Full Stack Data Science

Building Your First Smart Application

Who am I?

Chris Luiz

- Startups & Product Development
- Data Scientist & Engineer @ Progressive
- Technical Education Enthusiast
- Father, Beekeeper, and Hobby Farmer

Why This Topic?

“Unsolicited data science career advice:

Learn Flask + just enough React to be able to build functional front-ends for your DS work + just enough CSS to make it look not-horrible, it's like having a superpower (and it's not that hard)”

- Joel Grus, Author, Data Science from Scratch

Meet Your Neighbor!

Take the next couple minutes to say hello to the people on either side of you.

- What is their name?
- What do they do during the day?
- What do they hope to get out of this workshop?
- Offer a unique fact about yourself
    - Ie, You once climbed a mountain in a banana costume

Outline

- Environment Set Up
- Web Development Basics
- Data Acquisition & Processing
- Modelling: An Intro to scikit-learn
- Web Dev Part Two - Integration
- Stretch Options

Demo Time!

Part 1: Environment Set Up

- Install Script
- Run Script
- Config files

Part 1: Environment Set Up

Why should we use these tools?

Part 1: Environment Set Up

Why should we use these tools?

- Portability

Part 1: Environment Set Up

Why should we use these tools?

- Portability
- Flexibility

Part 1: Environment Set Up

Why should we use these tools?

- Portability
- Flexibility
- Reproducibility

Part 1: Environment Set Up

Why should we use these tools?

- Portability
  - Install using the install script, run with the runs script -> will run on any machine with the same hardware and permissions.
- Flexibility
  - No interference with existing system configuration or packages.
- Reproducibility
  - In DS, given the same data, you should produce the same results. This process is analogous at the application level.

Part 2: Web Development Basics

Tech stack:

- Flask

Part 2: Flask

- Open source python micro framework
- Easy to get started with
- “Flask can be everything you need and nothing you don’t.” - Flask Docs

Part 2: Web Development Basics

Tech stack:

- Flask
- Bootstrap 4

Part 2: Bootstrap

- Bootstrap 4 - January 2018
  - Open source, originally developed by twitter
  - Many resources, easy to use and adapt
    - https://hackerthemes.com/bootstrap-cheatsheet/

Part 2: Web Development Basics

Tech stack:

- Flask
- Bootstrap 4
- Jquery

Part 2: Jquery

- Jquery
  - Powerful javascript library used everywhere
  - Makes javascript easier to write
  - Simplifies dom interactions
  - https://htmlcheatsheet.com/jquery/

Part 2: Web Development Basics

Tech stack:

- Flask
- Bootstrap
- Jquery
- Bokeh

Part 2: Bokeh

Bokeh

- Bokeh is a Python interactive visualization library that targets modern web browsers for presentation.
  - https://bokeh.pydata.org/en/latest/docs/reference.html

Part 2: Web Development Basics

Tech stack:

- Flask
- Bootstrap,
- Jquery
- Bokeh

Other ingredients include...

- A simple & sane directory structure
  - Most files in main directory
  - Static, templates, logs, and data folders
- An emphasis on simplicity

Part 2: Web Development Basics

- Enough Talk! Let’s build something!

Part 3: Data Acquisition & Processing

- Problem definition, or, What the hell are we doing here?

Part 3: Data Acquisition & Processing

- Problem definition, or, What the hell are we doing here?
- Public Data Sets/Apis Overview

Part 3: Data Acquisition & Processing

- Problem definition, or, What the hell are we doing here?
- Public Data Sets/Apis Overview
  - Colorado open data: https://data.colorado.gov/
  - Denver open data: https://www.denvergov.org/opendata
  - Quandl : https://www.quandl.com/
  - https://github.com/awesomedata/awesome-public-datasets
  - https://github.com/datasets

Part 3: Data Acquisition & Processing

- Problem definition, or, What the hell are we doing here?
- Public Data Sets/Apis Overview
  - Colorado open data: https://data.colorado.gov/
  - Denver open data: https://www.denvergov.org/opendata
  - Quandl : https://www.quandl.com/
  - https://github.com/awesomedata/awesome-public-datasets
  - https://github.com/datasets
- Munging, merging, and feature engineering

Part 3: Data Acquisition & Processing

- We have a few datasets to pick through - where do we start?
  - https://data.colorado.gov/browse?q=marijuana&sortBy=relevance

Part 3: Data Acquisition & Processing

- We have a few datasets to pick through - where do we start?
  - https://data.colorado.gov/browse?q=marijuana&sortBy=relevance
- What should we be predicting?

Part 3: Data Acquisition & Processing

- We have a few datasets to pick through - where do we start?
  - https://data.colorado.gov/browse?q=marijuana&sortBy=relevance
- What should we be predicting?
- What data might support our prediction?

Part 3: Data Acquisition & Processing

- We have a few datasets to pick through - where do we start?
  - https://data.colorado.gov/browse?q=marijuana&sortBy=relevance
- What should we be predicting?
  - The most profitable county to open a rec dispensary
  - The county with the greatest expected revenue increase for the next rec store opened
- What data might support our prediction?
  - Revenue by county - med, rec, and total
  - Number of stores by county - med, rec, and total
  - Population of county (now & projected)
  - Other details on the shops or people to assist in segmentation

Part 3: Data Acquisition & Processing

- Steps
  - Get started with the dataset with the most to offer
  - Clean up and process the data to a useable form
  - Combine all of the data into one master dataset
  - Profit!

Part 4: Modelling With scikit-learn

- Why scikit-learn?

Part 4: Modelling With scikit-learn

- Why scikit-learn?
  - I don’t want to write a tensorflow install script and help everyone debug the resulting mess...

Part 4: Modelling With scikit-learn

- Why scikit-learn?
  - ~~I don't want to write a tensorflow install script and help everyone debug the resulting mess...~~

Part 4: Modelling With scikit-learn

- Why scikit-learn?
  - ~~I don't want to write a tensorflow install script and help everyone debug the install...~~
- Full suite of ML modelling techniques with a consistent api

Part 4: Modelling With scikit-learn

- Why scikit-learn?
  - ~~I don't want to write a tensorflow install script and help everyone debug the install...~~
- Full suite of ML modelling techniques with a consistent api

Part 4: Modelling With scikit-learn

- Why scikit-learn?
  - ~~I don't want to write a tensorflow install script and help everyone debug the install...~~
- Full suite of ML modelling techniques with a consistent api
- Many pre-built utility functions

Part 4: Modelling With scikit-learn

- Let’s see the code!

Part 5: Web Development Part Two; Integration

- Time to tie it all together!
  - View training data via web browser
  - Create a dynamically rendered visualization with Bokeh
  - Build a method to dynamically update our data
  - Pretty the place up a bit

Part 5: Web Development Part Two; Integration

- Time to tie it all together!

Part 5: Web Development Part Two; Integration

- Time to tie it all together!
  1. Pretty the place up a bit

Part 5: Web Development Part Two; Integration

- Time to tie it all together!
  1. Pretty the place up a bit
  2. Create a method to view training data in the browser

Part 5: Web Development Part Two; Integration

- Time to tie it all together!
  1. Pretty the place up a bit
  2. Create a method to view training data in the browser
  3. Create a dynamically rendered visualization with Bokeh

Part 5: Web Development Part Two; Integration

- Time to tie it all together!
  1. Pretty the place up a bit
  2. Create a method to view training data in the browser
  3. Create a dynamically rendered visualization with Bokeh
  4. Build a method to dynamically update our data

Stretch Options; aka We Still Have Time!!!??

- Stretch Options
  - Dig deeper on Bokeh
  - Add additional elements to the front end
  - Scheduled data updates
  - Expose trained model as an api