



38/100

30 JAN 24 | DAY - 38 | MACHINE LEARNING

# #100DAYSOFDATA SCIENCE

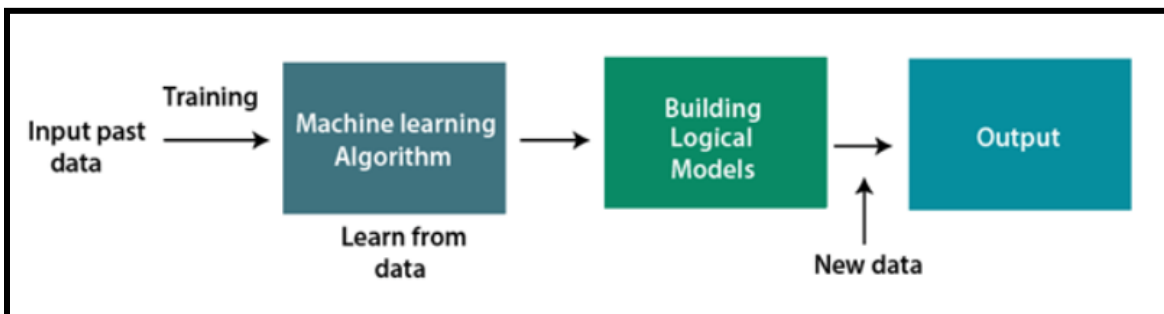
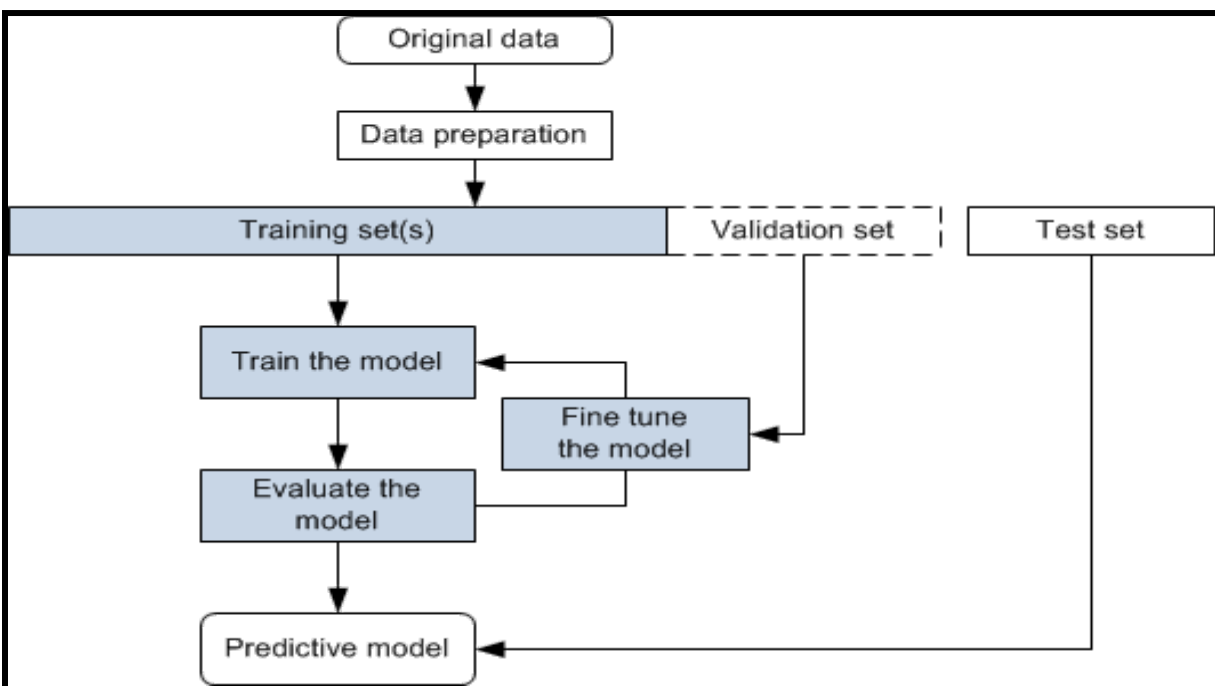
PYTHON | NUMPY | PANDAS | MATPLOTLIB | SEABORN | SQL | STATS | MACHINE LEARNING |

# Machine Learning Workflow

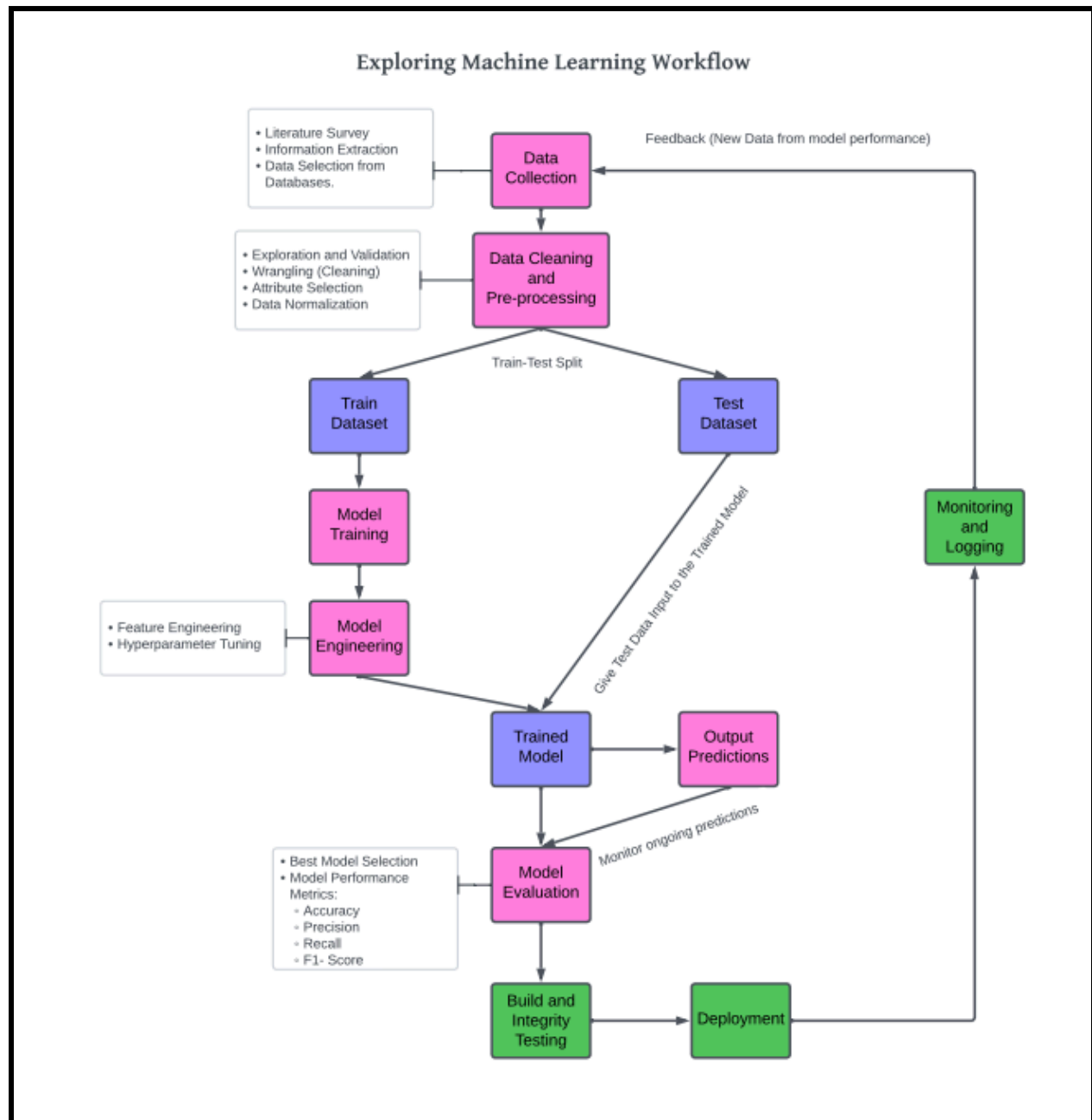
Machine learning workflows determine which phases are used during a machine learning project. An ML workflow describes the steps of a machine learning implementation.

Typically, the phases consist of data collection, data pre-processing, dataset building, model training and evaluation, and finally, deployment to production.

Machine learning workflows define the steps initiated during a particular machine learning implementation. Machine learning workflows vary by project, but four basic phases are typically included



## Visual Representation:



# The 7 Steps in a Machine Learning Workflow

## 1. Collect Data

Collecting data begins with defining the issue. An understanding of the issue is crucial for identifying the requirements – and the best solutions.

For example, a machine learning project that uses real-time data needs an IoT system that uses various data sensors. The first datasets can be gathered from different sources like a database, file, or sensor.

## 2. Prepare the Data

This means cleaning and formatting the raw data. Raw data can't be used for training the machine learning model. Also, machine learning models can only handle numbers, so ordinal and categorical data must be converted to numeric features.

## 3. Choose an ML Model

Considerations when choosing a model include performance (the quality of the model's results) and explainability (how easy it is to interpret the model's results).

Other considerations include dataset size (which affects how data is processed and synthesized) and training time and cost (of training a model).

## 4. Train the ML Model

There are 3 main steps to training the machine learning model:

1. Start with existing data
2. Analyze the data to find patterns
3. Make predictions

## **5. Evaluate the ML Model**

There are three main ways to evaluate the model:

1. Accuracy (the percentage of correct predictions for the test data)
2. Precision (the applicable cases that were predicted to belong to a particular class)
3. Recall (the cases that were predicted to belong to a class concerning all of the examples that legitimately belong in that class)

## **6. Perform Hyper Parameter Tuning**

Hyperparameters define the model architecture, so the process of trying to find the ideal model architecture is called hyperparameter tuning.

## **7. Deploy the ML Model To Make Predictions**

A prediction model is a container for the different versions of your ML model. To deploy a model, you first make a model resource in AI Platform Prediction (which runs your models in the cloud).

Next, develop a version of that model, and finally, link the model version to the model file that is stored in the cloud.

---

# What Can We Automate in a Machine Learning Workflow?

A machine learning workflow can be used much more efficiently if parts of it can be automated. So, what elements of the workflow are possible to automate?

## Data Ingestion

Automating data ingestion means freeing professionals up to be more productive at activities that need more manual attention.

This increased productivity allows processes to be optimized and resources to be managed more efficiently.

## Model Selection

Model selection means experimenting with various combinations of numbers and text data, plus various text processing methods – all done automatically.

## Feature Selection

This means automatically selecting the features of a dataset that are most valuable to your prediction variable – or most valuable to the output you are seeking.

## Hyperparameter Optimization

The goal of this is to find the hyperparameters with the lowest errors on the validation set so that these results generalize to the testing set.

---