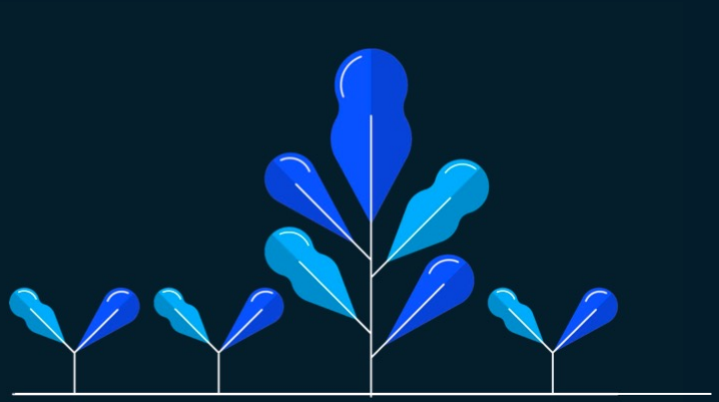
Misk مسك
مؤسسة محمد بن سلمان
Mohammed Bin Salman Foundation

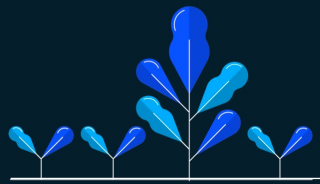


# Model Deployment in Python

Presenter: Shaimaa Alghamdi

# OUTLINES

- What is the Model Deployment?
- Criteria of ML Model Before Deployment.
- Steps for Deploying ML Models Into Production.
- Common Ways To Deploy ML Models.
- Fish Weight Prediction Application – DEMO
- References



# What Model Deployment?

To integrate a machine learning model into an existing production environment where it can take in an input and return an output.

# Criteria of ML Model Before Deployment

01

## Portability

To the ability of your software to be transferred from one machine or system to another.

02

## Scalability

Refers to how large your model can scale

# Four Steps for Deploying Machine Learning Models Into Production

**Prepare To Deploy the ML Model** 01

Train the model (This involves selecting an algorithm, setting its parameters, and training it on prepared, cleaned data)

**Validate the ML Model** 02

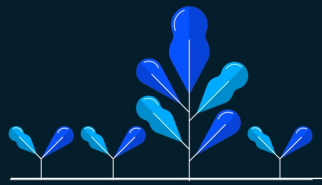
Validation includes testing the model on a fresh data set and comparing the results to its initial training.

**Deploy the ML Model** 03

Move the model to a production environment where resources are streamlined and controlled for safe and efficient performance.

**Monitor the ML Model** 04

Model monitoring ensures that the model is working properly and that its predictions are effective over time.

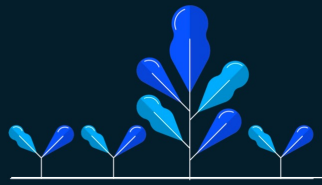


# Steps of Deploy The ML Model

01 The model needs to be moved into its deployed environment.

02 The model needs to be integrated into a process.

03 The people who will be using the model need to be trained in how to activate it, access its data, and interpret its output.



# Two Common Ways To Deploy ML Models

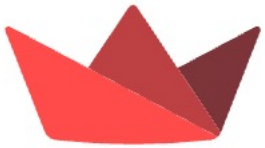
## Batch inference

- Also called offline inference, is the process of generating predictions on a batch of observations. The batch jobs are typically generated on some recurring schedule (e.g. hourly, daily).
- There is no constraint, you can deploy more complex models that can provide more accurate results.

## Online inference

- Also called real-time inference or dynamic inference. These predictions are generated on a single observation of data at runtime.
- Since it should provide results in real-time, you cannot use complex models with online inference.

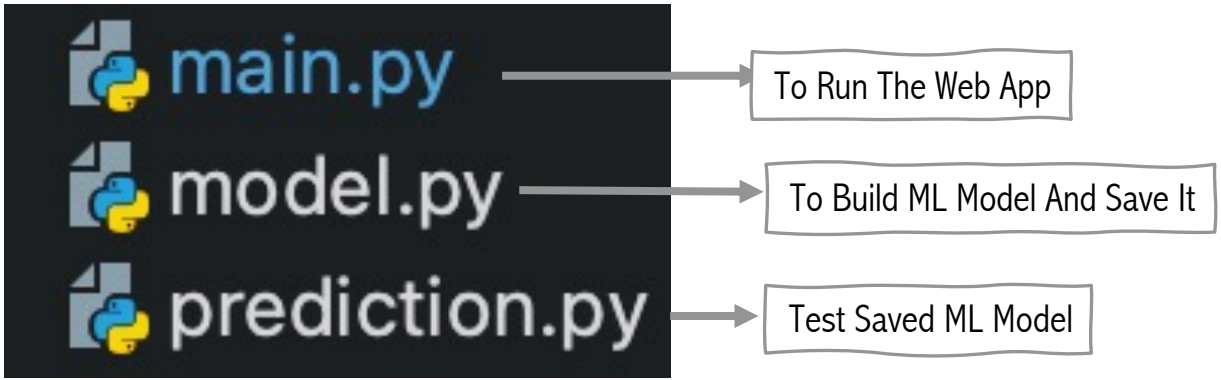
# Fish Weight Prediction Application - DEMO



STREAMLIT

- A **faster** way to build and share data apps.
- Streamlit turns data scripts into shareable **web apps in minutes**.
- **No** front-end experience required.

# Fish Weight Prediction Application - DEMO



# Fish Weight Prediction Application - DEMO

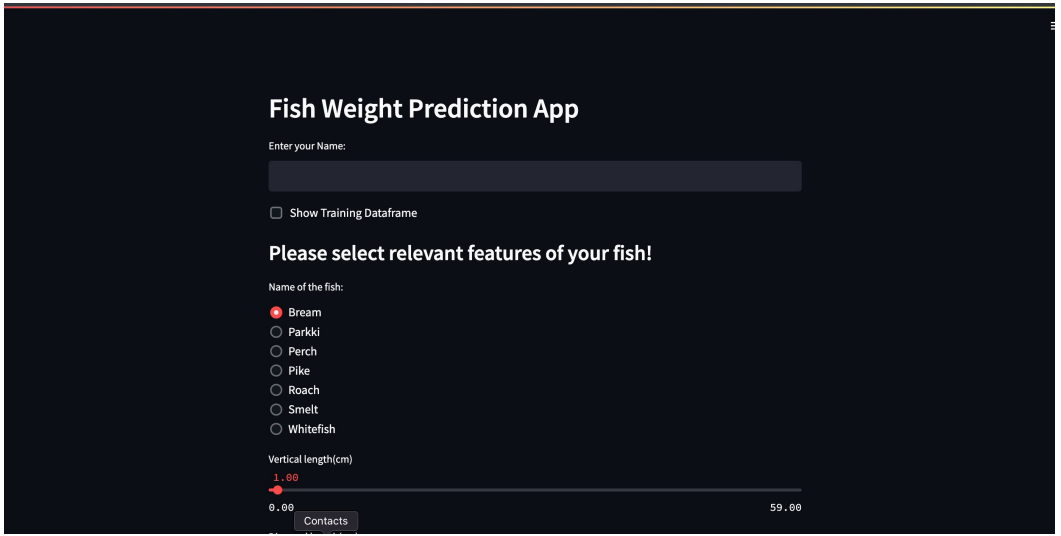
TO TEST THE DEMO, FOLLOW THESE STEPS :

**1. Cloning a repository [DEMO]**

**2. Using the Terminal:**

**1.1 Install necessary libraries**

*pip install requirements.txt*



# Fish Weight Prediction Application - DEMO

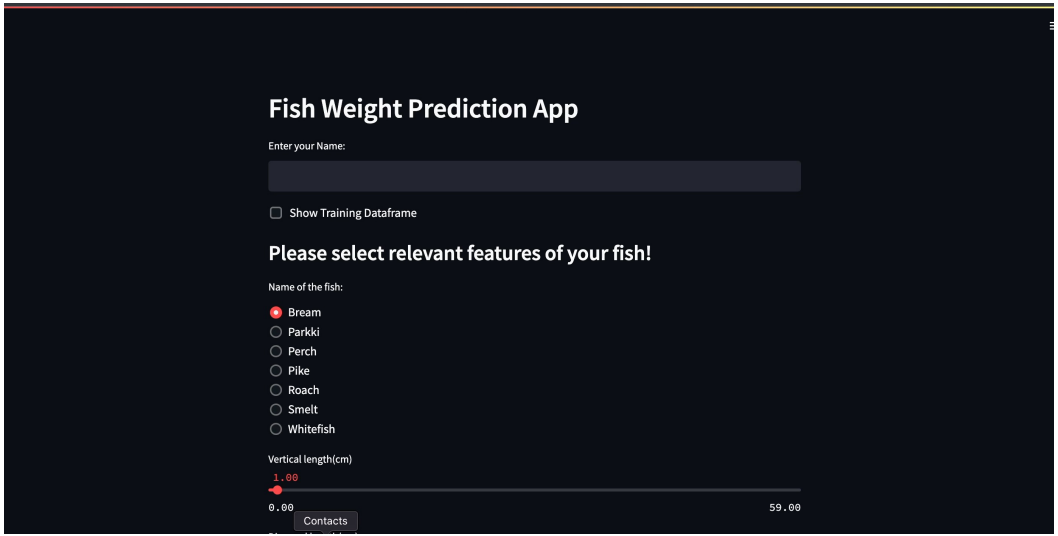
TO TEST THE DEMO, FOLLOW THESE STEPS:

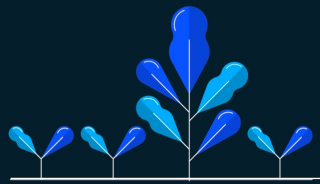
**1. Cloning a repository [DEMO]**

**2. Using the Terminal:**

**1.2 Run the Streamlit tool**

*streamlit run main.py*





# REFERENCES


**3 Ways to Deploy Machine Learning Models in Production**

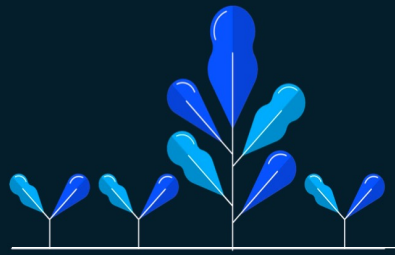
**How to Deploy Machine Learning Models**

**3 Steps To Efficient Machine Learning Model Deployment**

**A Guide to Machine Learning Model Deployment**

**Streamlit Tool**

**DEMO**



# THANKS
# ANY QUESTIONS