

Week O2 – Getting Started with Machine Learning

Overview of Modules 05, 06, 6.5, 07, 7.5, and 08



What Week 02 Is About

This week builds your machine learning foundation. You'll explore what ML is, learn data preparation for computers, and practice data exploration. By the end, you'll understand how to transform real-world data for machine learning models.

We'll cover key concepts and hands-on tasks, with each module building on the last.

Module O5 – Introduction to ML

What is Machine Learning?

Machine learning teaches computers to learn patterns from data, rather than explicit programming.

Three Main Types:

Supervised: Learning from labeled examples

Unsupervised: Finding hidden patterns

Reinforcement: Learning through trial and reward

Core Ideas

Features: Input data

Labels: Answers to predict

Training: Teaching the model

Testing: Checking performance

Generalization: Applying to new data



Data Collection

Gather data



Training

Teach model

Evaluation

Test performance

Common ML Challenges

Understanding these common pitfalls helps build better, more reliable models.

Bias

Training data doesn't represent the real world fairly, causing the model to learn incorrect patterns.

Overfitting

The model memorizes training data and noise instead of general patterns. Good training performance, but poor on new data.

Underfitting

The model is too simple to capture important patterns in the data. Poor performance on both training and new data.

Module O6 – Data Preprocessing & Feature Engineering

Raw data needs cleaning, transforming, and preparation for effective machine learning. This module covers these essential steps.

O1

Handling Missing Values

Address gaps in data by filling or removing incomplete entries.

O2

Encoding Categorical Features

Transform text categories into numerical formats for model processing.

O3

Scaling Features

Normalize feature ranges to ensure comparability for models.

O4

Detecting Outliers

Identify unusual data points.

O5

Feature Engineering

Generate new features from existing data to enhance model performance.

Why Preprocessing Matters

Quality Data = Quality Predictions

Machine learning models depend entirely on the quality of their input data. Poor, inconsistent, or biased data will lead to unreliable predictions, regardless of the model's sophistication. Preprocessing ensures clean, well-prepared data for better model performance and trustworthy results.

The Golden Rule:

Garbage in → Garbage out

80%
of ML work

is data preparation, not modeling

3X
improvement

possible with proper preprocessing

Module 6.5 – Mini Colab Task

A practical exercise to apply learned concepts in a real coding environment.



Load 3 Tiny CSVs

Practice reading different datasets.



Identify Features & Target

Determine input features and target variables.



Classify Task Type

Identify as classification or regression.



Sketch ML Pipeline

Outline the data processing and modeling steps.

 **Pro Tip:** Take your time with this task. Add comments in your notebook to explain your thinking.

Module O7 – Exploratory Data Analysis (EDA)

EDA is exploring data to understand its characteristics, uncover patterns, and identify potential issues before model building.

Key EDA Activities

Distributions

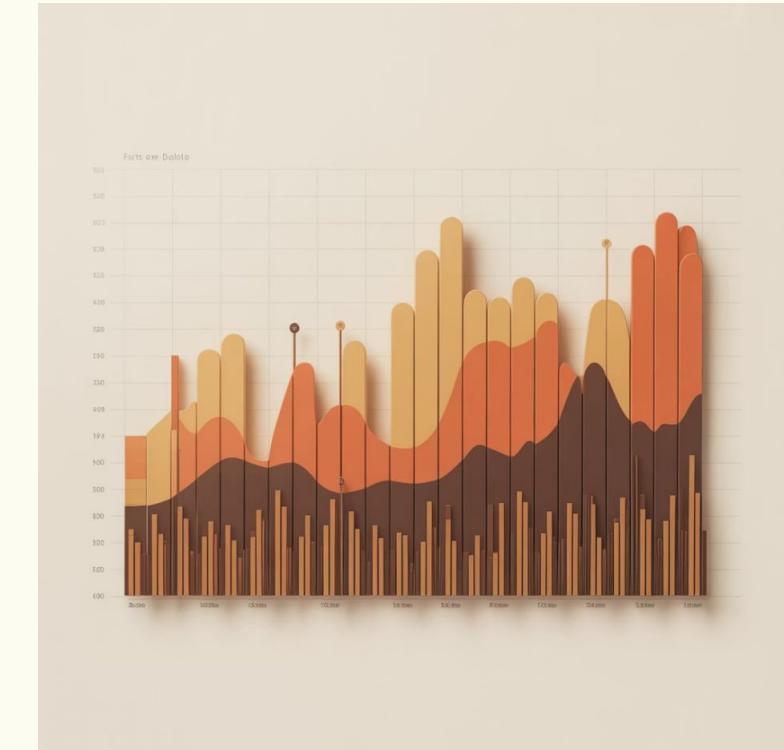
Analyze how data spreads to find balance, skew, or clusters.

Patterns & Correlations

Identify relationships between features and the target variable.

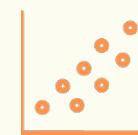
Imbalances

Detect uneven representation in categories for proper handling.



Histograms

Visualize single variable distributions.



Scatter Plots

Show relationships between two variables.



Correlation Heatmaps

Display all pairwise correlations.

Module 7.5 – EDA Audit in Colab

Perform a data audit in Python to inspect data quality and plan your preprocessing strategy.

1 Profile Missingness

Identify and count missing values in each column to understand data incompleteness.

2 Inspect Categorical Levels

Examine unique categories and their frequencies in text-based columns.

3 Create Preprocessing Checklist

Document necessary data cleaning steps, from missing values to feature encoding.

 **Remember:** Clear documentation saves time and guides future work.

Module O8 – Week O1 Review Quiz/Assignment

This quiz reinforces Week 02 concepts, focusing on practical application.



Quiz Topics Include:

- **Mini ML task classification**
- **EDA interpretation**
- **CSV preprocessing**
recommendations

Use this quiz as a learning tool to identify concepts needing review.

Test Your Understanding

What You'll Be Able to Do After Week 02

This week equips you with essential practical skills for your machine learning journey:

Understand ML Basics

Grasp ML fundamentals, task types, and the training process.

Preprocess & Clean Data

Handle missing values, encode categories, scale features, and detect outliers.

Explore Datasets Visually

Use plots to understand data distributions, find patterns, and identify problems.

Create Your First Pipeline

Map complete data processing workflows from raw inputs to model-ready data.

By Week 02, you'll have the toolkit and confidence to prepare real-world datasets for machine learning.



Ready to Begin Your ML Journey?

You've seen what Week 02 offers. Now it's time to dive in. Embrace experimentation and learning from mistakes.

ML becomes manageable by breaking it into steps. Soon, complex concepts will feel natural.

Your first step: **Start Module 05**. Let's build something amazing! 