

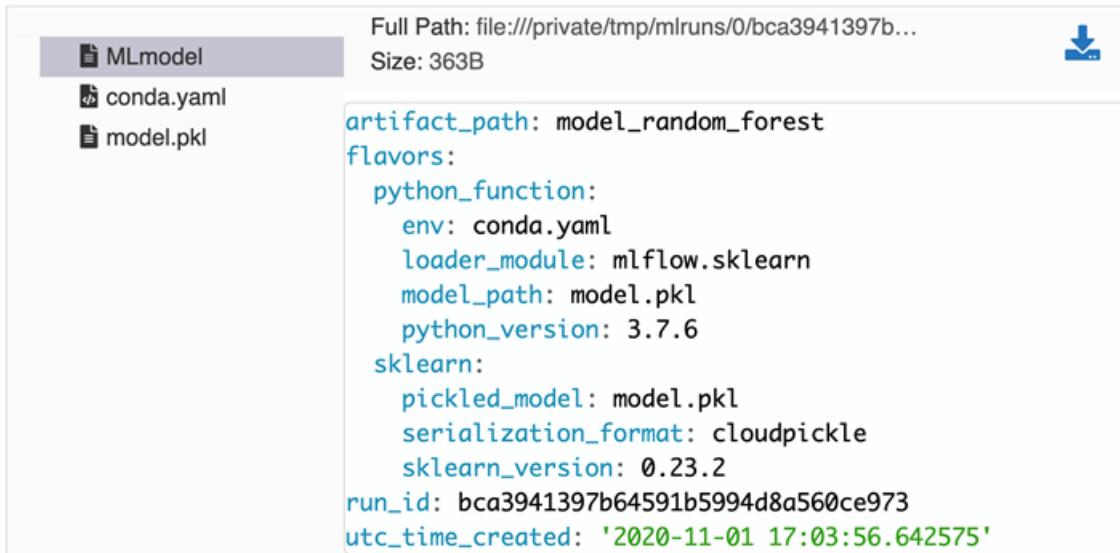
Chapter 1: Introducing MLflow



The screenshot shows the MLflow UI for a completed run. The top navigation bar includes 'mlflow', 'Experiments', and 'Models' tabs, and 'GitHub' and 'Docs' links. The main content area displays the following details for Run 442275f18d354564b6259e0188a12575:

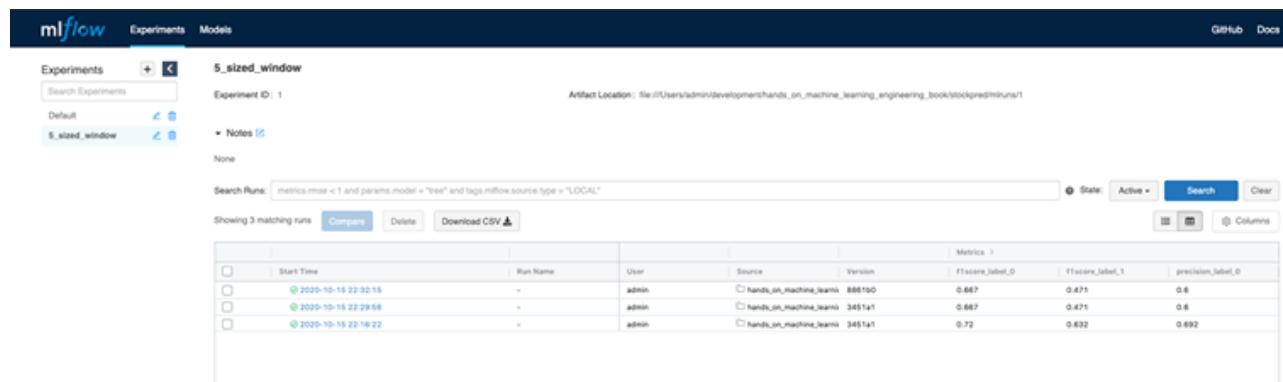
- Date:** 2020-10-15 19:19:35
- Entry Point:** main
- Status:** FINISHED
- Run Command:** mlflow run file:///Users/admin/development/hands_on_machine_learning_engineering_book#stockpred -v 3451a1f1f334d8863cba521d29cd6becb2788ed -b local
- Git Commit:** 3451a1f1f334d8863cba521d29cd6becb2788ed
- Duration:** 4.3s
- Notes:** Experiment with a window size for the prediction of 5 rolling days.
- Parameters:** A table showing Name and Value for parameters like 'threshold' (0.5), 'precision_label_0' (0.611), and 'precision_label_1' (0.389).
- Metrics:** A table showing Name and Value for metrics like 'f1score_label_0' (0.759), 'f1score_label_1' (0.364), and 'recall_label_0' (1.0).
- Tags:** A table showing Name, Value, and Actions. No tags found.

▼ Artifacts



The screenshot shows the artifacts for the run. The left sidebar lists 'MLmodel', 'conda.yaml', and 'model.pkl'. The main content area shows the following JSON configuration:

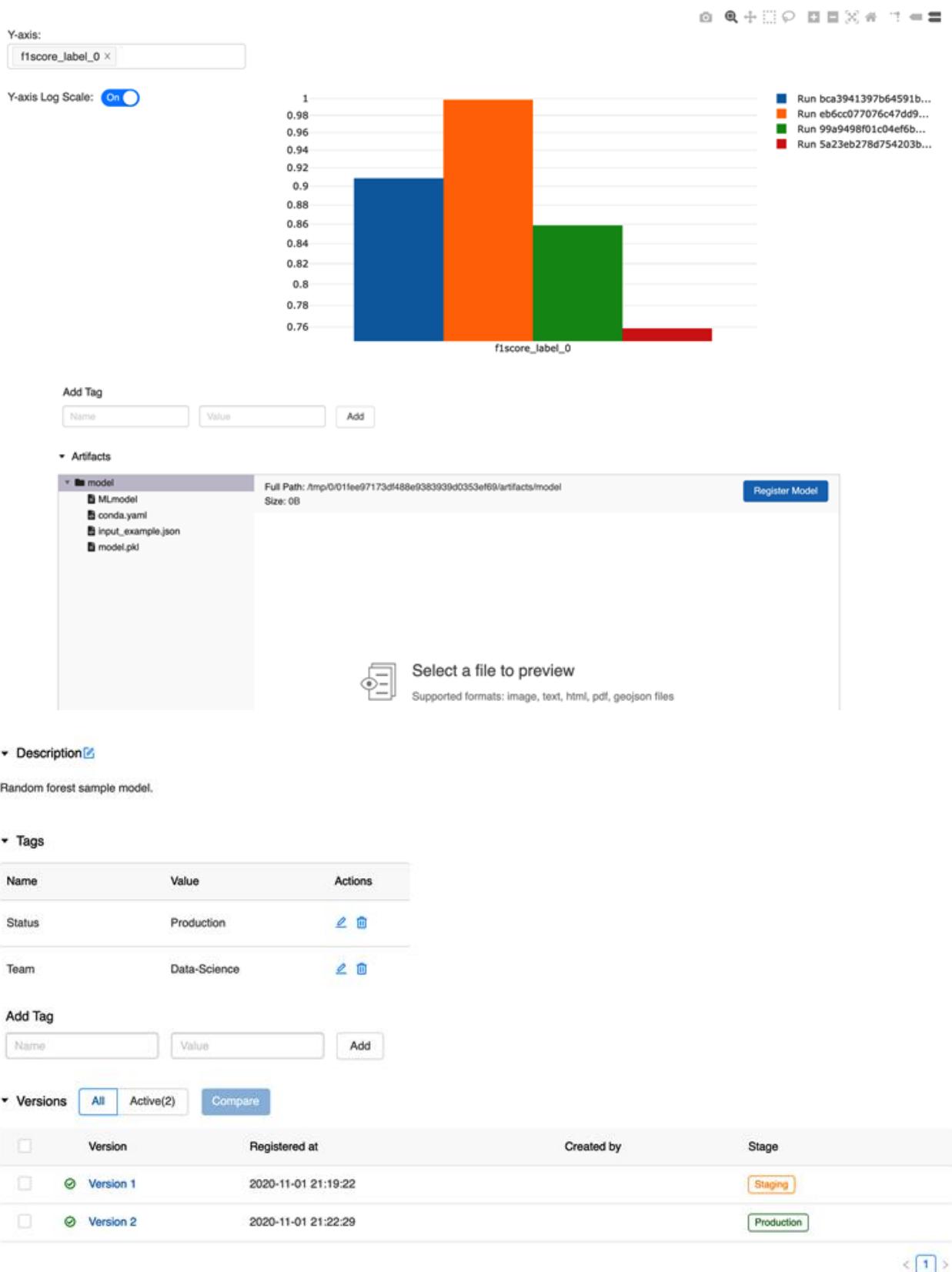
```
artifact_path: model_random_forest
flavors:
  python_function:
    env: conda.yaml
    loader_module: mlflow.sklearn
    model_path: model.pkl
    python_version: 3.7.6
  sklearn:
    pickled_model: model.pkl
    serialization_format: cloudpickle
    sklearn_version: 0.23.2
run_id: bca3941397b64591b5994d8a560ce973
utc_time_created: '2020-11-01 17:03:56.642575'
```



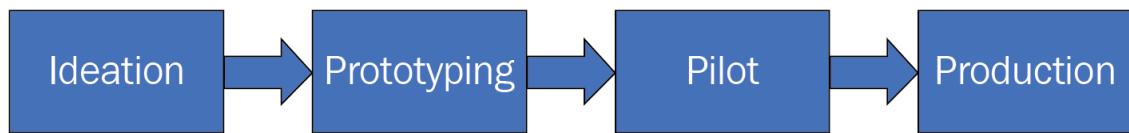
The screenshot shows the 'Experiments' page. The search bar includes 'Search Experiments' and 'Artifact Location: file:///Users/admin/development/hands_on_machine_learning_engineering_book#stockpred/mlruns/1'. The table lists three matching runs:

Start Time	Run Name	User	Source	Version	f1score_label_0	f1score_label_1	precision_label_0
2020-10-15 22:32:15	-	admin	hands_on_machine_learn...	8861a0	0.667	0.471	0.6
2020-10-15 22:29:56	-	admin	hands_on_machine_learn...	3451a1	0.667	0.471	0.6
2020-10-15 22:16:22	-	admin	hands_on_machine_learn...	3451a1	0.72	0.632	0.692

Default > Comparing 4 Runs > f1score_label_0



Chapter 2: Your Machine Learning Project

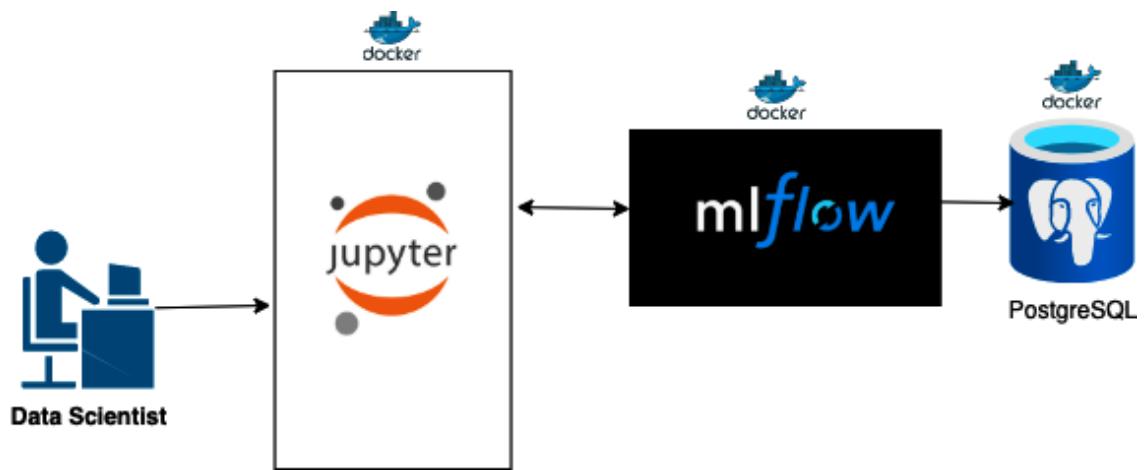
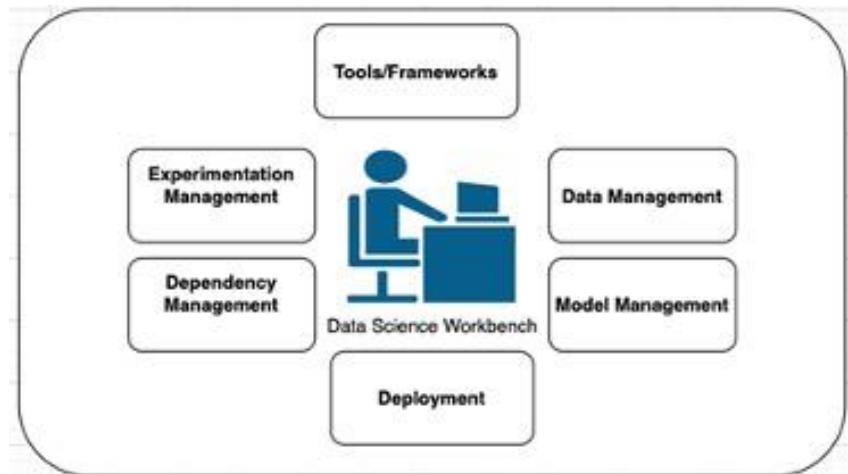


Sepal Length – Input	Sepal Length - Input	Sepal Length - Input	Sepal Length – Input	Class(Label)
Integer	Integer	Integer	Integer	String(Label)
10	20	30	40	Iris Setosa

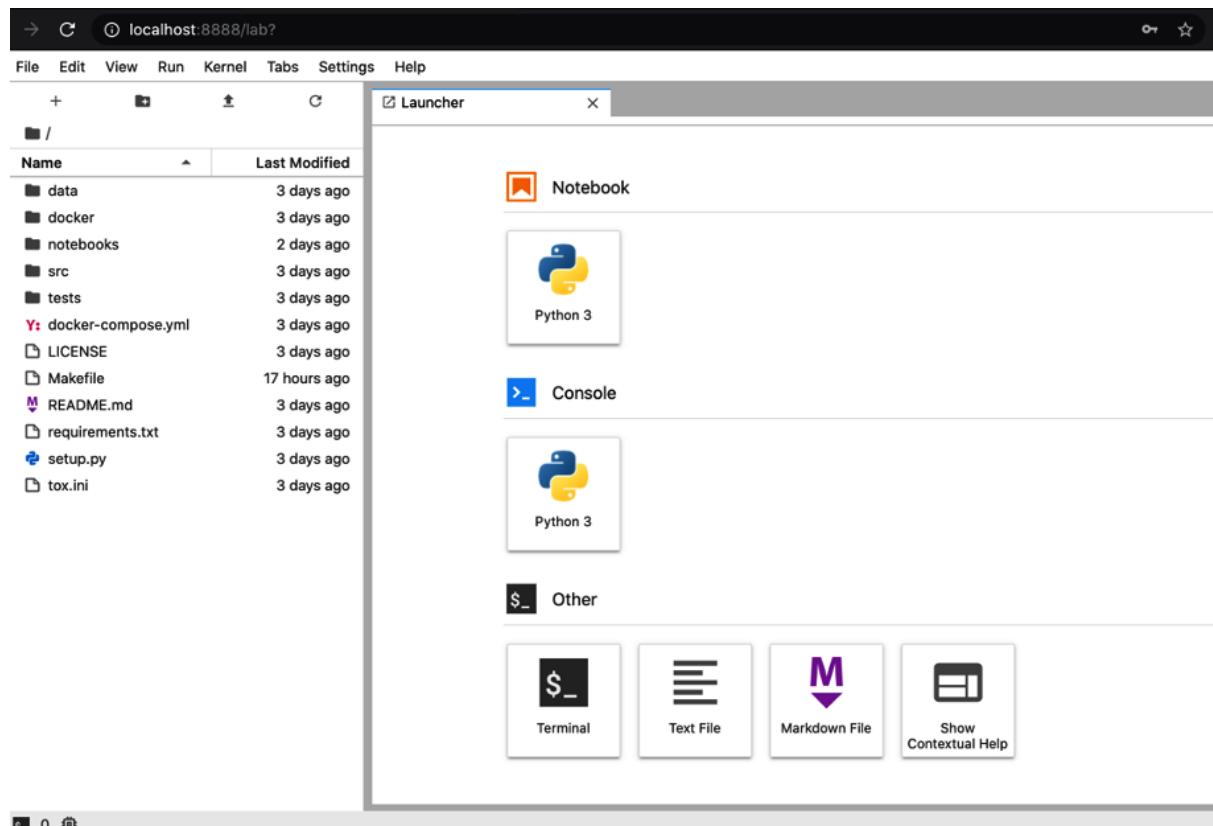
Day N-N Input	Day N-9 Input	...	Day N-1 Input	Class(Label)
Float	Float	Float	Float	Integer
103.2	203.1	...	200.1	1

Tweet Input	Class(Label)
String	Integer
“It seems that \$AMZN will break up positively”	Positive

Chapter 3: Your Data Science Workbench



CONTAINER ID	IMAGE
0dcf246e0aa5	gradflow/workbench/jupyter:0.1.0
pyter_1	
4ea4277255d0	gradflow/workbench/mlflow:0.1.0
flow_1	
98a0ce9ff504	gradflow/workbench/postgres:0.1.0
stgres_1	
(base) → Desktop	



The screenshot shows the mlflow UI. The top navigation bar includes 'mlflow', 'Experiments', 'Models', 'GitHub', and 'Docs'. The 'Experiments' tab is selected, showing a list with 'default' highlighted. A search bar and a 'Create Experiment' button are visible. A tooltip message says: 'Track machine learning training runs in an experiment. [Learn more](#)' with a close button 'X'. Below the experiment list, it shows 'Experiment ID: 3' and 'Artifact Location: /data/artifacts/3'. A 'Notes' section is expanded, showing 'None'. A search bar for 'Search Runs' with a placeholder 'metrics.rmse < 1 and params.model = "tree" and ...', and buttons for 'Filter', 'Search', and 'Clear' are present. A table header for 'Showing 0 matching runs' includes columns for 'Start Time', 'Run Name', 'User', 'Source', 'Version', and 'Models'. A message at the bottom of the table says: 'No runs yet. [Learn more](#) about how to create ML model training runs in this experiment.'

```
[1]: import mlflow
mlflow.set_experiment('mlflow_experiment')
with mlflow.start_run():
    mlflow.log_param("name", "mlflow_test")
    mlflow.log_param("category", "algorithm")
    mlflow.log_param("type", "classification")

    for i in range(5):
        mlflow.log_metric("i", i, step=i)
        mlflow.log_artifact("mlflow_example.ipynb")
INFO: 'mlflow_experiment' does not exist. Creating a new experiment
```

Experiments + <

mlflow_experiment

Search Experiments

Experiment ID: 2 Artifact Location: /data/artifacts/2

Default + <

mlflow_experiment + <

▼ Notes + <

None

Search Runs: metrics.rmse < 1 and params.model = "tree" and tags + < State: Active + < Search + < Clear

Showing 1 matching run + < Compare + < Delete + < Download CSV + < + < + < Columns

					Parameters		Metrics		
	Start Time	Run Name	User	Source	Version	category	name	type	i
	2020-12-28 10:11	-	admin	ipykernel	-	algorit...	mlflow...	classif...	4

mlflow_experiments > Run 20193c2881c741b98e0e0a8652465c3b

Date: 2021-07-02 19:01:10 Source: ipykernel_launcher.py User: admin

Duration: 184ms Status: FINISHED

► Notes + <

► Parameters

▼ Metrics

Name	Value
i	4

► Tags

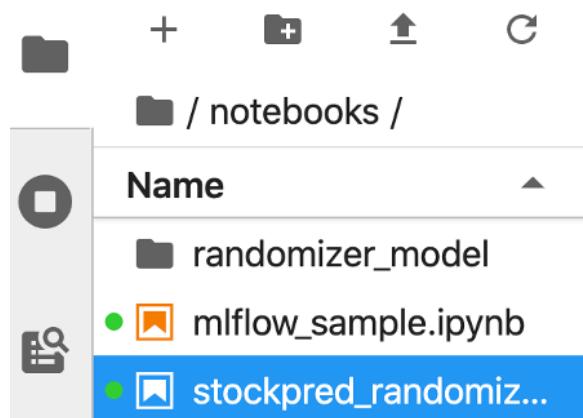
▼ Artifacts

+ < mlflow_sample.ipynb	Full Path: /data/artifacts/4/20193c2881c741b98e0e0a8652465c3b/artifacts/mlflow_s...	+ <
---------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------	-------------------------------------------------------------------

```
import random
import mlflow
from mlflow.pyfunc.model import PythonModel
```

```
class RandomPredictor(PythonModel):  
    def __init__(self):  
        pass  
  
    def predict(self, context, model_input):  
        return model_input.apply(lambda column: random.randint(0,1))
```

```
# Construct and save the model  
model_path = "randomizer_model"  
r = RandomPredictor()  
mlflow.pyfunc.save_model(path=model_path, python_model=r)
```



```
# Load the model in `python_function` format  
loaded_model = mlflow.pyfunc.load_model(model_path)
```

```
import pandas as pd  
model_input = pd.DataFrame([range(10)])  
  
random_predictor = RandomPredictor()  
  
mlflow.set_experiment('stockpred_experiment_days_up')  
with mlflow.start_run():  
    model_output = loaded_model.predict(model_input)  
  
    mlflow.log_metric("Days Up",model_output.sum())  
    mlflow.log_artifact("stockpred_randomizer.ipynb")
```

Experiments  

Search Experiments

Experiment ID: 4 Artifact Location: /data/artifacts/4

Default   stockpred_experiment   stockpred_experiment  stockpred_experiment... 

▼ Notes 

None

Search Runs: metrics.rmse < 1 and params.model = "tree" and tags  State: Active 

Showing 6 matching runs    

	Start Time	Run Name	User	Source	Version	Days Down	Days Up
<input type="checkbox"/>	2020-12-28 13:06:44	-	admin	ipykernel_launcher.py	-	-	8
<input type="checkbox"/>	2020-12-28 13:04:27	-	admin	ipykernel_launcher.py	-	-	5
<input type="checkbox"/>	2020-12-28 13:02:44	-	admin	ipykernel_launcher.py	-	1	0.5
<input type="checkbox"/>	2020-12-28 13:05:41	-	admin	ipykernel_launcher.py	-	-	-
<input type="checkbox"/>	2020-12-28 13:01:52	-	admin	ipykernel_launcher.py	-	-	-
<input type="checkbox"/>	2020-12-28 13:00:41	-	admin	ipykernel_launcher.py	-	-	-

mlflow Experiments Models GitHub Docs

stockpred_experiment_day5_up > Run 3201f594109a45d582a01151be8023fb ▾

Date: 2021-07-02 19:40:23 Source: ipykernel_launcher.py User: admin

Duration: 87ms Status: FINISHED

▼ Notes 

None

► Parameters

▼ Metrics

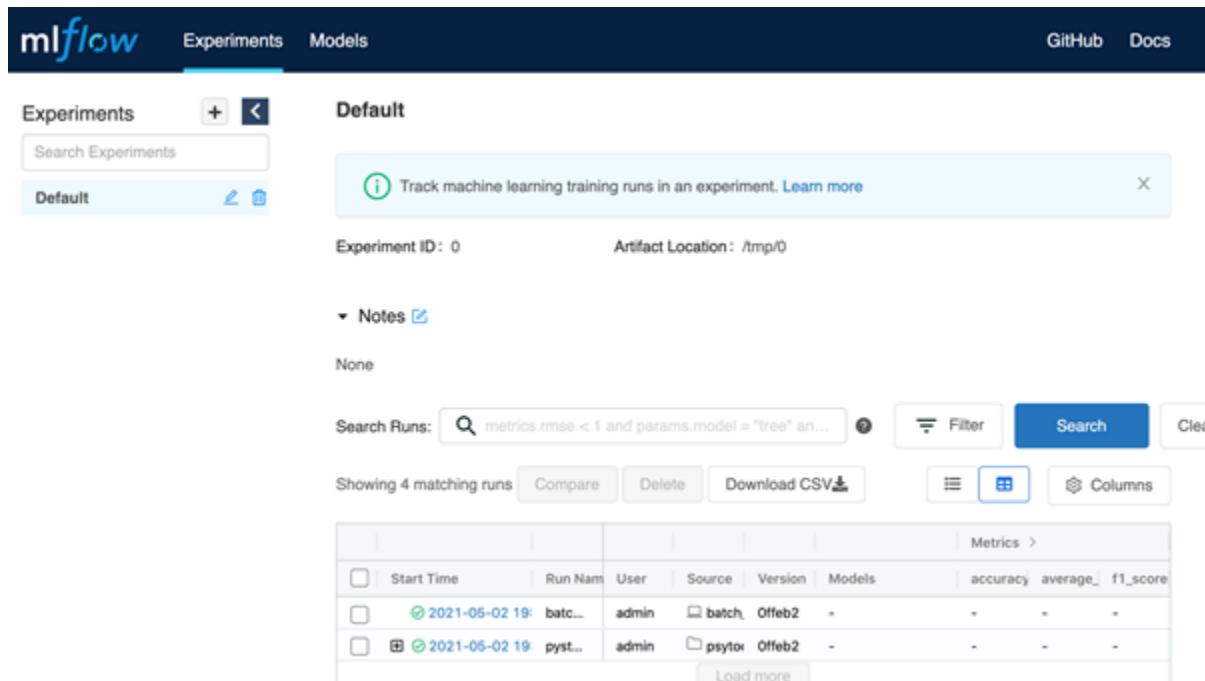
Name	Value
Days Up 	6

► Tags

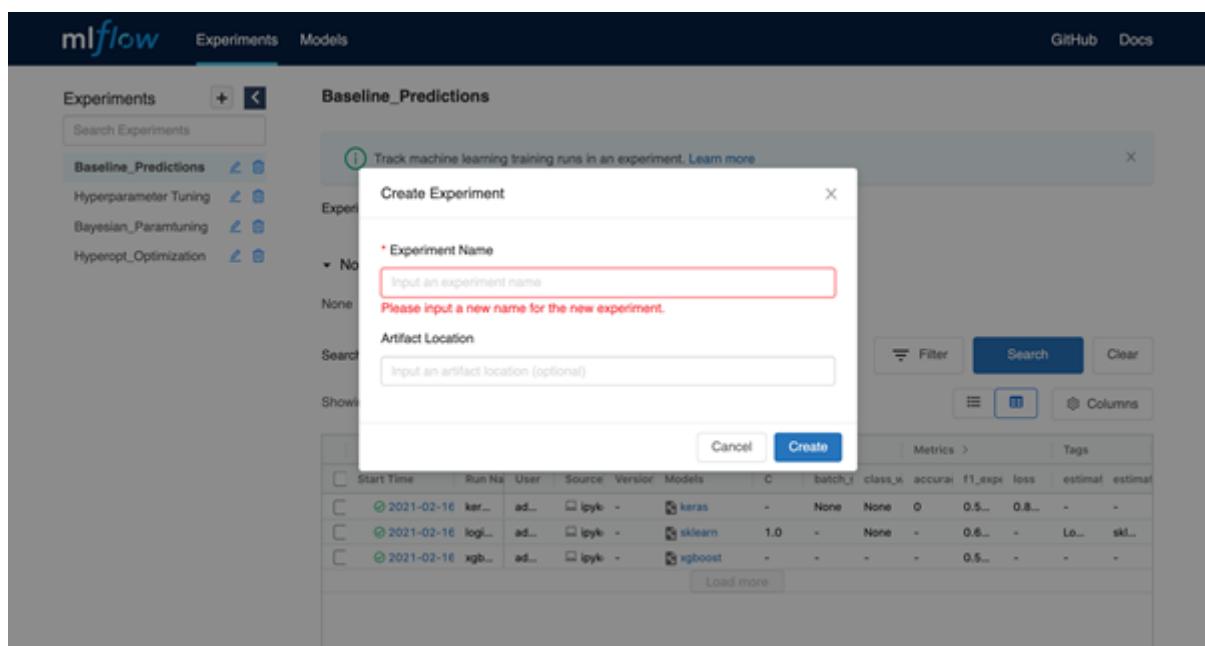
▼ Artifacts

 stockpred_randomizer.ipynb

Chapter 4: Experiment Management in MLflow



The screenshot shows the MLflow Experiment Management interface. The top navigation bar includes 'mlflow', 'Experiments', 'Models', 'GitHub', and 'Docs'. The 'Experiments' tab is selected. The main area is titled 'Default' with a sub-section 'Experiment ID: 0' and 'Artifact Location: /tmp/0'. A note says 'Track machine learning training runs in an experiment. Learn more'. Below this, a 'Notes' section is collapsed. A search bar shows a query: 'metrics.rmse < 1 and params.model = "tree" an...'. Buttons for 'Filter', 'Search', and 'Clear' are available. A table displays 'Showing 4 matching runs' with columns: Start Time, Run Name, User, Source, Version, Models, accuracy, average_, f1_score. The runs are: 2021-05-02 19:54:55 batch_1 admin batch OffFeb2 - - - - - - - -; 2021-05-02 19:54:55 pyst_1 admin psytor OffFeb2 - - - - - - - -.

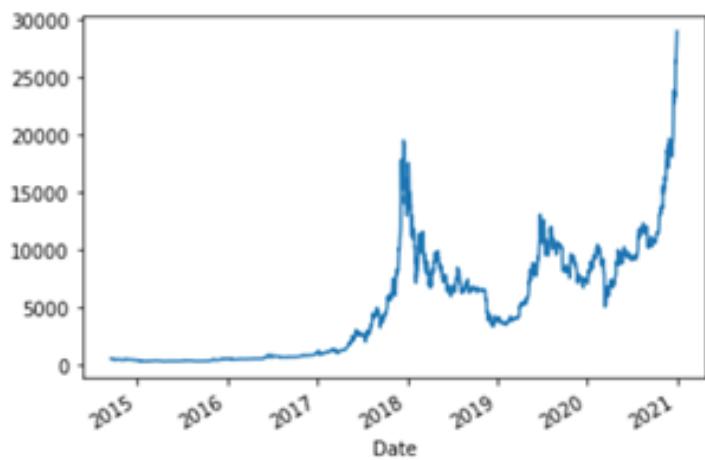


The screenshot shows the MLflow Experiment Management interface. The top navigation bar includes 'mlflow', 'Experiments', 'Models', 'GitHub', and 'Docs'. The 'Experiments' tab is selected. The main area is titled 'Baseline_Predictions' with a note 'Track machine learning training runs in an experiment. Learn more'. A 'Create Experiment' dialog box is open, prompting for 'Experiment Name' (input field: 'Input an experiment name') and 'Artifact Location' (input field: 'Input an artifact location (optional)'). Below the dialog, a table shows 'Showing 3 matching runs' with columns: Start Time, Run Name, User, Source, Version, Models, C, batch_size, class_w, accuracy, f1_score, loss, estimat, estimat. The runs are: 2021-02-16 keras_1 ad_... psytor - keras - None None 0 0.5... 0.8... - -; 2021-02-16 logit_1 ad_... psytor - sklearn 1.0 - None - 0.6... - Log... skil...; 2021-02-16 xgb_1 ad_... psytor - xgboost - - - - 0.5... - - - -.

[3]:

	High	Low	Open	Close	Volume	Adj Close
Date						
2014-09-16	468.174011	452.421997	465.864014	457.334015	2.105680e+07	457.334015
2014-09-17	456.859985	413.104004	456.859985	424.440002	3.448320e+07	424.440002
2014-09-18	427.834991	384.532013	424.102997	394.795990	3.791970e+07	394.795990
2014-09-19	423.295990	389.882996	394.673004	408.903992	3.686360e+07	408.903992
2014-09-20	412.425995	393.181000	408.084991	398.821014	2.658010e+07	398.821014
...
2020-12-28	27389.111328	26207.640625	26280.822266	27084.808594	4.905674e+10	27084.808594
2020-12-29	27370.720703	25987.298828	27081.810547	27362.437500	4.526595e+10	27362.437500
2020-12-30	28937.740234	27360.089844	27360.089844	28840.953125	5.128744e+10	28840.953125
2020-12-31	29244.876953	28201.992188	28841.574219	29001.720703	4.675496e+10	29001.720703
2021-01-01	29600.626953	28803.585938	28994.009766	29374.152344	4.073030e+10	29374.152344

[2]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8928e74850>



Baseline_Predictions > logistic_regression_model_baseline

Date: 2021-02-16 11:53:44

Source: ipykernel_launcher.py

Duration: 0.7s

Status: FINISHED

Notes

None

Parameters

Name	Value
C	1.0
class_weight	None
dual	False
fit_intercept	True
intercept_scaling	1
l1_ratio	None
max_iter	100
multi_class	auto
n_jobs	None

Artifacts

model

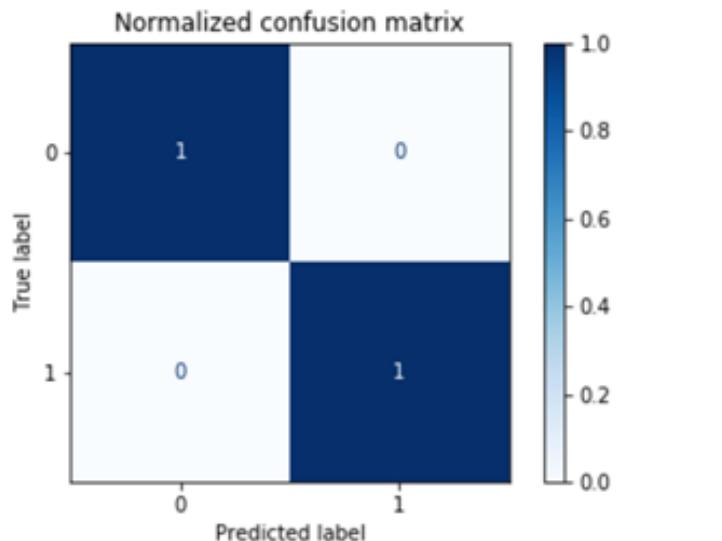
training_confusion_matrix.png

Full Path: /data/artifacts/3/14a8e87829524ff8b9f3ab0f70366271/artifacts/training_confusion_ma...

Size: 9.96KB

training_precision_recall_curve.png

training_roc_curve.png



Baseline_Predictions > xgboost_model_baseline ▾

Date: 2021-02-16 11:53:17

Source: ipykernel_launcher.py

Duration: 492ms

Status: FINISHED

▼ Notes ↗

None

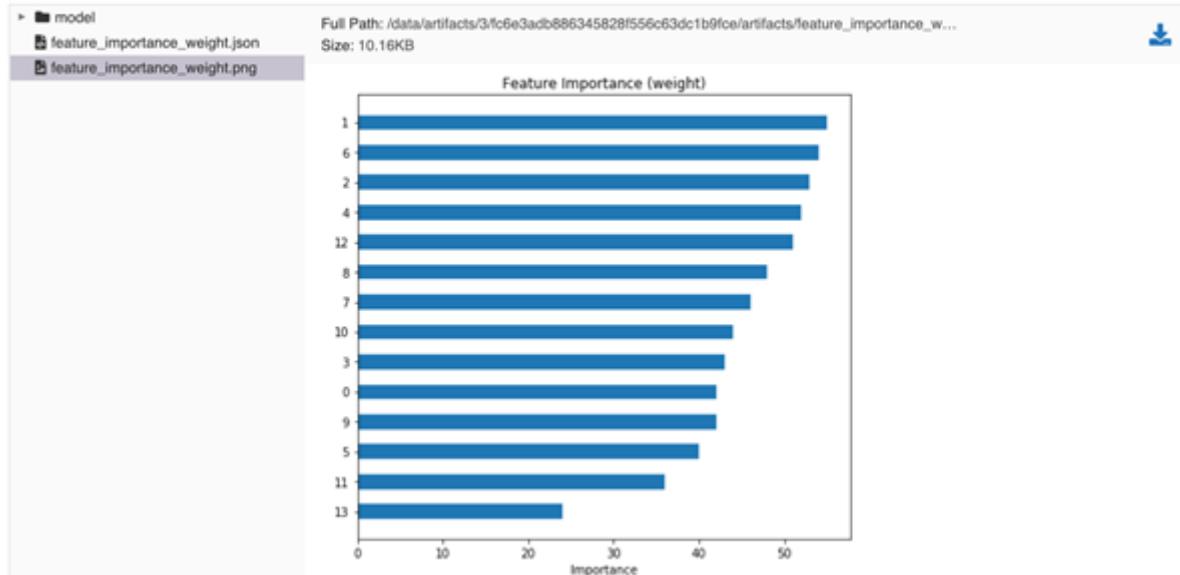
▼ Parameters

Name	Value
early_stopping_rounds	None
maximize	None
num_boost_round	10
verbose_eval	True

▼ Metrics

Name	Value
f1_experiment_score ↗	0.574

▼ Artifacts



Baseline_Predictions > keras_model_baseline ▾

Date: 2021-02-16 12:01:43

Source: ipykernel_launcher.py

Duration: 5.1s

Status: FINISHED

▼ Notes ↗

None

▼ Parameters

Name	Value
batch_size	None
class_weight	None
epochs	20
initial_epoch	0
max_queue_size	10
opt_amsgrad	False
opt_beta_1	0.9

			Metrics
	Start Time	Run Name	Models
<input type="checkbox"/>	2021-02-16 12:01:43	keras_model_baseline	f1_experiment_score
<input checked="" type="checkbox"/>	2021-02-16 12:01:43	keras_model_baseline	keras 0.583
<input checked="" type="checkbox"/>	2021-02-16 11:53:16	logistic_regression	sklearn 0.663
<input checked="" type="checkbox"/>	2021-02-16 11:53:16	xgboost	xgboost 0.574

[Load more](#)

Metrics

f1_experiment_score ↗	0.663	0.574
training_accuracy_score ↗	0.562	
training_f1_score ↗	0.52	
training_log_loss ↗	0.682	
training_precision_score ↗	0.552	
training_recall_score ↗	0.562	
training_roc_auc_score ↗	0.564	
training_score ↗	0.562	

Hyperopt_Optimization

i Track machine learning training runs in an experiment. [Learn more](#) X

Experiment ID: 6 Artifact Location: /data/artifacts/6

▼ Notes ↗

None

Search Runs: ... Filter Search Clear

Showing 11 matching runs Compare Delete Download CSV Columns

			Parameters >			Metrics >			Tags	
	Start Time	Source	Models	C	class_wei	dual	training_accu	training_f	training_I	estimat...
<input type="checkbox"/>	2021-02-17 19:...	ipykernel_Jaunc	-	-	-	-	-	-	-	-
<input type="checkbox"/>	2021-02-17 19:...	ipykernel_Jaunc	sklearn	2.99...	None	False	0.569	0.535	0.68	Logis...
<input type="checkbox"/>	2021-02-17 19:...	ipykernel_Jaunc	sklearn	2.73...	None	False	0.569	0.535	0.68	Logis...
<input type="checkbox"/>	2021-02-17 19:...	ipykernel_Jaunc	sklearn	2.73...	None	False	0.569	0.535	0.68	Logis...
<input type="checkbox"/>	2021-02-17 19:...	ipykernel_Jaunc	sklearn	1.65...	None	False	0.568	0.533	0.68	Logis...
<input type="checkbox"/>	2021-02-17 19:...	ipykernel_Jaunc	sklearn	2.96...	None	False	0.572	0.532	0.681	Logis...
<input type="checkbox"/>	2021-02-17 19:...	ipykernel_Jaunc	sklearn	1.91...	None	False	0.568	0.534	0.68	Logis...
<input type="checkbox"/>	2021-02-17 19:...	ipykernel_Jaunc	sklearn	0.48...	None	False	0.567	0.526	0.681	Logis...
<input type="checkbox"/>	2021-02-17 19:...	ipykernel_Jaunc	sklearn	1.86...	None	False	0.57	0.53	0.681	Logis...

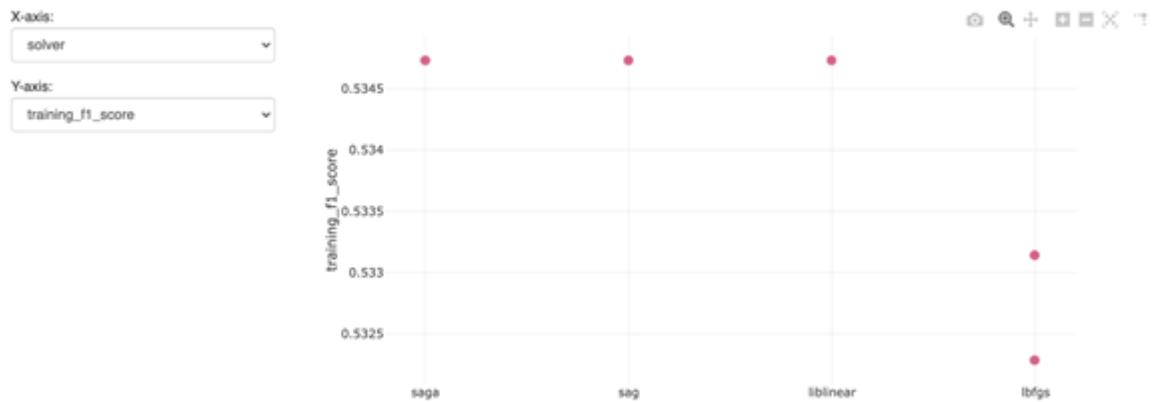
Parameters

C	2.995901125450704	2.7337497772805657	2.7345602113532017	1.6543252750173758	2.9604141428659534
class_weight	None	None	None	None	None
dual	False	False	False	False	False
fit_intercept	True	True	True	True	False
intercept_scaling	1	1	1	1	1
l1_ratio	None	None	None	None	None
max_iter	816	249	742	925	672
multi_class	auto	auto	auto	auto	auto
n_jobs	None	None	None	None	None
penalty	l2	l2	l2	l2	l2
random_state	0	0	0	0	0
solver	saga	sag	lbfgs	lbfgs	lbfgs
tol	5.699651203911781e-05	2.262203680356659e-05	7.578158461245823e-05	8.659147542059882e-05	9.34241158286789e-05
verbose	0	0	0	0	0
warm_start	False	False	True	False	True

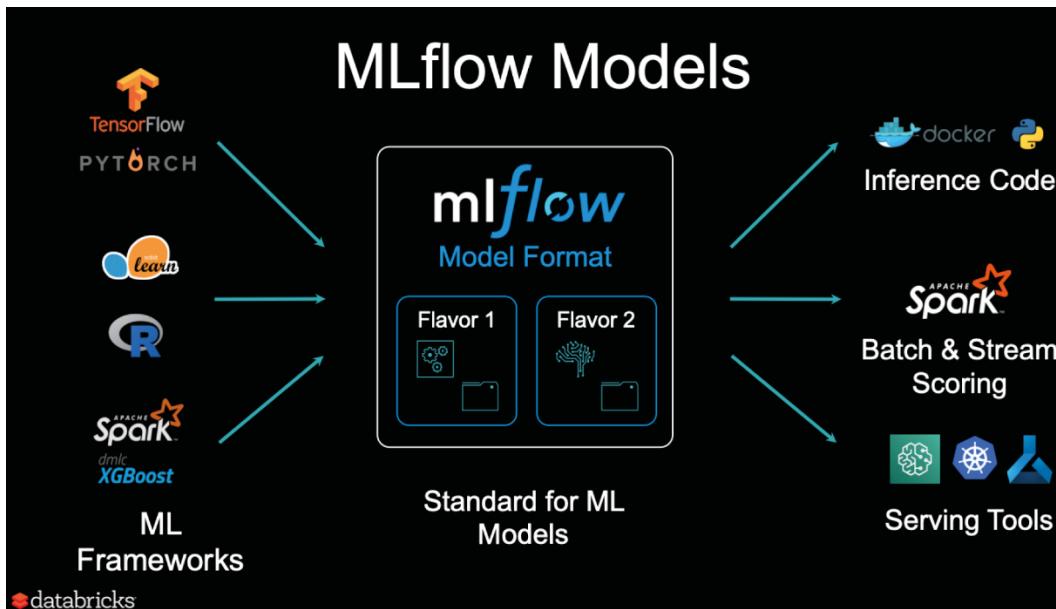
Metrics

training_accuracy_score	0.569	0.569	0.569	0.568	0.572
training_f1_score	0.535	0.535	0.535	0.533	0.532
training_log_loss	0.68	0.68	0.68	0.68	0.681

Scatter Plot Contour Plot Parallel Coordinates Plot



Chapter 5: Managing Models with MLflow



```
artifact_path: model
flavors:
  python_function:
    env: conda.yaml
    loader_module: mlflow.sklearn
    model_path: model.pkl
    python_version: 3.7.6
  sklearn:
    pickled_model: model.pkl
    serialization_format: cloudpickle
    sklearn_version: 0.22.2.post1
run_id: 75c2c826870d4d6082b3c6e10934a99f
signature:
  inputs: '[{"type": "double"}, {"type": "double"}]'
  outputs: '[{"type": "long"}]'
utc_time_created: '2021-02-22 15:47:40.557303'
```

```
mlflow.set_experiment("Baseline_Predictions_Mlflow_Check")
mlflow.tensorflow.autolog()

model = keras.Sequential([
    keras.layers.Dense(
        units=36,
        activation='relu',
        input_shape=(X_train.shape[-1],)
    ),
    keras.layers.BatchNormalization(),
    keras.layers.Dense(units=1, activation='sigmoid'),
])

model.compile(
    optimizer=keras.optimizers.Adam(lr=0.001),
    loss="binary_crossentropy",
    metrics="Accuracy"
)
with mlflow.start_run(run_name='keras_model_baseline') as run:
    model.fit(X_train, y_train, epochs=20, validation_split=0.05, shuffle=True, verbose=)
```

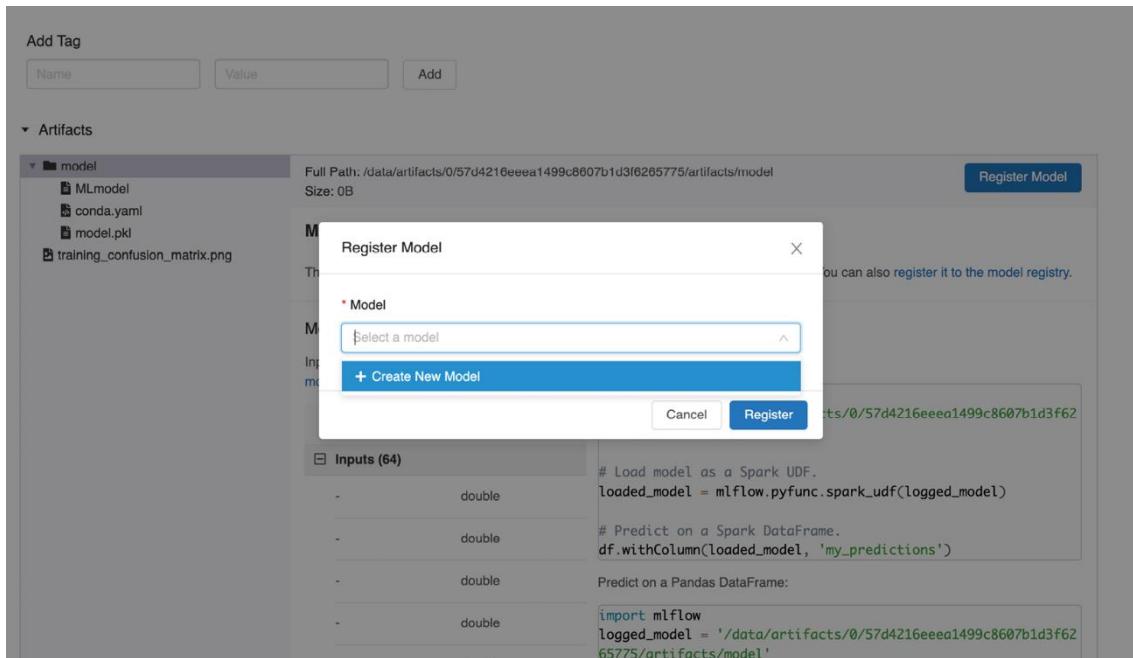
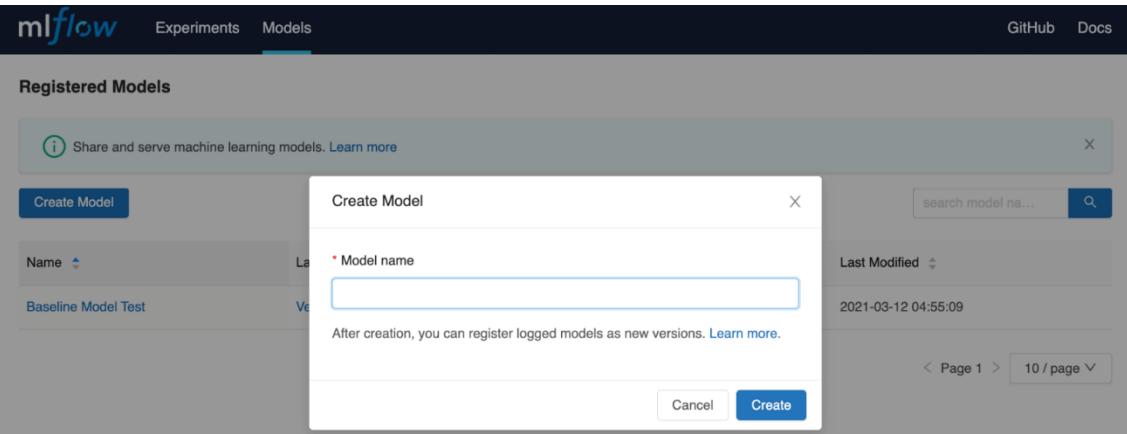
▼ Artifacts

```
model
  data
    keras_module.txt
    model.h5
    save_format.txt
  MLmodel
    conda.yaml
  tensorboard_logs
  model_summary.txt

Full Path: /data/artifacts/1/132e6fa332f2412d85f3cb9e6d6bc933/artifacts/model/ML...
Size: 334B
artifact_path: model
flavors:
  keras:
    data: data
    keras_module: tensorflow.keras
    keras_version: 2.4.0
    save_format: h5
  python_function:
    data: data
    env: conda.yaml
    loader_module: mlflow.keras
    python_version: 3.7.6
run_id: 132e6fa332f2412d85f3cb9e6d6bc933
utc_time_created: '2021-03-11 16:22:31.127098'
```

▼ Artifacts

model	Full Path: /data/artifacts/1/132e6fa332f2412d85f3cb9e6d6bc933/artifacts/model/con...
data	Size: 127B
keras_module.txt	
model.h5	
save_format.txt	
MLmodel	
conda.yaml	channels: - defaults - conda-forge dependencies: - python=3.7.6 - pip - pip: - mlflow - tensorflow==2.4.1 name: mlflow-env
tensorboard_logs	
model_summary.txt	



Search Runs: ? Filter Search Clear

Showing 11 matching runs Compare Delete Download CSV Columns

				Metrics <		
	Start Time	User	Models	training_accuracy	↓ training_f1_score	training_log_loss
<input type="checkbox"/>	2021-03-12 05:15:03	admin	sklearn	0.569	0.535	0.68
<input type="checkbox"/>	2021-03-12 05:14:42	admin	sklearn	0.569	0.535	0.68
<input type="checkbox"/>	2021-03-12 05:14:38	admin	sklearn	0.569	0.535	0.68
<input type="checkbox"/>	2021-03-12 05:14:59	admin	sklearn	0.568	0.534	0.68
<input type="checkbox"/>	2021-03-12 05:15:06	admin	sklearn	0.572	0.532	0.681
<input type="checkbox"/>	2021-03-12 05:14:45	admin	sklearn	0.568	0.531	0.68
<input type="checkbox"/>	2021-03-12 05:14:49	admin	sklearn	0.569	0.529	0.681
<input type="checkbox"/>	2021-03-12 05:14:52	admin	sklearn	0.568	0.528	0.681
<input type="checkbox"/>	2021-03-12 05:14:55	admin	sklearn	0.567	0.527	0.681
<input type="checkbox"/>	2021-03-12 05:14:34	admin	sklearn	0.565	0.517	0.681
<input type="checkbox"/>	2021-03-12 05:14:34	admin	-	-	-	-

▼ Artifacts

Full Path: /data/artifacts/2/719221d3ea7943458bef6b622dc66970/artifacts/model
Size: 0B

Register Model

Register Model

You can also register it to the model registry.

* Model: + Create New Model

* Model Name: BTC StockPrediction

Cancel Register

```
1           long
# Predict on a Spark DataFrame.
df.withColumn(loaded_model, 'my_predictions')
```

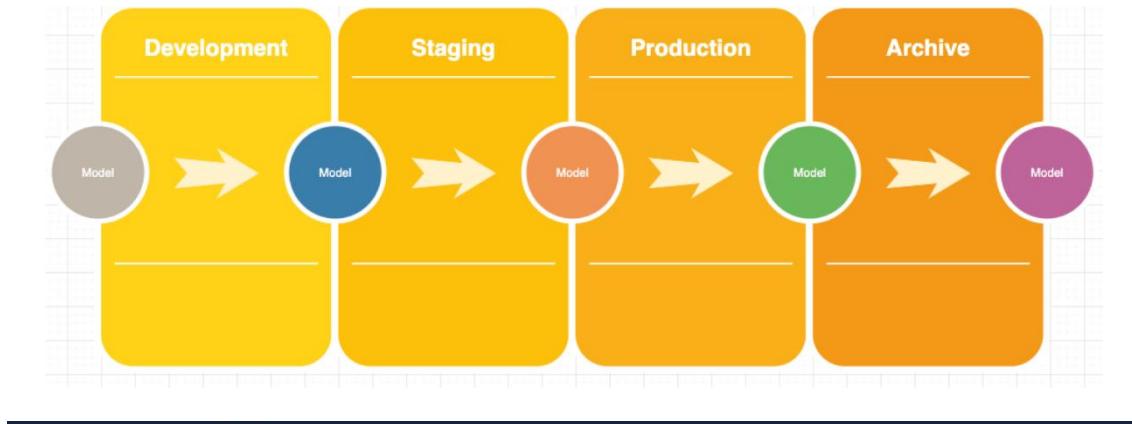
Registered Models

Share and serve machine learning models. [Learn more](#)

Create Model Search

Name	Latest Version	Staging	Production	Last Modified
Baseline Model Test	Version 2	-	-	2021-03-12 04:55:09
BTC StockPrediction	Version 1	-	-	2021-03-12 05:27:17

< Page 1 > 10 / page ▾



Registered Models > BTC StockPrediction > Version 1 ▾

Registered At: 2021-03-12 05:27:17

Creator:

Stage: None ▾

Last Modified: 2021-03-12 05:41:12

Source Run: Run 719221d3ea7943458bef6b622dc66970

Transition to → Staging
 Transition to → Production
 Transition to → Archived

▼ Description

First model to predict movement of bitcoin. Hyperparameter optimization was used and there is relative confidence to move the model to a Staging stage.

▼ Tags

Name	Value	Actions
Ticker	BTC	 

▼ Add Tag

<input type="text" value="Name"/>	<input type="text" value="Value"/>	<input type="button" value="Add"/>
-----------------------------------	------------------------------------	------------------------------------

▼ Schema

Name	Type
<input checked="" type="checkbox"/> Inputs (14)	
<input checked="" type="checkbox"/> Outputs (1)	

Version 1 ▾

Stage Transition

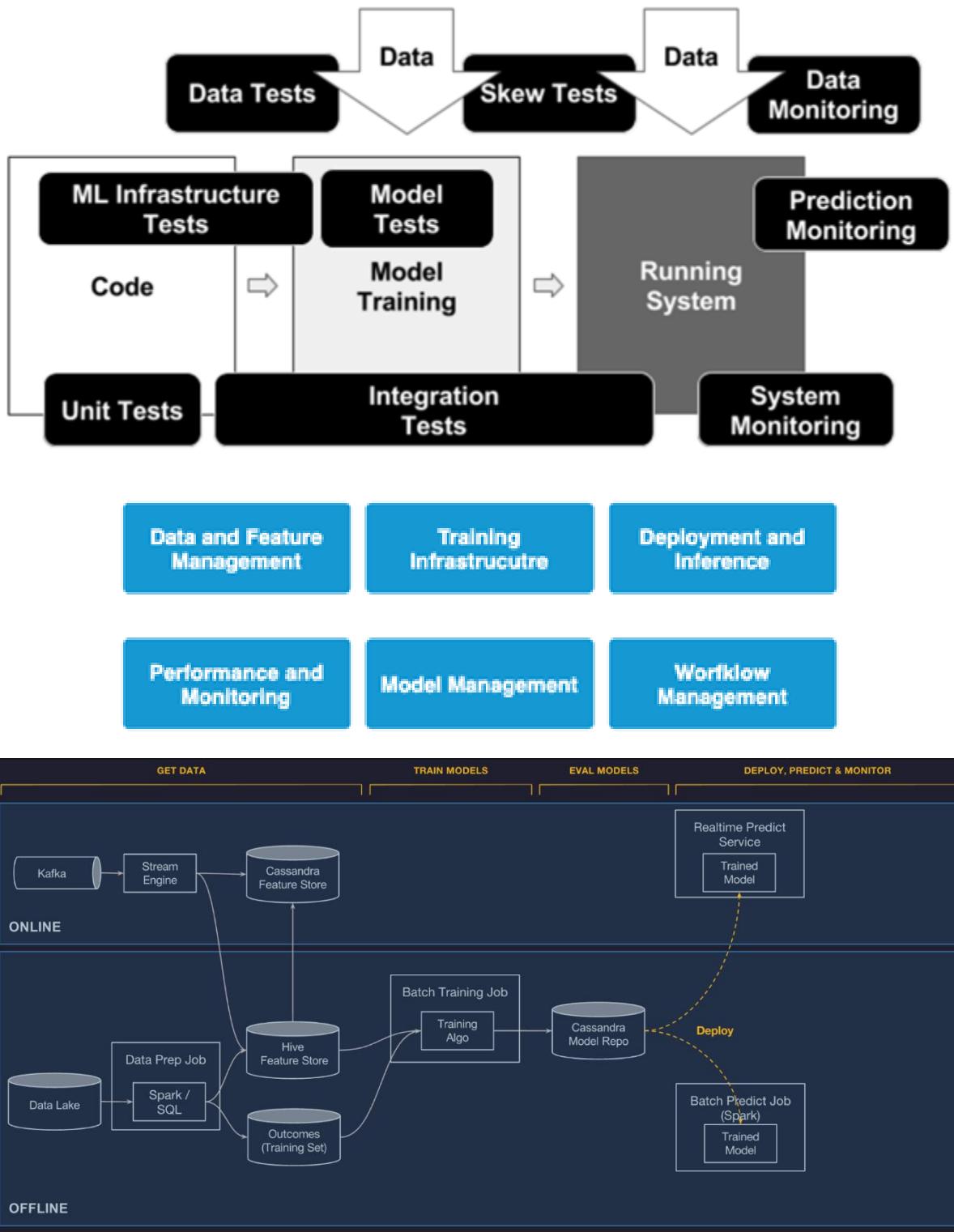
Stage: None ▾

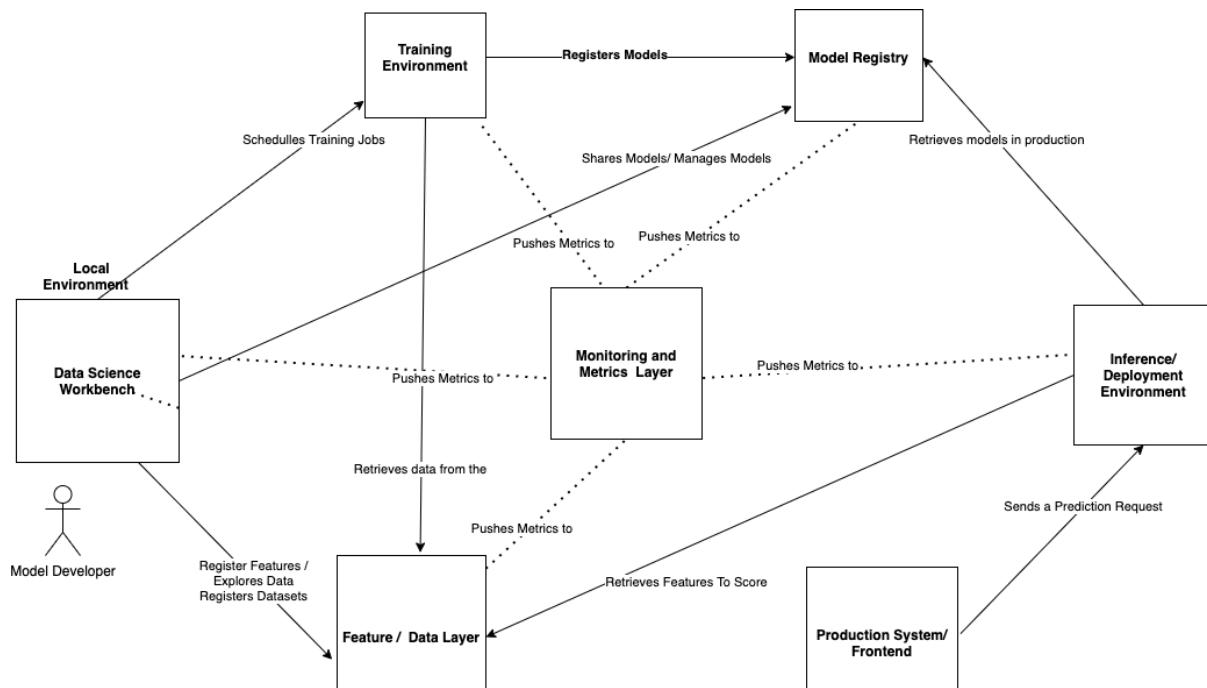
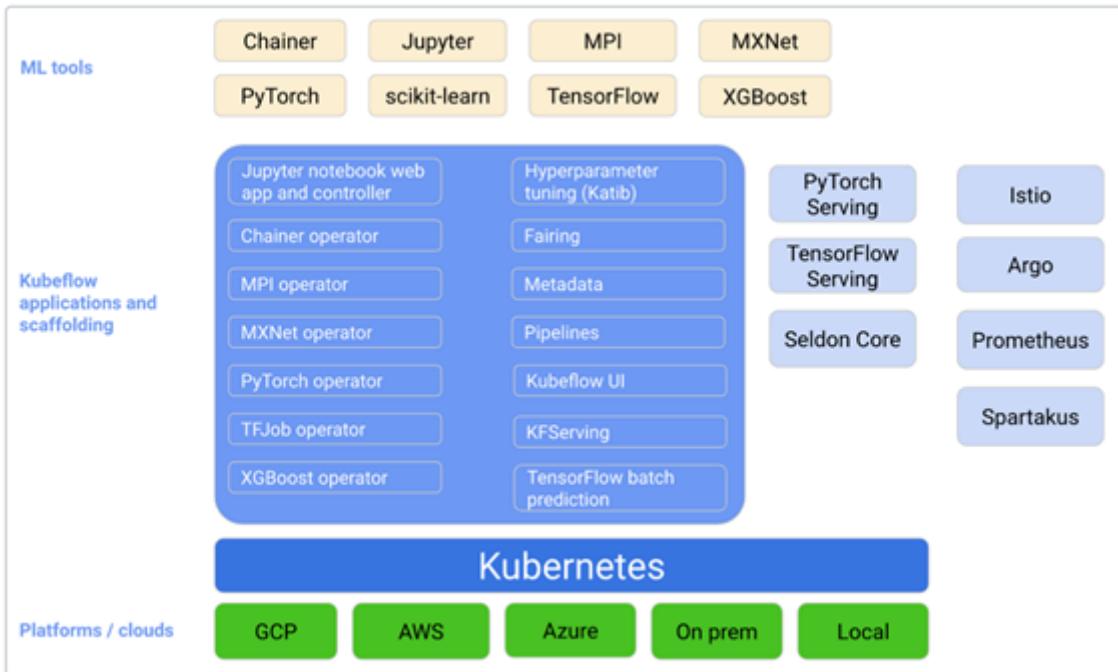
Transition to → Staging

Transition existing Staging model versions to Archived

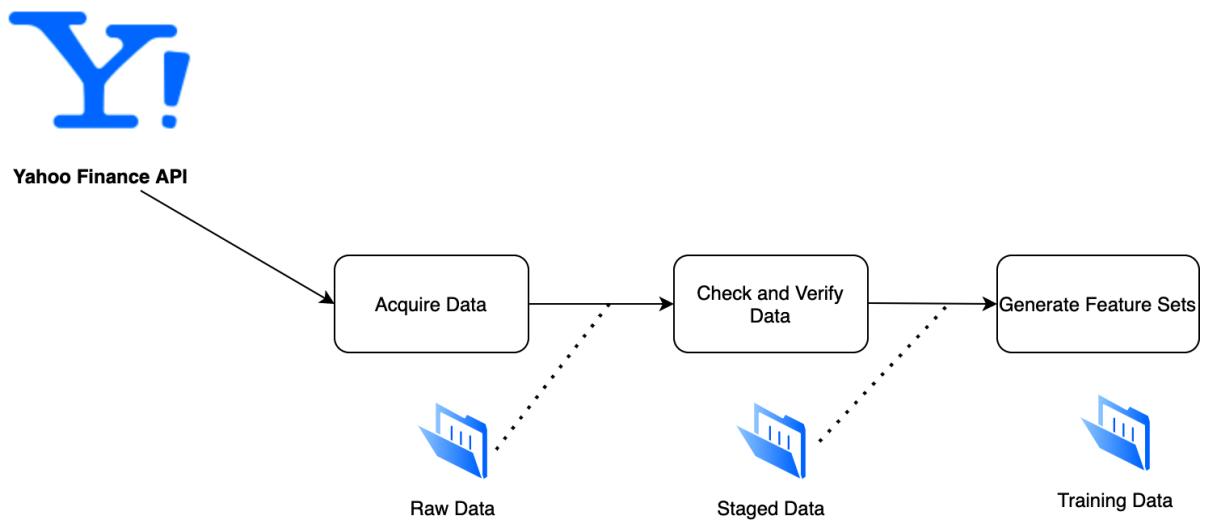
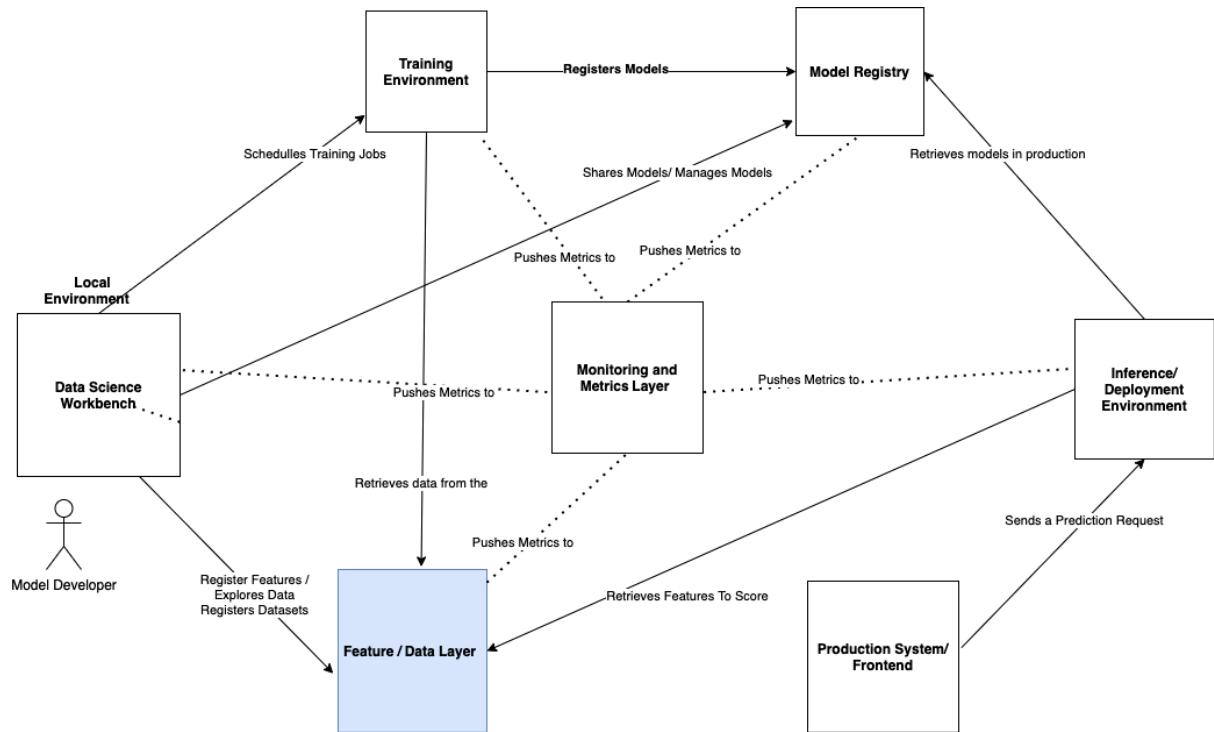
Hyperparameter optimization was used and there is relative confidence to move the model to a Staging stage.

Chapter 6: Introducing ML Systems Architecture





Chapter 7: Data and Feature Management



..	
data	Add folders
.gitignore	Fix gitignore
LICENSE	Add chapter 6 code
MLproject	Add chapter 6 code
README.md	Add chapter 6 code
check_verify_data.py	Add chapter 6 code
conda.yaml	Add chapter 6 code
feature_set_generation.py	Add chapter 6 code
load_raw_data.py	Add chapter 6 code
main.py	Add chapter 6 code

mlflow Experiments Models GitHub Docs

Experiments [+](#) [psystock_data_pipelines](#)

Search Experiments

Default [psystock_data_pipelines](#) [Edit](#) [Delete](#)

Experiment ID: 1

Artifact Location: file:///Users/admin/development/Machine-Learning-Engineering-with-Mlflow/chapter_7/psytock-data-features-main/mlruns/1

▼ Notes [Edit](#)

None

Search Runs: [?](#) [Filter](#) [Search](#) [Clear](#)

Showing 4 matching runs [Compare](#) [Delete](#) [Download CSV](#) [Columns](#)

	Start Time	Run Name	User	Source
<input type="checkbox"/>	2021-07-03 10:47:04	psytock-data-pip...	admin	psytock-data-features-main
<input type="checkbox"/>	2021-07-03 10:47:15	feature_set_gener...	admin	psytock-data-features-main
<input type="checkbox"/>	2021-07-03 10:47:11	check_verify_data	admin	psytock-data-features-main
<input type="checkbox"/>	2021-07-03 10:47:07	load_raw_data	admin	psytock-data-features-main

psystock_data_pipelines > load_raw_data

Date: 2021-07-03 10:47:07

Source: [psytock-data-features-main:load_raw_data](#) Git Commit: 55d140dfe8b9c09bf354d5328b15aee743d0d31b

Entry Point: load_raw_data

User: admin Duration: 3.7s

Status: FINISHED

Parent Run: [99d658fecab84723befcbda19dc1a5cb](#)

▼ Run Command

```
mlflow run file:///Users/admin/development/Machine-Learning-Engineering-with-Mlflow#chapter_7/psytock-data-features-main -v 55d140dfe8b9c09bf354d5328b15aee743d0d31b -e load_raw_data -b local
```

psystock_data_pipelines > check_verify_data -

Date: 2021-07-03 10:47:11	Source: psylock-data-features-main:clean_validate_data	Git Commit: 55d140dfe8b9c09bf354d5328b15aee743d0d31b
Entry Point: clean_validate_data	User: admin	Duration: 3.9s
Status: FINISHED	Parent Run: 99d658fecab84723befcbda19dc1a5cb	
<ul style="list-style-type: none"> ▶ Run Command ▶ Notes 🔗 ▶ Parameters ▶ Metrics ▶ Tags ▼ Artifacts 		

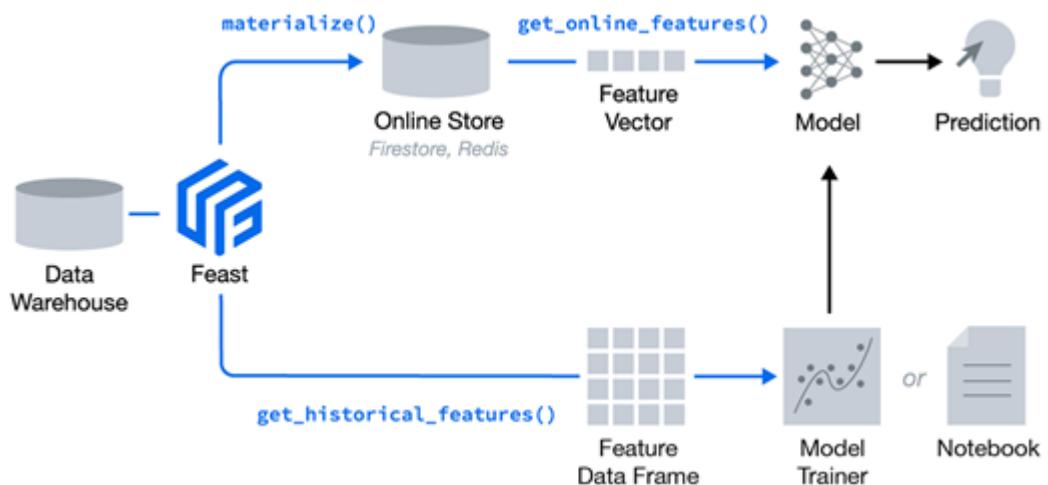
[describe_data.json](#)

Full Path: file:///Users/admin/development/Machine-Learning-Engineering-with-Mlflow/chapter_7/psystock-data-...
Size: 1.38KB

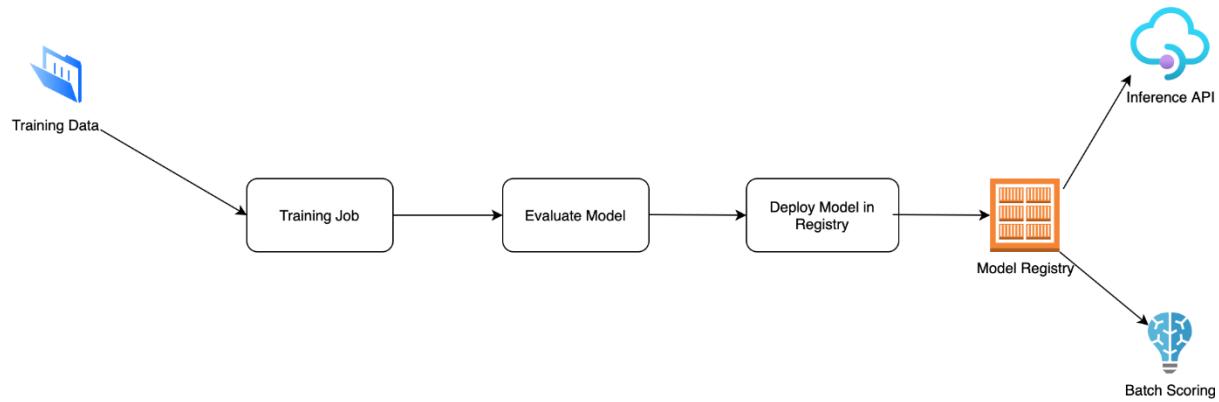
```
{
  "High": {
    "count": 92,
    "mean": 47448.515816066574,
    "std": 10467.57427637064,
    "min": 32637.587890625,
    "25%": 37590.8642578125,
    "50%": 44679.287109375,
    "75%": 58039.20703125,
    "max": 64863.09765625
  },
}
```

<input type="checkbox"/>	Start Time	Run Name	Source	Version
<input type="checkbox"/>	⌚ 2021-07-03 11:12:50	pystock-data-pipeline	psylock-data-features-	55d140
<input type="checkbox"/>	⌚ 2021-07-03 11:13:01	feature_set_generation	psylock-data-features-	55d140
<input type="checkbox"/>	⌚ 2021-07-03 11:12:56	check_verify_data	psylock-data-features-	55d140
<input type="checkbox"/>	⌚ 2021-07-03 11:12:53	load_raw_data	psylock-data-features-	55d140

Load more



Chapter 8: Training Models with MLflow



▼ Artifacts

MLmodel	Full Path: ./tmp/0/ed0a40810bbb44d883b6b7bb54776e12/artifacts/model	Register Model
conda.yaml	Size: 0B	
model.xgb		
feature_importance_weight.json		
feature_importance_weight.png		

MLflow Model

The code snippets below demonstrate how to make predictions using the logged model. You can also [register it to the model registry](#).

Model schema	Make Predictions												
Input and output schema for your model. Learn more	Predict on a Spark DataFrame: <pre>import mlflow logged_model = './tmp/0/ed0a40810bbb44d883b6b7bb54776e12/artifacts/model' # Load model as a Spark UDF. loaded_model = mlflow.pyfunc.spark_udf(logged_model) # Predict on a Spark DataFrame. df.withColumn(loaded_model, 'my_predictions')</pre>												
<table border="1"><thead><tr><th>Name</th><th>Type</th></tr></thead><tbody><tr><td colspan="2">Inputs (14)</td></tr><tr><td>-</td><td>long</td></tr><tr><td>1</td><td>long</td></tr><tr><td>2</td><td>long</td></tr><tr><td>3</td><td>long</td></tr></tbody></table>	Name	Type	Inputs (14)		-	long	1	long	2	long	3	long	Predict on a Pandas DataFrame: <pre>import mlflow logged_model = './tmp/0/ed0a40810bbb44d883b6b7bb54776e12/artifacts/model'</pre>
Name	Type												
Inputs (14)													
-	long												
1	long												
2	long												
3	long												

▼ Metrics

Name	Value
accuracy_score ↗	0.423
average_precision_score ↗	0.431
f1_score ↗	0.444
jaccard_score ↗	0.286
log_loss ↗	19.93
matthews_corrcoef ↗	-0.144
precision_score ↗	0.4
recall_score ↗	0.5
zero_one_loss ↗	0.577



Experiments Models

[GitHub](#) [Docs](#)

Registered Models



Share and serve machine learning models. [Learn more](#)

X

[Create Model](#)

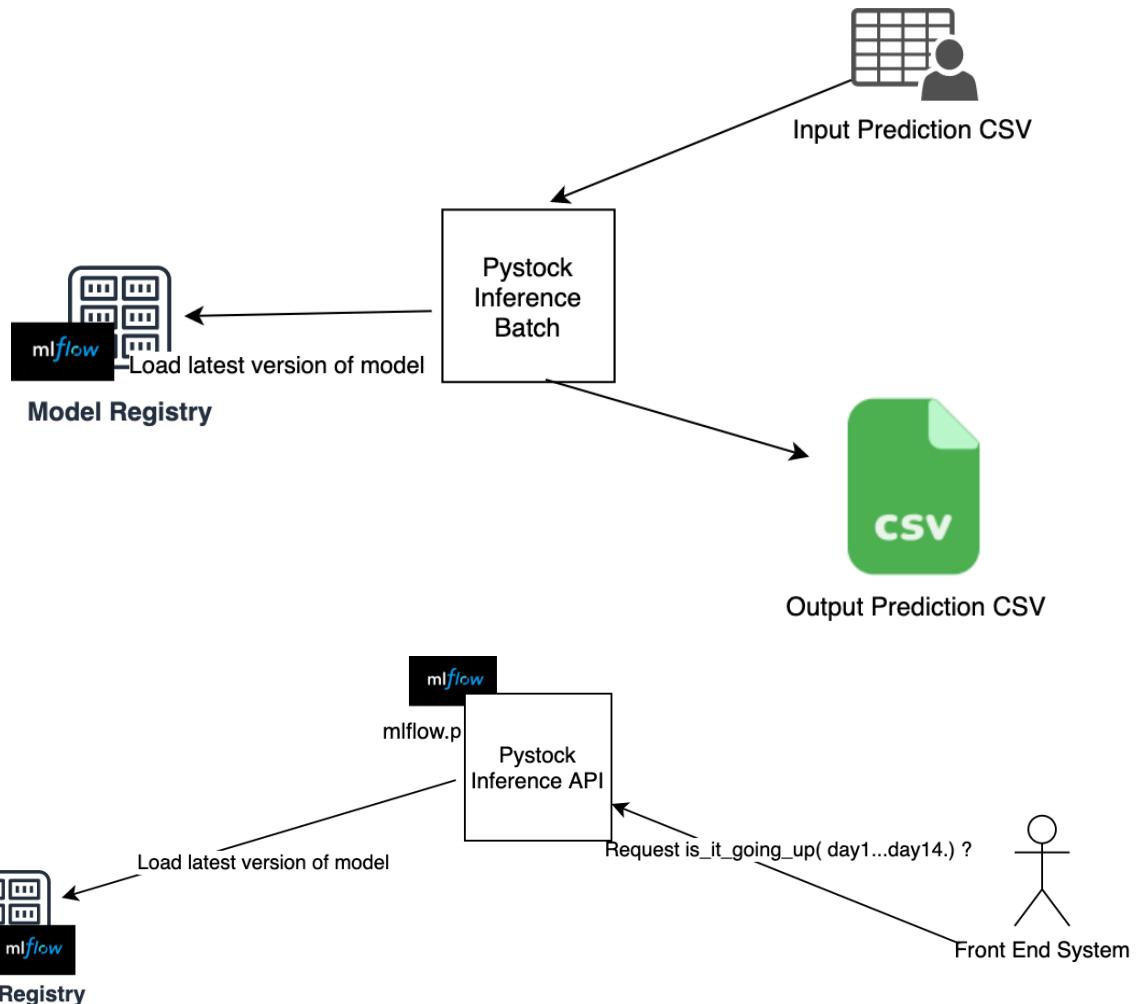
search model na...



Name	Latest Version	Staging	Production	Last Modified
training-model-psystock	Version 1	-	-	2021-07-05 19:49:22

◀ Page 1 ▶ 10 / page ▾

Chapter 9: Deployment and Inference with MLflow



Screenshot of the AWS ECR (Amazon Container Registry) console, showing a private repository named "mlflow-pyfunc".

The ECR interface includes:

- Services:** AWS Services dropdown.
- Search bar:** Search for services, features, marketplace products, and docs.
- User:** Natu Lauchande.
- Region:** Ireland.
- Support:** Support dropdown.
- Amazon Container Services:** Left sidebar.
- ECR > Repositories:** Current view.
- Private:** Repository type.
- Public:** Repository type.
- Private repositories (1):** mlflow-pyfunc.
- Actions:** View push commands, Delete, Edit, Create repository.
- Filters:** Find repositories.
- Table Headers:** Repository name, URI, Created at, Tag immutability, Scan on push, Encryption type.
- Table Data:** mlflow-pyfunc, 492299092739.dkr.ecr.eu-west-1.amazonaws.com/mlflow-pyfunc, 29 May 2021 17:52:24, Disabled, Disabled, AES-256.

Monitor

Access CloudWatch logs to view your Jupyter notebook's debugging and progress reporting. [Learn more](#)

[View invocation metrics](#)  [View instance metrics](#) 

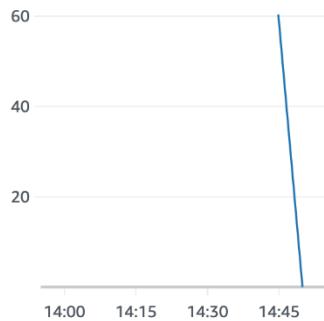
[View logs](#) 

[Add to dashboard](#)

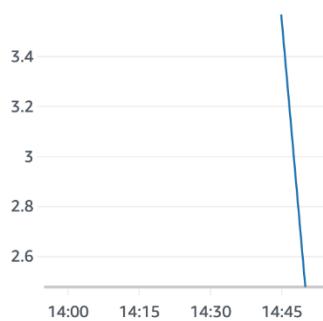
 3h 12h 1d 3d 1w



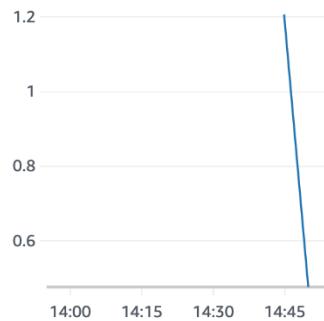
CPUUtilization



MemoryUtilization

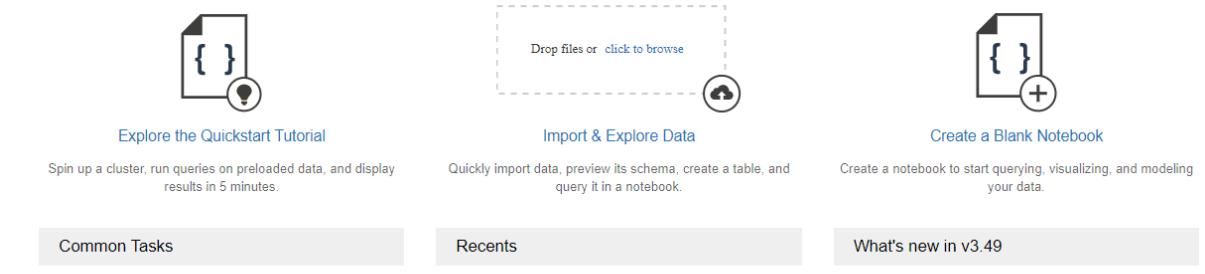


DiskUtilization



Chapter 10: Scaling Up Your Machine Learning Workflow

Welcome to databricks



Explore the Quickstart Tutorial
Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.

Import & Explore Data
Quickly import data, preview its schema, create a table, and query it in a notebook.

Create a Blank Notebook
Create a notebook to start querying, visualizing, and modeling your data.

Common Tasks

- New Notebook
- Create Table
- New Cluster
- New Job
- New MLflow Experiment
- Import Library

Recents
Recent files appear here as you work.

What's new in v3.49
Databricks Status
View latest release notes

Create New Table

Data source [?](#)

[Upload File](#) [S3](#) [DBFS](#) [Other Data Sources](#) [Partner Integrations](#)

DBFS Target Directory [?](#)
 [Select](#)

Files uploaded to DBFS are accessible by everyone who has access to this workspace. [Learn more](#)

Files [?](#)

 input_prediction 2.2 KB Remove file

✓ File uploaded to /FileStore/tables/input_prediction-5.csv

[Create Table with UI](#) [Create Table in Notebook](#) [?](#)

Create Cluster

New Cluster | [Cancel](#) | [Create Cluster](#) | **0 Workers**:0 GB Memory, 0 Cores, 0 DBU | **1 Driver**:15.3 GB Memory, 2 Cores, 1 DBU

[Cluster Name](#) [UI](#) | [JSON](#)
bitpred_poc_cluster

[Databricks Runtime Version](#) [Runtime: 8.3 \(Scala 2.12, Spark 3.1.1\)](#)

[Note](#) Databricks Runtime 8.x uses Delta Lake as the default table format. [Learn more](#)

[Instance](#)
Free 15GB Memory: As a Community Edition user, your cluster will automatically terminate after an idle period of two hours. For [more configuration options](#), please [upgrade your Databricks subscription](#).

[Instances](#) [Spark](#)

[Availability Zone](#) [auto](#)

Welcome

Create Notebook

Name

Default Language [Python](#)

Cluster [My Cluster](#)

[Cancel](#) | [Create](#)

Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.

Drop files or [click to browse](#)

Experiment Revision history

bitpred_poc

Experiment Runs Date ↴ ⏪ ⏪

2021-05-11 21:01:09 SAST ⏪ ⏪

⊕ C: 1.0, class_weight: None, dual: ...
⊕ training_accuracy_score: 0.7, ...

Models  sklearn

bitpred_poc (Python)

Detached Cmd 1

```
1 import pandas
2 import numpy
3 import mlflow
4 from sklearn import
5 from sklearn import
6 from sklearn
```

Command took 3.54s

Cmd 2

```
1 df = (spark
2 pandas_df = df.toPandas()
3 X=pandas_df.iloc[:, :-1]
4 Y=pandas_df.iloc[:, -1]
5 X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.33, random_state=4284, stratify=Y)
```

▶ (2) Spark Jobs

▶ df: pyspark.sql.dataframe.DataFrame = [10: string, 11: string ... 13 more fields]

Command took 3.56 seconds -- by plauchande@gmail.com at 11/05/2021 20:55:55 on bitpred_poc cluster

Notebook Published

The notebook was published successfully. Please copy the url and save it (it may take a minute or two for your updates to be publicly available). The link will remain valid for 6 months.

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e2>

Done

Chapter 11: Performance Monitoring

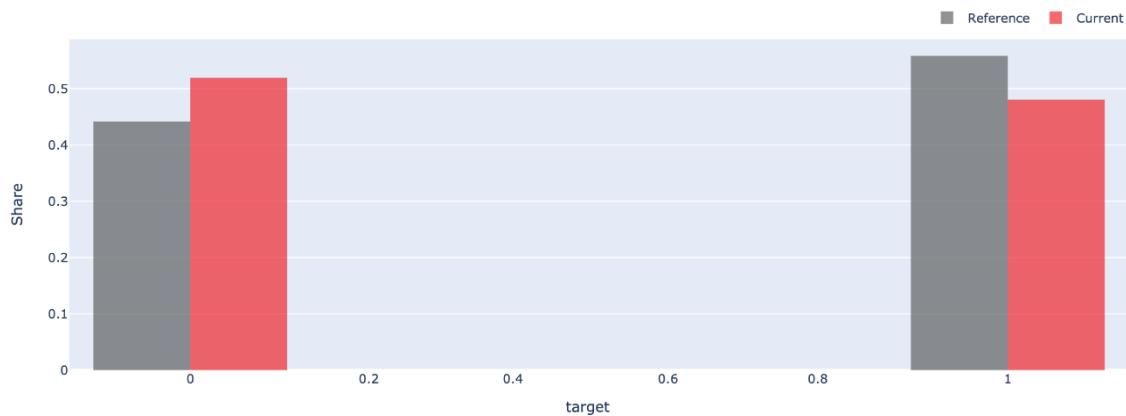
▼ Artifacts

input_data_drift.html
input_data_drift.json
Full Path: file:///Users/admin/development/Machine-Learning-Engineering-with-Mlflow...
Size: 9.46MB [Download](#)

Feature	Type	Reference Distribution	Current Distribution	Data drift	P-Value for Similarity Test ↑	
day9	num				Not Detected	1

DATA DRIFT **DATA DISTRIBUTION**

Target Drift: not detected, p_value=0.171135



Target (Prediction) Behavior By Feature

Feature

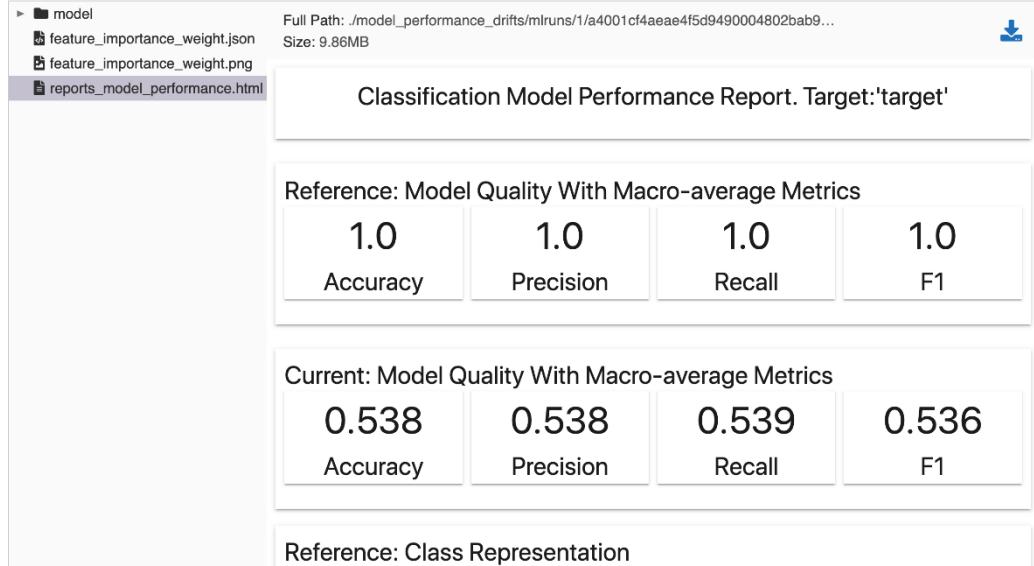
> day3

Search [X](#)

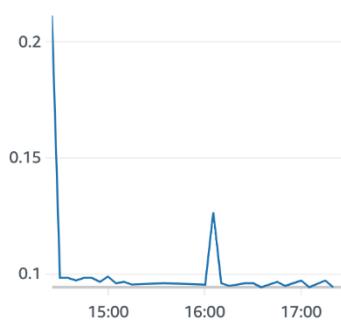
▼ Artifacts



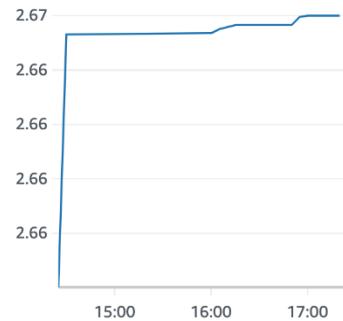
▼ Artifacts



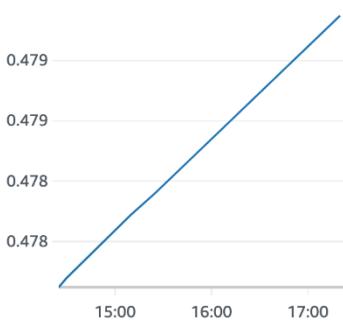
CPUUtilization



MemoryUtilization



DiskUtilization





All metrics Graphed metrics (6) Graph options Source

Ireland All > /aws/sagemaker/Endpoints > EndpointName, VariantName Graph search

EndpointName (6)	VariantName	Metric Name
pystock-api	pystock-api-model-o3g0aafabqlgdtuwys2buw	CPUUtilization
pystock-api	pystock-api-model-o3g0aafabqlgdtuwys2buw	MemoryUtilization
pystock-api	pystock-api-model-o3g0aafabqlgdtuwys2buw	DiskUtilization
pystock-api	pystock-api-model-pf20mijdriqe6yqpz0uakw	CPUUtilization
pystock-api	pystock-api-model-pf20mijdriqe6yqpz0uakw	MemoryUtilization
pystock-api	pystock-api-model-pf20mijdriqe6yqpz0uakw	DiskUtilization

aws Services ▾ [Option+S]

Step 1 Specify metric and conditions

Step 2 Configure actions

Step 3 Add name and description

Step 4 Preview and create

Specify metric and conditions

Metric

Graph
This alarm will trigger when the blue line goes above the red line for 1 datapoints within 5 minutes.

Percent

0.25
0.202 (red line)
0.15
0.10
0.05
0.00

07:00 07:05 15:17 16:00 17:00

2021-07-05 15:20 UTC

1. CPUUtilization 0.0968228

Namespace /aws/sagemaker/Endpoints

Metric name

EndpointName

VariantName

Statistic

Period

Chapter 12: Advanced Topics with MLflow

Create Cluster

New Cluster

Cancel Create Cluster

0 Workers:0 GB Memory, 0 Cores, 0 DBU
1 Driver:15.3 GB Memory, 2 Cores, 1 DBU

Cluster Name: bitpred_poc_cluster

Databricks Runtime Version: Runtime: 8.4 Beta (Scala 2.12, Spark 3.1.2)

Note: Databricks Runtime 8.x uses Delta Lake as the default table format. [Learn more](#)

Instance: Free 15GB Memory: As a Community Edition user, your cluster will automatically terminate after an idle period of two hours. For more configuration options, please [upgrade your Databricks subscription](#).

Availability Zone: us-west-2c

Welcome to  databricks

Explore the Quickstart Tutorial

Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.

Import & Explore Data

Drop files or click to browse

Create a Blank Notebook

Common Tasks

New Notebook

Create Table

New Cluster

New Job

New MLflow Experiment

Import Library

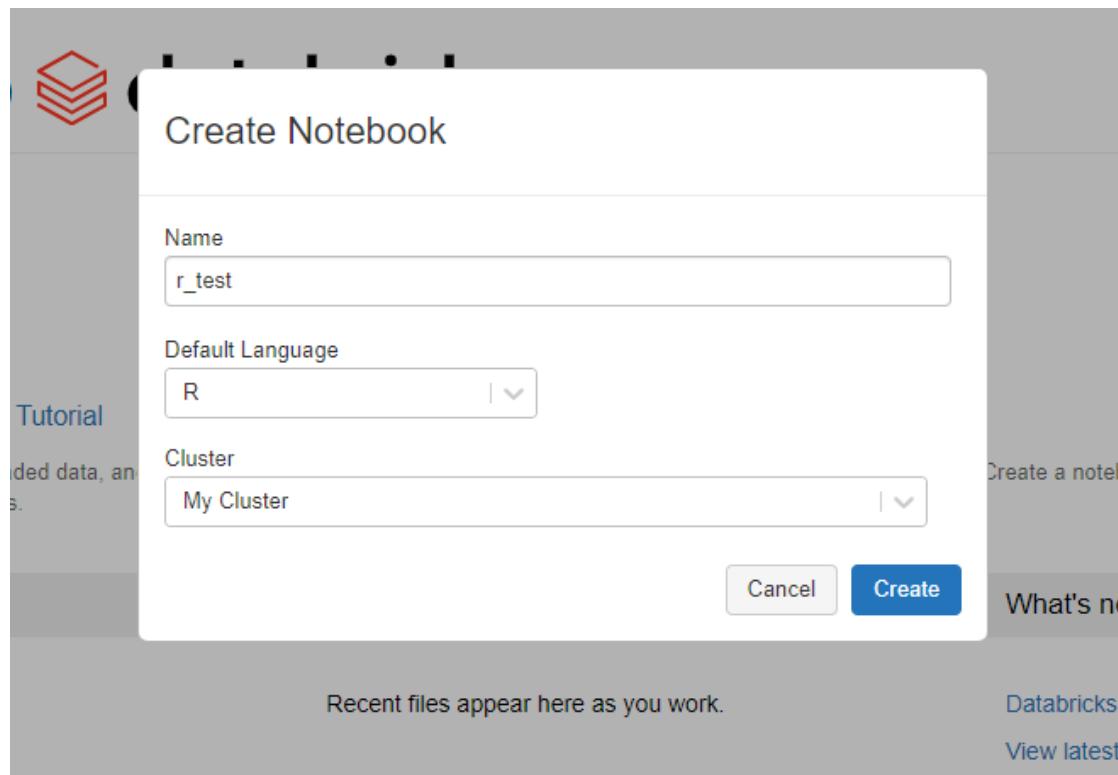
Recents

Recent files appear here as you work.

What's new in v3.49

Databricks Status

View latest release notes



Model		Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
dt	Decision Tree Classifier	0.6133	0.6083	0.6333	0.6683	0.6212	0.2126	0.2314	0.0050
lr	Logistic Regression	0.5733	0.6167	0.6667	0.6367	0.6281	0.1186	0.1474	0.3750
nb	Naive Bayes	0.5267	0.5667	0.6000	0.5667	0.5752	0.0270	0.0391	0.0060
lightgbm	Light Gradient Boosting Machine	0.5233	0.5778	0.6000	0.6117	0.5695	0.0186	0.0507	0.0410
knn	K Neighbors Classifier	0.5200	0.4556	0.7000	0.5367	0.6024	-0.0226	-0.0145	0.0080
ada	Ada Boost Classifier	0.5167	0.6111	0.5333	0.5683	0.5055	0.0538	0.0789	0.0330
xgboost	Extreme Gradient Boosting	0.5033	0.5278	0.5667	0.5333	0.5419	-0.0201	-0.0255	0.0910
ridge	Ridge Classifier	0.4967	0.0000	0.5333	0.5533	0.4948	-0.0008	0.0176	0.0050
qda	Quadratic Discriminant Analysis	0.4967	0.4722	0.6000	0.5250	0.5450	-0.0285	-0.0522	0.0060
lda	Linear Discriminant Analysis	0.4967	0.6167	0.5333	0.5533	0.4948	-0.0008	0.0176	0.0060
catboost	CatBoost Classifier	0.4500	0.3667	0.5333	0.4750	0.4943	-0.1054	-0.1242	0.7460
gbc	Gradient Boosting Classifier	0.4467	0.4944	0.5667	0.4500	0.5000	-0.1321	-0.1537	0.0240
rf	Random Forest Classifier	0.4333	0.3917	0.5333	0.4417	0.4640	-0.1281	-0.1614	0.0880
svm	SVM - Linear Kernel	0.4300	0.0000	0.4667	0.4200	0.4038	-0.1594	-0.1742	0.0050
et	Extra Trees Classifier	0.4300	0.4389	0.5333	0.4533	0.4748	-0.1767	-0.2021	0.0770

Experiment ID: 1

Artifact Location:
file:///Users/admin/development/Machine-Learning-Engineering-with-Mlflow/chapter_12/automl_pycaret/mlruns/1▼ Notes [🔗](#)

None

Search Runs: [?](#) [Filter](#) [Search](#) [Clear](#)Showing 33 matching runs [Compare](#) [Delete](#) [Download CSV](#)[☰](#) [grid](#) [Columns](#)

	Run Name	Parameters >			Metrics >			Tags	
		C	CPU Jobs	Categorical Feature	AUC	Accuracy	F1	USI	Run Time
<input type="checkbox"/>	CatBoost Classifier	-	-	-	0.372	0.38	0.421	796d	7.76
<input type="checkbox"/>	Random Forest Classifier	-	-	-	0.361	0.38	0.396	796d	1.3
<input type="checkbox"/>	K Neighbors Classifier	-	-	-	0.406	0.4	0.411	796d	0.25
<input type="checkbox"/>	Decision Tree Classifier	-	-	-	0.408	0.427	0.452	796d	0.22
<input type="checkbox"/>	Gradient Boosting Classifier	-	-	-	0.383	0.43	0.438	796d	0.39
<input type="checkbox"/>	Light Gradient Boosting M...	-	-	-	0.439	0.433	0.408	796d	0.59
<input type="checkbox"/>	Extra Trees Classifier	-	-	-	0.386	0.437	0.469	796d	0.86
<input type="checkbox"/>	Ada Boost Classifier	-	-	-	0.506	0.453	0.458	796d	0.49
<input type="checkbox"/>	Linear Discriminant Analy...	-	-	-	0.406	0.47	0.428	796d	0.2
<input type="checkbox"/>	Quadratic Discriminant An...	-	-	-	0.456	0.477	0.578	796d	0.21
<input type="checkbox"/>	Naive Bayes	-	-	-	0.456	0.477	0.485	796d	0.19
<input type="checkbox"/>	Extreme Gradient Boosting	-	-	-	0.467	0.49	0.48	796d	1.08
<input type="checkbox"/>	Ridge Classifier	-	-	-	0	0.49	0.468	796d	0.18
<input type="checkbox"/>	Logistic Regression	1.0	-	-	0.422	0.49	0.468	796d	4.88

V9	...	V20	V21	V22	V23	V24	V25	V26	V27	V28	Amount
0.363787	...	0.251412	-0.018307	0.277838	-0.110474	0.066928	0.128539	-0.189115	0.133558	-0.021053	149.62
-0.255425	...	-0.069083	-0.225775	-0.638672	0.101288	-0.339846	0.167170	0.125895	-0.008983	0.014724	2.69
-1.514654	...	0.524980	0.247998	0.771679	0.909412	-0.689281	-0.327642	-0.139097	-0.055353	-0.059752	378.66
-1.387024	...	-0.208038	-0.108300	0.005274	-0.190321	-1.175575	0.647376	-0.221929	0.062723	0.061458	123.50
0.817739	...	0.408542	-0.009431	0.798278	-0.137458	0.141267	-0.206010	0.502292	0.219422	0.215153	69.99
...
1.914428	...	1.475829	0.213454	0.111864	1.014480	-0.509348	1.436807	0.250034	0.943651	0.823731	0.77
0.584800	...	0.059616	0.214205	0.924384	0.012463	-1.016226	-0.606624	-0.395255	0.068472	-0.053527	24.79
0.432454	...	0.001396	0.232045	0.578229	-0.037501	0.640134	0.265745	-0.087371	0.004455	-0.026561	67.88
0.392087	...	0.127434	0.265245	0.800049	-0.163298	0.123205	-0.569159	0.546668	0.108821	0.104533	10.00
0.486180	...	0.382948	0.261057	0.643078	0.376777	0.008797	-0.473649	-0.818267	-0.002415	0.013649	217.00

ID	Name	Reference
abod	Angle-base Outlier Detection	pyod.models.abod.ABOD
cluster	Clustering-Based Local Outlier	pyod.models.cblob.CBLOF
cof	Connectivity-Based Local Outlier	pyod.models.cof.COF
iforest	Isolation Forest	pyod.models.iforest.IForest
histogram	Histogram-based Outlier Detection	pyod.models.hbos.HBOS
knn	K-Nearest Neighbors Detector	pyod.models.knn.KNN
lof	Local Outlier Factor	pyod.models.lof.LOF
svm	One-class SVM detector	pyod.models.ocsvm.OCSVM
pca	Principal Component Analysis	pyod.models.pca.PCA
mcd	Minimum Covariance Determinant	pyod.models.mcd.MCD
sod	Subspace Outlier Detection	pyod.models.sod.SOD
sos	Stochastic Outlier Selection	pyod.models.sos.SOS

psystock_anomaly > Isolation Forest ▾

Date : 2021-06-01 22:25:37

Source : ipykernel_launcher.py

Duration : 166ms

Status : FINISHED

▼ Notes ↗

None

▼ Parameters

Name	Value
behaviour	new
bootstrap	False
contamination	0.1
max_features	1.0
max_samples	auto
n_estimators	100
n_jobs	-1
random_state	8155
verbose	0

Experiment ID: 2101485559179462

Artifact Location: dbfs:/databricks/mlflow-tracking/2101485559179462

▶ Notes 

Showing 1 matching run

⟳ Refresh Compare Delete Download CSV 

☰ Columns 🔍 Search: metrics.rmse < 1 and params.model = "tree" Filter Clear

				Metrics		Parameters		
	Start Time	Run Name	Source	Models	sensitivity	specificity	mtry	ntree
<input type="checkbox"/>	 1 minute ago	-	 RTest	 crate	0.825	0.569	3	100

Load more