

Machine Learning in Finance

Prepared by Dr. Sumaya Alkubaisi

FNT633

Machine Learning in Finance

Lecture 0



Agenda

Introduction

- Introduction to topics we will cover in this course.
- High-level overview to build a roadmap.

Chapter Overviews

- **a. Chapter 1: Introduction to Machine Learning and Finance** - Explore how machine learning is applied in finance and understand the life cycle of data-driven solutions, including key financial data types and sources.

Agenda

- **b. Chapter 2: Data Preparation I** - Learn how to clean and preprocess financial data to ensure it's accurate, consistent, and ready for analysis.
- **c. Chapter 3: Data Processing and Feature Engineering II** - Discover techniques to integrate multiple datasets and transform features for better model performance.

Agenda

- **d. Chapter 4: Statistical Foundations for Data Science** - Understand the statistical methods and visualization tools used to summarize, explore, and interpret financial data.
- **e. Chapter 5: Supervised Learning Fundamentals** - Learn how to train models to predict financial outcomes using classification and regression techniques.

Agenda

- **f. Chapter 6: Decision Trees and Random Forests** - Explore tree-based algorithms for prediction, risk assessment, and decision-making in finance.
- **g. Chapter 7: Association Rule Learning and Naïve Bayes** - Discover patterns in data and predict outcomes using probabilities for applications like fraud detection and market analysis.

Agenda

- **h. Chapter 8: Unsupervised Learning in Finance** - Understand how to group financial data into patterns or segments without labeled outcomes.
- **i. Chapter 9: Time Series Analysis I** – Models and Forecasting: Explore methods to analyze and predict financial data that changes over time, like stock prices or sales.
- **j. Chapter 10: Time Series Analysis II** – Models and Forecasting: Continue learning advanced time series techniques for accurate forecasting and trend analysis in finance.

Course Roadmap & Expectations

Closing Remarks

Course Description

- This course provides a comprehensive introduction to machine learning and its applications in the financial sector.
- Students will learn about various machine learning techniques, from fundamental concepts to advanced algorithms, and how these can be applied to solve real-world economic problems.

Course Description

- The course covers data processing, feature engineering, supervised and unsupervised learning, focusing on their relevance to financial forecasting, credit scoring, risk management, algorithmic trading, and regulatory compliance.

Course Description

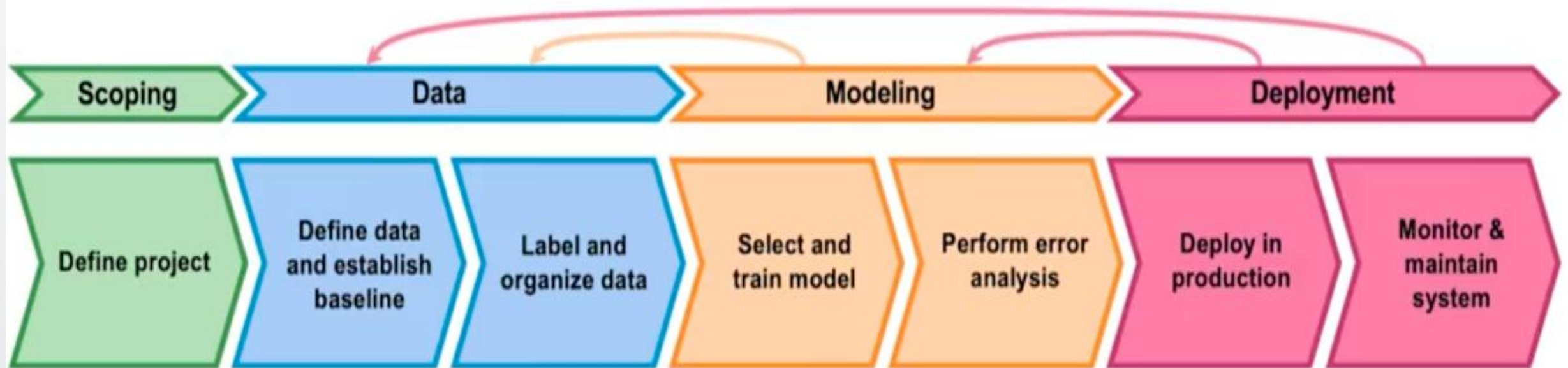
- Through lectures and case studies, participants will gain practical experience implementing machine learning models using economic data. The course aims to equip students with the skills to analyze market behaviors, enhance decision-making processes, and innovate with new financial technologies.

Introduction to ML & Finance

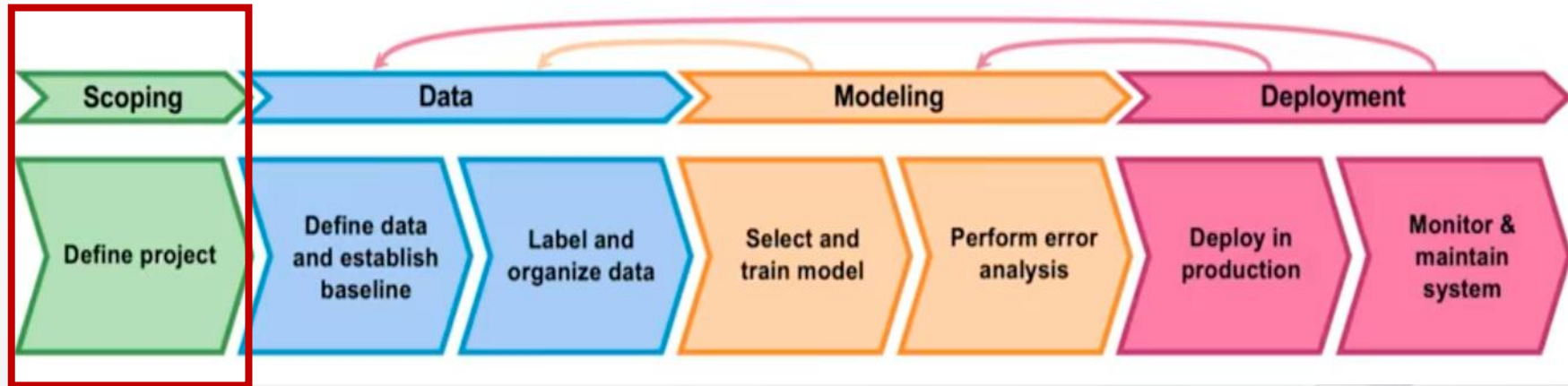
- **Machine Learning (ML)** = computers learn from data, not fixed rules.
- Used in finance for predictions, automation, and fraud detection.

Example: Predict loan approval chances based on income & credit score.

The ML project lifecycle

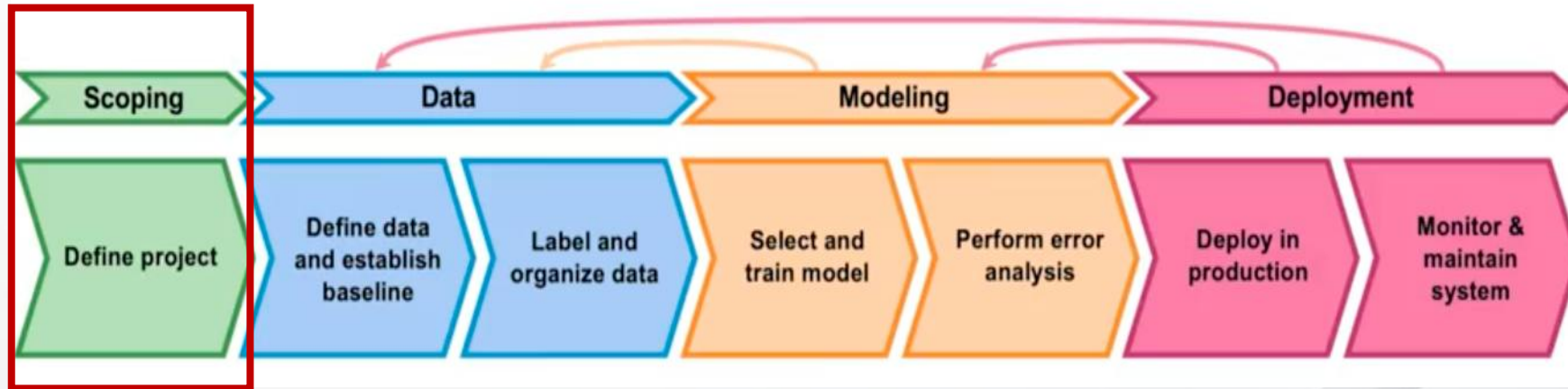


The ML project lifecycle



- What is Scoping? The Scoping stage is the foundational phase in which the following are defined:
- What problem are we trying to solve? Is it a classification, regression, or prediction problem?
- What is the practical objective of the model? Examples: Predicting loan default, Detecting financial fraud, Classifying emails as Spam or Not Spam

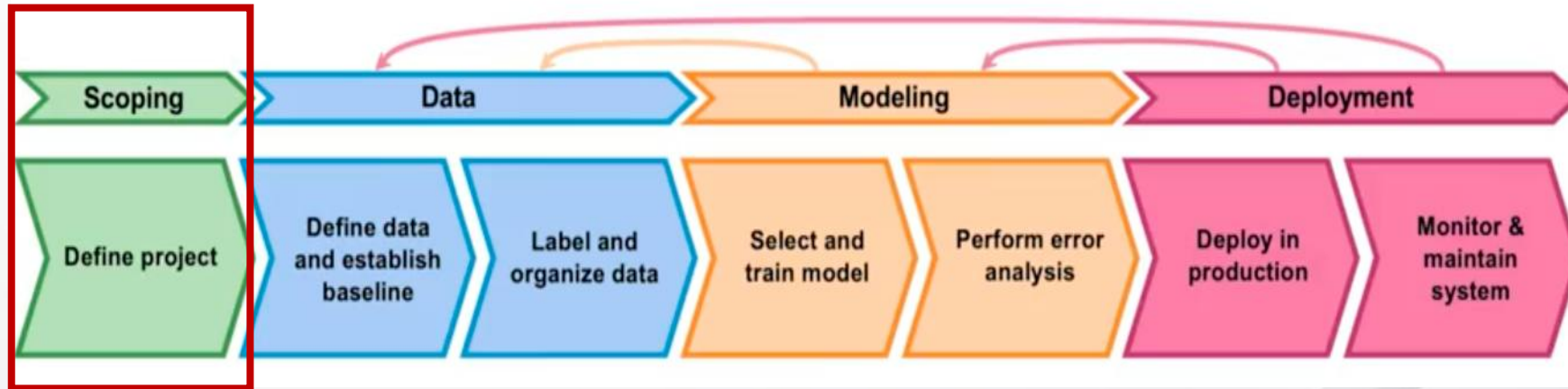
The ML project lifecycle



Key Decisions in Scoping: In this stage, we determine:

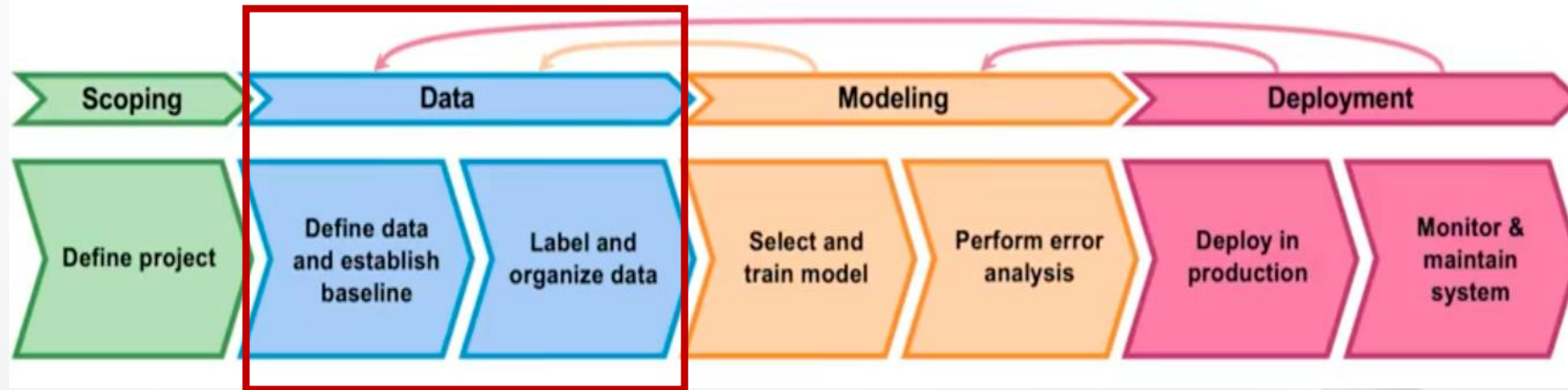
- Input features (Features)
- Output variables (Labels)
- Success criteria (Evaluation Metrics)
- Constraints (time, data availability, resources)

The ML project lifecycle



- Any mistake made at this stage will affect all subsequent stages.

The ML project lifecycle

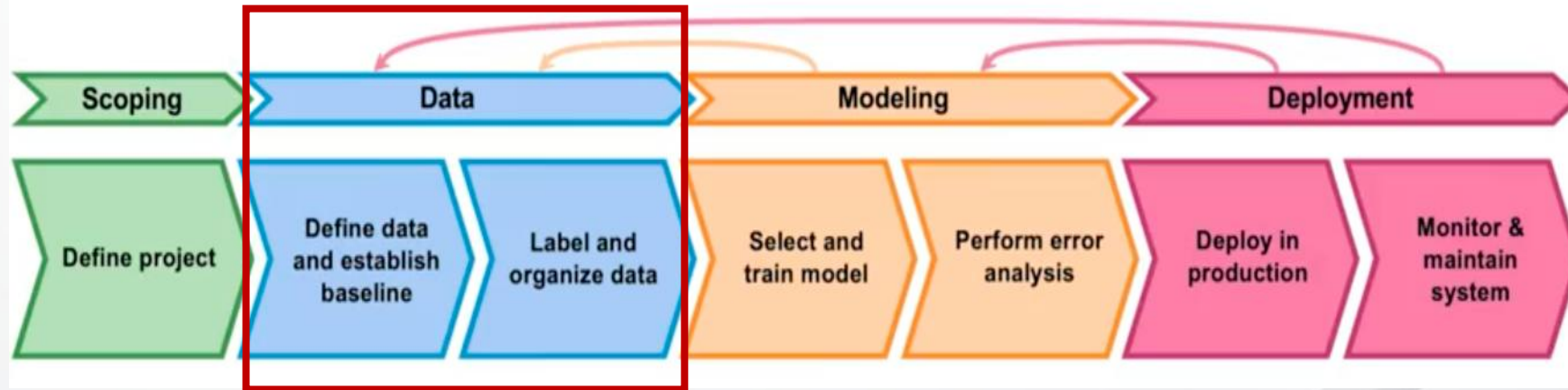


Data Collection

This stage includes:

- Collecting data from multiple sources
- Assessing data quality
- Handling: Missing values, Outliers

The ML project lifecycle

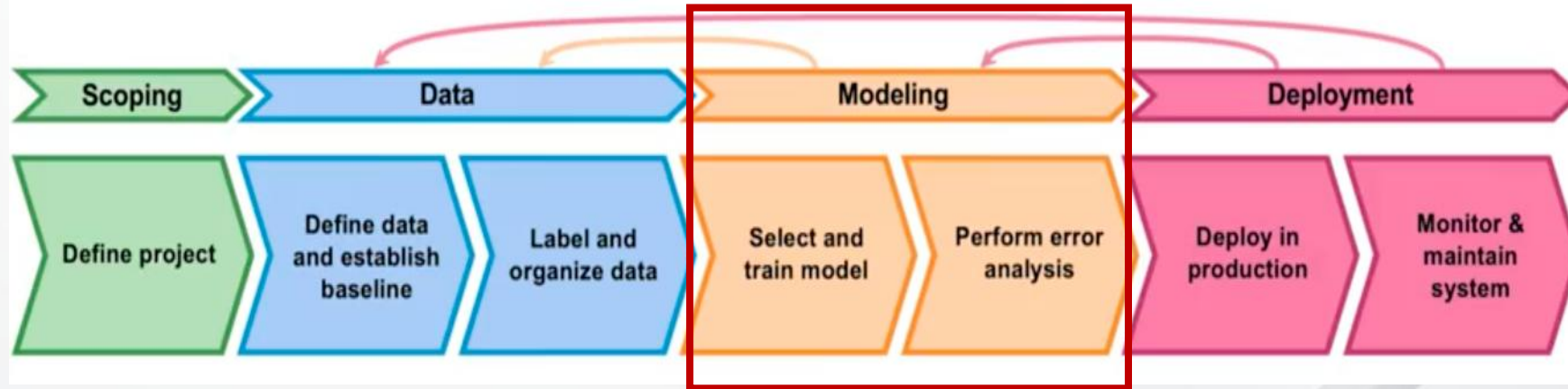


Data Labeling and Organization

- Organizing data into a format suitable for modeling
- Encoding data
- Splitting the dataset into: Training set and Test set

Data quality is often more important than model complexity.

The ML project lifecycle

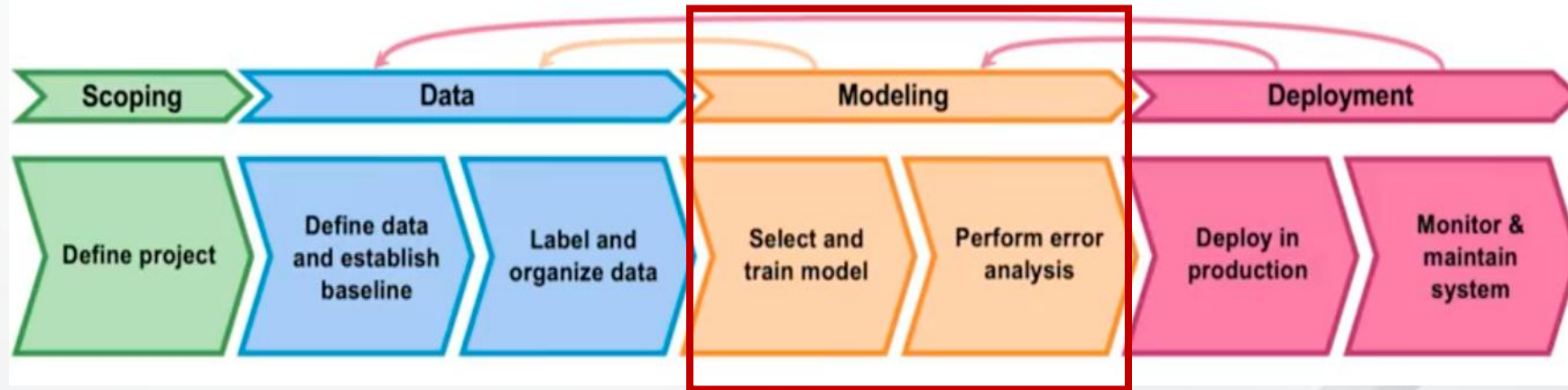


Model Selection and Training

In this stage:

- Selecting an appropriate algorithm (e.g., Logistic Regression, Decision Trees, Neural Networks)
- Training the model using the training data
- Tuning hyperparameters

The ML project lifecycle



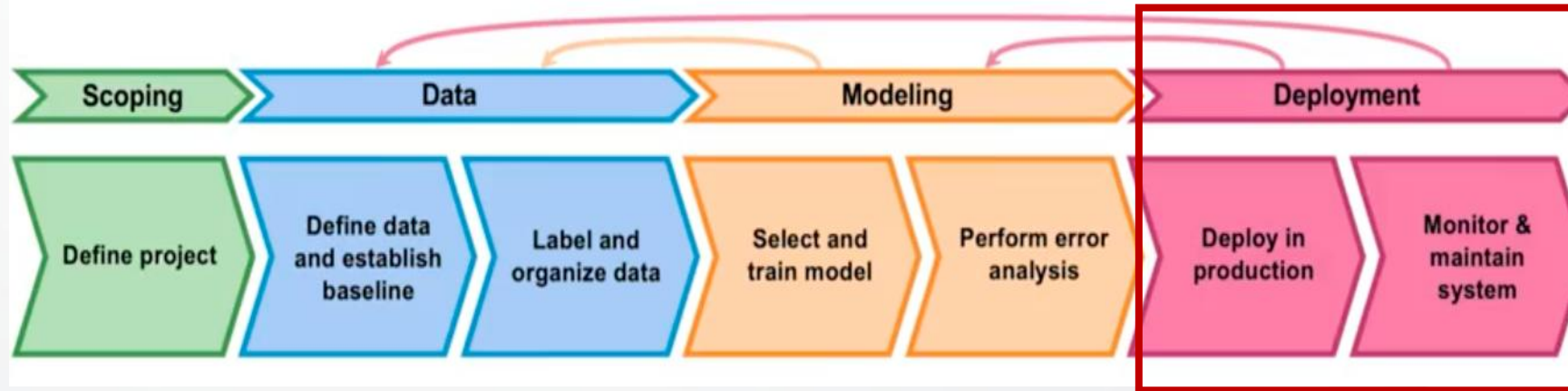
Error Analysis and Evaluation

We perform:

- Error analysis (False Positives / False Negatives)
- Performance evaluation using: Accuracy, Precision, Recall, F1-score

This is where the true understanding of the model's strengths and weaknesses emerges.

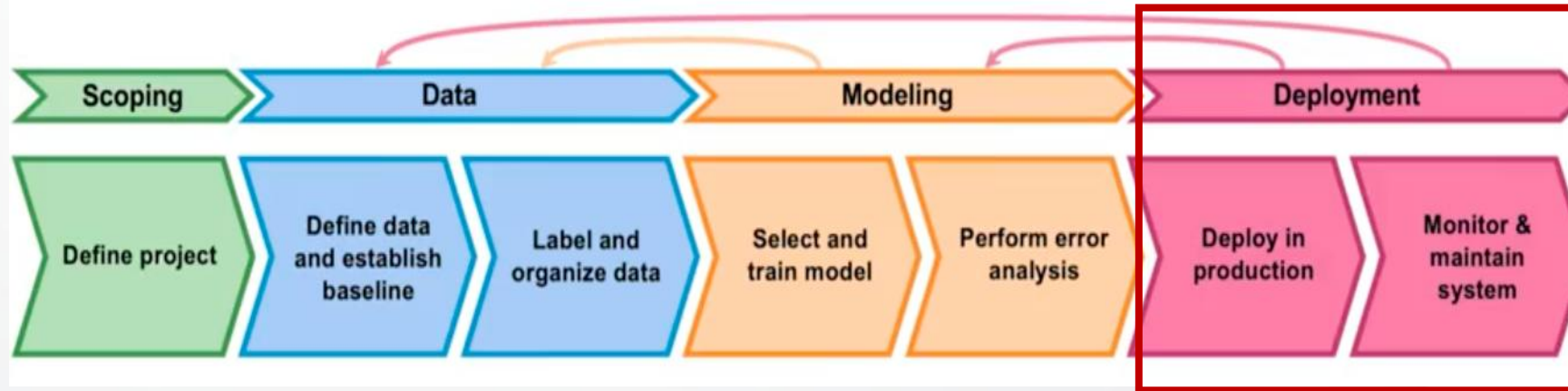
The ML project lifecycle



What Does Deployment Mean? – (Out of Scope for This Course)

- In real-world projects, deployment involves:
- Integrating the model into a real system
- Running the model on live data
- Ensuring system performance and stability

The ML project lifecycle



What We Will Do Instead?

In this course: We will not deploy models in a production environment,
Instead, we will:

- Run models on Google Colab
- Analyze results
- Interpret performance and make improvement decisions

Data Preparation (Cleaning & Preprocessing)

- Raw financial data often has missing values, duplicates, and errors.
- Preprocessing ensures clean, consistent datasets.

Example: Remove empty entries in customer loan applications.

Data Processing & Feature Engineering

- **Integration:** Combine data from multiple sources (e.g., stock + economic reports).
- **Normalization:** Scale values for fair comparison.

Example: Normalize transaction amounts to detect unusual spending patterns.

Statistical Foundations

- Use descriptive statistics (mean, median, variance) and inferential statistics (hypothesis tests).
- Visualize data with charts before modeling.

Example: Plot stock price distribution to check volatility.

Supervised Learning Fundamentals

- Train models with labeled data (input → output).
- Algorithms: K-Nearest Neighbors (KNN), Decision Trees.

Example: Predict if a customer will default on a loan (Yes/No).

Decision Trees & Random Forests

- **Decision Tree:** Splits data into rules (if-else).
- **Random Forest:** Combines multiple trees for stronger predictions.

Association Rule Learning & Naïve Bayes

- **Association Rules:** Find relationships in data (e.g., products bought together).
- **Naïve Bayes:** Simple, fast classifier based on probability.

Example: Detect unusual stock movements linked to news headlines.

Unsupervised Learning & Time Series

- **Unsupervised Learning:** Group unlabeled data (e.g., customer segmentation).
- **Time Series Analysis:** Forecast values based on time patterns.

Examples:

- Cluster investors by trading behavior.
- Predict future stock prices from past trends.

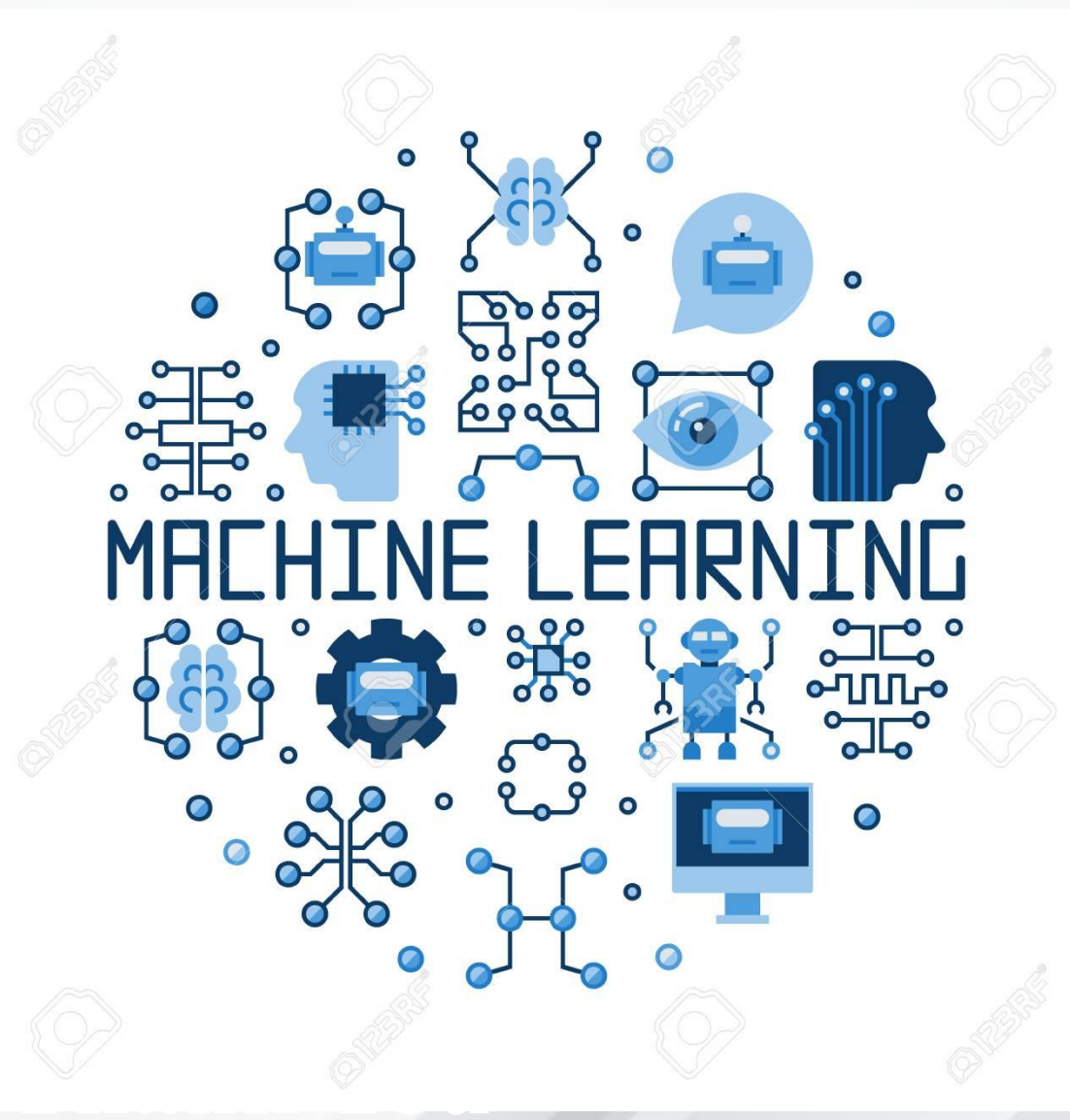
Course Roadmap & Expectations

- We will cover all these methods step by step.
- Focus on finance-specific applications.
- By the end: students can collect, prepare, analyze, and model financial data using Python ML.

Closing Remarks

Over the coming weeks, we'll explore key machine learning concepts and their applications in finance, from predicting risks to detecting fraud and forecasting trends.

This course is designed to build both your understanding and practical skills.

A circular graphic composed of various blue icons related to machine learning and artificial intelligence. The icons include a brain with circuitry, a gear, a robot, a computer monitor, a speech bubble, an eye, a head with circuitry, a brain, a gear, a robot, a computer monitor, a speech bubble, an eye, a head with circuitry, and a brain. The text "MACHINE LEARNING" is written in a bold, blue, sans-serif font across the center of the graphic.

MACHINE LEARNING

Thank you!