



# Data Foundations and Machine Learning Applications (Level 2)

## Learning Mode:

Online via Zoom

## Duration:

5 Days

## This course is specifically designed to provide participants with:

- A solid foundation in data handling, cleaning, and visualization using Python
- A clear understanding of machine learning workflows, from problem definition to model evaluation
- Practical experience with classification, regression, and clustering models for real-world datasets
- Exposure to deep learning concepts, including CNNs for image classification tasks
- Applied knowledge of time series forecasting, including feature engineering and sequence models (LSTM/GRU)
- Hands-on experience in building and evaluating end-to-end ML and DL pipelines
- The ability to interpret model performance and communicate insights effectively

## Course Overview

### SUMMARY

Data Foundations and Machine Learning Applications (Level 2) is a comprehensive, hands-on course designed to equip participants with practical skills in data analysis, machine learning, and selected deep learning applications using Python.

The course begins with strong data foundations, enabling participants to understand different data types, perform data cleaning, visualization, and basic exploration. It then progressively introduces core machine learning concepts, covering supervised learning for classification and regression, as well as unsupervised learning through clustering techniques.

Participants are further exposed to deep learning fundamentals, with a focused introduction to Convolutional Neural Networks (CNNs) for image-based applications. The course also addresses one of the most in-demand real-world use cases—time series forecasting, introducing both traditional baseline methods and modern sequence models such as LSTM and GRU.

The program concludes with an end-to-end time series forecasting project, allowing participants to integrate data preparation, model development, evaluation, and result interpretation in a structured, real-world scenario. This ensures participants not only understand theoretical concepts but can confidently apply them to practical problems.

## Course Content

### MODULE 1

#### Data Foundations & Introduction To Machine Learning

- Course overview and learning outcomes
- Types of data: numerical, categorical, time series, images
- Basic data handling in Python (load, inspect, summarize)
- Data cleaning: missing values, duplicates, simple transformations
- Data visualization: distributions, relationships, correlations
- What is Machine Learning and Deep Learning
- ML workflow: problem, data, model, evaluation, iteration

### MODULE 2

#### Supervised Learning – Classification

- Supervised learning concepts: features, labels, train/test split
- Classification problem types and examples
- Basic classification models:
  - Logistic Regression (concept + simple example)
  - k-Nearest Neighbors (k-NN)
  - Decision Tree
  - Tree Ensemble
- Model evaluation for classification:
  - Accuracy, confusion matrix
  - Precision, recall, F1-score (concept)
- Hands-on end-to-end classification examples

### MODULE 3

#### Regression & Unsupervised Learning (Clustering)

- Regression problem definition and use cases
- Basic regression models:
  - Linear Regression
  - k-Nearest Neighbors (k-NN)
  - Tree Ensemble
- Regression evaluation metrics: MAE, MSE, RMSE, R<sup>2</sup>
- Introduction to unsupervised learning
- Clustering with k-Means: intuition, how it works
- Clustering example with 2D visualization and interpretation





# Data Foundations and Machine Learning Applications (Level 2)

## Course Overview

### MODULE 4

#### Deep Learning With Cnn For Images

- From traditional ML to Deep Learning: why neural networks
- Neuron, layers, activation functions
- CNN fundamentals:
  - Convolution, filters, feature maps
  - Pooling, flattening, fully connected layers
- Image data preparation: resizing, normalization, train/validation split
- Building a simple CNN for image classification
- Training and monitoring performance (loss, accuracy curves)
- Overfitting and basic regularization (dropout, simple augmentation – concept)

### MODULE 5

#### Time Series Forecasting

- What is time series data and where it appears (energy, finance, sensors)
- Key concepts: trend, seasonality, noise, forecast horizon
- Train/validation/test split for time series (no shuffling)
- Feature construction for time series: lags, windows
- Baseline forecasting methods (naive last value, simple moving average)
- Introduction to sequence models for time series:
  - RNN concept (idea of memory)
  - LSTM and GRU (high-level gates and intuition)
- LSTM/GRU examples for one-step/multi-step forecasting

### MODULE 5

#### Recap & Time Series Forecasting Project

- Structured recap of the course:
- data → classification → regression & clustering → CNN → time series
- End-to-end ML/DL pipeline revisited using a time series case
- Time series project work:
  - Problem definition and dataset overview
  - Short EDA and preprocessing review
  - Model selection (baseline vs LSTM/GRU)
  - Training, evaluation, and plotting predictions vs actual
- Group or individual mini-presentations / discussion of project results
- Real-world considerations: data quality, domain knowledge, deployment concept
- Next steps for participants and Q&A