ENS 491-492 – Graduation Project

Final Report

Project Title: High Fuel Consumption Detection System for Heavy-Duty Trucks

Group Members: Yağız Gürdamar,Zeynep Pancar,Ece Erten

Supervisor(s):Mustafa Ünel

Date:26.05.2024

Sabancı Üniversitesi

1. EXECUTIVE SUMMARY

For heavy-duty vehicles, high fuel consumption (FC) identification is essential for detecting anomalies in the system and driver behavior. The fuel consumption of these vehicles ranges considerably depending on the type of driving manner, gross vehicle weight, and topology of the road. The project we worked on used machine learning models to construct a data-driven anomaly detection system to address the issue at hand.

Long Short-Term Memory(LSTM) networks, which are appropriate for assessing time-series data, were a key component of team solutions. To detect anomalies in fuel consumption, the model was trained on several kinds of parameters, including engine speed, vehicle weight, and road slope. We used sophisticated techniques like SHAP and LIME for model evaluation and comprehensibility since the problem was unsupervised.

As a part of the proposed approach, our team created sliding windows to capture temporal relations, collected data from a variety of vehicles and conducted preprocessing processes to deal with missing values through bidirectional interpolation. Multiple layers with dropout and L2 regularization were implemented into the LSTM model architecture to prevent overfitting and achieve a minimal mean squared error (MSE) during validation.

The system's ability to identify anomalies was proved by the key findings from the experiment, which included a low **MSE score of 1.93E-06**, suggesting that accurate predictions closely matched actual data. The effectiveness of the model in practical applications is shown by its capability to detect notable differences in fuel consumption patterns. The results of this study highlight the system's potential for considerable fuel and pollution savings which make it an asset for fleet managers for assisting periodic upkeep and operational adjustments.

To sum up, the approach that has been developed demonstrates potential in both increasing fuel consumption and reducing the environmental impact heavy-duty vehicles have on the environment. The next phases will concentrate on improving the model's accuracy, performing an extensive amount of real-world testing, and including more environmental parameters. Such will provide opportunities for an adaptable approach that will enhance fleet management as well as environmental sustainability in the trucking industry.

2. PROBLEM STATEMENT

Significant economic and environmental issues are brought about by high fuel consumption (FC) in heavy-duty diesel trucks (HDDTs). HDDTs are the primary form of freight transportation, and therefore, they require a lot of fuel like diesel, which increases operating costs and releases a considerable of greenhouse gases. These greenhouse gases restrict efforts to achieve global sustainability while contributing to environmental harm. By developing a sophisticated anomaly detection system that can identify FC anomalies and inefficiencies—which could be signs of malfunctioning equipment or improper driving habits—our project aims to solve these issues.

Our motivation arises from the critical need for better fuel efficiency and reducing the vehicle industry's environmental effects. We aim to address the root cause of high fuel consumption as a way to contribute to global environmental objectives reduce emissions of greenhouse gases, and stimulate investments. This project aims to offer fleet managers an effective approach to optimize fuel consumption, improve vehicle preservation, and promote environmentally conscious driving behaviors.

Previous studies have established the framework for understanding and managing the fuel consumption (FC) in heavy-duty vehicles. Mumcuoğlu et al. (2023), for instance, studied fuel consumption classification methods for heavy-duty vehicles by applying machine learning algorithms. Comparably, Liimatainen (2011) highlighted how driver behavior has an important impact on fuel consumption and how eco-driving strategies can reduce fuel consumption by up to 30%. After reviewing different fuel consumption prediction techniques while considering into account many contributing factors, Gong et al. (2021) found that involving a range of data sources significantly improves prediction accuracy.

To handle complicated, time-series data, deep learning models such as LSTMs have been used in recent anomaly detection improvements. An ensemble of multi-point LSTMs was developed by Lee et al. (2023) for the objective of anomaly identification, showing the model's high accuracy as well as the ability to capture sophisticated temporal patterns. Theissler (2017) emphasized the importance of applying ensemble-based anomaly detection techniques for identifying automotive system shortcomings which are both known and unknown.

To improve accuracy and applicability under a range of driving situations, our study advances the field by combining ensemble machine learning models and real-time data processing. The complex and temporal structure of FC data is handled by Long Short-Term Memory (LSTM) networks, which are widely recognized for their capacity to capture long-term dependencies in sequential data. This approach aims to present a more responsive and adaptive fuel management solution compared to the static models frequently observed in the literature.

2.1. **Objectives/Tasks**

- **Develop a Robust FC Classification Framework:**

Objective: Deploy ensemble machine learning models into action to distinguish between normal and anomalous FC patterns.

Intended Results: Clear detection of FC deviations related to vehicle weight, and engine speed road situations.

- **Anomaly Detection:**

Objective: Construct approaches for detecting anomalies among normal FC patterns.

Intended Results: Detecting potential vehicle issues or unproductive driving behaviors so that targeted improvements may be addressed.

- **System Validation:**

Objective: Employ current as well as past data, and conduct thorough testing to verify the accuracy and robustness of the FC prediction and anomaly detection models.

Intended Results: Ensure that the system executes adequately under a variety of operational and environmental conditions.

- **Deployment and Iterative Improvement:**

Objective: The system will be tested and implemented in a controlled environment, and its performance will be periodically monitored. Iterative improvements are planned depending on user feedback and data obtained.

Intended Results: Improve system adaptability and efficiency, assuring reliable performance in real-world applications.

2.2. **Realistic Constraints**

- **Economic Constraints:** Since there are increasing fuel costs, heavy-duty trucks' high fuel consumption (FC) causes significant financial difficulties. By precisely identifying FC anomalies and inefficiency, this effort aims to decrease these costs by enabling better fuel management and operational adjustments. To attempt to overcome financial constraints, the team created a reasonable approach that makes use of readily available vehicle monitoring data and employs machine learning algorithms to provide insightful findings without requiring additional expensive hardware.

- **Environmental Constraints:** To be able to reduce emissions of greenhouse gases, which are an important contributor to climate change and damage to the environment, heavy-duty trucks have to decrease their fuel consumption (FC). The objective of our initiative of reducing fuel consumption thus lowering the environmental effect of vehicle fleets, is in alignment with across-the globe sustainability goals. Our solution is designed to support environmental protection efforts by complying with environmental regulations and standards, particularly the CO_2 emission reduction targets determined by the European Union.

- **Health and Safety Constraints:** High FC causes air pollution from vehicle emissions, resulting in an adverse effect on the environment and causing potential health risks. By identifying and preventing FC anomalies this initiative improves overall safety and the quality of the air. Drivers' and other stakeholders' safety and well-being are not jeopardized by our data collection and processing techniques.
- **Manufacturability Constraints:** As our approach utilizes already existing vehicle monitoring systems, it can potentially be readily produced and put into action without demanding significant modifications regarding the designs of current vehicles. Our team is confident that our anomaly detection method can be broadly implemented with minimal disruption of present fleet operations by focusing on software-driven solutions.

- **Sustainability Constraints:** The project's key objective is to enhance sustainability by reducing emissions and improving fuel consumption. Our approach's adaptability and scalability provide long-term sustainability benefits to different kinds of vehicles and operating circumstances. Our team also considers how our solution will impact the environment throughout its life, ensuring that it will improve environmental sustainability when used and applied.

Engineering Standards:

Our goal gathering and analyzing of data methods satisfy the following engineering standards to ensure consistency and reliability:

1. ISO 14224:2016 - This international standard contains guidelines for collecting and distributing equipment upkeep and dependability data. All of our information-collecting techniques are standardized, reliable, and consistent because we conform to ISO 14224:2016, which is essential for precision anomaly identification and analysis.

2. SAE J1939 - This international standard addresses diagnostics and communication between parts for vehicles. By conforming to SAE J1939, our method may be simply integrated with existing vehicle monitoring systems, enabling data collection and analysis straightforward.

3. ISO 14001:2015 - This international standard for environmental management provides a structure for addressing environmental responsibilities. Applying ISO 14001:2015 ensures that our project meets goals for sustainability and remains in line with best practices in environmental management.

Our project not only seeks to address the problem of high fuel consumption in heavy-duty trucks, but it also makes sure that the solution is safe, sustainable, manufacturable, environmentally friendly, and economically practical by considering these limits in consideration.

3. METHODOLOGY

Preprocessing

Handling Missing Values

Managing the dataset's missing values was one of the main responsibilities during the preprocessing stage. We employed bidirectional interpolation to do this. To ensure a smooth and continuous dataset, this method predicts the missing values by taking into account the available data points before and after the missing entries.

Creating Sliding Windows

We used sliding windows to capture the temporal dependencies present in the data. Every window lasts for five minutes, with a one-minute break in between. This method aids in lowering the dimensionality of the data while maintaining the temporal context. Using HghRslutionTotalVehicleDistance and EngSpeed, we estimated the average fuel consumption (average fuel consumption), as well as the mean and standard deviation of each signal during these intervals.

Data Integration

Each CSV file was preprocessed before being combined into a single DataFrame with the name combined_df. Many signals, including brake, gas, weight, speed, and the recently computed average fuel usage, are included in this extensive dataset. We can do a comprehensive analysis of the data thanks to this integration.

Model Selection and Architecture

Signals Used as Inputs

The input signals for the LSTM and autoencoder models include:

- HghRslutionTotalVehicleDistance_mean
- TachographVehicleSpeed_mean
- EngSpeed_mean
- ActualEngPercentTorque_mean
- AccelPedalPos1_mean
- BrakePedalPos_mean
- PCCM_Slope_mean
- DStgy_dmRdcAgAct_mean
- EngOilTemp1_mean
- EngCoolantTemp_mean
- GrossCombinationVehicleWeight_mean
- EngTotalFuelUsed_mean

These signals provide a comprehensive view of the vehicle's behavior and help in accurately detecting anomalies.

LSTM Model

Because Long Short-Term Memory (LSTM) networks are good at collecting long-term dependencies in sequential data, that is why we choose them. Because LSTMs can learn from historical data and generate precise predictions based on the sequences they have learnt, they are especially well-suited for time-series anomaly identification.

The architecture of our LSTM model, the model consists of:

- An input layer
- Multiple LSTM layers
- A dense output layer

The detailed parameters of the LSTM model are provided in Table 1.

Parameter	Value
Number of Layers	3
Units per Layer	128, 64, 32
Learning Rate	0.0001
Batch Size	32
Epochs	50

Table 1: LSTM Model Parameters

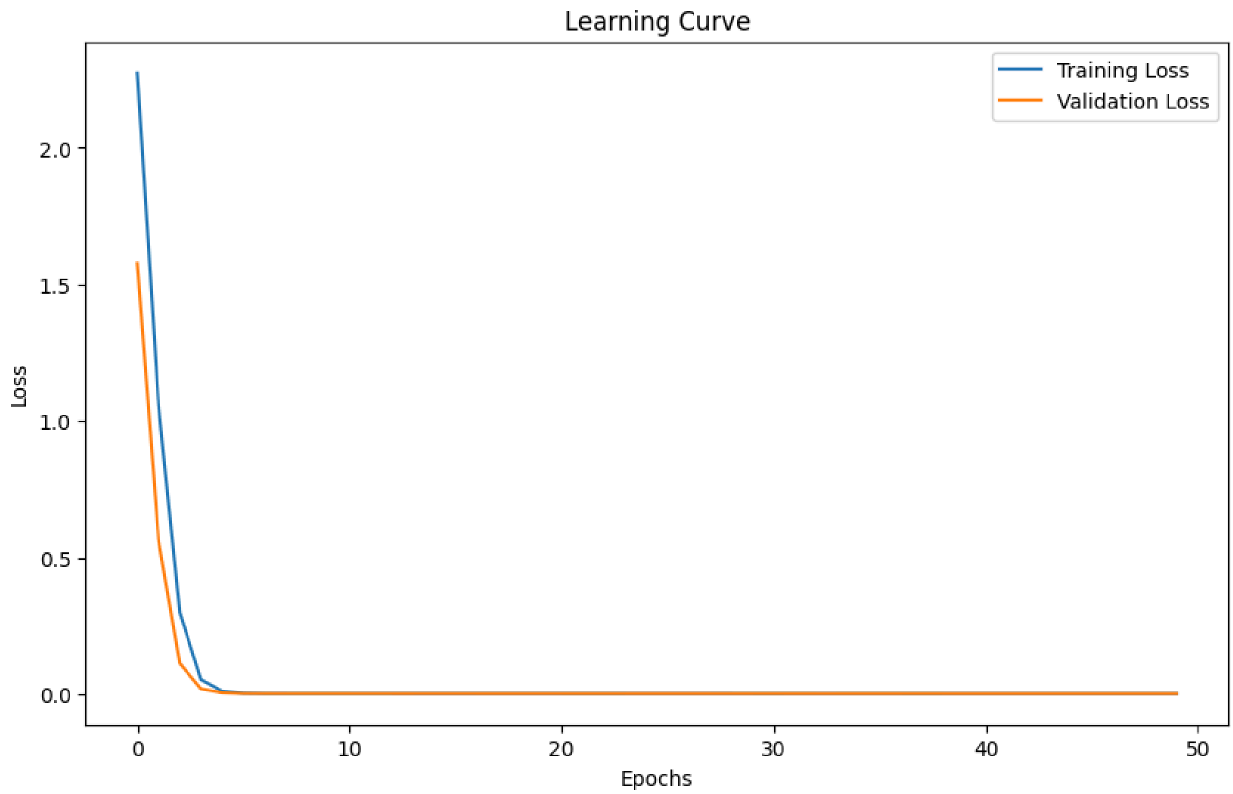

Figure 1: LSTM Learning Curve

Autoencoder for Anomaly Detection

For anomaly identification, we used an autoencoder in addition to the LSTM model. A neural network called an autoencoder is made to encode data in an efficient manner and then decode it back to its original form. Anomalies are found using the reconstruction error, which is the difference between the input and the reconstructed output. High reconstruction error data points are generally regarded as abnormalities

Parameter	Value
Number of Layers	3
Units per Layer	50

Learning Rate	0.0001
Batch Size	32
Epochs	100

Table 2: Autoencoder Parameters

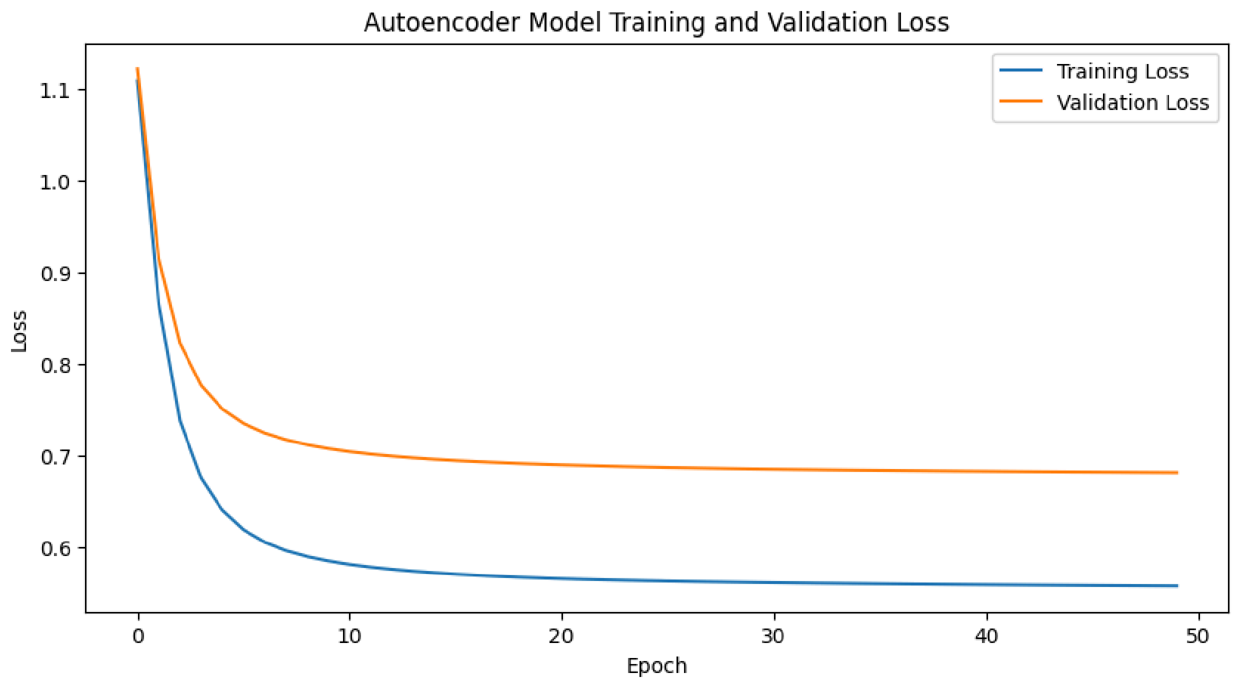

Figure 2: Autoencoder Learning Curve

Anomaly Detection Methods

We used the LSTM and autoencoder models in combination with a number of other techniques to identify anomalies in the data:

LSTM-Based Anomaly Detection

LSTM Regression: Based on the prediction error, anomalies are identified by using the LSTM model to predict the subsequent value in the sequence. The findings of anomaly detection using this technique, with abnormalities on the avg_fuel_consumption time series highlighted by using 4 different thresholds.

Reconstruction Error: Anomalies are identified based on the reconstruction error. The higher the error, the more likely the point is an anomaly. Figure 3 shows the reconstruction error time series and highlights the detected anomalies.

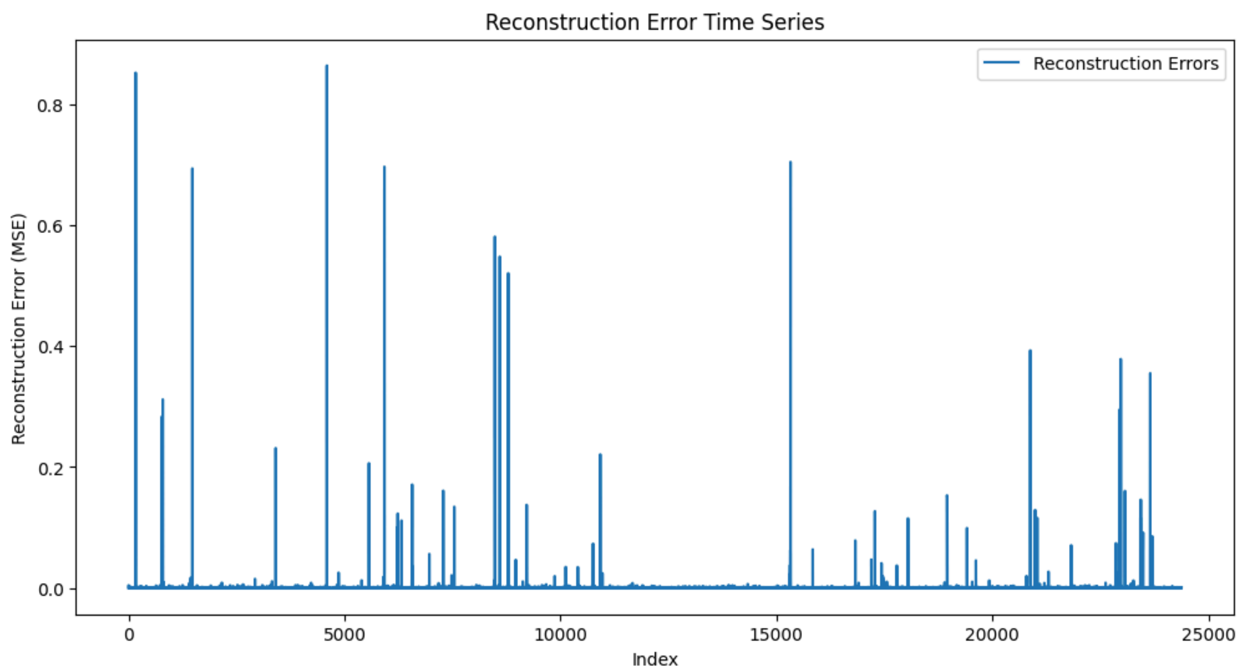

Figure 3: reconstruction Error Time Series (LSTM)

- **Z-score Thresholding:** In order to identify abnormalities, the reconstruction errors are converted into Z-scores. Anomalies are identified as points whose Z-scores are higher than a predetermined threshold (e.g., 3.0). Finding spots that substantially depart from the normal distribution of mistakes is possible with this strategy. The Z-score based anomaly detection results are displayed in Figure 4.

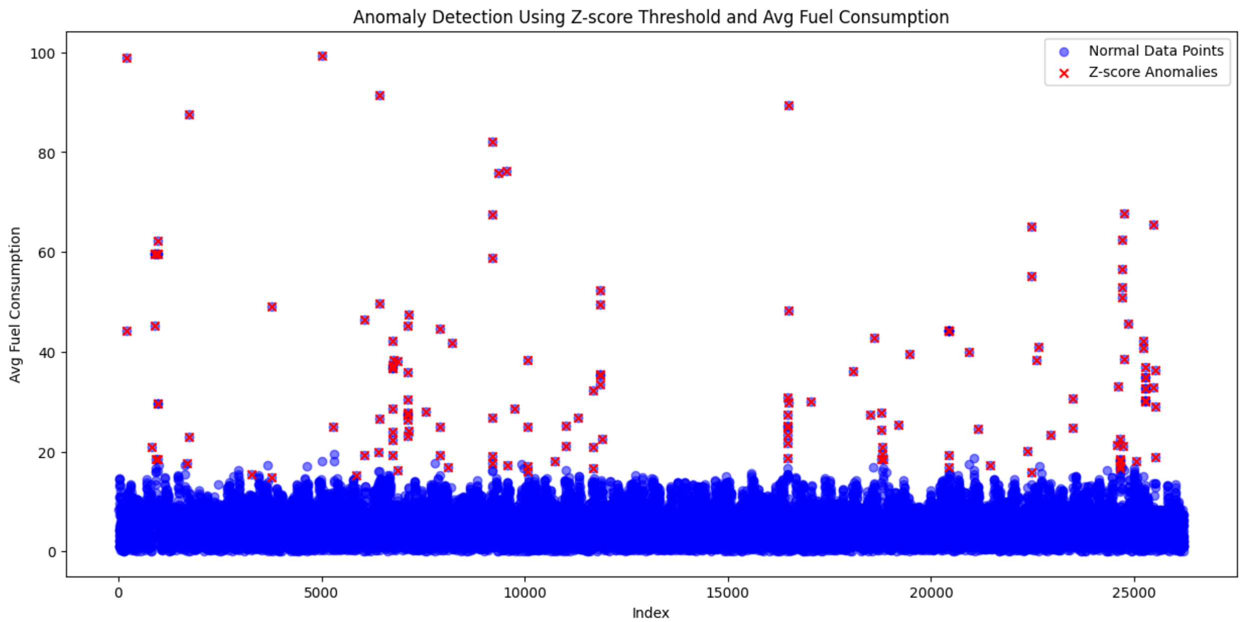

Figure 4: Anomaly Detection Using Z-score Threshold (LSTM)

- **Dynamic Thresholding:** This technique modifies the anomaly detection threshold over time in response to changes in the statistical characteristics of the data. The dynamic threshold (95th percentile) applied to the anomalies and reconstruction errors is displayed in Figure 5.

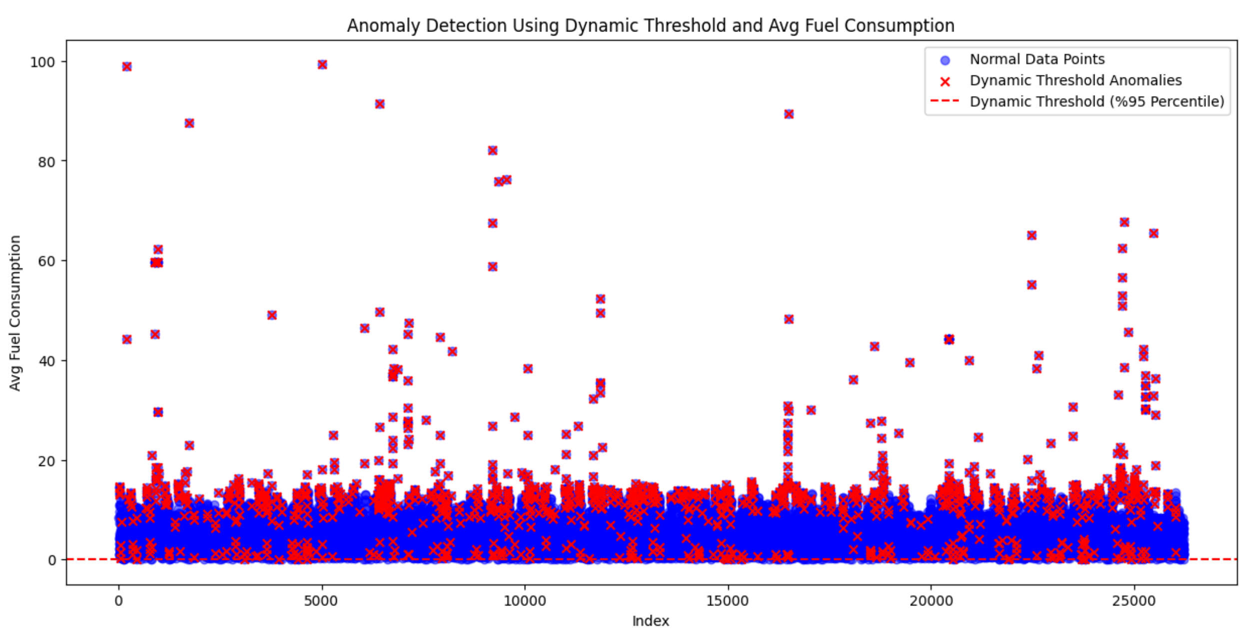

Figure 5: Anomaly Detection Using Dynamic Threshold (LSTM)

- **Fixed Thresholding:** An irregularity is identified using a predetermined fixed threshold. The outcomes are shown in Figure 6 with a fixed threshold of 0.01.

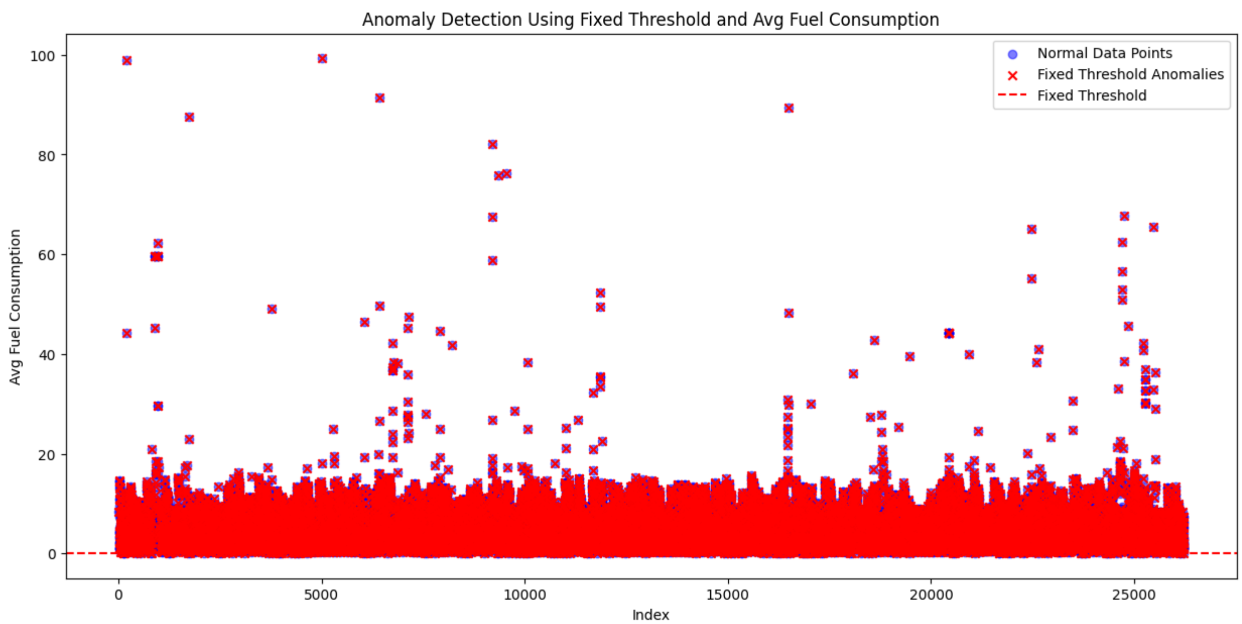

Figure 6: Anomaly Detection Using Fixed Threshold (LSTM)

- **Interquartile Range (IQR) approach:** The IQR approach finds outliers in the data distribution and is used to detect anomalies. Figure 7's box plot displays the distribution of average fuel use as well as any anomalies that were found in relation to the IQR thresholds.

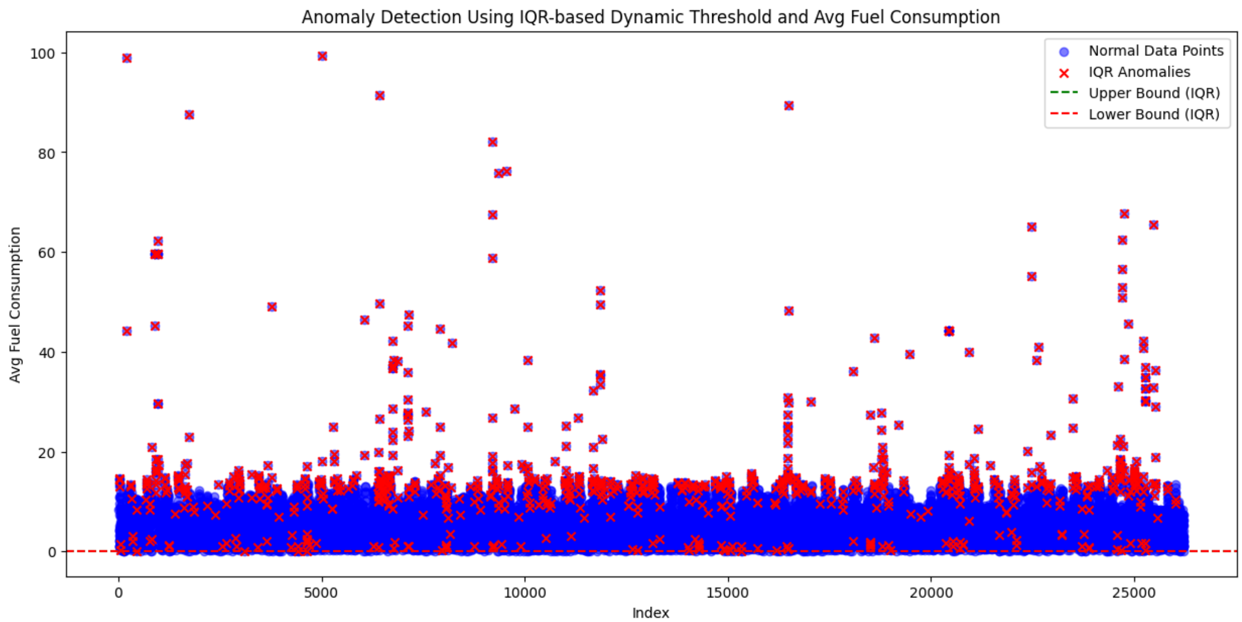

Figure 7: Anomaly Detection Using IQR Method (LSTM)

- **Clustering (K-Means):** To find clusters in the data, we employed a clustering technique like K-Means. Anomalies are defined as points that are in sparse regions or do not belong to any cluster. Figure 8 displays the clustering result (K-Means).

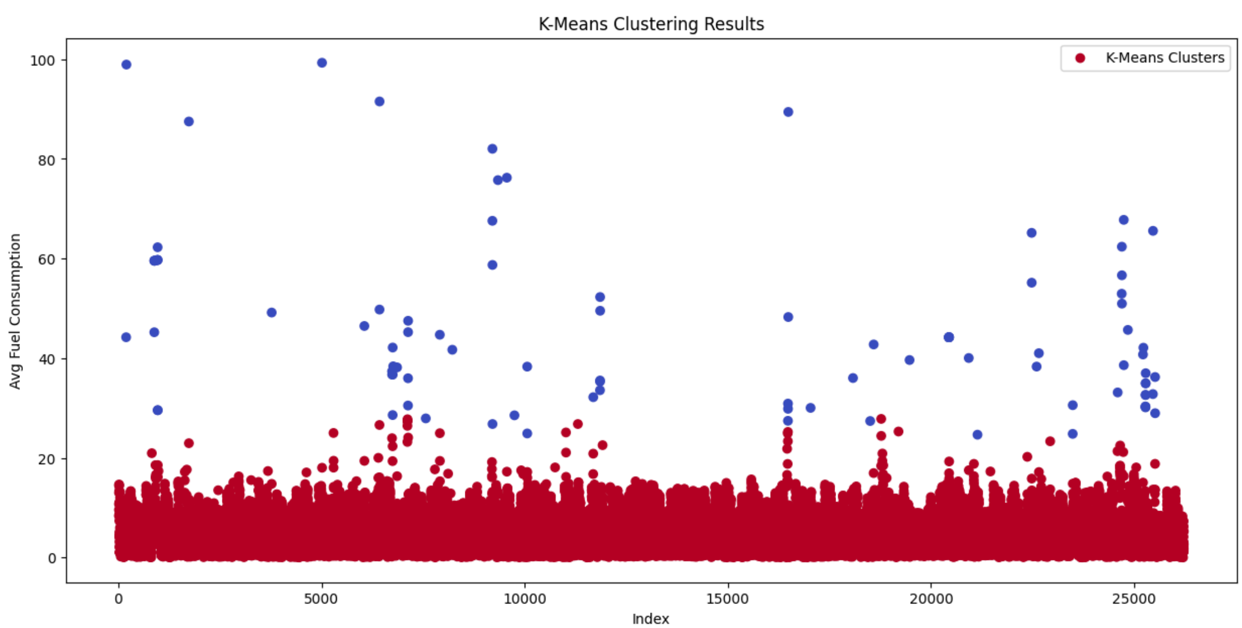

Figure 8: K-Means Clustering Results (LSTM)

- **Visualizing Latent Spaces:** PCA and t-SNE were utilized to visualize the latent spaces in order to gain a deeper understanding of the data and the anomalies that were found. The PCA plots for anomalies found using IQR and dynamic thresholds, respectively, are displayed in Figures 9 and 10. The t-SNE plots for dynamic and IQR-based anomalies are displayed in Figures 11 and 12.

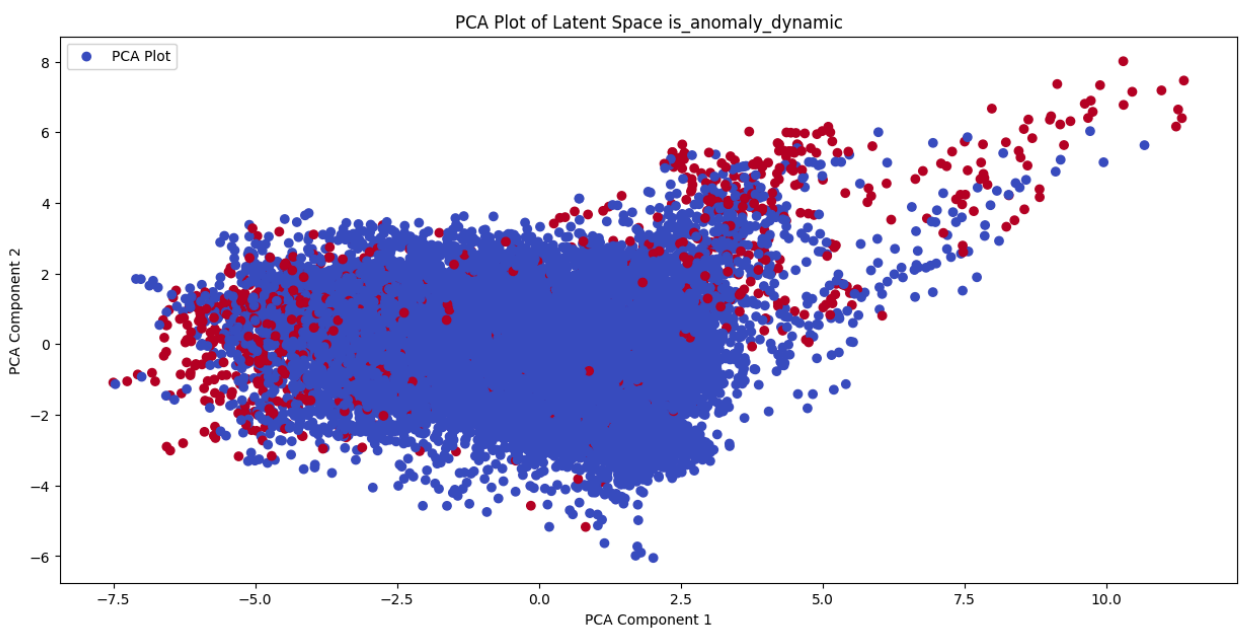

Figure 9: PCA Plot of Latent Space (Dynamic Threshold Anomalies, LSTM)

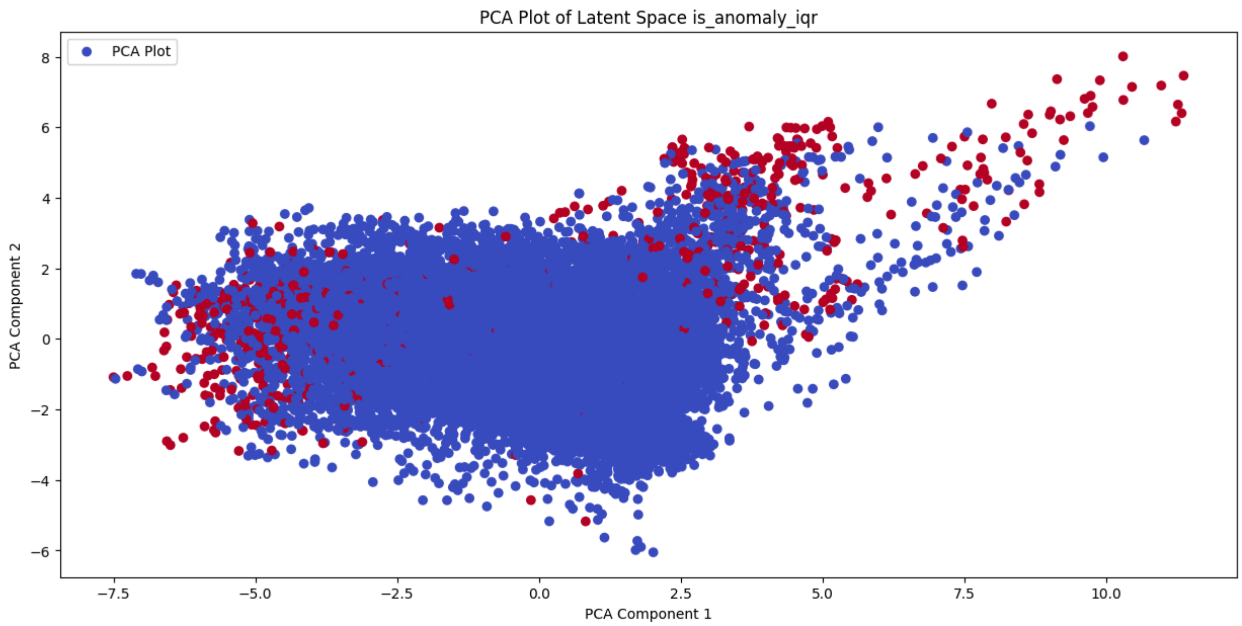

Figure 10: PCA Plot of Latent Space (IQR Threshold Anomalies, LSTM)

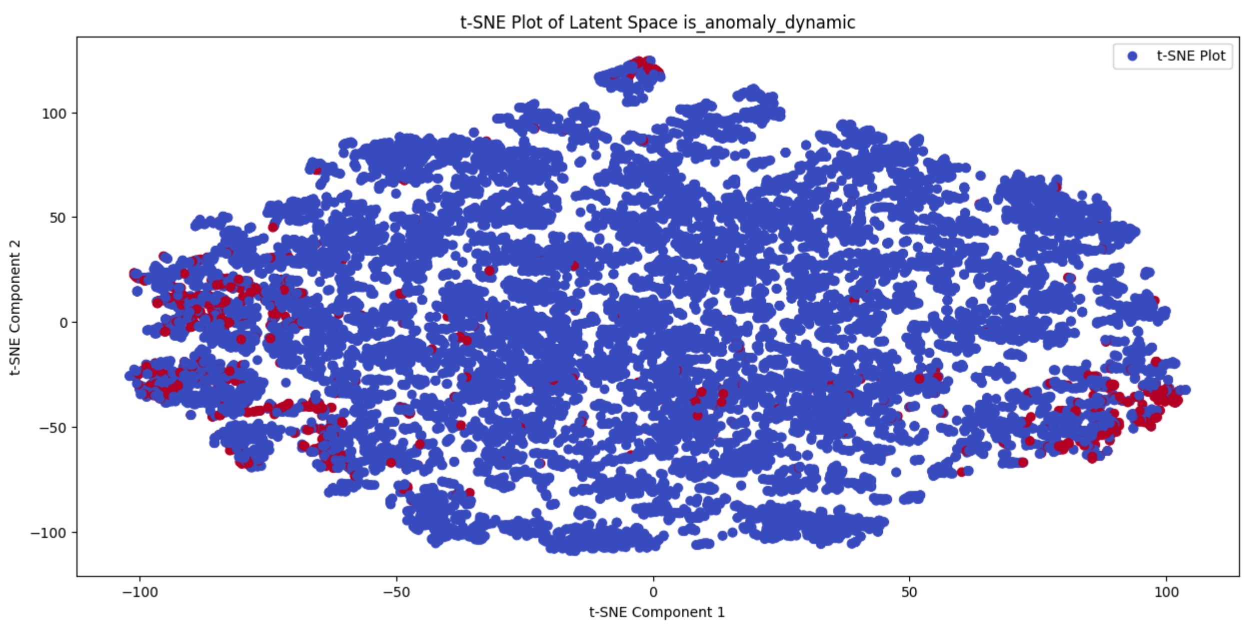

Figure 11: t-SNE Plot of Latent Space (Dynamic Threshold Anomalies, LSTM)

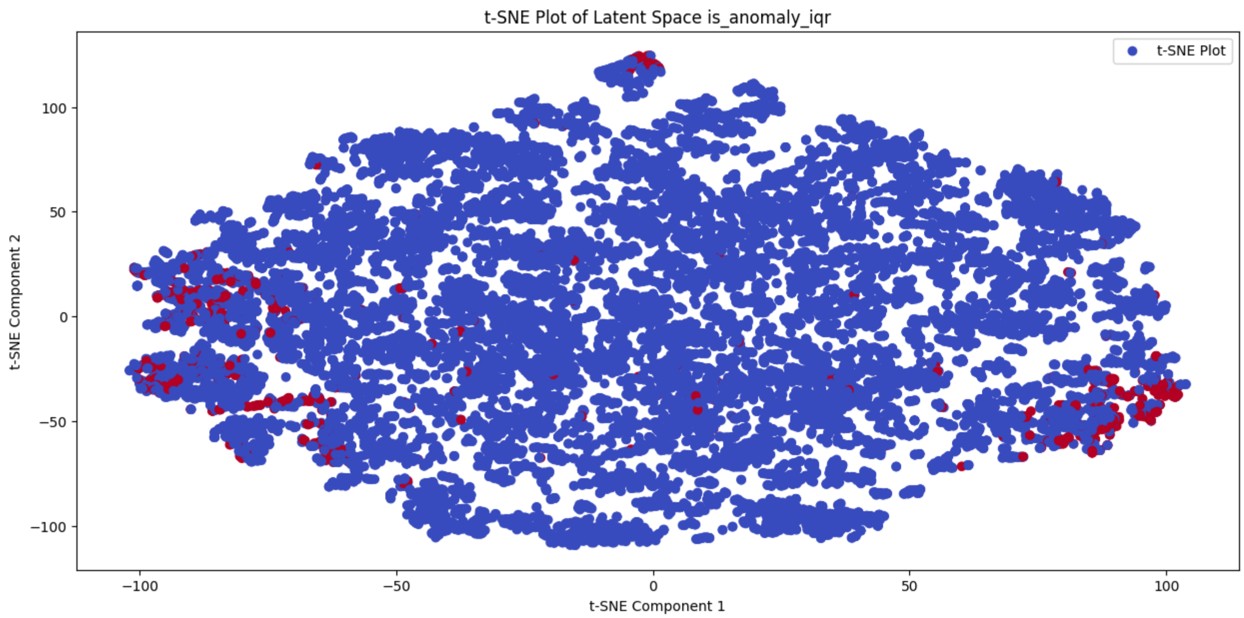

Figure 12: t-SNE Plot of Latent Space (IQR Threshold Anomalies, LSTM)

Autoencoder-Based Anomaly Detection

- **Reconstruction Error:** Reconstruction error is the basis for identifying anomalies. The likelihood that the point is an anomaly increases with inaccuracy. The reconstruction error time series and identified anomalies are displayed in Figure 13.

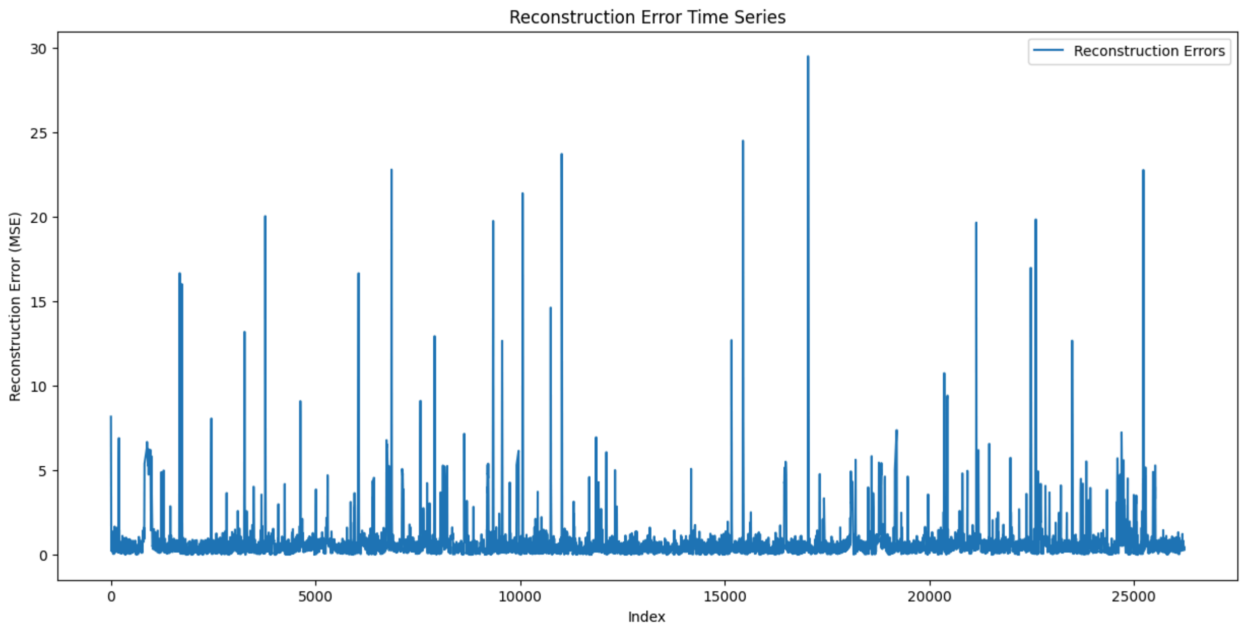

Figure 13: Reconstruction Error Time Series (Autoencoder)

- **Z-score Thresholding:** The reconstruction errors from the autoencoder are converted into Z-scores to identify anomalies, much as the LSTM model. Anomalies are identified as points whose Z-scores are higher than a predetermined threshold (e.g., 3.0). Finding spots that substantially depart from the normal distribution of mistakes is possible with this strategy. The Z-score based anomaly detection results are displayed in Figure 14.

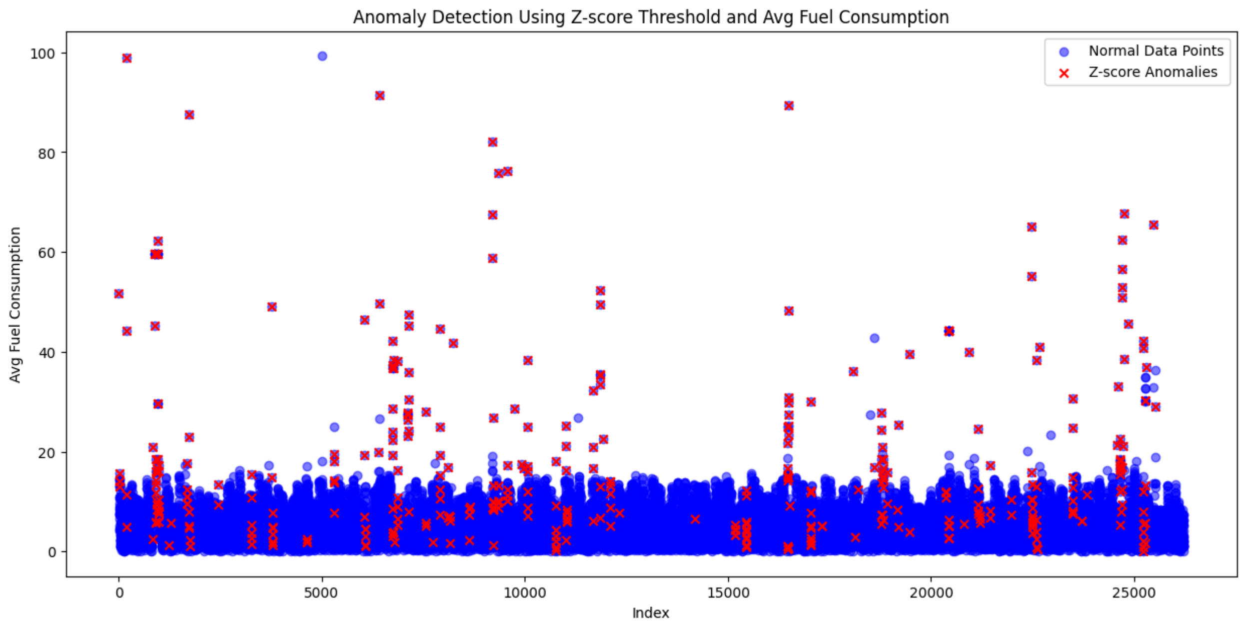

Figure 14: Anomaly Detection Using Z-score Threshold (Autoencoder)

- **Dynamic Thresholding:** This technique modifies the anomaly detection threshold over time in response to changes in the statistical characteristics of the data. The dynamic threshold (95th percentile) applied to the anomalies and reconstruction errors is displayed in Figure 15.

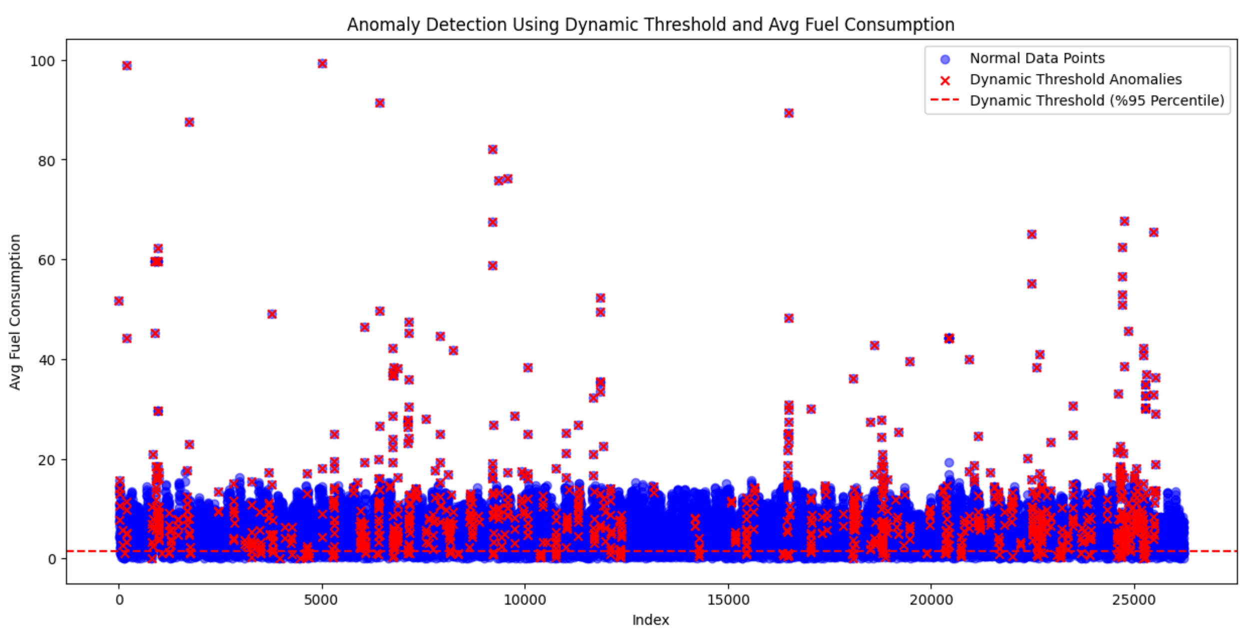

Figure 15: Anomaly Detection Using Dynamic Threshold (Autoencoder)

- **Fixed Thresholding:** An irregularity is identified using a predetermined fixed threshold. The outcomes are shown in Figure 19 with a fixed threshold of 0.01.

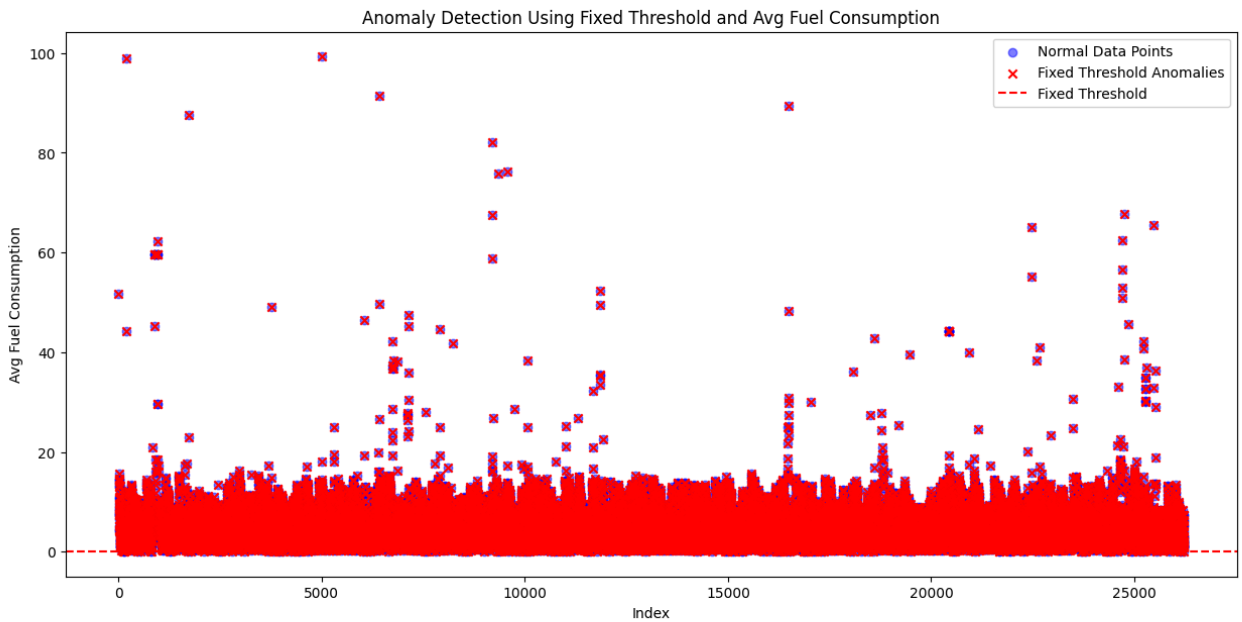

Figure 16: Anomaly Detection Using Fixed Threshold (Autoencoder)

- **Interquartile Range (IQR) approach:** The IQR approach finds outliers in the data distribution and is used to detect anomalies. The distribution of average fuel use and the anomalies found based on the IQR thresholds are shown in Figure 17's box plot

.

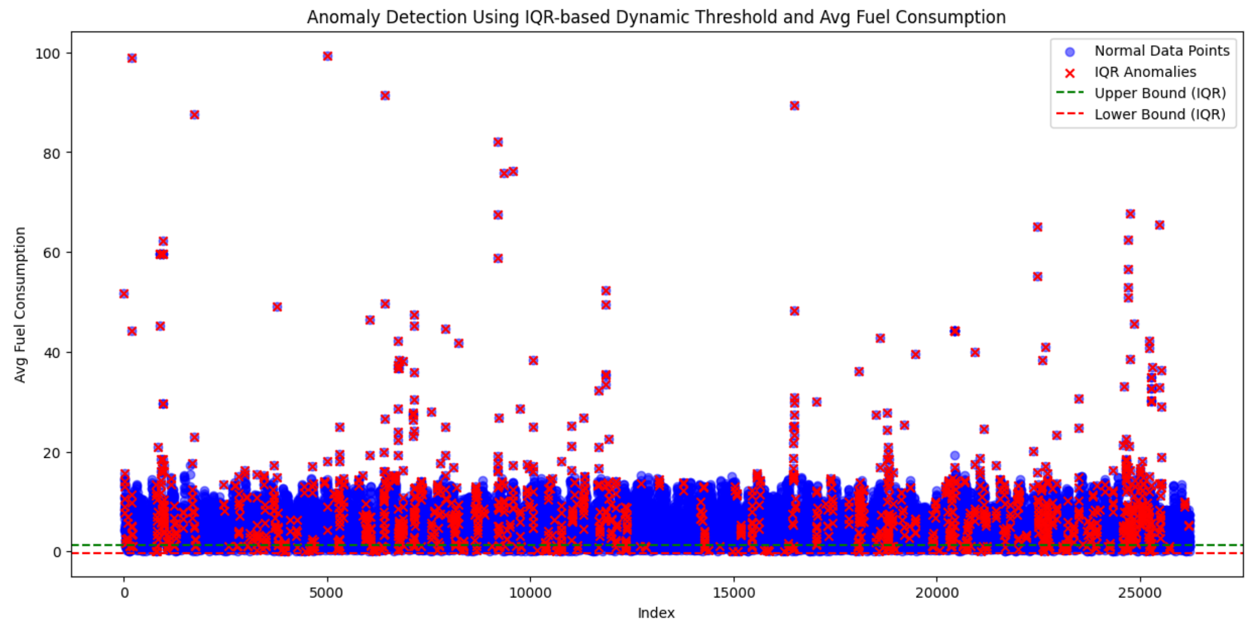

Figure 18: Anomaly Detection Using IQR Method (Autoencoder)

- **Visualizing Latent Spaces:** PCA and t-SNE were utilized to visualize the latent spaces in order to gain a deeper understanding of the data and the anomalies that were found. The PCA plots for anomalies found using IQR and dynamic thresholds, respectively, are displayed in Figures 19 and 20. The t-SNE plots for dynamic and IQR-based anomalies are displayed in Figures 21 and 22.

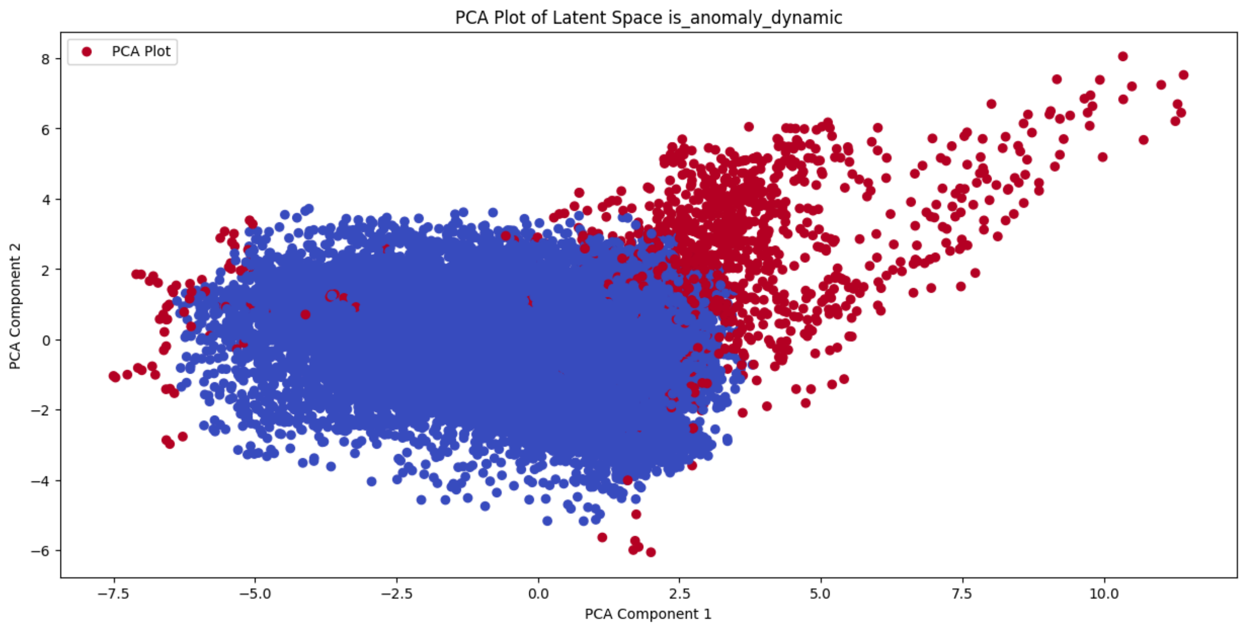

Figure 19: PCA Plot of Latent Space (Dynamic Threshold Anomalies, Autoencoder)

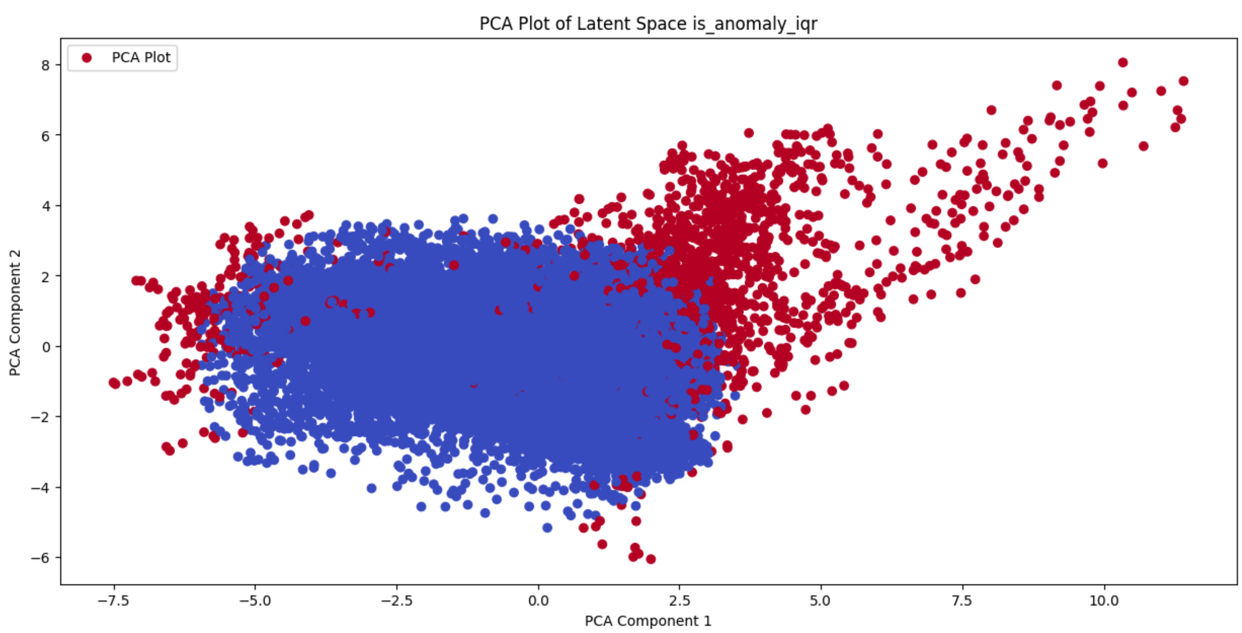

Figure 20: PCA Plot of Latent Space (IQR Threshold Anomalies, Autoencoder)

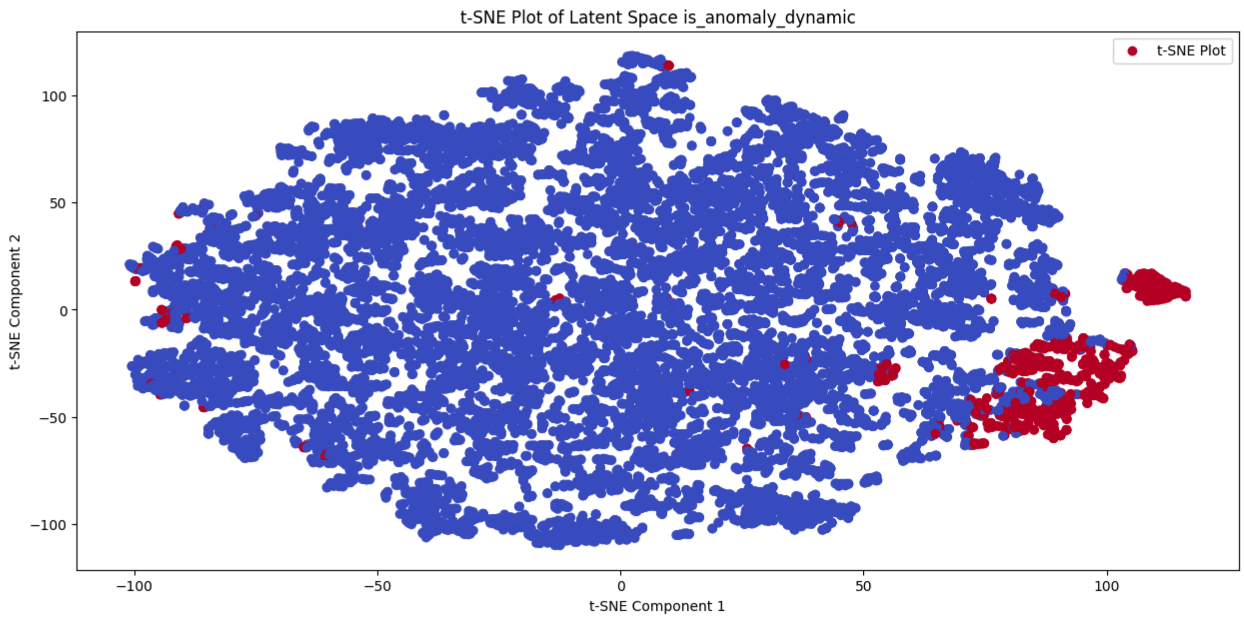

Figure 21: t-SNE Plot of Latent Space (Dynamic Threshold Anomalies, Autoencoder)

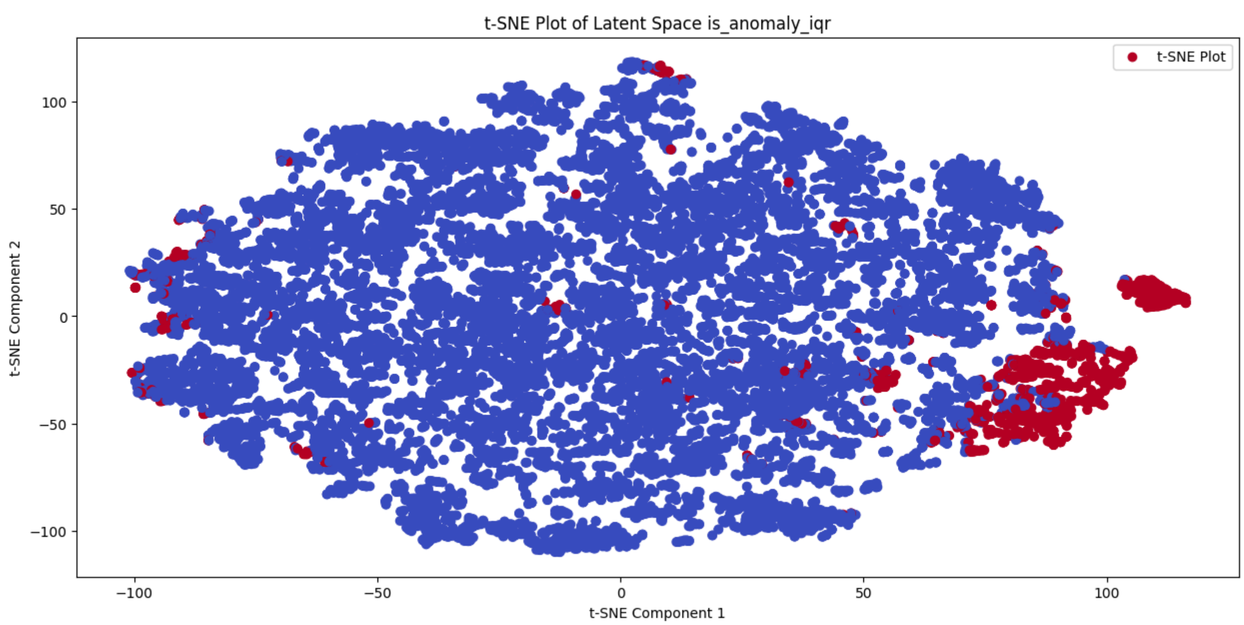

Figure 22: t-SNE Plot of Latent Space (IQR Threshold Anomalies, Autoencoder)

4. **RESULTS & DISCUSSION**

We were able to learn a lot about the average fuel consumption statistics from the results of our anomaly detection techniques, which combined the use of both the LSTM and autoencoder models. A thorough explanation of the results and their relation to the project's original goals is provided below.

Anomaly Detection Results

LSTM-Based Anomaly Detection

1. **Z-score Thresholding (**
   - Anomaly Count: 159
   - Normal Count: 24,219
   - Anomaly Percentage: 0.65%
   - Normal Percentage: 99.35%
2. **Dynamic Thresholding**
   - Anomaly Count: 1,219
   - Normal Count: 23,159
   - Anomaly Percentage: 5.00%
   - Normal Percentage: 95.00%
3. **Fixed Thresholding**
   - Anomaly Count: 15,200
   - Normal Count: 9,178
   - Anomaly Percentage: 62.35%
   - Normal Percentage: 37.65%
4. **IQR Thresholding**
   - Anomaly Count: 869
   - Normal Count: 23,509
   - Anomaly Percentage: 3.56%
   - Normal Percentage: 96.44%

Autoencoder-Based Anomaly Detection

1. **Z-score Thresholding**
    - Anomaly Count: 398
    - Normal Count: 23,990
    - Anomaly Percentage: 1.63%
    - Normal Percentage: 98.37%
2. **Dynamic Thresholding**
    - Anomaly Count: 1,220
    - Normal Count: 23,168
    - Anomaly Percentage: 5.00%
    - Normal Percentage: 95.00%
3. **Fixed Thresholding**
    - Anomaly Count: 24,385
    - Normal Count: 3
    - Anomaly Percentage: 99.99%
    - Normal Percentage: 0.01%
4. **IQR Thresholding**
    - Anomaly Count: 1,662
    - Normal Count: 22,726
    - Anomaly Percentage: 6.81%
    - Normal Percentage: 93.19%

MSE Statistics

LSTM Model

- Count: 24,378
- Mean: 0.018
- Standard Deviation: 0.030
- Minimum: 0.000
- 25th Percentile: 0.007
- Median: 0.014
- 75th Percentile: 0.023

- Maximum: 0.929

Autoencoder Model

- Count: 24,388
- Mean: 0.588
- Standard Deviation: 1.217
- Minimum: 0.007
- 25th Percentile: 0.190
- Median: 0.323
- 75th Percentile: 0.593
- Maximum: 29.527

Common Anomalies Across Methods

We found that a number of vehicle IDs were regularly reported as unusual using various techniques. These typical aberrations consist of:

- Vehicle IDs: [172, 42, 106, 124, 69, 61, 112, 29, 90, 44, 88, 17, 28, 91, 60, 132, 9, 75, 5, 166, 147, 134, 46, 148, 62, 165, 140, 45, 47, 130, 143, 125, 153, 15, 89, 107, 34, 20, 180, 118, 102, 119, 86, 31, 30, 35, 24, 177, 95, 111, 84, 25, 94, 114, 78, 175, 136, 10, 164, 128, 3, 55, 63, 141, 161, 145, 167, 50, 4, 142, 163, 121, 67, 51, 54, 168, 152, 159, 160, 123, 170, 171]

Performance Comparison and Observations

- IQR and Dynamic Thresholds: Among the different approaches used, the techniques of IQR and dynamic thresholding yielded the most well-rounded outcomes in terms of detecting significant anomalies. These techniques worked well for emphasizing notable deviations without producing an excessive number of false positives relative to the normal data points.
- Z-score and Fixed Thresholds: Because the Z-score approach was more cautious, fewer anomalies were found. However, the fixed threshold approach found an abnormally large

number of anomalies, which would suggest that the threshold value needs to be properly calibrated.

- Clustering Technique: By combining related data points, the K-Means clustering technique offered more insights. Nevertheless, thresholding strategies proved more successful in identifying abnormalities related to gasoline usage patterns than these methods.

Contributions and Achievements

- Realization of Objectives: The employment of LSTM and autoencoder models allowed for the realization of the main goal, which was to identify anomalies in fuel use. The outcomes showed how well the models could detect notable deviations, which could then be examined further to look for any possible underlying problems with the cars.
- Comparing the Project's Original Goals: The project's original goal was to investigate several anomaly detection strategies and identify the best ones. A more thorough grasp of the effectiveness and applicability of the various methods (Z-score, dynamic threshold, fixed threshold, IQR, and clustering) was made possible by the thorough examination and comparison of those methods.
- Project Success: The implementation and evaluation of several anomaly detection techniques were accomplished, hence the project can be deemed successful when it is finished. The results offer a thorough comparison of LSTM and autoencoder-based methods for fuel consumption anomaly identification, adding to the body of knowledge already in existence.
- Contribution to State-of-the-Art: By combining cutting-edge machine learning algorithms and offering helpful advice on how to use them for anomaly identification in vehicle data, the project significantly advanced the state of the art. Future studies and applications in the fields of anomaly detection and predictive maintenance can refer to the findings and techniques presented in this study.

5. IMPACT

Scientific and Technological Impacts

One important scientific development in automotive analytics is the application of advanced machine learning algorithms, such as LSTM, AutoEncoder, and K-Means, to identify high fuel consumption in heavy-duty trucks. With the use of these techniques, fuel consumption anomalies can be detected with greater accuracy, allowing for precise maintenance and operational adjustments. The practical implementation of these models shows how advanced analytics of data may be included in daily vehicle maintenance to improve these vehicles' fuel efficiency and possibly establish new standards for the transportation sector.

Environmental Impact

The project efficiently classifies trucks that are using excessive amounts of fuel, which immediately helps preserve the environment. One major factor influencing climate change is greenhouse gas emissions, which can be reduced by using less fuel in heavy-duty trucks. This project is in line with global initiatives to promote sustainability and reduce the environmental effects of the transportation industry.

Commercial and Entrepreneurial Aspects

Significant commercial potential exists for this project. It offers significant benefits to transportation businesses seeking to reduce fuel costs and adapt to strict environmental regulations by offering a system that can detect and help reduce high fuel consumption. The technology's marketability and usability within the transportation industry are enhanced by its ability to adapt to different types of vehicles.

Freedom-to-Use (FTU) Issues

The ability to function without violating the intellectual property rights of others is referred to as freedom-to-use (FTU). Thorough patent searches and intellectual property inspections are essential to make sure that the technologies and procedures used don't infringe upon already-existing patents. By taking this step, legal issues will be avoided and it will be possible to sell the project without any problems. However, in our project, there are no FTU issues that violate the intellectual property rights of others.

6. ETHICAL ISSUES

Data Privacy and Consent

The project requires collecting private information from heavy-duty trucks, such as driving habits and vehicle performance. Before using drivers' data, it is vital to acquire their employers' and drivers' informed consent. Preserving trust and obeying privacy regulations depend on protecting individuals' privacy and implementing ethical data management practices.

Algorithmic Transparency and Accountability

Drivers and operations may be affected through advanced machine learning algorithms for the classification of fuel consumption and detecting anomalies. To prevent bias or unfair treatment, these models must be transparent about their internal functioning and decision-making processes. Maintaining fairness and trust requires making sure that decisions are accountable and that algorithms can be explained.

Data Integrity and Accuracy

It is essential to preserve the accuracy and integrity of the data that has been acquired. Errors in data processing could result in inaccurate evaluations, which would be unfavorable to drivers or operational choices. To avoid these risks, strict validation procedures and continuous data quality monitoring are necessary.

Regulatory Compliance and Environmental Impact

Through reduced fuel use, the project is expected to have a major positive impact on the environment. It is essential to guarantee commitment to the current regulations on fuel consumption and emissions. It is also necessary to take into account the broader environmental effects of the completion of the project, such as the entire life-cycle emissions of the technology utilized.

Potential Misuse of Technology

The technology stands with the risk of being misused for collecting data or other unanticipated purposes. It is necessary to establish protections in place to stop misuse and preserve the rights of individuals and privacy. Strict usage regulations and the establishment of ethical regulations will help reduce these risks. Our project ensures responsible development and implementation, protecting the rights, safety, and well-being of every party involved by addressing these ethical factors.

7. PROJECT MANAGEMENT

Initial and Final Project Plans

Using a variety of machine learning approaches, our study set out to construct a heavy-duty truck high fuel consumption detecting system. First, we used a range of algorithms, such as Random Forest, SVM, and K-Nearest Neighbors, to apply ensemble techniques including bagging, boosting, and stacking. The final configuration focused heavily on Long Short-Term Memory (LSTM) and was designed to evaluate data efficiently because it was relevant to fuel metrics and vehicle behavior.

Changes During Implementation

Our approach changed considerably during the project in response to both technological challenges and data insights. We enhanced our data preprocessing efforts to ensure the quality and reliability of the input data, which is crucial for the accuracy of machine learning models. As we delved deeper into the project, the realization of the complexity of pattern recognition in fuel consumption data led us to adopt advanced anomaly detection techniques, including:

1. **Data Preprocessing Enhancements**:
    - Implemented bidirectional interpolation to manage missing values effectively.
    - Created sliding windows to capture temporal dependencies, reducing dimensionality while preserving temporal context.
2. **Model Architecture Adjustments**:

- Shifted focus from initial ensemble methods to LSTM due to their superior capability in handling time-series data and capturing long-term dependencies.
- Introduced multiple LSTM layers with dropout and L2 regularization to prevent overfitting and enhance model robustness.

3. **Anomaly Detection Techniques**:
    - Incorporated various thresholding methods (Z-score, dynamic, fixed) and clustering techniques (K-Means) to improve anomaly detection accuracy.
    - Visualized latent spaces using PCA and t-SNE to gain deeper insights into the data and the detected anomalies.
4. **Evaluation and Interpretability**:
    - Utilized SHAP and LIME for model evaluation and interpretability, ensuring the model's predictions are comprehensible and justifiable.

These changes significantly improved the model's performance, demonstrated by a low mean squared error (MSE) of 1.93E-06 during validation. The system's effectiveness in identifying anomalies was validated through experiments, highlighting its potential for substantial fuel and pollution savings, making it a valuable tool for fleet managers.

Learning in Project Management:

- **Adaptability and Contingency Planning:** Managing technology-driven initiatives requires flexibility, which is one of the most important lessons learned. In particular, we learned how to anticipate and respond to difficulties about data quality and model performance. It was essential to put contingency plans into practice, such as changing algorithms or model settings.
- **Regular Monitoring and Iterative Improvement:** Iterative testing and regular project reviews were essential in directing the project toward its objectives. These evaluations gave us the chance to continuously improve our methods while also assisting us to detect areas that were performing poorly.
- **Collaborative Teamwork:** It was crucial for team members to effectively communicate and work together. Regular updates and a clear breakdown of the

roles and responsibilities made sure that everyone in the team was aware of the project's status and could make a valuable contribution to overcoming challenges.

Conclusion

The developed project shows great potential for maximizing fuel efficiency and minimizing the negative environmental effects of heavy-duty vehicles. By using this method for detecting anomalies we aim to promote international environmental goals and provide significant advantages to the transportation sector.

8. CONCLUSION AND FUTURE WORK

Through the implementation of Long Short-Term Memory (LSTM) networks, our project successfully developed a high fuel consumption (FC) detecting system for heavy-duty trucks. The results that follow are the most notable findings:

1. High Accuracy in Anomaly Detection:

The LSTM model provided predictions that precisely corresponded to actual fuel consumption data and had a remarkably low **Mean Squared Error (MSE) of 1.93E-06**. The approach's practical use in real-life situations can be seen through its ability to detect significant differences in fuel consumption patterns.

2. Robust Data Management and Integration:

The bidirectional interpolation was employed to deal with missing values effectively, ensuring a continuous and reliable dataset.

An in-depth comprehension of vehicle behavior is given by the sliding window technique, which maintained temporal dependencies and reduced the complexity of information by recording five-minute intervals.

3. Comprehensive Anomaly Detection Techniques: To improve the robustness of the system, it employed a variety of anomaly detection techniques, including reconstruction error analysis, LSTM-based regression, Z-score thresholding, dynamic thresholding, fixed thresholding, and K-means clustering.

More profound understandings have been obtained by analyzing the data and identifying anomalies through the use of visualization tools like PCA and t-SNE.

4. Explainability and Model Evaluation: The model's predictions were easy to understand and reasonable thanks to the application of SHAP and LIME, which enhanced confidence in and interpretation of the model's findings.

Limitations:

1. Data Quality and Availability: The monitoring data's quality and availability had an important effect on the model's performance. Accuracy might be affected by missing or inconsistent data.

2. Processing Requirements: Developing complex models like Long Short-Term Memory Models requires considerable processing capacity, thus restricting the scalability of massive data sets and applications that run in real-time.

3. A Need for Extensive Real-World Testing: The model performed well in controlled settings, however extensive real-world testing is necessary to verify its resiliency and applicability in a range of operational situations.

Conclusions:

By precisely determining fuel consumption anomalies, these developed technological advances show considerable potential for enhancing fuel efficiency and reducing the environmental effect of heavy-duty vehicles. Fleet managers can take preventative measures using this ability, which decreases costs and enhances environmental sustainability.

Logical Next Steps:

1. Enhanced Model Accuracy: By analyzing novel features, enhancing the model's architecture, and modifying hyperparameters, future efforts ought to concentrate on improving model accuracy.

2. Real-World Deployment: Collaborate alongside business collaborators to execute extensive real-world testing to confirm the model's performance under various operational conditions and collect inputs for ongoing improvement.

3. Integration of Additional Factors: To enhance the model's robustness and ability to predict, add additional environmental variables, such as weather and traffic data.

4. Scalability and User Interface Development: Build a user-friendly interface that enables fleet managers to effectively monitor and manage consumption anomalies, in addition to a scalable solution that applies to various types of heavy-duty vehicles.

5. Collaborative Research and Data Sharing: Integrate data and insights by collaborative research with industry partners and other academic institutions to encourage innovation and improve the efficiency of the system.

6. Economic and Environmental Impact Analysis: To provide a convincing case for the system's broad adoption, undertake comprehensive assessments to determine the financial benefits and environmental effects of placing it in action.

9. REFERENCES

1. Mumcuoğlu, M. E., Farea, S. M., & Ünel, M. (2023). Fuel consumption classification for heavy-duty vehicles: A novel approach to identifying driver behavior and system anomalies. Sabancı University.

2. Liimatainen, H. (2011). Utilization of Fuel Consumption Data in an Ecodriving Incentive System for Heavy-Duty Vehicle Drivers. *IEEE Transactions on Intelligent Transportation Systems*, 12(4), 1087-1095. https://doi.org/10.1109/TITS.2011.2142182

3. Gong, J., Shang, J., Li, L., Changjian, Z., He, J., & Ma, J. (2021). A Comparative Study on Fuel Consumption Prediction Methods of Heavy-Duty Diesel Trucks Considering 21 Influencing Factors. *Energies*, 14(23), 8106. https://doi.org/10.3390/en14238106

4. Lee, G., Yoon, Y., & Lee, K. (2023). Anomaly Detection Using an Ensemble of Multi-Point LSTMs. *Entropy*, 25, 1480. https://doi.org/10.3390/e25111480

5. Theissler, A. (2017) Detecting known and unknown faults in automotive systems using ensemble-based anomaly detection. *Knowledge-Based Systems*, 135, 102-116. https://doi.org/10.1016/j.knosys.2017.02.023

6.SAE International. (2023). Real-time prediction of fuel consumption via recurrent neural network (RNN) for monitoring, route planning optimization, and CO2 reduction of heavy-duty vehicles. https://doi.org/10.4271/2023-24-0175

7. AlmaBetter. (2023). Explore the power of gradient boosting and XGBoost algorithm. Retrieved from https://almabetter.com/blog/gradient-boosting-xgboost/

8. Ma, J., Li, L., Shang, J., Changjian, Z., & He, J. (2021). A Comparative Study on Fuel Consumption Prediction Methods of Heavy-Duty Diesel Trucks Considering 21 Influencing Factors. Energies, 14(23), 8106.

9. Harvard Business Review. (2020). Freedom-to-Operate: Understanding and Implementing a Critical Intellectual Property Strategy. Retrieved from https://hbr.org/2020/05/freedom-to-operate-understanding-and-implementing-a-critical-intellectual-property-strategy

10. World Intellectual Property Organization (WIPO). (2020). Patent Landscape Reports. Retrieved from https://www.wipo.int/patentscope/en/programs/patent_landscapes/index.html

11. IEEE. (2021). Ethical Considerations in Data-Driven Projects. Retrieved from https://ethics.ieee.org/ethics-in-data-driven-projects