

INTERMEDIATE PROJECT

Credit Card Fraud Detection

PROGRAM:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, roc_auc_score

# Simulated example (mock dataset since actual CSV is not available)
data = pd.DataFrame({
    'CardID': [1, 2, 1, 2, 3],
    'TransactionTime': ['2024-01-01 10:00:00', '2024-01-01 11:00:00', '2024-01-02 10:00:00',
'2024-01-02 12:00:00', '2024-01-03 10:00:00'],
    'TransactionAmount': [100, 200, 150, 300, 250],
    'MerchantCategory': ['A', 'B', 'A', 'C', 'B'],
    'Fraud': [0, 1, 0, 1, 0]
})

# Preprocess data
data['TransactionTime'] = pd.to_datetime(data['TransactionTime'])
data['TimeSinceLastTransaction'] =
data.groupby('CardID')['TransactionTime'].diff().fillna(pd.Timedelta(seconds=0)).dt.total_seconds
()

# Encode categorical variables
data = pd.get_dummies(data, columns=['MerchantCategory'], drop_first=True)

# Split data
X = data.drop(columns=['Fraud'])
y = data['Fraud']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Standardize numerical features
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

```
# Display the results
X_train, X_test, y_train, y_test
```

OUTPUT :(preprocessed data)

	CardID	TransactionTime	TransactionAmount	Fraud
0	1	2024-01-01 10:00:00	100	0
1	2	2024-01-01 11:00:00	200	1
2	1	2024-01-02 10:00:00	150	0
3	2	2024-01-02 12:00:00	300	1
4	3	2024-01-03 10:00:00	250	0

Split data

X:

	CardID	TransactionTime	TransactionAmount	TimeSinceLastTransaction
0	1	2024-01-01 10:00:00	100	0.0
1				0

PREPARED BY GAMPALA SUREKHA