MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3403-3407

FEATURE-DRIVEN ANOMALIES DETECTION IN E-COMMERCE

Dr.P.Sivakumar[1*],
Assistant Professor, Department of Information Technology,
Manakula Vinayagar Institute of Technology, Puducherry, India.
Email: hodit@mvit.edu.in
Siva.I[2],
UG Student, Department of Information Technology,
Manakula Vinayagar Institute of Technology, Puducherry, India.
Sakthi SundaraMoorthi.P[3],
UG Student, Department of Information Technology,
Manakula Vinayagar Institute of Technology, Puducherry, India.
KarthikKaran.K[4],
UG Student, Department of Information Technology,
Manakula Vinayagar Institute of Technology, Puducherry, India.
Thoufiq Shariff.S[5],
UG Student, Department of Information Technology,
Manakula Vinayagar Institute of Technology, Puducherry, India.

Abstract
The project introduces an AI-driven deviation system specially designed for e-commerce platforms to increase transaction security and operational efficiency. In an environment where fraudulent patterns are becoming increasingly complicated, traditional rules-based systems are unable to detect micro or new deviations. To address this, our system utilizes trained machine learning models on structured transaction data to assess the risk level for each transaction. It predicts a constant risk score, classifies the type of deviation, provides a risk category, and proposes appropriate mitigation actions. The solution is designed with a modular architecture, which has a fixed-API-based backend and a React frontend that work together to provide a smooth user experience. Users can upload CSV files containing transaction items, treat them via a full pipeline—including prepricing, functional technique, and prediction—and download the final output with practical visualization. The visualization module includes risk distribution maps to help users quickly understand the nature and extent of potential dangers. This system enables active risk management and helps e-commerce companies to detect fraud, make informed decisions, and build more secure and reliable platforms.

Keywords-Anomaly Detection, E-Commerce Security, Machine Learning,Ensemble Learning, Risk Mitigation, Supervised Learning

I. INTRODUCTION:

E-commerce transactions are the backbone of online shopping platforms, which enable companies and consumers to participate in digital transactions. These systems include a wide range of features, including product selection, CART management, payment processing, and order supply. They alsolude back-end processes such as warehouse tracking and customer assistance and ensure that the transaction is performed evenly. With the growing laps in online shopping, e-commerce platforms have become necessary for all sizes of companies, providing a practical, sharp, and accessible way to buy and sell products globally. However, the rapid development of e-commerce transactions presents many challenges. One of the most pressing questions is how to detect fraud, as scammers develop their techniques to benefit from weaknesses in the constant online systems. Payment fraud, identity theft, and procurement of accounts are common questions that traditional security systems struggle to address effectively. In addition, under high transaction versions, treatment of real-time treatment and maintaining system performance can lead to significant delays or errors. E-commerce platforms should follow privacy data, protect customer information, and provide a safe environment to promote consumer choice. To overcome these challenges, advanced technologies and adaptive safety measures are needed to ensure the integrity and safety of e-commerce transactions.

II.LITERATURE REVIEW:

This research examined ways of detecting fraud online using machine learning algorithms. Various approaches, including monitored, unprotected, and semi-revised learning, were analyzed to identify fraudulent activities. The study focuses on specific model choices for evaluation criteria, functional engineering, and e-commerce. Challenges such as misleading strategy and manipulated data were addressed. The machine learning model was strictly tested on the actual world dataset, which demonstrates their efficiency and adaptability to develop the scam pattern. The study ended by discussing practical challenges and potential future instructions to detect machine learning-based fraud in e-commerce. [1]

This study suggested the method of detecting non-commercial data based on a variational autoencoder (VAE) to address the boundaries of traditional methods in handling complex and large data. VAE effectively retrieved functions and identified deviations by learning the latent distribution of data. The model incorporated an adaptive threshold adjustment mechanism and a mild classification network, which improved the accuracy and strength of the detection. Experimental results demonstrated that this approach improved traditional methods in larger matrices, such as procedures, recalls, F1-score, and ROC-AUC. The findings helped to improve e-commerce safety and provided insight to detect deviations in other domains. [2]A balanced dataset, it proposes a monitored YOLOv3 model with an ROI classifier to detect deviations. For unbalanced datasets with some deviation images, it introduces a semi-converted fast-en and model trained on normal samples using WAGAN-GP. This model detects deviations by analyzing the difference between generated and testing images. Evaluation in real industrial surroundings shows that both models achieve high accuracy and performance in real time. [3]Time series analysis has been a research hot spot in data mining, and it has been very important to identify outliers with time chain data. The letter introduced the principles of detection of the deviations of time, analyzed in detail the algorithms to detect the three times chain-based deviation, and compared the advantages and disadvantages. Finally, it briefly presented the most important developmental directions to detect deviations in the time chain. [4]Detection of deviations for the streaming time chain has been an important issue in real applications, especially in e-commerce as an IT industry. Instead of using traditional border-based methods, this article suggested a limit-free approach using deep learning. This introduced two parallel pipelines: a wise baseline (a nerve network with adaptation stages) and uncontrolled identity (a combination of nervous networks and machine learning algorithms). Intelligent baseline performed well with a periodic time chain, while uncontrolled identity was more effective for less periodic data. This supplement design ends the need for careful threshold setting. Experiments showed that the approach detected accurate predictions and reliable deviations. [5]

III.PROPOSED SYSTEM:

The proposed system is an AI-operated framework for deviations and risk-reducing frameworks designed for e-commerce platforms to analyze transaction data and identify unusual or potentially fraudulent activities. With the increasing complexity and volume of digital transactions, traditional rule-based systems are often inadequate to identify micro or developed fraud. To address this limit, our system utilizes machine learning models to analyze data on transactions and provide intelligent, automated insights to determine. The data on growth pipeline begins with data preparation, which involves handling data cleaning, generalization, and lack of values to ensure high-quality input for modeling. This is followed

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3403-3407

by a functional engineer, where relevant functions are achieved to capture deep patterns and conditions in the data. The system trains many machine learning models—imports XGBOOST and Multi-Layer Perceptron (MLP)—to perform three main functions: predict risk score (regression), classify deviation types, and recommend appropriate molding actions. In order to increase prediction stability and accuracy, a cloth approach is planned and combines production from individual models using strategies such as average and majority mood. This merger ensures strong and generalized results, especially effective in real-world scenarios where cases of deviations are rare and diverse.An important feature of the system is its mitigating motor, which wisely recommends reference-specific tasks such as 'flag for review,' 'inform user,' and 'hold transactions,' and reference-specific tasks such as 'block account.' These reactions detected are chosen based on the species and the severity of the deviation, which ensures targeted intervention without unnecessary disturbance. By combining automation, adaptability, and lecturers, the proposed system provides a scalable and intelligent solution to reduce the risk and increase operating flexibility in modern e-commerce ecosystems.

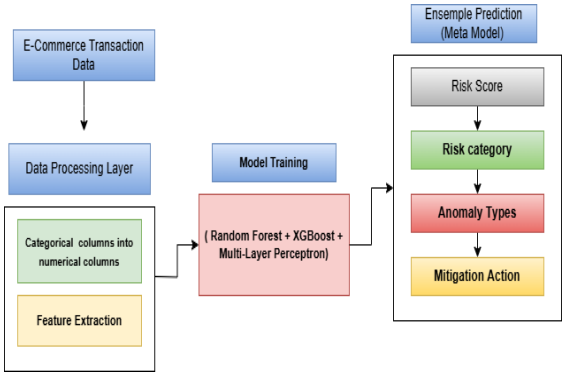


Figure 3.1 Architectural Workflow of E-Commerce Anomaly Detection System

A)Data Collection:The project uses an artificially generated e-commerce dataset designed to reflect the actual world behavior to detect fraud. This includes the main functions of using ID, transaction amount, payment method, units used, order time, and distribution space. These features help identify unusual patterns, such as abnormal units, outdoor transactions, unusual expenses, or unpredictable distribution sites, such as unusual expenses. Although artificial, the dataset effectively imitates the complex behavior required to detect modeling fraud and deviations.

B)Data processing module:The Data Preprocessing and Feature Engineering Module is the important foundation of the paradox detection pipeline, transforming raw CSV transaction facts right into an easy, based, and insightful format suitable for machine studying. Initially, the preprocessing thing ensures record consistency and excellence with the aid of managing missing values (either through imputing or removing them), correcting statistics types, disposing of outliers, and making use of normalization or scaling to numerical capabilities. These steps standardize the dataset, ensuring it is stable and geared up for evaluation. Following this, the feature engineering component enriches the wiped-clean statistics by way of generating meaningful capabilities that enhance version studying. This includes encoding express variables using strategies like one-hot or label encoding, creating interplay capabilities, aggregating transactional information, and applying area-particular logic to derive new, informative attributes. Together, this included module prepares a super, feature-rich dataset that substantially improves the accuracy and robustness of downstream device studying fashions.

C)Model Training Module:The model training module is responsible for creating a future model that uses data sets. It appoints three separate and powerful machine learning algorithms: Random Forest, XGBOOST, and Multi-Layer Perceptron (MLP), contributing unique power to each pipeline. Random forest is a versatile outfit model known for its strength, simple interpretation, and efficiency in both classification and regression functions. XGBOOST provides improved technique, high performance, and efficiency for an advanced shield, especially when large data sets and complex features deal with interactions. MLP, a type of intensive teaching model, captures complex non-session data and provides more flexibility to learn deep patterns. These models are trained to individually make three important forecasts: (1) to score the risk level for each transaction (regression), (2) identify the type of deviation or non-discomfort flow (classification), and (3) recommend appropriate fusion (classification). When the training is completed, the model is sorted using Gablib and stored as .pkl files. This ensures that trained models can be used effectively under estimates without the need to eradicate, which may be compatible with the runtime performance.

D)Prediction Module: The prediction module benefits from an ensemble technique to generate final predictions through a combination of already trained models—random forests, XGBOOST, and MLP output. This hybrid declaration method improves predicting accuracy, stability, and generalization in real interpreters with noise or indeterminate items. Along with determining the types of deviations and suggesting mitigation movements for class tasks, the module uses the majority voting system, where the maximum predicted labels are selected as the final output. For regression tasks such as transaction-determining danger scoring, the average of estimated values from all fashions is calculated to get a balanced and strong chance assessment. This outfit reduces general knowledge of personal fashion weaknesses and complements the system's normal future. By integrating multiple version ideas, the module ensures more reliable, accurate, and general results, which is essential in high-point applications such as detection or analysis of match threats.

IV.RESULT AND DISCUSSION

The Risk Category model shows strong performance overall with high scores: Accuracy (0.9914), Precision (0.9915), Recall (0.9914), and F1 Score (0.9914). Though effective, it might encounter occasional wrong classifications because of overlapping features or slight class imbalance. The Anomaly Type model does better than the Risk Category model scoring higher on all metrics: Accuracy (0.9936), Precision (0.9939), Recall (0.9936), and F1 Score (0.9936). This shows it can identify different types of anomalies with few errors and catch most issues. The Mitigation Action model comes out on top, with steady and excellent results: Accuracy, Precision, Recall, and F1 Score all hit 0.9941. This indicates it makes mistakes when proposing ways to fix problems proving its high-quality features and strong ability to learn.

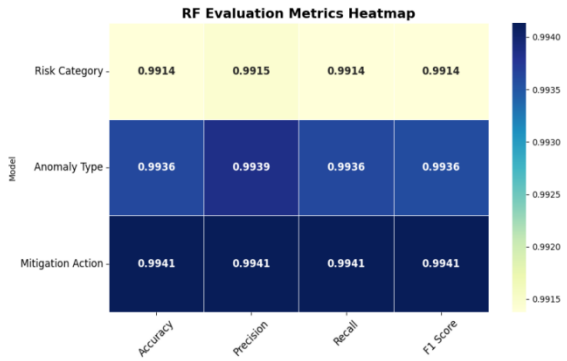
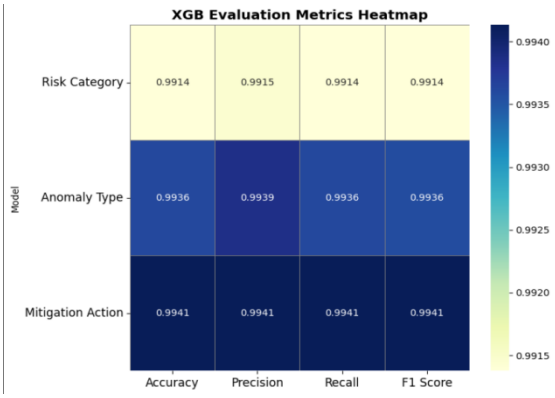


Figure:4.1 : RF Evaluation Metrics Heatmap

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3403-3407

2)XGBOOST:



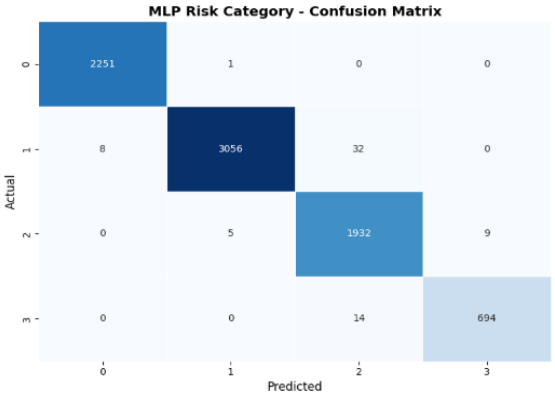
Heatmap reflects the performance of an XGBOOST (XGB) model that was tried on three prediction tasks: risk category, type of deviation and mitigating action. All three tasks are compared to four important matrixes - accuracy, accurate, recall and F1 score - which is much higher than 0.99,indicating the excellent performance of the model. Of the three, the mitigation action model scores the highest on all matrix of 0.9941, which means it has the highest future indicative capacity. The deviation type of work is also very strong, unless, but still strong scores, especially the accuracy of 0.9939. Risk category work, for comparison, which is somewhat backwards, is still very strong with a subsequent high score of approximately 0.9914. As a whole, the XGBOOST model is surprisingly well executed, not much variation in metric values, and shows that it is a very balanced and stable model.

Figure:4.2 : XGB Evaluation Metrics Heatmap

3)Multi-layer Perceptron (MLP):

Confusion matrix gives a wide view of how good MLP (multilayer perception) model predicts different risk categories, with sections marked from 0 to 3 .. There is a high number of correct predictions with diagonal 2.251 for class 0, class 1 for class 1, 3.056 for classes of 1,93, and 3.93, for example, for example. 2, and 14 examples of class 3 are also predicted as class 2. These patterns suggest that models sometimes struggle to distinguish between classes that can share overlapping features or behavioral patterns. Despite these minor errors, the number of abortion is less than the total number of samples, highlighting the model's general efficiency.

Figure:4.3: MLP Risk Category – Confusion Matrix



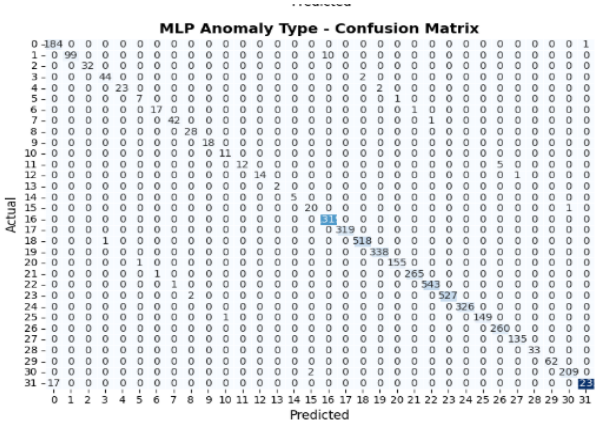


Figure:4.4: MLP Anomaly Type – Confusion Matrix

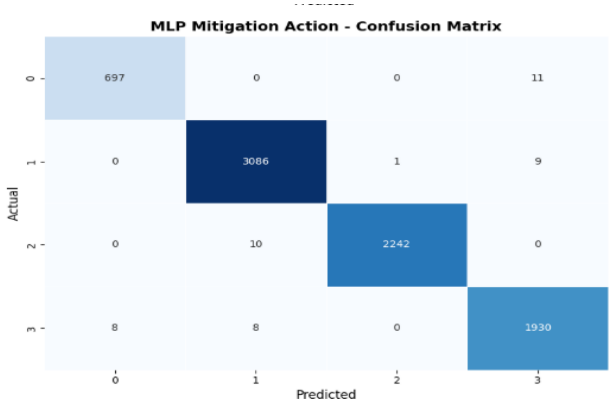


Figure:4.5 : MLP Mitigation Action – Confusion Matrix

Table 4 .1: Evaluation Table:

Models	Overall Accuracy	Risk Category Accuracy	Anomaly Type Accuracy	Mitigation Action Accuracy
Random Forest (RF)	0.9784	0.9723	0.9908	0.9723
XGBoost (XGB)	0.9678	0.9696	0.9625	0.9714
MLP (Neural Net)	0.9921	0.9916	0.9940	0.9908
Ensemble Model	0.9977	0.9970	0.9988	0.9974

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3403-3407

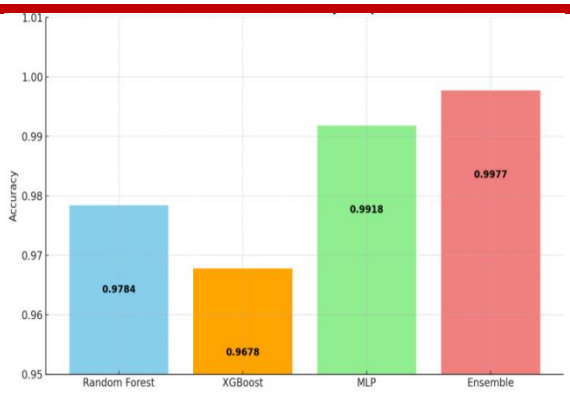


Figure 4.6: Overall Model Accuracy Comparison

PREDICTED OUTPUT:

a)Risk Category

The bar chart visually represents the distribution of transactions in four estimated risk categories: Low risk, no risk, moderate risk and high risk. The x-axis reflects risk categories, while Y-Axha indicates counting of transactions in each. Most transactions occur at low risk (38.8%), followed by no risk (28.1%) and moderate risk (24.3%), with a high risk of 9%. This helps visual stakeholders to explain classification results quickly from the model, consider which categories need to pay more attention, and must make informed resource allocation decisions, especially for the management of high -risk transactions.

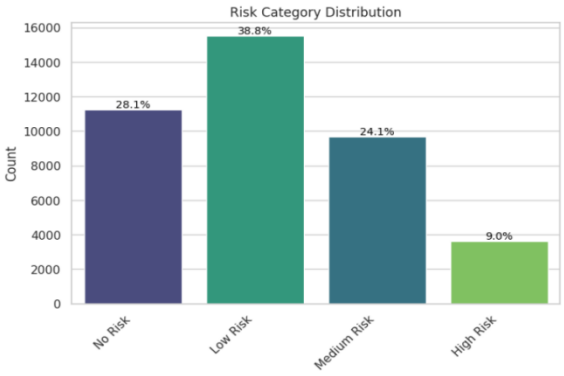


Figure 4.7: Visualization Of Risk Category Bar Chart

b)Anomaly Type

This pie diagram helps stakeholders to understand the distribution of deviations in the dataset. Various deviations in the largest part (31.3%) are classified as "other", while general transactions form 28.1%. The rest are types of specific, action -rich deviations, such as unusual orders or place. This makes it possible to prioritized efforts on the basis of insights and prefer the frequency of discovered deviations.

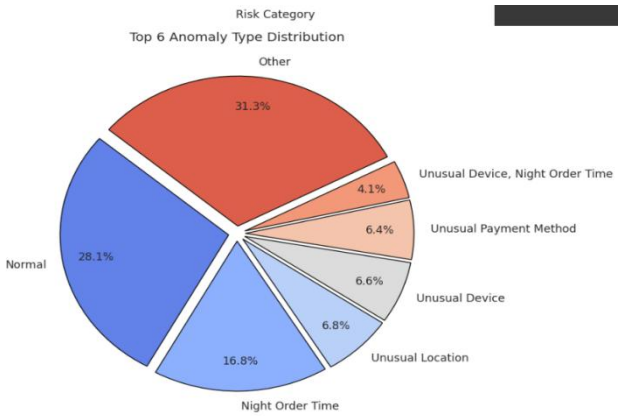


Figure 4.8 : Visualization Of Anomaly Type Pie Chart

c)Mitigation Action:

The "Mitigation vs. Risk" stacked bar diagram suggests how different mitigation actions correspond to different risk levels in transactions. Most actions "monitor transactions and inform the user," mainly for low- and moderate-risk cases. "No action necessary" applies to most non-risk transactions, while "blockage transactions and alert users" are used for high-risk. "Review OTP verification" is marginally used for moderate- and low-risk cases. The diagram has emphasized how the effort for mitigation to ensure proper safety responses is combined with the severity of risk.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3403-3407

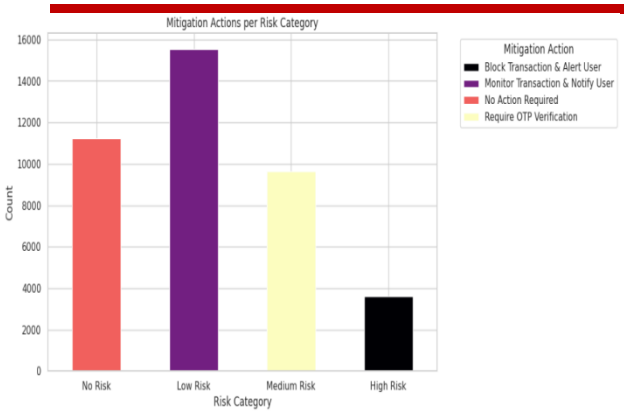


Figure 4.9: Visualization Of Mitigation Vs Risk - Stacked Bar

V.Conclusion and future work

In summary, the project shows an irregular or suspected behavior in transactions to detect an AI-based system with a high ability to match e-commerce platforms. By using advanced machine learning techniques such as random forests, XGBOOST, and multi-layer perceptrons, work in a common or dress approach to improve the quality of system prediction and capture several types of deviations. It classifies each transaction in categories based on the risk level in a smart way and recognizes what kind of non-affecting or fraud it can represent and suggests how to answer or reduce the problem. The model shows very high accuracy—the cline should be perfect in tests—and proves effective in evaluation. The design of the system is modular, which means that each part of the process (such as data cleaning, convenience choice, decision-making, and results visualization of the model) is separated and can easily be updated or scaled. Furthermore, the system can be improved by adding live data handling with devices such as Kafka or Kick streaming to monitor real-time transactions. In addition, adding clarification tools such as size or LIME can help users or analysts understand why the model makes some decisions. In the future, the model can help make the automatic setting system smart in the settings without much effort by using devices such as Optuna or Gridsearchcv. The system can also be expanded to use a variety of inputs, such as text-based patterns or user behavior trends, and may include learning methods that are adjusted over time to detect new or changed fraud styles. These will better equip future update systems, open decisions, and determine and support safe and intelligent digital trading platforms.

Reference

[1.] Geetha Manoharan, S Dada Noor Hayath Ali, Dr. Manoj Sathe, A Karthik, Amandeep Nagpal, Ajay Sidana, “Fraud Detection in E-commerce Transactions: A Machine Learning Perspective” International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2024.

[2.] Jia Li, Shaojiang Liu, Jiajun Zou, “E-Commerce Data Anomaly Detection Method Based on Variational Autoencoder” 3rd International Conference on Artificial Intelligence, Internet of Things and Cloud Computing Technology (AIoTC), 2024.

[3.] Yu Jiang, Wei Wang, Chunhui Zhao, “A Machine Vision-based Realtime Anomaly Detection Method for Industrial Products Using Deep Learning” Chinese Automation Congress (CAC), 2019.

[4.] Zhiyang Zhao, Yang Zhang, XianXun Zhu, Jiancun Zuo “Research on Time Series Anomaly Detection Algorithm and Application” Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2019.

[5.] Jing Zhang, Chao Wang, Zezhou Li, Xianbo Zhang “Threshold-free Anomaly Detection for Streaming Time Series through Deep Learning” International Conference on Machine Learning and Applications (ICMLA), 2021.

[6.] Qingqing Fang, Qinliang Su “Weakly Supervised Anomaly Detection by Utilizing Incomplete Anomaly Information” International Conference on Algorithms, Computing and Artificial Intelligence (ACAI), 2024.

[7.] Yu Jiang, Wei Wang, Chunhui Zhao “A Machine Vision-based Realtime Anomaly Detection Method for Industrial Products Using Deep Learning” Chinese Automation Congress (CAC), 2019.

[8.] Cheong Hee Park, “Anomaly Pattern Detection on Data Streams” International Conference on Big Data and Smart Computing (BigComp), 2018.

[9.] Zhuang Li, Ye Zhang “Hyperspectral Anomaly Detection Based on Improved RX with CNN Framework” IEEE International Geoscience and Remote Sensing Symposium, 2019.

[10.] Sahrul Mulia Siregar, Yudha Purwanto, Suryo Adhi Wibowo “Enhancing Network Anomaly Detection with Optimized One-Class SVM (OCSVM)” International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), 2023.