INT 395
Supervised Learning

Lecture #0

Course details

- **LTP – 2 0 2**
- **Credits – 3.0**
- **Mode – BYOD**

Course Assessment Model

- **Marks break up***
- **Attendance** 5
- **CA** 25

- **MTE (MCQ based)** **20**
- **ETE (MCQ+ Subjective)** 50
- **Total** **100**

Course Assessment

CA Category: A0202

- CA1: Project
- CA2: Code Based Test

Project Rubrics

1. Report / Documentation -10
2. Execution / Implementation -10
3. Presentation / Viva - 10

Course Outcomes

CO1: apply preprocessing techniques to prepare and transform data for machine learning models.

CO2: describe standard classification algorithms and interpret evaluation metrics for model assessment.

CO3: explain ensemble learning techniques and evaluate hyperparameter tuning methods for model optimization.

CO4: analyze and apply regression techniques for predictive modeling.

CO5: examine time series data characteristics and evaluate classical forecasting models.

CO6: use pipeline techniques and model evaluation strategies to streamline machine learning workflows.

Program Outcomes

PO1::Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2:: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3::Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for public health and safety, and cultural, societal, and environmental considerations.

PO4::Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5::Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

Program Outcomes

PO6::Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7:Understand the impact of professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8::Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9::Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10::Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

Program Outcomes

PO11::Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

Program Educational Objectives (PEO)

- The graduates shall demonstrate professional advancement through expanded leadership capabilities and technical accomplishment providing solutions to local and global societal issues through mindful engagement.
- The graduates shall undertake higher education or global certifications or exhibit impactful research accomplishments.
- The graduates shall extend global technology development and deployment expertise by becoming entrepreneurs, consultants, and innovators.
- Graduates shall embrace ethics and lifelong learning to adapt to a fast-changing world and enhance global employability in diverse work environments.

PSO’s

PSO1::Apply acquired skills in software engineering, networking, security, databases, intelligent systems, cloud computing, and operating systems to adapt and deploy innovative software solutions for diverse applications.

PSO2::Apply diverse IT skills to design, develop, and evaluate innovative solutions for business environments, considering risks, and utilizing interdisciplinary knowledge for efficient real-time projects benefiting society.

Vision

Vision

To become one of the leading Schools globally in Computer Science and Engineering recognized for its academics and innovations by nurturing professionals, researchers and entrepreneurs for sustainable growth of industry and society.

Mission

Mission

M1: To provide a learning-based environment on technical concepts applied to real-life situations with measurable outcomes.
M2: To establish connections with the industry for curriculum design, and creating internship cum career opportunities.
M3: To address societal issues related to regional, national and global challenges through meaningful research.
M4: To inspire graduates for pursuing lifelong learning in professional careers.
M5: To develop leadership potential in ethically competent entrepreneurs.

The course contents: Unit 1

- Introduction : overview of supervised learning and use cases, types of supervised learning, setting up python and scikit-learn environment, types of data, loading datasets, data exploration with Pandas and matplotlib/ seaborn
- Data Preprocessing : common data issues, handling missing values and outliers, handling class imbalance, scaling features, encoding categorical variables, Feature engineering, Feature selection, data splitting, dimensionality reduction

The course contents: Unit 2

- Classification with scikit-learn : Overview of classification, evaluation metrics- Accuracy, Precision, Recall, F1-Score, ROC-AUC, confusion matrix, perceptron, logistic regression, k-Nearest Neighbors, Decision Tree, Support Vector Machine, Naïve Bayes

The course contents: Unit 3

- Ensemble Methods and Hyperparameter Tuning : Overview of ensemble models, how ensembles can reduce error, Bagging, Boosting, Stacking, need for hyperparameter tuning, grid search, random search

The course contents: Unit 4

- Regression with scikit-Learn : Introduction to regression, exploratory data analysis, evaluation metrics, linear regression, RANSAC model, polynomial regression, regularized regression, support vector regression, decision tree regression, random forest regression

The course contents: Unit 5

- Time Series Regression : Time series data characteristics, univariate vs multivariate, regular vs irregular intervals, preparing Time Series data, Data splitting strategies, autoregressive models, moving average models, ARIMA, SARIMA

The course contents: Unit 6

- Pipelines and Model Evaluation : Streamlining workflows with Pipelines, Cross Validation strategies, debugging algorithms with learning curves and validation curves
- Model Deployment : importance of model deployment, model serialization and deserialization, local deployment, Web service deployment, Serverless options

List of Practical

- Write a program to explore and visualize a dataset using Pandas and Seaborn.
- Write a program to preprocess data by handling missing values, outliers, scaling, and encoding.
- Write a program to perform feature engineering and feature selection using statistical methods and PCA.
- Write a program to split data into training and testing sets and apply k-fold cross-validation.
- Write a program to implement logistic regression for binary classification.
- Write a program to build a decision tree classifier and evaluate it using a confusion matrix.
- Write a program to compare the performance of k-NN, SVM, and Naïve Bayes classifiers.

- Write a program to evaluate a classification model using precision, recall, F1-score, and ROC-AUC.
- Write a program to implement Random Forest classification using Bagging technique.
- Write a program to apply boosting algorithms like AdaBoost and Gradient Boosting for classification.
- Write a program to tune hyperparameters using Grid Search and Random Search methods.
- Write a program to implement linear, polynomial, and ridge regression models for prediction.
- Write a program to forecast time series data using ARIMA or SARIMA models.
- Write a program to deploy a machine learning model using Pickle and Streamlit.

Text and Reference Books

Text Books:

1. MACHINE LEARNING-I by CHANDRA S.S, VINOD, PHI Learning

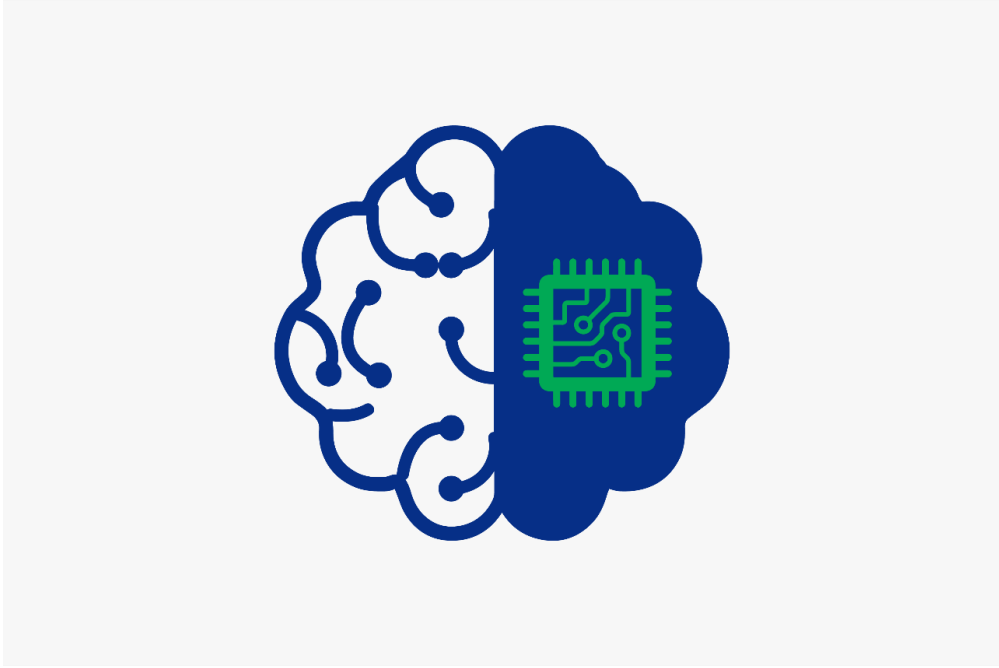
References:

1. MACHINE LEARNING by ETHEM ALPAYDIN, MIT Press

MOOC

- Introduction to Machine Learning
- By Prof. Balaraman Ravindran, IIT Madras
- https://onlinecourses.nptel.ac.in/noc23_cs18/preview

Machine Learning



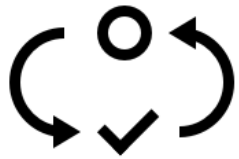
TYPES OF MACHINE LEARNING

Supervised Learning

Train an algorithm to perform classifcation and regression with a labelled data set.

Unsupervised Learning

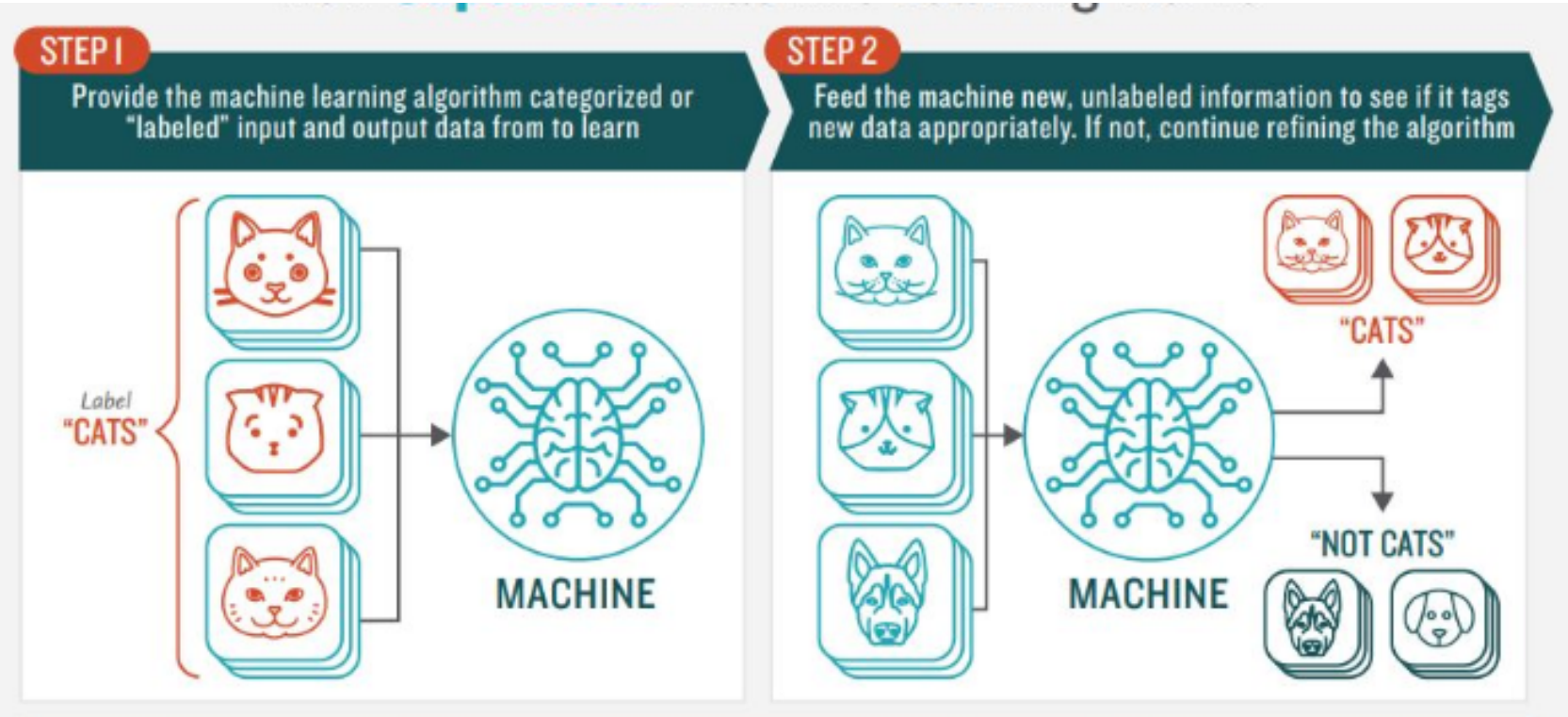
Train an algorithm to find clusters and associations in an unlabelled data set.

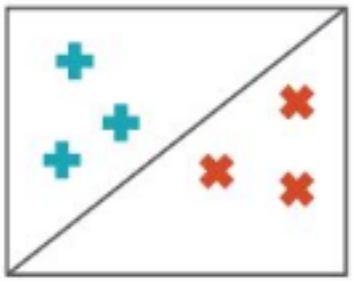


Reinforcement Learning

Train an agent to take certain actions in an environment without a data set.

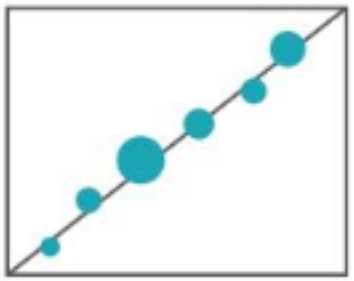
Classification





CLASSIFICATION

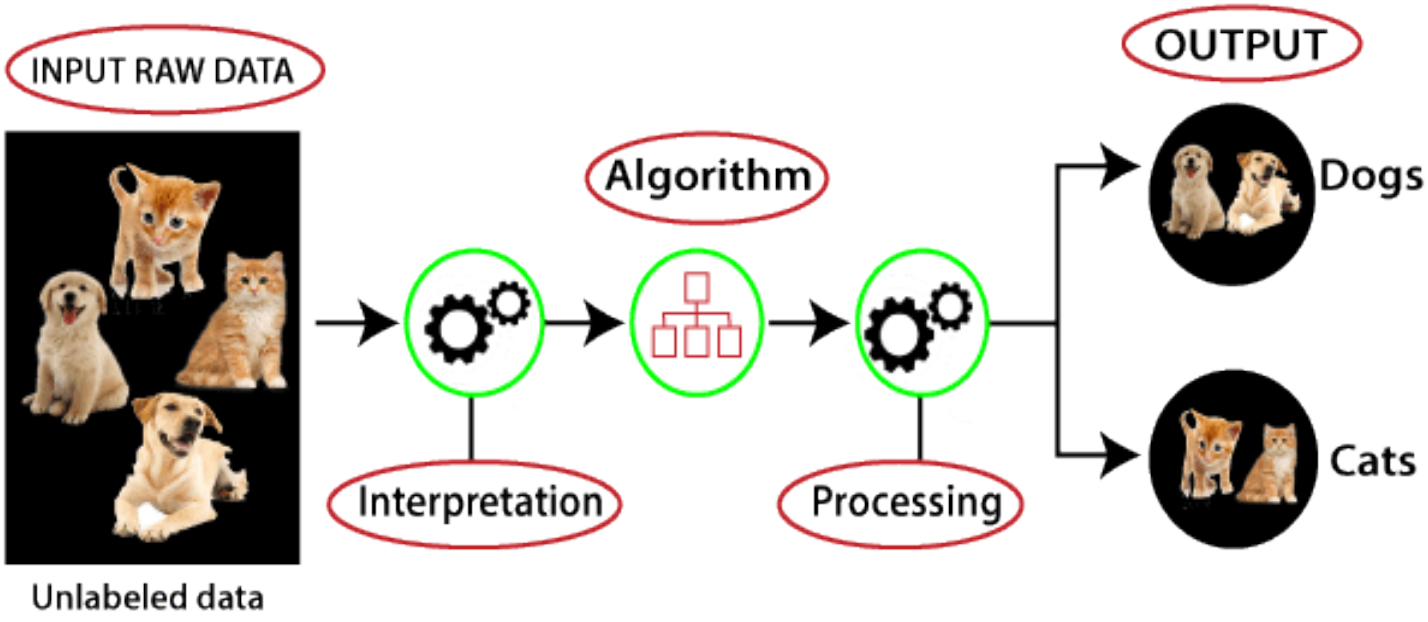
Sorting items into categories

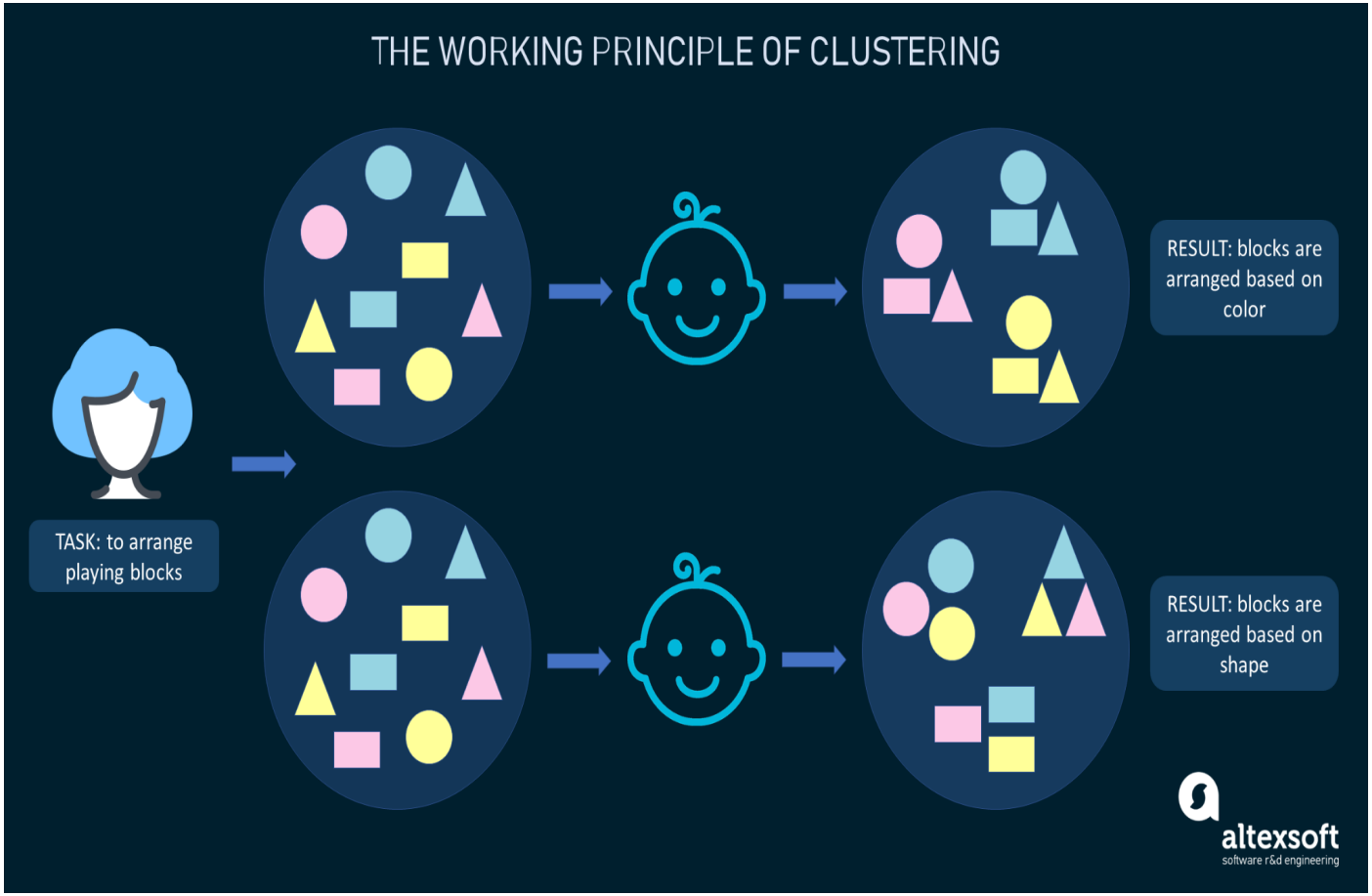


REGRESSION

Identifying real values (dollars, weight, etc.)

Unsupervised Machine Learning Example




THE WORKING PRINCIPLE OF CLUSTERING
TASK: to arrange playing blocks
RESULT: blocks are arranged based on color
RESULT: blocks are arranged based on shape
altexsoft
software r&d engineering

Edu-Revolution: Be the Change Initiatives

Category	Criteria
Recognition of Prior Learning (RPL)	To avail the benefits under RPL, the student must possess a certificate related to a Machine Learning course. Once the certificate is approved by the Machine Learning domain team, the student will be required to appear for a Benchmark test to qualify for the benefit.
MOOC: NPTEL/SWAYAM	https://onlinecourses.nptel.ac.in/noc23_cs18/preview Introduction to Machine Learning By Prof. Balaraman Ravindran, IIT Madras

Innovative Pedagogies

Strategy	Application in the Course
Case-based Learning	Unit III, IV and V To implement real-world examples like Parkinson Disease detection using Classifiers or Weather Prediction using regressor
Flipped Classrooms	Pre-recorded videos and reading material shared before class; live sessions for hands-on implementation and doubt clearance.
Quiz	At the end of each Unit
Events/ Competitions / Workshops	Hackathons like Smart India Hackathon, Kaggle competitions, and in-house ML events

Open Educational Resource(OER)

Unit Mapped	Broad Topic	Sub Topic	Source Type	Source Title	%age Mapped (approx)	Source URL
Unit I	Introduction to Supervised Learning & Data Preprocessing	Overview of supervised learning and use cases, types of supervised learning	Web Link	Introduction to Machine Learning (Prof. Balaraman Ravindran, IIT Madras) - Week 1: Introduction	100%	https://onlinecourses.nptel.ac.in/noc23_cs18/preview
		Setting up Python and scikit-learn environment	Web Link	Python for Data Science (Prof. Ragunathan Rengasamy, IIT Madras) - Covers Python basics and relevant libraries	70%	https://nptel.ac.in/courses/106106212
		Types of data, loading datasets, data exploration with Pandas and matplotlib/seaborn	Web Link	Python for Data Science (Prof. Ragunathan Rengasamy, IIT Madras) - Week 2-4: Data Handling with Pandas & Visualization	90%	https://nptel.ac.in/courses/106106212
		Common data issues, handling missing values and outliers, handling class imbalance, scaling features, encoding categorical variables, Feature engineering, Feature selection, data splitting, dimensionality reduction	Web Link	Machine Learning (Prof. S. Sarkar, IIT Kharagpur) - Data Preprocessing Module	80%	https://nptel.ac.in/courses/106105152
		Dimensionality Reduction (PCA)	Web Link	Introduction to Machine Learning (Prof. Balaraman Ravindran, IIT Madras) - Week 2: Principal Component Analysis	90%	https://onlinecourses.nptel.ac.in/noc23_cs18/preview

Unit Mapped	Broad Topic	Sub Topic	Source Type	Source Title	%age Mapped (approx)	Source URL
Unit II	Classification with scikit-learn	Overview of classification, evaluation metrics (Accuracy, Precision, Recall, F1-Score, ROC-AUC, confusion matrix)	Web Link	Introduction to Machine Learning (Prof. Balaraman Ravindran, IIT Madras) - Week 1: Evaluation Metrics	90%	https://onlinecourses.nptel.ac.in/noc23_cs18/preview
		Perceptron, Logistic Regression, k-Nearest Neighbors, Decision Tree, Support Vector Machine, Naïve Bayes	Web Link	Introduction to Machine Learning (Prof. Balaraman Ravindran, IIT Madras) - Weeks 3-8: Classification Algorithms	80%	https://onlinecourses.nptel.ac.in/noc23_cs18/preview
		Classification Algorithms (Alternative Course)	Web Link	Machine Learning (Prof. S. Sarkar, IIT Kharagpur) - Classification Module	90%	https://nptel.ac.in/courses/106105152

Unit Mapped	Broad Topic	Sub Topic	Source Type	Source Title	%age Mapped (approx)	Source URL
Unit III	Ensemble Methods and Hyperparameter Tuning	Overview of ensemble models, how ensembles can reduce error, Bagging, Boosting, Stacking	Web Link	Introduction to Machine Learning (Prof. Balaraman Ravindran, IIT Madras) - Week 7: Ensemble Methods	90%	https://onlinecourses.nptel.ac.in/noc23_cs18/preview
		Need for hyperparameter tuning, grid search, random search	Web Link	Introduction to Machine Learning (Prof. Balaraman Ravindran, IIT Madras) - Week 7: Cross Validation & Hyperparameter Tuning	80%	https://onlinecourses.nptel.ac.in/noc23_cs18/preview

Unit Mapped	Broad Topic	Sub Topic	Source Type	Source Title	%age Mapped (approx)	Source URL
Unit IV	Regression with scikit-Learn	Introduction to regression, exploratory data analysis, evaluation metrics	Web Link	Introduction to Machine Learning (Prof. Balaraman Ravindran, IIT Madras) - Week 1-2: Regression & Evaluation	90%	https://onlinecourses.nptel.ac.in/noc23_cs18/preview
		Linear regression, RANSAC model, polynomial regression, regularized regression, support vector regression, decision tree regression, random forest regression	Web Link	Introduction to Machine Learning (Prof. Balaraman Ravindran, IIT Madras) - Week 2: Linear Regression, Week 6: Regression Trees	90%	https://onlinecourses.nptel.ac.in/noc23_cs18/preview
		Regression (Alternative Course)	Web Link	Machine Learning (Prof. S. Sarkar, IIT Kharagpur) - Regression Module	80%	https://nptel.ac.in/courses/106105152

Unit Mapped	Broad Topic	Sub Topic	Source Type	Source Title	%age Mapped (approx)	Source URL
Unit V	Time Series Regression	Time series data characteristics, univariate vs multivariate, regular vs irregular intervals, preparing Time Series data, Data splitting strategies, autoregressive models, moving average models, ARIMA, SARIMA	Web Link	Applied Time Series Analysis (Prof. Arun K. Tangirala, IIT Madras) - Comprehensive course on Time Series	90%	https://nptel.ac.in/courses/103106147
		Time Series Forecasting	Web Link	Introduction to Machine Learning (Prof. Balaraman Ravindran, IIT Madras) - May touch upon basic time series concepts in advanced topics	60%	https://onlinecourses.nptel.ac.in/noc23_cs18/preview

Unit Mapped	Broad Topic	Sub Topic	Source Type	Source Title	%age Mapped (approx)	Source URL
Unit VI	Pipelines and Model Evaluation & Deployment	Streamlining workflows with Pipelines	Web Link	While not a dedicated lecture on scikit-learn pipelines, the concept of workflow management is covered in practical ML discussions. Refer to the overall ML courses for best practices.	50%	https://onlinecourses.nptel.ac.in/noc23_cs18/preview
		Cross Validation strategies, debugging algorithms with learning curves and validation curves	Web Link	Introduction to Machine Learning (Prof. Balaraman Ravindran, IIT Madras) - Week 7: Cross Validation, Learning Curves	90%	https://onlinecourses.nptel.ac.in/noc23_cs18/preview

THANK YOU