

These are the steps towards deploying a machine learning application:

1. Problem Definition

identify the problem you want to solve

define the success criteria

find the requirements for the deployment (so real-time predicitons, scalability)

2. Data Collection

collect the data you need to solve the problem

clean the data

preprocess the data

3. Model Development

choose the right model (linear regression, neural network, etc.)

train the model using the training dataset

evaluate the model using the test dataset and capture accuracy, precision, recall

4. Model Deployment

save the model in a file

create a preprocessing pipeline for data input

create a prediction pipeline

5. Deployment architecture planning

choose the right infrastructure

local servers

cloud

IoT devices