ASSIGNMENT

Install mlflow and required libraries.(5points)

```
import warnings
import math
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics
from sklearn.metrics import accuracy_score,f1_score,recall_score,prec
import mlflow
import mlflow.sklearn
from mlflow import log_metric, log_param, log_artifacts, log_metrics
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import LinearRegression , Ridge , Lasso
from sklearn.ensemble import RandomForestRegressor
```

```
Model trained
(Titanic) maheshnukala@Maheshs-MacBook-Pro PartB % conda activate Titanic
(Titanic) maheshnukala@Maheshs-MacBook-Pro PartB % pip install numpy pandas matplotlib seaborn scikit-learn mlflow d
vc
Requirement already satisfied: numpy in /opt/anaconda3/envs/titanic/lib/python3.8/site-packages (1.23.4)
Requirement already satisfied: pandas in /opt/anaconda3/envs/titanic/lib/python3.8/site-packages (1.5.1)
Requirement already satisfied: matplotlib in /opt/anaconda3/envs/titanic/lib/python3.8/site-packages (3.6.2)
Requirement already satisfied: seaborn in /opt/anaconda3/envs/titanic/lib/python3.8/site-packages (0.12.1)
Requirement already satisfied: scikit-learn in /opt/anaconda3/envs/titanic/lib/python3.8/site-packages (1.1.3)
```

Load the given data, perform basic EDA, and build a classification model to predict the passengers who survived the titanic shipwreck. (Note:- you can use any classification algorithm you have learned in your previous modules). (20 points)

```
if __name__ == '__main__':
    print('Starting the experiment')

    #mlflow.set_tracking_uri("http://127.0.0.0:5000")
    mlflow.set_experiment(experiment_name = 'Titanic')
    df = pd.read_csv("/Users/maheshnukala/Desktop/IIIT/PartB/Week2/download.csv")
    df.head()
    df.isnull().sum()
    df=df.replace('male','1').replace('female','0')
    df['Sex']=df['Sex'].astype(int)
    df=df.replace('S','1').replace('C','3').replace('Q','2')
    df['Embarked'] = df['Embarked'].replace(np.nan, '3')
    df['Embarked']=df['Embarked'].astype(int)
    df.describe()
    df['Age'] = df['Age'].replace(np.nan, 29.6)
    df = df.drop(['Name'], axis=1)
    df = df.drop(['Ticket'], axis=1)
    df = df.drop(['PassengerId'], axis=1)
    df = df.drop(['Cabin'], axis=1)
    df_feature = df.drop(['Survived'], axis=1)
    df_target = df.Survived
    df_feature.shape, df_target.shape
    corre=df.corr()
    corre
    from sklearn.model_selection import train_test_split

    X_train, X_validation, y_train, y_validation = train_test_split(df_feature, df_targ
    X_train.shape, X_validation.shape, y_train.shape, y_validation.shape
    log_param("Train shape",X_train.shape )
    log_param("Test shape",X_validation.shape )
    #model_entropy = DecisionTreeClassifier(criterion = "gini",
                                           #    max_depth=10, min_samples_leaf=3)
```

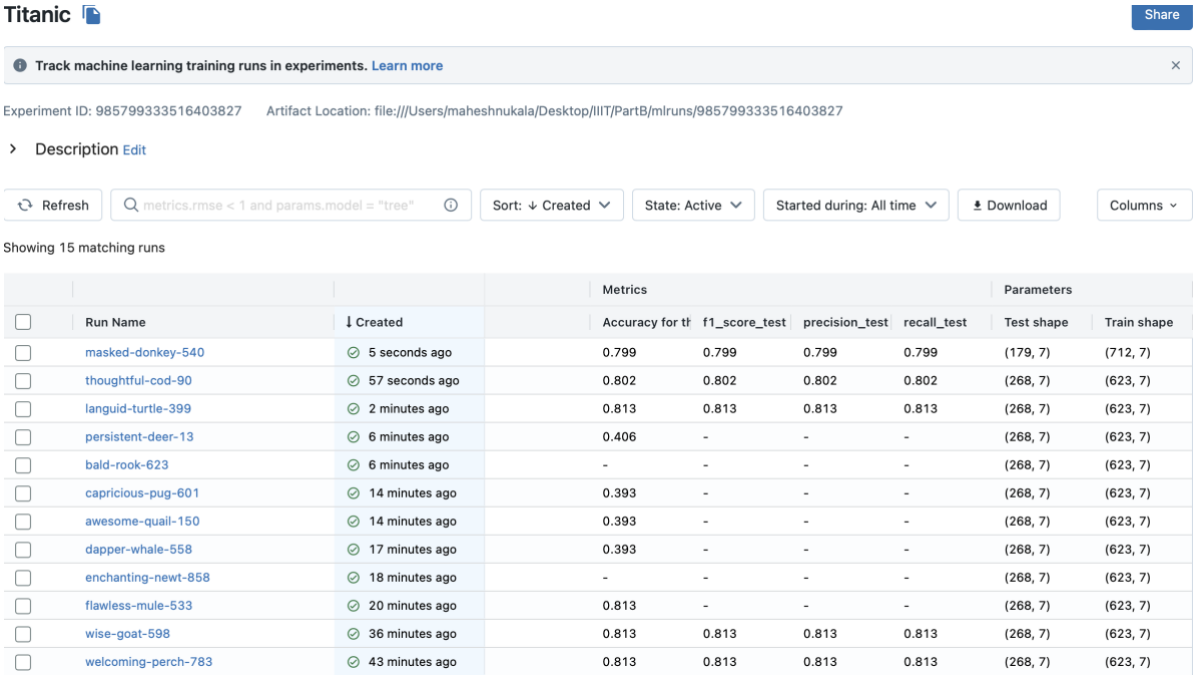
- Log the following parameters using mlflow.(5points)

```
log_param("Train shape",X_train.shape )
log_param("Test shape",X_validation.shape )
```

```
log_metrics(perf_metrics)
mlflow.sklearn.log_model(classifier, "Decision Tree Model")
```

Run the above model and track the logged parameters in the mlflowui.

```
(base) maheshnukala@Maheshs-MacBook-Pro ~ % conda activate Titanic
(Titanic) maheshnukala@Maheshs-MacBook-Pro ~ % cd /Users/maheshnukala/Desktop/IIIT/PartB/
(Titanic) maheshnukala@Maheshs-MacBook-Pro PartB % python Titanic.py
Starting the experiment
Model trained
```



			Metrics				Parameters	
	Run Name	Created	Accuracy for th	f1_score_test	precision_test	recall_test	Test shape	Train shape
	masked-donkey-540	5 seconds ago	0.799	0.799	0.799	0.799	(179, 7)	(712, 7)
	thoughtful-cod-90	57 seconds ago	0.802	0.802	0.802	0.802	(268, 7)	(623, 7)
	languid-turtle-399	2 minutes ago	0.813	0.813	0.813	0.813	(268, 7)	(623, 7)
	persistent-deer-13	6 minutes ago	0.406	-	-	-	(268, 7)	(623, 7)
	bald-rook-623	6 minutes ago	-	-	-	-	(268, 7)	(623, 7)
	capricious-pug-601	14 minutes ago	0.393	-	-	-	(268, 7)	(623, 7)
	awesome-quail-150	14 minutes ago	0.393	-	-	-	(268, 7)	(623, 7)
	dapper-whale-558	17 minutes ago	0.393	-	-	-	(268, 7)	(623, 7)
	enchanting-newt-858	18 minutes ago	-	-	-	-	(268, 7)	(623, 7)
	flawless-mule-533	20 minutes ago	0.813	-	-	-	(268, 7)	(623, 7)
	wise-goat-598	36 minutes ago	0.813	0.813	0.813	0.813	(268, 7)	(623, 7)
	welcoming-perch-783	43 minutes ago	0.813	0.813	0.813	0.813	(268, 7)	(623, 7)

Make some changes in the model training(like-change a few hyperparameters, train size, etc) and train the updated model.

	Run Name	Created	Accuracy for th	f1_score_test	precision_test	recall_test	Test shape	Train shap
	persistent-ox-164	1 minute ago	0.803	0.803	0.803	0.803	(223, 7)	(668, 7)
	masked-donkey-540	19 minutes ago	0.799	0.799	0.799	0.799	(179, 7)	(712, 7)
	thoughtful-cod-90	20 minutes ago	0.802	0.802	0.802	0.802	(268, 7)	(623, 7)
	languid-turtle-399	21 minutes ago	0.813	0.813	0.813	0.813	(268, 7)	(623, 7)
	persistent-deer-13	25 minutes ago	0.406	-	-	-	(268, 7)	(623, 7)
	bald-rook-623	26 minutes ago	-	-	-	-	(268, 7)	(623, 7)
	capricious-pug-601	33 minutes ago	0.393	-	-	-	(268, 7)	(623, 7)

Track the logged parameters of the second version of the model and compare it with the first model.

Titanic ›

Comparing 3 Runs from 1 Experiment

Visualizations

Parallel Coordinates Plot | Scatter Plot | Box Plot | Contour Plot

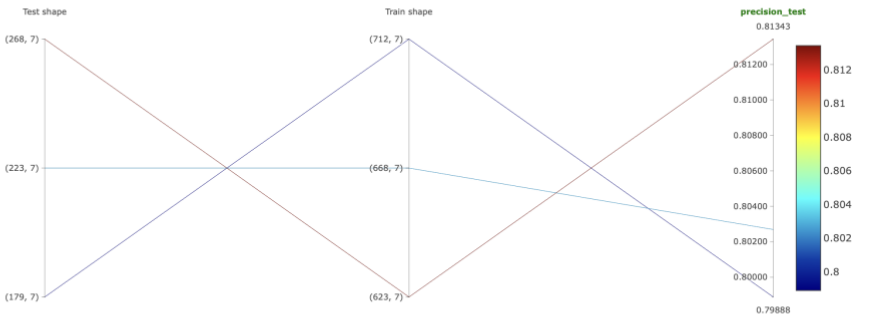
Parameters:

Test shape ✕ Train shape ✕

Metrics:

precision_test ✕

Clear All



Run details

Run ID:	76253d67f6044b44a668b2f62e72c46c	2fa00fe3cba24f4e8be56151753b8444	3d262a4dd909401eb6d03a9cb5353418
Run Name:	persistent-ox-164	masked-donkey-540	wise-goat-598
Start Time:	2022-11-19 10:19:29	2022-11-19 10:01:25	2022-11-19 09:25:11
End Time:	2022-11-19 10:19:32	2022-11-19 10:01:27	2022-11-19 09:25:13
Duration:	2.4s	2.3s	1.6s

Parameters

how diff only

Test shape	(223, 7)	(179, 7)	(268, 7)
Train shape	(668, 7)	(712, 7)	(623, 7)

Metrics

how diff only

Accuracy for this run	0.803	0.799	0.813
f1_score_test	0.803	0.799	0.813
precision_test	0.803	0.799	0.813
recall_test	0.803	0.799	0.813

Tags