Project 1A: Anomaly Detection System for Market Fraud Detection

Raushan Kumar
Department of Electronics and Instrumentation Engineering
NIT Silchar

January 20, 2026

Abstract

Financial markets are highly dynamic and vulnerable to fraudulent activities and price manipulation. Traditional rule-based monitoring systems often fail to adapt to evolving trading behaviors and generate a large number of false positives. This project presents a machine learning-based anomaly detection system designed to identify unusual trading behavior in near real-time. The system employs statistical feature engineering and unsupervised learning techniques to detect price and volume anomalies, generate anomaly scores, and prioritize alerts. The proposed solution demonstrates an end-to-end workflow from data ingestion to live anomaly detection and alert generation.

1 Introduction

Market manipulation and financial fraud pose significant threats to the stability and integrity of financial systems. With the increasing volume and speed of electronic trading, it has become challenging to manually monitor transactions for suspicious activity. Automated anomaly detection systems provide an effective solution by identifying patterns that deviate from normal market behavior.

This project aims to design and implement a machine learning-based anomaly detection system capable of identifying unusual trading patterns using price and volume data. The system is designed to operate in near real-time and assist analysts in prioritizing high-risk alerts.

2 Objectives

The main objectives of this project are:

- Detect abnormal price and volume movements in financial markets
- Assign anomaly scores to suspicious observations
- Prioritize alerts based on risk level
- Enable live or near real-time anomaly detection
- Reduce false positives through statistical filtering

3 Project Timeline

The project was completed over a duration of 15 days, from initial design to deployment.

Phase	**Duration**
Requirement Analysis & Design	Day 1–2
Data Collection & Preprocessing	Day 3–5
Model Development	Day 6–8
Alerting & Prioritization	Day 9–10
Live Detection Integration	Day 11–12
Testing & Optimization	Day 13–14
Deployment & Documentation	Day 15

4 Development Approach

The project followed an AI-assisted development methodology:

- **80% AI-generated code:** Model scaffolding, preprocessing pipelines, and baseline implementations using tools such as ChatGPT, Claude.ai, and Cursor.
- **20% Customization:** Feature engineering, threshold tuning, alert prioritization, and false positive handling.

5 Tools and Technologies

- Programming Language: Python
- Machine Learning Libraries: Scikit-learn, PyOD
- Data Processing: Pandas, NumPy
- AI Assistance Tools: ChatGPT, Claude.ai, Cursor

6 System Architecture

The anomaly detection system follows a modular pipeline:

1. Market data ingestion
2. Feature engineering and normalization
3. Machine learning-based anomaly detection
4. Anomaly scoring
5. Alert prioritization
6. Live monitoring and reporting

7 Methodology

7.1 Data Processing

Simulated financial market data consisting of price and volume information was used. The dataset was cleaned, normalized, and prepared for feature extraction.

7.2 Feature Engineering

The following features were derived:

- Price percentage change
- Volume percentage change

- Rolling mean of price
- Rolling standard deviation of price

7.3 Anomaly Detection Model

Isolation Forest, an unsupervised machine learning algorithm, was employed for anomaly detection. The model is well-suited for financial data due to its robustness to noise and ability to isolate outliers efficiently.

8 Alert Prioritization

Detected anomalies were assigned risk levels based on anomaly scores:

- **High Risk:** Strong deviation from normal behavior
- **Medium Risk:** Moderate deviation
- **Low Risk:** Mild deviation

9 Live Anomaly Detection

The system supports near real-time detection by sequentially processing incoming data points. Anomaly scores and alert levels are generated instantly, enabling timely investigation of suspicious market activity.

10 False Positive Management

False positives were reduced using:

- Rolling statistical context
- Threshold tuning
- Risk-based alert categorization

11 Results

The system successfully detected injected anomalies in both price and volume data. Alerts were generated with appropriate risk levels, demonstrating the effectiveness of the anomaly detection and prioritization approach.

12 Success Metric

The project successfully achieved the following:

- Functional live anomaly detection system
- Accurate identification of unusual trading behavior
- Reduced false positives through alert prioritization

13 Conclusion

This project demonstrates the application of machine learning techniques for detecting anomalous behavior in financial markets. The proposed system is scalable, adaptable, and suitable for real-world fraud detection and market surveillance applications.

14 Future Enhancements

- Integration with real-time stock market APIs
- Deep learning-based autoencoders
- Dashboard visualization using Streamlit
- Cloud-based deployment