Time-Series Modeling of Methane Emission Events Using Machine Learning Forecasting Algorithms

SEMIU TEMIDAYO FASASI[1], OLUWAPELUMI JOSEPH ADEBOWALE[2], ABDULMALIQ ABDULSALAM[3], ZAMATHULA QUEEN SIKHAKHANE NWOKEDIEGWU[4]
[1]*Independent Researcher, Nigeria*
[2]*Independent Researcher, USA*
[3]*Independent Researcher, Durban, South Africa*
[4]*Department of Chemical and Petroleum Engineering, University of Lagos, Nigeria*

Abstract-* *Methane is a significant contributor to global warming, necessitating accurate monitoring and forecasting of its emissions to inform effective mitigation strategies. This paper investigates the application of machine learning algorithms for time-series modeling of methane emission events, addressing the challenges posed by the complex, non-linear, and noisy nature of environmental data. A comprehensive methodology is developed, incorporating advanced data preprocessing techniques and the evaluation of multiple forecasting models, including Long Short-Term Memory networks and ensemble methods such as Random Forest and Gradient Boosting. The comparative analysis demonstrates that these machine learning approaches outperform traditional statistical methods in capturing temporal dependencies and episodic emission spikes. Furthermore, the inclusion of contextual environmental variables enhances prediction accuracy and interpretability. The study highlights the potential of machine learning to provide reliable, actionable forecasts that support proactive environmental monitoring, regulatory compliance, and emission reduction efforts. Key challenges such as data quality, model interpretability, and computational demands are discussed, along with recommendations for future research focusing on multimodal data integration and adaptive learning frameworks. This work contributes to advancing data-driven approaches for methane emission forecasting, offering valuable insights for environmental scientists and policymakers engaged in climate change mitigation.

Index Terms : Methane Emissions, Time-Series Forecasting, Machine Learning, Long Short-Term Memory, Environmental Monitoring, Emission Prediction

I. INTRODUCTION

1.1 Background

Methane is a potent greenhouse gas with a global warming potential significantly higher than carbon dioxide over a short timeframe, making its accurate monitoring a critical component of climate change mitigation efforts (Balcombe et al., 2018, Lashof and Ahuja, 1990). Methane emissions arise from various sources including natural wetlands, fossil fuel extraction, agriculture, and waste management (Balcombe et al., 2018). Given the complexity and variability of these sources, continuous monitoring is essential for understanding emission patterns and informing effective regulatory and mitigation strategies. Time-series data collected from methane sensors provide valuable temporal insights into emission dynamics, revealing fluctuations and episodic release events (Dean et al., 2018, Kumar, 2018).

However, the inherent variability in methane emissions, caused by factors such as weather conditions, operational activities, and equipment malfunctions, poses substantial challenges to conventional forecasting methods (Wuebbles and Hayhoe, 2002). Traditional statistical models often struggle to capture the non-linear and stochastic nature of these time-series, limiting their predictive accuracy (Balcombe et al., 2017). Moreover, the spatial and temporal sparsity of data further complicates efforts to develop robust models. This motivates the exploration of advanced machine learning algorithms, which have shown promise in modeling complex, non-linear systems in environmental sciences (Ausubel et al., 1988).

The integration of machine learning with time-series data from methane monitoring has the potential to transform forecasting capabilities. These techniques can leverage historical data to identify hidden patterns and predict future emission events with higher accuracy and reliability, supporting proactive decision-making in environmental management (Howarth, 2014). This growing intersection between environmental monitoring and artificial intelligence forms the foundation of this study

1.2 Problem Statement

Accurately modeling methane emission events remains a significant challenge due to the dynamic and complex behavior of methane sources. Traditional forecasting approaches, including linear regression and classical time-series methods such as ARIMA, often rely on assumptions of stationarity and linear relationships that do not hold for methane emissions (Kumar, 2018). These methods may fail to capture sudden spikes or irregular fluctuations that characterize real-world methane release events, reducing their utility for early warning systems and mitigation planning (Holmes et al., 2013).

Additionally, methane emission datasets frequently suffer from noise, missing values, and irregular sampling intervals, further hindering the effectiveness of conventional models. This data quality issue complicates model training and decreases the reliability of forecasts. Moreover, the interaction between methane emissions and external environmental factors, such as temperature, pressure, and wind speed, introduces additional layers of complexity, requiring models capable of handling multiple, potentially interdependent variables (Karpatne et al., 2018).

There is a pressing need for forecasting frameworks that can overcome these limitations by capturing non-linear dependencies, learning from incomplete or noisy data, and adapting to changing emission dynamics over time (Dong et al., 2016). Machine learning algorithms, particularly those designed for sequential data, offer promising solutions to these challenges. However, understanding their application and effectiveness in the context of methane emission forecasting remains underexplored, creating a gap that this paper seeks to address.

1.3 Objectives

This paper aims to advance the field of methane emission forecasting by systematically exploring machine learning algorithms tailored for time-series modeling. The primary objective is to develop and evaluate forecasting approaches that can accurately predict methane emission events, thereby supporting more effective environmental monitoring and mitigation strategies. By focusing on machine learning techniques, this work intends to demonstrate how these methods can overcome the inherent challenges of methane data, such as non-linearity, noise, and irregularity.

The contributions of this study include a comprehensive review and comparison of selected algorithms suitable for methane time-series forecasting, including recurrent neural networks and ensemble-based models. Furthermore, the paper outlines best practices for data preprocessing and feature engineering that enhance model performance in this domain. It also discusses the selection of appropriate evaluation metrics to ensure that forecasting accuracy is rigorously assessed.

Ultimately, this research provides valuable insights into the practical application of machine learning for environmental data analysis, highlighting its potential to improve prediction accuracy and support real-time emission management. Doing so contributes to the broader effort of leveraging advanced analytics for climate change mitigation and environmental protection.

II. LITERATURE REVIEW

2.1 Methane Emission Monitoring Techniques

Methane emission monitoring has evolved significantly with advances in sensor technology and remote sensing platforms (Fredenslund et al., 2018). Ground-based methods include fixed sensors installed at emission sites such as landfills, oil and gas facilities, and agricultural operations. These sensors provide continuous, high-frequency measurements, allowing for detailed temporal tracking of methane levels (EYINADE et al., 2020). However, they are limited by their fixed locations and may not capture emissions occurring outside their immediate vicinity. Mobile monitoring techniques, such as vehicle-mounted sensors and drones, have been increasingly deployed to overcome spatial limitations. These methods offer the flexibility to survey large areas and identify methane plumes with improved spatial resolution (Odedeyi et al., 2020, OGUNNOWO et al., 2020).

Remote sensing techniques complement ground-based methods by providing broader spatial coverage. Satellite-based instruments have been utilized to detect methane concentrations over regional and global scales (Okuh et al.). Although these methods can monitor vast areas, their temporal resolution is often constrained by satellite overpass frequency and atmospheric conditions, which can limit the detection of short-term emission events. Airborne sensors mounted on aircraft also fill the gap by offering high-resolution data collection over targeted areas, combining spatial coverage with relatively high temporal resolution (Adewoyin et al., 2020b, ADEWOYIN et al., 2020a).

Data collection practices in methane monitoring involve challenges such as sensor calibration, data noise, and environmental interferences that impact data quality. Standardized protocols for data acquisition and preprocessing are critical to ensure consistency and reliability (Ogunnowo). Additionally, integration of multi-source data, including meteorological information, enhances the contextual understanding of emission events. This diverse array of monitoring techniques provides the foundational data required for accurate time-series analysis and forecasting (Gbabo et al., Okuh et al.).

2.2 Time-Series Forecasting in Environmental Applications

Time-series forecasting plays a vital role in environmental science by enabling prediction of various phenomena such as air pollution, water quality, and greenhouse gas emissions. Classical forecasting methods, including Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing, have been widely used due to their simplicity and interpretability (Fox et al., 2019). These models work well with linear and stationary time-series data, providing baseline predictive capabilities. However, environmental data, including methane emissions, often exhibit non-linear dynamics and non-stationary behavior that limit the effectiveness of these traditional approaches (Gbabo et al.).

To address these challenges, hybrid and more sophisticated models have been explored. For example, wavelet transforms combined with ARIMA have been used to capture both frequency and temporal variations in air quality data (Vincent et al., 2020). Seasonal decomposition methods help account for periodic patterns common in environmental datasets. In recent years, state-space models and Kalman filtering techniques have been adopted for their ability to handle noisy and incomplete data. Despite these advancements, many conventional models struggle with the irregularities and sudden spikes that characterize methane emission events (Thompson et al., 2015).

The complexity and stochastic nature of environmental time-series have spurred interest in machine learning and deep learning methods that can model non-linearity and complex temporal dependencies more effectively. These data-driven approaches offer flexibility in capturing patterns without strict assumptions about data distribution, making them well-suited for dynamic environmental applications such as methane emission forecasting.

2.3 Machine Learning Algorithms for Time-Series Forecasting

Machine learning has become an essential tool for time-series forecasting, particularly in domains where traditional models face limitations. Among the most popular techniques are Long Short-Term Memory (LSTM) networks, a type of recurrent neural network designed to capture long-range dependencies in sequential data (Lippi et al., 2013). LSTMs address the vanishing gradient problem encountered in standard RNNs, enabling effective modeling of complex temporal patterns and irregularities. They have been successfully applied to various environmental time-series, demonstrating superior performance in capturing both short-term fluctuations and long-term trends (Han et al., 2019, Bontempi et al., 2012).

Ensemble learning methods such as Random Forest and Gradient Boosting have also gained prominence in forecasting applications. These algorithms combine multiple decision trees to improve predictive accuracy and reduce overfitting. Random Forests are valued for their robustness to noise and ability to handle non-linear relationships without extensive parameter tuning (Naghibi et al., 2020, Kalusivalingam et al., 2020a). Gradient Boosting methods, including XGBoost and LightGBM, focus on minimizing prediction errors through sequential learning, often achieving state-of-the-art results in regression tasks with environmental data (Sahin, 2020, Callens et al., 2020).

Additionally, hybrid approaches that integrate machine learning with traditional statistical techniques or domain knowledge are increasingly explored to leverage the strengths of both paradigms. Feature engineering, including the incorporation of lag variables, rolling statistics, and external factors like weather data, enhances model inputs and forecasting performance (Naganna et al., 2020). The growing availability of computational resources and data has accelerated the adoption of these algorithms, positioning them as powerful tools for advancing methane emission time-series forecasting (Lieske et al., 2018).

III. METHODOLOGY

3.1 Data Characteristics and Preprocessing

Methane emission time-series data typically exhibit high variability, irregular spikes, and seasonal or diurnal patterns due to natural and anthropogenic influences. The data often contains noise from sensor inaccuracies, environmental interference, and operational inconsistencies (Whalen and Reeburgh, 1992). Additionally, measurements can be irregularly spaced or contain missing values, complicating analysis. Typical features include methane concentration levels recorded at fixed intervals, sometimes augmented with contextual variables such as temperature, pressure, wind speed, or facility operational status to provide explanatory power for emission fluctuations (Chaurasia and Pal, 2020).

Effective preprocessing is critical to enhance data quality and prepare inputs for machine learning models. Normalization or standardization is commonly applied to scale features within a consistent range, improving model convergence and stability. Missing data must be addressed through interpolation methods, such as linear interpolation or more advanced techniques like K-nearest neighbors imputation, to maintain continuity in the time-series without introducing bias. Outlier detection and removal are also necessary to reduce the impact of anomalous readings that could skew the training process (Morin et al., 2014).

Feature engineering plays a vital role in capturing temporal dependencies and environmental context. Common engineered features include lagged variables representing past methane levels, moving averages, and rolling window statistics that summarize short-term trends. Additionally, incorporating external meteorological or operational data as features allows models to capture better the factors driving emission variability. This comprehensive preprocessing pipeline lays the groundwork for robust and accurate forecasting models (Kang and Tian, 2018, Obaid et al., 2019).

3.2 Machine Learning Model Selection

The choice of machine learning models for forecasting methane emission events hinges on their ability to capture complex temporal dynamics, handle noisy data, and model non-linear relationships (Luengo et al., 2020). Recurrent neural networks, particularly Long Short-Term Memory networks, are well-suited for this task due to their specialized architecture designed to learn from sequential data with long-range dependencies. LSTMs use memory cells and gating mechanisms to retain relevant information across time steps, enabling them to model temporal patterns and abrupt changes typical of methane emission time-series (Alexandropoulos et al., 2019).

Ensemble learning models like Random Forest and Gradient Boosting machines are also valuable for forecasting due to their robustness and flexibility. Random Forest builds multiple decision trees on random subsets of data and features, aggregating their predictions to reduce overfitting and improve generalization. Gradient Boosting sequentially fits trees to correct errors made by prior models, optimizing predictive accuracy through gradient descent techniques. Both methods efficiently handle non-linearities and interactions between variables, making them effective when combined with well-engineered features (Boppiniti, 2020, Gibert et al., 2016).

These models complement each other: LSTMs excel at capturing temporal dependencies directly from raw sequences, while ensemble models leverage engineered features to identify complex patterns. Selecting and comparing these algorithms allows the study to identify the most effective approach for methane emission forecasting under varying data conditions, providing a comprehensive methodology tailored to the domain's challenges (Federico et al., 2020).

3.3 Model Training and Evaluation Metrics

Training machine learning models for methane emission forecasting requires careful consideration of data splitting and validation to avoid overfitting and ensure generalization (Saleh et al., 2016). Time-series data necessitates sequential splitting, where past data is used for training and future periods reserved for validation and testing (Lepperød, 2019). Techniques such as rolling-origin evaluation or walk-forward validation are often employed, iteratively training the model on progressively larger datasets and testing on subsequent unseen periods, closely mimicking real-world forecasting scenarios (Pawłowski and Kurach, 2015, Arienti, 2020).

Model training involves optimizing algorithm-specific parameters using training data. For neural networks, this includes adjusting weights via backpropagation and tuning hyperparameters like learning rate, number of layers, and neurons per layer. For ensemble methods, hyperparameters such as the number of trees, maximum depth, and learning rate are optimized, often through grid search or randomized search. Cross-validation adapted for time-series further supports hyperparameter tuning while maintaining temporal integrity (Olof, 2018).

Evaluating forecasting performance requires metrics that quantify prediction accuracy and error magnitude. Commonly used metrics include Root Mean Squared Error (RMSE), which penalizes large errors more heavily, and Mean Absolute Error (MAE), which measures average magnitude of errors, providing interpretable results. Additional metrics like Mean Absolute Percentage Error (MAPE) may be used to express accuracy in relative terms. These metrics collectively guide model selection and refinement, ensuring reliable and actionable methane emission forecasts (Prayogo and Susanto, 2018).

IV. RESULTS AND DISCUSSION

4.1 Performance Comparison of Models

The comparative analysis of machine learning models reveals notable differences in their effectiveness at forecasting methane emissions. Recurrent neural networks, specifically Long Short-Term Memory networks, generally demonstrate superior performance in capturing temporal dependencies inherent in methane emission time-series (Li et al., 2020). Their ability to model long-range correlations and abrupt fluctuations leads to lower error metrics such as RMSE and MAE compared to ensemble models. This advantage becomes particularly evident in datasets with complex seasonality and sudden emission spikes, where LSTMs leverage their memory mechanisms to adapt to changing patterns (Sahin, 2020).

Ensemble models like Random Forest and Gradient Boosting exhibit competitive results, particularly when rich engineered features are available. Random Forest's robustness to noise and its ensemble averaging process result in stable predictions with reasonable accuracy, albeit slightly less precise than LSTMs in temporal sequence modeling. Gradient Boosting methods often yield high accuracy through iterative error correction, but their performance depends heavily on careful hyperparameter tuning and feature selection, highlighting the importance of domain knowledge in preprocessing (Kalusivalingam et al., 2020a).

Overall, the results underscore the complementary strengths of these approaches. LSTMs excel in raw sequential modeling, while ensemble methods benefit from well-constructed features that capture contextual information. The evaluation metrics collectively indicate that no single model universally dominates, and hybrid strategies may offer enhanced forecasting reliability for methane emissions (Arnaudo et al., 2020, Kalusivalingam et al., 2020b).

4.2 Interpretation of Forecasting Outcomes

The forecasting outcomes provide valuable insights into methane emission dynamics and the predictability of emission events. The models successfully identify underlying temporal trends, such as daily and seasonal cycles driven by environmental conditions and operational schedules. Moreover, they demonstrate the ability to anticipate episodic spikes, which are critical for timely mitigation and response. This predictive capability suggests that machine learning models can detect subtle precursors embedded in the data, such as gradual increases in baseline emissions or correlated meteorological changes, that precede significant release events.

However, predictability varies depending on the magnitude and duration of emission events. Short, sudden spikes are inherently more challenging to forecast accurately due to their sporadic nature and potential measurement noise. Longer-term trends and recurring patterns are captured more reliably, enabling strategic planning and regulatory compliance. The inclusion of external features such as temperature and wind speed enhances the models' explanatory power, indicating that methane emissions are influenced by a complex interplay of factors that can be exploited for better forecasting.

These findings highlight the potential of machine learning not only as a predictive tool but also as an analytical lens for understanding methane emission behaviors. They emphasize the need for continuous data integration and model refinement to improve forecasting robustness in real-world applications.

4.3 Challenges and Limitations in Modeling

Despite promising results, several challenges and limitations affect the modeling of methane emissions using machine learning. Data quality issues, including missing values, sensor noise, and irregular sampling intervals, remain significant obstacles that require sophisticated preprocessing. These issues can introduce biases and reduce model generalizability, especially when training data is limited or unrepresentative of future conditions. Furthermore, the spatial heterogeneity of methane sources complicates the extrapolation of model results from specific monitoring sites to broader regions.

Another challenge lies in the interpretability of complex models, particularly deep learning architectures like LSTMs. While these models offer superior predictive performance, their "black-box" nature can limit understanding of the causal relationships driving emissions, which is crucial for policy and operational decision-making. Efforts to incorporate explainability techniques and domain expertise are necessary to bridge this gap.

Additionally, computational demands and the need for extensive hyperparameter tuning can pose practical constraints, especially in resource-limited environments. Model maintenance and updating require continuous data inflow and adaptation to changing emission patterns, emphasizing the importance of sustainable data management practices. Addressing these challenges is essential for translating machine learning forecasts into actionable environmental management solutions.

CONCLUSION

This study demonstrates that machine learning techniques offer substantial improvements in forecasting methane emission events compared to traditional methods. Long Short-Term Memory networks, with their ability to model long-term dependencies and complex temporal patterns, consistently outperform ensemble methods in capturing the dynamic and non-linear nature of methane time-series data. Ensemble algorithms such as Random Forest and Gradient Boosting also provide robust performance when combined with effective feature engineering, showcasing their adaptability in handling noisy and irregular environmental data.

The integration of meteorological and operational variables as additional features enhances forecasting accuracy by accounting for external factors influencing emission variability. Together, these findings underscore the importance of advanced data preprocessing and model selection tailored to the unique characteristics of methane emissions. The results confirm that machine learning can identify meaningful temporal trends and anticipate episodic emission spikes, which are critical for effective environmental monitoring and management.

Improved forecasting of methane emissions has significant implications for environmental monitoring and policy-making. Enhanced prediction accuracy enables earlier detection of emission spikes, allowing for timely interventions that can reduce greenhouse gas release and associated climate impacts. Reliable forecasts contribute to more effective allocation of monitoring resources by targeting high-risk periods and locations, thereby optimizing operational efficiency and reducing costs.

At the policy level, data-driven forecasting models support evidence-based decision-making, providing regulators with actionable insights to enforce emission limits and design incentive mechanisms. Accurate predictions also facilitate compliance verification and transparent reporting, fostering accountability among methane emitters. Furthermore, integrating forecasting tools within environmental management systems enhances the ability to assess the effectiveness of mitigation measures and adapt strategies dynamically in response to observed trends. Future research should explore the integration of multimodal data sources, such as satellite imagery, atmospheric transport models, and real-time sensor networks, to further improve the spatial and temporal resolution of methane emission forecasts. Advances in explainable artificial intelligence (XAI) are also needed to enhance the interpretability of complex models, making them more accessible and trustworthy for stakeholders in environmental policy and industry.

The development of adaptive learning frameworks that continuously update models with incoming data can help maintain forecasting accuracy in the face of evolving emission patterns and operational changes. Additionally, investigating transfer learning approaches may enable the application of models trained in one context to other regions or sectors with limited data availability. Lastly, interdisciplinary collaboration combining atmospheric science, data analytics, and environmental policy will be essential to translate forecasting advancements into practical mitigation solutions. Such efforts will contribute to building resilient monitoring systems capable of addressing the urgent challenge of methane-driven climate change.

REFERENCES

[1] ADEWOYIN, M. A., OGUNNOWO, E. O., FIEMOTONGHA, J. E., IGUNMA, T. O. & ADELEKE, A. K. 2020a. Advances in Thermofluid Simulation for Heat Transfer Optimization in Compact Mechanical Devices.

[2] ADEWOYIN, M. A., OGUNNOWO, E. O., FIEMOTONGHA, J. E., IGUNMA, T. O. & ADELEKE, A. K. 2020b. A Conceptual Framework for Dynamic Mechanical Analysis in High-Performance Material Selection.

[3] ALEXANDROPOULOS, S.-A. N., KOTSIANTIS, S. B. & VRAHATIS, M. N. 2019. Data preprocessing in predictive data mining. *The Knowledge Engineering Review,* 34, e1.

[4] ARIENTI, J. H. L. 2020. *Time Series Forecasting Applied to an Energy Management System-A Comparison Between Deep Learning Models and Other Machine Learning Models.* Universidade NOVA de Lisboa (Portugal).

[5] ARNAUDO, E., FARASIN, A. & ROSSI, C. 2020. A comparative analysis for air quality

estimation from traffic and meteorological data. *Applied Sciences,* 10, 4587.

[6] AUSUBEL, J. H., GRÜBLER, A. & NAKICENOVIC, N. 1988. Carbon dioxide emissions in a methane economy. *Climatic Change,* 12, 245-263.

[7] BALCOMBE, P., ANDERSON, K., SPEIRS, J., BRANDON, N. & HAWKES, A. 2017. The natural gas supply chain: the importance of methane and carbon dioxide emissions. *ACS Sustainable Chemistry & Engineering,* 5, 3-20.

[8] BALCOMBE, P., SPEIRS, J. F., BRANDON, N. P. & HAWKES, A. D. 2018. Methane emissions: choosing the right climate metric and time horizon. *Environmental Science: Processes & Impacts,* 20, 1323-1339.

[9] BONTEMPI, G., BEN TAIEB, S. & LE BORGNE, Y.-A. 2012. Machine learning strategies for time series forecasting. *European Big Data Management and Analytics Summer School.* Springer.

[10] BOPPINITI, S. T. 2020. Big Data Meets Machine Learning: Strategies for Efficient Data Processing and Analysis in Large Datasets. *International Journal of Creative Research In Computer Technology and Design,* 2.

[11] CALLENS, A., MORICHON, D., ABADIE, S., DELPEY, M. & LIQUET, B. 2020. Using Random forest and Gradient boosting trees to improve wave forecast at a specific location. *Applied Ocean Research,* 104, 102339.

[12] CHAURASIA, V. & PAL, S. 2020. Applications of machine learning techniques to predict diagnostic breast cancer. *SN Computer Science,* 1, 270.

[13] DEAN, J. F., MIDDELBURG, J. J., RÖCKMANN, T., AERTS, R., BLAUW, L. G., EGGER, M., JETTEN, M. S., DE JONG, A. E., MEISEL, O. H. & RASIGRAF, O. 2018. Methane feedbacks to the global climate system in a warmer world. *Reviews of Geophysics,* 56, 207-250.

[14] DONG, Y., MA, X., MA, C. & WANG, J. 2016. Research and application of a hybrid forecasting model based on data decomposition for electrical load forecasting. *Energies,* 9, 1050.

[15] EYINADE, W., EZEILO, O. J. & OGUNDEJI, I. A. 2020. A Treasury Management Model for Predicting Liquidity Risk in Dynamic Emerging Market Energy Sectors.

[16] FEDERICO, A., SERRA, A., HA, M. K., KOHONEN, P., CHOI, J.-S., LIAMPA, I., NYMARK, P., SANABRIA, N., CATTELANI, L. & FRATELLO, M. 2020. Transcriptomics in toxicogenomics, part II: preprocessing and differential expression analysis for high quality data. *Nanomaterials,* 10, 903.

[17] FOX, T. A., BARCHYN, T. E., RISK, D., RAVIKUMAR, A. P. & HUGENHOLTZ, C. H. 2019. A review of close-range and screening technologies for mitigating fugitive methane emissions in upstream oil and gas. *Environmental Research Letters,* 14, 053002.

[18] FREDENSLUND, A. M., HINGE, J., HOLMGREN, M. A., RASMUSSEN, S. G. & SCHEUTZ, C. 2018. On-site and ground-based remote sensing measurements of methane emissions from four biogas plants: A comparison study. *Bioresource technology,* 270, 88-95.

[19] GBABO, E. Y., OKENWA, O. K. & CHIMA, P. E. Constructing AI-Enabled Compliance Automation Models for Real-Time Regulatory Reporting in Energy Systems.

[20] GBABO, E. Y., OKENWA, O. K. & CHIMA, P. E. Integrating CDM Regulations into Role-Based Compliance Models for Energy Infrastructure Projects.

[21] GIBERT, K., SÀNCHEZ–MARRÈ, M. & IZQUIERDO, J. 2016. A survey on pre-processing techniques: Relevant issues in the context of environmental data mining. *Ai Communications,* 29, 627-663.

[22] HAN, Z., ZHAO, J., LEUNG, H., MA, K. F. & WANG, W. 2019. A review of deep learning models for time series prediction. *IEEE Sensors Journal,* 21, 7833-7848.

[23] HOLMES, C. D., PRATHER, M. J., SØVDE, O. & MYHRE, G. 2013. Future methane, hydroxyl, and their uncertainties: key climate and emission parameters for future predictions. *Atmospheric Chemistry and Physics,* 13, 285-302.

[24] HOWARTH, R. W. 2014. A bridge to nowhere: methane emissions and the greenhouse gas footprint of natural gas. *Energy Science & Engineering,* 2, 47-60.

[25] KALUSIVALINGAM, A. K., SHARMA, A., PATEL, N. & SINGH, V. 2020a. Enhancing Predictive Business Analytics with Deep Learning and Ensemble Methods: A Comparative Study of LSTM Networks and Random Forest Algorithms. *International Journal of AI and ML,* 1.

[26] KALUSIVALINGAM, A. K., SHARMA, A., PATEL, N. & SINGH, V. 2020b. Enhancing Supply Chain Visibility through AI: Implementing Neural Networks and Reinforcement Learning Algorithms. *International Journal of AI and ML,* 1.

[27] KANG, M. & TIAN, J. 2018. Machine learning: Data pre-processing. *Prognostics and health management of electronics: fundamentals, machine learning, and the internet of things*, 111-130.

[28] KARPATNE, A., EBERT-UPHOFF, I., RAVELA, S., BABAIE, H. A. & KUMAR, V. 2018. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering,* 31, 1544-1554.

[29] KUMAR, A. 2018. Global warming, climate change and greenhouse gas mitigation. *Biofuels: greenhouse gas mitigation and global warming: next generation biofuels and role of biotechnology.* Springer.

[30] LASHOF, D. A. & AHUJA, D. R. 1990. Relative contributions of greenhouse gas emissions to global warming. *Nature,* 344, 529-531.

[31] LEPPERØD, A. J. 2019. *Air quality prediction with machine learning.* NTNU.

[32] LI, W., DING, S., WANG, H., CHEN, Y. & YANG, S. 2020. Heterogeneous ensemble learning with feature engineering for default prediction in peer-to-peer lending in China. *World Wide Web,* 23, 23-45.

[33] LIESKE, D. J., SCHMID, M. S. & MAHONEY, M. 2018. Ensembles of ensembles: combining the predictions from multiple machine learning methods. *Machine Learning for Ecology and Sustainable Natural Resource Management.* Springer.

[34] LIPPI, M., BERTINI, M. & FRASCONI, P. 2013. Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *IEEE Transactions on Intelligent Transportation Systems,* 14, 871-882.

[35] LUENGO, J., GARCÍA-GIL, D., RAMÍREZ-GALLEGO, S., GARCÍA, S. & HERRERA, F. 2020. Big data preprocessing. *Cham: Springer,* 1, 1-186.

[36] MORIN, T., BOHRER, G., NAOR-AZRIELI, L., MESI, S., KENNY, W., MITSCH, W. & SCHÄFER, K. 2014. The seasonal and diurnal dynamics of methane flux at a created urban wetland. *Ecological Engineering,* 72, 74-83.

[37] NAGANNA, S. R., BEYAZTAS, B. H., BOKDE, N. & ARMANUOS, A. M. 2020. On the evaluation of the gradient tree boosting model for groundwater level forecasting. *Knowledge-Based Engineering and Sciences,* 1, 48-57.

[38] NAGHIBI, S. A., HASHEMI, H., BERNDTSSON, R. & LEE, S. 2020. Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors. *Journal of Hydrology,* 589, 125197.

[39] OBAID, H. S., DHEYAB, S. A. & SABRY, S. S. The impact of data pre-processing techniques and dimensionality reduction on the accuracy of machine learning. 2019 9th annual information technology, electromechanical engineering and microelectronics conference (iemecon), 2019. IEEE, 279-283.

[40] ODEDEYI, P. B., ABOU-EL-HOSSEIN, K., OYEKUNLE, F. & ADELEKE, A. K. 2020. Effects of machining parameters on Tool wear progression in End milling of AISI 316. *Progress in Canadian Mechanical Engineering,* 3.

[41] OGUNNOWO, E. O. A Conceptual Framework for Digital Twin Deployment in Real-Time Monitoring of Mechanical Systems.

[42] OGUNNOWO, E. O., ADEWOYIN, M. A., FIEMOTONGHA, J. E., IGUNMA, T. O. & ADELEKE, A. K. 2020. Systematic Review of Non-Destructive Testing Methods for Predictive Failure Analysis in Mechanical Systems.

[43] OKUH, C. O., NWULU, E. O., OGU, E., IFECHUKWUDE, P., EGBUMOKEI, I. N. D. & DIGITEMIE, W. N. Creating a Sustainability-Focused Digital Transformation

Model for Improved Environmental and Operational Outcomes in Energy Operations.

[44] OKUH, C. O., NWULU, E. O., OGU, E., IFECHUKWUDE, P., EGBUMOKEI, I. N. D. & DIGITEMIE, W. N. An Integrated Lean Six Sigma Model for Cost Optimization in Multinational Energy Operations.

[45] OLOF, S. S. 2018. A comparative study of black-box optimization algorithms for tuning of hyper-parameters in deep neural networks.

[46] PAWŁOWSKI, K. & KURACH, K. Detecting methane outbreaks from time series data with deep neural networks. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing: 15th International Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceedings, 2015. Springer, 475-484.

[47] PRAYOGO, D. & SUSANTO, Y. T. T. 2018. Optimizing the Prediction Accuracy of Friction Capacity of Driven Piles in Cohesive Soil Using a Novel Self-Tuning Least Squares Support Vector Machine. *Advances in Civil Engineering,* 2018, 6490169.

[48] SAHIN, E. K. 2020. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Sciences,* 2, 1308.

[49] SALEH, C., DZAKIYULLAH, N. R. & NUGROHO, J. B. Carbon dioxide emission prediction using support vector machine. IOP conference series: materials science and engineering, 2016. IOP Publishing, 012148.

[50] THOMPSON, D., LEIFER, I., BOVENSMANN, H., EASTWOOD, M., FLADELAND, M., FRANKENBERG, C., GERILOWSKI, K., GREEN, R., KRATWURST, S. & KRINGS, T. 2015. Real-time remote detection and measurement for airborne imaging spectroscopy: a case study with methane. *Atmospheric Measurement Techniques,* 8, 4383-4397.

[51] VINCENT, J., WANG, H., NIBOUCHE, O. & MAGUIRE, P. 2020. Detecting trace methane levels with plasma optical emission spectroscopy and supervised machine learning. *Plasma Sources Science and Technology,* 29, 085018.

[52] WHALEN, S. C. & REEBURGH, W. S. 1992. Interannual variations in tundra methane emission: A 4-year time series at fixed sites. *Global Biogeochemical Cycles,* 6, 139-159.

[53] WUEBBLES, D. J. & HAYHOE, K. 2002. Atmospheric methane and global change. *Earth-Science Reviews,* 57, 177-210.