

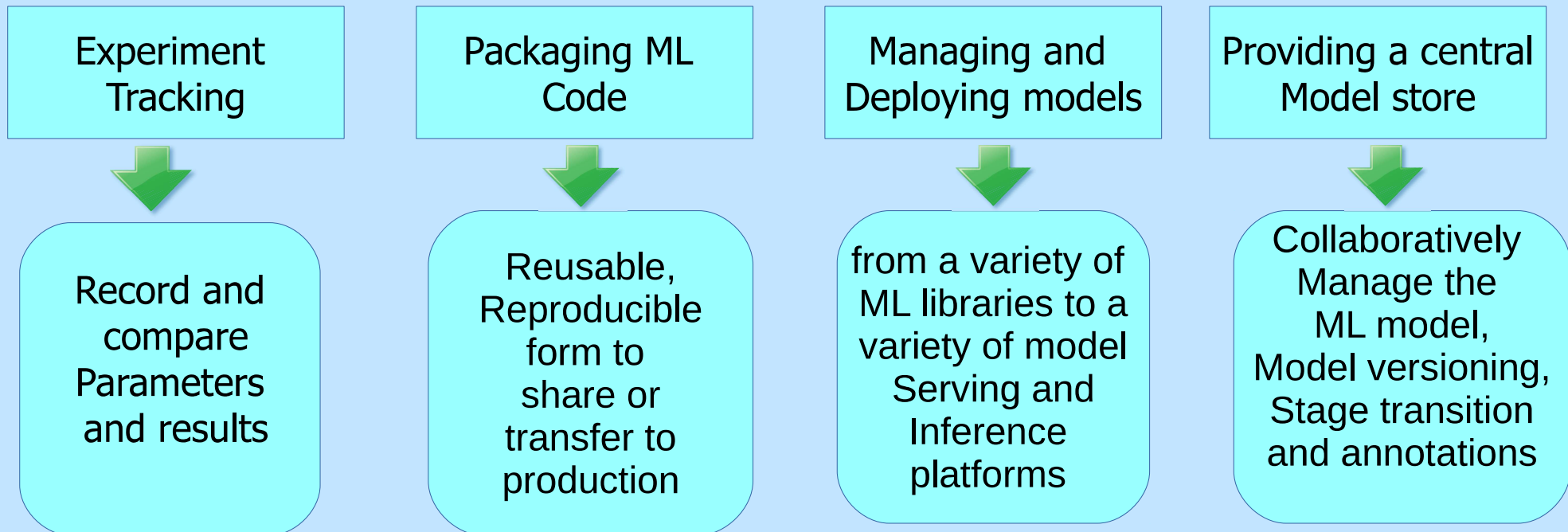
# MLflow

Using MLflow for Experiment Tracking and Model Management for Diamond data set

# Introduction to MLflow

MLflow is an open source platform for managing the end-to-end machine learning lifecycle.

It tackles four primary functions:



# Introduction to Experiment Tracking

## Terminologies:

- 1.Experiment
- 2.Run
- 3.Metadata (i.e. Tags, Parameters, Metrics)
- 4.Artifacts (i.e. Output files associated with experiment runs)

## What do you want to track for each Experiment Run?

- 1.Training and Validation Data Used
- 2.Hyperparameters
- 3.Metrics
- 4.Models

## Why Track?

Organization Optimization Reproducibility

# Uses of ML flow

## Tool - MLflow

MLFlow helps you to organize your experiments into runs.

MLflow keeps track of:

- Tags

- Parameters

- Metrics

- Models

- Artifact

- Source code, Start and End Time, Authors etc

# Working on MLflow

## Step 1 - Import MLFlow

```
import mlflow
```

## Step 2 - Set the tracker and experiment

```
mlflow.set_tracking_uri(DATABASE_URI)
```

```
mlflow.set_experiment("EXPERIMENT_NAME")
```

## Step 3 - Start a experiment run

```
with mlflow.start_run():
```

## Step 4 - Logging the metadata

```
mlflow.set_tag(KEY, VALUE)
```

```
mlflow.log_param(KEY, VALUE) mlflow.log_metric(KEY, VALUE)
```

## Step 5 - Logging the model and other files (2 ways)

```
Way 1 – mlflow.<FRAMEWORK>.log_model(MODEL_OBJECT,artifact_path="PATH")
```

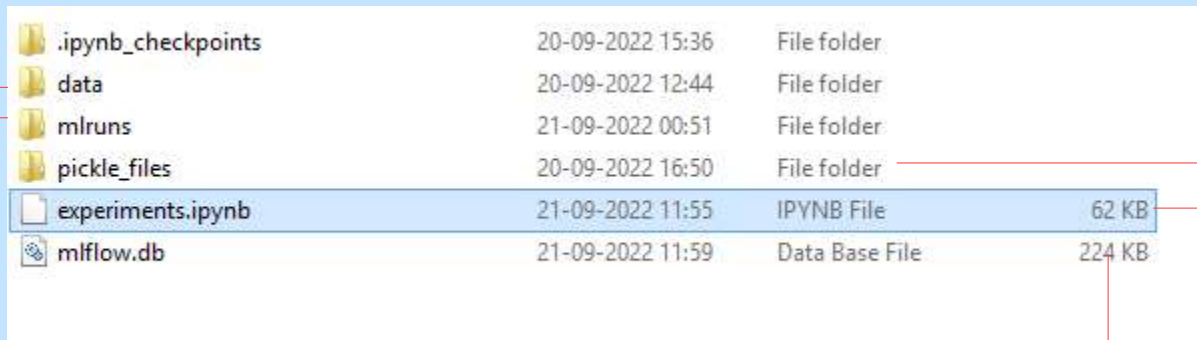
```
Way 2 - mlflow.log_artifact(LOCAL_PATH, artifact_path="PATH")
```

# Working on MLflow

Requirements :

Source:

[https://github.com/KusumaChalla/Innomatics\\_Research\\_Lab\\_Intenship2022/tree/main/Machine%20Learning/Experiments](https://github.com/KusumaChalla/Innomatics_Research_Lab_Intenship2022/tree/main/Machine%20Learning/Experiments)



.ipynb_checkpoints	20-09-2022 15:36	File folder	
data	20-09-2022 12:44	File folder	
mlruns	21-09-2022 00:51	File folder	
pickle_files	20-09-2022 16:50	File folder	
experiments.ipynb	21-09-2022 11:55	IPYNB File	62 KB
mlflow.db	21-09-2022 11:59	Data Base File	224 KB

Contains  
diamond.csv

Contains the experiment id  
folder i.e set\_experiment( )  
and logs of artifacts files and  
created while we run mlflow

Contains mlflow code

Contain pickle files like  
labelencoder, scaler  
pickle files

Created when  
mlflow ui --backend-store-uri sqlite:///mlflow.db is  
run on cmd terminal after mlflow installation and its the  
sqlite database that logs the models

# 1. Install mlflow

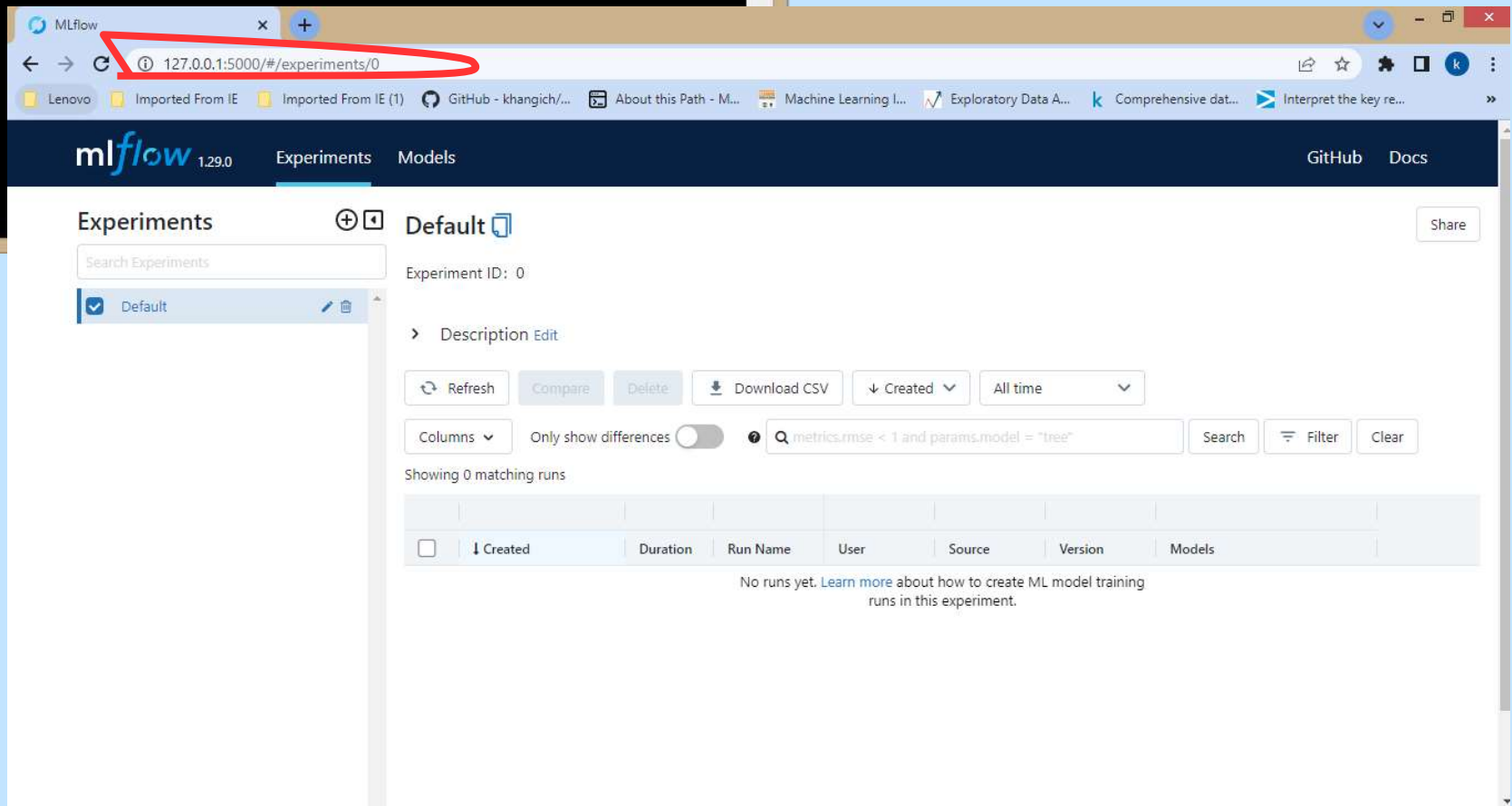
`pip install mlflow`

`mlflow ui --backend-store-uri sqlite:///mlflow.db`

```
C:\WINDOWS\system32\cmd.exe - mlflow ui --backend-store-uri sqlite:///mlflow.db
Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.

C:\Users\kusuma>cd C:\Users\kusuma\Desktop\Innomatics Lab Intership\AppDev\Diamendprice_StreamlitApp\Experiments

C:\Users\kusuma\Desktop\Innomatics Lab Intership\AppDev\Diamendprice_StreamlitApp\Experiments>mlflow ui --backend-store-uri sqlite:///mlflow.db
INFO:waitress:erving on http://127.0.0.1:5000
```



## 2.Run the MLFlow code

### Experiment run

```
import mlflow
```

```
mlflow.set_tracking_uri("sqlite:///mlflow.db")
```

```
mlflow.set_experiment("Diamond Price Prediction")
```



The screenshot shows the MLflow web interface in a browser. The address bar shows the URL `127.0.0.1:5000/#/experiments/2`. The interface has a dark blue header with the MLflow logo and version 1.29.0, and tabs for 'Experiments' and 'Models'. On the left, under 'Experiments', there is a search bar and a list of experiments. The 'Diamond\_Price Prediction' experiment is selected and highlighted with a red rectangle. The main panel shows details for this experiment, including 'Experiment ID: 2', a 'Description Edit' button, and a list of actions: 'Refresh', 'Compare', 'Delete', 'Download CSV', 'Created' (with a dropdown), and 'All time' (with a dropdown). Below these are filters for 'Columns', 'Only show differences' (a toggle), and a search bar containing the query `metrics.rmse < 1 and params.model = "tree"`. The results section shows 'Showing 0 matching runs' and a table with columns: 'Created', 'Duration', 'Run Name', 'User', 'Source', 'Version', and 'Models'. A message at the bottom states: 'No runs yet. Learn more about how to create ML model training runs in this experiment.'



### 3.Run the experiments/models

Ex :KneighborsRegressor model



#### Experiment 2 -Training KNeighborsRegressor

```
In [41]: with mlflow.start_run():  
    mlflow.set_tag("dev", "Kusuma")  
    mlflow.set_tag("algorithm", "KNN")  
    # Log the data for each run using log_param, log_metric, log_model  
    mlflow.log_param("data-path", "data/diamonds.csv")  
    k = 5  
    mlflow.log_param("n_neighbors", k)  
    knn_regressor = KNeighborsRegressor(n_neighbors=k)  
    knn_regressor.fit(X_train_transformed, y_train)  
    y_test_pred = knn_regressor.predict(X_test_transformed)  
    MAE = metrics.mean_absolute_error(y_test, y_test_pred)  
    mlflow.log_metric("Mean Absolute Error", MAE)  
    mlflow.sklearn.log_model(knn_regressor, artifact_path="models")  
    mlflow.log_artifact("pickle_files/label_encoderfile.pkl")  
    mlflow.log_artifact("pickle_files/standard_scaler.pkl")
```

MLflow 1.29.0 Experiments Models GitHub Docs

### Experiments

Search Experiments

- ☐ Default
- ☒ Diamond\_Price Prediction

### Diamond\_Price Prediction

Experiment ID: 2

Description Edit

Refresh Compare Delete Download CSV Created All time

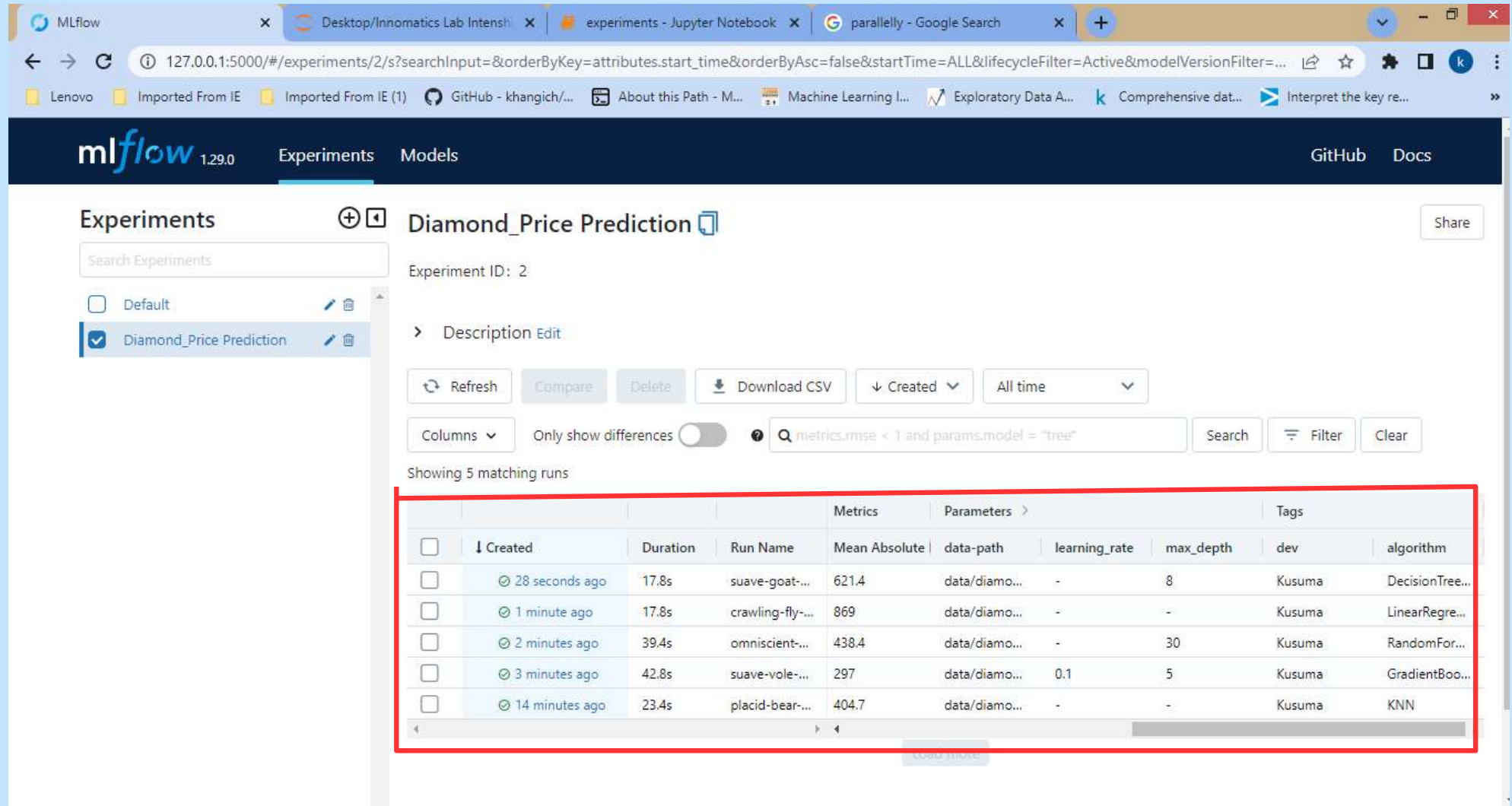
Columns Only show differences ☐ metrics.rmse < 1 and params.model = "tree" Search Filter Clear

Showing 1 matching run

	Created	Duration	Run Name	User	Sc	Ve	Models	Mean	data-p	n_neig	dev	algorithm
<input type="checkbox"/>	36 seconds ago	23.4s	placid-be...	kusuma			sklea	40...	da...	5	Kusuma	KNN

Load more

## 4.Run all the experiments and check parallelly the Mlflow UI to track the experiments



The screenshot displays the MLflow Experiments interface in a web browser. The browser's address bar shows the URL: `127.0.0.1:5000/#/experiments/2/s?searchInput=&orderByKey=attributes.start_time&orderByAsc=false&startTime=ALL&lifecycleFilter=Active&modelVersionFilter=...`. The MLflow logo and version 1.29.0 are visible in the top left. The 'Experiments' tab is selected, and the experiment name 'Diamond\_Price Prediction' is shown. The experiment ID is 2. A search bar on the left contains 'Diamond\_Price Prediction'. The main area shows a list of 5 runs, which are highlighted with a red box. The runs are sorted by 'Created' time, showing the most recent first. The table columns include 'Created', 'Duration', 'Run Name', 'Metrics', 'Parameters', and 'Tags'. The 'Metrics' column shows the 'Mean Absolute' error for each run. The 'Parameters' column shows the 'data-path', 'learning\_rate', and 'max\_depth' for each run. The 'Tags' column shows the 'dev' and 'algorithm' for each run.

Experiments

Search Experiments

Default

Diamond\_Price Prediction

Diamond\_Price Prediction

Experiment ID: 2

Description Edit

Refresh Compare Delete Download CSV Created All time

Columns Only show differences ☐ ☐ metrics.rmse < 1 and params.model = "tree" Search Filter Clear

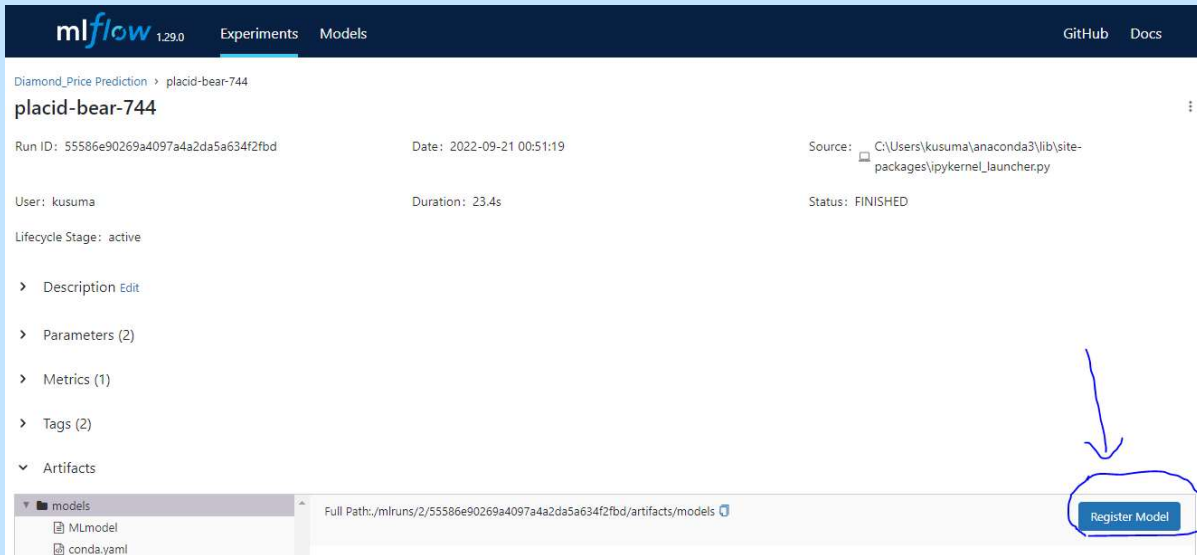
Showing 5 matching runs

				Metrics	Parameters	Tags			
	↓ Created	Duration	Run Name	Mean Absolute	data-path	learning_rate	max_depth	dev	algorithm
<input type="checkbox"/>	28 seconds ago	17.8s	suave-goat-...	621.4	data/diamo...	-	8	Kusuma	DecisionTree...
<input type="checkbox"/>	1 minute ago	17.8s	crawling-fly-...	869	data/diamo...	-	-	Kusuma	LinearRegre...
<input type="checkbox"/>	2 minutes ago	39.4s	omniscient-...	438.4	data/diamo...	-	30	Kusuma	RandomFor...
<input type="checkbox"/>	3 minutes ago	42.8s	suave-vole-...	297	data/diamo...	0.1	5	Kusuma	GradientBoo...
<input type="checkbox"/>	14 minutes ago	23.4s	placid-bear-...	404.7	data/diamo...	-	-	Kusuma	KNN

Load more

## 5. Select the experiments you want to register model

EX: Registered the experiments under a model name



mlflow 1.29.0 Experiments Models GitHub Docs

Diamond\_Price Prediction > placid-bear-744

Run ID: 55586e90269a4097a4a2da5a634f2fbd Date: 2022-09-21 00:51:19 Source: C:\Users\kusuma\anaconda3\lib\site-packages\ipykernel\_launcher.py

User: kusuma Duration: 23.4s Status: FINISHED

Lifecycle Stage: active

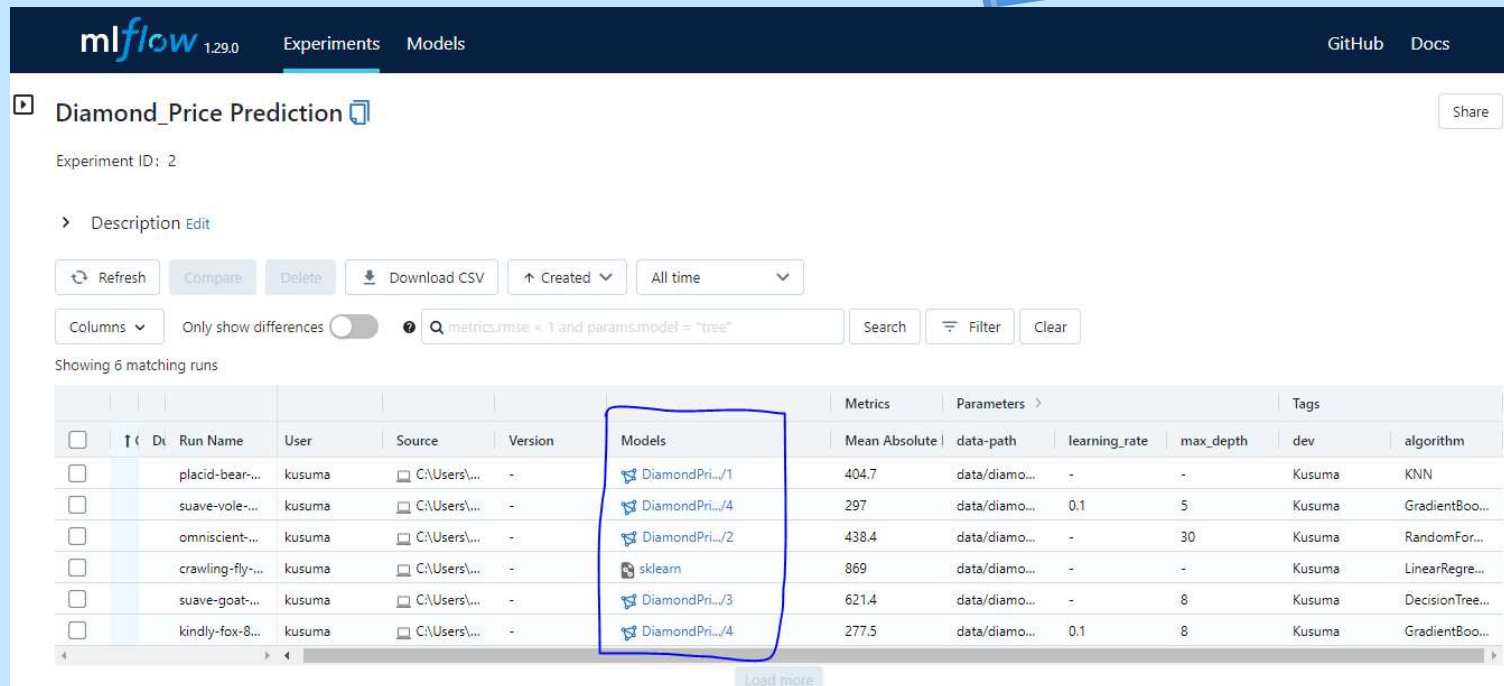
- > Description Edit
- > Parameters (2)
- > Metrics (1)
- > Tags (2)
- > Artifacts

models

- MLmodel
- conda.yaml

Full Path: mlruns/2/55586e90269a4097a4a2da5a634f2fbd/artifacts/models

Register Model



mlflow 1.29.0 Experiments Models GitHub Docs

Diamond\_Price Prediction Experiment ID: 2 Share

> Description Edit

Refresh Compare Delete Download CSV Created All time


Columns Only show differences metrics.rmse < 1 and params.model = "tree" Search Filter Clear

Showing 6 matching runs

							Metrics	Parameters	Tags			
	Run Name	User	Source	Version	Models		Mean Absolute	data-path	learning_rate	max_depth	dev	algorithm
	placid-bear-...	kusuma	C:\Users\...	-	DiamondPri.../1	404.7	data/diamo...	-	-	-	Kusuma	KNN
	suave-vole-...	kusuma	C:\Users\...	-	DiamondPri.../4	297	data/diamo...	0.1	5	-	Kusuma	GradientBoo...
	omniscient-...	kusuma	C:\Users\...	-	DiamondPri.../2	438.4	data/diamo...	-	30	-	Kusuma	RandomFor...
	crawling-fly-...	kusuma	C:\Users\...	-	sklearn	869	data/diamo...	-	-	-	Kusuma	LinearRegre...
	suave-goat-...	kusuma	C:\Users\...	-	DiamondPri.../3	621.4	data/diamo...	-	8	-	Kusuma	DecisionTree...
	kindly-fox-8...	kusuma	C:\Users\...	-	DiamondPri.../4	277.5	data/diamo...	0.1	8	-	Kusuma	GradientBoo...

Load more

## 6. Check the registered models

 1.29.0 [Experiments](#) [Models](#) [GitHub](#) [Docs](#)

Registered Models > [DiamondPrice\\_predict](#)

### DiamondPrice\_predict

Created Time: 2022-09-21 01:24:31 Last Modified: 2022-09-21 01:28:08

> Description [Edit](#)

> Tags

> Versions 

All Active 0 Compare

<input type="checkbox"/>	Version	Registered at	Created by	Stage	Description
<input type="checkbox"/>	<input checked="" type="checkbox"/> Version 1	2022-09-21 01:24:32		None	
<input type="checkbox"/>	<input checked="" type="checkbox"/> Version 2	2022-09-21 01:25:05		None	
<input type="checkbox"/>	<input checked="" type="checkbox"/> Version 3	2022-09-21 01:25:39		None	
<input type="checkbox"/>	<input checked="" type="checkbox"/> Version 4	2022-09-21 01:28:08		None	

1

## 7. Name the registered models with their respective functional name and assign a stage transaction for the model

The screenshot shows the MLflow interface for a registered model. The top navigation bar includes the MLflow logo (1.29.0), 'Experiments', and 'Models' (which is selected). On the right, there are links for 'GitHub' and 'Docs'.

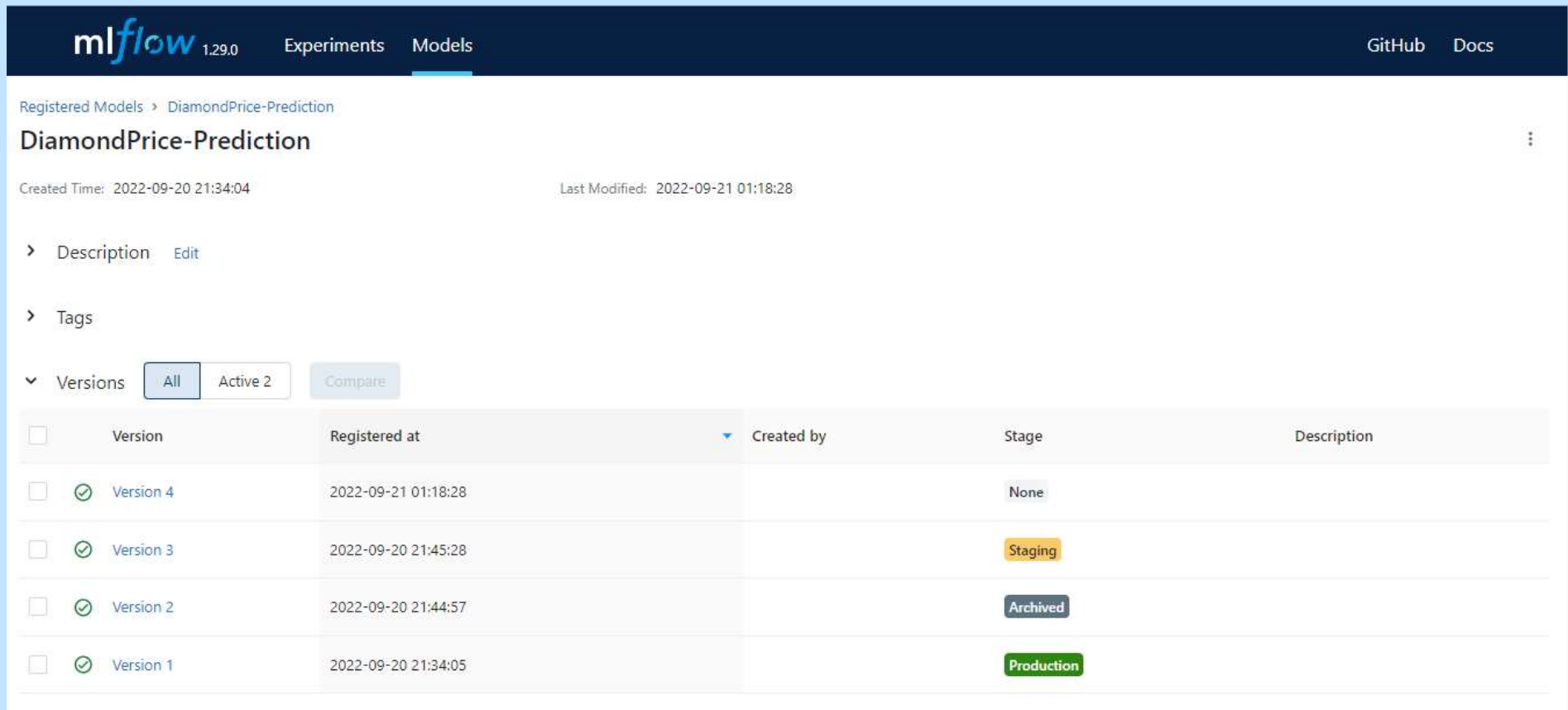
The breadcrumb trail is 'Registered Models > DiamondPrice\_predict > Version 4'. The main heading is 'Version 4'. Below this, it shows 'Registered At: 2022-09-21 01:28:08' and 'Source Run: kindly-fox-825'. There are links for 'Description Edit' and a 'Tags' section.

A red box highlights the 'Stage' dropdown menu, which is currently set to 'None'. The dropdown options are: 'Transition to → Staging' (yellow button), 'Transition to → Production' (green button), and 'Transition to → Archived' (blue button).

Below the stage menu, there is a table with columns 'Name', 'Value', and 'Actions'. The table is currently empty, with the text 'No tags found.' displayed. A red box highlights the input area for adding a tag, which includes a text input field containing 'algorithm', a text input field containing 'GradientBoostingRegressi', and an 'Add' button.

Below the tag input area, there is a 'Schema' section. It contains a table with columns 'Name' and 'Type'. The table is currently empty, with the text 'No schema. See MLflow docs for how to include input and output schema with your model.' displayed.

## 8. Check the registered model and staging state



mlflow 1.29.0 Experiments Models GitHub Docs

Registered Models > DiamondPrice-Prediction

### DiamondPrice-Prediction

Created Time: 2022-09-20 21:34:04 Last Modified: 2022-09-21 01:18:28

> Description Edit

> Tags

▼ Versions All Active 2 Compare

<input type="checkbox"/>	Version	Registered at	Created by	Stage	Description
<input type="checkbox"/>	✓ Version 4	2022-09-21 01:18:28		None	
<input type="checkbox"/>	✓ Version 3	2022-09-20 21:45:28		Staging	
<input type="checkbox"/>	✓ Version 2	2022-09-20 21:44:57		Archived	
<input type="checkbox"/>	✓ Version 1	2022-09-20 21:34:05		Production	

### References:

<https://www.mlflow.org/docs/latest/index.html>

<https://github.com/bansalkanav>