Initial Model Training Code, Model Validation and Evaluation Report

Project Name : Amazon Instrument Analysis

The dataset was split into **training (75%)** and **testing (25%)** sets.
 Features were extracted using **TF-IDF Vectorization** (unigrams, bigrams, trigrams with 5000 features).

Two machine learning models were trained:

- **Logistic Regression**

- **Random Forest Classifier**

Initial Model Training Code :

```python
#Splitting dataset

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X_final,y_final,test_size=0.25,random_state=42)
```
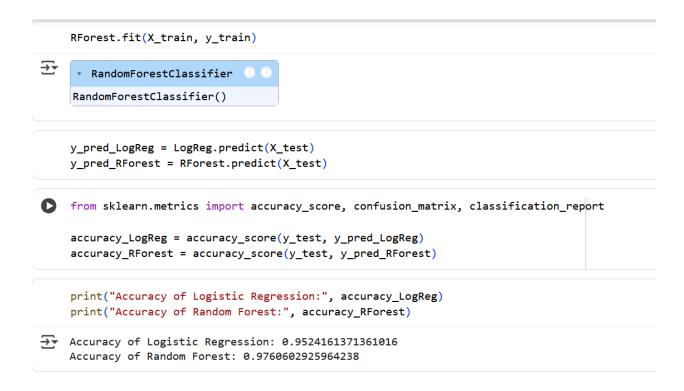
```python
# Model Selection & Evaluation


from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
```

```python
LogReg = LogisticRegression()
RForest = RandomForestClassifier()
```

```python
LogReg.fit(X_train, y_train)
```

```
    ▾  LogisticRegression  ⓘ ⓘ
LogisticRegression()
```

```python
RForest.fit(X_train, y_train)
```

▾ RandomForestClassifier ⓘ ❓
RandomForestClassifier()

```python
y_pred_LogReg = LogReg.predict(X_test)
y_pred_RForest = RForest.predict(X_test)
```

```python
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report

accuracy_LogReg = accuracy_score(y_test, y_pred_LogReg)
accuracy_RForest = accuracy_score(y_test, y_pred_RForest)
```

```python
print("Accuracy of Logistic Regression:", accuracy_LogReg)
print("Accuracy of Random Forest:", accuracy_RForest)
```

Accuracy of Logistic Regression: 0.9524161371361016
Accuracy of Random Forest: 0.9760602925964238

Model Validation and Evaluation Report :

MODEL	Summary	Training and Validation Performance Metrics
Model - 1 : Logistic Regression		

Model - 2 : Random Forest

```python
# Model Selection & Evaluation

from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier

LogReg = LogisticRegression()
RForest = RandomForestClassifier()
```

```python
RForest.fit(X_train, y_train)
```

```
▼ RandomForestClassifier
RandomForestClassifier()
```

```python
y_pred_LogReg = LogReg.predict(X_test)
y_pred_RForest = RForest.predict(X_test)

from sklearn.metrics import accuracy_score, confusion_matrix, classification_report

accuracy_LogReg = accuracy_score(y_test, y_pred_LogReg)
accuracy_RForest = accuracy_score(y_test, y_pred_RForest)
```

```python
print("Accuracy of Logistic Regression:", accuracy_LogReg)
print("Accuracy of Random Forest:", accuracy_RForest)
```

```
Accuracy of Logistic Regression: 0.9524161371361016
Accuracy of Random Forest: 0.9760602925964238
```