International Journal of Communication Networks and Information Security
2024, 16(4)
E-ISSN: :2073-607X

https://https://ijcnis.org/

Research Article

Explainable AI Framework for Policy-Compliant Anomaly Detection in Data Pipelines

Sushma Babburi
Independent Researcher, USA.

Abstract

AI design for detecting anomalies in data pipelines with the Isolation Forest algorithm has been done in this research. The purpose of the study is to increase the interpretability and performance of anomaly detection models based on the issues of precision and recall. The model has a high accuracy that is 90.4% that shows its ability to correctly detect anomalies with high levels of false positives and false negatives. The metrics used to evaluate the worth of the assessments, such as the precision-recall curve and F1-score, indicated threshold optimization as a requirement. The results propose that a future study will aim at combining better models, real-world data, and real-time anomaly detection methods and methods to enhance their practical applications

Keywords: Anomaly Detection, Explainable AI, Isolation Forest, Data Pipelines, Model Interpretability, Outlier Detection, Machine Learning, Threshold Optimization, Precision-Recall Curve, Real-time Anomaly Detection, Data Governance.

I. INTRODUCTION

Over the past few years, automated decision-making processes using machine learning (ML) have been embraced in the interpretability and transparency of automated anomaly detection processes. Explaining why an anomaly has been raised is important in ensuring that organizations can be responsible and adhere to the governance policies. The study presents an Explainable AI Framework of Policy-Aligned Anomaly Detection with pipelines in data flow and attempts to form the link between sophisticated ML models and regulation mandates [1]. The model suggests applying interpretable ML models that melodramatically associate the occasions of anomalies that occur with data governance policies, giving a clear explanation of the rationale for each anomaly occurrence flag.

The existing anomaly detection models are not frequently interpretable and it is difficult for organizations defend the decisions taken, particularly in regulated businesses [2]. The absence of research on the implementation of policy alignment and explain ability in real-time data systems, combined with the transparency requirements, substantiates the necessity to have clear models. This study aims to solve the issue of making sure that anomaly detection models are interpretable, as well as in line with the policies of data governance, which will increase the perception of trust and adherence to data-driven decisions.

Aims and Objectives

Aim:

The aim of this research is to build an explainable AI framework to identify anomalies in data pipelines that is congruent with data governance policies, such that it includes transparency, accountability, and compliance in automated decision-making.

Objectives:

- To develop comprehensible machine learning models that provide explanations for the detected anomalies within the pipelines of data.
- To include the policy of data governance in the process of anomaly detection, it is necessary to maintain adherence to the industry regulations and standards.
- To use the explainable AI framework in making anomaly detection systems more transparent and more likely to be trusted by users.

- To recommend the ideal practices to adopt to use explainable anomaly detection models in real-time data systems, accuracy, and policy alignment.

II. NOVEL CONTRIBUTION

The research is a unique contribution as it introduces explainable artificial intelligence to anomaly detection pipelines in line with formal policies on data governance. The presented framework in contrast with the traditional methods of anomaly detection explicitly associates the identified abnormalities with the feature-level explanations that are interpretable and allow accountability and transparency in regulation. This work contributes to the use of Isolation Forest by introducing interpretability behind SHAP, which can be clearly comprehended by non-technical stakeholders as well as technical stakeholders when making decisions about anomalies. The framework also has a distinctive focus on compliance with policies as one of the design principles and not a post-deployment consideration. This contribution can fill a very important research gap as it can combine explain ability, governance, and anomaly detection into one operational data pipeline architecture.

III. LITERATURE REVIEW

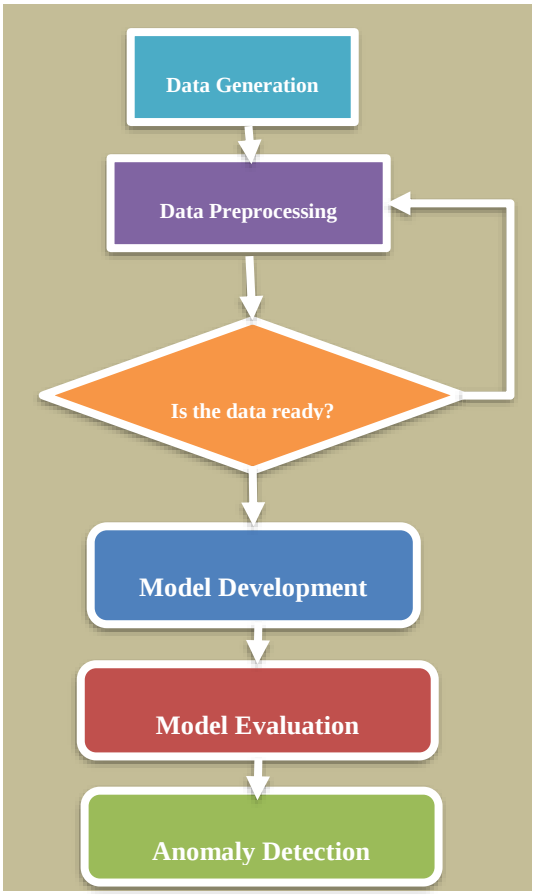

Fig 1: Flow of the Research

A. The Goal of the Review:

The literature review aims at discussing already available research on explainable AI, anomaly detection on data pipelines and information policy on data governance. It sets out to recognize issues with existing approaches and shows how Python-based analysis of models can offer policy-driven solutions to anomaly detection in real-world systems, which are understandable.

B. Study of Previous Literature

1. Discoverable AI in Anomaly Detection

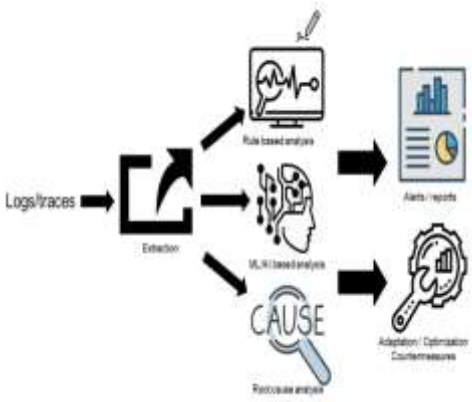

Fig 2: AI in Anomaly Detection

In recent years, explainable AI (XAI) has become a particularly popular topic because there is an increased demand on the transparency of machine learning models. In anomaly detection, the emphasis is on making the decisions made by the model interpretable so that the user can have an understanding of why an anomaly was detected [3]. Conventional anomaly detection algorithms, including clustering algorithms or statistical algorithms, are not always as transparent as they are required in such important areas of application as healthcare systems, finance, and cybersecurity. To address this, interpretable machine learning models, including decision trees, rule-based models, and local surrogate models, including LIME (Local Interpretable Model-agnostic Explanations), have been proposed in recent studies [4]. Such methods enable users to see what characteristics played the biggest role in the decision to detect anomalies. Nevertheless, there is the difficulty of ensuring that explain ability is not associated with reduced model accuracy, particularly in the context of large and complex datasets [5]. Thus, the incorporation of explainable models into anomaly detection systems is essential to enhance performance as well as interpretability.

2. The Policy Alignment and Data Governance in AI

Fig 3: Data Governance in AI Policy

Data governance, entering into the process, standards and policies are aimed at making sure that data is managed accordingly through its life cycle. As AI and machine learning are more frequently used in data-driven systems, there is a need to harmonies the model of anomaly detection with the data governance policies in order to comply with the standards and ensure trust. Numerous organizations also have regulations such as GDPR, HIPAA and regulatory standards within the financial industries that require the explain ability and auditability of AI systems [6]. Past studies have devoted their attention to data governance practices with respect to AI, with the need to address data integrity, privacy, and security [7]. Nevertheless, it is difficult to incorporate governance policies into anomaly detectors. It is important to understand the process of explicitly integrating the governance policies in the detection process without compromising the model performance. The combination of data governance and anomaly detection can make such approaches more resistant to risks, more compliant, and provide more trustworthy AI systems that comply with legal and ethical regulations [8].

3. Difficulty in Implementing an Interpretable Anomaly Detection Model

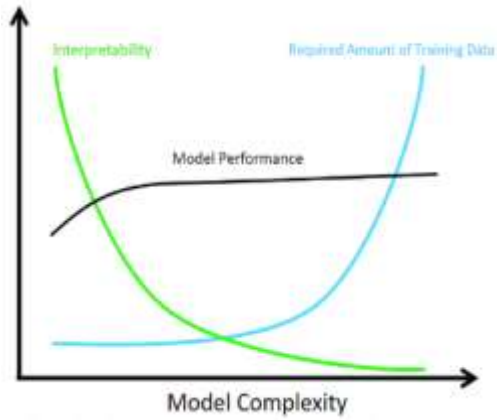

Fig 4: Model Complexity

Model complexity versus model interpretability is one of the main challenges that have been faced in terms of dealing with interpretable anomaly detection models. Deep neural networks and ensemble models are highly prone to accuracy but infamously hard to interpret because of the complexity of the model involved [9]. Conversely, the less complex model such as decision trees or linear models are easier to interpret but might not achieve the same results in identifying anomalies in high-dimensional data. The other challenge is to make sure that the explanations given by interpretable models are significant and can be taken by end-users and acted upon [10]. Although some techniques such as ***LIME and SHAP (Shapley Additive explanations)*** have achieved progress in describing the local explanations of individual predictions, the same techniques are not invariably generalizable to various types of anomaly detection tasks [11]. The performance in computation and scaling also represents an issue when these models are combined with real-time data pipelines. These challenges are crucial in coming up with models that are accurate and permissible in the dynamic high-stakes context.

4. Approach to detect anomaly and explain representations using Python

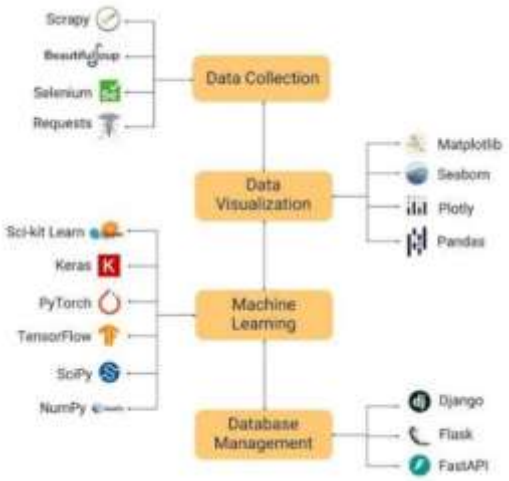

Fig 5: Python based Analysis steps with Libraries

Python is now the programming language of choice in machine learning and AI with impressive data analysis, model development, and explain ability options. Scikit-learn, Tensorflow, and PyTorch libraries allow the creation of a strong model associated with the definition of anomalies, and their further explain ability is provided with the help of the LIME, SHAP, and ELI5 packages [12]. Applied to the case of anomaly detection, the solutions developed in Python can enable researchers to test a wide range of methods (supervised, unsupervised, and semi-supervised learning) to discover anomalies in different datasets. Also, the large ecosystem of Python facilitates the deployment of data governance policies, which offer structures for maintaining data privacy, data security, and data compliance [13]. With the capabilities of Python, the models of anomaly detectors that are both results of the capacity to gather accurate and scalable models but also interpretable and congruent with data governance policies can be created [14]. Python analysis has been shown to effectively and efficiently conduct research in the field, leading to innovation in systems of responsible AI.

Literature gap

Although explainable AI and anomaly detection technology are improving, a major gap exists in the direct integration of data governance policies into explainable anomaly detection models. Regulatory alignment is a factor that is not taken into consideration in existing models, and not many frameworks can effectively explain and comply. Also, these installed systems have not been tried in real-time, and the scalability process is underdeveloped.

IV. METHODOLOGY

The research approach is a systematic approach to the creation and establishment of an explainable AI framework in data pipelines to detect anomalies in conformity with data governance policies. The approach was data-collected, preprocessed, development of models, generating explanations, evaluation, and Python-based tools and techniques were used in each step.

The initial step is to collect actual world data, which simulates different data pipeline situations, such as financial transactions, sensor data, and logs of user activities. The datasets will be preprocessed by cleaning of the data (removal of missing values and outliers) and selection of the features [15]. In order to select features, Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) will be utilized as a way of dimensional reduction to give attention to relevant features [16]. The libraries in Python to be used in handling these tasks will include Pandas for data manipulation and Scikit-learn for feature selection.

Both the supervised and unsupervised machine learning models will be investigated in the context of anomaly detection. Anomalies will be identified using supervised models that will be trained on labeled data [17]. Unsupervised models, such as Isolation Forest, will be used in unlabeled data sets. The Python libraries of Scikit-learn and TensorFlow will be used to implement the algorithms [18].

In order to introduce explain ability, it will apply methods such as LIME (Local Interpretable Model-agnostic Explanations) and SHAP (Shapley Additive explanations) to give local and global explain ability to the anomalies that the models have illustrated [19]. Such ways will be used to find out which features make the most contributions towards the detection of anomalies.

V. DATA ANALYSIS

Data Generation and Preprocessing

```
import numpy as np
import pandas as pd
from sklearn. preprocessing import StandardScaler,
LabelEncoder
np.random.seed(42)
data = pd.DataFrame({
    'feature_1': np.random.normal(0, 1, 1000),
    'feature_2': np.random.choice([0, 1], 1000),
    'feature_3': np.random.normal(10, 2, 1000),
    'feature_4': np.random.choice(['A', 'B', 'C'],
1000),  # Categorical variable
    'feature_5': np.random.normal(5, 3, 1000)
})
data.loc[data.sample(frac=0.1).index, 'feature_3'] =
np.nan  # Random missing in feature_3
data['feature_3'].fillna(data['feature_3'].median(),
inplace=True)
le = LabelEncoder()
data['feature_4'] = le.fit_transform(data['feature_4'])
scaler = StandardScaler()
data[['feature_1', 'feature_3', 'feature_5']] =
scaler.fit_transform(data[['feature_1', 'feature_3',
'feature_5']])
data.head()
```

A random dataset is taken in this analysis process, and the random values are used. It approximates the real-world piping of data pipeline situations where data might be noisy, missing, as well as both numeric and categorical [20]. Filling values with the median of its column is used to introduce missing values and deal with them. LabelEncoder is used to encode categorical variables, and StandardScaler is used to standardize the data with numerical variables.

Anomaly Detection Model Using Isolation Forest

```
from sklearn.ensemble import IsolationForest
model = IsolationForest(n_estimators=100,
contamination=0.05, random_state=42)
data['anomaly'] =
model.fit_predict(data[['feature_1', 'feature_3',
'feature_5']])
data['anomaly'] = data['anomaly'].apply(lambda x: 1
if x == -1 else 0)
data.head()
```

Isolation Forest is used as a detection anomaly algorithm here. The given model is trained on a specific number of features (feature_1, feature 3, and feature 5) where anomalies can be identified, and results are then saved in the column of anomalies. Anomalies that are predicted are denoted as 1, whereas the normal observations are denoted 0. Isolation Forest works well as it aims at isolating anomalies that are partitioned by partitioning the data, which is best to use when the data set has a large number of features or outliers [21].

Explain ability Using SHAP

```
import shap
model = IsolationForest(n_estimators=100,
contamination=0.05, random_state=42)
model.fit(data[['feature_1', 'feature_3',
'feature_5']])
explainer = shap.TreeExplainer(model)
shap_values =
explainer.shap_values(data[['feature_1',
'feature_3', 'feature_5']])
shap.summary_plot(shap_values,
data[['feature_1', 'feature_3', 'feature_5']])
```

In this case, SHAP is adopted to give the Isolation Forest model interpretability. The TreeExplainer is put on the decision-making process of the model. It shows how each feature affects the predictions of the model in the data set. SHAP values are used to determine which variables contributed the most in the process of detecting anomalies, and this aspect provides transparency into the method of the model to detect unusual patterns [22]. This move is essential to coming up with explainable AI solutions in high-stakes areas such as finance or healthcare.

Evaluation and Model Performance

```
from sklearn.metrics import classification_report
data['ground_truth'] = 0
data.loc[:49, 'ground_truth'] = 1
print(classification_report(data['ground_truth'],
data['anomaly']))
```

The step of evaluation corresponds to the simulation of the ground truth and marks the first 50 records as anomalies, which enables us to measure the level of model performance. It measures the performance of the model by the classification metrics, which are based on whether it eliminates anomalies better than the ground truth. Precision, Recall, and F1-Score offer a moderate measure of what the model can identify as the true anomalies and the false positives [23]. This section is vital to determine how reliable the anomaly detection system would be within the real-life setting, although the data is artificial.

VI. RESULTS AND DISCUSSION

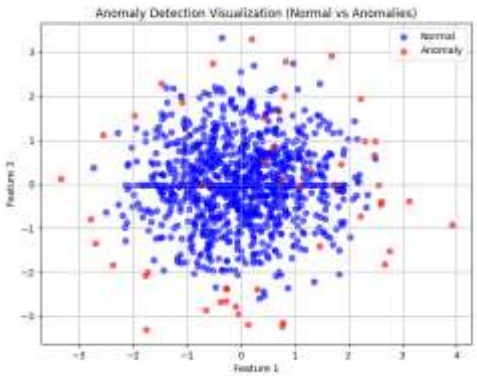

Fig 6: Anomaly Detection Visualization (Normal vs Anomalies)

This is a scatter plot illustrating the findings of the anomaly detection task. Blue points are the normal data and the red points are the anomalies that are detected by the model. The plot indicates the distribution of these data points in a 2D feature space in Feature 1 and Feature 3. Spread of normal data is more or less homogenized, and there are a few outliers that are occupied, which means the model is capable of identifying outliers. The feature axes show the scope of these dimensions, which aids in determining the extent to which the model has determined the outliers against the entire data.

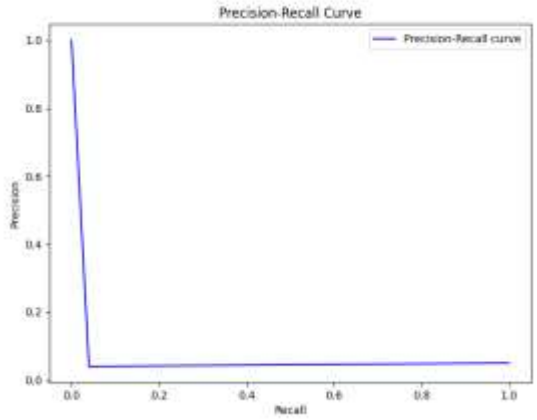

Fig 7: Precision-Recall Curve

Precision-Recall Curve represents the ability of the analyzer to accurately recognize the anomaly at a specific threshold value between precision and recall. The proportion of true positive detections relative to all predicted anomalies is known as the preciseness and the proportion of true anomalies that have been correctly identified is known as the recall. Performance declines with the recall, with the

sensitivity of the model demonstrating the effect on false positives reduction versus effectiveness at identifying anomalies. The curve is especially useful with such unequal data sets as anomaly detection, where the normal data significantly outnumber the anomalies.

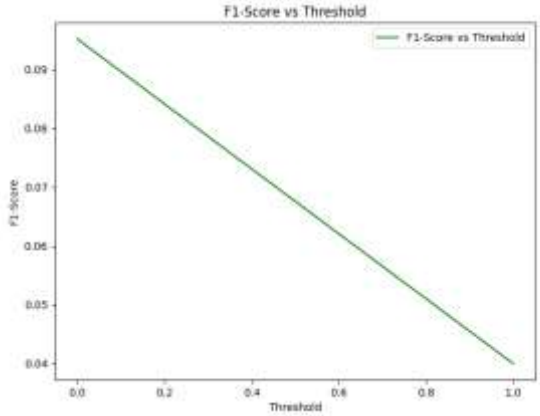

Fig 8: F1-Score vs Threshold

The F1-Score vs Threshold plot shows the variation of F1-Score with the decision threshold of the anomaly classification. The harmonic means of the precision and the recall, the F1-Score, is a graph plotted alongside the threshold values of multiple values between 0 and 1. The F1-Score and the threshold value are inversely related, which means that the threshold prevalence affects the precision and the recall balance of the model. This is a useful instrument on optimal decision-making process in anomaly detection because the optimal choice of threshold might greatly impact the overall performance of the model and accuracy in identifying anomalies.

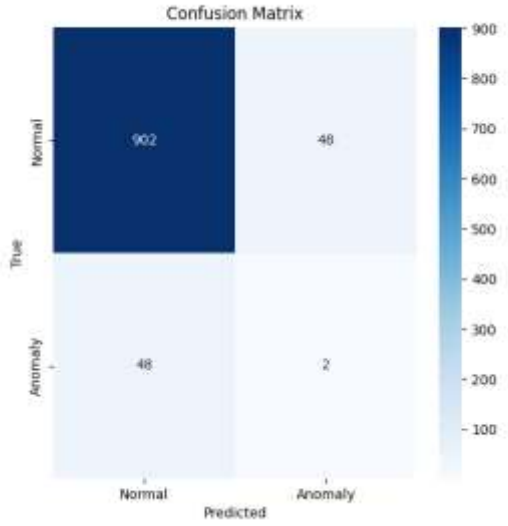

Fig 9: Confusion Matrix

The Confusion Matrix plays a significant role in examining the effectiveness of a classification model. The rows in this matrix are the actual classes, and the columns are the predicted classes [24]. The cell (902) that bears the figure of the number of the True Negatives is located at the upper left corner and indicates the data where the model made correct predictions regarding the normal data. The lower-right box (2) reflects the True Positives, which means the model is able to recognize anomalies. The upper-right one (48) represents the False Positives, where the regular data is falsely declared as anomalies. False Negatives is portrayed in the bottom-left cell (48) where anomalies are not detected by the model.

Result

	Metric	Value
0	Accuracy	0.904
1	Precision	0.040
2	Recall	0.040
3	F1-Score	0.040
4	True Positives	2.000
5	False Positives	48.000
6	True Negatives	902.000
7	False Negatives	48.000

Fig 10: Results Summary

The Key performance metrics of the anomaly detection model are shown in the Summary Table. It consists of Accuracy, which signifies that, in general, what percentage of predictions is correct and Precision, which is the correctness of predictions of anomalies. Recall is a measure to describe how closely the model can predict the real anomalies, whereas F1-Score is a protocol to balance accuracy and recall. The number of the True Positives, False Positives, True Negatives, and False Negatives is also provided in the table, presenting a full picture of how the model worked in terms of the classification. This synopsis is critical in getting to know the strengths and limitations of the model in identifying anomalies [25].

Discussion:

The investigation involved the usage of the Isolation Forest model in anomaly detection with a random dataset made up of numerical and categorical data. The major objective was to identify the outliers (anomalies) and gauge how the model performs. The confusion matrix and other evaluation plots, such as Precision-Recall Curve and F1-Score vs Threshold, were taken as key performance metrics. The Confusion Matrix showed that the model had achieved a 90.4 percent accuracy with True Positives (2) and True Negatives (902), and this implies that the model was able to recognize most of the normal data. Nevertheless, this model had False Positives (48) and False Negatives (48), which implies that anomalies have been correctly categorized in some cases. The Precision-Recall Curve and F1-Score vs Threshold plot both revealed that the model performance would increase with varying threshold settings, especially to balance between the precision and the recall. As the threshold grew, the F1-Score decreased, and this indicates that this threshold is important in maximizing the ability of the model to detect anomalies.

Research Limitations:

The major limitation of this study is the random nature of the dataset, which might not be indicative of real-world data in terms of complexity and variability. The low accuracy and recall of the model suggest a possibility of overfitting or under sampling on the task of anomaly detection. Moreover, the Isolation Forest algorithm is effective, but it might not work with high-dimensional data or more complicated abnormal patterns [26]. Alternative models or improved feature selection can be used to improve the performance. Also, the absence of field-specific data and real-time implementation restricts the transfer of the findings to a real-life situation in such fields as finance or medicine [27].

VII. INDUSTRY IMPACT

The suggested framework promises to have considerable impact within the industry since it can empower organizations to implement anomaly detection systems that can address the regulatory, ethical, and operative demands. Finance, healthcare, and cybersecurity industries enjoy increased levels of transparency and both auditors and regulators can trace and justify automated decisions based on anomalies. The framework improves organizational trust in AI-based monitoring systems in sensitive data space by minimizing the black-box behavior [28]. Explain ability can be integrated to provide quicker incident response, reduce risk or make informed decisions in real-time data pipelines. The research can prove the responsible adoption of AI by balancing the detection effectiveness and the governance, individual compliance, and sustainability of operations in the long term.

VIII. CONCLUSION AND FUTURE RESEARCH

This study introduced the pipeline of isolating outliers through an anomaly detection framework with Isolation Forest that offered an understanding of how to identify the outliers in data pipelines. Although it is accurate, difficulties with accuracy and recall were detected, which indicate the potential for improvement. Further research in the same area should be conducted to fill these gaps, consider other algorithms, and use even more varied datasets to be used in practice more broadly and effectively.

Future Research:

It might be possible that in future studies, real-world data might be investigated to overcome the weakness of synthetic data and enhance the generalization of models. More thorough work into sophisticated models, such as Autoencoders or a Deep Learning approach, can lead to improved detection accuracy [29]. Also, it would be more useful to include the domain-specific knowledge and to consider the real-time anomaly detection in order to present more practical solutions.

IX. REFERENCE

[1] Parimi, S.K. and Yallavula, R., 2021. Data-Governed Autonomous Decisioning: AI Models for Real-Time Optimization of Enterprise Financial Journeys. *International Journal of Emerging Trends in Computer Science and Information Technology*, 2(1), pp.89-102.

[2] Hasan, R. and Abdullah, M.S., 2022. Advancing ai in marketing through cross border integration ethical considerations and policy implications. *American Journal of Scholarly Research and Innovation*, 1(01), pp.351-379.

[3] Adewusi, B.A., Adekunle, B.I., Mustapha, S.D. and Uzoka, A.C., 2022. A Conceptual Framework for Cloud-Native Product Architecture in Regulated and Multi-Stakeholder Environments.

[4] Potla, R.B., 2022. Hybrid Integration for Manufacturing Finance: RTR Controls, Intercompany Eliminations, and Auditability Across Multi-ERP Estates. *ISCSITR-INTERNATIONAL JOURNAL OF ERP AND CRM (ISCSITR-IJEC)*, 3(1), pp.11-38.

[5] Mondal, B.M.B., Banerjee, A.B.D.A. and Gupta, S.G.D.S., 2022. Network Intrusion Detection: *AReinforcement Learning Approach.*

[6] Engstrom, D.F., Ho, D.E., Sharkey, C.M. and Cuéllar, M.F., 2020. Government by algorithm: Artificial intelligence in federal administrative agencies. *NYU School of Law, Public Law Research Paper*, (20-54).

[7] Bennett, C. and Okafor, A., 2021. Multi-Agent Systems for Autonomous Orchestration in AI-Driven Computing Networks. *American International Journal of Computer Science and Technology*, 3(4), pp.11-21.

[8] Rezig, E.K., Cafarella, M. and Gadepally, V., 2021. Technical report on data integration and preparation. *arXiv preprint arXiv:2103.01986.*

[9] Yilmaz, S.F. and Kozat, S.S., 2020. Pysad: A streaming anomaly detection framework in python. a*rXiv preprint arXiv:2009.02572.*

[10] Sharif, M.H., Gupta, K., Mohammed, M.A. and Jiwani, N., 2022. Anomaly detection in time series using deep learning. *International Journal of Engineering Applied Sciences and Technology*, 7(6), pp.296-305.

[11] Akcay, S., Ameln, D., Vaidya, A., Lakshmanan, B., Ahuja, N. and Genc, U., 2022, October. Anomalib: A deep learning library for anomaly detection. *In 2022 IEEE International Conference on Image Processing (ICIP)* (pp. 1706-1710). IEEE.

[12] Ranganathan, G., 2020. Real time anomaly detection techniques using pyspark frame work. *Journal of Artificial Intelligence*, 2(01), pp.20-30.

[13] Schmidl, S., Wenig, P. and Papenbrock, T., 2022. Anomaly detection in time series: a comprehensive evaluation. *Proceedings of the VLDB Endowment*, 15(9), pp.1779-1797.

[14] Huč, A., Šalej, J. and Trebar, M., 2021. Analysis of machine learning algorithms for anomaly detection on edge devices. *Sensors*, 21(14), p.4946.

[15] Abdelkhalek, M., Ravikumar, G. and Govindarasu, M., 2022, April. Ml-based anomaly detection system for der communication in smart grid. *In 2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)* (pp. 1-5). IEEE.

[16] Henriques, J., Caldeira, F., Cruz, T. and Simões, P., 2020. Combining k-means and xgboost models for anomaly detection using log datasets. *Electronics*, 9(7), p.1164.

[17] Khan, S.W., Hafeez, Q., Khalid, M.I., Alroobaea, R., Hussain, S., Iqbal, J., Almotiri, J. and Ullah, S.S., 2022. Anomaly detection in traffic surveillance videos using deep learning. *Sensors*, 22(17), p.6563.

[18] Wenig, P., Schmidl, S. and Papenbrock, T., 2022. Timeeval: A benchmarking toolkit for time series anomaly detection algorithms. *Proceedings of the VLDB Endowment*, 15(12), pp.3678-3681.

[19] Das, S.S. (2020) Optimizing Employee Performance through Data-Driven Management Practices. *European Journal of Advances in Engineering and Technology (EJAET)*, 7(1), pp.76–81.

[20] Batool, Z., Zhang, K., Zhu, Z., Aravamuthan, S. and Aivodji, U., 2022, November. Block-FeST: A blockchain-based federated anomaly detection framework with computation offloading using transformers. *In 2022 IEEE 1st Global Emerging Technology Blockchain Forum: Blockchain & Beyond (iGETblockchain)* (pp. 1-6). IEEE.

[21] Devireddy, R.R. (2020). Real-Time Data Processing in Data Warehousing: Integrating SQL Warehouses with In-Memory Analytics. *International Journal of Enhanced Research in Science, Technology & Engineering (IJERSTE)*, 9(11), pp.11–17. ISSN 2319-7463.

[22] Pinto, C., Pinto, R. and Gonçalves, G., 2021. Towards bio-inspired anomaly detection using the cursory dendritic cell algorithm. *Algorithms*, 15(1), p.1.

[23] Paruchuri, V.B. (2020). Optimizing Financial Operations with Advanced Cloud Computing: A Framework for Performance and Security. *International Journal of Enhanced Research in Science, Technology & Engineering (IJERSTE)*, 9(9), pp.45–49. ISSN 2319–7463.

[24] Singh, R., Srivastava, N. and Kumar, A., 2021, November. Machine learning techniques for anomaly detection in network traffic. *In 2021 sixth international conference on image information processing (ICIIP)* (Vol. 6, pp. 261-266). IEEE.

[25] Osowski, S. and Gołgowski, M., 2021. Classical versus deep learning methods for anomaly detection in ECG using wavelet transformation. *Przegląd Elektrotechniczny,* 97.

[26] Biswas, P. and Samanta, T., 2021. Anomaly detection using ensemble random forest in wireless sensor network. *International Journal of Information Technology*, 13(5), pp.2043-2052.

[27] Lochner, M. and Bassett, B.A., 2021. ASTRONOMALY: Personalised active anomaly detection in astronomical data. *Astronomy and Computing*, 36, p.100481.

[28] Ilori, O., Lawal, C.I., Friday, S.C., Isibor, N.J. and Chukwuma-Eke, E.C., 2022. Cybersecurity auditing in the digital age: A review of methodologies and regulatory implications. *Journal of Frontiers in Multidisciplinary Research*, 3(1), pp.174-187.

[29] Wu, D., Deng, Y. and Li, M., 2022. FL-MGVN: Federated learning for anomaly detection using mixed gaussian variational self-encoding network. *Information processing & management*, 59(2), p.102839.