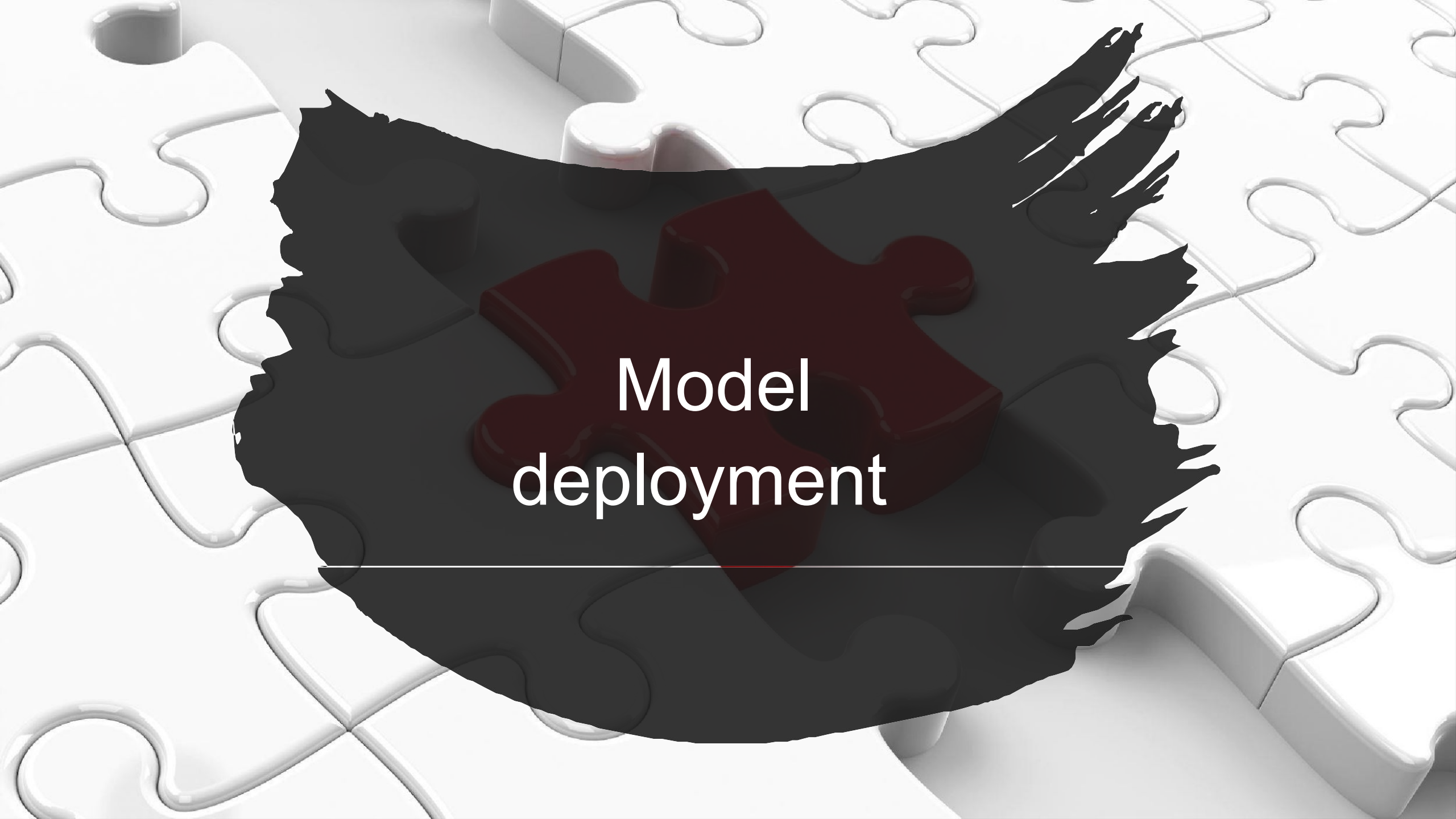
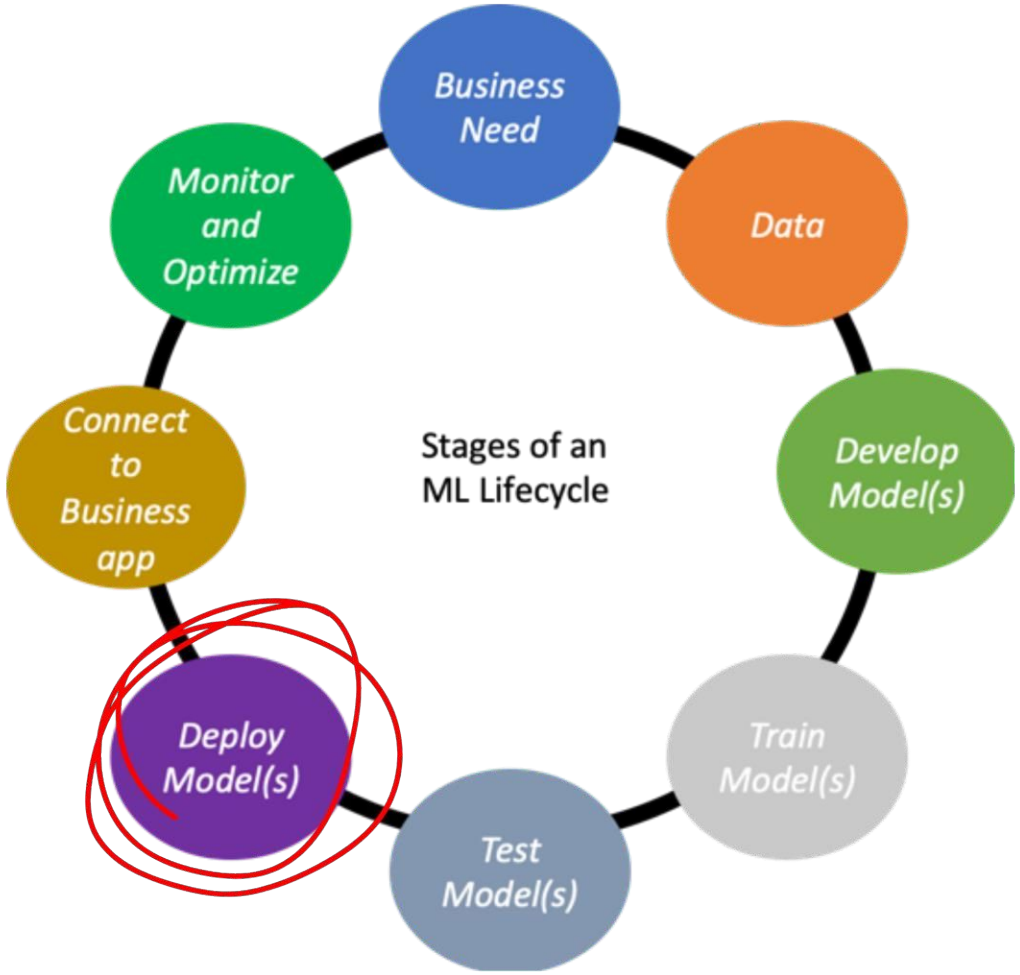

Model
deployment

Content

- What is model deployment
- Why model deployment
- Train the model and save it
- Why flask
- Why Heroku
- Creating a web app
- Commit code into GitHub
- Link GitHub to Heroku

What is model deployment?

- One of the last stages in the ML life cycle.
- Integrating ML models into an existing production environment
- It should be available to the end users.
- Or to make business decisions based on data.

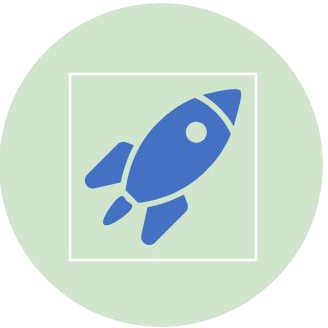


Why it is important?

"No machine learning model is valuable, unless it’s deployed to production." – Luigi Patruno

TO MAKE MOST VALUE OUT OF ML MODELS

LOT OF CHALLENGES



FUTURE-PROOF

Things to consider

- Input data and output data: Data storage, Data pre- processing pipeline, Input data stream,Output data stream
- Frequency and urgency
- Batch or real-time predictions
- Latency: how fast output should be
- Privacy: User privacy
- Computing costs

Frameworks and tooling

Frameworks

Tensorflow
Pytorch
Scikit-learn

Programming languages

Python
R
jav a

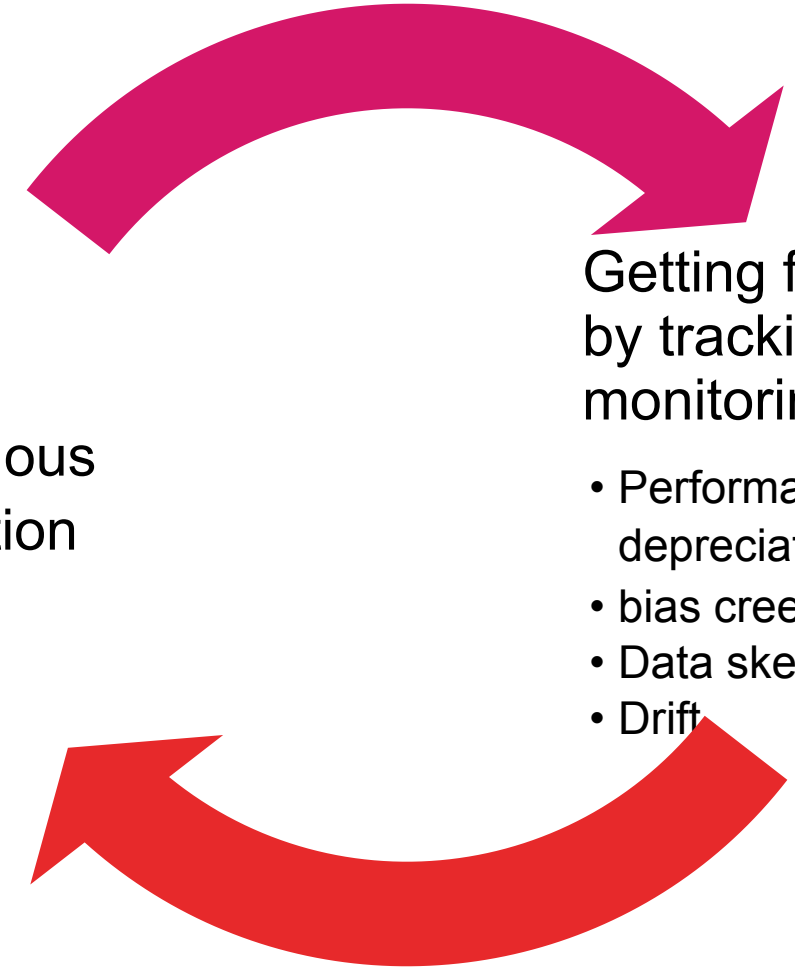
Cloud environment

AWS
GCP
Azure

How to choose the best one?

Efficiency
Popularity
Support

Feedback and iteration

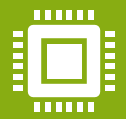


Getting feedback by tracking and monitoring

- Performance depreciation/decay
- bias creep
- Data skew
- Drift

Continuous integration

Things to consider while building models



Portability: Is the ability of software to be transferred from one machine or system to another.

Scalability: Is the ability of a program to scale.

Operationalization: Refers to the deployment of models to be consumed by business applications.

Test: Refers to the validation of output to processes and input.

System Architecture

- Architecture can be defined as the way software components are arranged and the interactions between them.
  - Modularity
  - Reproducibility
  - Scalability
  - Extensibility
  - Testing

High-Level Architecture of an ML System

Data Layer: Access to all the data sources.

Feature Layer: Generating feature data. Which should be transparent, reusable and scalable

Scoring Layer: the scoring layer transforms features into predictions.

Evaluation Layer: Monitor and compare how closely the training predictions match the predictions on live traffic.

Different methods to deploy

- Train
  - Batch: Ad-hoc training
  - Real-time: Learning on fly
    - Data doesn’t fit into memory
    - If data distribution drift over time
    - Data is a function of time
- Serve
  - Batch
  - Realtime

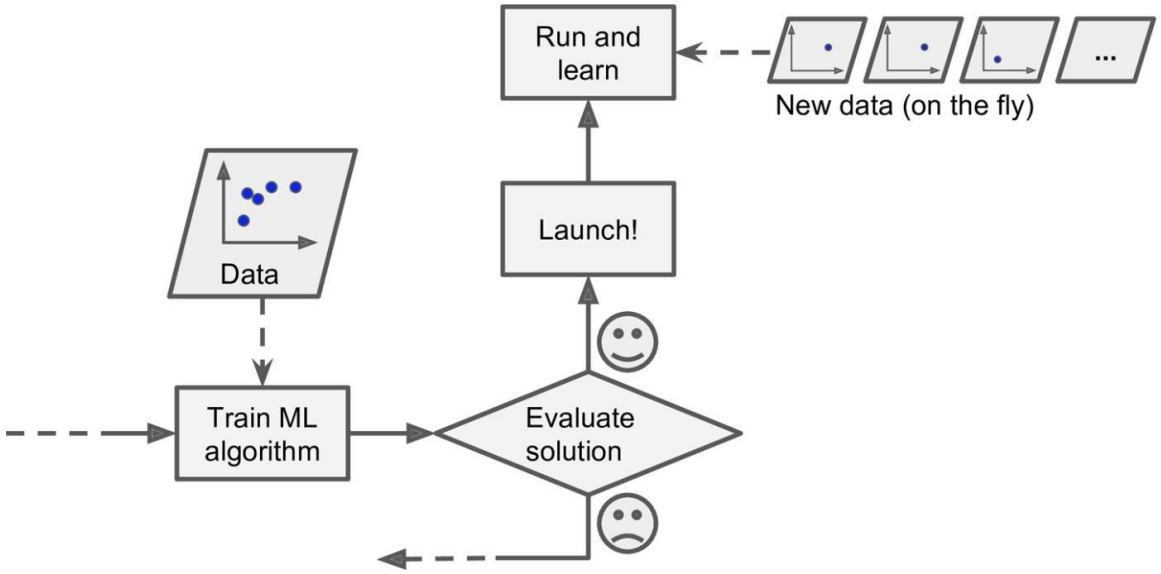


Figure 1-13. Online learning

Ask your questions?

Train the model
and save it

Flask

- API-first approach
- It is a framework that allows you to build web applications.
- Other frameworks like Django, Falcon, Hug and many more.
- For R, we have a package called plumber.
- Pip install flask

Steps

- Function to load the saved model
- Create root path
- Create a route path to predict the class
- Return the result

Ask your questions?

Heroku

- Heroku is a platform as a service (PaaS) that enables the deployment of web apps based on a managed container system, with integrated data services and a powerful ecosystem.

Steps

- Create Procfile
- Create requirements.txt
- 2 ways to deploy
  - From UI
    - Commit Files to GitHub
    - Deploying With Github on heroku
  - With CLI

With CLI

- heroku login
- heroku create appname
- git init
- git add .
- git commit -m 'commit'
- git push heroku master
- heroku open

Ask your questions?

Slide Download Link

You can download the slides from the below link:

https://docs.google.com/presentation/d/12niH7HnkhC5OpdgO3kJ-kR--ziizbw_p7Bv2FGZwDe8/edit?usp=sharing

Thank you