Diamond Dataset ML Pipeline Orchestration and Experiment Tracking.

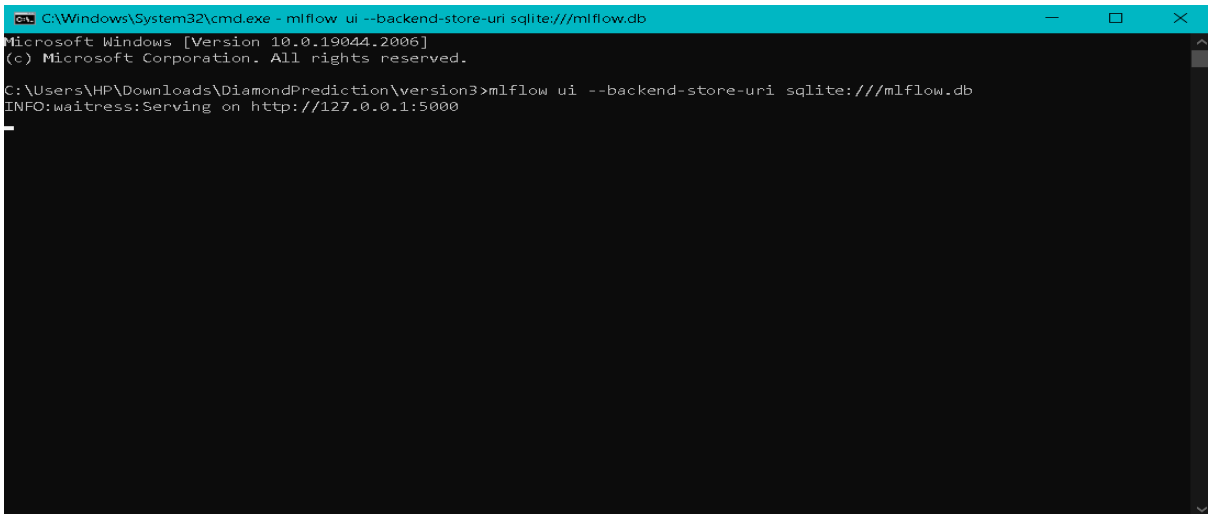
Experiment Tracking: -

Experiment tracking is the process of saving all experiment related information that you care about for every experiment you run. This “metadata you care about” will strongly depend on your project, but it may include:

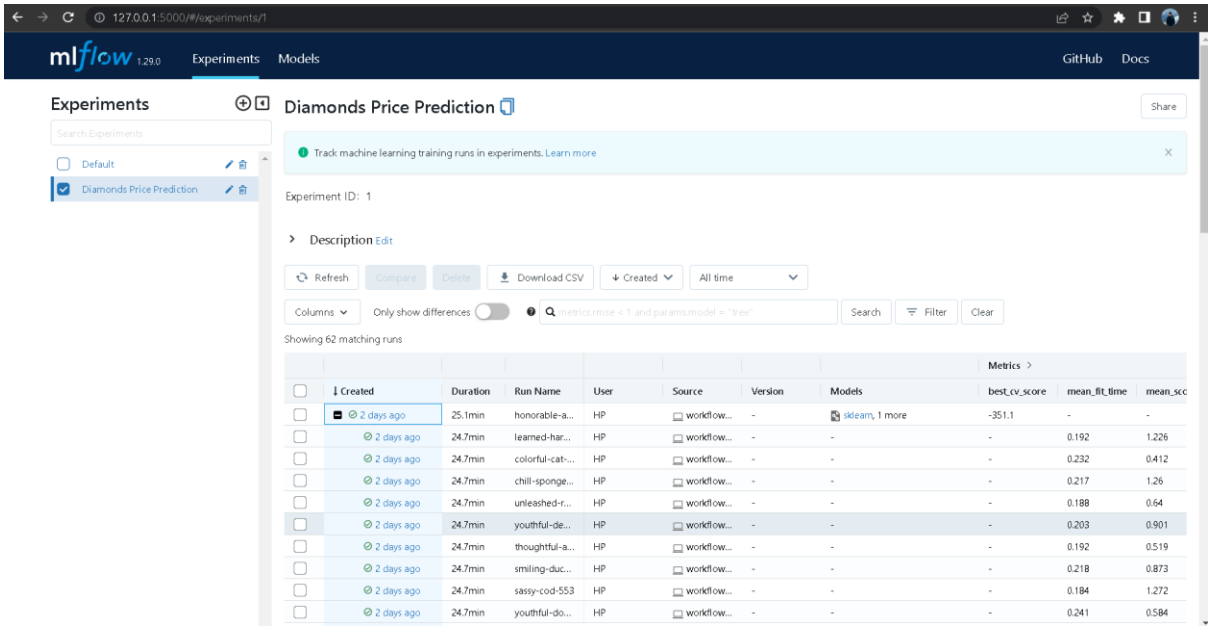
- Scripts used for running the experiment
- Environment configuration files
- Versions of the data used for training and evaluation
- Parameter configurations
- Evaluation metrics
- Model weights
- Performance visualizations

Experiment tracking focuses on the iterative model development phase when you try many things to get your model performance to the level you need.

- In this project I used ‘*MLFlow*’ tool for experiment tracking.



Command for running MLFLow.



MLFlow Dashboard

ML Pipeline Orchestration: -

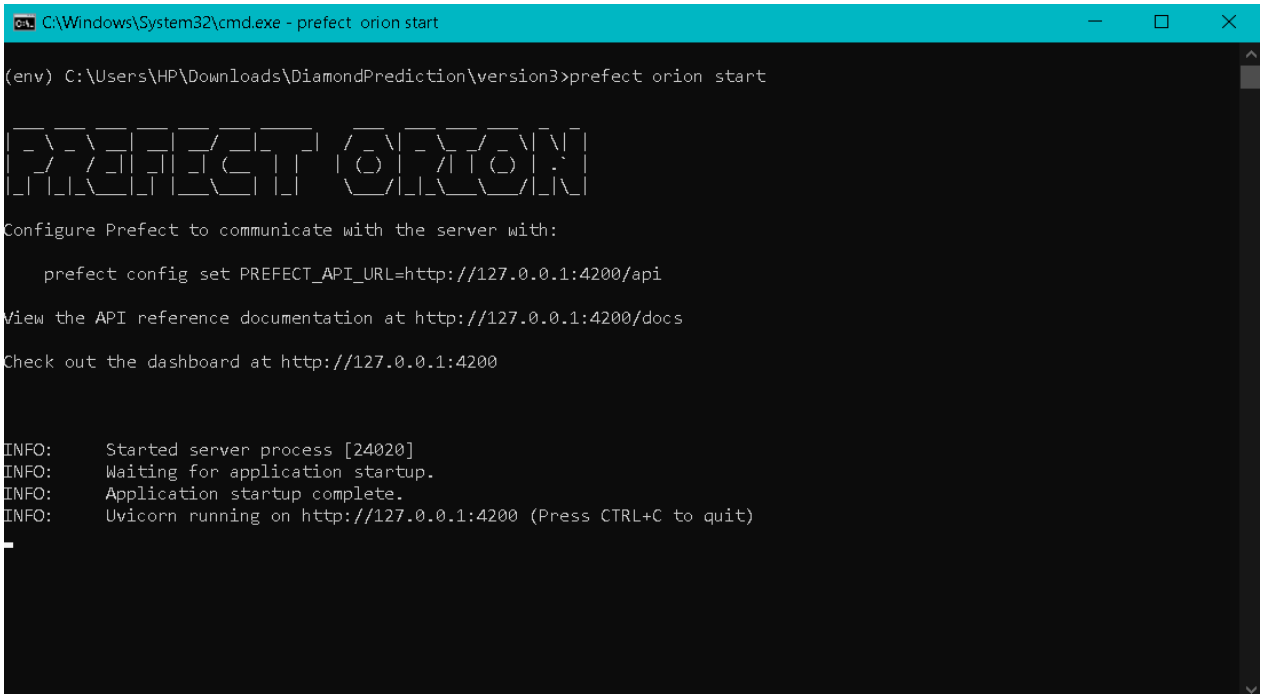
ML Pipeline Orchestration basically means the workflow of our Machine Learning Project in which we convert python jupyter notebook to Python Script for Production ready code.

Pipelines in machine learning are an infrastructural medium for the medium for the entire ML workflow. Pipelines help automate the overall MLOps workflow, from data gathering, EDA, data arguments, to model building and deployment. After the deployment, it also supports reproduction, tracking and monitoring.

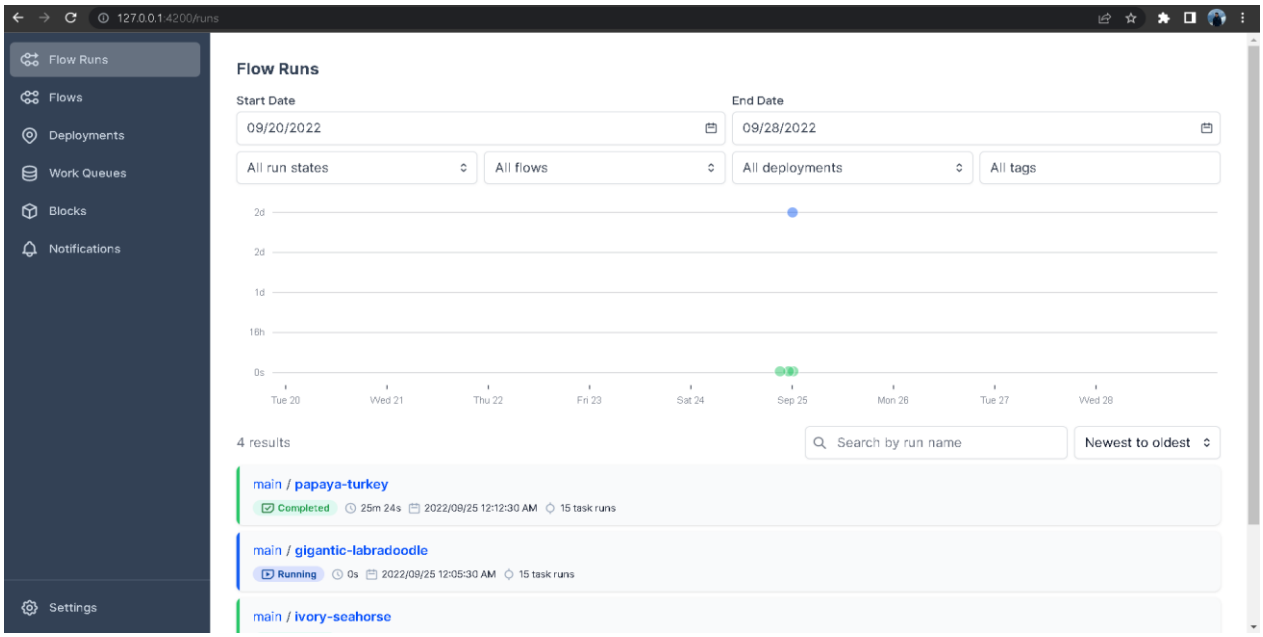
ML pipelines help improve the performance and management of the entire model, resulting in quick and easy deployment.

Machine Learning Orchestration tool :- ML orchestration tools are used to automate and manage workflows and pipeline infrastructure, with a simple, collaborative interface, Along with management and creation of custom workflows and their pipelines, these tools also help us track and monitor models for further analysis.

- In this project I used '*Prefect*' tool for pipeline orchestration.



cmd for running prefect



Prefect dashboard