



INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Sentimental Analysis Experimentation using MLFlow Integration

Objective of the project

The key aims of this project include:

1. Integrate MLflow into machine learning project.
2. Customizing MLflow UI with run names.
3. Demonstration of log parameters, metrics, and artifacts using MLflow tracking APIs.
4. Demonstration of metric plots and hyperparameter plots.
5. Demonstration on registering models and managing them by tagging.

Introduction

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, productionization, 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

Integrating MLFlow

Run below commands to install mlflow on your system:

1. Create Virtual Environment
python -m venv .env
2. Activate Virtual Environment
.env\Scripts\activate
3. Install MLFlow
pip install mlflow

4. Install Suptools (Optional)
pip install setuptools

5. To open MLflow Dashboard
mlflow ui

The screenshot shows the MLflow Experiments dashboard. The top navigation bar includes 'mlflow 2.11.3', 'Experiments', and 'Models'. The left sidebar shows a search bar and a list of experiments, with 'sentimental_analysis_prediction_experimentation' selected. The main content area displays the details for this experiment, including a search bar, filters (Time created, State: Active, Datasets), and a table of runs. The table has columns for Run Name, Created, Dataset, Duration, Source, and Models. The runs listed are: bald-wren-566, defiant-quail-332, whimsical-duck-796, traveling-shoat-28, gifted-goat-280, and luminous-duck-761. At the bottom, it indicates '230 matching runs'.

Run Name	Created	Dataset	Duration	Source	Models
bald-wren-566	7 minutes ago	-	6.4min	C:\Users\...	sklearn, 1 more
defiant-quail-332	18 minutes ago	-	11.2min	C:\Users\...	sklearn, 1 more
whimsical-duck-796	24 minutes ago	-	5.7min	C:\Users\...	sklearn, 1 more
traveling-shoat-28	39 minutes ago	-	14.9min	C:\Users\...	sklearn, 1 more
gifted-goat-280	40 minutes ago	-	1.2min	C:\Users\...	sklearn, 1 more
luminous-duck-761	40 minutes ago	-	19.0s	C:\Users\...	sklearn, 1 more

Customizing MLflow

1. Click on Run Name “bald-wren-566”

The screenshot shows the MLflow Experiments dashboard with the 'bald-wren-566' run selected. The left sidebar shows the search bar and the list of experiments. The main content area displays the details for this run, including a search bar, filters (Time created, State: Active, Datasets), and a table of runs. The table has columns for Run Name, Created, Dataset, Duration, Source, and Models. The runs listed are: bald-wren-566, defiant-quail-332, whimsical-duck-796, traveling-shoat-28, gifted-goat-280, and luminous-duck-761. At the bottom, it indicates '230 matching runs'.

Run Name	Created	Dataset	Duration	Source	Models
bald-wren-566	7 minutes ago	-	6.4min	C:\Users\...	sklearn, 1 more
defiant-quail-332	18 minutes ago	-	11.2min	C:\Users\...	sklearn, 1 more
whimsical-duck-796	24 minutes ago	-	5.7min	C:\Users\...	sklearn, 1 more
traveling-shoat-28	39 minutes ago	-	14.9min	C:\Users\...	sklearn, 1 more
gifted-goat-280	40 minutes ago	-	1.2min	C:\Users\...	sklearn, 1 more
luminous-duck-761	40 minutes ago	-	19.0s	C:\Users\...	sklearn, 1 more

2. Click on 3 dots → Select Rename

The screenshot shows the MLflow Experiments dashboard with the 'bald-wren-566' run selected. The left sidebar shows the search bar and the list of experiments. The main content area displays the details for this run, including a search bar, filters (Time created, State: Active, Datasets), and a table of runs. The table has columns for Run Name, Created, Dataset, Duration, Source, and Models. The runs listed are: bald-wren-566, defiant-quail-332, whimsical-duck-796, traveling-shoat-28, gifted-goat-280, and luminous-duck-761. At the bottom, it indicates '230 matching runs'.

Run Name	Created	Dataset	Duration	Source	Models
bald-wren-566	7 minutes ago	-	6.4min	C:\Users\...	sklearn, 1 more
defiant-quail-332	18 minutes ago	-	11.2min	C:\Users\...	sklearn, 1 more
whimsical-duck-796	24 minutes ago	-	5.7min	C:\Users\...	sklearn, 1 more
traveling-shoat-28	39 minutes ago	-	14.9min	C:\Users\...	sklearn, 1 more
gifted-goat-280	40 minutes ago	-	1.2min	C:\Users\...	sklearn, 1 more
luminous-duck-761	40 minutes ago	-	19.0s	C:\Users\...	sklearn, 1 more

3. Enter new run name → click Save

The screenshot shows the mlflow 2.11.3 interface. A modal dialog titled 'Rename Run' is open, asking for a 'New run name'. The input field contains 'random_forest_run' and a 'Save' button is at the bottom right. In the background, the 'Details' tab of a run named 'bald-wren-566' is visible, showing metadata like 'Created at', 'Created by', 'Status', 'Run ID', 'Duration', 'Tags', 'Source', 'Logged models', and 'Registered models'.

4. Repeat step 1 to 3 for remaining runs

The screenshot shows the 'Experiments' page in mlflow. The experiment 'sentimental_analysis_prediction_experimentation' is selected. A table lists several runs with columns for Run Name, Created, Dataset, Duration, Source, and Models. The runs listed are 'random_forest_run', 'support_vector_mach...', 'k-nearest_neighbor_r...', 'logistic_regression_run', 'decision_tree_run', and 'naive_bayes_run'. At the bottom, it indicates '230 matching runs'.

Run Name	Created	Dataset	Duration	Source	Models
random_forest_run	24 minutes ago	-	6.4min	C:\Users\...	sklearn, 1 more
support_vector_mach...	35 minutes ago	-	11.2min	C:\Users\...	sklearn, 1 more
k-nearest_neighbor_r...	41 minutes ago	-	5.7min	C:\Users\...	sklearn, 1 more
logistic_regression_run	55 minutes ago	-	14.9min	C:\Users\...	sklearn, 1 more
decision_tree_run	57 minutes ago	-	1.2min	C:\Users\...	sklearn, 1 more
naive_bayes_run	57 minutes ago	-	19.0s	C:\Users\...	sklearn, 1 more

Log Parameters, Metrics and Artifacts

1. Click on Run name

This screenshot is identical to the one above, showing the 'Experiments' page with the 'sentimental_analysis_prediction_experimentation' experiment selected and a table of runs.

Run Name	Created	Dataset	Duration	Source	Models
random_forest_run	24 minutes ago	-	6.4min	C:\Users\...	sklearn, 1 more
support_vector_mach...	35 minutes ago	-	11.2min	C:\Users\...	sklearn, 1 more
k-nearest_neighbor_r...	41 minutes ago	-	5.7min	C:\Users\...	sklearn, 1 more
logistic_regression_run	55 minutes ago	-	14.9min	C:\Users\...	sklearn, 1 more
decision_tree_run	57 minutes ago	-	1.2min	C:\Users\...	sklearn, 1 more
naive_bayes_run	57 minutes ago	-	19.0s	C:\Users\...	sklearn, 1 more

2. Click on **Overview** Tab and Scroll down to View Log **Parameters** and **Metrics** of that run

sentimental_analysis_prediction_experimentation > random_forest_run

Overview Model metrics System metrics Artifacts

Parameters (12)

Parameter	Value
best_classifier_n_estimators	200
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	f1
verbose	1

Metrics (9)

Metric	Value
best_cv_score	0.9553744768582643
GridSearchCV_score_X_test_clean	0.964424514200299
training_accuracy_score	0.979741842954464
training_f1_score	0.9793721059023888
training_log_loss	0.10364627918608836
training_precision_score	0.9795428422561445
training_recall_score	0.979741842954464
training_roc_auc	0.9745213023968051
training_score	0.9883396966257352

3. Click on **Artifacts** Tab to view the artifacts which includes best_estimator, MLflow Model etc.

sentimental_analysis_prediction_experimentation > random_forest_run

Overview Model metrics System metrics Artifacts

best_estimator

Path: file:///C:/Users/hanna/Desktop/Anu/INMOMATICS/Assignments/08_Using%20MLflow%20for%20Experiment%20Tracking%20and%20Model%20Management/mlruns/19618087473214084

MLflow Model

The code snippets below demonstrate how to make predictions using the logged model. You can also register it to the model registry to version control

Model schema

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

Name	Type
Inputs (1)	
review_text (required)	string
Outputs (1)	
- (required)	Tensor (dtype: int64, shape: [-1])

Make Predictions

Predict on a Spark DataFrame:

```
import mlflow
from pyspark.sql.functions import struct, col
logged_model = 'runs:/adf6814d26ee4d538acc16db3dc87d/best_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=logged_model, result_type='double')

# Predict on a Spark DataFrame.
df.withColumn('predictions', loaded_model(struct(*map(col, df.columns))))
```

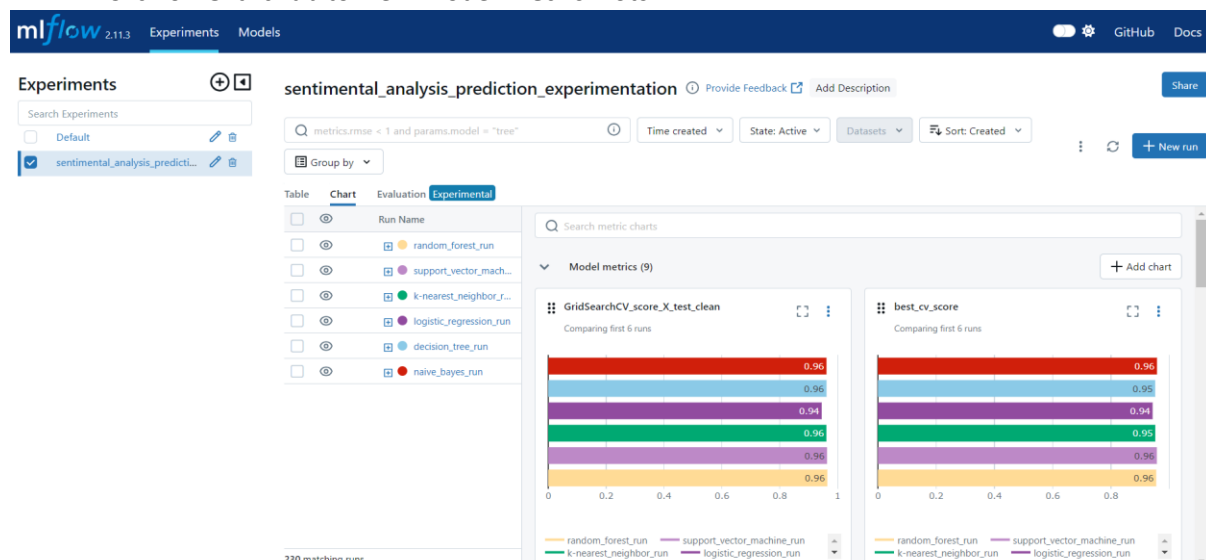
Predict on a Pandas DataFrame:

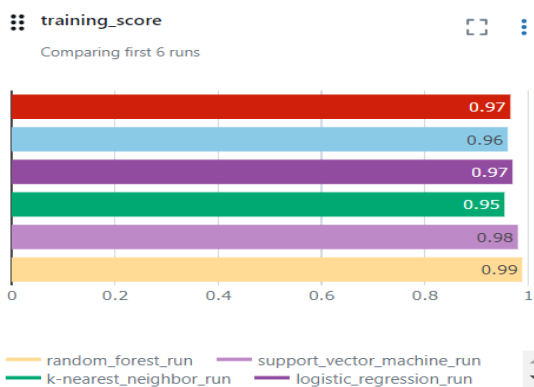
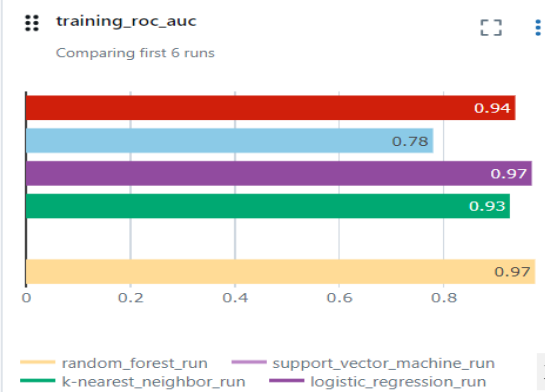
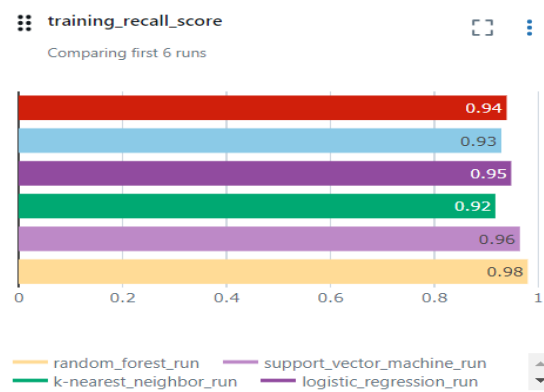
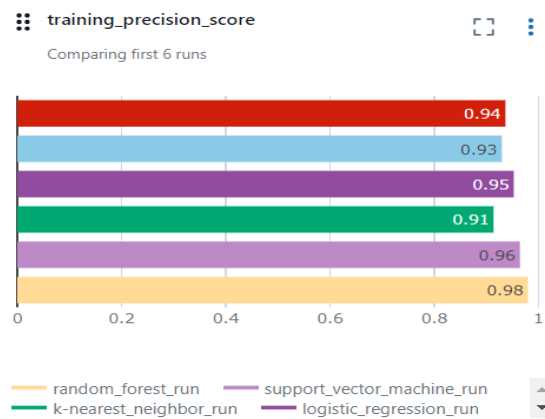
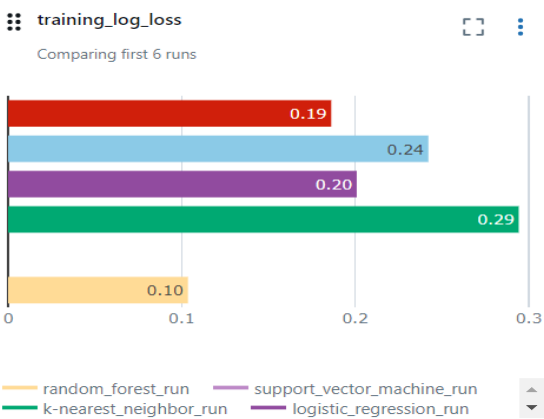
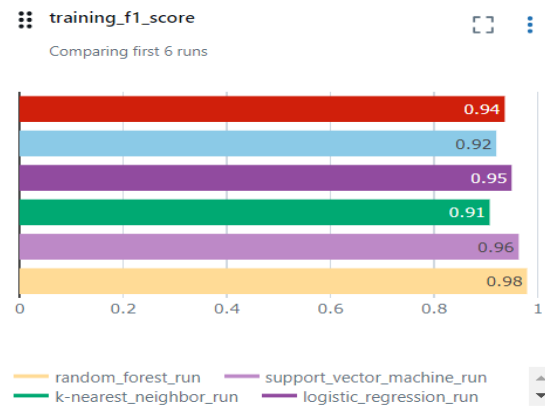
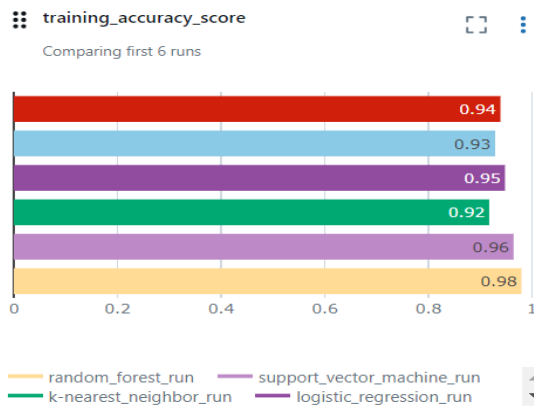
```
import mlflow
logged_model = 'runs:/adf6814d26ee4d538acc16db3dc87d/best_estimator'

# Load model as a PyFuncModel.
loaded_model = mlflow.pyfunc.load_model(logged_model)
```

Metric Plots

1. Click on **Chart** Tab to view Model **Metric** Plots





Hyperparameter Plots

1. Click on **Add Section**

The screenshot shows the mlflow Experiments page for the experiment 'sentimental_analysis_prediction_experimentation'. The 'Experimental' tab is selected, and the 'Add section' button is visible at the bottom of the chart area. The left sidebar shows the experiment list with 'sentimental_analysis_prediction_experimentation' selected. The top navigation bar includes the mlflow logo, version 2.11.3, and links to Experiments, Models, GitHub, and Docs. The search bar shows 'metrics.rmse < 1 and params.model = "tree"'. The 'Group by' dropdown is set to 'Run Name'. The table lists several runs, including 'random_forest_run', 'support_vector_mach...', 'k-nearest_neighbor_f...', 'logistic_regression_run', 'decision_tree_run', and 'naive_bayes_run'. The chart area shows 'Model metrics (9)' and 'System metrics (0)'.

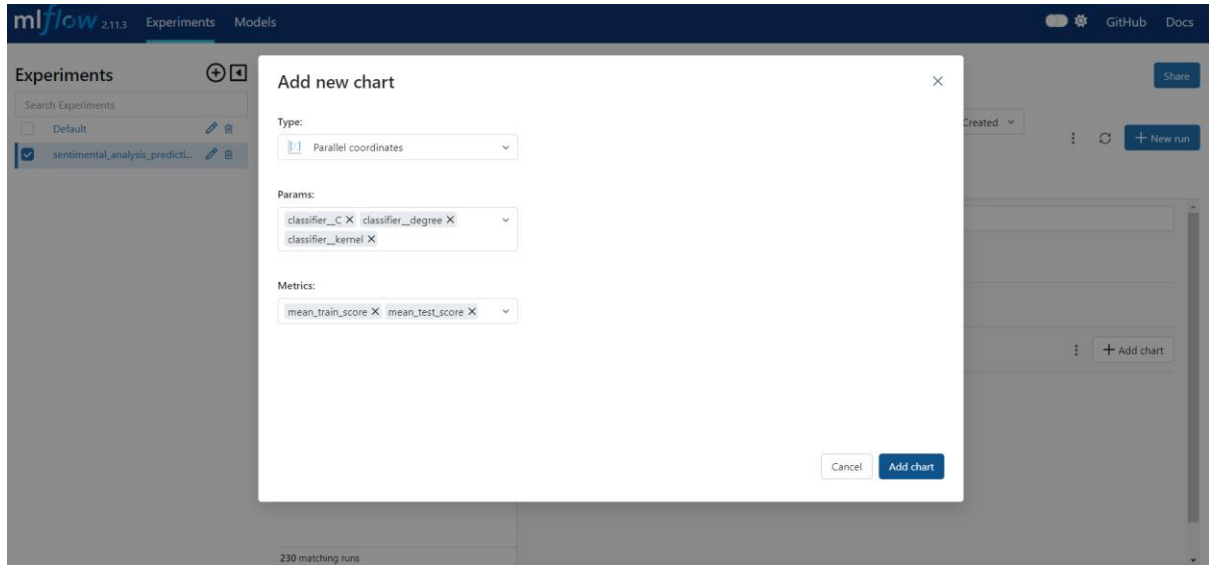
2. Enter the **Name** to the section

The screenshot shows the mlflow Experiments page for the experiment 'sentimental_analysis_prediction_experimentation'. The 'Experimental' tab is selected, and the 'Add section' button is visible at the bottom of the chart area. The left sidebar shows the experiment list with 'sentimental_analysis_prediction_experimentation' selected. The top navigation bar includes the mlflow logo, version 2.11.3, and links to Experiments, Models, GitHub, and Docs. The search bar shows 'metrics.rmse < 1 and params.model = "tree"'. The 'Group by' dropdown is set to 'Run Name'. The table lists several runs, including 'random_forest_run', 'support_vector_mach...', 'k-nearest_neighbor_f...', 'logistic_regression_run', 'decision_tree_run', and 'naive_bayes_run'. The chart area shows 'Model metrics (18)', 'System metrics (0)', and 'svm_visuals (0)'. The 'Add section' button is now labeled with the name 'svm_visuals (0)'.

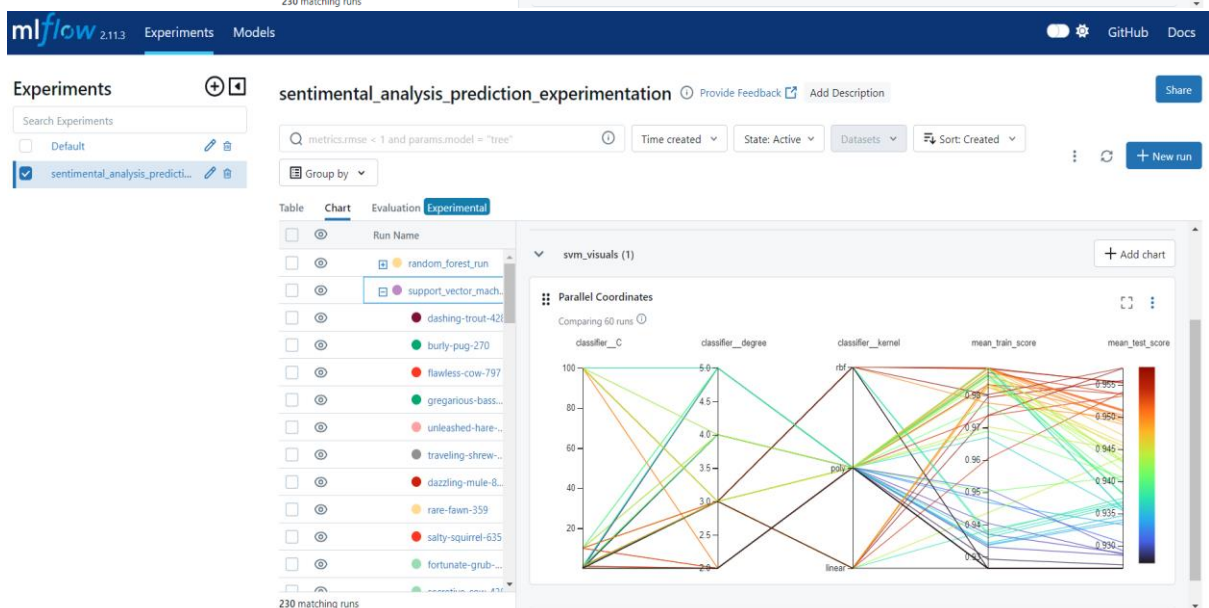
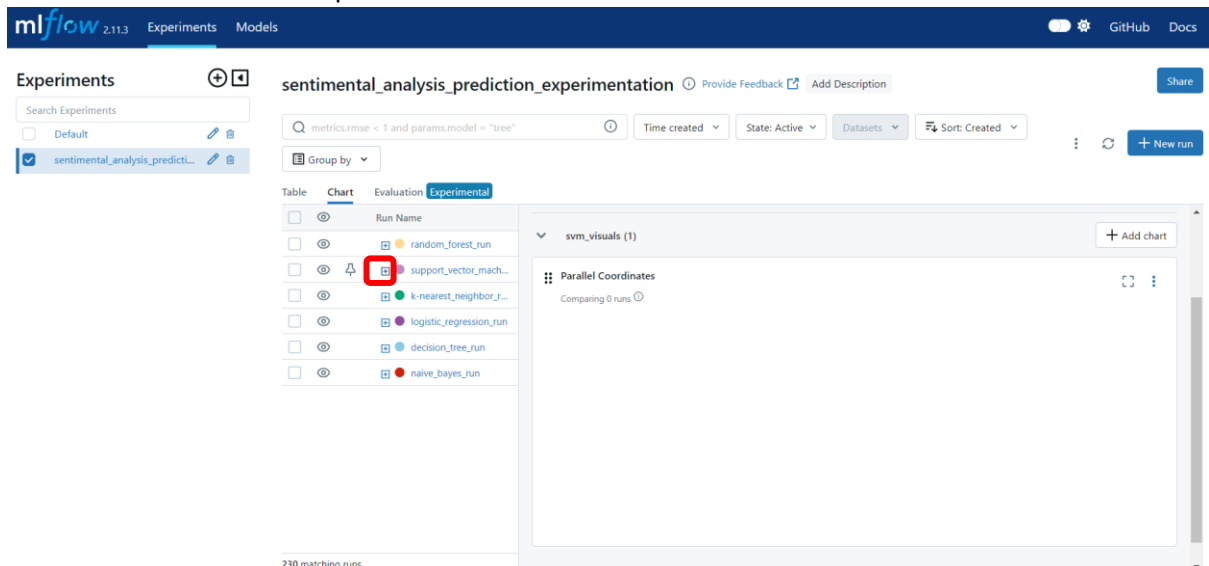
3. Click on **Add chart** → Select **Parallel coordinates**

The screenshot shows the mlflow Experiments page for the experiment 'sentimental_analysis_prediction_experimentation'. The 'Experimental' tab is selected, and the 'Add chart' button is visible at the bottom of the chart area. The left sidebar shows the experiment list with 'sentimental_analysis_prediction_experimentation' selected. The top navigation bar includes the mlflow logo, version 2.11.3, and links to Experiments, Models, GitHub, and Docs. The search bar shows 'metrics.rmse < 1 and params.model = "tree"'. The 'Group by' dropdown is set to 'Run Name'. The table lists several runs, including 'random_forest_run', 'support_vector_mach...', 'k-nearest_neighbor_f...', 'logistic_regression_run', 'decision_tree_run', and 'naive_bayes_run'. The chart area shows 'Model metrics (18)', 'System metrics (0)', and 'svm_visuals (0)'. The 'Add chart' button is now labeled with the name 'svm_visuals (0)'. A dropdown menu is open, showing options for 'Bar chart', 'Line chart', 'Parallel coordinates', 'Scatter chart', and 'Contour chart'. The 'Parallel coordinates' option is selected.

4. Select the **Params** and **Metrics** from drop-down → click on **Add chart**

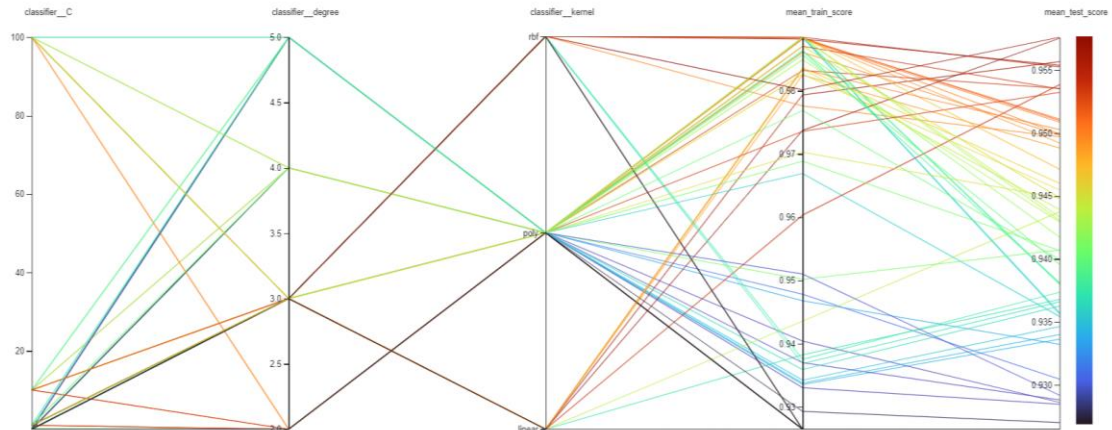


5. Click on '+' beside specified run name under **Chart** Tab



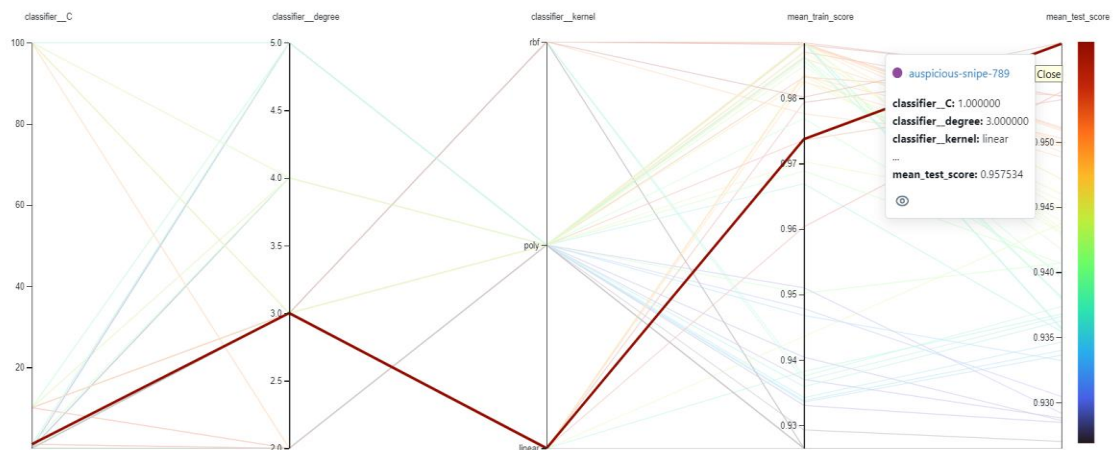
Parallel Coordinates

Comparing 60 runs



Parallel Coordinates

Comparing 60 runs



Registering Models

1. Click on run name → Click on Register model → Select best_estimator

mlflow 2.11.3 Experiments Models

sentimental_analysis_prediction_experimentation > **naive_bayes_run**

Overview Model metrics System metrics Artifacts

Description [🔗](#)
No description

Details

Created at	2024-03-26 13:46:30
Created by	hanna
Status	🟢 Finished
Run ID	639bbdf11e0e4b70b0693a4fb7554855
Duration	19.0s
Datasets used	—
Tags	estimator_class: sklearn.model_selection_search.G... estimator_name: GridSearchCV 🔗
Source	C:\Users\hanna\anaconda3\Lib\site-packages\ipykernel_launcher.py
Logged models	🔗 sklearn >1
Registered models	—

Parameters (13) Metrics (9)

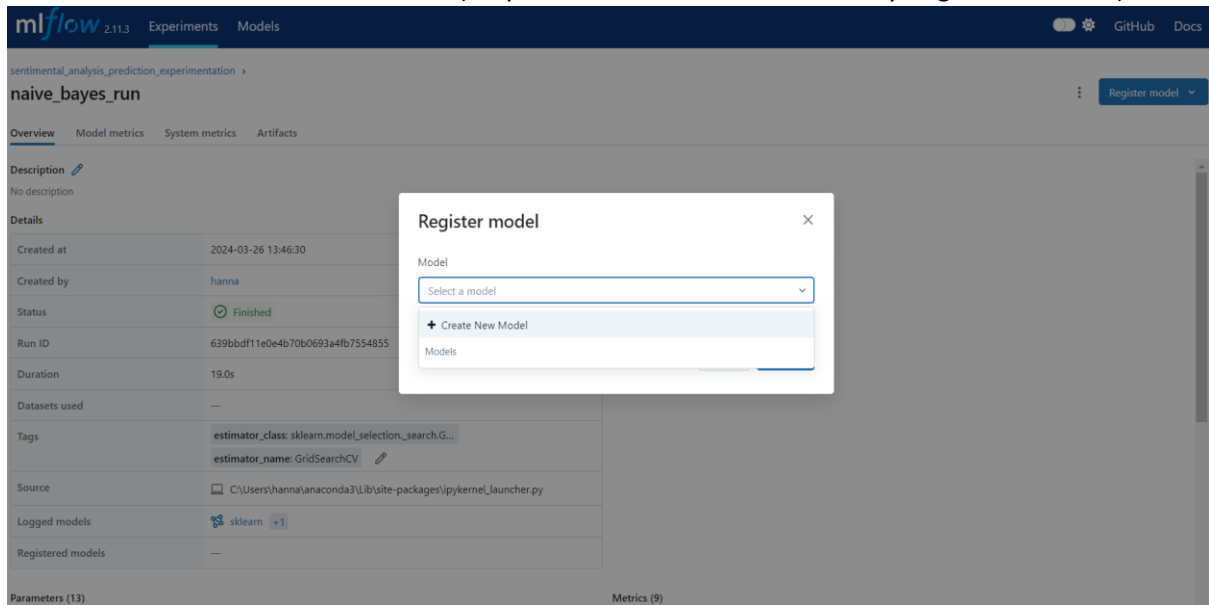
Register model

Unregistered models

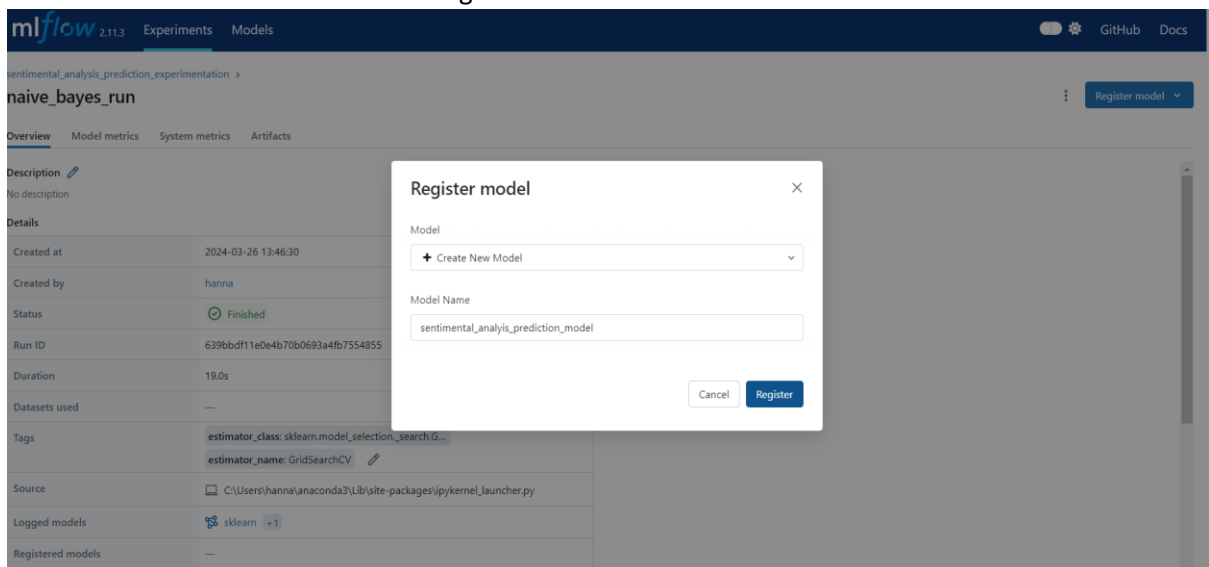
best_estimator View model [🔗](#)

model View model [🔗](#)

2. Click on + Create New Model (only for first time as I don't have any Registered Model)



3. Enter Model Name → click Register



We can view under Models column which models are registered and its version.

The screenshot shows the mlflow 'Models' page. A table lists the registered models. The 'Models' column is circled in red, showing the registered models and their versions.

Run Name	Created	Duration	Source	Models	Metrics
random_forest_run	1 hour ago	6.4min	C:\Users\...	sklearn, 1 more	0.964424514...
support_vector_mach...	1 hour ago	11.2min	C:\Users\...	sklearn, 1 more	0.963447251...
k-nearest_neighbor_f...	1 hour ago	5.7min	C:\Users\...	sentimental..., 4, 1 more	0.961254066...
logistic_regression_run	1 hour ago	14.9min	C:\Users\...	sentimental..., 3, 1 more	0.944993819...
decision_tree_run	1 hour ago	1.2min	C:\Users\...	sentimental..., 2, 1 more	0.961606615...
naive_bayes_run	1 hour ago	19.0s	C:\Users\...	sentimental..., 1, 1 more	0.961515333...

Managing and Tagging the Models

1. Click on 'Models' Tab on the top of the dashboard.

The screenshot shows the mlflow 2.11.3 interface. The 'Models' tab is highlighted in the top navigation bar. Below the navigation bar, the 'Experiments' section is visible on the left, and the main area displays the 'sentimental_analysis_prediction_experimentation' experiment. The 'Experimental' tab is selected, showing a table of runs. The table has columns for Run Name, Created, Dataset, Duration, Source, and Models. Three runs are listed: 'random_forest_run', 'support_vector_mach...', and 'k-nearest_neighbor_r...'. Each run has a 'Created' timestamp of '1 hour ago' and a 'Dataset' of '-'. The 'Duration' column shows '6.4min', '11.2min', and '5.7min' respectively. The 'Source' column shows 'C:\Users\...' for each run. The 'Models' column shows 'sentimental_.../5, 1 more', 'sentimental_.../5, 1 more', and 'sentimental_.../4, 1 more' respectively. A 'Load more' button is at the bottom right of the table.

2. Click on the Registered model name i.e., sentimental_analysis_prediction_model.

The screenshot shows the mlflow 2.11.3 interface with the 'Models' tab selected. The 'Registered Models' section is displayed, showing a table of registered models. The table has columns for Name, Latest version, Staging, Production, Created by, Last modified, and Tags. One model is listed: 'sentimental_analis_prediction_model' with 'Version 6' as the latest version. The 'Staging' and 'Production' columns are empty. The 'Created by' column is empty. The 'Last modified' column shows '2024-03-26 15:46:00'. The 'Tags' column is empty. A 'Create Model' button is at the top right.

3. Choose the version → click on Stage menu, Select the appropriate option as 'Staging', 'Production' and 'Archived'.

Archived: These versions are no longer in active use.

Staged: These versions are ready for deployment pending final validation.

Production: These versions are actively serving users in live environments.

The screenshot shows the mlflow 2.11.3 interface with the 'Models' tab selected. The 'Registered Models' section is displayed, showing a table of registered models. The table has columns for Version, Registered at, Created by, Stage, and Description. Six versions are listed: 'Version 6', 'Version 5', 'Version 4', 'Version 3', 'Version 2', and 'Version 1'. Each version has a 'Registered at' timestamp and a 'Created by' field. The 'Stage' column shows 'None' for all versions. The 'Description' column is empty. A dropdown menu is open for the 'Stage' column, showing options: 'None', 'Staging', 'Production', and 'Archived'. The 'Staging' option is highlighted. A 'New model registry UI' button is at the top right.

The screenshot shows the mlflow 2.11.3 interface with the 'Models' tab selected. The 'Version 6' details are displayed. The 'Registered At' is '2024-03-26 15:46:00', the 'Last Modified' is '2024-03-26 16:05:18', and the 'Source Run' is 'random_forest_run'. The 'Stage' is 'None'. A dropdown menu is open for the 'Stage' column, showing options: 'None', 'Staging', 'Production', and 'Archived'. The 'Staging' option is highlighted. The 'Description' field is empty. The 'Tags' field is empty. The 'Schema' section is expanded, showing a table with columns for Name and Type. The 'Inputs (1)' section is expanded, showing a table with columns for Name and Type. The 'Outputs (1)' section is expanded, showing a table with columns for Name and Type.

Registered Models >

sentimental_analysis_prediction_model

⋮

Created Time: 2024-03-26 15:38:53

Last Modified: 2024-03-26 16:06:20

> Description

Edit

> Tags

▼ Versions

AllActive 5Compare

New model registry UI

☐

Version	Registered at ↕	Created by	Stage	Description
✔ Version 6	2024-03-26 15:46:00		Production	
✔ Version 5	2024-03-26 15:45:42		Staging	
✔ Version 4	2024-03-26 15:40:30		Staging	
✔ Version 3	2024-03-26 15:40:01		Archived	
✔ Version 2	2024-03-26 15:39:34		Staging	
✔ Version 1	2024-03-26 15:38:53		Staging	