Building AI solutions for a billion users

From Notebooks to MLOps

Disclaimer

- Thoughts and ideas are my own. They do not represent my team or my company.
- The purpose of this presentation to drive discussion and learning. Any infringement is not intended.

A billion customer website - Yosemite

- E-Retail website which offers instant delivery to its customers
- Provides great offers regularly to customers
- Periodically offers sale over short period of time - great steal deals
- Drives more than a billion transactions annually
- Needs to manage its infrastructure to:
  - Optimize the cost
  - Provide reliability to customers

Requirements

A model which can understand the patterns of traffic and predict the traffic every hour.

Available data:

1. Hourly aggregated data of traffic for the past year
2. Infrastructure deployment over the past year
3. Aggregated performance of infrastructure over the past year

How do you experiment today?

Challenges

1. Data is too large to fit any local machine
2. Reproducibility is required in results of the model
3. Access of data is business sensitive
4. Need to able to pinpoint changes in case of failure and roll back
5. The model should be explainable

How to overcome these challenges?

MLOps!

MLOps

- Experimentation as a continuous process
- Continuous Versioning - Data, Model, Steps
- Continuous Validation - Metrics of the model
- Provides visibility into every step
- Enables natural evolution of models - just like software
- Enables A/B experimentation for further data driven decision making

Continuous experimentation

1. Various training methodologies
2. Various datasets
3. Various parameters and hyper parameters
4. Various environments

Continuous versioning

1. What data did the model work with? What was the quality of the data? What was the characteristic of the data? - Data versioning

2. What is the difference between the previous model and the current model? Can we revert back to/ reuse one of the previous models? - Model versioning

3. Can we change the step X of the this model to Y? Can we reuse this step from that model? - Step versioning

Continuous validation

- How much more efficient is the new model over the old one?
- What is the accuracy of the new model?
- What is the rate of false positives?

Visibility

1. How is the data transformed after every step?
2. How does this step contribute towards the model performance?
3. Can I reuse this step for other model?
4. Can I update this step?

Natural evolution of models

1. Model design bottlenecks
2. Technological advancements
3. Evolution of business requirements

A/B experimentation

- Model confidence in training, but production scenarios?
- Incremental rollout?

Other benefits

1. Data Governance
2. Audits
3. Repeatability
4. Reusability
5. Extensibility
6. Scalability

Questions

Thank you