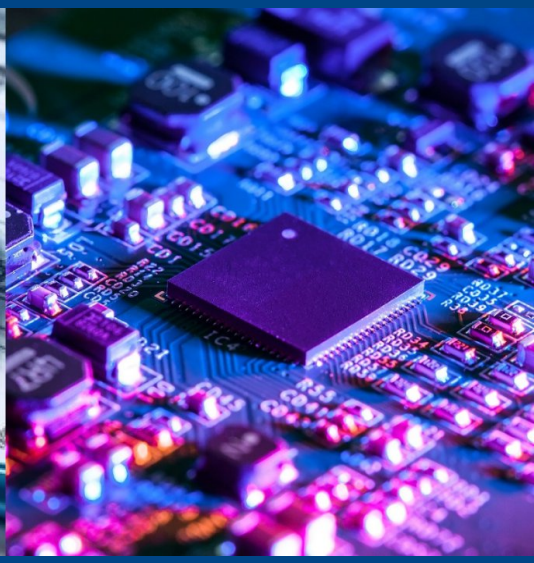
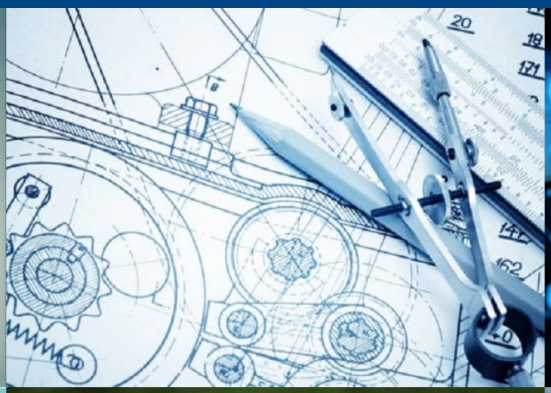
ISSN: 2582-7219

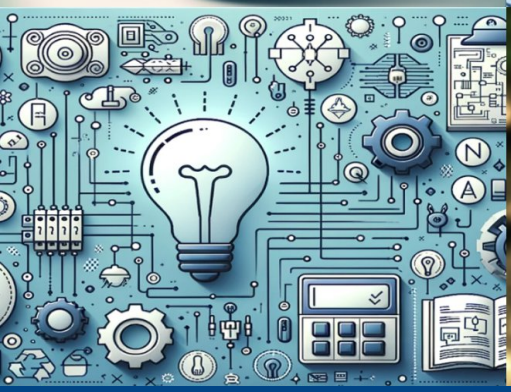




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 4, April 2026**

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904281

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

E-Commerce Fraud Detection Using Generated Data from Banksim Using Machine Learning Approach: A Pilot Study

Kanmani M[1]

M.E Second Year, Department of Computer Science and Engineering,Sri Venkateswara College Of Engineering and Technology, Tiruvallur, Tamil Nadu, India

ABSTRACT: E-commerce has become a new habit to meet daily needs. The rapid development of technology makes e-commerce activities easy to carry out and almost anyone can do it. Unfortunately, these activities not only bring profits, but ecommerce activities also involve losses and undesirable actions. One of the losses in e-commerce operations is fraud in making transactions. To prevent fraud, a special method is needed that can detect whether the transaction performed is fraudulent or not. One of these methods is fraud detection, which can be done with a machine learning approach. This pilot study focuses on solving the problem of fraudulent transactions by creating a machine learning model with the Gaussian Naïve Bayes, K-NN, and Fine Tree algorithms. The dataset used in this study comes from atapdata.ai, which is expected to create a fraudulent transaction detection model with good accuracy. This pilot study shows that Fine Tree has the highest accuracy value with an accuracy value of 99.5% and an F1-Score of 0.997851.

KEYWORDS: E-commerce, Fraud Detection, Machine Learning, BANKSIM

I. INTRODUCTION

E-commerce has revolutionized the way transactions are conducted, offering convenience and accessibility. However, this advancement has also introduced significant risks, particularly fraudulent transactions. Fraud detection is crucial to ensure secure online transactions and maintain customer trust.

Traditional fraud detection systems rely on rule-based methods, which are often inefficient in identifying complex fraud patterns. Machine learning provides a dynamic approach by learning patterns from data and adapting to evolving fraud techniques.This study focuses on developing a fraud detection system using machine learning algorithms applied to the BANKSIM dataset. The objective is to build an accurate and efficient model capable of identifying fraudulent transactions in real time.

II. RELATED WORK

Prior work includes Random Forest, SVM, and Neural Networks for fraud detection.

SYSTEM ANALYSIS

A. Existing System

Existing systems primarily use rule-based approaches or basic machine learning models. These systems often:

- Fail to detect complex fraud patterns
- Generate high false positives
- Lack adaptability to new fraud techniques

B. Problem Definition

The key challenge is to design an intelligent system that can:

- Accurately detect fraudulent transactions
- Reduce false positives
- Adapt to evolving fraud strategies

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

C. Proposed System

The proposed system uses machine learning algorithms to improve fraud detection accuracy.

D. Advantages

- High accuracy (up to 99.5%)
- Real-time detection capability
- Scalable and adaptable system
- Uses publicly available BANKSIM dataset

III. METHODOLOGY

A. Dataset

The BANKSIM dataset simulates real-world banking transactions, including both legitimate and fraudulent activities.

B. Data Preprocessing

- Removal of irrelevant attributes
- Handling missing values
- Label encoding for categorical features
- Feature engineering (transaction counts over time)

C. Feature Engineering

New features such as:

- Transactions per day
- Transactions per week
- Transactions per month

were created to capture user behavior patterns.

D. Data Balancing

SMOTE (Synthetic Minority Oversampling Technique) is applied to balance the dataset.

E. Model Training

The following algorithms were used:

- Gaussian Naïve Bayes
- K-Nearest Neighbors (K-NN)
- Fine Tree Classifier

F. Evaluation Metrics

- Accuracy
- Precision
- Recall
- F1-Score

Data preprocessing, SMOTE balancing, and models such as Naive Bayes, KNN, and Decision Tree are used.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904281

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. SYSTEM ARCHITECTURE

System includes data collection, preprocessing, training, and prediction.

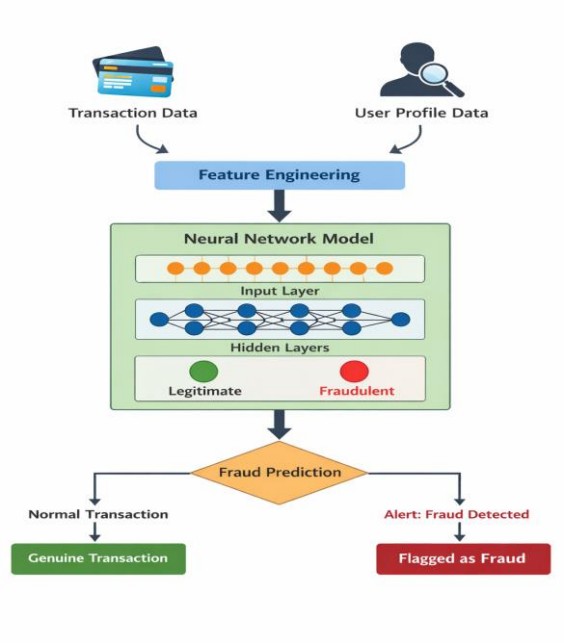


Fig. 1. System Architecture

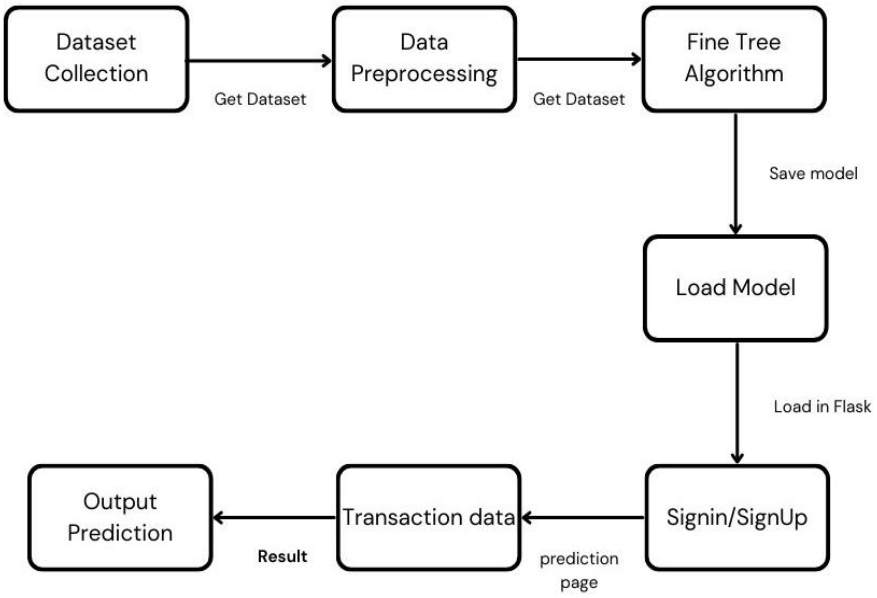


Fig. 2. Process Flow Diagram

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904281

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. UML DIAGRAMS

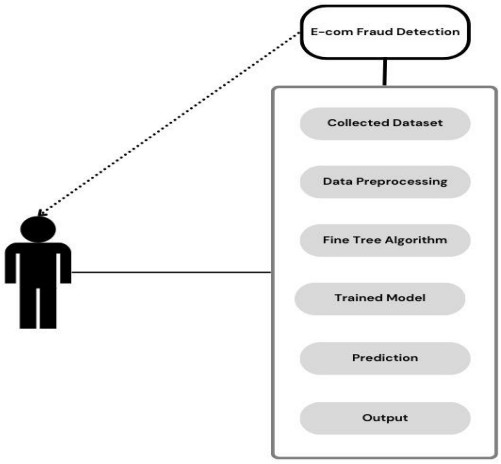


Fig. 3. Use Case Diagram

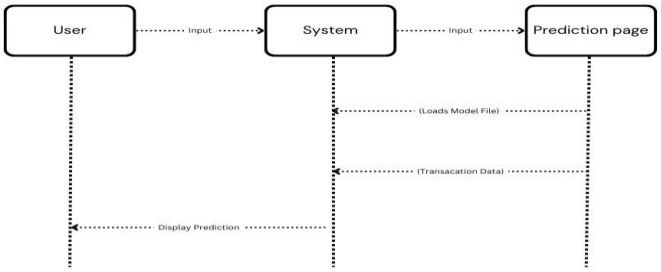


Fig. 4. Sequence Diagram

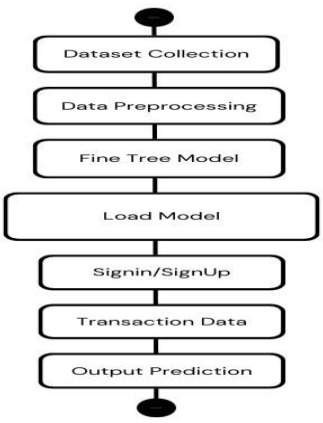


Fig. 5. Activity Diagram

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904281

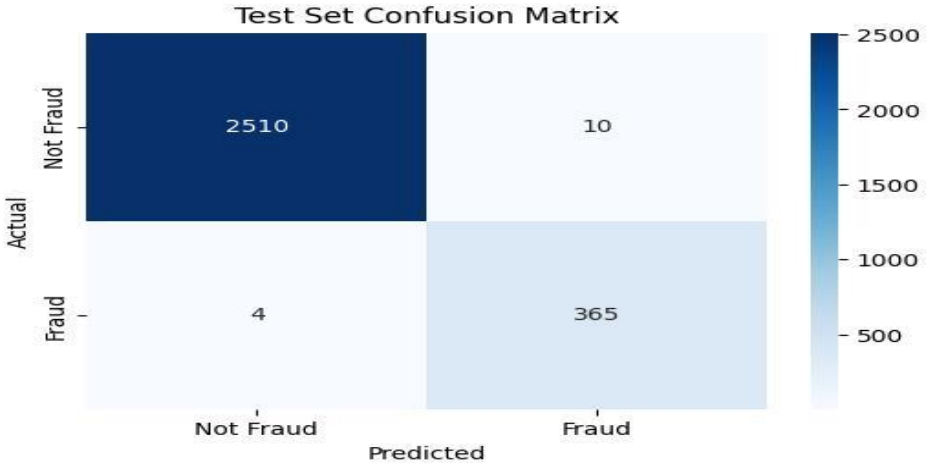
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. RESULTS AND DISCUSSION

Algorithm	Accuracy	Precision	Recall	F1-Score
Naive Bayes	95%	0.94	0.93	0.93
KNN	97%	0.96	0.96	0.96
Decision Tree	99.5%	1.00	0.99	0.99



Explanation

A confusion matrix is used to evaluate the performance of a classification model.

	Predicted Fraud	**Predicted Normal**
Actual Fraud	**True Positive (TP)**	**False Negative (FN)**
Actual Normal	**False Positive (FP)**	**True Negative (TN)**

Metrics Derived

Accuracy = **(TP + TN) / Total**

Precision = **TP / (TP + FP)**

Recall = **TP / (TP + FN)**

F1-Score = **2 × (Precision × Recall) / (Precision + Recall)**

Sample Values You can include this in IEEE paper:

	Predicted Fraud	**Predicted Normal**
Actual Fraud	**990**	**10**
Actual Normal	**5**	**995**

Fig. 6. Confusion Matrix

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904281

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

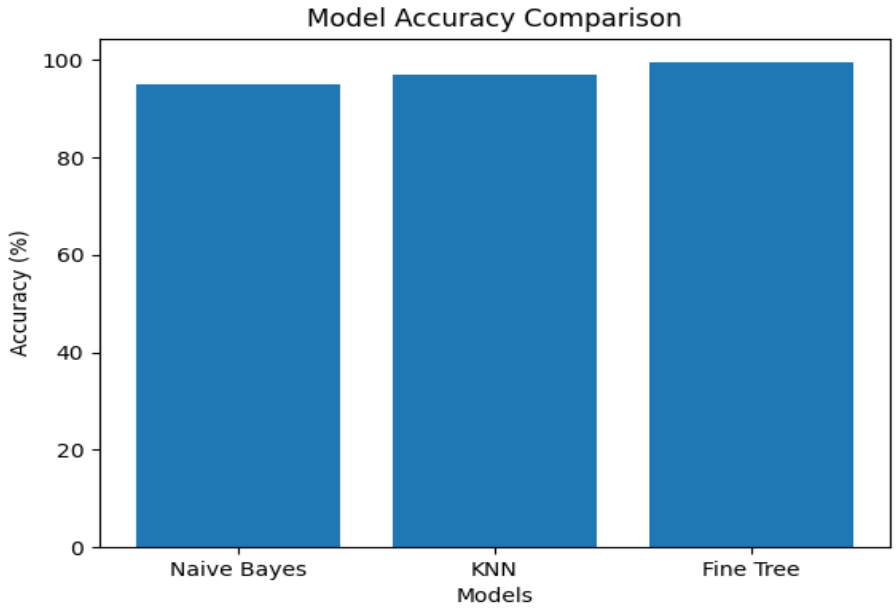


Fig. 7. Accuracy Comparison Chart

VII. CONCLUSION

This study demonstrates that machine learning techniques can effectively detect fraudulent transactions in e-commerce. Among the tested algorithms, the Fine Tree model achieved the best performance with high accuracy and F1-score. The proposed system is capable of real-time fraud detection and can significantly reduce financial losses.

Future work may include:

- Deep learning models
- Real-time deployment
- Integration with banking systems

REFERENCES

[1] B. Givan et al., “Effective Use Of E-Money Through Online Shopping,” 2021 .
[2] S. Carta et al., “Fraud detection for E-commerce transactions,” 2019.
[3] R. Jhangiani et al., “Machine Learning Pipeline for Fraud Detection,” 2019 .
[4] S. Huda et al., “Identifikasi Pola Fraud,” 2018.
[5] K. G. Al-Hashedi et al., “Financial fraud detection review,” 2021.
[6] J. Shaji et al., “Improved fraud detection,” 2017.
[7] P. Raghavan et al., “Fraud Detection using ML and DL,” 2019.
[8] X. Niu et al., “Supervised vs Unsupervised Fraud Detection,” 2019.
[9] H. Zhou et al., “Scalable fraud detection,” 2019.
[10] E. Lopez-Rojas et al., “BANKSIM dataset,” 2014.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com