DS Development Best Practices

What to expect ?

- DS practice dos & don’ts
- Some advice from experience

Best Practice with Facts:

- Keep It Simple: you’ll live longer
- First target MVP: you & your Product lead will live longer
- DS is a continuous process: no job loss
- Always Upgrade yourself: you’ll be in competitive stage
- Be Humble & Accept the Fact: you’ll have peace
- Be an Innovator & Initiative taker: you’ll be on edge
- Always stay connected with Basics: you’ll save a lot time
- Follow Best Practice in Software Engineering.
- Don’t get hyped & overwhelmed by latest Gen-AI news
- Commercial Product matters most

Python Project Structure

```
Project/
    Config/
    Data/
    Logs/
    Ml_Experiments/
    Models/
    Src/
        base_class/
        logger/
        db/
        s3/
        ml/
        main.py
    Test/
       test_base_class.py
       test_logger.py
       test_db.py
       test_main.py
    README.md
    Requirements.txt
    Setup.py
    Dockerfile
```

Time Saver Tips

- Always create virtual environments for your specific project
- Use git bash for execution of code & management of code
- Python packages for coding standard ‘Pylint, Black, Isort, Flake8..etc’
- Python Packages for data encapsulation & data I/O validation – ‘Typing, dataclasses & pydantic’
- Communication between functions or methods with JSON
- Following SOLID principles for functional design
- Follow Concept Principles for simple structure
- Use of Logging instead Print()
- Use Global configurations throughout application
- Add DocString in module/class/function and hash-comment before each complex block
- Be careful with python namespaces(specifically main name space)
- Do experiment in whichever IDE, but once done , add to your project structure
- Run unit-test cases for your components & entire application
- Update README.md file with Steps-to-execute, version , new changes & unit test result
- If it can be used as package, then use setuptools with setup.py to make a .wheel file
- Push code to BitBucket after above steps complete. Raise PR & make sure that’s gets reviewed and approved.
- Add Documentations of your Application usage, test outputs..etc in confluence
- In a Sprint work, when working on task, it’s always good practice to task description , findings & outcomes

Data & Data-Science

- Limited Data for your Project is Fact & Accept it
- Plan & Strategize your work
- Dos:
  - Observe Data: Identify Patterns(Any Decision Boundary is present?)
  - Create Data Label if not provided
  - EDA
  - Model Selection
  - Train & Test
  - Process Continues… (MlOps)
  - Automation(adding to project, script to run experiments & getting metrics)
- Don’ts:
  - Blind Selection of Model based on Hype
  - Not understanding Algorithm before using it
  - Using directly any pre-trained model before understanding on what data , model is being trained
  - Not understanding problem statement & it’s source
  - Not understanding pros & cons of each Algorithm
  - Not making independent data & model evaluation

Conclusion

- Have Patience & Continue Learning
- Be Transparent & Kind to your Colleagues
- Discuss your Learning, Approach with colleagues
- Don’t limit yourself to one skill
- Don’t shy to take help
- Look for Mentor