<https://www.coursera.org/learn/introduction-to-machine-learning-in-production/exam/pXfsA/deployment/view-attempt>

Shadow mode is a type of deployment where the ML algorithm runs in parallel with the user but it’s output isn’t used for any decision making.

Canary deployment allows you to monitor the performance of an algorithm on a subset of all traffic and then either ramp up to more traffic or rollback if an issue is detected.

Changing the wording will result in a concept or data drift.

Don't plan or expect to perfect the deployment on your first attempt, it's much more reasonable and efficient to iterate on the process and optimize its performance over time.

<https://www.coursera.org/learn/introduction-to-machine-learning-in-production/exam/XQM8q/selecting-and-training-a-model/view-attempt>

**Data-centric** means you focus your efforts on improving the data to raise the system's performance, while keeping the code fixed.

High average test set accuracy is a great achievement, but there is more work to be done to ensure the algorithm works well on real-world data, is fair, and performs well on rare classes of diseases.

For most unstructured data problems, human-level performance is a great estimate of Bayes error - an upper limit to your system's potential

Something is clearly wrong with the implementation if the algorithm is unable to overfit to a single training example! Find the root cause, fix the problem, and then move onto larger datasets.

Tracking tool :- But some examples include Weight &Biases, Comet, MLflow, Sage Maker Studio, Verta.ai Granseal also has its own experiment tracking tool focusing on computed vision and manufacturing applications

data drift, where the distribution of the data you trained on maybe eventually become very different from the distribution of the data that you're running inference on

TFX is an open-sourced end to end ML platform that we'll be using in this course. We could push to TensorFlow Lite and use our model in a mobile application or on an IOT device

ML Model Deployments and Pipelines :-

- GraphPipe (https://oracle.github.io/graphpipe/#/) from Oracle

- Tensorflow Serving, TFX (https://www.tensorflow.org/tfx/) from Google

- Mlflow (https://mlflow.org) from databricks

- kubeflow (https://www.kubeflow.org/) from Google

- Seldon (<https://www.seldon.io/>)

For some applications, having and tracking metadata,

data provenance, and data lineage can be a big help