



INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Using **MLflow** for Experiment Tracking and Model
Management-
Sentiment Analysis of Flipkart Reviews

Prepared by Eleshala Pravalika

Objective of the Report:

- The objective is to introduce MLflow for experiment tracking, model management, and reproducibility in **Sentiment Analysis of Flipkart reviews**.

MLFlow:

MLFlow: Unified Platform for Experiment Tracking and Model Registry

MLflow is an open-source platform for managing the end-to-end machine learning lifecycle. It provides a suite of tools and components designed to streamline the development, experimentation, productionisation, and collaboration aspects of machine learning projects. MLflow is widely used by data scientists, machine learning engineers, and researchers to track experiments, package and share code, and deploy models at scale.

Key Features:

1. Experiment Tracking
2. Model Registry

Integration of MLflow into projects:

```
pip install mlflow  
mlflow ui
```

```
import mlflow  
mlflow.set_experiment("Sentiment_Analysis_Flipkart_Reviews")
```

```
mlflow.sklearn.autolog(max_tuning_runs=None)
```

```
with mlflow.start_run() as run:  
    %time grid_search.fit(X_train, y_train)
```

MLflow dashboard:

mlflow 2.11.3 Experiments Models

Search Experiments

☐ Default

☒ Sentiment_Analysis_Fli...

Sentiment_Analysis_Flipkart_Reviews

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

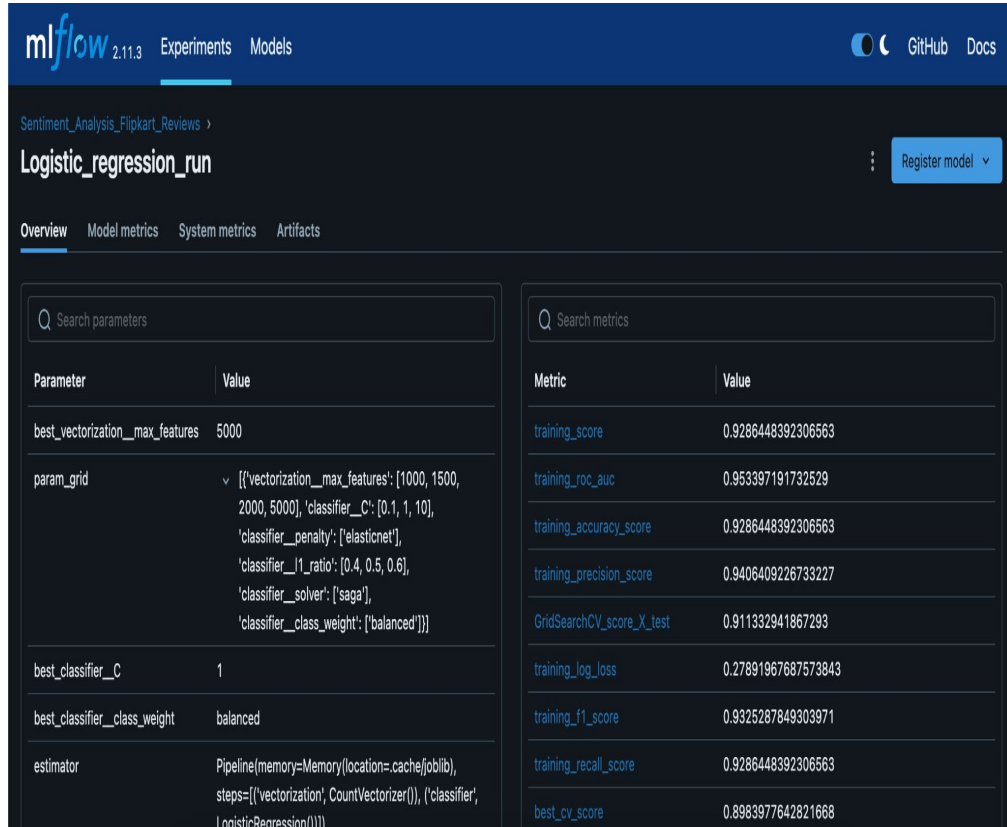
Time created State: Active

Datasets Sort: Created Columns Group by

Table Chart Evaluation Experimental

		Run Name	Created	Duration	Source	Models	Metrics
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> Logistic_regression...	4 hours ago	11.3min	ipykerne...	Sentiment_.../1, 1 more	0.9113329...
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> Random_Forest_run	4 hours ago	5.8min	ipykerne...	Sentiment_.../2, 1 more	0.9183793...
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> DecisionTree_Run	4 hours ago	9.5s	ipykerne...	Sentiment_.../3, 1 more	0.9148561...
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> naivebayes_run	4 hours ago	6.4s	ipykerne...	Sentiment_.../4, 1 more	0.9242513...

Demonstration of logging parameters, metrics, and artifacts using MLflow tracking APIs:



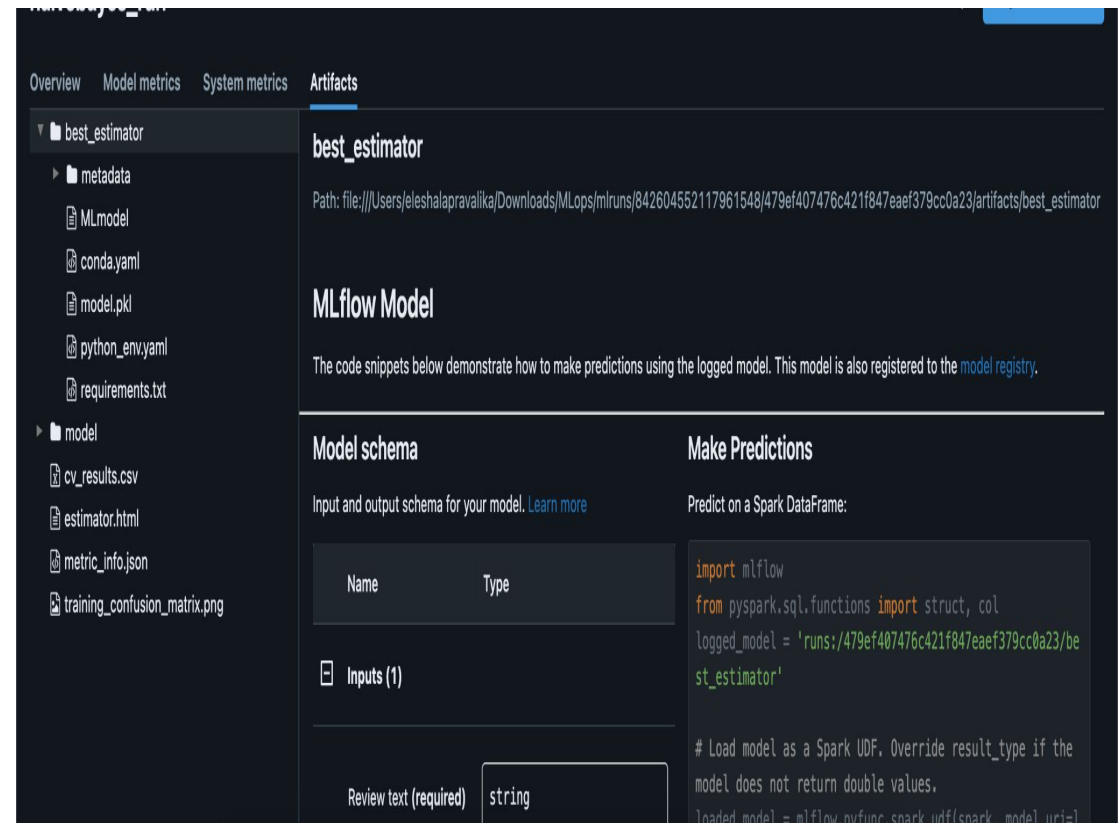
The screenshot shows the MLflow web interface for a run named 'Logistic_regression_run'. The 'Overview' tab is active, displaying a table of parameters and a table of metrics.

Parameters:

Parameter	Value
best_vectorization__max_features	5000
param_grid	[{'vectorization__max_features': [1000, 1500, 2000, 5000], 'classifier__C': [0.1, 1, 10], 'classifier__penalty': ['elasticnet'], 'classifier__l1_ratio': [0.4, 0.5, 0.6], 'classifier__solver': ['saga'], 'classifier__class_weight': ['balanced']}]
best_classifier__C	1
best_classifier__class_weight	balanced
estimator	Pipeline(memory=Memory(location=cache/joblib), steps=[('vectorization', CountVectorizer()), ('classifier', LogisticRegression())])

Metrics:

Metric	Value
training_score	0.9286448392306563
training_roc_auc	0.953397191732529
training_accuracy_score	0.9286448392306563
training_precision_score	0.9406409226733227
GridSearchCV_score_X_test	0.911332941867293
training_log_loss	0.27891967687573843
training_f1_score	0.9325287849303971
training_recall_score	0.9286448392306563
best_cv_score	0.8983977642821668



The screenshot shows the MLflow web interface for a run named 'best_estimator'. The 'Artifacts' tab is active, displaying a list of artifacts and a section for the MLflow Model.

Artifacts:

- best_estimator
 - metadata
 - MLmodel
 - conda.yaml
 - model.pkl
 - python_env.yaml
 - requirements.txt
 - model
 - cv_results.csv
 - estimator.html
 - metric_info.json
 - training_confusion_matrix.png

MLflow Model

The code snippets below demonstrate how to make predictions using the logged model. This model is also registered to the [model registry](#).

Model schema

Input and output schema for your model. [Learn more](#)

Name	Type
Inputs (1)	
Review text (required)	string

Make Predictions

Predict on a Spark DataFrame:

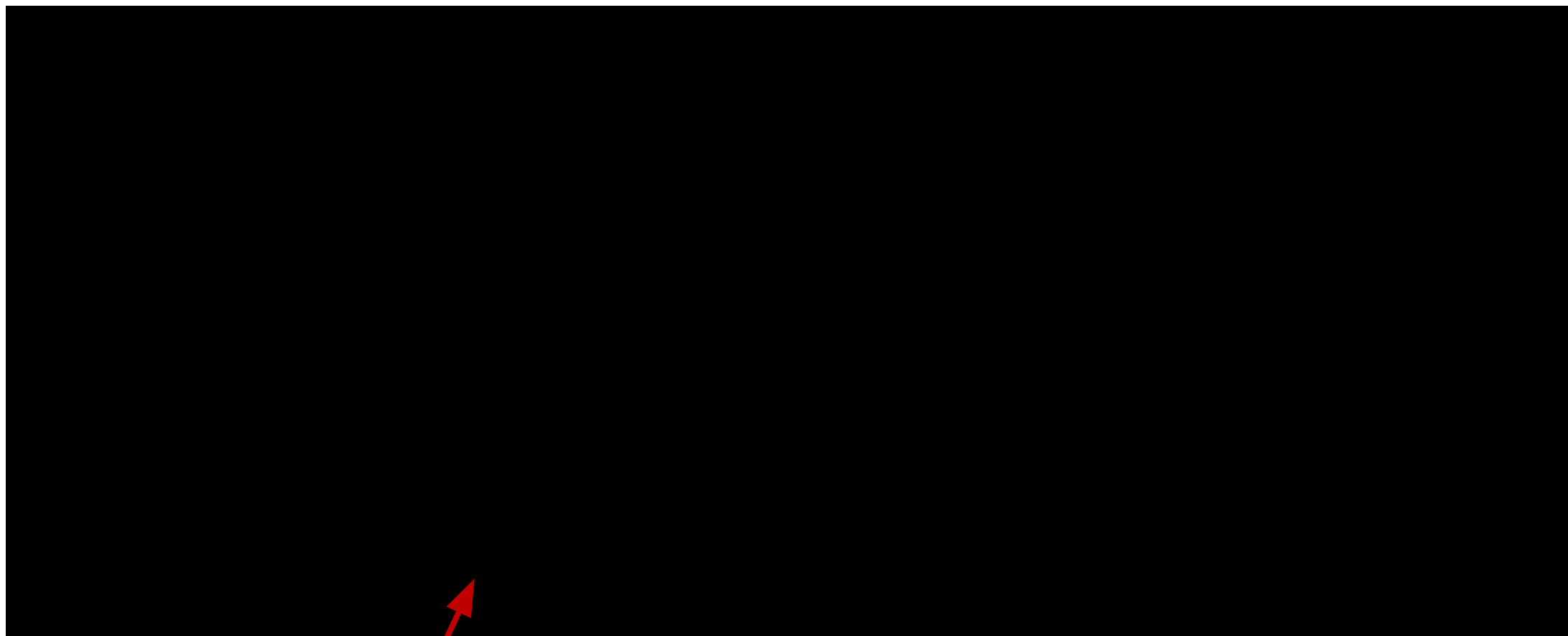
```
import mlflow
from pyspark.sql.functions import struct, col
logged_model = 'runs:/479ef407476c421f847eaf379cc0a23/be
st_estimator'

# Load model as a Spark UDF. Override result_type if the
model does not return double values.
loaded_model = mlflow.pyfunc.spark_udf(spark, model_uri=
```

Demonstration of logging parameters, metrics, and artifacts using MLflow tracking APIs:

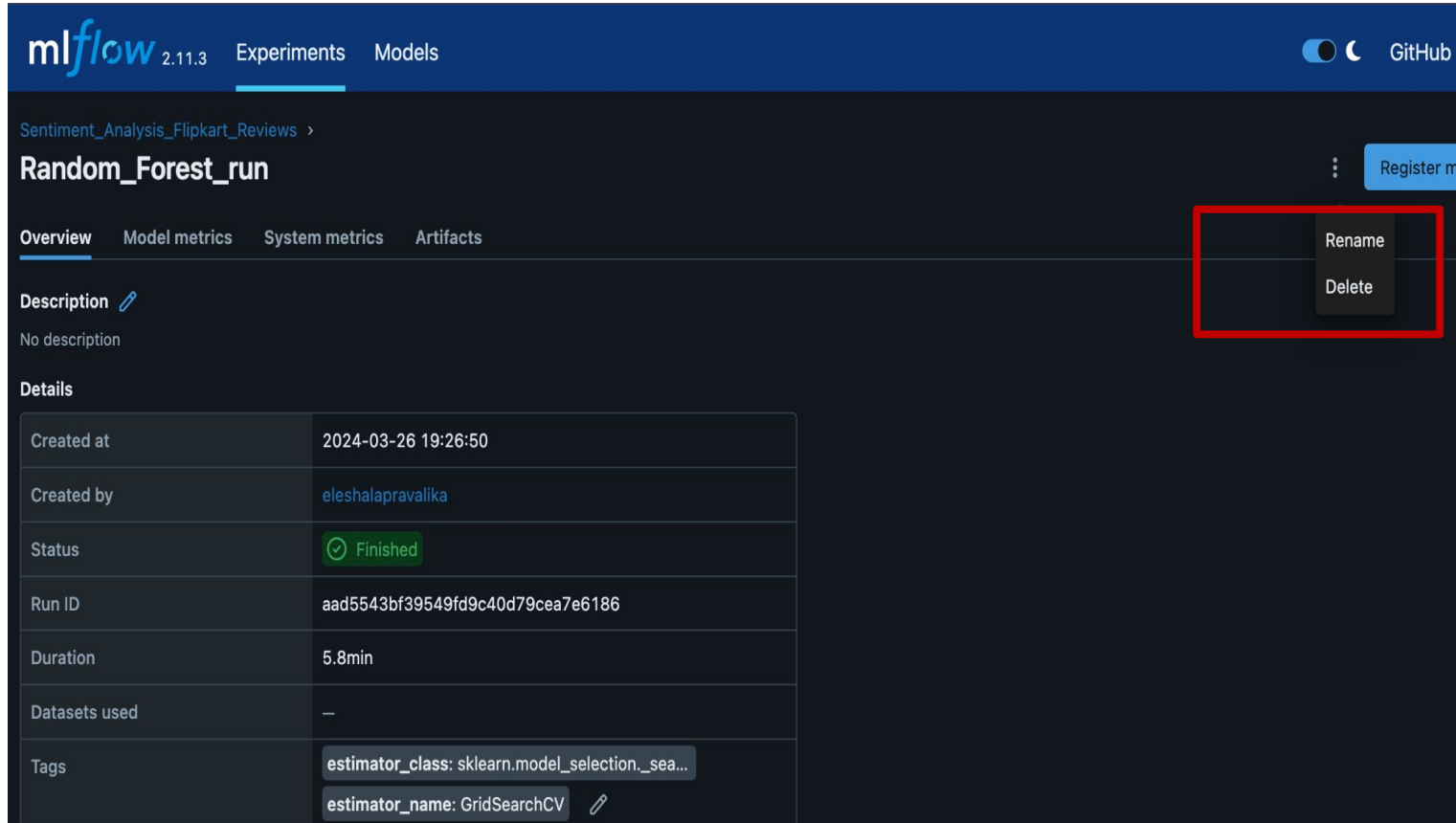
*Click on Experiment_name → Run name → (scroll down for) →
Parameters , Metrics, Artifacts.*

Demonstration of logging parameters, metrics, and artifacts using MLflow tracking APIs:



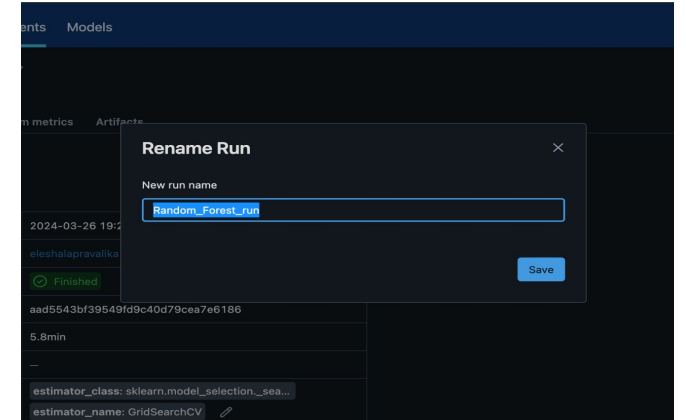
[Click on the attachment for video](#)

Customizing Mlflow UI with run names:

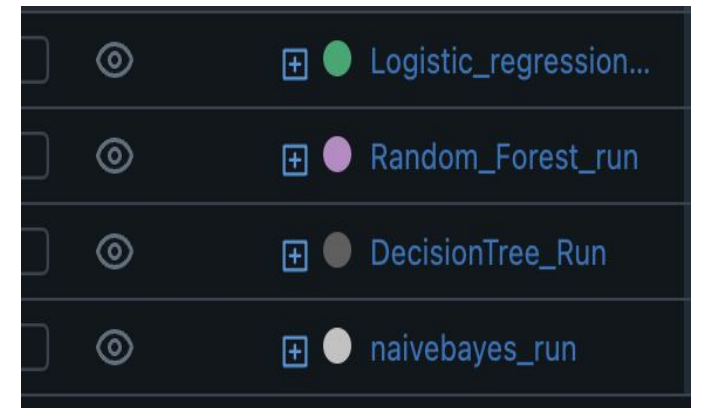


The screenshot shows the Mlflow UI interface for an experiment named 'Random_Forest_run'. The top navigation bar includes the Mlflow logo (version 2.11.3), 'Experiments', and 'Models' tabs. A 'Register model' button is visible in the top right. The main content area has tabs for 'Overview', 'Model metrics', 'System metrics', and 'Artifacts'. The 'Overview' tab is active, displaying a description (No description) and a details table.

Details	
Created at	2024-03-26 19:26:50
Created by	eleshalapavalika
Status	Finished
Run ID	aad5543bf39549fd9c40d79cea7e6186
Duration	5.8min
Datasets used	—
Tags	<code>estimator_class: sklearn.model_selection_sea...</code> <code>estimator_name: GridSearchCV</code>



The screenshot shows a 'Rename Run' dialog box. It has a title bar with a close button. The main content area has a label 'New run name' followed by a text input field containing 'Random_Forest_run'. A 'Save' button is located at the bottom right of the dialog.

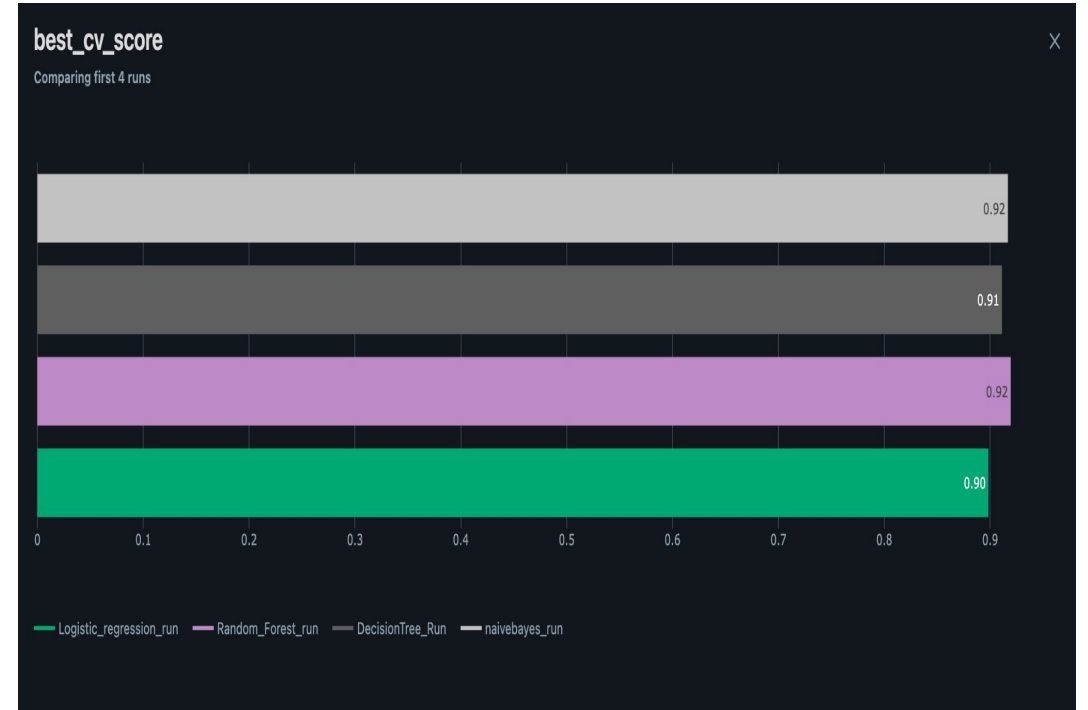
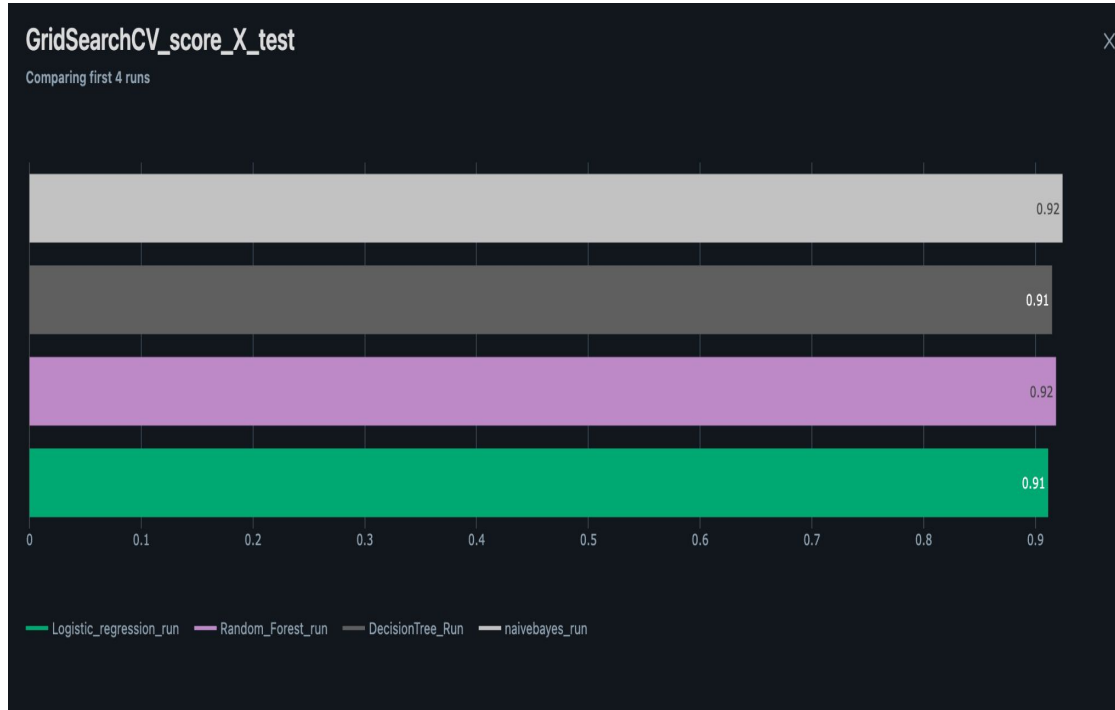


The screenshot shows a list of runs in the Mlflow UI. Each run entry consists of a status icon (a square with a circle inside), a colored circle, and the run name. The runs listed are: 'Logistic_regression...' (green circle), 'Random_Forest_run' (purple circle), 'DecisionTree_Run' (grey circle), and 'naivebayes_run' (grey circle).

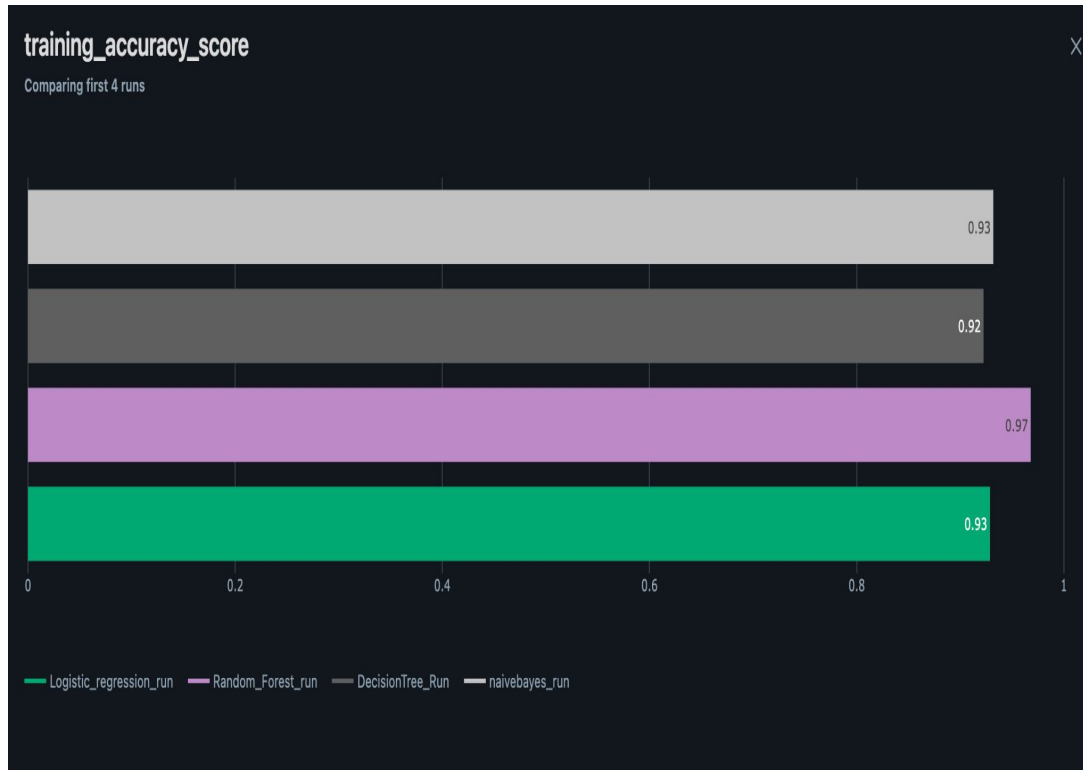
Customizing Mlflow UI with run names:

Click on Run name → Right corner : 3 dots → Rename → Give the name for run → Save.

Metric Plots :



Metric Plots :



HyperParameter Plots Creation:

Click on charts → Add Section → Name hyperparameter with model → Add Parallel coordinates chart → give parameters and metrics of particular run → click on the runname's "+" : This will pop the interactive chart

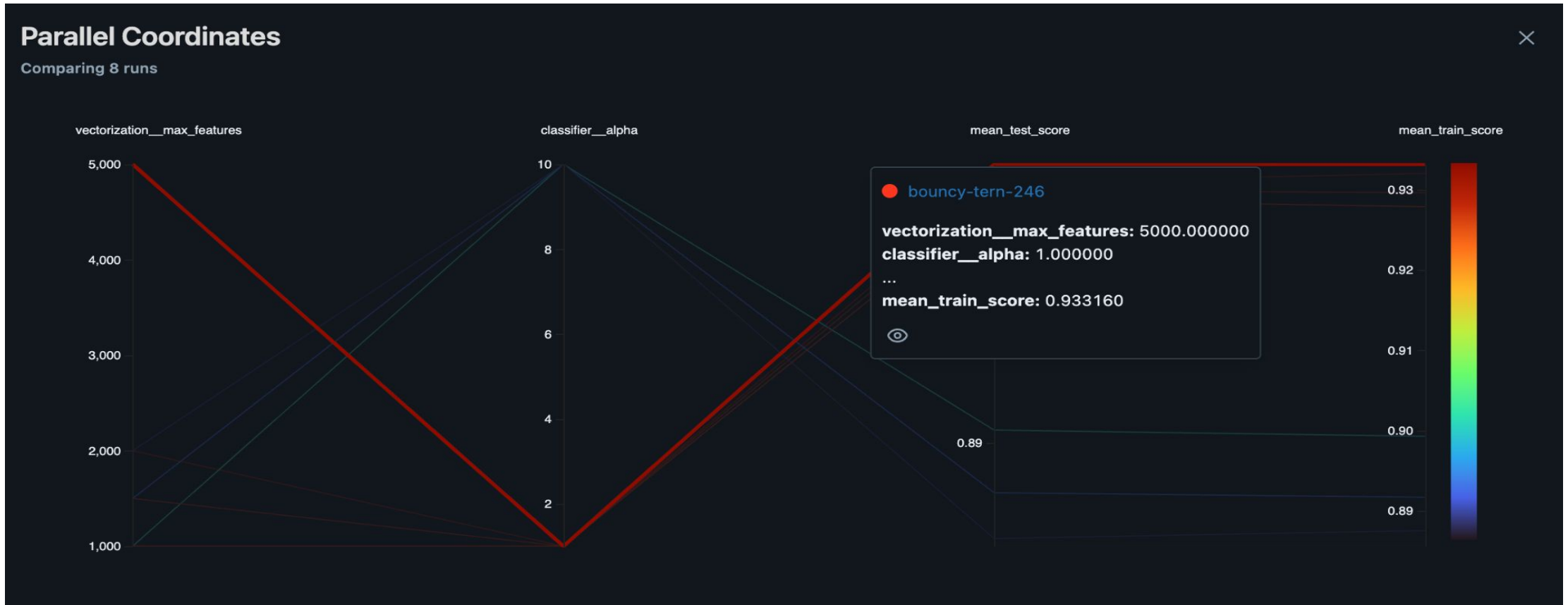
HyperParameter Plots :



HyperParameter Plots :



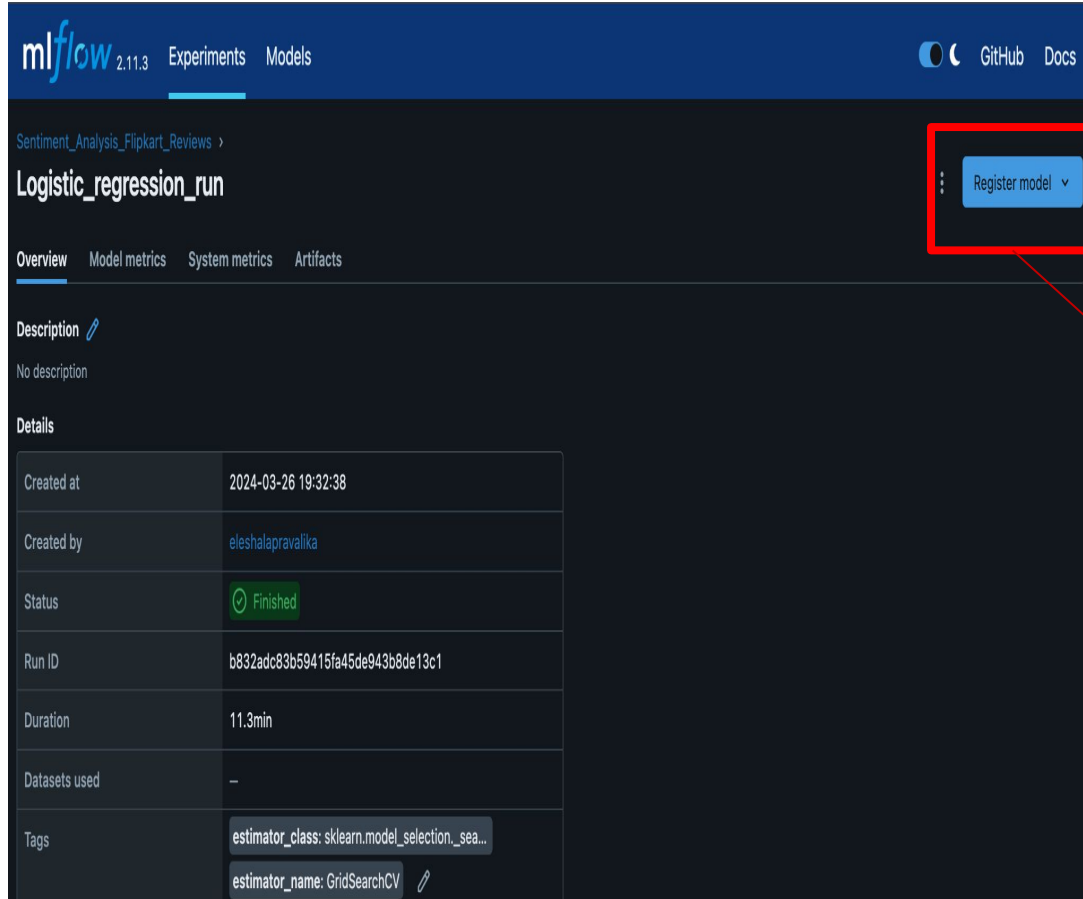
HyperParameter Plots :



HyperParameter Plots :



Registering models and Managing with tags:



mlflow 2.11.3 Experiments Models

Sentiment_Analysis_Flipkart_Reviews >

Logistic_regression_run

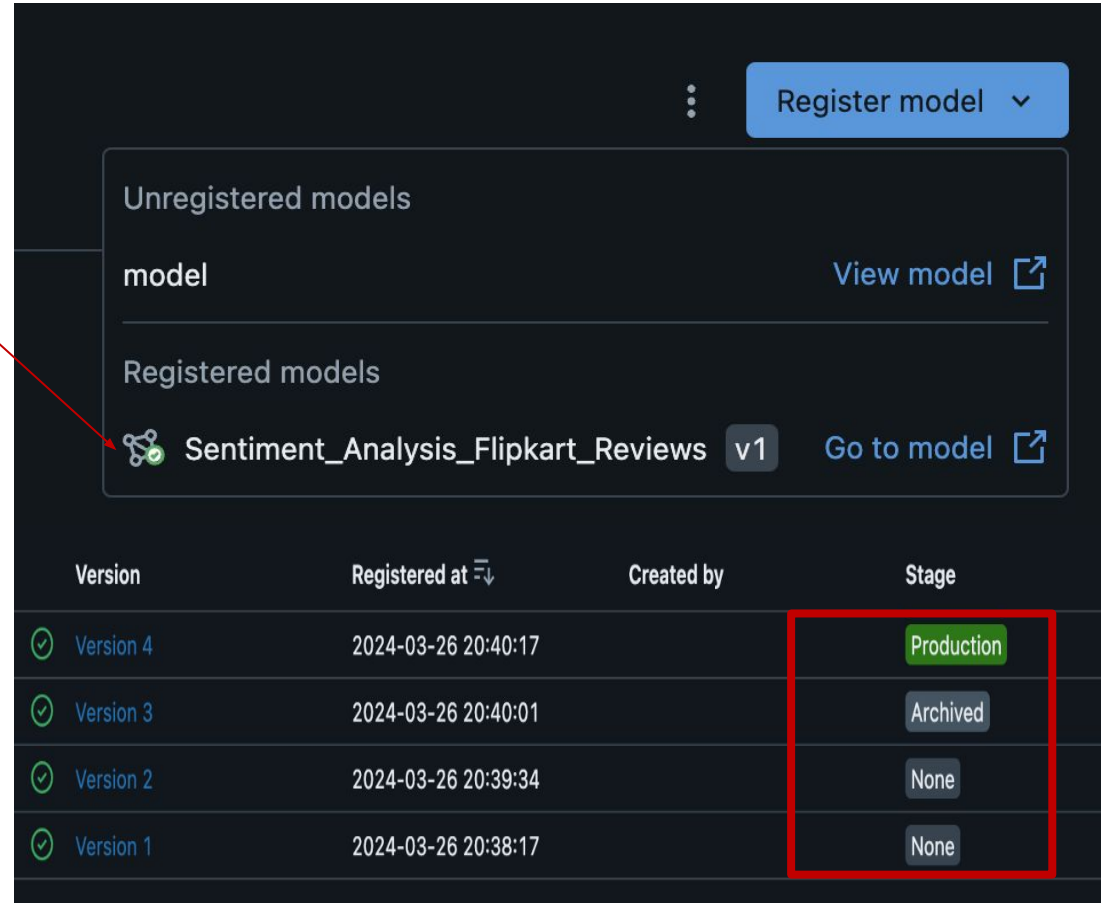
Overview Model metrics System metrics Artifacts

Description

No description

Details

Created at	2024-03-26 19:32:38
Created by	eleshalapavalika
Status	Finished
Run ID	b832adc83b59415fa45de943b8de13c1
Duration	11.3min
Datasets used	—
Tags	<code>estimator_class: sklearn.model_selection_sea...</code> <code>estimator_name: GridSearchCV</code>



Register model

Unregistered models

model View model

Registered models

Sentiment_Analysis_Flipkart_Reviews v1 Go to model

Version	Registered at	Created by	Stage
Version 4	2024-03-26 20:40:17		Production
Version 3	2024-03-26 20:40:01		Archived
Version 2	2024-03-26 20:39:34		None
Version 1	2024-03-26 20:38:17		None

MLflow Experiment Tracking and Model Management

mlflow 2.11.3 Experiments Models

Search Experiments

- ☒ Default
- ☐ Sentiment_Analysis_Fli...

Default [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	Dataset	Duration	Source
No runs logged				

No runs have been logged yet. [Learn more](#) about how to create ML model training runs in this experiment.

0 matching runs

Click on the attachment for video

THANK
YOU

