A primer guide

Machine Learning project steps

Introduction

AI (Artificial Intelligence) refers to the development of systems/machines that can perform tasks that historically only humans could do e.g.

- Reasoning
- Decision making
- Problem-solving
- Language Translation
- Perceptions

---Industry-wise impact---

Healthcare: AI is making significant contributions in diagnosing diseases, drug discovery, personalized treatment plans, etc.

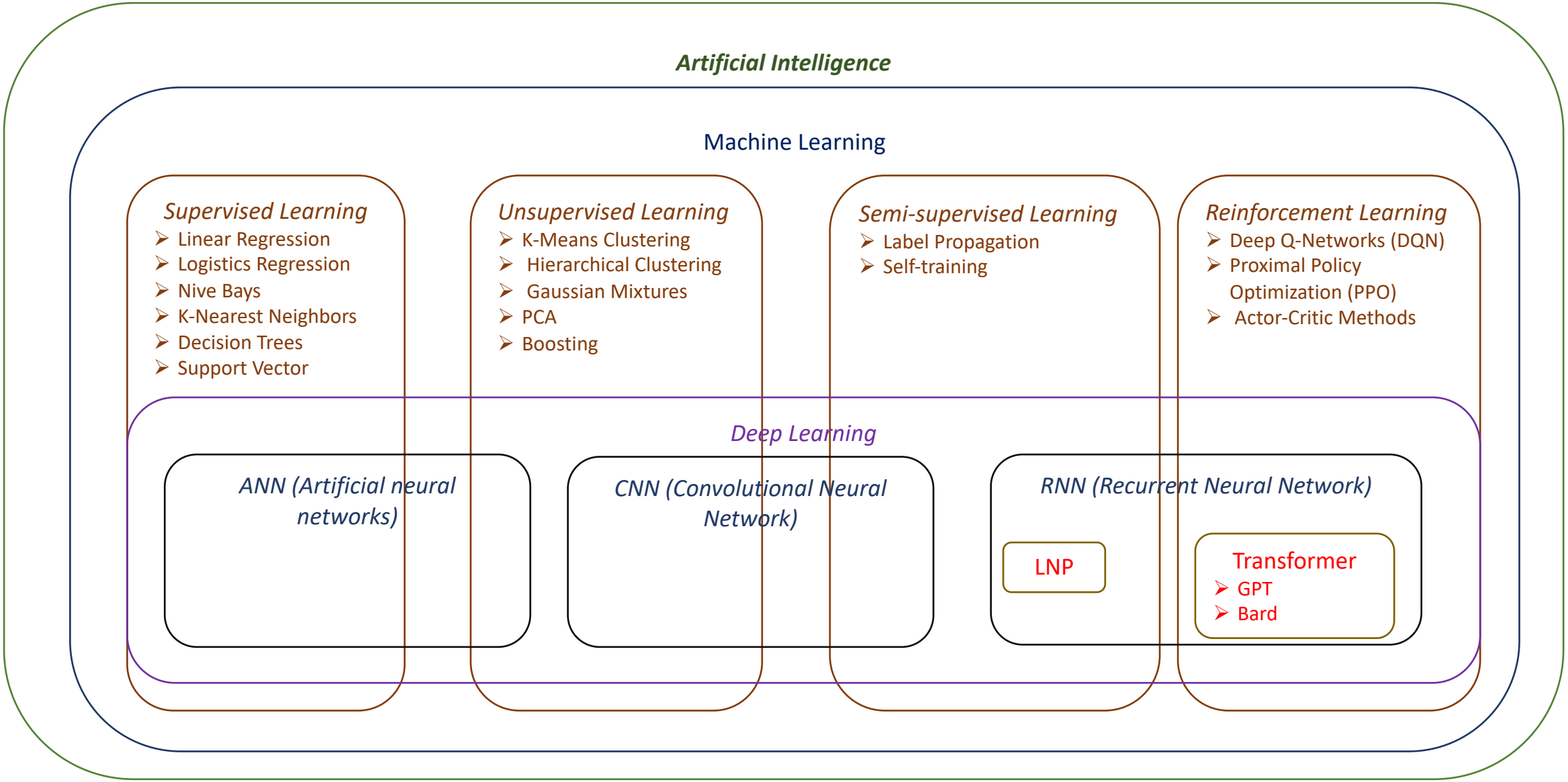
Finance: The financial industry uses AI for fraud detection, algorithmic trading, risk assessment, customer service, etc.

Transportation: Autonomous vehicles, and traffic management systems utilize AI to improve safety and efficiency in transportation.

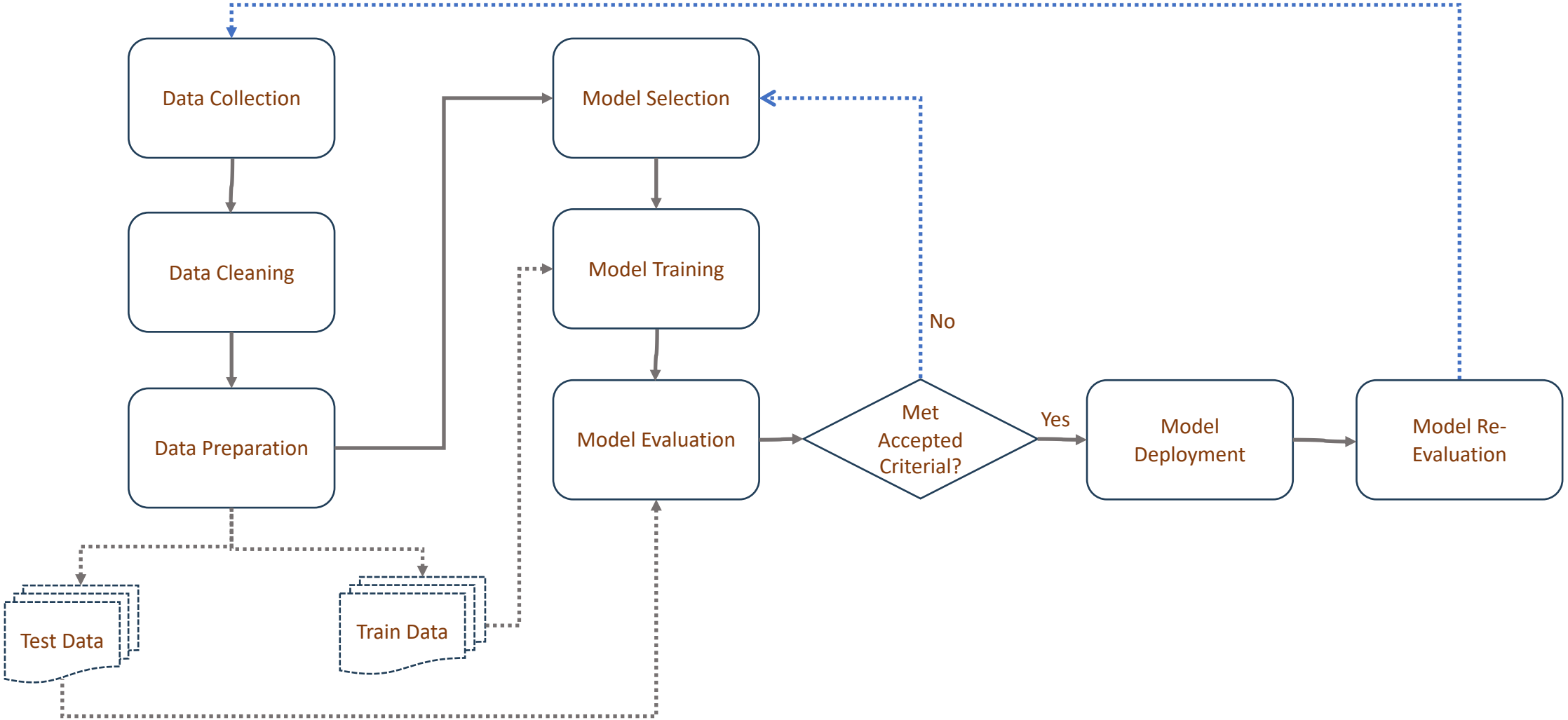
Entertainment: content recommendation, game design, special effects where AI is making significant contributions

Retail: Product recommendations, personalized user journey, customer support where AI is making a major improvement

Types & Classification



ML project steps



ML project steps cont..

Data Collection

- In this step data for the problem statement are collected . Data can be collected from various data source e.g.
  - Domain specific transactional data
  - Web Analytics
  - GPS Sensor
  - Social network

Data Cleaning

- This step refers to the process of
  - Selection of feature variable (in our example use-case; House age, Distance from nearest MRT satiation, Number of convenience stores)
  - Removal of unwanted data (No, Transaction Date, Latitude, Longitude)
  - Filling missing values in feature columns with mean /median of the columns/ KNN
  - Feature encoding (turning text into numbers)
  - Feature normalization (scaling / standardization)

Data Preparation

- In Data preparation steps we generally split our sample data in 2 categories
  1. Train data (70% ~ 80%) – used for training the selected model
  2. Test data (20% ~ 30%) – used for validate the accuracy of the model

Model Selection

- Model selection is a process of fitting the dataset into a standard ML model algorithm . Our current dataset is best fit for Linear Regression algorithm. Python sklearn library provides an exhaustive list to ML model algorithm .

Model Training

- In this process model is trained by using Training data created during “Data Preparation” step

Model Evaluation

- In this process model is evaluated buy using Test data created during “Data Preparation step”. Versus matrix are captured to see the performance of the model.

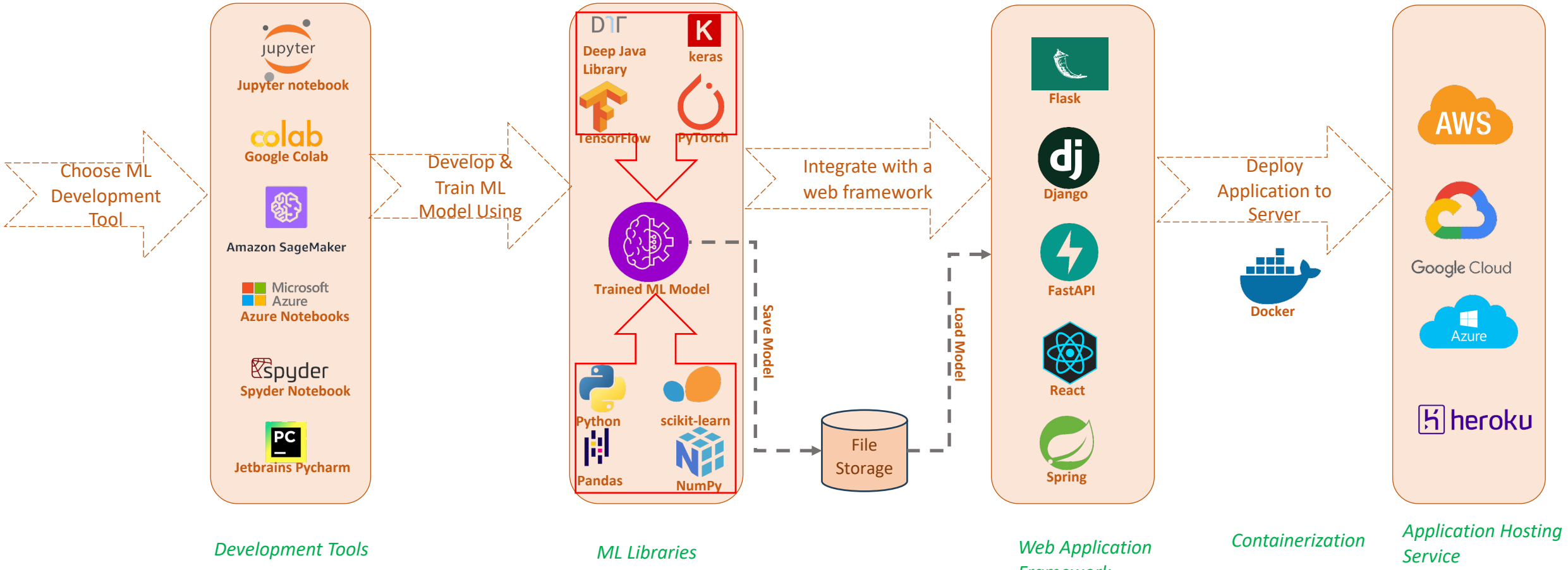
Model Deployment

- Once the model is tuned and evaluated model needs to be deployed. One common deployment methodology is to expose the model as web services API. Steps to expose as API.
  - Export the trained model as a pickle/joblib/HDF5 file
  - Read the model in Python flask/fastapi and expose it as restful web services

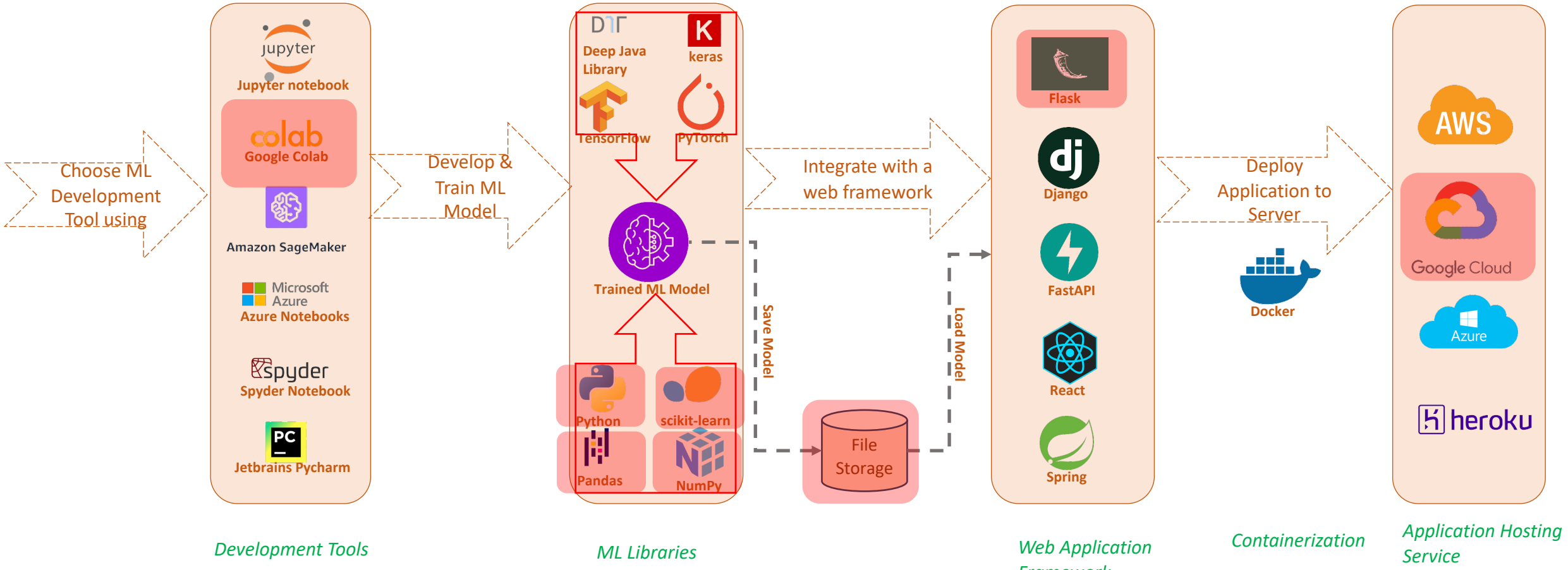
Model Re-Evaluation

- In general, a model needs to be re-evaluated again
  - If the model was trained in a smaller dataset
  - If a new set of data is accumulated
  - If the model is outdated

Development & Deployment process.



Development & Deployment process. (used in hands-on example)



Hands-on code sample.

To explain various steps in ML project. We will use "Real estate price prediction" data set . The Kaggle link for the data set
https://www.kaggle.com/datasets/quantbruce/real-estate-price-prediction
*Note ** this dataset is collected from the public domain and not related to any project (or any proprietary data)*

This dataset has 414 sample data . We will create a ML model to predict the price per unit area of a house based on -
1.Age of the house
2.Distance from nearest MRT satiation
3.Number of convenience stores

No	transaction date	house age	distance to the nearest MRT station	number of convenience stores	latitude	X6 longitude	house price of unit area
1	2012.917	32	84.87882	10	24.98298	121.54024	37.9
2	2012.917	19.5	306.5947	9	24.98034	121.53951	42.2
3	2013.583	13.3	561.9845	5	24.98746	121.54391	47.3
4	2013.5	13.3	561.9845	5	24.98746	121.54391	54.8
5	2012.833	5	390.5684	5	24.97937	121.54245	43.1

Please refer the attachment for code sample

Model development code in pdf	**RealState_Prediction.pdf**
Model deployment code in pdf	ML_Web.pdf

THANK YOU