Developing a Machine Learning Model

This lesson provides an overview of the key steps involved in developing a machine learning model and explains important performance measurement metrics.

Define the Problem

01 Understand how the solution will be used

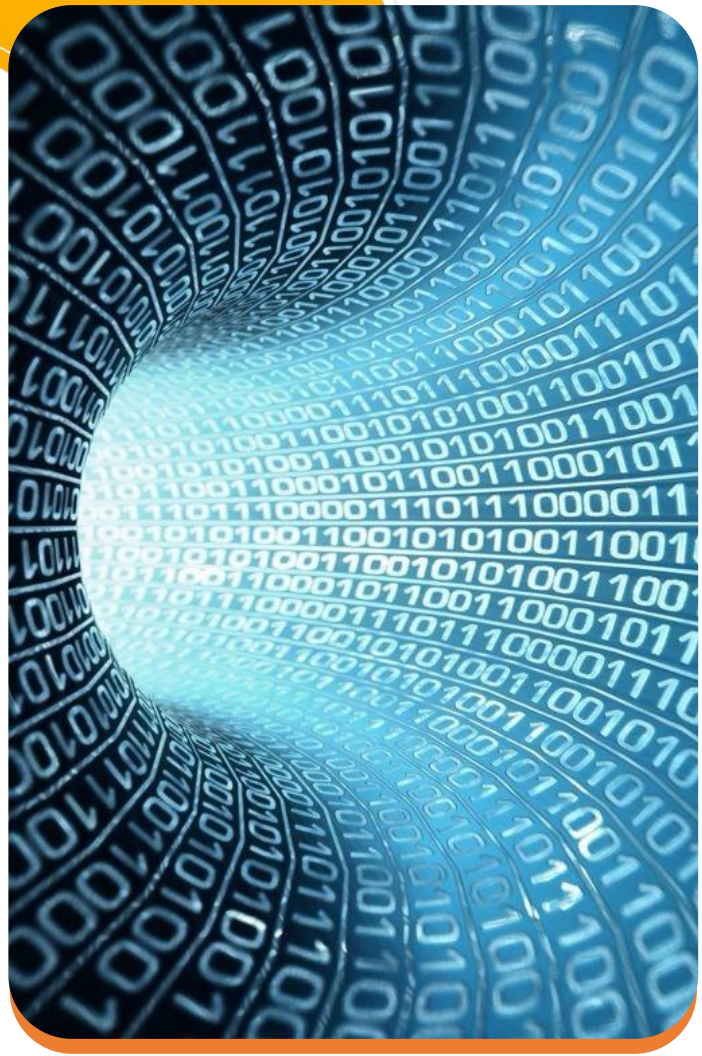
02 Identify the goal, inputs, and outputs of the model

Collect and Prepare Data

01 Gather data from various sources

02 Clean the data to handle missing values, outliers, and errors

03 Perform feature engineering for improved model performance



Choose a Model

- Select an appropriate ML algorithm based on problem type
- Consider data size, quality, and problem complexity

Train the Model

01 Split data into training and testing sets

02 Train the model using the training data

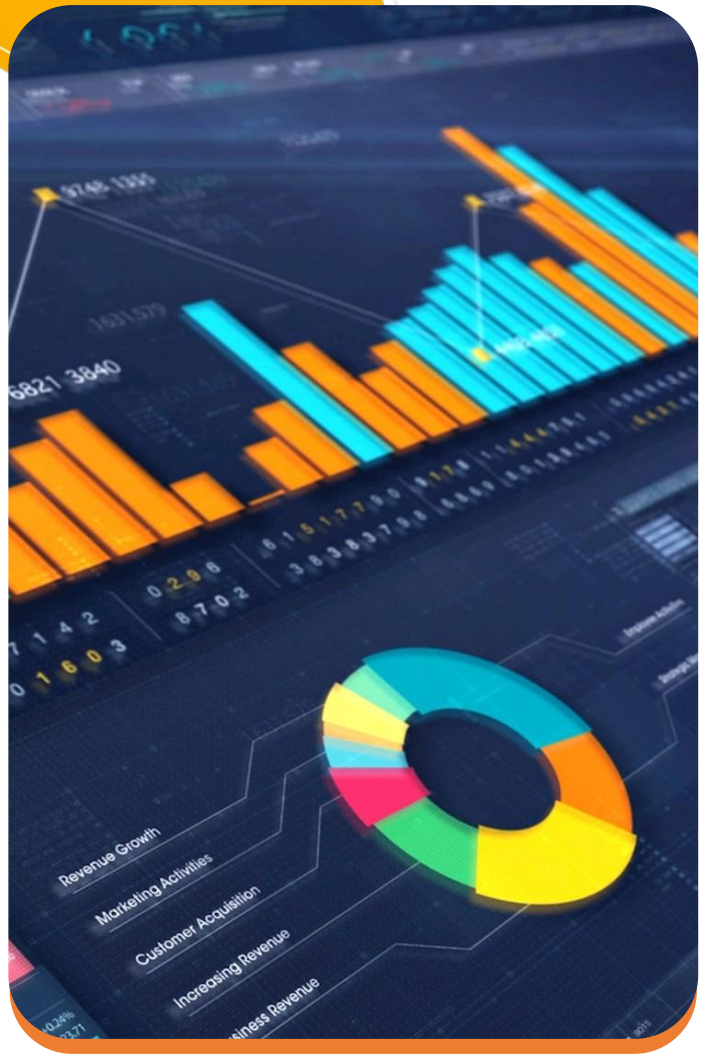
03 Evaluate model performance using the testing data

Evaluate the Model

01 Use metrics like accuracy score for classification problems

02 Use metrics like MSE or MAPE for regression problems

03 Adjust the model or preprocess data based on results



Deploy the Model

01 Integrate the model into existing software applications or systems

02 Deploy the model in a production environment

Monitor and Update the Model

- Continuously monitor model performance over time
- Retrain or update the model with new data periodically

Performance Measurement Metrics

Accuracy Score: Measures overall correctness of the model. Accuracy ranges between 0 and 1 where 1 means 100% accurate.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

Performance Measurement Metrics

MSE (Mean Squared Error): Measures average squared difference between estimated and actual values.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Where y is the actual value and y-hat is the predicted value for the i-th observation, and n is the number of observations.

Performance Measurement Metrics

MAPE (Mean Absolute Percentage Error): Measures average magnitude of errors as a percentage

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Where y is the actual value and y-hat is the predicted value for the i-th observation, and n is the number of observations.

Thank you. 😄