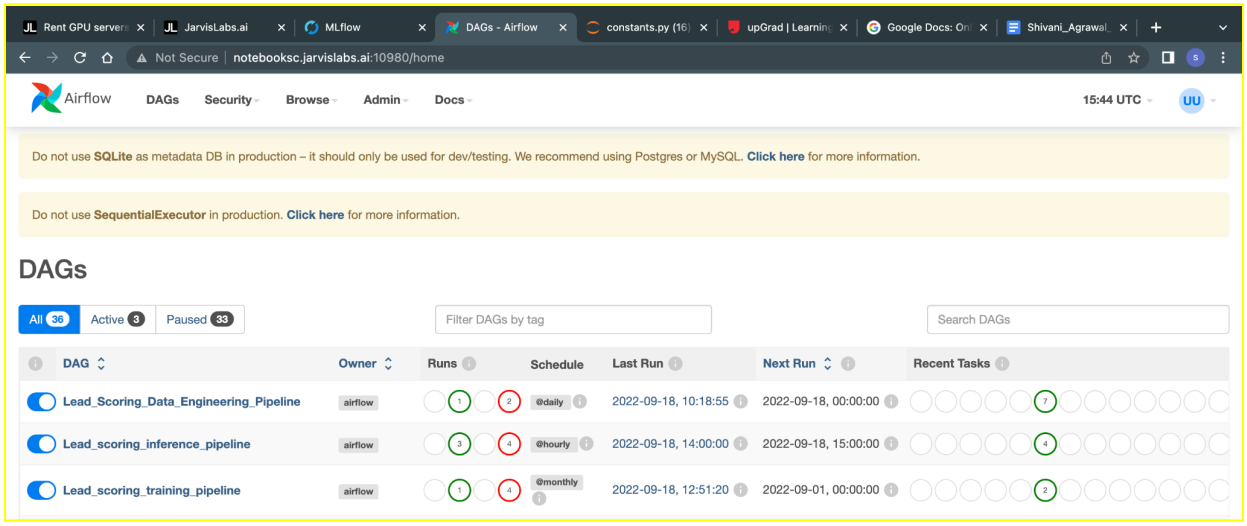
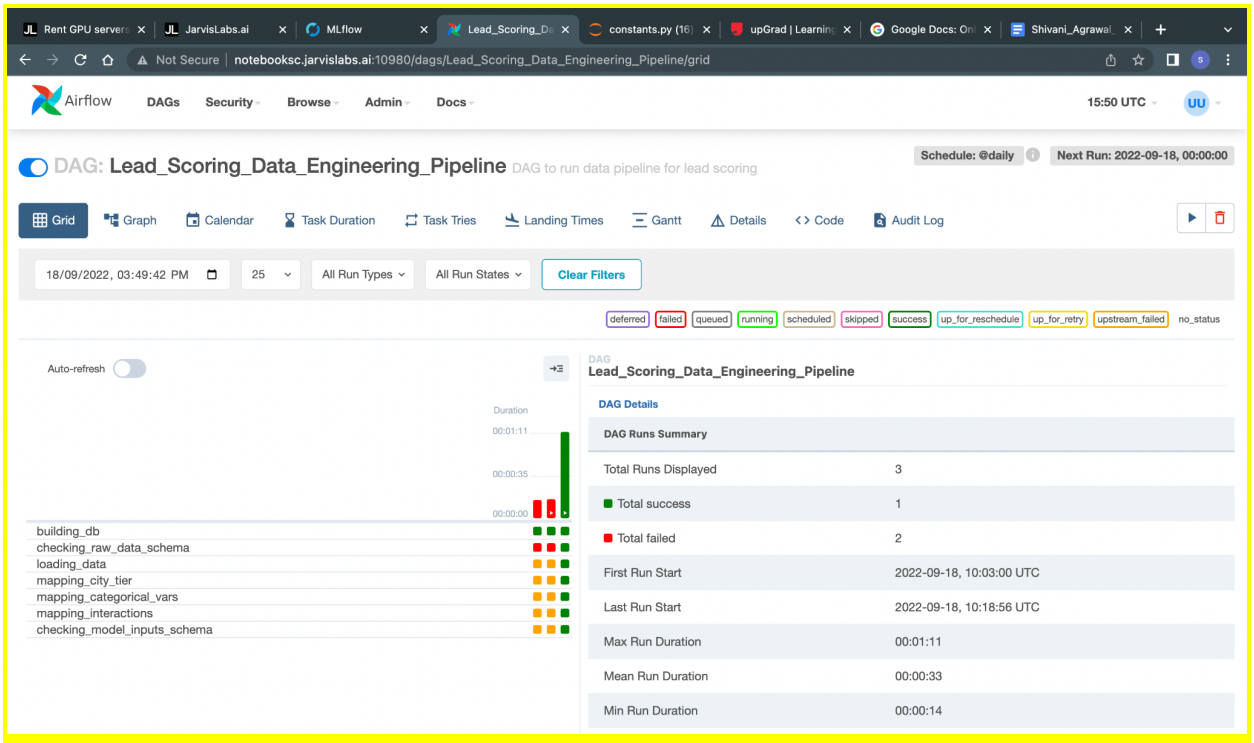
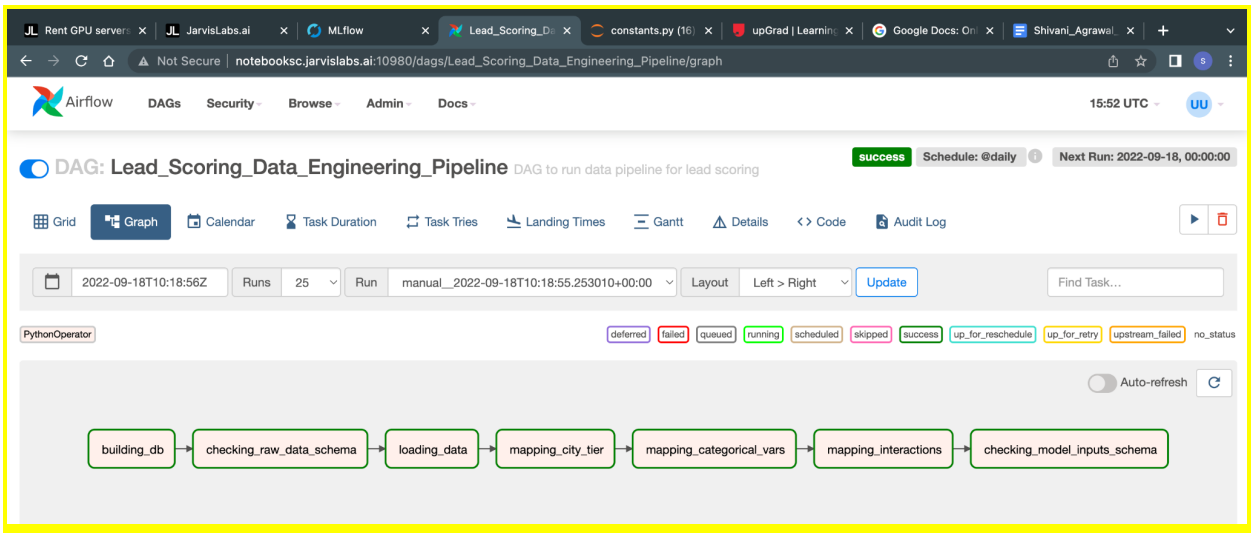
Airflow DAGs page:

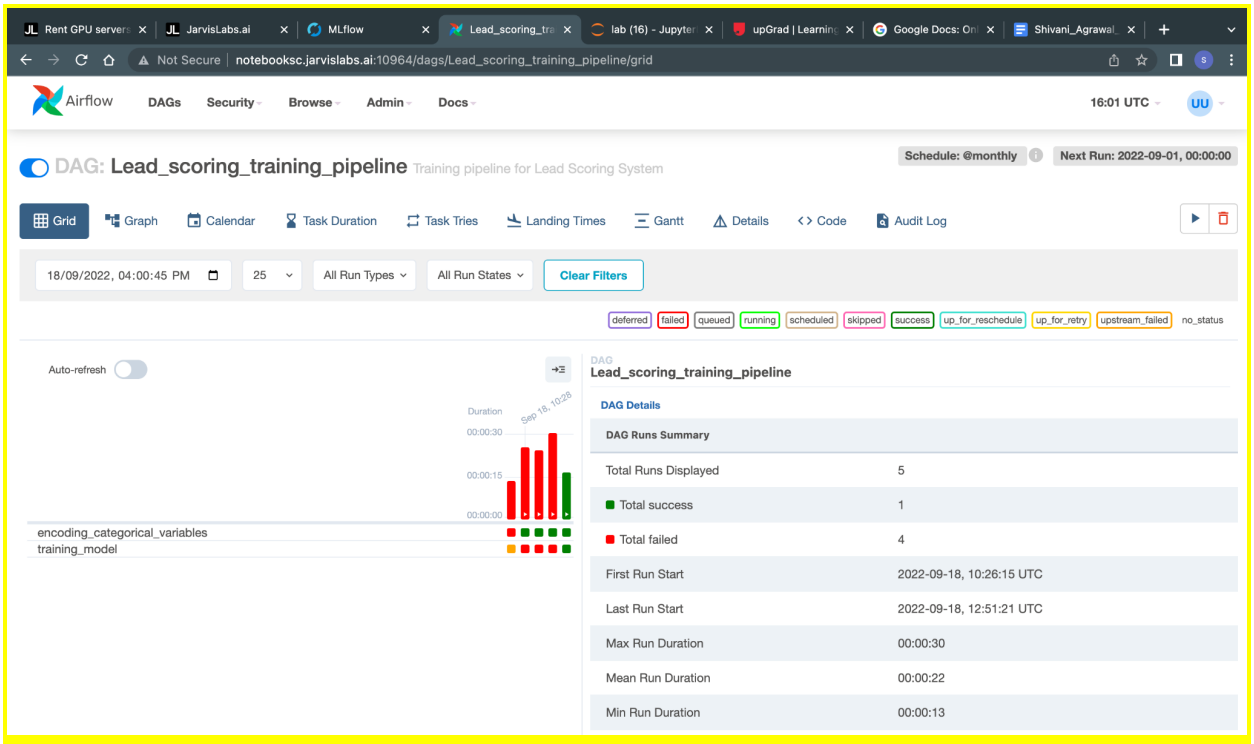


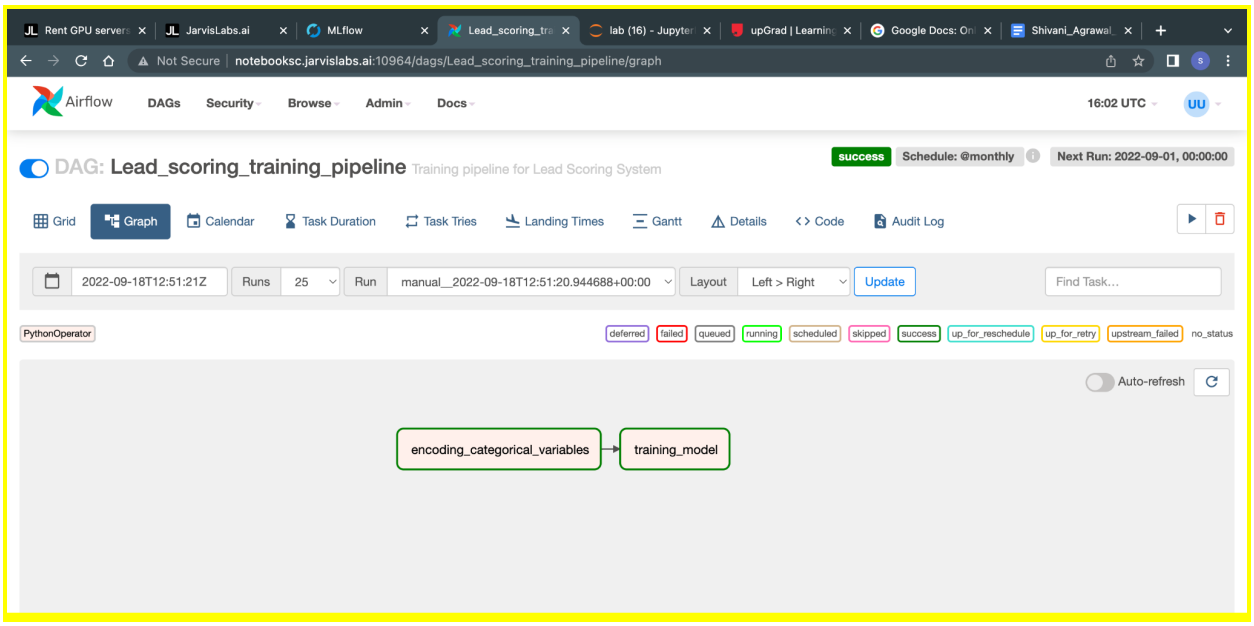
Lead_Scoring_Data_Engineering_Pipeline DAG:



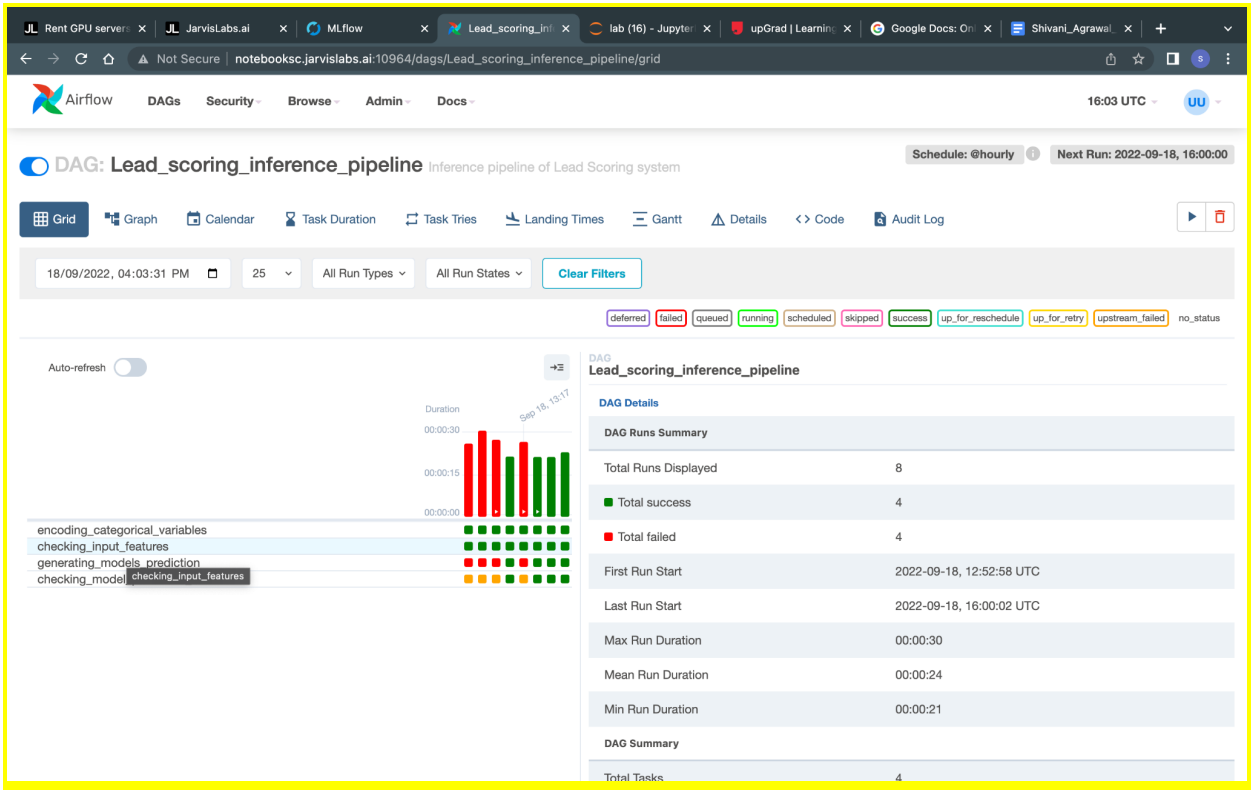


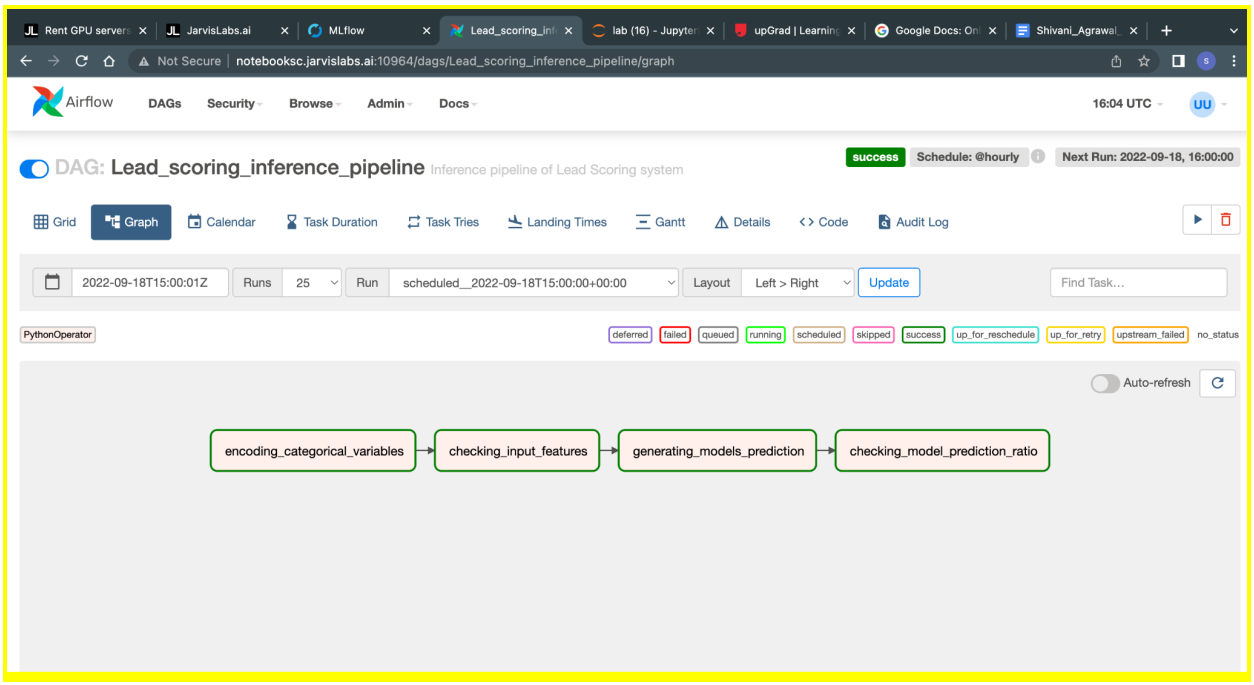
Lead_scoring_training_pipeline:



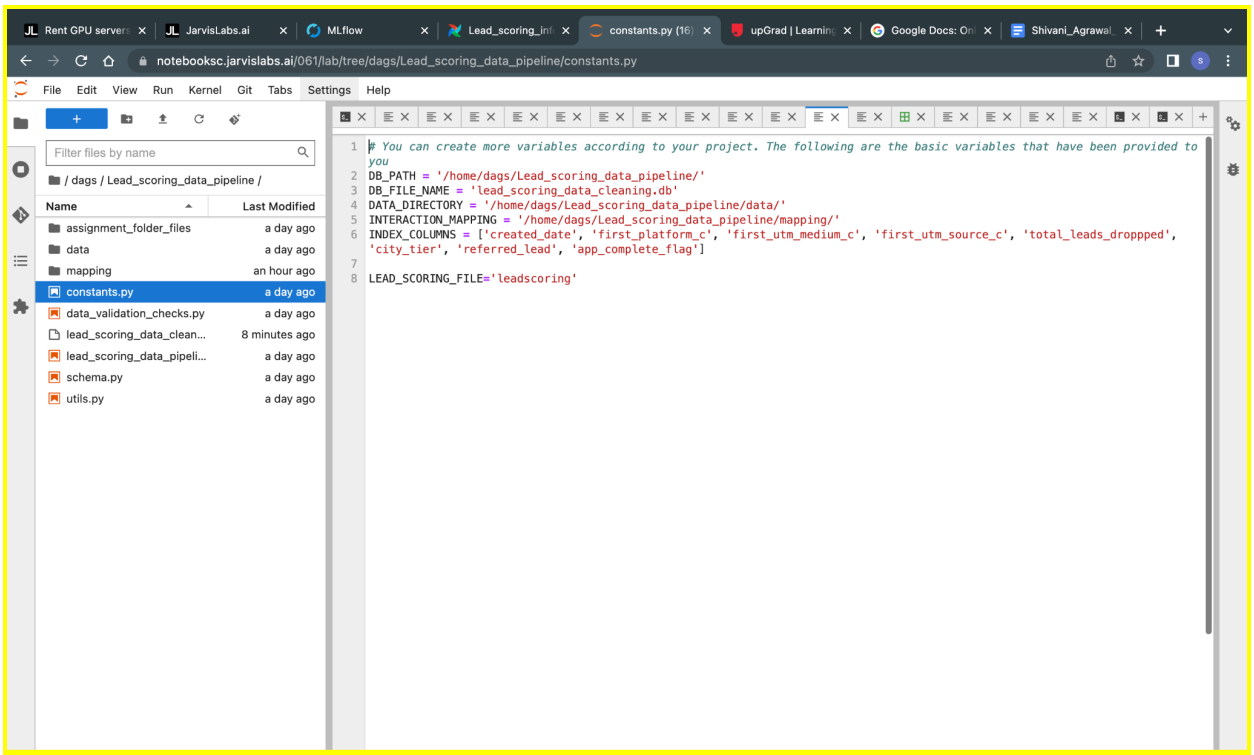


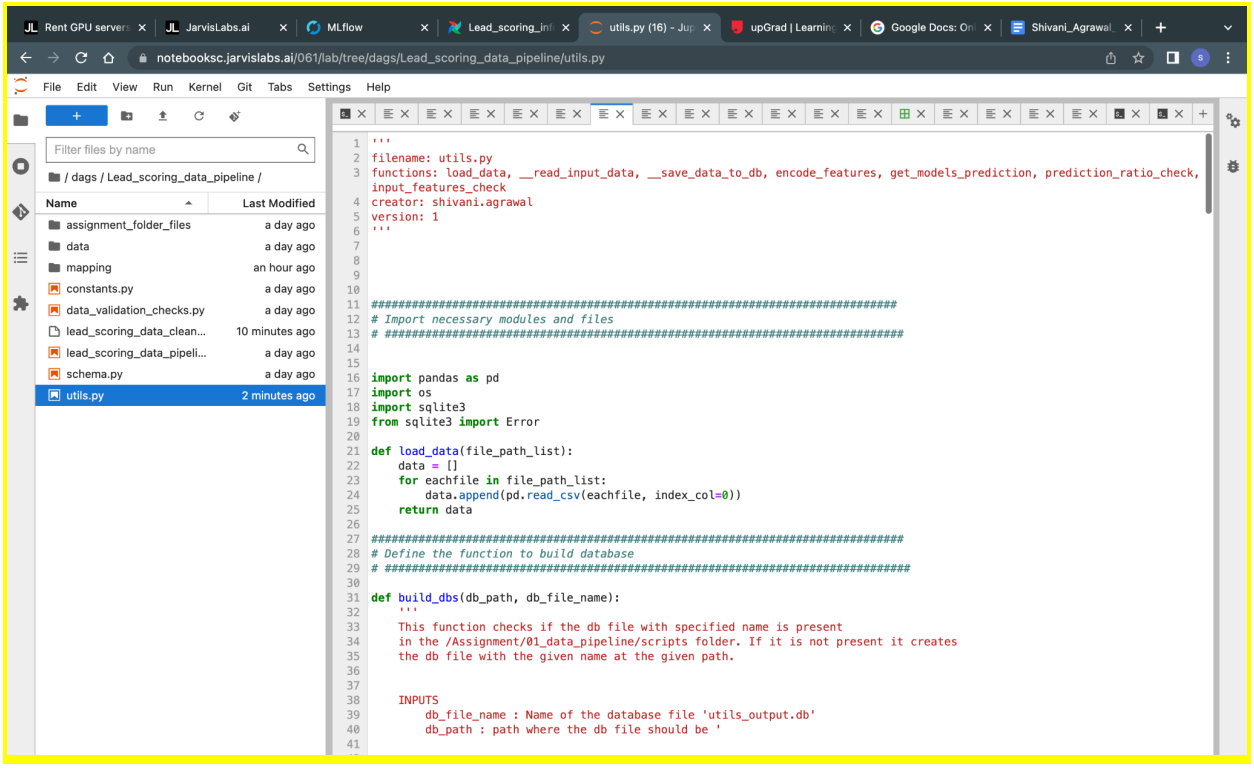
Lead_scoring_inference_pipeline:

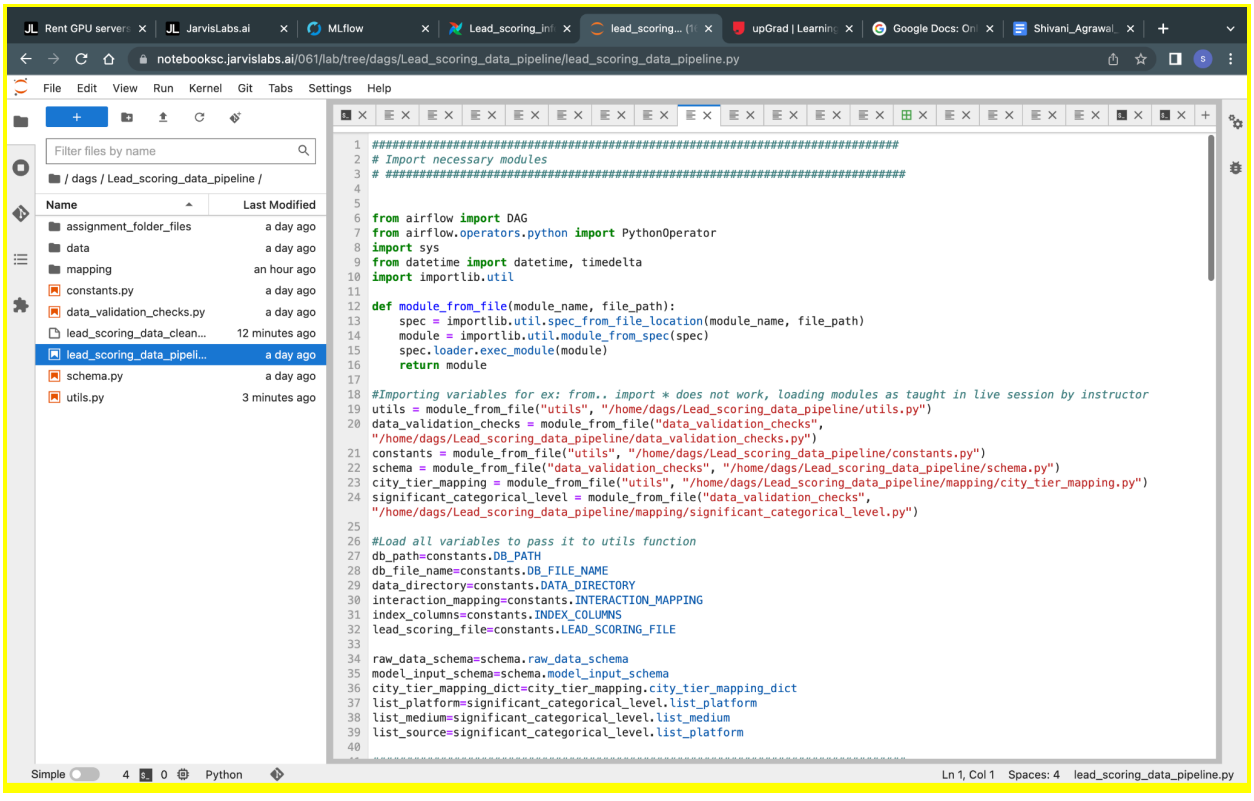


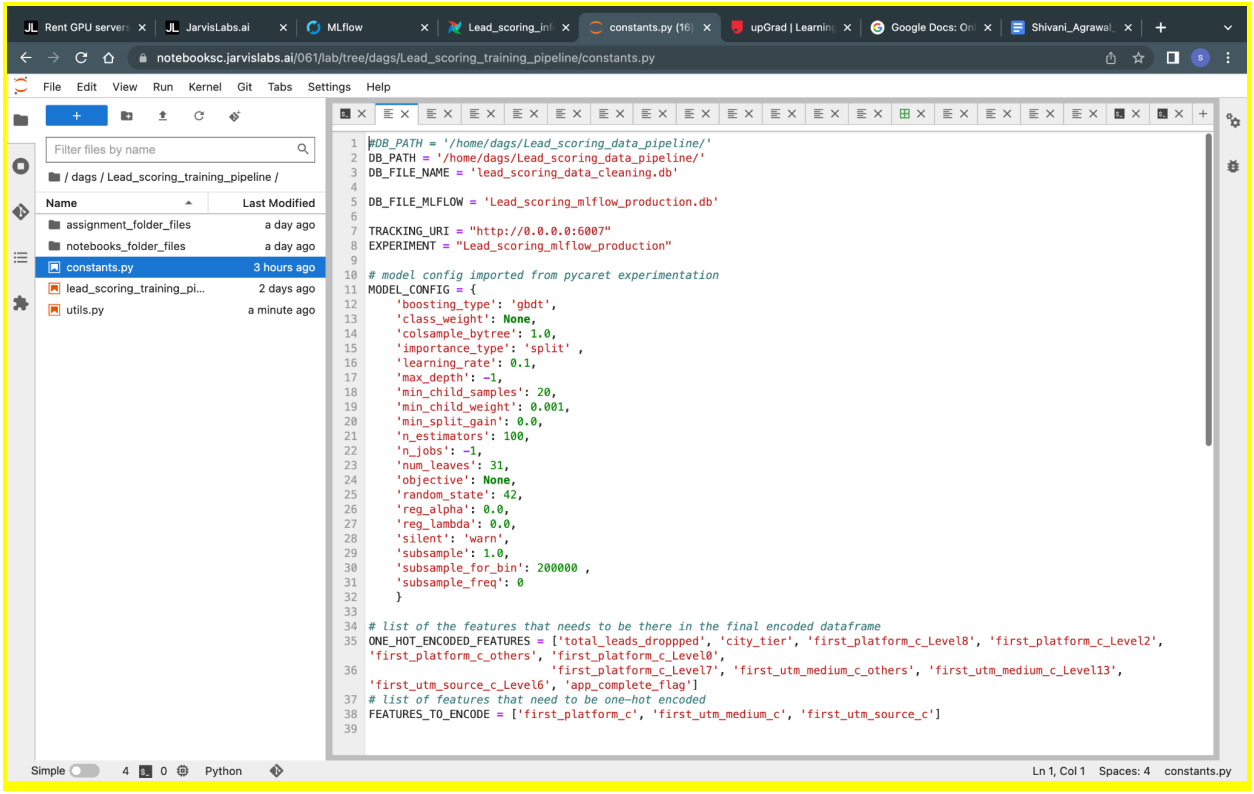


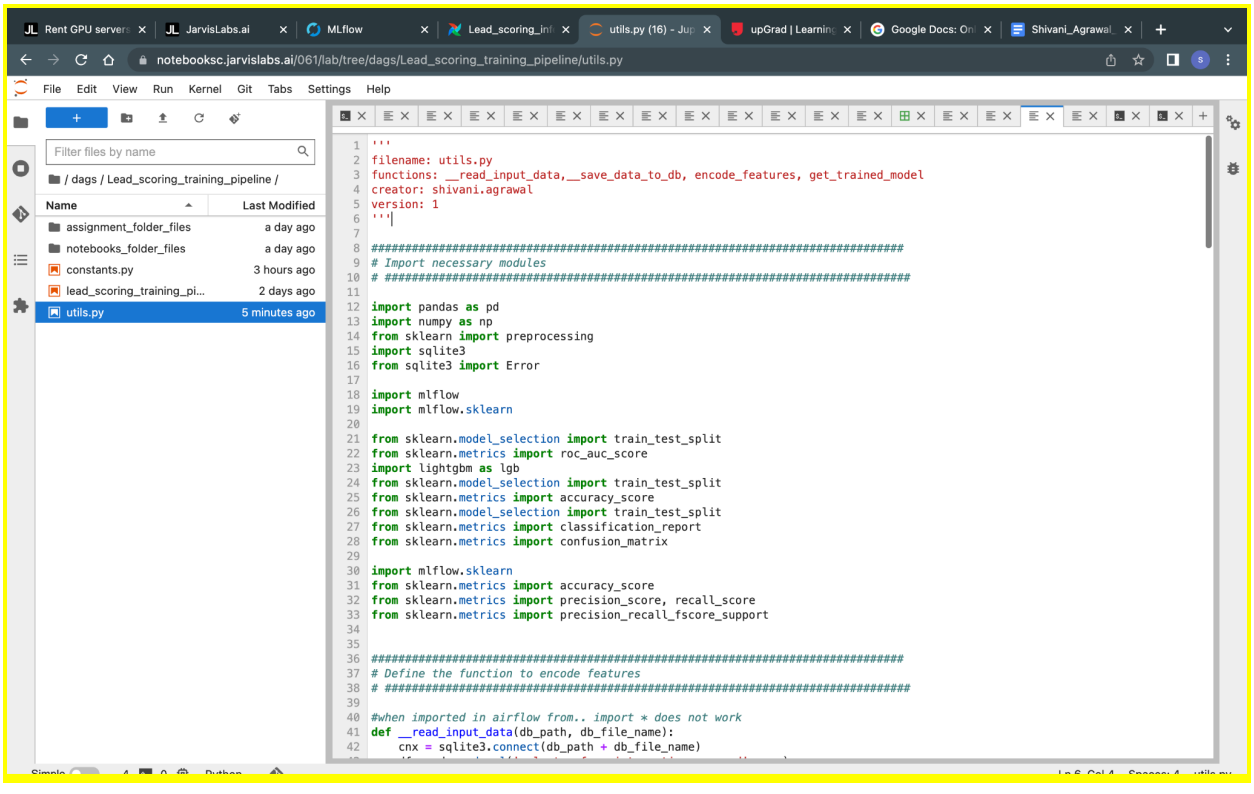
Some Jarvis jupyter code snaps:

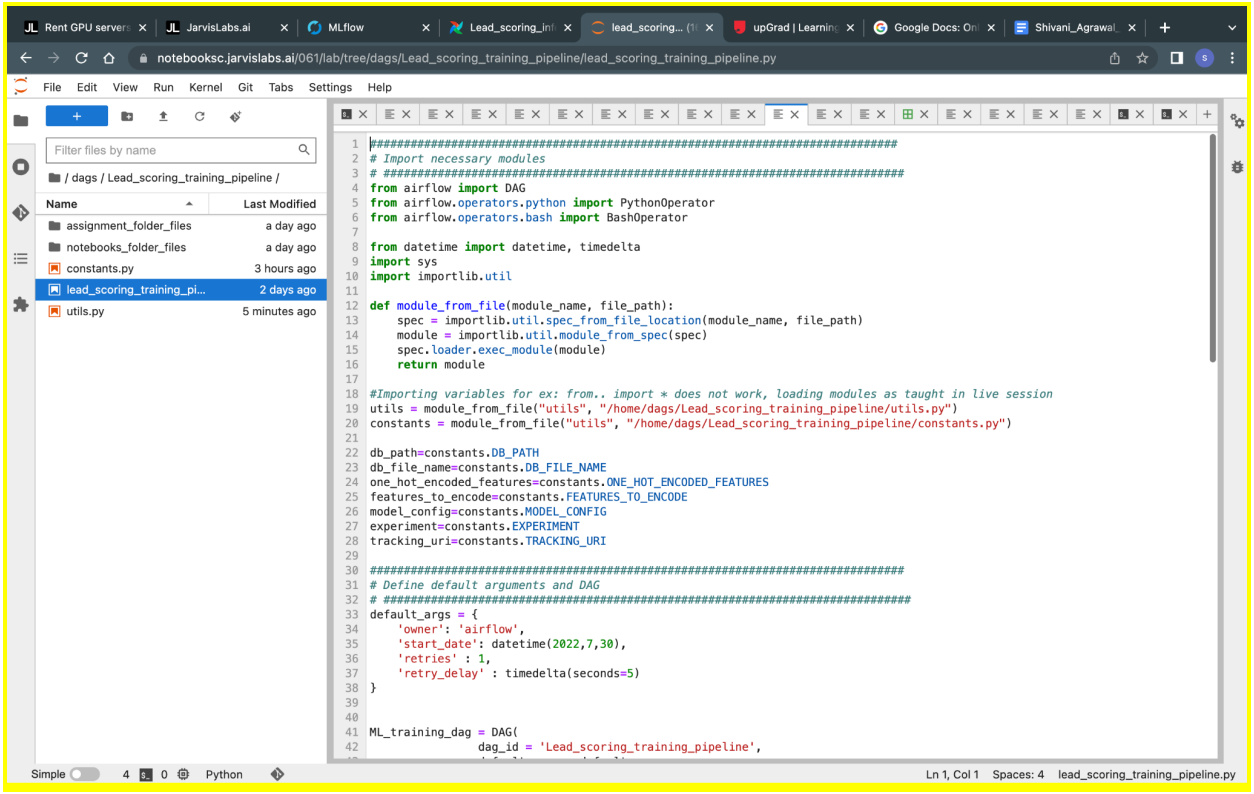



notebooksc.jarvislabs.ai/061/lab/tree/dags/Lead_scoring_data_pipeline/utils.py
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ dags / Lead_scoring_data_pipeline /
Name
Last Modified
assignment_folder_files a day ago
data a day ago
mapping an hour ago
constants.py a day ago
data_validation_checks.py a day ago
lead_scoring_data_clean... 10 minutes ago
lead_scoring_data_pipeli... a day ago
schema.py a day ago
utils.py 2 minutes ago
'''
filename: utils.py
functions: load_data, __read_input_data, __save_data_to_db, encode_features, get_models_prediction, prediction_ratio_check, input_features_check
creator: shivani.agrawal
version: 1
'''
# Import necessary modules and files
import pandas as pd
import os
import sqlite3
from sqlite3 import Error
def load_data(file_path_list):
data = []
for eachfile in file_path_list:
data.append(pd.read_csv(eachfile, index_col=0))
return data
# Define the function to build database
def build_dbs(db_path, db_file_name):
'''
This function checks if the db file with specified name is present
in the /Assignment/01_data_pipeline/scripts folder. If it is not present it creates
the db file with the given name at the given path.
INPUTS
db_file_name : Name of the database file 'utils_output.db'
db_path : path where the db file should be '

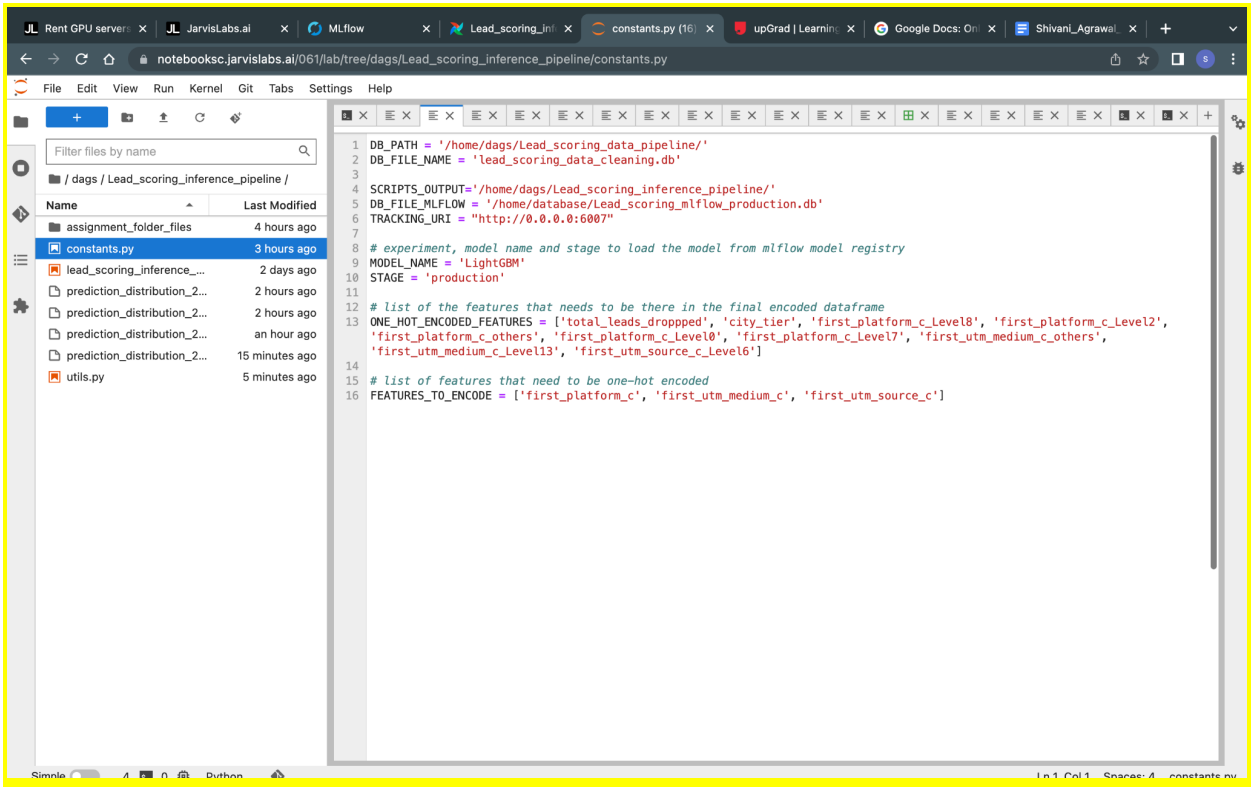

notebooksc.jarvislabs.ai/061/lab/tree/dags/Lead_scoring_data_pipeline/lead_scoring_data_pipeline.py
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ dags / Lead_scoring_data_pipeline /
Name
Last Modified
assignment_folder_files a day ago
data a day ago
mapping an hour ago
constants.py a day ago
data_validation_checks.py a day ago
lead_scoring_data_clean... 12 minutes ago
lead_scoring_data_pipeli... a day ago
schema.py a day ago
utils.py 3 minutes ago
# Import necessary modules
from airflow import DAG
from airflow.operators.python import PythonOperator
import sys
from datetime import datetime, timedelta
import importlib.util
def module_from_file(module_name, file_path):
spec = importlib.util.spec_from_file_location(module_name, file_path)
module = importlib.util.module_from_spec(spec)
spec.loader.exec_module(module)
return module
#Importing variables for ex: from.. import * does not work, loading modules as taught in live session by instructor
utils = module_from_file("utils", "/home/dags/Lead_scoring_data_pipeline/utils.py")
data_validation_checks = module_from_file("data_validation_checks", "/home/dags/Lead_scoring_data_pipeline/data_validation_checks.py")
constants = module_from_file("utils", "/home/dags/Lead_scoring_data_pipeline/constants.py")
schema = module_from_file("data_validation_checks", "/home/dags/Lead_scoring_data_pipeline/schema.py")
city_tier_mapping = module_from_file("utils", "/home/dags/Lead_scoring_data_pipeline/mapping/city_tier_mapping.py")
significant_categorical_level = module_from_file("data_validation_checks", "/home/dags/Lead_scoring_data_pipeline/mapping/significant_categorical_level.py")
#Load all variables to pass it to utils function
db_path=constants.DB_PATH
db_file_name=constants.DB_FILE_NAME
data_directory=constants.DATA_DIRECTORY
interaction_mapping=constants.INTERACTION_MAPPING
index_columns=constants.INDEX_COLUMNS
lead_scoring_file=constants.LEAD_SCORING_FILE
raw_data_schema=schema.raw_data_schema
model_input_schema=schema.model_input_schema
city_tier_mapping_dict=city_tier_mapping.city_tier_mapping_dict
list_platform=significant_categorical_level.list_platform
list_medium=significant_categorical_level.list_medium
list_source=significant_categorical_level.list_platform
Simple 4 0 Python
Ln 1, Col 1 Spaces: 4 lead_scoring_data_pipeline.py


Rent GPU servers
JarvisLabs.ai
MLflow
Lead_scoring_inf
constants.py (16)
upGrad | Learning
Google Docs: Online
Shivani_Agrawal
notebooksc.jarvislabs.ai/061/lab/tree/dags/Lead_scoring_training_pipeline/constants.py
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ dags / Lead_scoring_training_pipeline /
Name
Last Modified
assignment_folder_files a day ago
notebooks_folder_files a day ago
constants.py 3 hours ago
lead_scoring_training_pi... 2 days ago
utils.py a minute ago
#DB_PATH = '/home/dags/Lead_scoring_data_pipeline/'
DB_PATH = '/home/dags/Lead_scoring_data_pipeline/'
DB_FILE_NAME = 'lead_scoring_data_cleaning.db'
DB_FILE_MLFLOW = 'Lead_scoring_mlflow_production.db'
TRACKING_URI = "http://0.0.0.0:6007"
EXPERIMENT = "Lead_scoring_mlflow_production"
# model config imported from pycaret experimentation
MODEL_CONFIG = {
'boosting_type': 'gbdt',
'class_weight': None,
'colsample_bytree': 1.0,
'importance_type': 'split' ,
'learning_rate': 0.1,
'max_depth': -1,
'min_child_samples': 20,
'min_child_weight': 0.001,
'min_split_gain': 0.0,
'n_estimators': 100,
'n_jobs': -1,
'num_leaves': 31,
'objective': None,
'random_state': 42,
'reg_alpha': 0.0,
'reg_lambda': 0.0,
'silent': 'warn',
'subsample': 1.0,
'subsample_for_bin': 200000 ,
'subsample_freq': 0
}
# list of the features that needs to be there in the final encoded dataframe
ONE_HOT_ENCODED_FEATURES = ['total_leads_droppped', 'city_tier', 'first_platform_c_Level8', 'first_platform_c_Level2',
'first_platform_c_others', 'first_platform_c_Level0',
'first_platform_c_Level7', 'first_utm_medium_c_others', 'first_utm_medium_c_Level13',
'first_utm_source_c_Level6', 'app_complete_flag']
# list of features that need to be one-hot encoded
FEATURES_TO_ENCODE = ['first_platform_c', 'first_utm_medium_c', 'first_utm_source_c']
Simple 4 0 Python
Ln 1, Col 1 Spaces: 4 constants.py

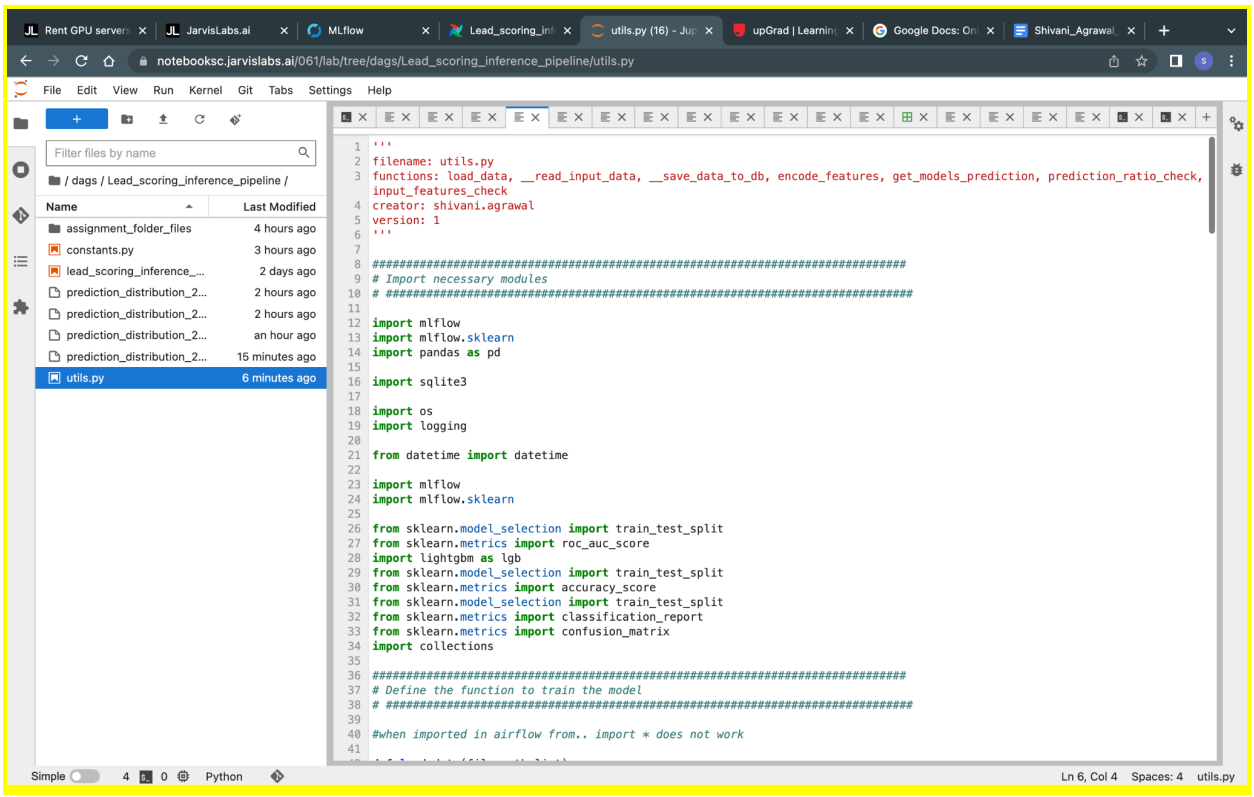

Rent GPU servers
JarvisLabs.ai
MLflow
Lead_scoring_inf
utils.py (16) - Jup
upGrad | Learning
Google Docs: Online
Shivani_Agrawal
notebooksc.jarvislabs.ai/061/lab/tree/dags/Lead_scoring_training_pipeline/utils.py
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ dags / Lead_scoring_training_pipeline /
Name
Last Modified
assignment_folder_files a day ago
notebooks_folder_files a day ago
constants.py 3 hours ago
lead_scoring_training_pi... 2 days ago
utils.py 5 minutes ago
'''
filename: utils.py
functions: __read_input_data,__save_data_to_db, encode_features, get_trained_model
creator: shivani.agrawal
version: 1
'''
###
# Import necessary modules
# ###
import pandas as pd
import numpy as np
from sklearn import preprocessing
import sqlite3
from sqlite3 import Error
import mlflow
import mlflow.sklearn
from sklearn.model_selection import train_test_split
from sklearn.metrics import roc_auc_score
import lightgbm as lgb
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
import mlflow.sklearn
from sklearn.metrics import accuracy_score
from sklearn.metrics import precision_score, recall_score
from sklearn.metrics import precision_recall_fscore_support
###
Define the function to encode features
###
#when imported in airflow from.. import * does not work
def __read_input_data(db_path, db_file_name):
cnx = sqlite3.connect(db_path + db_file_name)



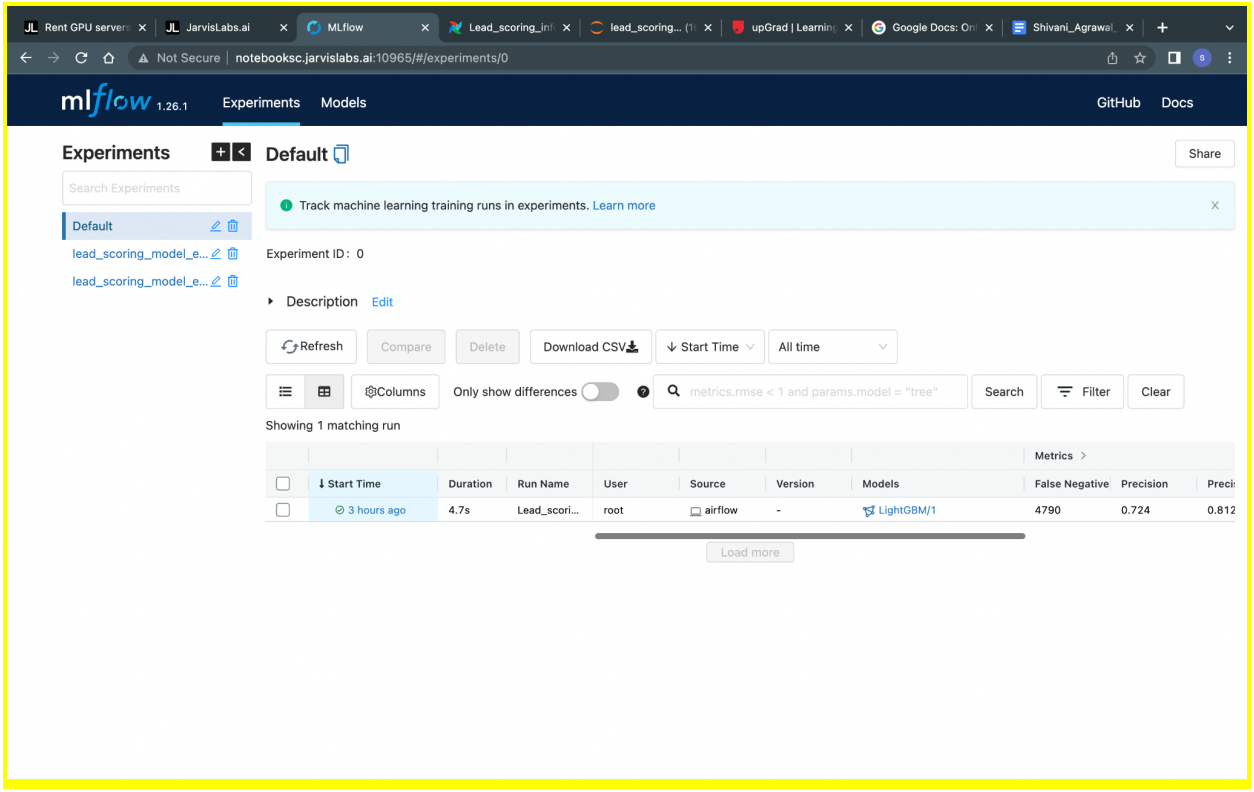

```python
##############################################################################
# Import necessary modules
# ##############################################################################
from airflow import DAG
from airflow.operators.python import PythonOperator
from airflow.operators.bash import BashOperator

from datetime import datetime, timedelta
import sys
import importlib.util

def module_from_file(module_name, file_path):
    spec = importlib.util.spec_from_file_location(module_name, file_path)
    module = importlib.util.module_from_spec(spec)
    spec.loader.exec_module(module)
    return module

#Importing variables for ex: from.. import * does not work, loading modules as taught in live session
utils = module_from_file("utils", "/home/dags/Lead_scoring_training_pipeline/utils.py")
constants = module_from_file("utils", "/home/dags/Lead_scoring_training_pipeline/constants.py")

db_path=constants.DB_PATH
db_file_name=constants.DB_FILE_NAME
one_hot_encoded_features=constants.ONE_HOT_ENCODED_FEATURES
features_to_encode=constants.FEATURES_TO_ENCODE
model_config=constants.MODEL_CONFIG
experiment=constants.EXPERIMENT
tracking_uri=constants.TRACKING_URI

##############################################################################
# Define default arguments and DAG
# ##############################################################################
default_args = {
    'owner': 'airflow',
    'start_date': datetime(2022,7,30),
    'retries' : 1,
    'retry_delay' : timedelta(seconds=5)
}


ML_training_dag = DAG(
                dag_id = 'Lead_scoring_training_pipeline',
```

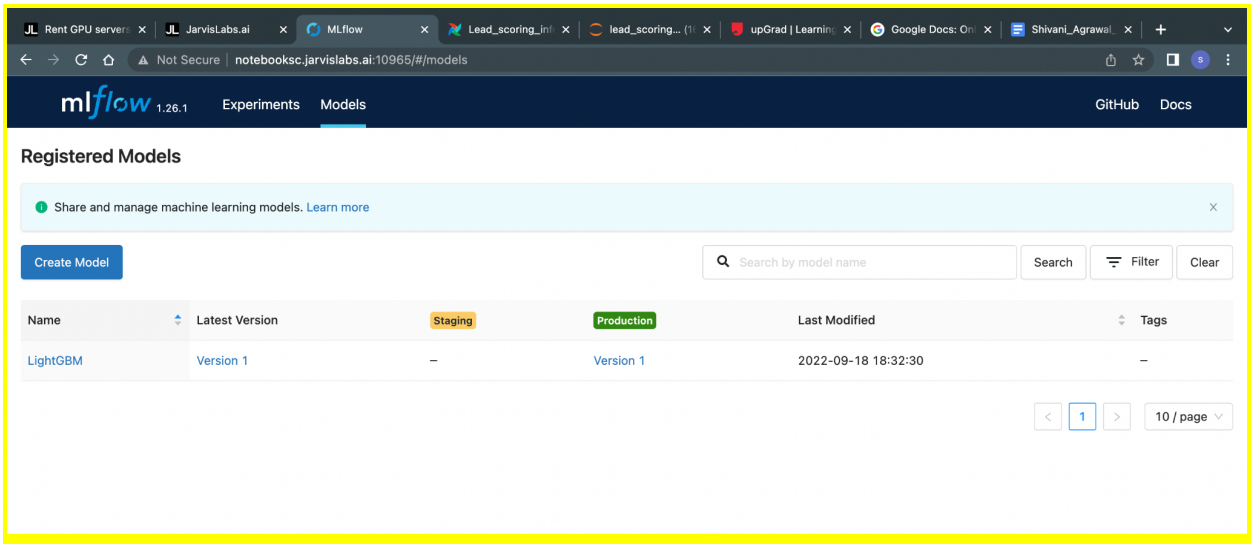


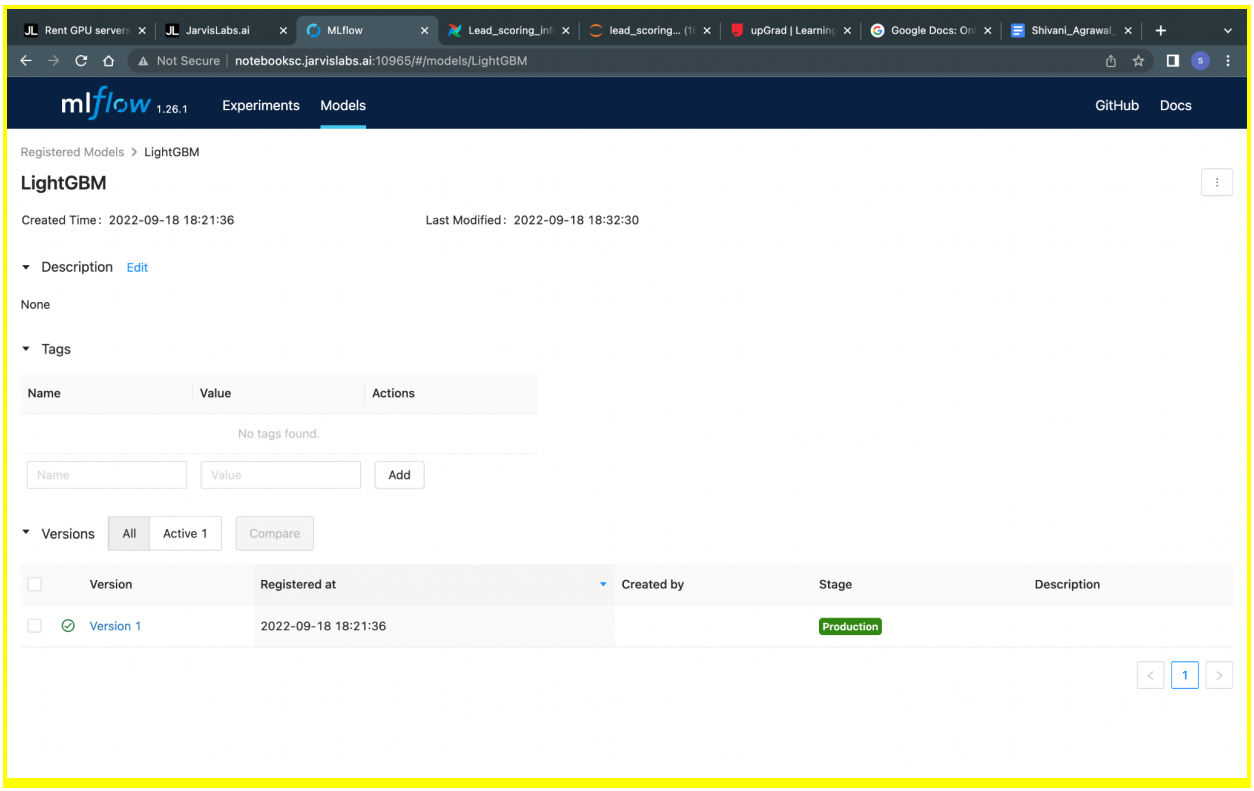

```python
DB_PATH = '/home/dags/Lead_scoring_data_pipeline/'
DB_FILE_NAME = 'lead_scoring_data_cleaning.db'

SCRIPTS_OUTPUT='/home/dags/Lead_scoring_inference_pipeline/'
DB_FILE_MLFLOW = '/home/database/Lead_scoring_mlflow_production.db'
TRACKING_URI = "http://0.0.0.0:6007"

# experiment, model name and stage to load the model from mlflow model registry
MODEL_NAME = 'LightGBM'
STAGE = 'production'

# list of the features that needs to be there in the final encoded dataframe
ONE_HOT_ENCODED_FEATURES = ['total_leads_droppped', 'city_tier', 'first_platform_c_Level8', 'first_platform_c_Level2',
'first_platform_c_others', 'first_platform_c_Level0', 'first_platform_c_Level7', 'first_utm_medium_c_others',
'first_utm_medium_c_Level13', 'first_utm_source_c_Level6']

# list of features that need to be one-hot encoded
FEATURES_TO_ENCODE = ['first_platform_c', 'first_utm_medium_c', 'first_utm_source_c']
```


Rent GPU servers
JarvisLabs.ai
MLflow
Lead_scoring_inf
utils.py (16) - Jup
upGrad | Learning
Google Docs: Onl
Shivani_Agrawal_
notebooksc.jarvislabs.ai/061/lab/tree/dags/Lead_scoring_inference_pipeline/utils.py
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ dags / Lead_scoring_inference_pipeline /
Name
Last Modified
assignment_folder_files 4 hours ago
constants.py 3 hours ago
lead_scoring_inference_... 2 days ago
prediction_distribution_2... 2 hours ago
prediction_distribution_2... 2 hours ago
prediction_distribution_2... an hour ago
prediction_distribution_2... 15 minutes ago
utils.py 6 minutes ago
'''
filename: utils.py
functions: load_data, __read_input_data, __save_data_to_db, encode_features, get_models_prediction, prediction_ratio_check, input_features_check
creator: shivani.agrawal
version: 1
'''
##
Import necessary modules
##
import mlflow
import mlflow.sklearn
import pandas as pd
import sqlite3
import os
import logging
from datetime import datetime
import mlflow
import mlflow.sklearn
from sklearn.model_selection import train_test_split
from sklearn.metrics import roc_auc_score
import lightgbm as lgb
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
import collections
##
Define the function to train the model
##
#when imported in airflow from.. import * does not work
Simple 4 0 Python
Ln 6, Col 4 Spaces: 4 utils.py

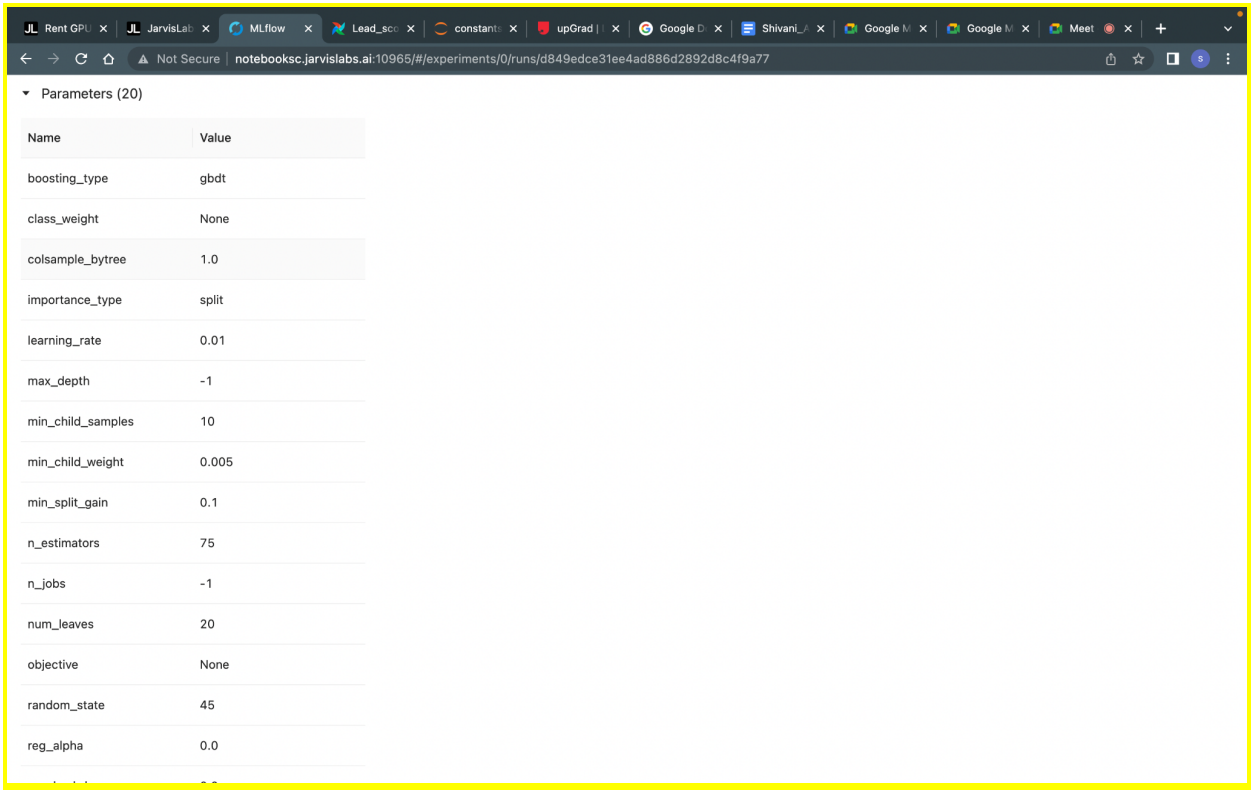
MLflow:


Rent GPU servers
JarvisLabs.ai
MLflow
Lead_scoring_Inf
lead_scoring... (1
upGrad | Learning
Google Docs: Onl
Shivani_Agrawal
Not Secure | notebooksc.jarvislabs.ai:10965/#/experiments/0
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Experiments
Default
Share
Search Experiments
Default
lead_scoring_model_e...
lead_scoring_model_e...
Track machine learning training runs in experiments. Learn more
Experiment ID: 0
Description
Edit
Refresh
Compare
Delete
Download CSV
Start Time
All time
Columns
Only show differences
metrics.rmse < 1 and params.model = "tree"
Search
Filter
Clear
Showing 1 matching run
Metrics
Start Time
Duration
Run Name
User
Source
Version
Models
False Negative
Precision
3 hours ago
4.7s
Lead_scori...
root
airflow
-
LightGBM/1
4790
0.724
0.812
Load more


Rent GPU servers
JarvisLabs.ai
MLflow
Lead_scoring_Inf
lead_scoring... (1
upGrad | Learning
Google Docs: Onl
Shivani_Agrawal
Not Secure | notebooksc.jarvislabs.ai:10965/#/models
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Registered Models
Share and manage machine learning models. Learn more
Create Model
Search by model name
Search
Filter
Clear
Name
Latest Version
Staging
Production
Last Modified
Tags
LightGBM
Version 1
–
Version 1
2022-09-18 18:32:30
–
1
10 / page

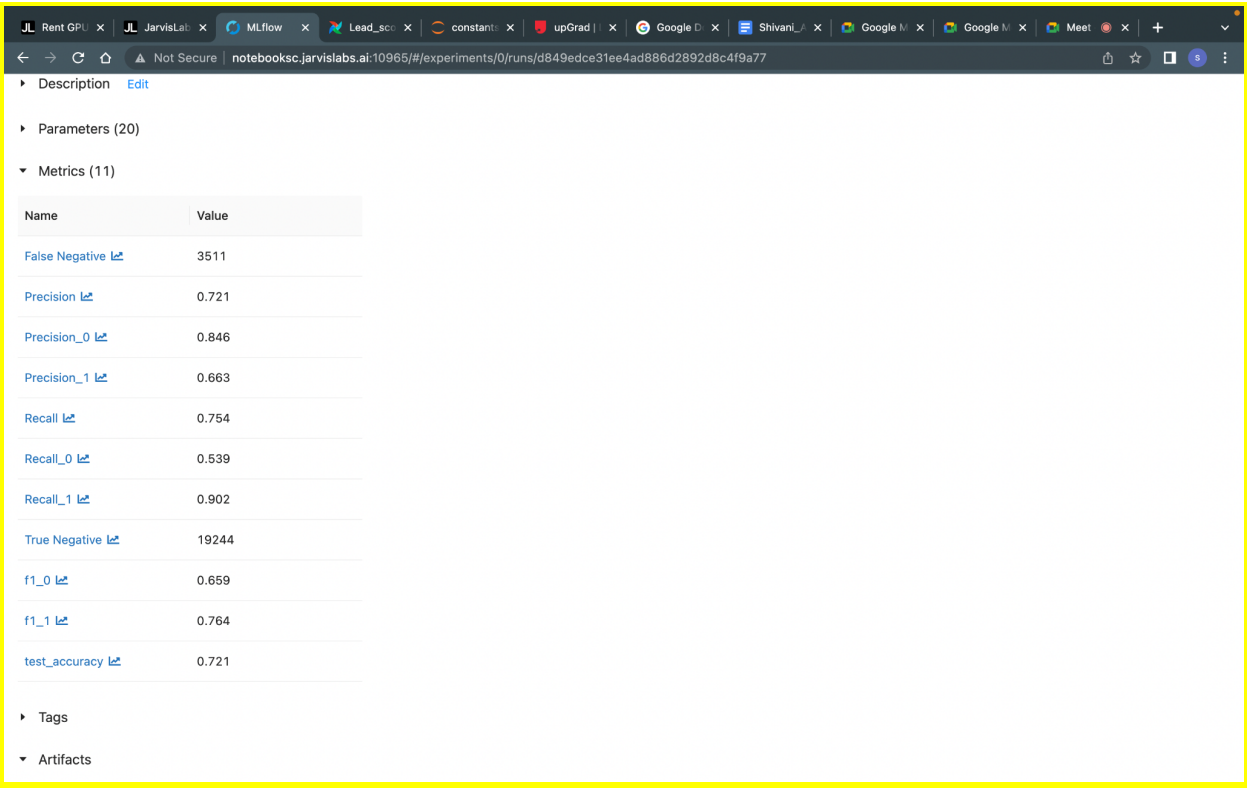

Not Secure | notebooksc.jarvislabs.ai:10965/#/models/LightGBM
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Registered Models > LightGBM
LightGBM
Created Time: 2022-09-18 18:21:36
Last Modified: 2022-09-18 18:32:30
Description Edit
None
Tags
Name
Value
Actions
No tags found.
Name
Value
Add
Versions
All
Active 1
Compare
Version
Registered at
Created by
Stage
Description
Version 1
2022-09-18 18:21:36
Production
1

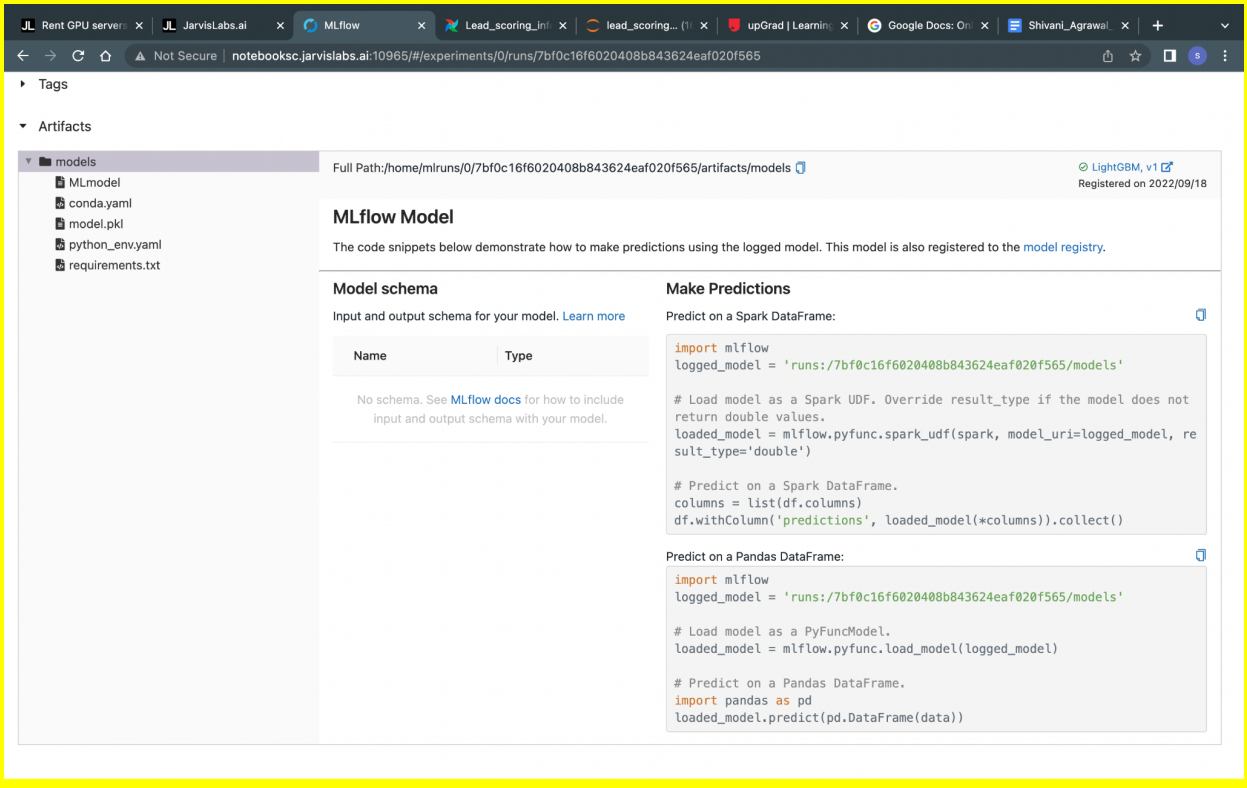
Parameters:

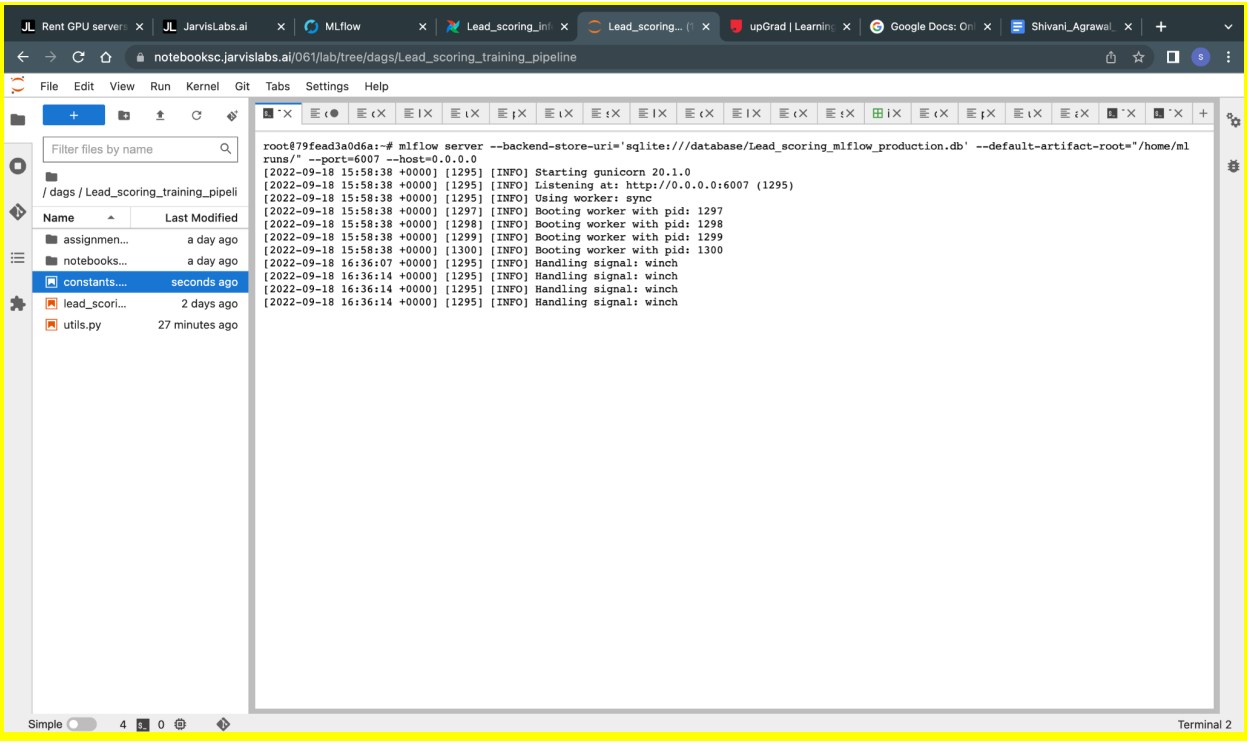

Not Secure | notebooksc.jarvislabs.ai:10965/#/experiments/0/runs/d849edce31ee4ad886d2892d8c4f9a77
Parameters (20)
Name
Value
boosting_type
gbdt
class_weight
None
colsample_bytree
1.0
importance_type
split
learning_rate
0.01
max_depth
-1
min_child_samples
10
min_child_weight
0.005
min_split_gain
0.1
n_estimators
75
n_jobs
-1
num_leaves
20
objective
None
random_state
45
reg_alpha
0.0

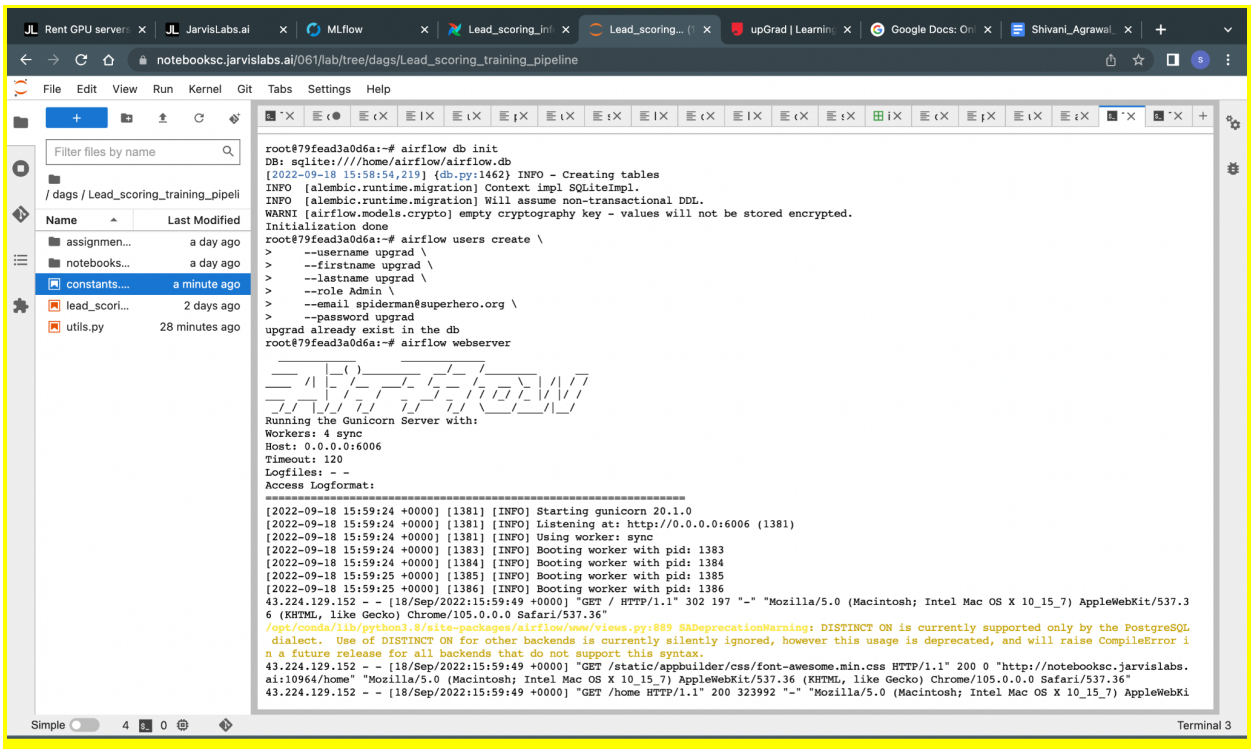
reg_lambda	0.0
silent	warn
subsample	1.0
subsample_for_bin	2000
subsample_freq	0

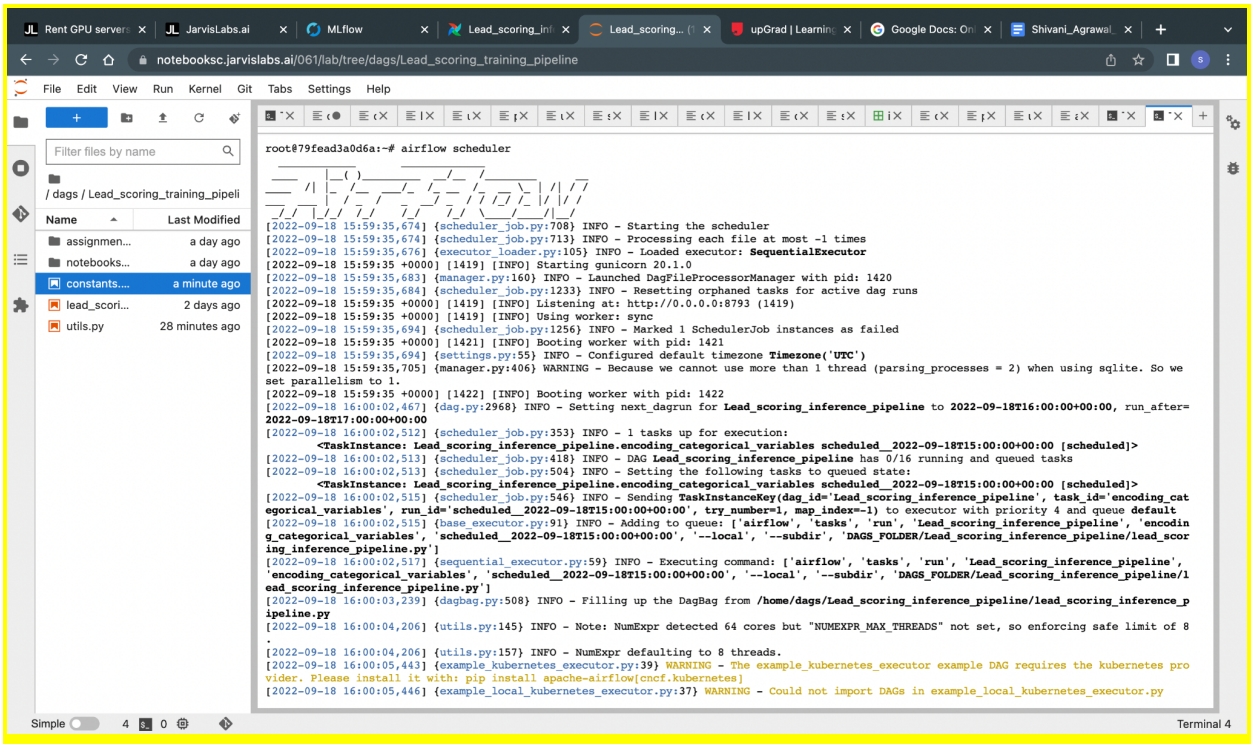
Metrics :




Rent GPU servers
JarvisLabs.ai
MLflow
Lead_scoring_Inf
lead_scoring...
upGrad | Learning
Google Docs: Onl
Shivani_Agrawal
Not Secure
notebooksc.jarvislabs.ai:10965/#/experiments/0/runs/7bf0c16f6020408b843624eaf020f565
Tags
Artifacts
models
MLmodel
conda.yaml
model.pkl
python_env.yaml
requirements.txt
Full Path:/home/mlruns/0/7bf0c16f6020408b843624eaf020f565/artifacts/models
LightGBM, v1
Registered on 2022/09/18
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
No schema. See MLflow docs for how to include input and output schema with your model.
Make Predictions
Predict on a Spark DataFrame:
import mlflow
logged_model = 'runs:/7bf0c16f6020408b843624eaf020f565/models'
# 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.
columns = list(df.columns)
df.withColumn('predictions', loaded_model(*columns)).collect()
Predict on a Pandas DataFrame:
import mlflow
logged_model = 'runs:/7bf0c16f6020408b843624eaf020f565/models'
# Load model as a PyFuncModel.
loaded_model = mlflow.pyfunc.load_model(logged_model)
# Predict on a Pandas DataFrame.
import pandas as pd
loaded_model.predict(pd.DataFrame(data))


Rent GPU servers
JarvisLabs.ai
MLflow
Lead_scoring_Inf
Lead_scoring...
upGrad | Learning
Google Docs: Onl
Shivani_Agrawal
notebooksc.jarvislabs.ai/061/lab/tree/dags/Lead_scoring_training_pipeline
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ dags / Lead_scoring_training_pipeli
Name
Last Modified
assignmen... a day ago
notebooks... a day ago
constants.... seconds ago
lead_scori... 2 days ago
utils.py 27 minutes ago
root@79fead3a0d6a:~# mlflow server --backend-store-uri='sqlite:///database/Lead_scoring_mlflow_production.db' --default-artifact-root="/home/mlruns/" --port=6007 --host=0.0.0.0
[2022-09-18 15:58:38 +0000] [1295] [INFO] Starting gunicorn 20.1.0
[2022-09-18 15:58:38 +0000] [1295] [INFO] Listening at: http://0.0.0.0:6007 (1295)
[2022-09-18 15:58:38 +0000] [1295] [INFO] Using worker: sync
[2022-09-18 15:58:38 +0000] [1297] [INFO] Booting worker with pid: 1297
[2022-09-18 15:58:38 +0000] [1298] [INFO] Booting worker with pid: 1298
[2022-09-18 15:58:38 +0000] [1299] [INFO] Booting worker with pid: 1299
[2022-09-18 15:58:38 +0000] [1300] [INFO] Booting worker with pid: 1300
[2022-09-18 16:36:07 +0000] [1295] [INFO] Handling signal: winch
[2022-09-18 16:36:14 +0000] [1295] [INFO] Handling signal: winch
[2022-09-18 16:36:14 +0000] [1295] [INFO] Handling signal: winch
[2022-09-18 16:36:14 +0000] [1295] [INFO] Handling signal: winch
Simple
Terminal 2


Rent GPU servers
JarvisLabs.ai
MLflow
Lead_scoring_inf
Lead_scoring...
upGrad | Learning
Google Docs: Online
Shivani_Agrawal
notebooksc.jarvislabs.ai/061/lab/tree/dags/Lead_scoring_training_pipeline
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ dags / Lead_scoring_training_pipeli
Name
Last Modified
assignmen... a day ago
notebooks... a day ago
constants.... a minute ago
lead_scori... 2 days ago
utils.py 28 minutes ago
root@79fead3a0d6a:~# airflow db init
DB: sqlite:////home/airflow/airflow.db
[2022-09-18 15:58:54,219] {db.py:1462} INFO - Creating tables
INFO [alembic.runtime.migration] Context impl SQLiteImpl.
INFO [alembic.runtime.migration] Will assume non-transactional DDL.
WARNI [airflow.models.crypto] empty cryptography key - values will not be stored encrypted.
Initialization done
root@79fead3a0d6a:~# airflow users create \
> --username upgrad \
> --firstname upgrad \
> --lastname upgrad \
> --role Admin \
> --email spiderman@superhero.org \
> --password upgrad
upgrad already exist in the db
root@79fead3a0d6a:~# airflow webserver
Running the Gunicorn Server with:
Workers: 4 sync
Host: 0.0.0.0:6006
Timeout: 120
Logfiles: - -
Access Logformat:
[2022-09-18 15:59:24 +0000] [1381] [INFO] Starting gunicorn 20.1.0
[2022-09-18 15:59:24 +0000] [1381] [INFO] Listening at: http://0.0.0.0:6006 (1381)
[2022-09-18 15:59:24 +0000] [1381] [INFO] Using worker: sync
[2022-09-18 15:59:24 +0000] [1383] [INFO] Booting worker with pid: 1383
[2022-09-18 15:59:24 +0000] [1384] [INFO] Booting worker with pid: 1384
[2022-09-18 15:59:25 +0000] [1385] [INFO] Booting worker with pid: 1385
[2022-09-18 15:59:25 +0000] [1386] [INFO] Booting worker with pid: 1386
43.224.129.152 - - [18/Sep/2022:15:59:49 +0000] "GET / HTTP/1.1" 302 197 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/105.0.0.0 Safari/537.36"
/opt/conda/lib/python3.8/site-packages/airflow/www/views.py:889 SADeprecationWarning: DISTINCT ON is currently supported only by the PostgreSQL dialect. Use of DISTINCT ON for other backends is currently silently ignored, however this usage is deprecated, and will raise CompileError in a future release for all backends that do not support this syntax.
43.224.129.152 - - [18/Sep/2022:15:59:49 +0000] "GET /static/appbuilder/css/font-awesome.min.css HTTP/1.1" 200 0 "http://notebooksc.jarvislabs.ai:10964/home" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/105.0.0.0 Safari/537.36"
43.224.129.152 - - [18/Sep/2022:15:59:49 +0000] "GET /home HTTP/1.1" 200 323992 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKi
Simple
Terminal 3


Rent GPU servers
JarvisLabs.ai
MLflow
Lead_scoring_inf
Lead_scoring...
upGrad | Learning
Google Docs: Online
Shivani_Agrawal
notebooksc.jarvislabs.ai/061/lab/tree/dags/Lead_scoring_training_pipeline
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ dags / Lead_scoring_training_pipeli
Name
Last Modified
assignmen... a day ago
notebooks... a day ago
constants.... a minute ago
lead_scori... 2 days ago
utils.py 28 minutes ago
root@79fead3a0d6a:~# airflow scheduler
[2022-09-18 15:59:35,674] {scheduler_job.py:708} INFO - Starting the scheduler
[2022-09-18 15:59:35,674] {scheduler_job.py:713} INFO - Processing each file at most -1 times
[2022-09-18 15:59:35,676] {executor_loader.py:105} INFO - Loaded executor: SequentialExecutor
[2022-09-18 15:59:35 +0000] [1419] [INFO] Starting gunicorn 20.1.0
[2022-09-18 15:59:35,683] {manager.py:160} INFO - Launched DagFileProcessorManager with pid: 1420
[2022-09-18 15:59:35,684] {scheduler_job.py:1233} INFO - Resetting orphaned tasks for active dag runs
[2022-09-18 15:59:35 +0000] [1419] [INFO] Listening at: http://0.0.0.0:8793 (1419)
[2022-09-18 15:59:35 +0000] [1419] [INFO] Using worker: sync
[2022-09-18 15:59:35,694] {scheduler_job.py:1256} INFO - Marked 1 SchedulerJob instances as failed
[2022-09-18 15:59:35 +0000] [1421] [INFO] Booting worker with pid: 1421
[2022-09-18 15:59:35,694] {settings.py:55} INFO - Configured default timezone Timezone('UTC')
[2022-09-18 15:59:35,705] {manager.py:406} WARNING - Because we cannot use more than 1 thread (parsing_processes = 2) when using sqlite. So we set parallelism to 1.
[2022-09-18 15:59:35 +0000] [1422] [INFO] Booting worker with pid: 1422
[2022-09-18 16:00:02,467] {dag.py:2968} INFO - Setting next_dagrun for Lead_scoring_inference_pipeline to 2022-09-18T16:00:00+00:00, run_after=2022-09-18T17:00:00+00:00
[2022-09-18 16:00:02,512] {scheduler_job.py:353} INFO - 1 tasks up for execution:
<TaskInstance: Lead_scoring_inference_pipeline.encoding_categorical_variables scheduled__2022-09-18T15:00:00+00:00 [scheduled]>
[2022-09-18 16:00:02,513] {scheduler_job.py:418} INFO - DAG Lead_scoring_inference_pipeline has 0/16 running and queued tasks
[2022-09-18 16:00:02,513] {scheduler_job.py:504} INFO - Setting the following tasks to queued state:
<TaskInstance: Lead_scoring_inference_pipeline.encoding_categorical_variables scheduled__2022-09-18T15:00:00+00:00 [scheduled]>
[2022-09-18 16:00:02,515] {scheduler_job.py:546} INFO - Sending TaskInstanceKey(dag_id='Lead_scoring_inference_pipeline', task_id='encoding_categorical_variables', run_id='scheduled__2022-09-18T15:00:00+00:00', try_number=1, map_index=-1) to executor with priority 4 and queue default
[2022-09-18 16:00:02,515] {base_executor.py:91} INFO - Adding to queue: ['airflow', 'tasks', 'run', 'Lead_scoring_inference_pipeline', 'encoding_categorical_variables', 'scheduled__2022-09-18T15:00:00+00:00', '--local', '--subdir', 'DAGS_FOLDER/Lead_scoring_inference_pipeline/lead_scoring_inference_pipeline.py']
[2022-09-18 16:00:02,517] {sequential_executor.py:59} INFO - Executing command: ['airflow', 'tasks', 'run', 'Lead_scoring_inference_pipeline', 'encoding_categorical_variables', 'scheduled__2022-09-18T15:00:00+00:00', '--local', '--subdir', 'DAGS_FOLDER/Lead_scoring_inference_pipeline/lead_scoring_inference_pipeline.py']
[2022-09-18 16:00:03,239] {dagbag.py:508} INFO - Filling up the DagBag from /home/dags/Lead_scoring_inference_pipeline/lead_scoring_inference_pipeline.py
[2022-09-18 16:00:04,206] {utils.py:145} INFO - Note: NumExpr detected 64 cores but "NUMEXPR_MAX_THREADS" not set, so enforcing safe limit of 8.
[2022-09-18 16:00:04,206] {utils.py:157} INFO - NumExpr defaulting to 8 threads.
[2022-09-18 16:00:05,443] {example_kubernetes_executor.py:39} WARNING - The example_kubernetes_executor example DAG requires the kubernetes provider. Please install it with: pip install apache-airflow[cncf.kubernetes]
[2022-09-18 16:00:05,446] {example_local_kubernetes_executor.py:37} WARNING - Could not import DAGs in example_local_kubernetes_executor.py
Simple
Terminal 4