

Agriculture Cyber Crime Fraud Detection System: A Machine Learning and Exploratory Data Analysis Framework

Parth Sanjay Apte¹, Prof. Tagade R. O.²

¹Student, Department of Computer Engineering, Sau. Sundarbai Manik Adsul Polytechnic, Chas, Ahilyanagar, Maharashtra, India

²Guide / Assistant Professor, Department of Computer Engineering, Sau. Sundarbai Manik Adsul Polytechnic, Chas, Ahilyanagar, Maharashtra, India
Corresponding Email: parthapte444@gmail.com

Abstract:

As digital banking, UPI services, and government subsidy portals become routine tools for farmers across rural India, online security has become a pressing concern. Limited cybersecurity awareness in farming communities has led to frequent financial scams, including fake fertilizer websites, deceptive SMS phishing links, OTP theft, and unauthorized payment transfers. This paper introduces a practical detection framework built to collect, clean, and analyze agricultural fraud incident records. By pairing structured data preprocessing in Python with machine learning classification models, the proposed system differentiates genuine agricultural transactions from fraudulent activity. In addition to flagging suspicious records, the framework compiles regional, bank-level, and attack-vector summaries, providing practical data that can help cybercrime investigators and rural banking institutions protect agricultural livelihoods.

Keywords: Agricultural Cybercrime, Fraud Detection, Machine Learning, Digital Farming Security, Phishing Mitigation, Data Analytics, Pandas.

I. INTRODUCTION

Agriculture is the economic backbone of rural India, supporting millions of families. The rapid expansion of mobile networks and digital platforms has changed how farmers manage daily transactions. Today, farmers regularly use smartphones for direct benefit transfer (DBT) subsidies, online banking, mobile recharges, and UPI apps.

While these tools offer convenience, they also present real risks. Because many rural users have limited familiarity with digital security practices, they are easily targeted by opportunistic cybercriminals. Scammers deploy fake subsidy alerts, deceptive agricultural retail portals, phishing messages, and social engineering tactics to siphon money from rural accounts. The financial losses cause immediate distress and weaken trust in helpful digital farming services.

This project develops a working data analysis and detection pipeline that processes incident records, cleans messy field data, and detects patterns behind common agricultural fraud schemes. Built using Python data science tools, the system provides a practical baseline for automated fraud detection and reporting.

II. LITERATURE REVIEW

A review of relevant literature highlights the challenges of securing digital systems in agriculture:

- **IoT and Agriculture System Vulnerabilities:** Vasilakos and Boulos surveyed cybersecurity threats in smart farming, finding that field-deployed systems often run without standard encryption or security monitoring. Studies by Alahe et al. and Demestichas et al. confirmed that rural network endpoints remain easy targets for malicious actors.
- **Machine Learning in Agriculture:** Liakos et al. and Tripathi & Jha showed that machine learning algorithms reliably extract clear patterns from noisy, unstructured agricultural data.
- **Transaction Fraud Analytics:** Dal Pozzolo et al. demonstrated that data-driven classification methods reliably isolate fraudulent credit card transactions. Adkisson et al. applied similar methods to detect anomalies in connected farm machinery.

While existing studies focus heavily on farm machinery, IoT sensors, or high-volume commercial banking, few look specifically at the payment fraud targeting individual farmers. This project addresses that gap by focusing on the scams affecting daily rural transactions.

III. PROBLEM STATEMENT AND OBJECTIVES

A. Problem Statement

Rural farming households are increasingly targeted by cyber fraud, including fake input suppliers, phishing links, and OTP interception. Because reporting mechanisms in rural areas are often manual and delayed, authorities struggle to spot emerging attack trends quickly. An automated system is needed to centralize incident reports, verify data integrity, and identify fraudulent patterns early to mitigate financial loss.

B. Objectives

1. Identify and categorize the most frequent cyber scams impacting farmers.
2. Build a structured dataset documenting victim demographics, loss amounts, and fraud techniques.
3. Implement a Python data pipeline to clean, format, and handle missing incident data.
4. Train classification models to flag suspicious transactional patterns accurately.
5. Produce practical visual reports (by village, bank, and fraud type) to support investigative action.

IV. SYSTEM ARCHITECTURE AND METHODOLOGY

A. Data Collection and Attributes

The empirical foundation of this study relies on a compiled dataset of 1,000 agricultural fraud records collected and curated for research and simulation purposes across rural administrative blocks. The dataset captures key demographic, transactional, and incident attributes as outlined in Table I.

TABLE I. DATASET SCHEMA SPECIFICATIONS

Attribute	Data Type	Operational Description
Farmer_ID	Integer	Unique anonymized identification code
Age & Village	Mixed	Demographic and regional identifier
Transaction_ID	String	Banking / UPI reference tracking

		token	
Bank_Name	String	Originating	or
		processing financial	
		institution	
Fraud_Type	String	Classification	of
		attack vector (UPI,	
		Phishing)	

B. Preprocessing Pipeline

- 1. Whitespace and Duplicate Removal: String fields are cleaned with .str.strip() and duplicate transactions are removed.
- 2. Imputation: Missing categorical values are imputed using mode values, while numerical outliers are handled via median boundaries.
- 3. Categorical Encoding: Bank names and fraud vectors are transformed into numerical representations for classifier training.

C. Model Training & Validation

The cleaned dataset is split into an 80% training set and a 20% validation set. Supervised learning models (Logistic Regression, Decision Trees, and Random Forest) are trained using scikit-learn. Hyperparameters are tuned to balance precision and recall, ensuring minimal false negatives.

V. RESULTS AND DISCUSSION

A. Incident Vector Distribution

Analysis of the 1,000 incident records identified the distribution of agricultural cyber scams:

- UPI and Payment Scams: 34% of cases (counterfeit QR codes / payment requests).
- SMS Phishing Links: 28% of cases (deceptive scheme enrollment notices).
- Fake Government Subsidies: 19% of cases (coinciding with state disbursement windows).
- Counterfeit Agri-Product Stores: 12% of cases (fake seed/fertilizer suppliers).
- Other Exploits: 7% of cases.

B. Classifier Evaluation Metrics

Performance evaluation confirms that ensemble tree-based models outperform linear baselines on tabular fraud records, as evidenced in Table II.

TABLE II. CLASSIFIER PERFORMANCE COMPARISON

Algorithm	Accuracy	Precision	Recall
Logistic Regression	84.2%	81.0%	79.4%
Decision Tree	88.6%	86.3%	85.1%
Random Forest	91.4%	90.2%	89.5%

C. Operational Insights

The Random Forest model attained the highest detection accuracy at 91.4% with a 90.2% precision rate, demonstrating that transaction amount, time velocity, and attack vector provide decisive discriminatory power. Regional mapping highlighted localized scam hotspots, enabling targeted cyber awareness interventions.

VI. FUTURE SCOPE

Upcoming enhancements include deploying the model as a lightweight regional mobile utility (supporting Marathi and Hindi) and integrating automated verification hooks with the National Cyber Crime Reporting Portal. Deep learning autoencoders will also be evaluated to flag novel, zero-day fraud variants.

VII. CONCLUSION

The Agriculture Cyber Crime Fraud Detection System offers an effective, data-driven approach to protecting farming households from financial fraud. By combining systematic data cleansing routines with machine learning classifiers, the pipeline identifies suspicious transactions with over 91% accuracy. This framework provides an accessible baseline to safeguard digital inclusion for rural communities.

ACKNOWLEDGEMENT

The authors express sincere gratitude to Sau. Sundarbai Manik Adsul Polytechnic, Chas, and the Department of Computer Engineering for providing the necessary facilities, academic resources, and technical guidance throughout this research work.

REFERENCES:

- [1] V. G. Vasilakos and M. N. K. Boulos, 'Cyber Security in Smart Agriculture: A Survey,' *Journal of Agricultural Informatics*, vol. 11, no. 2, pp. 45–58, 2020.
- [2] A. Dal Pozzolo, O. Caelen, and G. Bontempi, 'Data Science-Based Fraud Detection System,' *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3412–3424, 2018.
- [3] R. K. Tripathi and A. K. Jha, 'Artificial Intelligence for Agriculture: A Review,' *International Journal of Computer Applications*, vol. 178, no. 12, pp. 19–25, 2019.
- [4] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, 'Machine Learning in Agriculture: A Review,' *Sensors*, vol. 18, no. 8, p. 2674, 2018.
- [5] S. Wolfert, L. Ge, C. Verdouw, and M. J. Bogaardt, 'Big Data in Smart Farming: A Review,' *Agricultural Systems*, vol. 153, pp. 69–80, 2017.
- [6] K. Demestichas, N. Peppes, and T. Alexakis, 'Survey on Security Threats in Agricultural IoT,' *Computers & Security*, vol. 99, p. 102038, 2020.
- [7] M. A. Alahe, R. Islam, and T. Bhuiyan, 'Cybersecurity in Smart Agriculture,' in *Proc. International Conference on Information and Communication Technology*, pp. 112–117, 2021.
- [8] M. Adkisson et al., 'Autoencoder-based Anomaly Detection in Smart Farming Systems,' *Computers and Electronics in Agriculture*, vol. 182, p. 106012, 2021.
- [9] Indian Computer Emergency Response Team (CERT-In), 'Security Guidelines for Digital Banking and UPI Applications,' Technical Report, MeitY, New Delhi, India.
- [10] Ministry of Agriculture & Farmers Welfare, 'Digital Agriculture Mission Report,' Government of India, 2023.