



Project on

Sentiment Analysis of Real-time Flipkart Product Reviews Using MLflow for Experiment Tracking and Model Management

T. Mahesh Reddy

Intern ID: IN1240834

OBJECTIVE OF THE PROJECT

The objective of this task is to introduce you to MLflow for experiment tracking, model management, and reproducibility in machine learning projects for the Sentiment Analysis Project.

Experiments

Search Experiments

☐ Default

☒ Flipkart_sentiment_analysis

Flipkart_sentiment_analysis

Provide Feedback

Add Description

Share

metrics.rmse < 1 and params.model = "tree"

Time created

State: Active

Datasets

Sort: Created

Columns

Group by

+ New run

Table

Chart

Evaluation

Experimental

		Run Name	Created	Dataset	Duration	Source	Models
<input type="checkbox"/>	<input type="checkbox"/>	<div><div></div> Logistic_run</div>	<div><div></div> 36 minutes ago</div>	-	13.5min	C:\Users\...	<div><div></div> sklearn, 1 more</div>
<input type="checkbox"/>	<input type="checkbox"/>	<div><div></div> Random_run</div>	<div><div></div> 42 minutes ago</div>	-	6.4min	C:\Users\...	<div><div></div> sklearn, 1 more</div>
<input type="checkbox"/>	<input type="checkbox"/>	<div><div></div> Decision_run</div>	<div><div></div> 44 minutes ago</div>	-	1.7min	C:\Users\...	<div><div></div> sklearn, 1 more</div>
<input type="checkbox"/>	<input type="checkbox"/>	<div><div></div> Multinomial_run</div>	<div><div></div> 45 minutes ago</div>	-	54.2s	C:\Users\...	<div><div></div> sklearn, 1 more</div>
<input type="checkbox"/>	<input type="checkbox"/>	<div><div></div> SVC_run</div>	<div><div></div> 57 minutes ago</div>	-	11.9min	C:\Users\...	<div><div></div> sklearn, 1 more</div>
<input type="checkbox"/>	<input type="checkbox"/>	<div><div></div> Knn_run</div>	<div><div></div> 1 hour ago</div>	-	7.1min	C:\Users\...	<div><div></div> sklearn, 1 more</div>

MLFlow

MLflow is an open source platform for managing machine learning workflows. It is used by MLOPs teams and data scientists. MLflow has main components.

- **The experiment tracking** allows you to record machine model training sessions (called runs) and run queries using Python.
- **The model management** provides a standard unit for packaging and reusing machine learning models.
- **The deployment** is a standard format that packages a machine learning model with its metadata, such as dependencies and inference schema.

MLFlow Dashboard

- The runs were executed using different models, such as Logistic Regression, Random forest, Decision Tree, Multinomial Naive Bayes, Support Vector Classification (SVC), and K-Nearest Neighbors (Knn).

The screenshot displays the MLFlow Dashboard interface. The top navigation bar includes the MLFlow logo (version 2.11.3), a red-outlined 'Experiments' tab, and a 'Models' tab. On the right of the header are links for GitHub and Docs, along with a settings icon. The left sidebar shows a search bar and a list of experiments, with 'Flipkart_sentiment_analysis' selected. The main content area is titled 'Flipkart_sentiment_analysis' and includes a search filter 'metrics.rmse < 1 and params.model = "tree"', filter buttons for 'Time created', 'State: Active', 'Datasets', and 'Sort: Created', and a '+ New run' button. Below this is a table view showing experimental runs. The table has columns for Run Name, Created, Duration, Source, Models, and Metrics. The 'Experimental' tab is active, showing a list of runs with their respective metrics.

							Metrics				Par
<input type="checkbox"/>	<input type="checkbox"/>	Run Name	Created	Duration	Source	Models	best_cv_score	mean_score_tim	mean_test_score	clas	
<input type="checkbox"/>	<input type="checkbox"/>	Logistic_run	4 hours ago	13.5min	C:\Users\...	sklearn, 1 more	0.902652106...	-	-	-	
<input type="checkbox"/>	<input type="checkbox"/>	Random_run	4 hours ago	6.4min	C:\Users\...	Flipkart r.../3, 1 more	0.917316692...	-	-	-	
<input type="checkbox"/>	<input type="checkbox"/>	Decision_run	4 hours ago	1.7min	C:\Users\...	Flipkart r.../2, 1 more	0.914976599...	-	-	-	
<input type="checkbox"/>	<input type="checkbox"/>	Multinomial_run	4 hours ago	54.2s	C:\Users\...	Flipkart r.../1, 1 more	0.921684867...	-	-	-	
<input type="checkbox"/>	<input type="checkbox"/>	SVC_run	5 hours ago	11.9min	C:\Users\...	sklearn, 1 more	0.919656786...	-	-	-	
<input type="checkbox"/>	<input type="checkbox"/>	Knn_run	5 hours ago	7.1min	C:\Users\...	Flipkart r.../4, 1 more	0.907644305...	-	-	-	

230 matching runs

mlflow2.11.3ExperimentsModels

Flipkart_sentiment_analysis >Random_run

Register model

OverviewModel metricsSystem metricsArtifacts

No description

Details

Created at	2024-03-31 17:19:34
Created by	Mahes
Status	Finished
Run ID	a0c7042b05994050ab7e328614cabe98
Duration	6.4min
Datasets used	—
Tags	estimator_class: sklearn.model_selection_search.G... estimator_name: GridSearchCV
Source	C:\Users\Mahes\anaconda3\Lib\site-packages\ipykernel_launcher.py
Logged models	sklearn +1
Registered models	—

The "Random_run" has no description, and there are no logged or registered models associated with it. However, the run has a "sklearn" tag with a value of "+1", indicating that it used the scikit-learn library for machine learning

Flipkart_sentiment_analysis >Random_run

Register model

OverviewModel metricsSystem metricsArtifacts

Parameters (12)

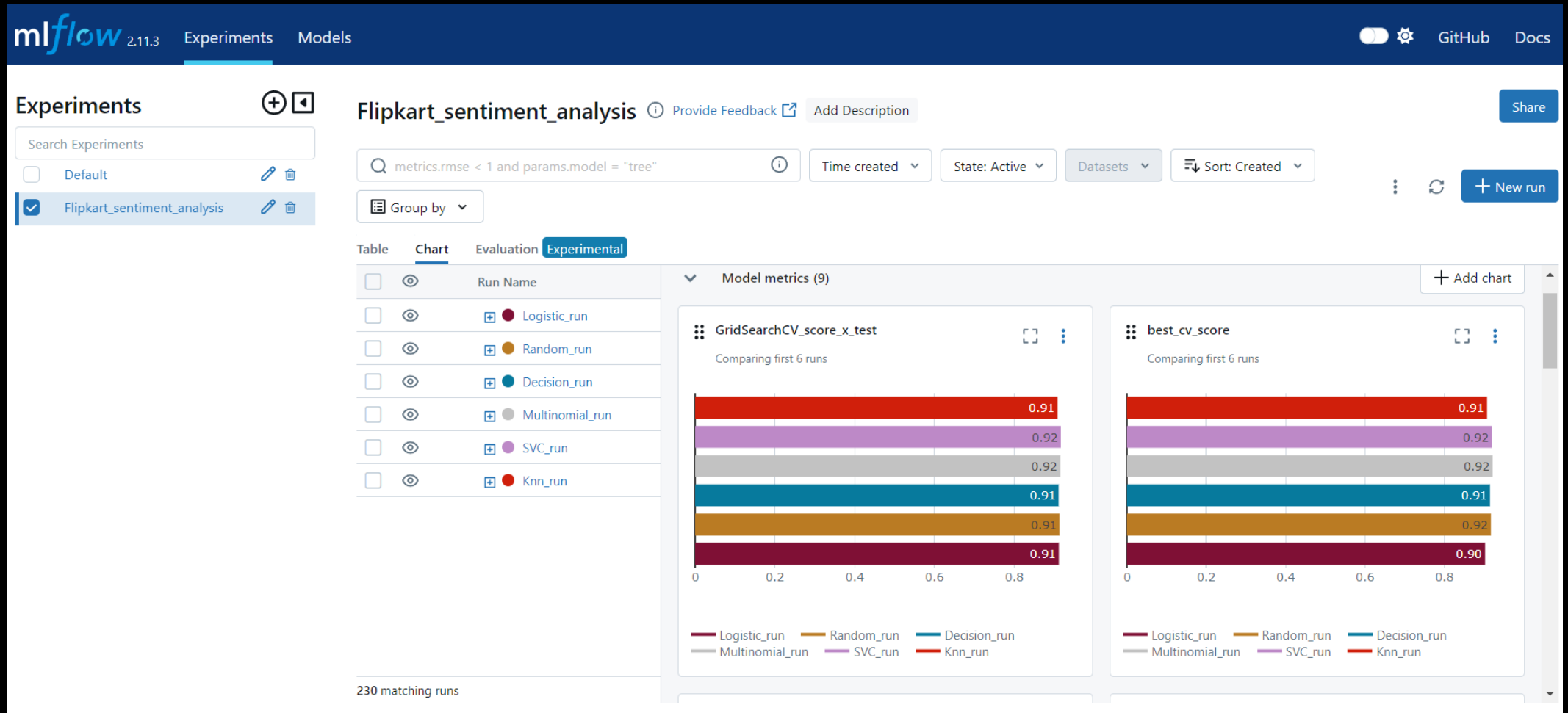
Parameter	Value
best_classifier__n_estimators	100
best_vectorization	TfidfVectorizer()
cv	5
error_score	nan
estimator	Pipeline(memory=Memory(location=.cache\joblib), steps=[('vectorization', CountVectorizer()), ('classifier', RandomForestClassifier())])
n_jobs	None
param_grid	[{'vectorization': [CountVectorizer(), TfidfVectorizer()], 'classifier__n_estimators': [50, 100, 200]}]
pre_dispatch	2*n_jobs
refit	True
return_train_score	True
scoring	accuracy
verbose	1

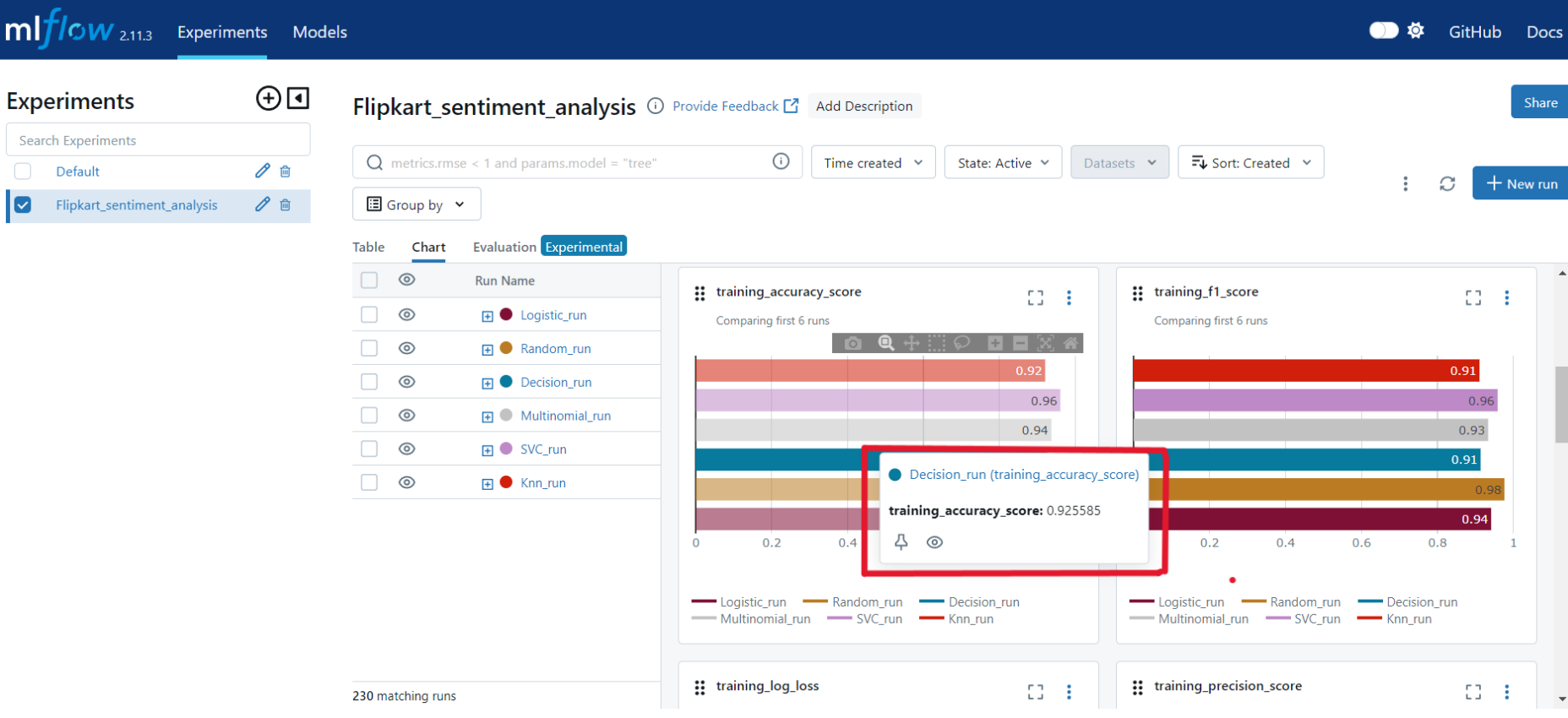
Metrics (9)

Metric	Value
best_cv_score	0.9173166926677068
GridSearchCV_score_x_test	0.9139114160948222
training_accuracy_score	0.9767550702028082
training_f1_score	0.9760709932385104
training_log_loss	0.10305937195879623
training_precision_score	0.9765241516001645
training_recall_score	0.9767550702028082
training_roc_auc	0.9735331835524605
training_score	0.9767550702028082

Model Metrics

- The table view shows the run name, model metrics, and a chart view. The user can filter the runs based on various parameters and add charts to visualize the performance of the models.
- Overall, MLflow provides a user-friendly interface to manage and track machine learning experiments, making it easy to compare and evaluate the performance of various models.

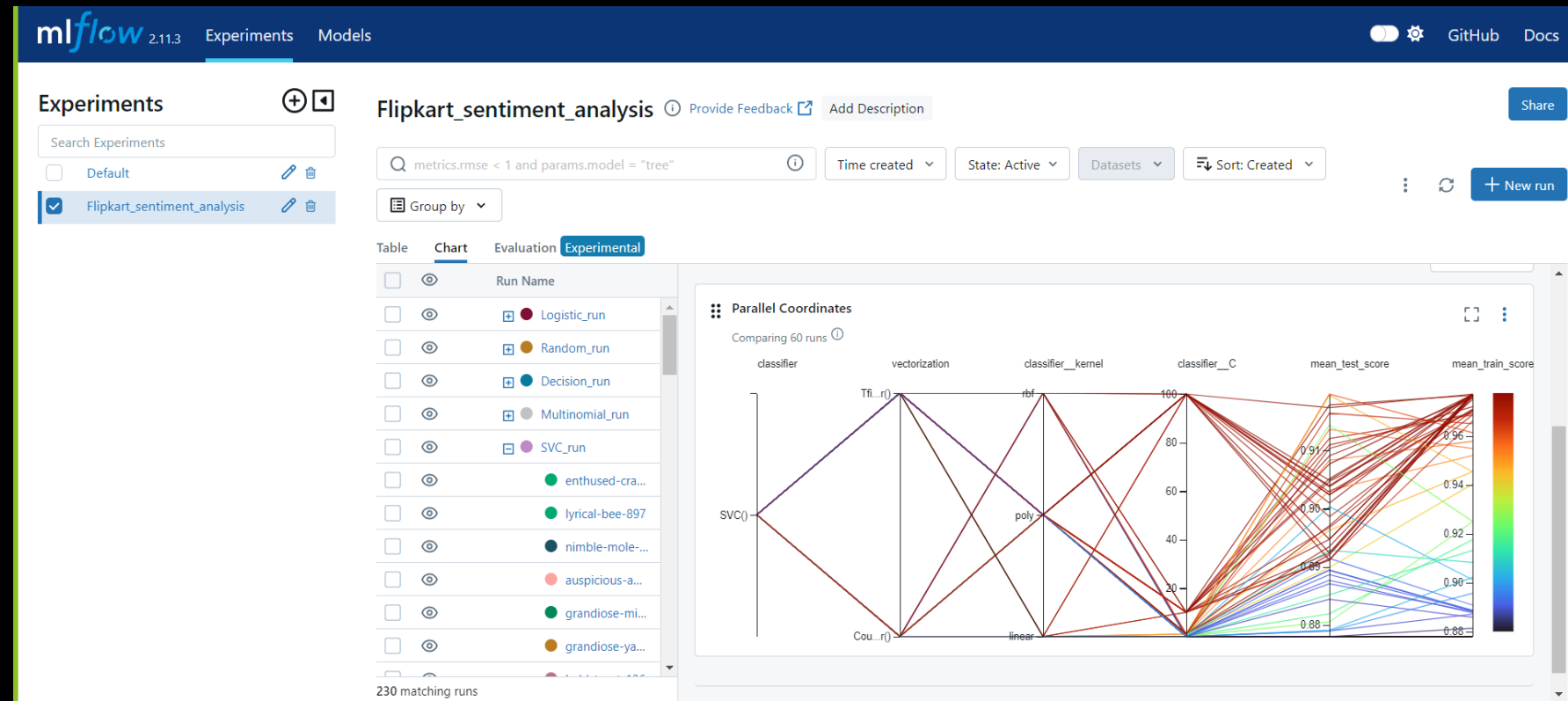


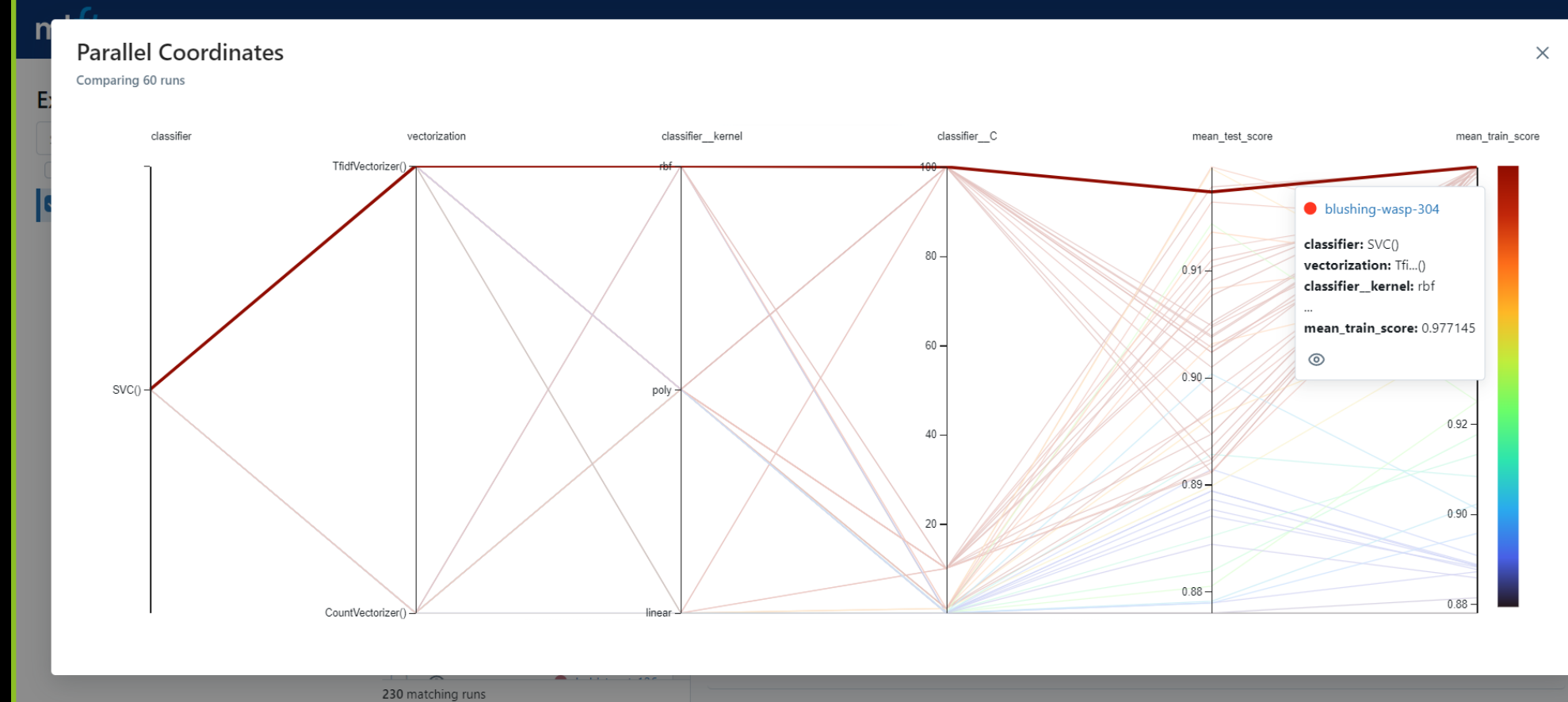


- The table shows the model metrics such as training_accuracy_score, training_f1_score. The user can view the values of these metrics for each run by clicking on the run name.
- For example, clicking on the run name "Logistic_run" shows the metrics values for the Logistic Regression model, such as training_accuracy_score: 0.925585.

Hyperparameters Plots

- The table can be grouped by various parameters, such as classifier, vectorization, and classifier kernel. The user can also sort the table based on various metrics, such as mean_train_score and mean_test_score.
- It shows the run name, time created, state, and datasets used. The user can click on the run name to view more details about the run.



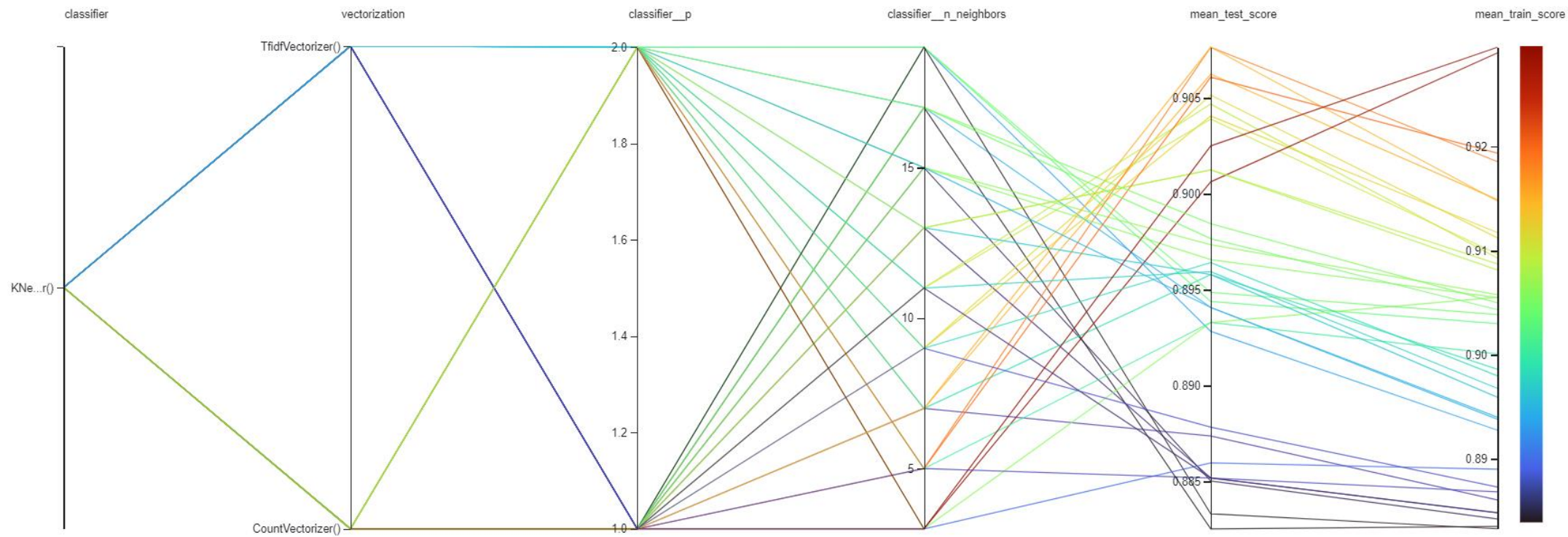


Each horizontal line represents a single run, and the intersection with the vertical axis shows the value of the parameter or metric for that run.

The top horizontal line shows that the classifier used in that run is SVC(), the vectorization is TfidfVectorizer(), and the kernel is rbf. The mean train score is 0.977145, and the mean test score is 0.91.

Parallel Coordinates

Comparing 36 runs

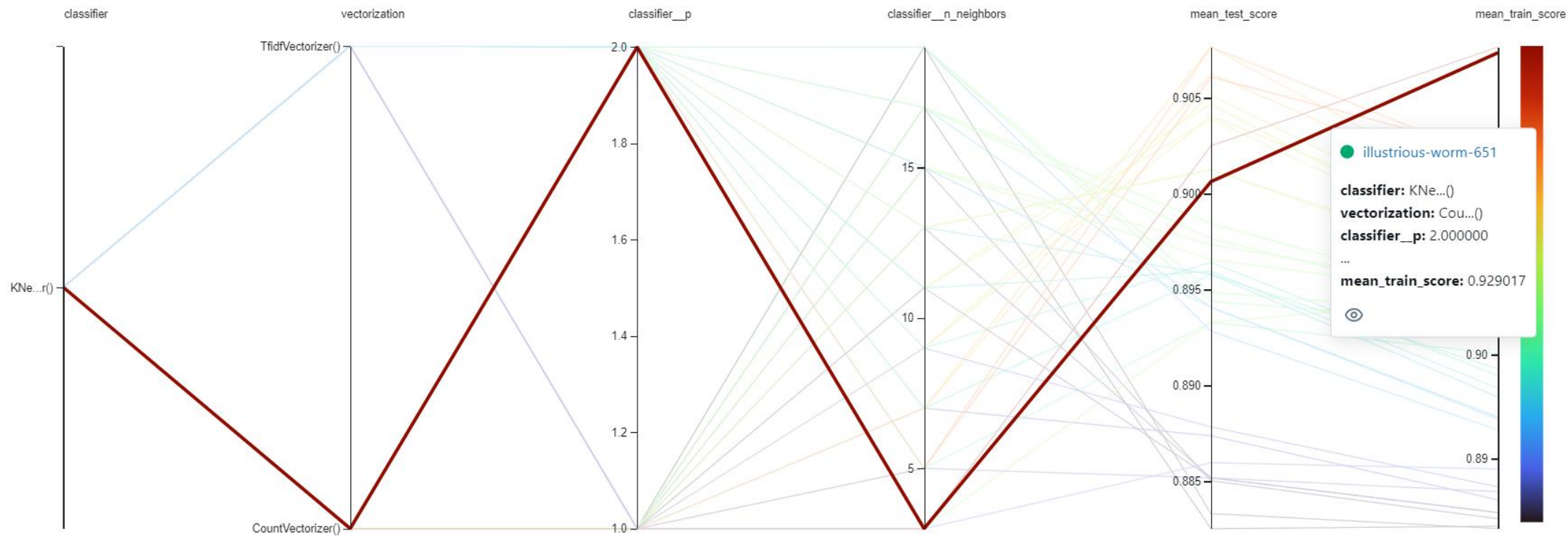


230 matching runs

+ Add section

Parallel Coordinates

Comparing 36 runs



230 matching runs

+ Add section

Registering Models

mlflow 2.11.3 Experiments Models

Experiments

Search Experiments

☐ Default

☒ Flipkart_sentiment_analysis

Flipkart_sentiment_analysis [Provide Feedback](#) [Add Description](#) [Share](#)

metrics.rmse < 1 and params.model = "tree" Time created State: Active Datasets Sort: Created

Columns Group by

Table Chart Evaluation Experimental

		Run Name	Created	Duration	Source	Models	best_cv_score	mean_score_tim	mean_test_score	clas
<input type="checkbox"/>	<input type="checkbox"/>	Logistic_run	4 hours ago	13.5min	C:\Users\...	sklearn, 1 more	0.902652106...	-	-	-
<input type="checkbox"/>	<input type="checkbox"/>	Random_run	4 hours ago	6.4min	C:\Users\...	Flipkart r.../3, 1 more	0.917316692...	-	-	-
<input type="checkbox"/>	<input type="checkbox"/>	Decision_run	4 hours ago	1.7min	C:\Users\...	Flipkart r.../2, 1 more	0.914976599...	-	-	-
<input type="checkbox"/>	<input type="checkbox"/>	Multinomial_run	4 hours ago	54.2s	C:\Users\...	Flipkart r.../1, 1 more	0.921684867...	-	-	-
<input type="checkbox"/>	<input type="checkbox"/>	SVC_run	5 hours ago	11.9min	C:\Users\...	sklearn, 1 more	0.919656786...	-	-	-
<input type="checkbox"/>	<input type="checkbox"/>	Knn_run	5 hours ago	7.1min	C:\Users\...	Flipkart r.../4, 1 more	0.907644305...	-	-	-

230 matching runs

The table below the Versions section provides information about each version of the model, including the version number, the date and time it was registered, the user who registered it, and its current stage.

Overall, the context shows the status and version history of the "Flipkart review production" experiment in the MLflow Models UI, including the current active version and the previous staging and archived versions.

mlflow

2.11.3

ExperimentsModels

Registered Models >

Flipkart review production

Created Time: 2024-03-31 21:34:47Last Modified: 2024-03-31 21:45:39

> Description Edit

> Tags

<> VersionsAllActive 2Compare

New model registry UI

Version	Registered at ↕	Created by	Stage	Description
✓ Version 4	2024-03-31 21:39:27		Staging	
✓ Version 3	2024-03-31 21:38:26		Archived	
✓ Version 2	2024-03-31 21:37:08		Archived	
✓ Version 1	2024-03-31 21:34:47		Production	

1

Prefect

The Prefect is also available as an open source, locally hosted orchestration engine, API server, and UI, giving you insight into the flows running with any local Prefect server instance.

Prefect installation

`pip install prefect`

`prefect server start`

```
-----  
|_ \  _ \  _ \| | | | | | /  _ \| | | | | | |
|_ /  _ /  _ \| | | | | | \  _ \| | | |  
|_| | | \_ _ \| | | | | | \_ _ \| | | |  
|_| | | \_ _ \| | | | | | \_ _ \| | | |
```

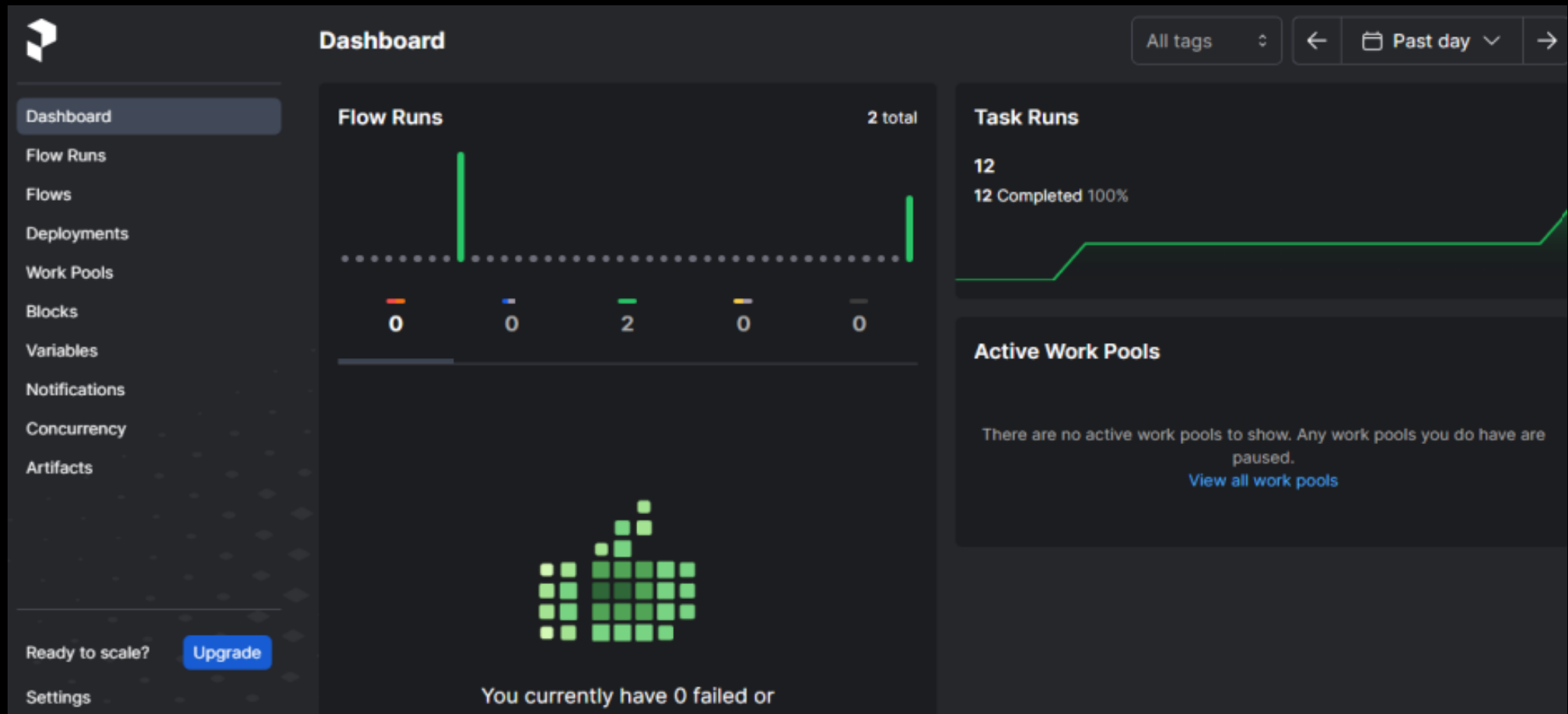
Configure Prefect to communicate with the server with:

```
prefect config set PREFECT_API_URL=http://127.0.0.1:4200/api
```

View the API reference documentation at <http://127.0.0.1:4200/docs>

Check out the dashboard at <http://127.0.0.1:4200>

Prefect dashboard





Dashboard

All tags ▾

←

📅 Past day ▾

→

- Dashboard
 - Flow Runs
 - Flows
 - Deployments
 - Work Pools
 - Blocks
 - Variables
 - Notifications
 - Concurrency
 - Artifacts
- Ready to scale? Upgrade

Flow Runs

2 total

0 0 2 0 0

Naive Bayes Flow

1h 26m ago

2 ^

Naive Bayes Flow > knowing-manul

Completed 2024/03/28 02:09:10 PM 5s 6 Task runs

Naive Bayes Flow > violet-ara

Completed 2024/03/28 08:36:41 AM 2s 6 Task runs

Task Runs

12

12 Completed 100%

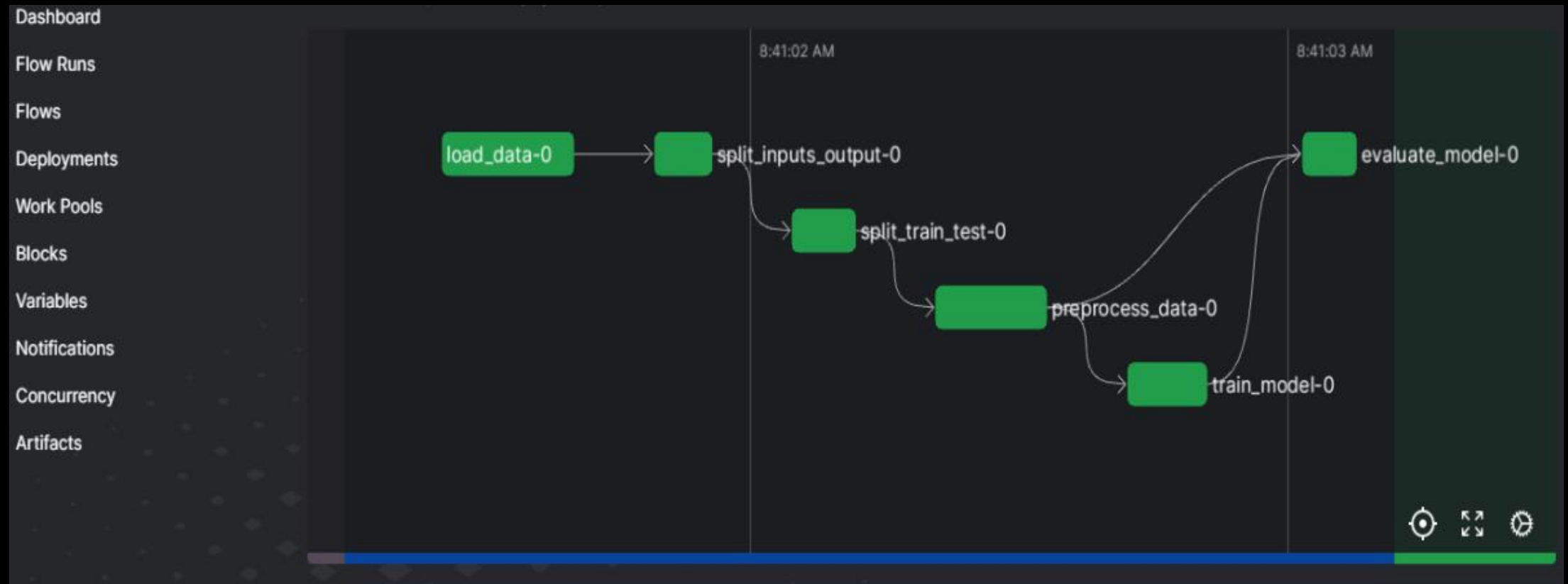
Active Work Pools

There are no active work pools to show. Any work pools you do have are paused.

View all work pools

Settings

Prefect Workflow



THANK YOU

