CONCEPTS IDEAS AND TODO

February 3, 2022

1 Topics, Skills, and Methods To Potentially Cover and/or Incorporate

- Docker
- Kubernetes
- Terraform
- multicore processing
- cuda/GPU computing
- Feature Selection in Data Science and ML
- Data roles and their differences: Data Scientist, Data Engineer, Machine Learning Engineer, Data Analyst, etc.
- Big Data fundamentals (mostly covered in my mini course on spark)
- Pipeline and different parts of a Data Science project.
- Biases and AI ethics.
- XAI methods.
- ML deployment and monitoring.
- Audio processing block
- Snowflake
- Tableau
- SAS (Afia?)
- Kibana, Confluence, TIcketing, Asana, and other productivity and progress seting/monitoring software
- Reveal.js and Slides.com for Presentations - Host Your Beautiful Slideshow on AWS
- Creating effective presentations:
  - Short Talks - (Internal to pitch ideas/projects to supervisor)
  - Pitches and presentations general
  - Stakeholder Communications and Progress reports
  - Conference talks
  - Constructing Effective Poster Presentations
- How to Create Meaningful and Complete README.md files (should be both informative and eye-catching w.r.t. markup aesthetic)
- How To Appropriately Document Projects: How-To, The Does, and The Don'ts
- Creating Short Project Demos (video presentations highlight functionality and significance)
- Good Coding Practices
  - Pep-8 Standards
    python: using Black and Lint to guide you

  - The importance of Type hinting, Commenting Lines, Doc Strings, and Usage Examples
  - Importance of Writing Quality and Robust Unit and Functional Tests
- Create Cheat sheets where applicable/useful
- Securely Storing API Credentials Locally and Accessing them Programmatically WHen Needed (created jupyter notebook)
- Essential Math
  - The Kernel Trick
  - MMD, Kernel Trick and Deep Learning

2 Approach

How do we incorporate concepts we feel important into a learning/self-learning framework? Need to discuss this with Chetan, Afia, Nishi, (Maybe Sergio?) and other potentially contributors.
For advanced topics it is less necessary to provide as. much strict direction and 'prereqs' since those already in the field will understand their current state of knowledge and could therefore pick and choose topic level concepts much more efficiently/meaningfully.

3 Accompanying/Supplental Documentation

In the case of Sergio's suggestion to include for example "Data roles and their differences: Data Scientist, Data Engineer, Machine Learning Engineer, Data Analyst, etc." Should we write a page or two on the topic and have a separate handbook on our interpretations and definitions we feel best describe an idea or concept? Also, I wouldn't mind created a 30-60 minute video for each case like this. It could be a pocket reference for core concepts with short accompanying videos. This can apply more to the and intermediate users, but we can easily think of a few cases where an advanced learner might benefit, i.e. A 1-2 page write up on Multi-armed bandit Problems.... with an Accompanying 30-60 minute discussion which can me technical or not... we can discuss the importance and usage cases of such concepts to motivate further exploration, but we don't necessarily have to do a deep dive on everything ourselves, just motivate and provide external resources for those wanting to pursue the topic we introduced them to.

4 Suggestion

- plus minus in sports
- sports analysis articles (by Nate Silver "the king of data science")
- soccer analytics book
- deep soccer analytics articles
- analytic basketball videos
- NFL combine analysis (1)(2)(3)