Streamlining ML model Deployment

Team2 - Ankitha Shetty, Kirtana Kirtivasan, Swetha Vijaya Raju, Tejaswini Kambhampati

About The Company (Fictitious)

- Lead is a GPS-enabled navigation application that uses real-time user location data and user-submitted reports to suggest optimized routes
- Company size : 100-150 employees
- Reference use case: Waze

Business Problem

- Lead's Data science teams generate tens of ml and dl models, (using python) from different verticals, on a daily basis which needs to be translated to job schedulers to deploy them into production environment
- Lead's Engineering and Operation teams are also required to monitor the health of these data pipelines
- This process is time consuming,extremely mechanical and iterative in nature
- They want to find the appropriate service/solution provider to standardize, streamline their model deployment and monitoring needs

Business Problem

- Lead being a small company, still at it's growing stage, has a small employee base of around 150.
- The roles are overlapped and stakeholders wear multiple hats most of the times.
- But, currently, the whole deployment process easily takes up to over 3 months for Lead.
- Why?
  - Multiple stakeholders involved
  - Multiple skill sets involved
- Lead cannot afford to spend so much time on deployment because they have already exceeded the labor cost for the project by 30%

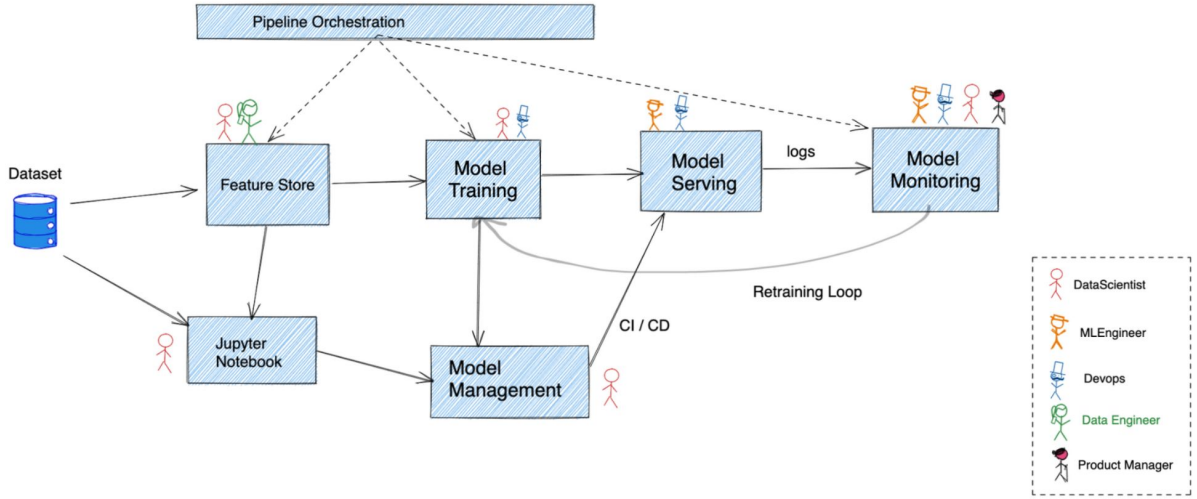
Reference - Impactful ML Model

Business Problem

- Lead has to deploy the ML model at the same time as they build a scalable framework to support future modeling activities.
- Based on the deployment cost spent for the model until now, Lead extrapolated a deployment and maintenance cost of $175,000 for the next five years.
- The allocated budget for the model deployment and maintenance for the next 5 years is $95,000
- It’s critical that we need a third party platform that we can purchase to streamline this whole deployment process end to end.
- This platform should essentially bring down the overall model deployment cost by 45%.

Reference - What is the Cost to Deploy and Maintain a Machine Learning Model?

WorkFlow



Overall Pipeline with different stakeholders

1. Project Ideation
2. Data Gathering
3. Data Analysis
4. Feature Engineering
5. Model Training
6. Model Serving
7. Product Integration
8. Model Monitoring
    a. System Monitoring
    b. Model Monitoring
9. Complete Automation

Our project sits here

Image from [Impactful ML Model]

Project Objectives

Technical Objectives:

- Easy to set up service, that decrease latency and speed up the deployment process by 10%
- Simplified logging, login time must be less than 10 minutes
- Version Control to track to track every individual change by each contributor,compare differences and avoid conflicts
- Initiate hourly batch requests to the same model, so hardware is used efficiently

Project Objectives

Political Objectives

- Promotes efficient collaboration among ML Engineers, Data Scientists, Product Managers and Site Reliability Engineers and increase the productivity by 5%
- Foster compatibility with the existing system

Financial Objectives

- Free to use open source platforms thus reducing the deployment cost by 45%
- Modular cost, based on the feature you use

Recommendations: Vertex AI

- Provider - Google
- Vertex AI Pipelines helps you to automate, monitor, and govern your ML systems by orchestrating your ML workflow in a serverless manner, and storing your workflow's artifacts using Vertex ML Metadata

Advantages:

- Provides manual labeling of pipeline
- Train and compare models using AutoML or custom code training and all your models are stored in one central model repository.[1]
- Vertex supports computer vision model explainability using a google approach XRAI[1]

Reference - Vertex AI

Recommendations: MLFlow

- Provider - Databricks
- The platform can be used for ML deployment by individual developers as well as teams
- It can be incorporated into any programming ecosystem
- The library is built to satisfy various technological needs and can be used with different machine learning libraries

Advantages:

- An open-source tool
- The logging is simplified, so it’s easy to run experiments
- When it comes to training, tuning and deploying ML models, it brings transparency and standardization

Project Benefits

- Overall customer satisfaction score should go up 5%
- Reduce manual labour of deployment by 10%
- Increase collaboration between the teams involved by 5%
- Enhance scalability and reproducibility