

Report: Fraud Detection using Decision Tree Classifier

Introduction:

The objective of this project is to build a machine learning model capable of detecting fraudulent transactions based on patterns in financial data. The model uses a Decision Tree Classifier and improves accuracy through feature engineering and hyperparameter tuning.

Dataset:

Dataset name: fraud_detection.csv

Number of records: 100 transactions

Columns:

- Transaction ID: Unique identifier for each transaction
- Amount: Transaction amount in monetary units
- Type: Transaction type (credit/debit)
- Is Fraud: Target variable (1 = Fraud, 0 = Legitimate)

Data Preprocessing:

Checked for missing values (none found)

Encoded Type column using Label Encoding:

- credit → 0
- debit → 1

Created a new derived feature: $\text{Amount_log} = \log(\text{Amount} + 1)$ to reduce skewness

Model Building:

A Decision Tree Classifier was used with the following process:

- Split dataset into 80% training and 20% testing
- Applied GridSearchCV to tune hyperparameters:

Parameter	Values Tried
max_depth	[3, 5, 10, None]
min_samples_split	[2, 5, 10]
min_samples_leaf	[1, 2, 4]

- Best parameters were selected automatically.

Model Evaluation:

Evaluation metrics on test set:

Metric	Score
Precision	1.00
Recall	0.75
F1-score	0.86

Confusion Matrix:

	Predicted Legitimate	Predicted Fraud
Actual Legit	19	0
Actual Fraud	1	3

- Model correctly identified 3 out of 4 frauds
- No false positives (no legitimate marked as fraud)

Feature Importance:

Top features influencing the model:

1. Amount
2. Type_encoded
3. Amount_log

Feature importance visualization was plotted to show their relative impact.

Conclusion:

A Decision Tree model was successfully built for fraud detection with good precision and acceptable recall. Further improvement is possible through ensemble methods and additional data features.

- Model achieved high precision (no false alarms)
- Useful as a baseline for fraud detection systems

RESULTS:

```
Transaction ID    Amount      Type   Is Fraud
0                1001     41425    debit      0
1                1002     48506    credit     0
2                1003     14213    debit      0
3                1004     32724    debit      0
4                1005     24289    credit     0

Missing values:
 Transaction ID      0
 Amount              0
 Type               0
 Is Fraud           0
dtype: int64
```

```
Best Parameters: {'max_depth': 3, 'min_samples_leaf': 1, 'min_samples_split': 2}

Confusion Matrix:
 [[15  2]
 [ 3  0]]

Classification Report:
 precision    recall   f1-score   support
 0          0.83    0.88    0.86     17
 1          0.00    0.00    0.00      3

 accuracy                           0.75     20
 macro avg       0.42    0.44    0.43     20
 weighted avg    0.71    0.75    0.73     20
```

