Project Development Phase
Model Performance Test

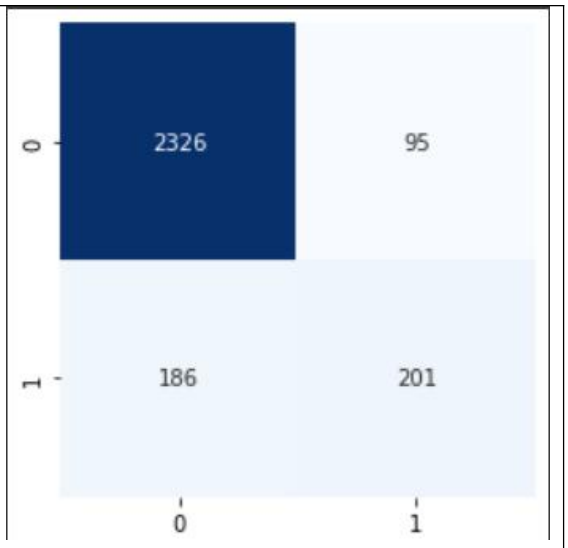
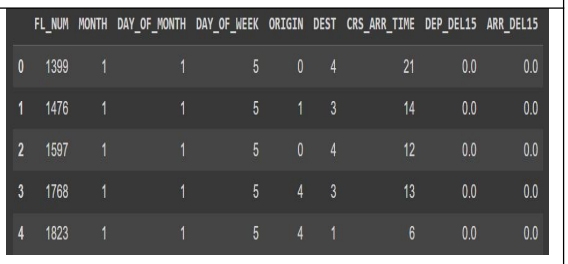
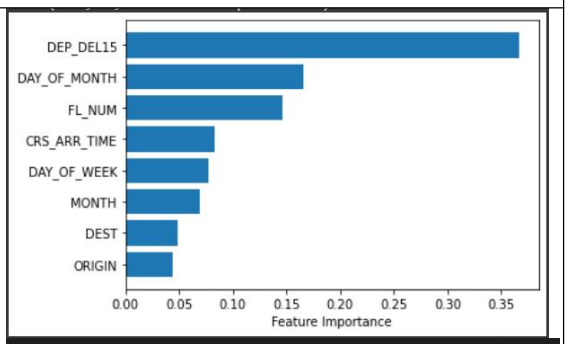
Date	10 November 2022
Team ID	PNT2022TMID17945
Project Name	Developing a Flight Delay Prediction Model using Machine Learning
Maximum Marks	10 Marks

Model Performance Testing:

S.No.	Parameter	Values	Screenshot

1.	Model Summary	Random Forest Classifier	Random Forest Classifier
			`forest_reg = RandomForestClassifier(n_estimators = 10, criterion = 'entropy',random_state = 42)`
			`forest_reg.fit(X_train,Y_train)`
			`model = forest_reg.fit(X_train , Y_train.ravel())`
			`pred = model.predict(X_test)`
			`print('Mean Absolute Error:', metrics.mean_absolute_error(Y_test, pred))`
			`print('Mean Squared Error:', metrics.mean_squared_error(Y_test, pred))`
			`print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(Y_test, pred)))`
			`print("Accuracy:",metrics.accuracy_score(Y_test, pred))`
			`#print("Report for test size = ", test_sz[i], "is shown below:")`
			`print(classification_report(Y_test, pred))`
			`matrix_confusion = confusion_matrix(y_true=Y_test, y_pred=pred)`
			`print(matrix_confusion)`
			`sns.heatmap(matrix_confusion, square=True, annot=True, cmap='Blues', fmt='d', cbar=False)`

2.	Accuracy	Training Accuracy - 89
		Classification Report
			`Mean Absolute Error: 0.10007122507122507`
			`Mean Squared Error: 0.10007122507122507`
			`Root Mean Squared Error: 0.3163403626969298`
			`Accuracy: 0.8999287749287749`
			`precision recall f1-score support`
			`0.0 0.93 0.96 0.94 2421`
			`1.0 0.68 0.52 0.59 387`
			`accuracy 0.90 2808`
			`macro avg 0.80 0.74 0.77 2808`
			`weighted avg 0.89 0.90 0.89 2808`

3.	Metrics	Confusion Matrix	
4.	Label Encoding	Encoding Results	
5.	Feature Selection	Random Forest Regressor	

	FL_NUM	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	ORIGIN	DEST	CRS_ARR_TIME	DEP_DEL15	ARR_DEL15
0	1399	1	1	5	0	4	21	0.0	0.0
1	1476	1	1	5	1	3	14	0.0	0.0
2	1597	1	1	5	0	4	12	0.0	0.0
3	1768	1	1	5	4	3	13	0.0	0.0
4	1823	1	1	5	4	1	6	0.0	0.0

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.inspection import permutation_importance
from matplotlib import pyplot as plt
from sklearn.model_selection import train_test_split

dataset.replace([np.inf, -np.inf], np.nan, inplace=True)
dataset.fillna(999, inplace=True)

X = dataset.iloc[:, 0:8].values
Y = dataset.iloc[:, 8:9].values

X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.25, random_state=42)

rf = RandomForestRegressor(n_estimators=150)
rf.fit(X_train, Y_train)

sort = rf.feature_importances_.argsort()
print(rf.feature_importances_)
plt.barh(dataset.columns[sort], rf.feature_importances_[sort])
plt.xlabel("Feature Importance")
```

6.	UI	Dashboard Design