**COURSE SYLLABUS**

| **Course Code:** CS2211 | **Course Name:** Introduction to Machine Learning 1 | |
| --- | --- | --- |
| **Semester:** 4 | **Category:** Major Computer Science Core (Programming) | |
| **Credits:** 2 | **L T P J:** 1 0 1 0 | **Contact Hours:** 15:0:30:0 = 45 |
| **Prerequisite (Course/Skill/Knowledge):**   * Sem 1: CS1200 – Data Analysis using Python | | |

# **Course Overview**

This course introduces students to the exciting field of machine learning, providing a comprehensive overview of key concepts and practical applications. This course aims to equip learners with the foundational knowledge and skills necessary to understand, fit, and evaluate machine learning models. This course will cover supervised, unsupervised, and deep learning, along with model evaluation at the basic level. Students are provided with real-time datasets and activities to inculcate the much-needed 21st century skills of problem solving and logical thinking.

# **Course Outcomes and Contents**

**CO1: Understand the history, definition, types and applications of machine learning in Natural Language Processing and Computer Vision 3 Hours**

Objective: Overview of Machine Learning - History of Machine Learning, Supervised, Unsupervised, and Reinforcement Learning Algorithms, definitions, and types of machine learning, Applications - Natural Language Processing: Sentiment Analysis, Machine Translation; Computer Vision: Object Recognition, Object Detection, and Object Generation.

**CO2: Perform categorical encoding, scaling, feature engineering, train test split after handling null values, duplicates and outliers 6 Hours**

Objective: Data preprocessing—handling missing data, scaling and normalization, dealing with outliers, encoding categorical variables. Terminology: feature, sample, training set, validation set, test set, feature engineering.

**CO3: Determine the relationship between dependent and independent variables encountered in specific domains using simple linear regression and its error metrics (MSE, MAE, RMSE) 6Hours** Objective: simple linear regression, loss functions, gradient descent, mean squared error, mean absolute error, root mean squared error.

**CO4: Perform model evaluation using cross-validation, confusion matrix and classification report for K-Nearest Neighbour in specific domains 6 Hours**

Objective: KNN, model evaluation metrics (confusion matrix, accuracy, precision, recall, F1-Score), cross-validation.

**CO5: Identify patterns in a provided dataset using k-means clustering with elbow method and dendrograms 3 Hours**

Objective: K-means clustering, elbow method and dendrograms

**CO6: Understand the concepts of perceptron, logistic regression (sigmoid activation function, binary classification), error backpropagation, feed-forward neural network 9 Hours**

Objective: Perceptron-Logistic regression, sigmoid activation function, binary classification, error back propagation algorithm, feed-forward neural network, Universal Approximation Theorem

**CO7: Solve a real-world problem belonging to health/finance by applying machine learning skills**

**12 Hours**

Project selection, project status review-1. Final project submission: source code and report, presentation, and Viva.

# **Textbooks**

1. Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc.".
2. Aurélien, Géron. "Hands-on machine learning with scikit-learn & tensorflow." *Geron Aurelien* 134 (2017): 145-150.

# **Reference Material**

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
2. VanderPlas, Jake. *Python data science handbook: Essential tools for working with data*. " O'Reilly Media, Inc.", 2016.
3. https://machinelearningmastery.com/machine-learning-algorithms-from-scratch/#packages

**Note:** Students completing Machine Learning specialization (nptel, coursera) and producing completion certificate will get corresponding assessment marks.

Example: <https://www.coursera.org/specializations/machine-learning-introduction>

# **Activities Plan:**

**Note: Please provide the details of HW or SW tools or platforms planned to be used for the lab activities.**

| **No:** | **Activity** |  |
| --- | --- | --- |
| **1** | **Given a dataset, apply data preprocessing methods** | **Anaconda or Google colab** |
| **2** | **Apply linear regression model to predict house prices in Bangalore** | **Anaconda or Google colab** |
|  | **TBD** |  |