4** | Interpret and evaluate models generated from data. | | | | | | |
| **5** | Optimize the models learned and report on the expected accuracy that can be attained by applying the algorithms to a real-world problem. | | | | | | |

**Contents of the Syllabus**

**UNIT-I [14h]**

**Unit 1 (14 hrs)**

**Fundamentals of Machine Learning:** Introduction to Machine Learning (ML), Different types of Machine Learning, Machine Learning Life Cycle: Data Discovery, Exploratory Analysis: Data Preparation, Model Planning, Model Building, Model Evaluation, Real World Case Study. Foundation of ML: ML Techniques.

**Statistics Learning and Exploratory Data Analysis** :  Mean Median, Mode, Correlation, Covariance, Quartile, Maximum Likelihood, Bayesian Inference, Bias, Variance, Distance metrics: Euclidean Distance, Manhattan Distance, Gaussian (or Normal) Distributions, statistical hypothesis testing. Missing Value treatment, Outlier Detection, Feature Engineering, Graphs and Plots.

## Supervised Learning with Regression and Classification techniques -1: Linear Regression, Multiple Regression, Bias-Variance Dichotomy, Model Validation Approaches, Evaluation of the performance of an algorithm: Mean Squared Error, Root Mean Squared Error.

**Unit 2 (15 hrs)**

## Supervised Learning with Regression and Classification techniques -2:

Logistic Regression, Support Vector Machine (SVM), Naive Bayesian Classifier, K-Nearest Neighbor (KNN), Cross-Validation, Confusion Matrix. Evaluation of the performance of an algorithm: Accuracy, Error Rate, Precision, Recall, Specificity, F1 Score. Decision Tree: Picking the best splitting attribute, entropy and information gain, over fitting and under fitting, noisy data and pruning.Ensemble Methods: Random Forest,

**Unit 3 (16 hrs)**

**Unsupervised Learning**: Clustering, Partitioning Method - K-means, K-medioids, Hierarchical Clustering- Agglomerative and divisive clustering, Evaluation of clustering algorithms.Principal Component Analysis (Eigen values, Eigen Vectors, Orthogonality). Association Rules: Association Rule mining, Apriori Algorithm, Support and Confidence Parameters, Lift and Leverage. Feature Reduction and Dimensionality Reduction.

**Semi-Supervised Learning**: Introduction, Assumptions, Working and Real-World Applications. Reinforcement Learning: Introduction, Applications and Examples, Challenges of applying reinforcement learning, reinforcement learning algorithm.

## Text Books:

1. Ethem Alpaydın, Introduction to Machine Learning (Adaptive Computation and Machine Learning), MIT Press, 2004.
2. Mitchell. T, Machine Learning, McGraw Hill, 1997.
3. Nilsson, Nils J. "Introduction to machine learning: An early draft of a proposed textbook." (1996).

## Reference Material:

1. Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
2. Ryszard S. Michalski, Jaime G. Carbonell, Tom M. Mitchell, and Machine Learning: An Artificial Intelligence Approach, Tioga Publishing Company, 1983.

| **CO-PO-Mapping** | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CO** | **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** | **PSO1** | **PSO2** |
| CO-1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CO-2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CO-3 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CO-4 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CO-5 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Student’s Outcomes** | | | | | | | | | | | | | | |
|  | **CAC--SO1** | **CAC--SO2** | **CAC--SO3** | **CAC--SO4** | **CAC--SO5** | **EAC-SO1** | **EAC-SO2** | **EAC-SO3** | **EAC-SO4** | **EAC-SO5** | **EAC-SO6** | **EAC-SO7** | **CAC--SO1** | **CAC--SO2** |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Chandigarh University, Gharuan**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Subject Code CSP-317** | | **Machine Learning Lab** | | **L** | **T** | **P** | **C** |
| Total Contact Hours : 48Hours | | **0** | **0** | **2** | **1** |
| Common to all Specializations of CSE 3rd Year | |
| **Prerequisite:**   * Knowledge of basic computer science principles and skills, at a level sufficient to write a reasonably non-trivial computer program. | | | | | |
| **Marks-100** | | | | | | | |
| Internal-60 | | | External-40 | | | | |
| **Unit** | **Course Outcome** | | | | | | |
| **1** | Classify fundamental of data analysis, machine learning algorithms as supervised learning or unsupervised learning. | | | | | | |
| **2** | Select and apply the appropriate machine learning algorithm to solve problems of moderate complexity. | | | | | | |
| **3** | Design and evaluate the unsupervised models through python / R in built functions. | | | | | | |
| **4** | Evaluate the machine learning models pre-processed through various feature engineering algorithms by python/ R programming. | | | | | | |
| **5** | Optimize the models learned and report on the expected accuracy that can be attained by applying the algorithms to a real-world problem. | | | | | | |

**List of Experiments**

**UNIT-I**

1. Implement Exploratory Data Analysis on any data set.
2. Implement Data Visualization.
3. Implement Linear Regression on any data set.

**UNIT-II**

1. Implement Support Vector Machine on any data set and analyze the accuracy with Logistic regression
2. Implement Naïve Bayes on any dataset.
3. Implement K-Nearest Neighbor on any data set
4. Implement Decision Tree and compare the performance with Random Forest on any data set.

**UNIT-III**

1. Implement K-means clustering algorithm (cluster some sample data set into disjoint clusters using K-means).
2. Implement Principle Component Analysis.
3. Implement Association Rule Mining.

| **CO-PO-Mapping** | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CO** | **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** | **PSO1** | **PSO2** |
| CO-1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CO-2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CO-3 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CO-4 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CO-5 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Student’s Outcomes** | | | | | | | | | | | | | | |
|  | **CAC--SO1** | **CAC--SO2** | **CAC--SO3** | **CAC--SO4** | **CAC--SO5** | **EAC-SO1** | **EAC-SO2** | **EAC-SO3** | **EAC-SO4** | **EAC-SO5** | **EAC-SO6** | **EAC-SO7** | **CAC--SO1** | **CAC--SO2** |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |