Streamlining ML Model Deployment - MLFlow vs Vertex AI

Team lead: Swetha Vijaya Raju
Team members: Ankita Shetty, Kirtana Kirtivasan, Krishna Sai Tejaswini Kambhampati

Company: Lead is a fictitious mid-size company with 170 employees. Lead's main mission is to provide a platform for communities, which not only provides real time navigation on roads, but also reports accidents, traffic jams, road works, constructions, community driven points of interests and many more activities that are the root cause of traffic congestion in cities.
Lead relies on large scale data solutions on a daily basis as they attempt to solve problems like- Predicting ETA, Matching Riders & Drivers and Serving the right advertisements. There are multiple engineers (Data scientists, Software engineers, reliability engineers, Data engineers and many more) working on opportunistically including ML solutions for their application.

Problem Summary

Lead uses multiple ML models to explore and develop their ML algorithms for various use cases; performing semi-manual operations for training, validation and deployment; limited monitoring and validation capabilities after a model is deployed; and a hideously long deployment cycle from idea to production – This is a challenge that manifested itself causing a chaotic state at Lead. Currently, Lead does not have a service to manage the ML lifecycle.

Solution Summary

The major objective of this project is to streamline the process and minimize the timelines for deployment by reducing the manual work wherever possible in the life cycle of the algorithm, and comparatively reduce overall costs of the system using managed services. Invest in deployment processes through managed services (such as MLFlow or Vertex AI) and ensure that, by using these streamlined systems, a single Data Scientist could be able to close the product cycle from research to production.

Recommendation:

Vertex AI

- Vertex AI provides a unified platform that can accommodate data readiness to ML model deployment streamlining the entire ML lifecycle while MLFlow will only help in Experimentation Model tracking.
- With Vertex AI, a single Data Scientist could close the product cycle from research to production.
- Vertex AI natively uses Google infrastructure and our system is majorly on Google Cloud. So the adoption of Vertex AI to our current system is compatible**.**

Success Criteria:

- The adoption of Vertex AI will bring down the operational cost of ML models in Lead by 34% in the next 5 years.
- With the adoption of Vertex AI, deployment timeline will be reduced by 22% approximately.