INTRODUCTION TO MACHINE LEARNING
23CSOE322

(COURSE HANDBOOK)

COURSE Offered by:
Dept. of CS&E

MANGALORE INSTITUTE OF TECHNOLOGY & ENGINEERING
(A unit of Rajalaxmi Education Trust ®, Mangalore)
Autonomous Institute Affiliated to V.T.U., Belagavi , Approved by AICTE, New Delhi
Accredited by NAAC with A+ Grade & an ISO 9001:2015 Certified Institution

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Course Hand Book – Introduction to Machine Learning

1. Course Information

- **Semester**: VI Semester (Open Elective)
- **Course Title**: Introduction to Machine Learning
- **Course Code**: 23CSOE322
- **Teaching Scheme (L:T:P)**: 3 : 0 : 0
- **Total Contact Hours**: 42
- **Credits**: 3
- **CIE Marks**: 50
- **SEE Marks**: 50
- **Exam Duration**: 3 Hours

2. Course Prerequisites

Students enrolling in this course are expected to have basic proficiency in Python programming, with emphasis on the following fundamentals:

- Python syntax and program structure
- Data types: int, float, string, list, tuple, set, dictionary
- Indexing and slicing of strings and lists
- Conditional statements and looping constructs (if, for, while)
- Functions and basic modular programming
- File handling operations (reading from and writing to files)
- Basic familiarity with Python libraries for data handling (such as NumPy and Pandas)

Note: Students lacking these prerequisites are advised to complete a short Python refresher module before or during the initial weeks of the course to ensure smooth progression through Machine Learning concepts and model implementation.

2.1.Python Refresher Course websites :

- https://www.w3schools.com/python/
- https://infyspringboard.onwingspan.com/web/en/login
- https://www.freecodecamp.org/
- https://www.tutorialspoint.com/python/index.htm

3. Course Description

This course offers a systematic and comprehensive introduction to Machine Learning (ML), focusing on the principles, methodologies, and practical techniques that enable computers to learn from data and make intelligent decisions. The course begins with foundational concepts, including the nature of learning problems, concept learning, inductive bias, and the overall framework for designing machine learning systems.

Students will gain in-depth exposure to data preprocessing and preparation techniques, such as data cleaning, feature selection, dimensionality reduction, and validation strategies, which are critical for building reliable and efficient ML models. The course then covers core supervised learning algorithms, including regression and classification methods, along with gradient-based optimization techniques and performance evaluation metrics.

In addition, the course introduces unsupervised learning and ensemble learning approaches, enabling students to understand clustering, model aggregation, and methods for improving prediction accuracy and generalization. Through real-world case studies and datasets, students will learn to build, train, evaluate, and fine-tune machine learning models using modern tools and libraries.

By the end of the course, learners will develop a strong conceptual foundation and practical competency in machine learning, preparing them for advanced AI courses, data-driven problem solving, industry applications, and research-oriented work in artificial intelligence and machine learning domains.

4. Course Learning Objectives (CLOs)

This Course Designed to

1. Impart the knowledge on core concepts and underlying principles of machine learning.
2. Familiarize the various data preprocessing techniques.
3. Enable to build basic Machine Learning models using classification, regression, gradient descent algorithms and ensemble methods.

5. Course Outcomes (COs)

At the end of the course, students will be able to:

- **CO1**: Describe the foundational concepts of Concept Learning & Machine learning..
- **CO2**: Applies the essential data preparation techniques for robust and efficient machine learning implementations.
- **CO3**: Apply regression model and gradient descent algorithm to various realistic dataset & evaluate the performance evaluation of models.
- **CO4**: Apply various Classification algorithm to realistic dataset & evaluate the performance evaluation of models.
- **CO5**: Apply ensemble approach, SVM & K-Means algorithms to realistic dataset and fine tune the model for performance increase.

6. Course Structure & Module-wise Breakdown

Module 1: Foundations of Machine Learning (8 Hours)

- Learning Problems and Designing a Learning System
- Perspectives and Issues in Machine Learning
- Concept Learning Task
- Concept Learning as Search
- Find-S Algorithm
- Version Spaces and Candidate Elimination Algorithm
- Inductive Bias
- Framework for Developing Machine Learning Models

Learning Focus: Understanding how machines learn and represent concepts.

Module I Textbook & Section Wise Coverage :

- Textbook :
  - Machine Learning by Tom M Mitchell
  - Machine Learning using Python by Manaranjan Pradhan and U Dinesh Kumar
- Coverage :
- Chapter 1.1, 1.2, 1.3, 2.1 to 2.5,2.7, by Machine Learning by Tom M Mitchell
- Chapter 1.1, 1.2, 1.3 from textbook "Machine Learning using Python" by Manaranjan Pradhan and U Dinesh Kumar

Module 2: Dataset Pre-processing (8 Hours)

- Data Preparation and Challenges
- Train-Test Split and K-Fold Cross Validation
- Data Cleaning: Outliers, Missing Values, Statistical Imputation
- Feature Selection: Categorical and Numerical Features
- Data Transforms: Scaling and Encoding
- Dimensionality Reduction: LDA, PCA, SVD

Learning Focus: Preparing high-quality datasets for machine learning models.

Module II Textbook & Section Wise Coverage :

- Text Book : "Data Preparation for Machine Learning – Data Cleaning, Feature Selection and data transform in python" by Jason Brownlee
- Coverage :
  - Chapter 3 : 3.1 to 3.7, Chapter 4 : 4.1 to 4.4, Chapter 5 : 5.1 to 5.8, Chapter 6 : 6.1 to 6.5, Chapter 7 : 7.1,7.2,7.3,7.5, Chapter 8 : 8.1 to 8.4.1, Chapter 11 : 11.2, Chapter 12 : 12.1, 12.2, 12.3, Chapter 13 : 13.1, 13.2, 13.3, Chapter 17: 17.1 to 17.3, Chapter 19 : 19.1,19.2, 19.3.1,19.3.2, Chapter 27 : 27.1 to 27.3, Chapter 28 : 28.1 to 28.4, Chapter 29 : 29.1 to 29.4, Chapter 30 : 30.1 to 30.4

Module 3: Linear Regression and Gradient Descent (9 Hours)

- Introduction to Linear Regression
- Steps in Building a Linear Regression Model
- Gradient Descent Algorithm
- Model Building using Scikit-learn
- Performance Metrics: R-Squared, RMSE
- Bias–Variance Trade-off
- K-Fold Cross Validation
- Regularization Techniques
- Case Study: Regression Model for IPL Dataset

Learning Focus: Predictive modeling using regression techniques.

Module III Textbook & Section Wise Coverage :

- **Textbook** : Machine Learning using Python by Manaranjan Pradhan and U Dinesh Kumar
- **Coverage :**
  - Chapter 4.1, 4.2,4.3, 6.1, 6.2, 6.3, 6.4 from Text book "Machine Learning using Python" by Manaranjan Pradhan and U Dinesh Kumar

Module 4: Classification (9 Hours)

- Overview of Classification Problems
- Binary Logistic Regression
- Credit Classification Case Study
- Model Evaluation: ROC Curve, AUC
- Confusion Matrix
- Optimal Classification Cut-off (Youden's Index)
- K-Nearest Neighbors (KNN)
- Bayes Theorem and Naïve Bayes Classifier
- Text Classification
- Bayesian Belief Networks
- EM Algorithm

Learning Focus: Decision-making and probabilistic classification models.

Module IV Textbook & Section wise Coverage :

- **Textbook :**
  - Machine Learning using Python by Manaranjan Pradhan and U Dinesh Kumar
  - Machine Learning by Tom M Mitchell
- **Coverage :**
  - Chapter 5.1, 5.2, 5.3, 6.5.3, from Text book Machine Learning using Python by Manaranjan Pradhan and U Dinesh Kumar
  - Chapter 6,2, 6.3(6.3.1), 6.7, 6.9, 6.10, 6.11(6.11.1,6.11.2,6.11.3), 6.12 from Text book : Machine Learning by Tom M Mitchell

Module 5: Advanced Machine Learning Algorithms (8 Hours)

- Ensemble Learning Concepts
- Voting Classifiers
- Bagging and Pasting
- Random Patches and Random Subspaces
- Random Forests
- Boosting and Stacking
- Clustering: K-Means Algorithm
- Support Vector Machines (SVM):
  - Linear SVM Classification
  - Nonlinear SVM Classification
  - SVM Regression
  - Decision Functions and Training Objectives

Learning Focus: Advanced models for improved accuracy and generalization.

Module V Textbook & Section wise Coverage :

- Textbook :
  - Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow by Aurelien Geron
  - Machine Learning using Python by Manaranjan Pradhan and U Dinesh Kumar
- Coverage :
  - Chapter 7 : Voting classifiers, Bagging & Pasting, Random Patches & Subspace, Gradient Boosting, Stacking from textbook "Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurelien Geron
  - Chapter 5 : SVM from textbook "Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurelien Geron
  - Random Forest (6.5.5), Adaboost (6.5.6.1) from textbook "Machine Learning using Python" by Manaranjan Pradhan and U Dinesh Kumar

7. Teaching and Learning Methodology

- Chalk-and-talk lectures
- PPT-based explanations
- Demonstration using Python and Scikit-learn
- Case studies and real-world datasets
- Continuous assessment through quizzes and assignments

8. Assessment Scheme : Activity-Based Assessment (ABA)

The following activities are suggested for Activity-Based Assessment (ABA):

- The Assignment component is implemented as a Mini Project, where students apply Machine Learning techniques to any real-world case study for **10 Marks**

- A Seminar on any approved subjective/topic related to Machine Learning is conducted as part of ABA for **10 Marks**

9. Marks Distribution (Integrated Course – CIE)

Table 1 : Marks Distribution

Sl. No.	Evaluation Method	Marks	Weightage (%)
1	Written Test-1	15	60
2	Written Test-2	15	
3	ABA-1	10	40
4	ABA-2	10	
	Total	50	100

10. **Assessment Rules:**

- Each written test shall be conducted for **50 marks and scaled down to 15 marks**.
- Each ABA component shall be conducted for **50 marks and scaled down appropriately**.

11. Attendance Requirements

- Students must maintain a **minimum attendance of 85%** in each registered course.
- The **Principal may condone attendance shortage up to 10%** in special cases.
- Students failing to meet attendance requirements shall be **declared NE (Not Eligible)** for SEE in the concerned course.
- Such students must **re-register and repeat the course** in a subsequent semester.
- Students may appear for SEE in other courses of the same semester if attendance requirements are satisfied.

12. Textbooks

1. Tom M. Mitchell, Machine Learning, McGraw Hill, 2003.
2. Jason Brownlee, Data Preparation for Machine Learning, 2020.
3. Dinesh Kumar, Machine Learning Using Python, Wiley, 2019.
4. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, O'Reilly, 2019.

13. Reference Books

1. Andreas C. Müller and Sarah Guido, *Introduction to Machine Learning with Python*, O'Reilly, 2016.

Online Resources

- NPTEL Course on Machine Learning : http://digimat.in/nptel/courses/video/106105152/L01.html
- YouTube: Machine Learning Full Course : https://www.youtube.com/watch?v=LcWFedjaR4Q

Graduate Attributes Addressed

- Engineering Knowledge
- Problem Analysis
- Modern Tool Usage

Remarks

This course equips students with essential machine learning skills required for advanced AI courses, industry projects, and research in data-driven domains.

Prepared by: Department of CS&E