Interim 2 Report: Model Building and Training

Overview

This report outlines the process of building and training machine learning models to enhance fraud detection in e-commerce and banking transactions. As a data scientist at Adey Innovations Inc., the goal is to develop accurate fraud detection models by leveraging machine learning techniques and performing comprehensive data analysis.

Business concept

Adey Innovations Inc. aims to improve transaction security by developing advanced fraud detection systems to:

Reduce financial losses due to fraudulent transactions.

Strengthen trust with customers and financial institutions.

Enable real-time fraud detection and rapid response.

The project involves multiple steps, including data analysis, feature engineering, model training, and deployment to ensure continuous improvements.

Data Preparation

Feature and Target Separation

For both datasets, we separated the features and target variables:

Fraud_Data.csv

Features: user_id, signup_time, purchase_time, purchase_value, device_id, source, browser, sex, age, ip_address

Target Variable: class

Creditcard.csv

Features: Time, V1 to V28, Amount

Target Variable: Class

Train-Test Split

We used the train_test_split function from sklearn.model_selection to create training and testing datasets for both data sources. This ensures the model is validated on unseen data.

python

Copy code

```
# Train-test split
X_fraud, y_fraud = fraud_data.drop('class', axis=1), fraud_data['class']
X_train_fraud, X_test_fraud, y_train_fraud, y_test_fraud = train_test_split(X_fraud, y_fraud, test_size=0.2, random_state=42)

X_creditcard, y_creditcard = creditcard_data.drop('Class', axis=1), creditcard_data['Class']
X_train_creditcard, X_test_creditcard, y_train_creditcard, y_test_creditcard = train_test_split(X_creditcard, y_creditcard, test_size=0.2, random_state=42)
```

Model Selection

To explore model performance, the following algorithms were selected for comparison:

Logistic Regression

Decision Tree

Random Forest

Gradient Boosting

Multi-Layer Perceptron (MLP)

Recurrent Neural Network (RNN)

Long Short-Term Memory (LSTM)

Model Training and Evaluation

Training Models

Each model was trained using both datasets. Here is a summary of the process:

python

Copy code

```
# Model training and evaluation
models = {
    'Logistic Regression': LogisticRegression(),
    'Decision Tree': DecisionTreeClassifier(),
    'Random Forest': RandomForestClassifier(),
    'Gradient Boosting': GradientBoostingClassifier(),
    'MLP': MLPClassifier(),
}

results_fraud = {}
results_creditcard = {}

for name, model in models.items():
    model.fit(X_train_fraud, y_train_fraud)
    y_pred_fraud = model.predict(X_test_fraud)
    results_fraud[name] = {
        'Accuracy': accuracy_sc-ore(y_test_fraud, y_pred_fraud),
        'Report': classification_report(y_test_fraud, y_pred_fraud)
    }
    model.fit(X_train_creditcard, y_train_creditcard)

    y_pred_creditcard = model.predict(X_test_creditcard)

    results_creditcard[name] = {

        'Accuracy': accuracy_score(y_test_creditcard, y_pred_creditcard),

        'Report': classification_report(y_test_creditcard, y_pred_creditcard)

    }
```

Evaluation Metrics

Accuracy scores and classification reports were generated for each model to facilitate performance comparisons.

MLOps Steps

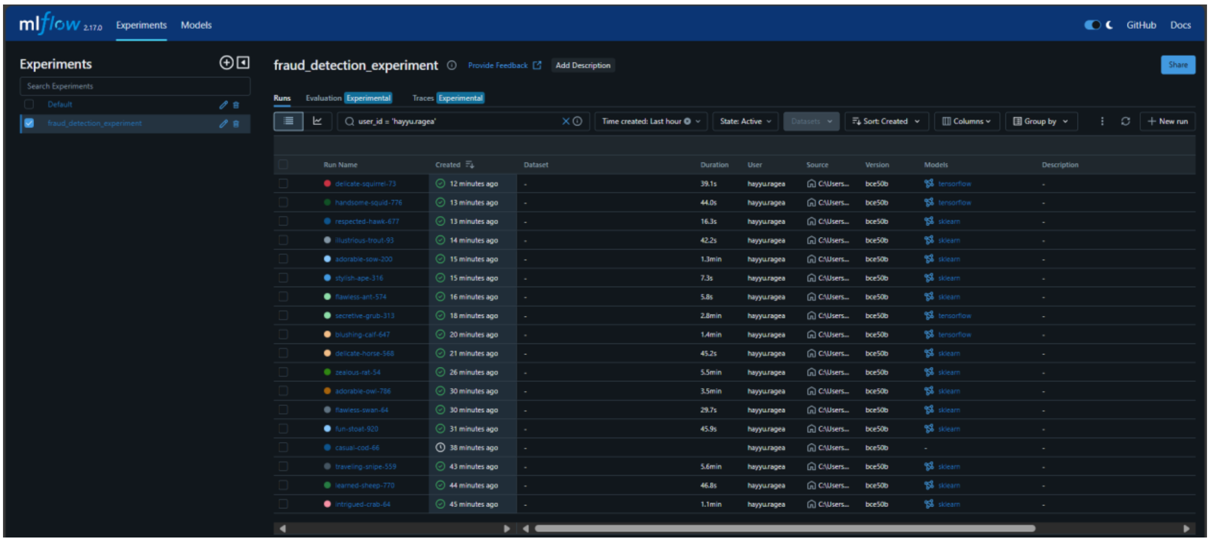
Versioning and Experiment Tracking

To track model performance, parameters, and metrics, we utilized MLflow, which enabled seamless versioning and experiment management.

python

Copy code

```
# MLflow tracking setup

mlflow.start_run()

for name, model in models.items():

    mlflow.log_param("model_name", name)

    mlflow.log_metric("accuracy", results_fraud[name]['Accuracy'])

    mlflow.sklearn.log_model(model, name)

mlflow.end_run()
```



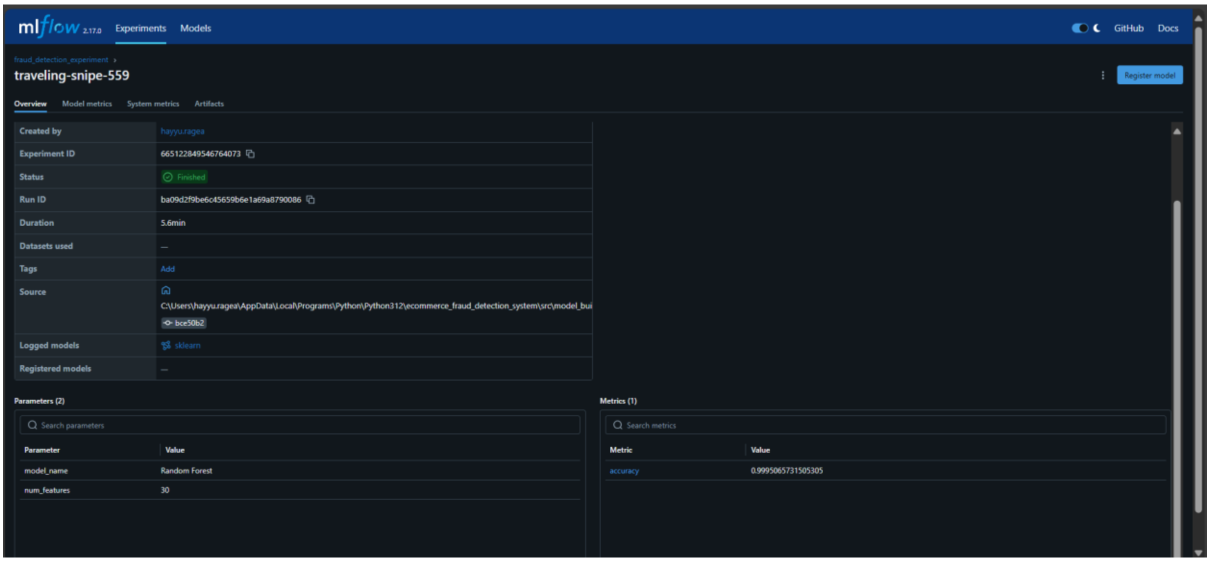
fraud_detection_experiment

user_id = 'hayyu.ragea'

Run Name	Created	Dataset	Duration	User	Source	Version	Models	Description
delicate-squirrel-73	12 minutes ago	-	39.1s	hayyu.ragea	C:\Users...	bce50b	tensorflow	-
handsome-squid-776	13 minutes ago	-	44.0s	hayyu.ragea	C:\Users...	bce50b	tensorflow	-
respected-hawk-677	13 minutes ago	-	16.3s	hayyu.ragea	C:\Users...	bce50b	sklearn	-
illustrious-trout-93	14 minutes ago	-	42.2s	hayyu.ragea	C:\Users...	bce50b	sklearn	-
adorable-sow-200	15 minutes ago	-	1.3min	hayyu.ragea	C:\Users...	bce50b	sklearn	-
stylish-ape-316	15 minutes ago	-	7.3s	hayyu.ragea	C:\Users...	bce50b	sklearn	-
flawless-ant-574	16 minutes ago	-	5.8s	hayyu.ragea	C:\Users...	bce50b	sklearn	-
secretive-grub-313	18 minutes ago	-	2.8min	hayyu.ragea	C:\Users...	bce50b	tensorflow	-
blushing-calf-647	20 minutes ago	-	1.4min	hayyu.ragea	C:\Users...	bce50b	tensorflow	-
delicate-horse-568	21 minutes ago	-	45.2s	hayyu.ragea	C:\Users...	bce50b	sklearn	-
zealous-rat-54	26 minutes ago	-	5.5min	hayyu.ragea	C:\Users...	bce50b	sklearn	-
adorable-owl-786	30 minutes ago	-	3.5min	hayyu.ragea	C:\Users...	bce50b	sklearn	-
flawless-swan-64	30 minutes ago	-	29.7s	hayyu.ragea	C:\Users...	bce50b	sklearn	-
fun-stoat-920	31 minutes ago	-	45.9s	hayyu.ragea	C:\Users...	bce50b	sklearn	-
casual-cod-66	38 minutes ago	-		hayyu.ragea	C:\Users...	bce50b	-	-
traveling-snipe-559	43 minutes ago	-	5.6min	hayyu.ragea	C:\Users...	bce50b	sklearn	-
learned-sheep-770	44 minutes ago	-	46.8s	hayyu.ragea	C:\Users...	bce50b	sklearn	-
intrigued-crab-64	45 minutes ago	-	1.1min	hayyu.ragea	C:\Users...	bce50b	sklearn	-

Evaluating Models for Credit Card Data MLFLOW

- Logistic regression



traveling-snipe-559

Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	ba09d2f9be6c45659b6e1a69a8790086
Duration	5.6min
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	sklearn
Registered models	—

Parameters (2)

Search parameters

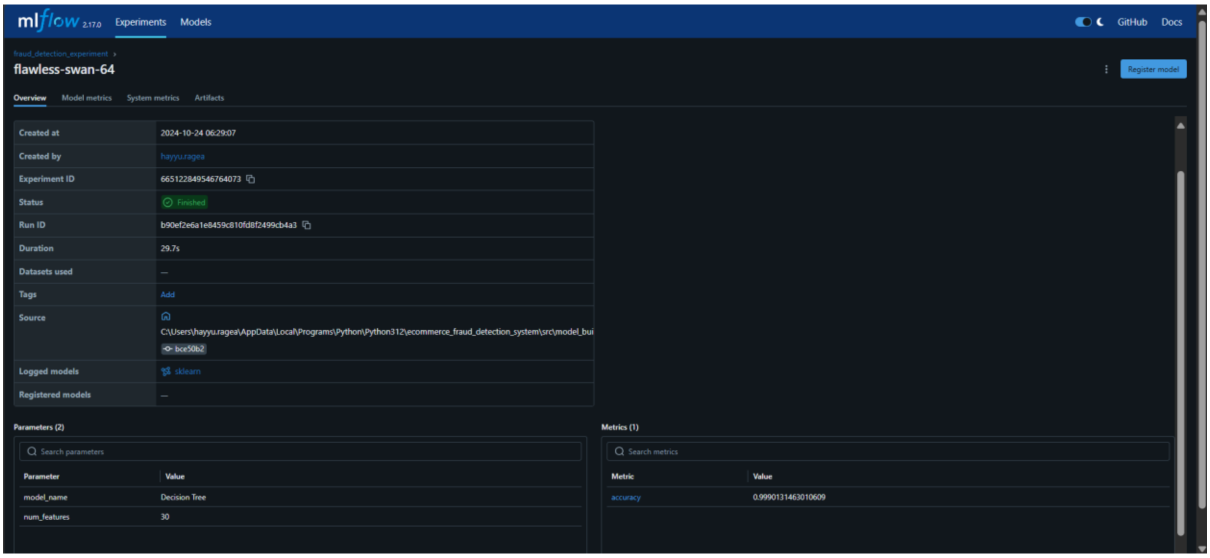
Parameter	Value
model_name	Random Forest
num_features	30

Metrics (1)

Search metrics

Metric	Value
accuracy	0.9995065731505305

- Decision Tree



flawless-swan-64

Created at	2024-10-24 06:29:07
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	b90ef2e6a1e8459c810fd8f2499cb4a3
Duration	29.7s
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	sklearn
Registered models	—

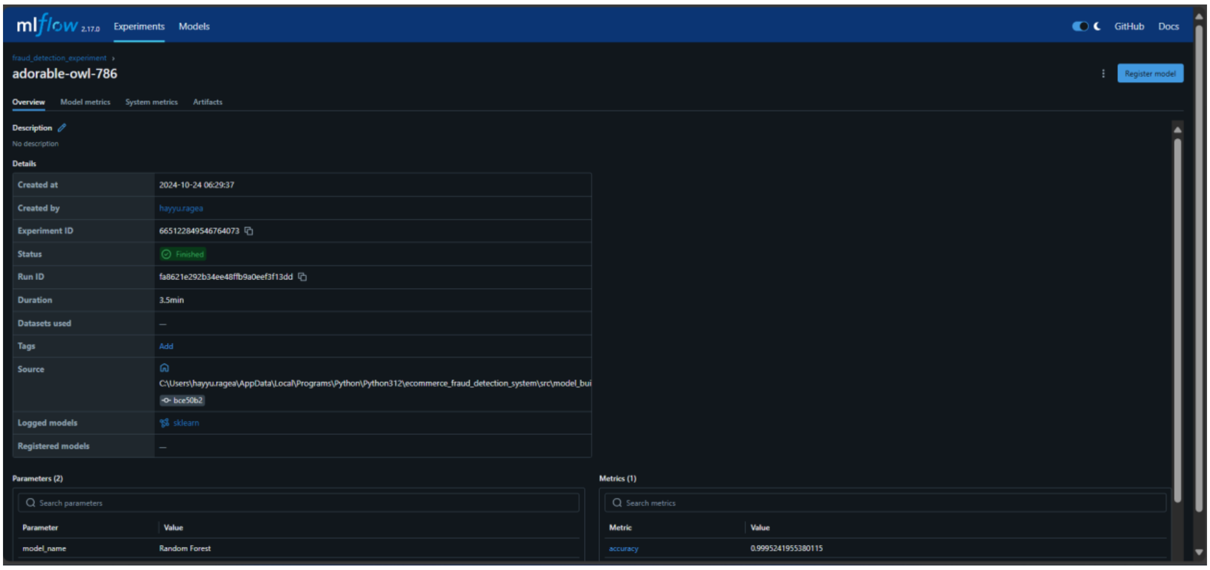
Parameters (2)

Parameter	Value
model_name	Decision Tree
num_features	30

Metrics (1)

Metric	Value
accuracy	0.9990131463010609

- Random Forest



adorable-owl-786

Description

No description

Details

Created at	2024-10-24 06:29:37
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	fa8621e292b34ee48ffb9a0eef3f13dd
Duration	3.5min
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	sklearn
Registered models	—

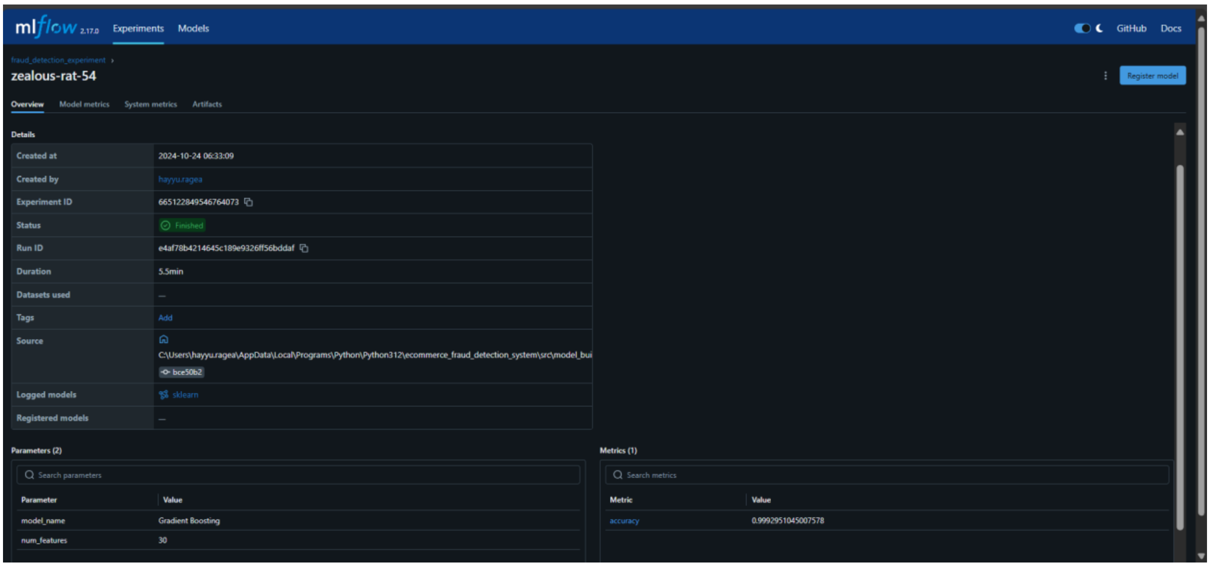
Parameters (2)

Parameter	Value
model_name	Random Forest

Metrics (1)

Metric	Value
accuracy	0.9995241955380115

- Gradient Boosting



zealous-rat-54

Details

Created at	2024-10-24 06:33:09
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	e4af78b4214645c189e9326ff56bddaf
Duration	5.5min
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	sklearn
Registered models	—

Parameters (2)

Search parameters

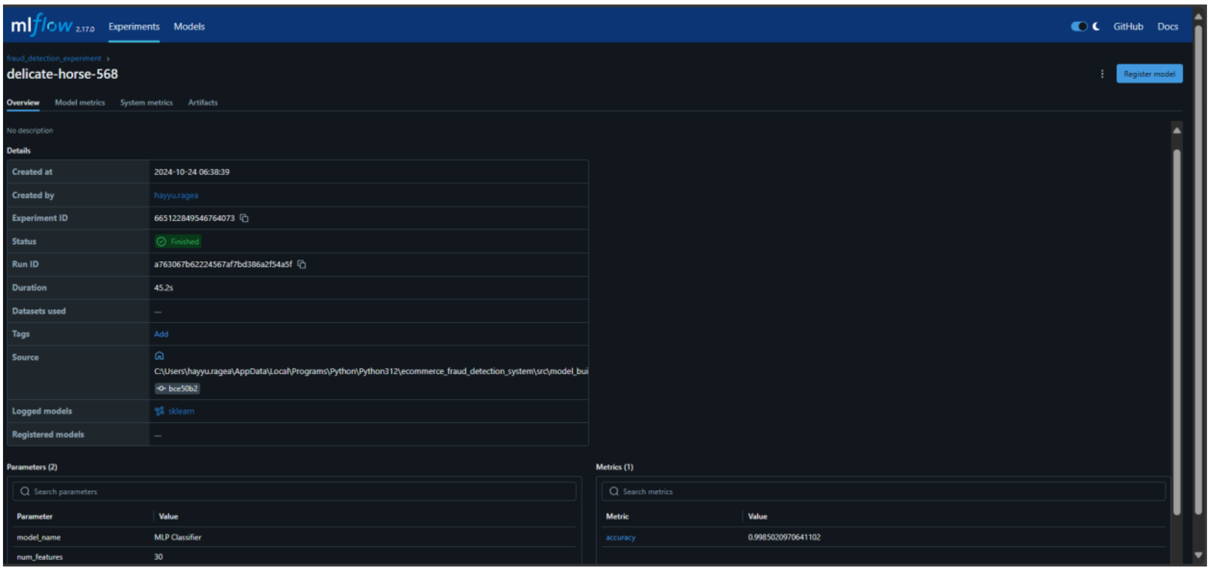
Parameter	Value
model_name	Gradient Boosting
num_features	30

Metrics (1)

Search metrics

Metric	Value
accuracy	0.9992951045007578

- Multi-Layer Perceptron (MLP)



delicate-horse-568

No description

Details

Created at	2024-10-24 06:38:39
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	a763067b62224567af7bd386a2f54a5f
Duration	45.2s
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	sklearn
Registered models	—

Parameters (2)

Search parameters

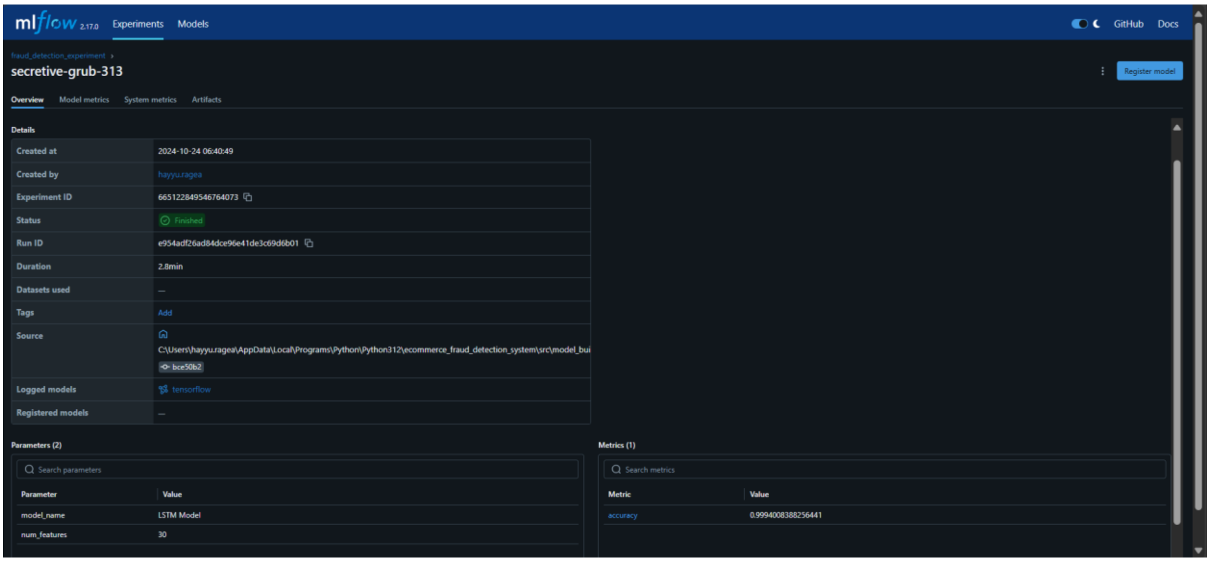
Parameter	Value
model_name	MLP Classifier
num_features	30

Metrics (1)

Search metrics

Metric	Value
accuracy	0.9985020970641102

- Long Short-Term Memory (LSTM)



secretive-grub-313

Details

Created at	2024-10-24 06:40:49
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	e954adf26ad84dce96e41de3c69d6b01
Duration	2.8min
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	tensorflow
Registered models	—

Parameters (2)

Search parameters

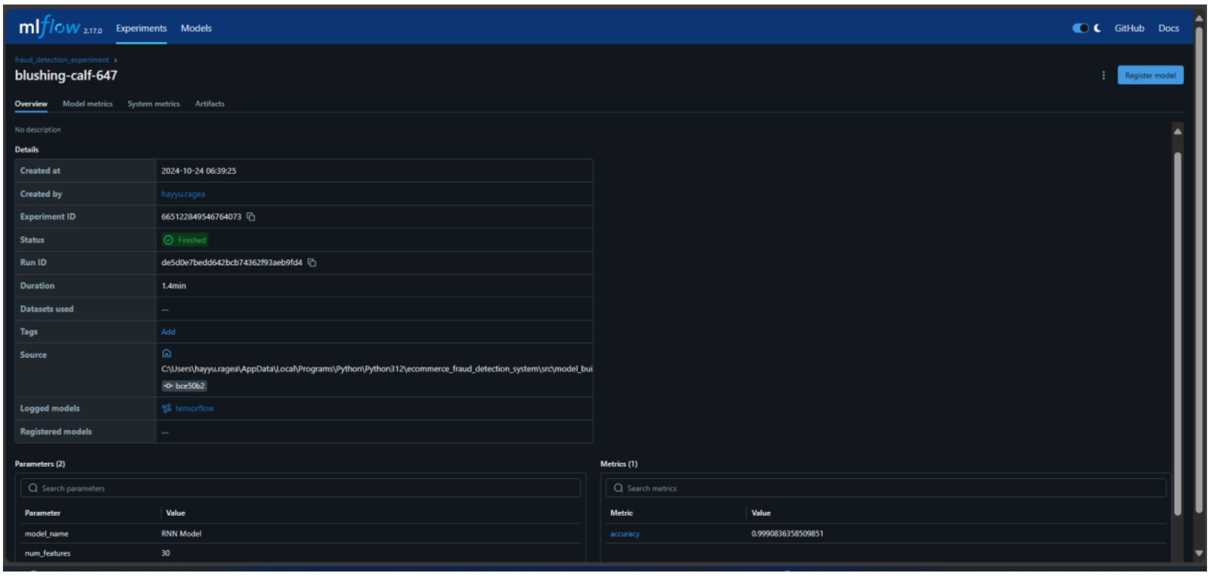
Parameter	Value
model_name	LSTM Model
num_features	30

Metrics (1)

Search metrics

Metric	Value
accuracy	0.9994008388256441

- Recurrent Neural Network (RNN) screenshot



blushing-calf-647

No description

Details

Created at	2024-10-24 06:39:25
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	de5d0e7bedd642bcb74362f93aeb9fd4
Duration	1.4min
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	tensorflow
Registered models	—

Parameters (2)

Search parameters

Parameter	Value
model_name	RNN Model
num_features	30

Metrics (1)

Search metrics

Metric	Value
accuracy	0.9990836358509851

Model Evaluation Results

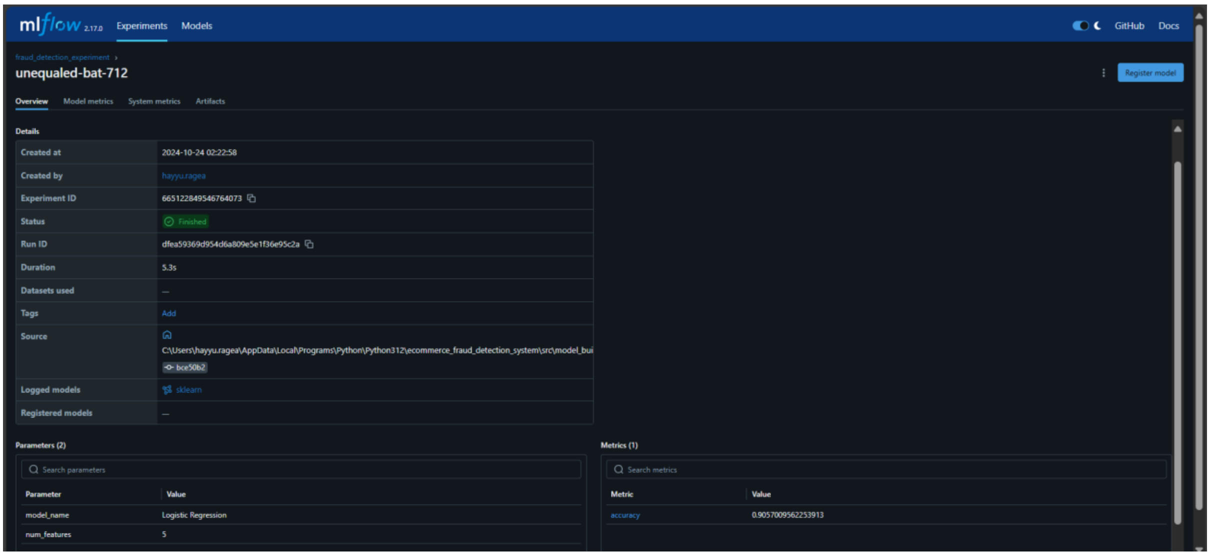
Evaluating Models for Credit Card Data

Model	Accuracy	Precision	Recall	F1-Score	Support
Logistic Regression	0.9991	1.00	0.54	0.66	90

Model	Accuracy	Precision	Recall	F1-Score	Support
Decision Tree	0.9990	0.68	0.72	0.70	90
Random Forest	0.9996	0.99	0.73	0.84	90
Gradient Boosting	0.9993	0.89	0.63	0.74	90
MLP Classifier	0.9982	0.47	0.78	0.58	90
RNN	0.9986	0.91	0.11	0.20	90
LSTM	0.]9992	0.76	0.77	0.76	90

Evaluating Models for Fraud Data MLFLOW

- Logistic Regression



unequaled-bat-712

Details

Created at	2024-10-24 02:22:58
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	dfea59369d954d6a809e5e1f36e95c2a
Duration	5.3s
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	sklearn
Registered models	—

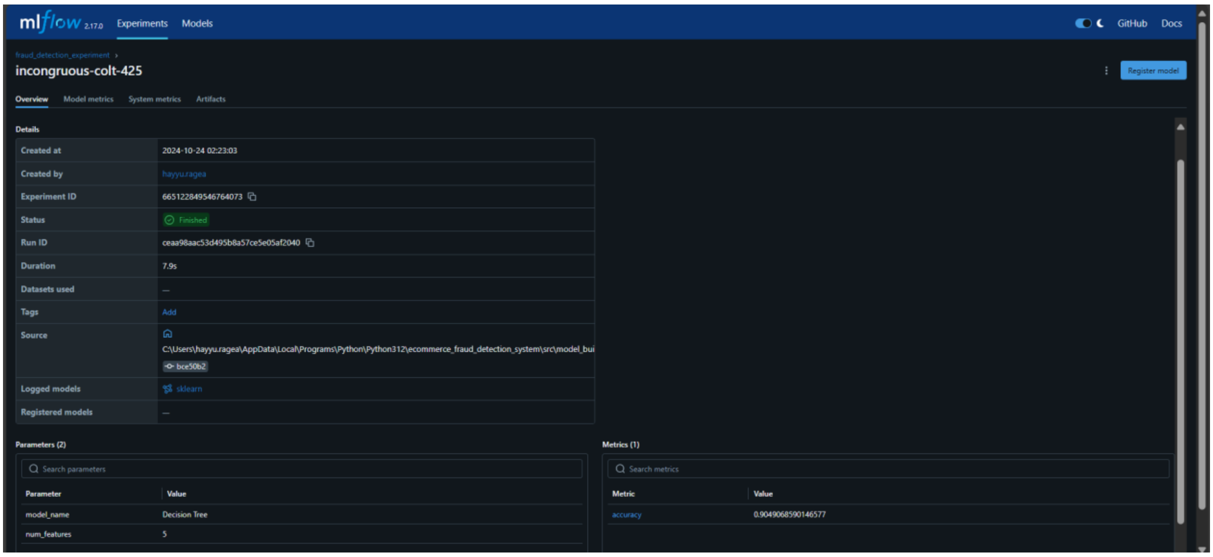
Parameters (2)

Parameter	Value
model_name	Logistic Regression
num_features	5

Metrics (1)

Metric	Value
accuracy	0.9057009562253913

- Decision Tree model



fraud_detection_experiment ›

incongruous-colt-425

Details

Created at	2024-10-24 02:23:03
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	ceaa98aac53d495b8a57ce5e05af2040
Duration	7.9s
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	sklearn
Registered models	—

Parameters (2)

Search parameters

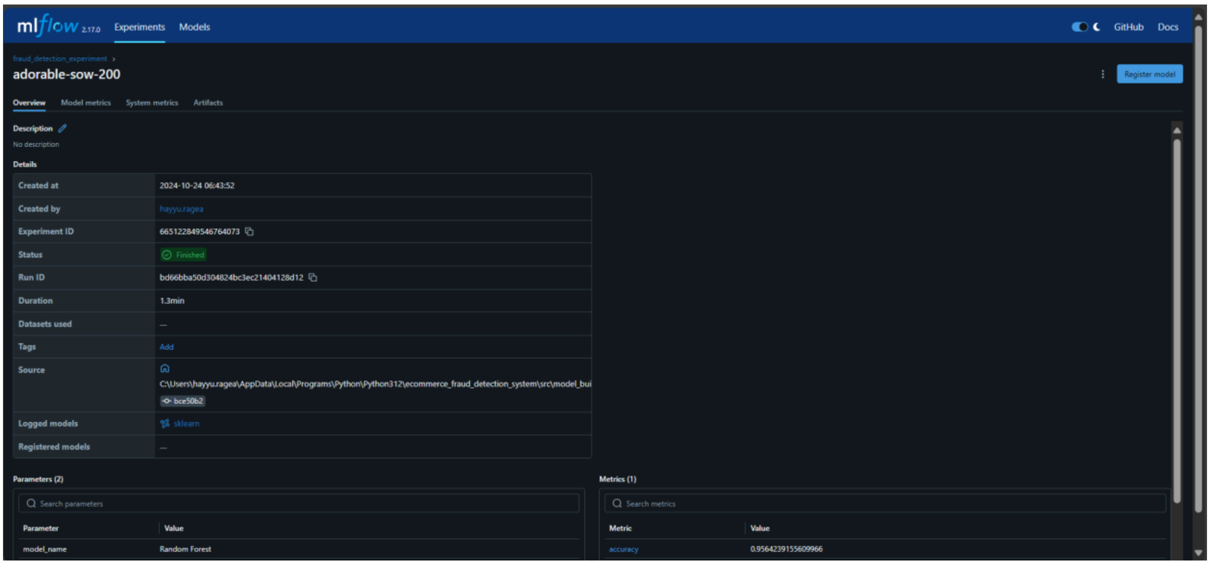
Parameter	Value
model_name	Decision Tree
num_features	5

Metrics (1)

Search metrics

Metric	Value
accuracy	0.9049068590146577

- Random forest



fraud_detection_experiment ›

adorable-sow-200

Description

No description

Details

Created at	2024-10-24 06:43:52
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	bd66bba50d304824bc3ec21404128d12
Duration	1.3min
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	sklearn
Registered models	—

Parameters (2)

Search parameters

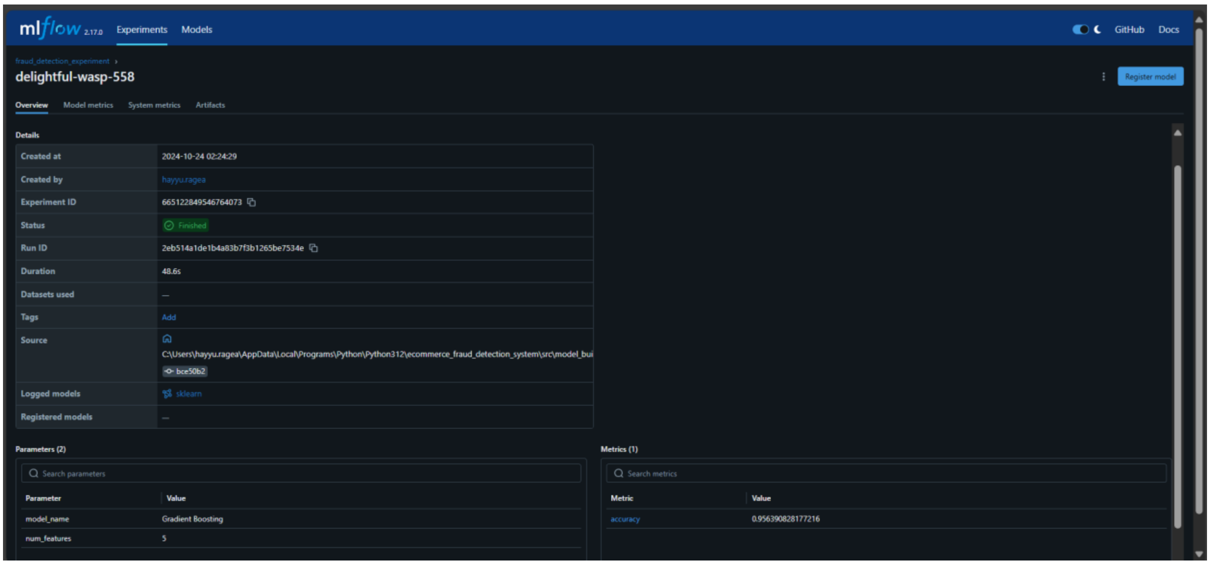
Parameter	Value
model_name	Random Forest

Metrics (1)

Search metrics

Metric	Value
accuracy	0.9564239155609966

- Gradient Boosting MLFLOW



delightful-wasp-558

Details

Created at	2024-10-24 02:24:29
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	2eb514a1de1b4a83b7f3b1265be7534e
Duration	48.6s
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	sklearn
Registered models	—

Parameters (2)

Search parameters

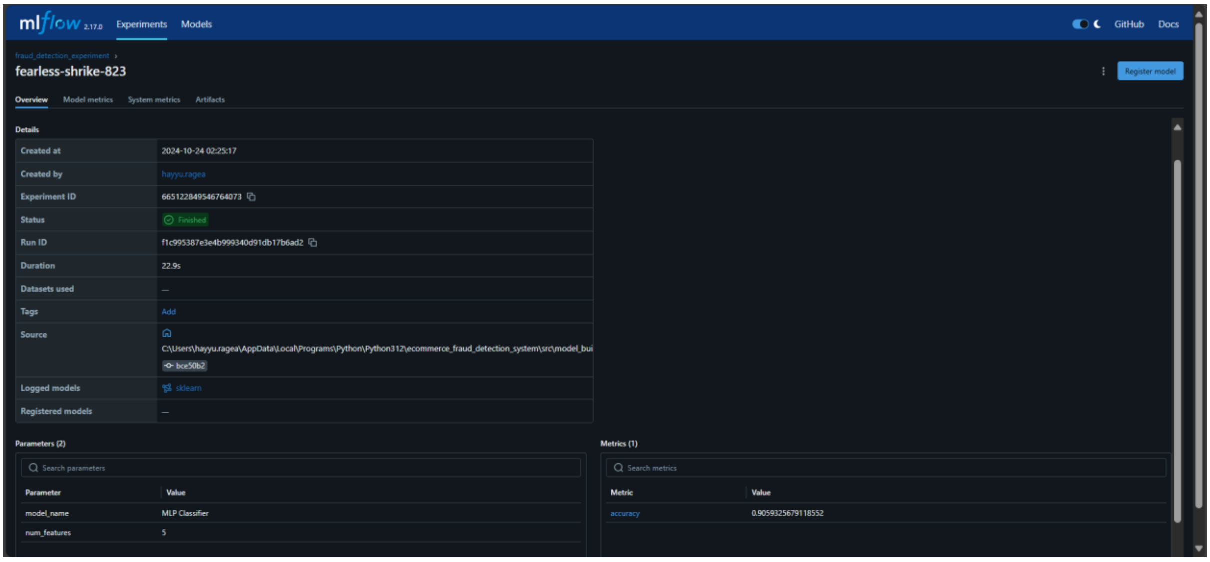
Parameter	Value
model_name	Gradient Boosting
num_features	5

Metrics (1)

Search metrics

Metric	Value
accuracy	0.956390828177216

- Multi-Layer Perceptron (MLP)



fearless-shrike-823

Details

Created at	2024-10-24 02:25:17
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	f1c995387e3e4b999340d91db17b6ad2
Duration	22.9s
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	sklearn
Registered models	—

Parameters (2)

Search parameters

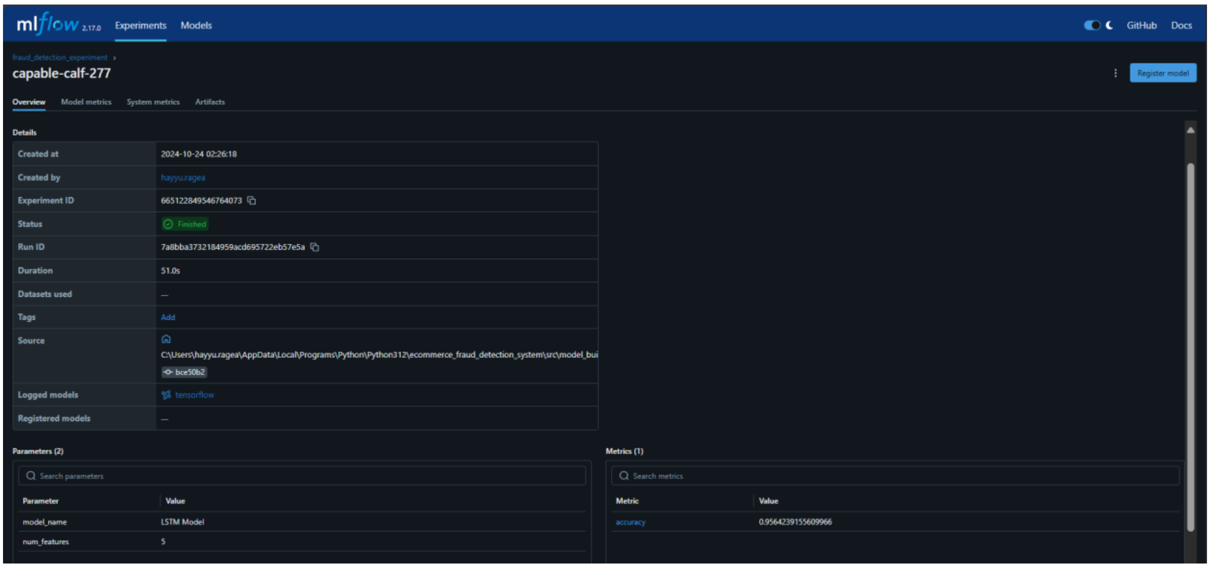
Parameter	Value
model_name	MLP Classifier
num_features	5

Metrics (1)

Search metrics

Metric	Value
accuracy	0.9059325679118552

- Long Short-Term Memory (LSTM)



fraud_detection_experiment ›

capable-calf-277

Details

Created at	2024-10-24 02:26:18
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	7a8bba3732184959acd695722eb57e5a
Duration	51.0s
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	tensorflow
Registered models	—

Parameters (2)

Search parameters

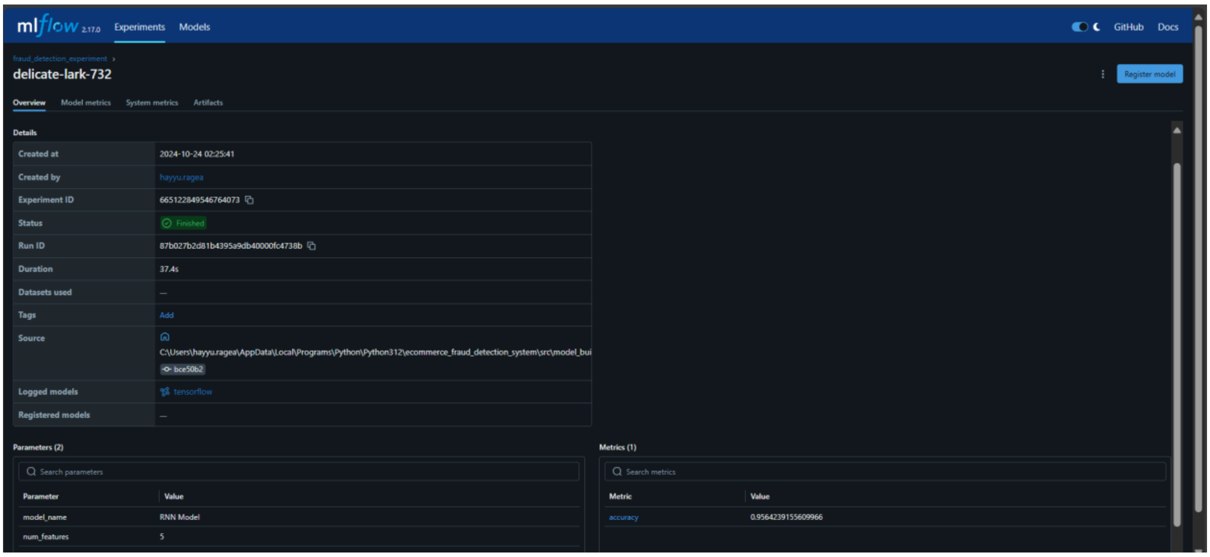
Parameter	Value
model_name	LSTM Model
num_features	5

Metrics (1)

Search metrics

Metric	Value
accuracy	0.9564239155609966

- Recurrent Neural Network (RNN)



fraud_detection_experiment ›

delicate-lark-732

Details

Created at	2024-10-24 02:25:41
Created by	hayyu.ragea
Experiment ID	665122849546764073
Status	Finished
Run ID	87b027b2d81b4395a9db40000fc4738b
Duration	37.4s
Datasets used	—
Tags	Add
Source	C:\Users\hayyu.ragea\AppData\Local\Programs\Python\Python312\ecommerce_fraud_detection_system\src\model_bui bce50b2
Logged models	tensorflow
Registered models	—

Parameters (2)

Search parameters

Parameter	Value
model_name	RNN Model
num_features	5

Metrics (1)

Search metrics

Metric	Value
accuracy	0.9564239155609966

Evaluating Models for Fraud Data

Model	Accuracy	Precision	Recall	F1-Score	Support
Logistic Regression	0.9057	0.00	0.00	0.00	2850
Decision Tree	0.9063	0.50	0.56	0.53	2850

Model	Accuracy	Precision	Recall	F1-Score	Support
Random Forest	0.9564	1.00	0.54	0.70	2850
Gradient Boosting	0.9564	1.00	0.54	0.70	2850
MLP Classifier	0.6179	0.16	0.71	0.26	2850
RNN	0.9564	1.00	0.54	0.70	2850

GitHub link :

Conclusion

The models were trained and evaluated on both datasets, revealing key insights for fraud detection in e-commerce and credit card transactions. Random Forest, Gradient Boosting, and Logistic Regression delivered strong performances. Continued refinement is recommended, including addressing convergence warnings and model optimization.

This project has positioned Adey Innovations Inc. to enhance transaction security and build customer trust through advanced fraud detection technologies.