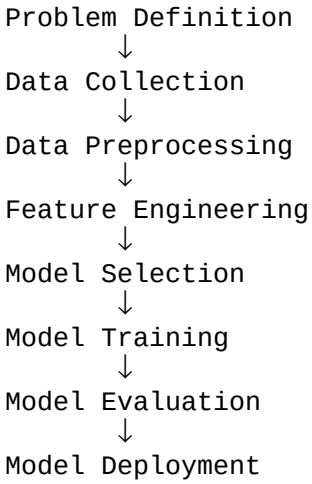
Machine Learning Workflow (Detailed & Diagram Based)

Introduction

The Machine Learning workflow is a systematic, step-by-step process used to build effective and reliable machine learning models. Each stage in the workflow contributes to the overall performance and real-world usability of the system.

Complete Workflow Diagram



1. Problem Definition

This step focuses on clearly defining the problem to be solved. A well-defined problem ensures the correct selection of data, algorithms, and evaluation metrics.

1. Identify objective
2. Decide ML task type
3. Define success criteria

Example: Predicting whether an email is spam or not.

2. Data Collection

Data collection involves gathering relevant data from reliable sources. High-quality data improves model accuracy.

1. Databases
2. APIs
3. Web scraping
4. Sensors and logs

3. Data Preprocessing

Data preprocessing cleans raw data and converts it into a usable format for machine learning algorithms.

1. Handle missing values
2. Remove duplicates
3. Encode categorical data
4. Normalize and scale features

Diagram:

```
[ Raw Data ] → Cleaning → Encoding → Scaling → [ Prepared Data ]
```

4. Feature Engineering

Feature engineering improves model performance by selecting or creating relevant features from data.

1. Feature selection
2. Feature extraction
3. Feature transformation

Example: Creating age categories from date of birth.

5. Model Selection

Choosing the right algorithm depends on the problem type, data size, and complexity.

1. Regression models for prediction
2. Classification models for labeling
3. Clustering models for grouping

6. Model Training

The model learns patterns from training data by adjusting internal parameters.

Diagram:

```
[ Training Data ] → [ Learning Algorithm ] → [ Trained Model ]
```

7. Model Evaluation

Model evaluation measures how well the model performs on unseen data.

1. Accuracy
2. Precision
3. Recall
4. F1-score
5. MSE / RMSE

Diagram:

```
[ Test Data ] → [ Model ] → [ Evaluation Metrics ]
```

8. Model Deployment

Deployment integrates the trained model into real-world applications to make live predictions.

Diagram:

```
[ Live Data ] → [ Deployed Model ] → [ Output / Prediction ]
```

Summary

Following a structured machine learning workflow ensures better performance, scalability, and reliability of ML systems.