WEEK_9_DAY_4_AFTER_NOON

- Monitoring
- Deployment
- Model monitoring

Monitoring

ML model monitoring is the practice of tracking the performance of ML models in production to identify potential issues that can add negative business value. These practices help proactively monitor prediction quality issues, data relevance, model accuracy, and bias.

ML monitoring constitutes the subset of AI observability where it showcases a bigger picture with testing, validation, explainability, and exploring unforeseen failure modes. The performance of ML models starts degrading over time. It can be due to data inconsistencies, skews, and drifts, making deployed models inaccurate and irrelevant. Appropriate ML monitoring helps identify precisely when the model performance started diminishing. Such proactive monitoring helps take required actions like retraining models or replacing models. It helps foster users' trust in ML systems.

Deployment

After training and analyzing the model, it's time to deploy the model. An ML model can be deployed in three ways, which are:

- Using the Model server,
- In a Browser
- On Edge device