International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly
Vol. 6 No. 22 March - April - 2025

AI-Driven Fraud Detection in Indian Banking: A Statistical Approach

Ruta Vaidya[1], Snehal Kulkarni[2], Nutan Dhame[3] & Dhanashri Korpad[4]
[1]*Assistant Professor, Haribhai V. Desai College of Arts, Science and Commerce, Pune*
[2]*Assistant Professor, Vishwakarma College of Arts, Commerce and Science, Pune*
[3] *Assistant Professor, Haribhai V. Desai College of Arts, Science and Commerce, Pune*
[4] *Assistant Professor, Vishwakarma College of Arts, Commerce and Science, Pune*
Corresponding Author – Ruta Vaidya
DOI - 10.5281/zenodo.15532628

Abstract:
Online banking is now a pillar of India's digital economy, with millions of transactions happening every day in Indian Rupees (INR). But this growth in digital transactions has also brought with it a growth in online banking fraud, and therefore the need for sophisticated detection systems. Artificial Intelligence (AI) has been found to be a key weapon against fraud by utilizing statistical techniques. This aper examines the way AI-powered systems apply statistical methods like anomaly detection, machine learning, and predictive modeling to identify fraudulent transactions in INR. Real-life examples and case studies of the Indian banking industry are presented to demonstrate the use of such approaches. The paper also discusses the statistical approaches in greater detail, discussing their mathematical basis and real-world implementations.

Keywords: Statistical Methods, Logistic Regression, Random Forest, Artificial Intelligence (AI), Regression Analysis.

Introduction:

India has seen a sharp increase in digital transactions due to efforts such as Unified Payments Interface (UPI), Aadhaar payments, and the spread of mobile banking. The Reserve Bank of India (RBI) reports that digital transactions in India exceeded ₹7,422 lakh crore during 2022-2023. But this growth has also drawn in fraudsters, with fraud cases reported at ₹60,414 crore in 2022. Conventional fraud detection mechanisms are usually not up to the task of dealing with the complexity and magnitude of such frauds.AI based on statistical techniques, provides a strong solution to this issue.

Statistical Methods in AI-Driven Fraud Detection:

Anomaly Detection:

Anomaly detection detects transactions that are far from normal behavior. In the Indian context, this is especially helpful in detecting suspicious transactions in INR.

Techniques Used:

1. Z-Score Analysis: The Z-score is a measure of how many standard deviations a data point is away from the mean. It is given by: $Z = \frac{X-\mu}{\sigma}$

where X is the data point, μ is the mean, and σ is the standard deviation.

Example: If the average transaction value is ₹1,500 with a standard deviation of ₹500, a transaction value of ₹50,000 would have a Z-score of: $Z = \frac{50,000-1500}{500} = 97$

This high Z-score means an anomaly.

2. Isolation Forest: It separates anomalies by choosing features at random and dividing the data. Anomalies are recognized as data points that take fewer splits to isolate.

Example, a transaction of ₹ 50,000 in a data set where all other transactions are under ₹ 5,000 would get isolated quickly, showing potential fraud.

3. Local Outlier Factor (LOF): LOF calculates the local deviation of a point from its neighbors. The greater the LOF, the greater the probability of being an outlier.

Example: A transaction of ₹ 1,00,000 in a neighborhood of transactions with an average of ₹2,000 would have a high LOF, which would be an indication of a fraud.

Machine Learning Algorithms:

Machine learning algorithms study patterns in transactional data to identify fraud. Machine learning algorithms are trained using historical information, such as legitimate and fraudulent transactions in INR.

Techniques Used:

1. Logistic Regression: Logistic regression estimates the probability of a binary outcome (e.g., fraudulent or not) based on one or more predictor variables. The logistic function is expressed as:

$$P(Y=1) = \frac{1}{e^{-(\beta 0 + \beta 1 X1 + \beta 2 X2 + \cdots\ldots\ldots\ldots\ldots + \beta n Xn)}}$$

Where P(Y=1) is the probability of the transaction being fraudulent, and $X_1, X_2, ., X_n$ are the predictor variables.

Example: A transaction of ₹1,00,000 at 3:00 AM on a new device might have a high chance of being fraudulent.

2. Random Forest: Random Forest is an ensemble learning algorithm that aggregates several decision trees to enhance accuracy. Each tree in the forest is trained on a random subset of data, and the overall prediction is obtained by averaging the predictions of all the trees.

Example: A payment of ₹75,000 to a newly added beneficiary can be flagged if the same patterns were seen in previous fraud cases.

3. Support Vector Machines (SVM): SVM determines the hyperplane that optimally separates the classes (fraudulent vs. legitimate) in the feature space. The optimization problem is expressed as:

minimizing $1/2 * ||w||^2$ subject to $y_i (w \cdot x_i + b) \geq 1$ for all i

subject to $y_i (w \cdot x_i + b) \geq 1$ for all i, where:

y_i is the class label of the itch data point (+1 or -1).

x_i is the i-th data point.

b is the bias term.

$w \cdot x_i$ is the dot product of the weight vector and the data point.

Example: A transaction of ₹2,00,000 for luxury items in a foreign country is alerted if the cardholder has no history of international transactions.

Predictive Modeling:

Predictive models employ past data to predict the probability of future fraudulent transactions.

Techniques Used:

1. **Time-Series Analysis:** Analysis of time-series data is concerned with analyzing the data points gathered or documented at regular time periods Application of methods such as ARIMA (Auto Regressive Integrated Moving Average) and Exponential Smoothing to model and forecast future values.

Example: If a customer's monthly average transaction is ₹20,000, a one-time jump to ₹5,00,000 would raise a red flag.

2. Regression Analysis: Regression models the connection between a dependent variable and one or more independent variables.

The linear regression model is as follows:

$Y=\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots.+\beta n Xn+\epsilon$

where Y is the dependent variable, $X_1, X_2, ., Xn$ are the independent

variables, $\beta_0,\beta_{1....},.,\beta n$ are the coefficients, and ϵ is the error term.

Example: A regression model can forecast fraud likelihood based on transaction amount, location, and time.

Real-World Examples in the Indian Context:

1. Case Study: UPI Fraud Detection:

Unified Payments Interface (UPI) is now a widely used payment method in India, with over ₹12.82 lakh crore worth of transactions in March 2023. Nevertheless, UPI Frauds, including phishing and SIM swap scams, are increasingly common.

AI Solution:

An AI system tracks UPI transactions in real-time through anomaly detection and machine learning.

Example: A user who usually transfers ₹1,000 through UPI suddenly transfers ₹50,000 to an unknown account. The system alerts this transaction and sends an OTP verification to the user.

2. Case Study: Credit Card Fraud:

Credit card fraud is a serious problem in India, with losses of ₹155 crore in 2022.

AI Solution:

A bank applies logistic regression to detect fraudulent credit card transactions.

Example: A credit card purchase of ₹2,00,000 worth of luxury items in a foreign country is indicated if the cardholder has no international transaction history.

3. Case Study: Phishing and Account Takeover:

Phishing attacks are usually employed by fraudsters to access bank accounts.

AI Solution:

An AI system examines patterns of logins and raises an alarm on suspicious login activity.

Example: A customer who normally logs in from Mumbai logs in from a different location and initiates a transfer of ₹10,00,000. The system sends an alert and blocks the transaction.

Indian Context Challenges:

1. Large Number of Low-Value Transactions:

India experiences a large volume of low-value transactions (e.g., ₹100 to ₹1,000), and it is difficult to identify legitimate and fraudulent transactions.

2. Unawareness:

Most customers in India are not aware of fraud prevention strategies, so they are susceptible to fraud.

3. Regulatory Compliance:

Banks need to comply with RBI regulations, which demand strong fraud detection systems without sacrificing customer convenience.

Future Directions:

1. Explainable AI:

Building AI models that give satisfactory explanations for risky transactions to foster customer confidence.

2. Integration with Aadhaar:

Riding Aadhaar-based authentication for stronger fraud identification.

3. Real-Time Alerts:

Introduction of real-time SMS and application alerts for high-risk transactions in INR

Conclusion:

Statistically powered, AI-based fraud detection systems are revolutionizing Indian banks' fight against online fraud. Through application of techniques such as anomaly detection, machine learning, and predictive modeling, the systems are capable of detecting and blocking fraudulent transactions in INR. Challenges such as voluminous transactions, unawareness, and regulation, though, need to be overcome for the success of these systems. With India increasingly adopting digital banking, AI

will be at the center of protecting the country's financial system.

References:

1. Reserve Bank of India. (2023). *Annual Report on Banking Fraud*. Retrieved from https://www.rbi.org.in
2. National Payments Corporation of India. (2023). *UPI Transaction Statistics*. Retrieved from https://www.npci.org.in
3. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 1-58.
4. Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
5. Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), 1464-1480.
6. Applied Linear Regression Models, Michael H. Kutner, Christopher J. Nachtsheim,
7. John Neter, and William Li, McGraw-Hill, 5th Edition
8. *Pattern Recognition and Machine Learning*, Christopher M. Bishop, Springer
9. *Statistics for Business and Economics, Paul Newbold, William L. Carver, and Betty Thorne,, Pearson, 9th Edition*