Machine Learning Deployments

Description

Productionizing your machine learning model to be used inside of a project at a hackathon is one of the most commonly faced challenges for hackers. This workshop covers the limitations of deploying machine learning models with Flask, and covers how to make more efficient deployment while also being able to monitor your machine learning model performance.

Learning Outcomes

This is what you will walk away from the workshop able to do:

- Learn how to deploy machine learning models wish Flask
- Learn how to monitor your machine learning performance post-deployment
- How to add version control to your machine learning models

Prerequisite Knowledge

In order to get the most out of this workshop you should be comfortable with the following concepts:

- Basic Python
- Basic Machine learning concepts: You should have built a very simple ML model yourself (takes 15 mins to do with the resource mentioned in Pre-workshop checklist) or you should have used a ML model available on the internet
- Some familiarity with TensorFlow or PyTorch
  You should know what a REST API is? (Read this to get up to speed)

Pre-Workshop Checklist

Before the workshop, please make sure you complete the following items (Recommended):

- [OPTIONAL] Install Python, Install TensorFlow, Install TensorFlow Serving
- Go through this very beginner tutorial for training a TensorFlow Model

Technical Jargon and Definitions

- Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed
- Machine Learning Model: A combination of weights and model architecture which also represents what your algorithm learned
- TensorFlow: An end-to-end open source machine learning platform
- TensorFlow Extended: The ecosystem of tools that allows deploying your models or doing tasks related to ML deployments easy
- ONNX: An open format built to represent machine learning models
- Deployment: You want your software to be used by your users and not just reside on your machine, to do so you deploy it
- API: APIs are the little pieces of code that make it possible for digital devices, software applications, and data servers to talk with each other, and they're the essential backbone of so many services we now rely on

Timeline (60 mins)

Time	Module	Description
2.5 mins	End to End lifecycle of an ML project [REFRESHER]	A refresher on how you can structure your Machine Learning projects: ideation, experimentation, data collection, model training and more.
2.5 mins	Training TensorFlow and PyTorch models [REFRESHER]	A refresher on training basic models in TensorFlow and Python.
5 mins	Why another step after training?	Now you have trained your model, or you got a model which was open-sourced, and you want to use it. Can you use the model in the exact same way you did while training? TL: DR No.
8 mins	What is needed for successful deployments?	What all do you need to make a successful deployment, maybe a server to host your model, maybe an API so people could consume your model, some specific things for ML deployments.
2.5 mins	Setting up the background for a model we want to deploy	Set up some background for a model we will be taking a look at throughout the workshop this could most certainly be your own model.
5 mins	Inefficient	Quite a lot of folks might have created APIs for their

	Deployments	ML models here, probably with Flask but all things specific to ML might not be covered with these deployment strategies. And we also briefly talk about serverless ML deployments.
15 mins	Make a Deployment with TFX especially using TF Serving	Let us now take a look at how you can serve your model to users, host it, get inferences from it, build consistent APIs and more!
15 mins	Model Deployment [DEMO]	We did use some code snippets earlier and took a look at parts of the model deployment process. Let us now take a popular model and serve it with TF Serving.
5 mins	Some of the other deployment features	For larger, more involved use cases you might need to support mini-batching, model versioning, polling, model monitoring and meta APIs. We also give you a quick overview of some out-of-scope but useful topics related to deploying models.
2 mins	Questions	Did we cover every single thing you could for deploying your model, well, no and we would love to take on questions.

Workshop Lead Contact

Shivay Lamba	Rishit Dagli
shivaylamba@gmail.com https://twitter.com/howdevelop	rishit.dagli@gmail.com https://twitter.com/rishit_dagli

Additional Resources

Hack the North Resources

Hack the North 2022 Event Schedule

Check this out to stay up-to-date on activities, workshops, and other key happenings this weekend.

Workshop-Specific Resources

1. TensorFlow: https://www.tensorflow.org/
2. TensorFlow Extended: https://www.tensorflow.org/tfx
3. TensorFlow Serving: https://www.tensorflow.org/tfx/guide/serving
4. Future learning: https://www.deeplearning.ai/courses/machine-learning-engineering-for-production-mlops/