

Unit 3.4 Graded Assignment:

Muhammad Khan (2303.khi.deg.027)
Qadeer Hussain (2303.KHI.DEG.006)

Daily Assignment :

Run your Classification model that you trained on the day dedicated to Supervised Learning in MLFlow.

Answer:

[train.py](#)

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.datasets import load_wine
from sklearn.svm import SVC
from sklearn.metrics import f1_score
from sklearn.preprocessing import StandardScaler

def preprocess_data(df: pd.DataFrame):
    scaler = StandardScaler()
    X_train = scaler.fit_transform(df)
    y_train = df.target
    return X_train, y_train

def train_svm(X_train, y_train):
    clf = SVC(kernel='linear')
    clf.fit(X_train, y_train)
    return clf

def evaluate_svm(clf, X_test, y_test):
    y_pred = clf.predict(X_test)
    f1 = f1_score(y_test, y_pred, average='micro')
    return f1

def main(file_name: str):
    df = preprocess_data(pd.read_csv(file_name))
    X_train, X_test, Y_train, Y_test = split_data(df)
    with mlflow.start_run():
        clf = train_svm(X_train, y_train)
        f1 = evaluate_svm(clf, X_test, y_test)
        # Log parameters and metrics with MLflow
        mlflow.log_param("file_name", file_name)
        mlflow.log_metric("f1_score", f1)
        # Save the trained model
        mlflow.sklearn.log_model(clf, "svm_model", registered_model_name="sklearn_svm")
        print(f"F1 Score: {f1}")
        print("Model saved with run_id:", mlflow.active_run().info.run_id)

if __name__ == "__main__":
    file_name = "wine_dataset.csv" # Replace with the actual path to your data file
    main(file_name)
```

Mlflow Production Screen

The screenshot displays the Mlflow Experiments interface. The top navigation bar includes the Mlflow logo, version 2.3.1, and tabs for Experiments and Models. The Experiments tab is active, showing a search bar and a list of experiments. The 'Default' experiment is selected, and its details are shown below the search bar. The 'Description' tab is active, displaying a table of runs. The table has columns for Run Name, Created, Duration, User, Source, Version, Models, Metrics (f1_score), and Parameters (file_name). The runs are sorted by Source and show f1_score values of 1.0 for all runs.

Run Name	Created	Duration	User	Source	Version	Models	Metrics	Parameters
sassy-rook-728	1 hour ago	1.1s	muhamma...	train2.py	10cefo	sklearn_sv_9	1	wine_datas...
secretive-sloth-927	1 hour ago	1.1s	muhamma...	train2.py	10cefo	sklearn_sv_8	1	wine_datas...
gentle-awk-251	1 hour ago	1.2s	muhamma...	train2.py	10cefo	sklearn_sv_7	1	wine_datas...
bright-dove-890	1 hour ago	1.2s	muhamma...	train2.py	10cefo	sklearn_sv_6	1	wine_datas...
legendary-jay-517	1 hour ago	1.2s	muhamma...	train2.py	10cefo	sklearn_sv_5	1	wine_datas...
auspicious-hog-989	1 hour ago	1.1s	muhamma...	train2.py	10cefo	sklearn_sv_4	1	wine_datas...
carefree-moth-6	1 hour ago	1.2s	muhamma...	train2.py	10cefo	sklearn_sv_3	1	wine_datas...
blushing-whale-204	1 hour ago	1.3s	muhamma...	train2.py	10cefo	sklearn_sv_2	1	wine_datas...

WE used the sklearn algorithm SVM.