



# Deploy ML models

by Pratik Parmar aka Pintudo





Hello!

I AM Pratik Parmar

And I am here to bore you with Machine Learning.



- App Idea: Dog Breed classifier



And dogs are like...

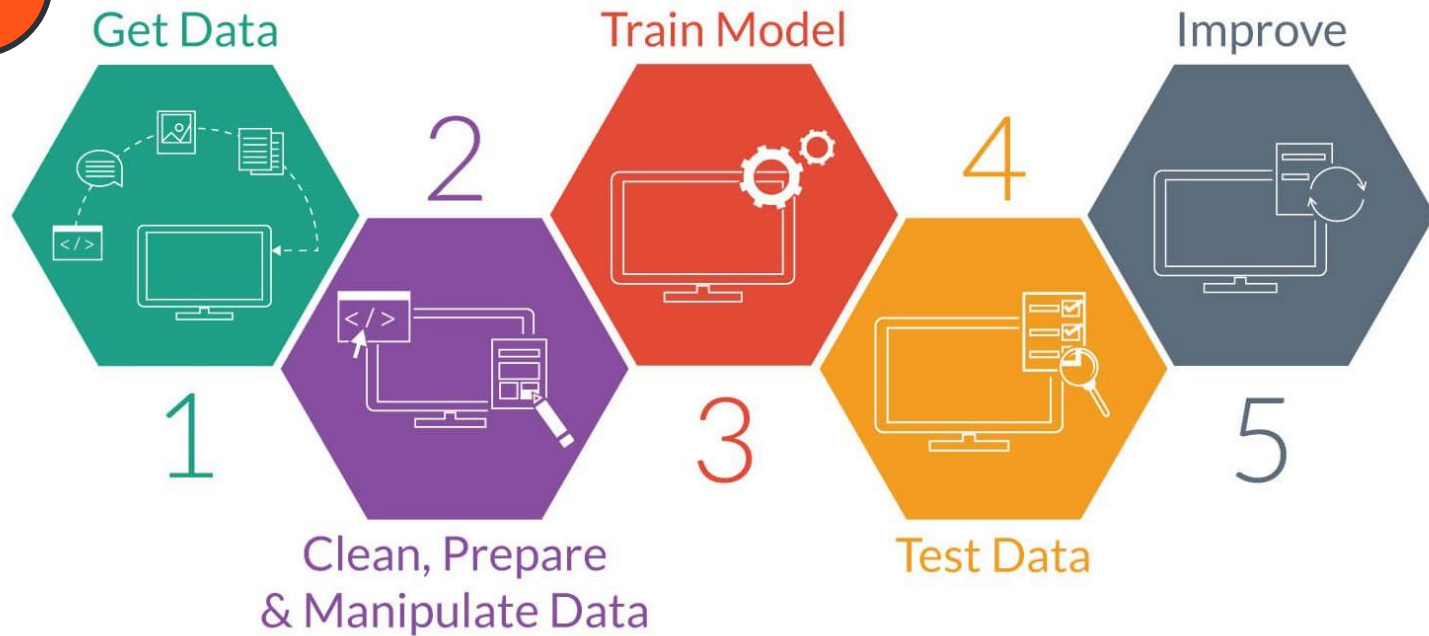


1

Let's talk about basic Machine Learning workflow



- General Machine Learning Workflow



- After evaluation

The workflow for building machine learning models often ends at the evaluation stage: you have achieved an acceptable accuracy, and “*ta-da! Mission Accomplished.*”



- Why ???

Maybe, going the extra mile to put your model into production is not always needed. And even when it is, this task is delegated to a system administrator.



“

*During my course at Coursera I was always asking myself—I have my model, which I can run in Jupyter Notebook and see the result, but what can I do with it? How can other use it?*





- Why Demo ??

No matter how silly your project is, demonstrating our work is generally great way to get a wider audience interest

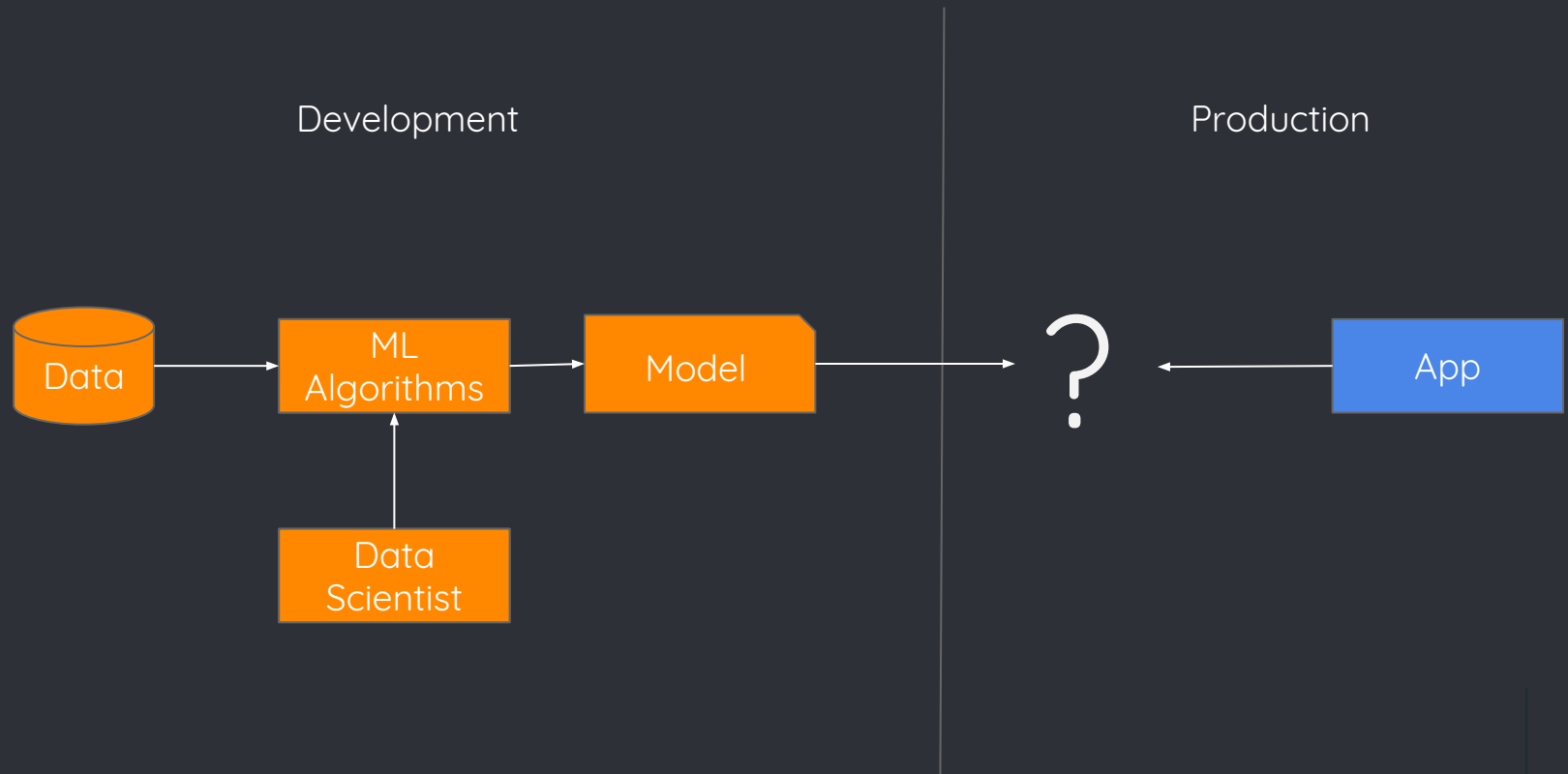


2

How to deploy Machine Learning models?



- How to deploy a ML model ?





*Any guess*





# Serving

Serving is how you *apply* a ML model, after you've trained it



- 3 options to implement Machine Learning models

- **Rewrite it**

Rewriting the whole code in the language that the software engineering folks work.

- API-first approach**

Create web API for your ML model using any web framework i.e. Flask or Django

- Tensorflow Serving**

TensorFlow Serving makes it easy to deploy new algorithms and experiments, while keeping the same server architecture and APIs. (API-first approach, but only for tensorflow)

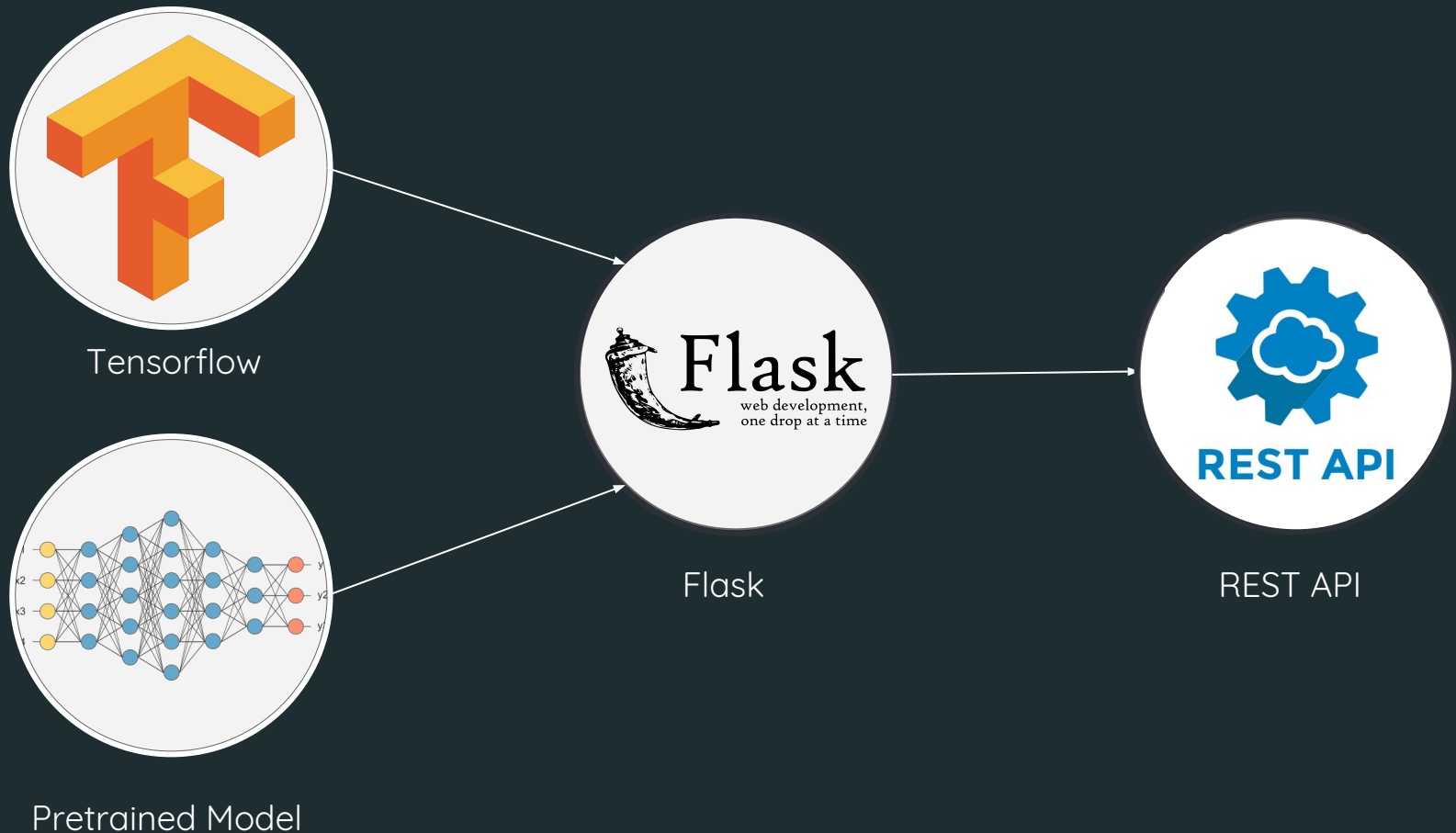


3

Let's get our hands dirty with the code



- Create simple Web API for Dog Image Classifier







Which model to choose?





## Pretrained model

For the sake of this session, we're gonna use retrained MobileNet Model



- Main steps to to deploy Model are

- Train the model and saving the checkpoints on the disk

- Load saved model and test that it works properly

- Export model into Protobuf format (.pb)

- Create the client to issue requests and make an API



## ● Protobuf

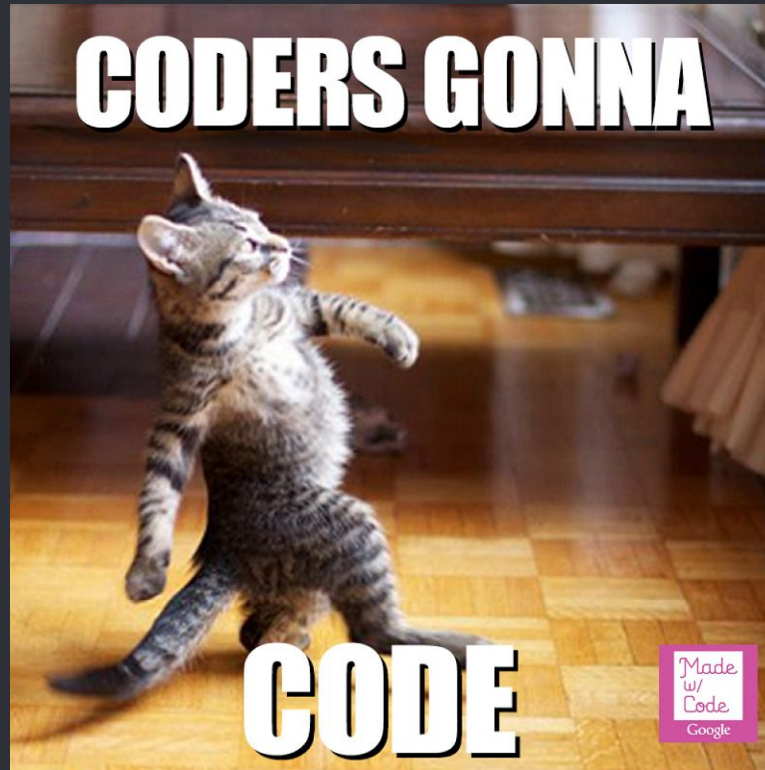
○ Protocol buffers are Google's language-neutral, platform-neutral, extensible mechanism for serializing structured data – think *XML*, but smaller, faster, and simpler.

## ● Export the model into Protobuf

○ Tensorflow serving provides *SavedModelBuild* class to save the model as Protobuf.



- Give some space because



- Machine Learning for everyone



- NO QUESTIONS PLEASE 🦴



- Thank you for bearing me 🐻

