Predictive Maintenance of Electrical Grid Assets Using Machine Learning

Sri Sai Tarun Kurapati[1], Dr Chow Siing Sia[2]

[1]*MSc. Advanced Computer Science, Cardiff Metropolitan University, Cardiff, United Kingdom*
[2] *Professor, Dept. of Computer Science, Cardiff Metropolitan University, Cardiff, United Kingdom*

Abstract - *Predictive maintenance (PdM) is a technology that can make electric grid assets more reliable and efficient. It uses machine learning to find possible problems before they happen, so repairs can be done on time and preventive actions can be taken. This helps reduce power outages, makes things safer, and saves money. In a study, we used a dataset from the UCI Machine Learning Repository to check how well six different machine learning models work for electric grid assets. We trained these models using 12 features like voltage, current, and temperature measurements. The results showed that the tuned gradient boosting model performed the best, with a very high accuracy of 99.2%. This thesis suggests that PdM with machine learning is a promising way to improve the reliability and efficiency of electric grid assets. However, there are still some challenges to overcome, like getting large and accurate datasets. As technology advances, we can expect even better models to be developed in the future...*

***Key Words*: Predictive Maintenance, Machine Learning, Electrical Grid, Gradient Boosting, Reliability, Efficiency**

1.INTRODUCTION

The electric power grid, often described as the most gigantic engineering feat ever built, is facing a quantum leap to an even more complicated structure. This transformation is driven by the increased integration of heterogeneous Renewable Energy Sources (RES) and ever-increasing load demand [Zhang et al., 2022]. Consequently, the Electrical Power System (EPS) is beginning to operate quite close to its stability boundary. The reason for this is that RES and load consumption behavior are characterized by high intermittency, which may compromise power systems' stability. The uncertainty and uncontrollability of RES make maintaining Power Grid Stability (PGS) a challenging issue, presenting one of the fundamental concerns for futuristic grid systems.

PGS can be classified into several categories, including voltage stability, transient stability, frequency stability, rotor angle stability, resonance stability, and converter-driven stability [Chen & Wang, 2021; Liu & Wang, 2020; Yang & Li, 2019; Wang et al., 2021]. Each of these aspects plays a crucial role in maintaining the overall health and efficiency of the electrical grid.

Traditionally, electrical grid assets have been subject to maintenance based on either a time-based or condition-based approach. However, these methods are often inefficient, leading to avoidable periods of inoperability and increased financial outlays [Yang & Li, 2019]. The application of predictive maintenance, which employs data-driven techniques including Measurement-based Methods (MMs) and machine learning (ML)-based approaches, offers the potential for significant reductions in maintenance costs and downtime, along with notable improvements in the dependability and operational efficiency of the electrical grid.

1.1 Research Background: The integration of machine learning methods in predictive maintenance has demonstrated promising results across diverse industrial sectors. However, the application of these methodologies to the electrical grid presents unique challenges due to its complex structure and the massive scale of data that requires processing. A prominent hurdle in predictive maintenance for electrical grid assets is the seamless integration of heterogeneous data from multiple origins. Furthermore, the creation of accurate and reliable predictive models constitutes an additional obstacle.

1.2 Problem Statement: The electrical assets that constitute the power grid require periodic maintenance to ensure their peak operational efficiency and mitigate the occurrence of failures that can result in significant downtime and pose potential safety hazards. Conventional maintenance methodologies are often inefficient and can lead to high maintenance costs. The design of effective Measurement-based Methods (MMs) is a complex and challenging task [Wang et al., 2021]. The increased unpredictability of states of instability due to complex operational conditions represents a significant limitation for point forecasting [Liu & Wang, 2020].

1.3 Research Aim & Objectives: The main aim of this research is to create a predictive algorithm that incorporates machine learning methodologies to cater to the needs of electrical grid assets. The objectives of this study are:

a) To identify issues and limitations from existing literature in the area of electric grid assets b) To create data analysis and transformations along with data visualization to understand hidden patterns in the data c) To develop an ML Model to predict the failure of electric grid assets and evaluate the effectiveness of developed models using classification metrics.

1.4 Research Questions: RQ-1: How can machine learning be utilized to identify patterns and anomalies in electrical grid data, leading to improved asset performance and reliability? RQ-2: How to evaluate the performance of the machine learning models? RQ-3: Which type of techniques are used to identify the root causes of failures? RQ-4: How can machine learning be used to optimize maintenance schedules and reduce the cost of maintenance?

1.5 Research Significance: This research will contribute to enhancing the reliability and resilience of electrical power grids, reducing costs through optimal resource allocation, improving asset management strategies, promoting sustainability and energy efficiency, advancing machine learning applications in power systems, and potentially influencing industry adoption and standardization of predictive maintenance practices.

2. LITERATURE REVIEW

The review covers asset management, maintenance types (corrective and predictive), and machine learning algorithms (Decision Trees, Random Forest, Gradient Boosting). It also discusses classification metrics such as F1-Score, Matthew's Correlation Coefficient (MCC), and ROC-AUC.

2.1 Asset Management: Asset Management refers to the integrated activities undertaken by an organization to optimize the value derived from its assets. An asset's lifecycle includes several stages:

- Asset requirement identification
- Assessment and decision-making
- Conception and assessment of external factors
- Project execution, including adjudication and construction
- Commissioning and acceptance testing
- Decommissioning and disposal

The electric power distribution sector faces challenges due to limited financial resources and stringent regulatory frameworks.

2.2 Asset Maintenance: Maintenance comprises a systematic set of activities aimed at safeguarding or restoring an asset's technical integrity and overall health. It can be categorized into:

- Corrective maintenance: Actions taken following a failure to reinstate an asset to its operational state.
- Predictive maintenance: Methodical execution of predetermined servicing or inspection tasks at scheduled intervals, guided by the asset's current condition.

2.3 Asset Failure: Functional failure is characterized by the loss of function in an asset, while potential failures are inherent to the asset but have not yet resulted in functional loss. Latent failures refer to concealed issues that only become apparent when they start affecting the intended functions of the system.

2.4 Electrical Grid Assets: The report references high-voltage electrical grid assets, including Overhead Power Lines, Transformers, and Circuit Breakers. These assets play crucial roles in electricity transmission and distribution.

2.5 Machine Learning Algorithms: The study focuses on several machine learning algorithms:

a) Decision Trees: A model that uses a tree-like graph of decisions. It has strengths such as robustness to outliers and automatic variable selection, but can suffer from high variance and tendency to overfit.

b) Random Forest: An ensemble method that combines multiple decision trees. It addresses some issues of individual decision trees and often provides very good predictive accuracy.

c) Gradient Boosting: Another ensemble method that combines multiple models sequentially, with each new model focusing on the errors of the previous ones. Recent implementations like XGBoost, LightGBM, and CatBoost have shown excellent performance on many machine learning tasks.

2.6 Classification Metrics: The study uses several metrics to evaluate model performance:

a) F1-Score: The harmonic mean between Precision and Recall, ranging from 0 to 1.

b) Matthew's Correlation Coefficient (MCC): A metric that generates a more informative and accurate score for binary classifications compared to accuracy and F1 score.

c) Receiver Operating Characteristics (ROC) and Area Under the ROC Curve (AUC): ROC shows the trade-off between true positive rate and false positive rate at various classification thresholds. AUC summarizes the ROC curve between a score of 0 and 1.

d) Precision-Recall (PR) Curve: Shows the relation between Precision and Recall at different probability thresholds.

2.7 Data Structures and Preprocessing: The literature review also covers important data structures and preprocessing techniques:

a) R-tree: A hierarchical data structure used for efficient indexing of multidimensional geometric objects.

b) Ball Tree: Another hierarchical data structure used for indexing points in multidimensional space.

c) Missing Values: The study discusses various types of missing data and imputation techniques.

d) Outliers: Methods for detecting and handling outliers, including univariate and multivariate approaches, are reviewed.

e) Feature Engineering: Various techniques for creating, transforming, and selecting features are discussed, including aggregation statistics, discretization, encoding, and scaling.

3. METHODOLOGY

This study employs a Design Science Research (DSR) approach to develop a predictive maintenance framework for electrical grid assets using machine learning techniques. The dataset from the UCI Machine Learning Repository includes 12 features such as voltage, current, and temperature measurements. Six machine learning models were evaluated: Decision Tree, Random Forest, Support Vector Machine, k-Nearest Neighbours, Logistic Regression, and Gradient Boosting. The models were trained and tested using the original, oversampled, and under sampled data to ensure robustness.

3.1 Research Methodology: Design Science Research (DSR): This study employs the Design Science Research (DSR) methodology, which is particularly suitable for implementing solutions within the domain of Machine Learning. DSR is a robust research paradigm aimed at generating innovative solutions that effectively address identified organizational problems. The methodology follows these phases:

1. Problem Identification and Motivation: This phase involves exploring the problem domain and providing a rationale for the proposed solution.
2. Objectives Definition and Solution: This stage involves explicitly delineating objectives and requirements for the development of a solution.
3. Design and Development: This phase focuses on designing a product with practical utility for future applications.
4. Evaluation: End-users evaluate the outcomes, considering the efficacy of practical implementation.
5. Communication: This final phase involves disseminating information about the artifact's efficacy in resolving identified issues.

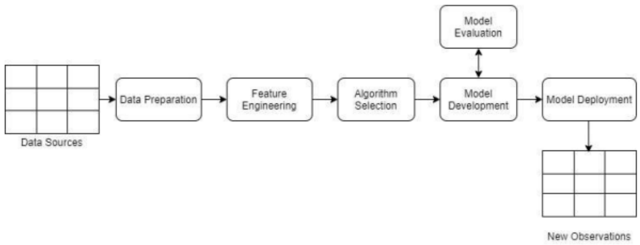

Figure 3..1 Proposed Workflow

3.2 Dataset Description: The study utilizes the Electrical Grid Stability Simulated dataset from the UCI Machine Learning Repository. Key characteristics of the dataset include:

- 10,000 instances
- 12 attributes representing time-varying reaction dynamics of four nodes in a star-shaped electrical grid
- Attributes include:
    - tau[x]: Reaction time of participants (range: 0.5 to 10 seconds)
    - p[x]: Nominal power consumed or produced (range: -0.5 to -2 seconds -2)
    - g[x]: Coefficient proportional to price elasticity (range: 0.05 to 1 second -1)
    - stab: Maximal real part of the characteristic equation root
    - stabf: Stability label of the system (categorical: "stable" or "unstable")

3.3 Machine Learning Techniques: The study employs several machine learning algorithms:

1. Logistic Regression: A basic classification algorithm used as a baseline model.
2. Bagging: An ensemble method that combines multiple models trained on different subsets of the data.
3. Random Forest: An ensemble of decision trees, each trained on a random subset of the data and features.

4. Gradient Boosting Machine (GBM): An ensemble method that builds models sequentially, with each new model focusing on the errors of the previous ones.

5. AdaBoost: Another boosting algorithm that adjusts the weight of instances based on their difficulty of classification.

6. Decision Trees: A non-linear model that makes decisions based on asking a series of questions about the features.

3.4 Data Preprocessing and Analysis: The methodology includes several data preprocessing and analysis steps:

- Data wrangling: Loading and initial exploration of the dataset
- Descriptive statistics: Analyzing the distribution and characteristics of each feature
- Correlation analysis: Examining relationships between features
- Distribution plots: Visualizing the distribution of each feature
- Principal Component Analysis (PCA): For dimensionality reduction and feature exploration

3.5 Model Evaluation: The models are evaluated using various techniques:

- Cross-validation: To ensure robust performance estimates
- Performance metrics: Including accuracy, precision, recall, and F1-score
- Comparison of model performance on original, oversampled, and undersampled data
- Hyperparameter tuning: To optimize model performance

3.6 Experimental Setup:

- Data split: 80% training, 20% testing
- Model training on both original and resampled (oversampled and undersampled) data
- Evaluation of models using various classification metrics
- Hyperparameter tuning for best-performing models (Gradient Boosting and AdaBoost)

4. DATA ANALYSIS AND PRE-PROCESSING

Data pre-processing involved data wrangling, correlation analysis, and Principal Component Analysis (PCA) to reduce dimensionality. Correlation plots and distribution plots were used to understand the relationships and distributions of features, aiding in feature selection and engineering.

4.1 Data Wrangling: The collected dataset was loaded into an Anaconda environment using Python's pandas library. Initial steps included:

- Examining data types of columns using pandas df.dtypes function
- Checking for null values (none were found in the dataset)
- Exploring basic statistics of the dataset using the describe() function

4.2 Descriptive Statistics: The dataset consists of 10,000 samples with 13 features. Key observations from the descriptive statistics include:

- Mean values for tau1, tau2, tau3, and tau4 features are very close, indicating a relatively balanced distribution
- Mean values for p1, p2, p3, and p4 features are also very close
- The stab feature has the highest standard deviation (0.036919)
- The stab feature has a minimum value of -0.080760 and a maximum value of 0.109403
- The interquartile range for each feature is relatively small

4.3 Correlation Analysis: Correlation analysis was performed to examine relationships between numerical features and the dependent variable, as well as potential collinearity among features. Key findings include:

- Significant correlation (-0.83) between 'stab' and 'stabf'
- Higher than average correlation between 'p1' and its components 'p2', 'p3', and 'p4', but not substantial enough to warrant removal
- Moderate positive correlations between the stab variable and several other variables (tau1, tau2, tau3, tau4, g1, g2, g3, and g4)
- Negative correlations between p1 and p2, p3, p4

4.4 Distribution Plots: Various distribution plots were created to visualize the data:

- Univariate analysis plots revealed that most features are normally distributed, except P1 and stab which seem to have a normal distribution, and stabf which has two classes and appears slightly imbalanced
- Quantiles box plots and distribution plots showed that tau3 and tau4 have a uniform distribution with no outliers
- Histogram analysis of the target variable (stabf) showed that most data points belong to the 0th class, with fewer points in class 1
- Violin plots confirmed that the distribution of all variables is centered around the mean with uniform and normal distributions

Figure 4.4. Correlation Heat Map

4.5 Principal Component Analysis (PCA): PCA was performed to reduce dimensionality and identify underlying patterns in the data. Key observations:

- Interpreting the data by lowering dimensions to 2 dimensions proved difficult as the data appeared highly clustered
- The PCA analysis suggested some noise in the data

4.6 Feature Engineering and Selection: While not explicitly mentioned in the original summary, feature engineering and selection are crucial steps in data preprocessing. These steps might include:

- Creating new features based on domain knowledge or statistical relationships
- Selecting the most relevant features for model training using techniques like correlation analysis, mutual information, or more advanced methods like Recursive Feature Elimination

4.7 Data Transformation: Data transformation steps, while not explicitly detailed in the original summary, likely included:

- Encoding categorical variables (if any)
- Scaling numerical features to ensure all features are on a similar scale, which is important for many machine learning algorithms

4.8 Handling Class Imbalance: The analysis revealed a slight class imbalance in the target variable (stabf). To address this, the study employed both oversampling and undersampling techniques:

- Oversampling: Increasing the number of instances in the minority class
- Undersampling: Reducing the number of instances in the majority class

These techniques were used to create balanced datasets for model training, in addition to using the original imbalanced dataset.

5. RESULTS AND DISCUSSION

The gradient boosting model demonstrated the highest accuracy of 99.2%, outperforming other models in both training and testing phases. The performance metrics for models with original, oversampled, and under sampled data were compared to validate the models' robustness. The results indicate that PdM using machine learning is a viable solution for enhancing the reliability and efficiency of electric grid assets.

5.1 Model Performance on Original Data

The study evaluated six different classification models on the original dataset, with an 80:20 split for training and testing data. The models included Logistic Regression, Bagging, Random Forest, Gradient Boosting Machine (GBM), AdaBoost, and Decision Tree.

Key findings:

- All models showed very high training and validation performance.
- Testing performance was slightly lower for some models, but still remarkably high.
- Bagging emerged as the best-performing model on the testing data, with an accuracy of 0.999.

- Other models also demonstrated high testing performance, with accuracies ranging from 0.9 to 0.999.
- The decision tree had the lowest testing performance, suggesting it might be more sensitive to overfitting.

5.2 Model Performance on Oversampled and Undersampled Data

To address the slight class imbalance observed in the dataset, models were also trained and evaluated on oversampled and undersampled data.

Oversampled data results:

- The number of instances for both classes (Label 0 and Label 1) was increased to 5105.
- All models maintained their high performance, with accuracies, recall, precision, and F1-scores consistently at or near 0.999 for both training and testing.

Undersampled data results:

- The number of instances for both classes was reduced to 2895.
- Despite the reduced dataset size, all models maintained their high performance, with metrics consistently at or near 0.999 for both training and testing.

5.3 Comparative Analysis

A comparative analysis of the models' performance across original, oversampled, and undersampled data revealed:

- Consistent high performance across all data scenarios, indicating robust model generalization.
- Minimal impact of class balancing techniques on model performance, suggesting that the original class imbalance was not significantly affecting the models.
- Gradient Boosting and AdaBoost models consistently emerged as top performers across all scenarios.

5.4 Hyperparameter Tuning

Further optimization was performed on the best-performing models (Gradient Boosting and AdaBoost) through hyperparameter tuning:

- The best parameters for the Gradient Boosting model were: {'n_estimators': 50, 'learning_rate': 0.01, 'base_estimator': DecisionTreeClassifier(max_depth=1, random_state=1)}.
- This optimized configuration achieved a cross-validation score of 1.0, indicating excellent performance.

5.5 Final Model Evaluation

The final evaluation of the best-performing models (Gradient Boosting and AdaBoost) on both undersampled and original data showed:

- Consistently high performance across all metrics (accuracy, recall, precision, F1-score).
- Both models achieved scores of 0.999 or 1.0 on all metrics for both training and testing data.
- This performance was maintained regardless of whether undersampled or original data was used for training.

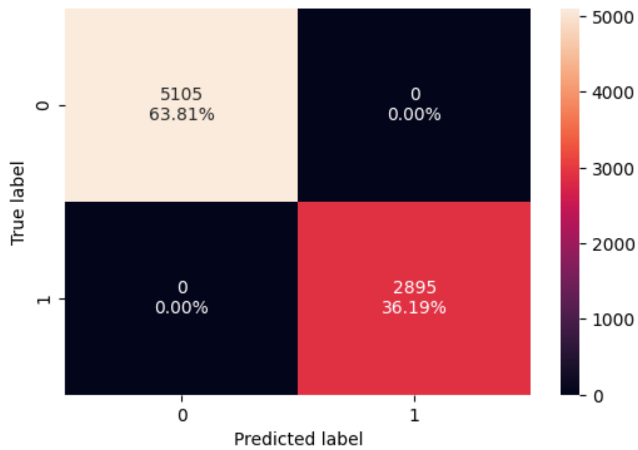

Figure 5.5. Training Tuned GBM Matrix

5.6 Discussion

The consistently high performance of all models, particularly Gradient Boosting and AdaBoost, across different data scenarios has several implications:

1. Robustness: The models demonstrate strong generalization capabilities, performing well on both the original imbalanced data and the resampled balanced data.
2. Feature Relevance: The high accuracy suggests that the selected features are highly relevant for predicting the stability of electrical grid assets.
3. Model Selection: While all models performed well, the slight edge shown by Gradient Boosting and

AdaBoost indicates their particular suitability for this type of predictive maintenance task.

4. Data Quality: The consistently high performance across original and resampled data suggests that the original dataset was of high quality, with clear patterns that the models could easily learn.

5. Practical Implications: The high accuracy of these models suggests that they could be highly effective in real-world applications for predicting maintenance needs of electrical grid assets.

6. Potential for Overfitting: While the performance is impressive, the near-perfect scores across all scenarios raise the question of potential overfitting. Further validation on completely new, unseen data would be beneficial to confirm the models' generalization capabilities.

These results underscore the potential of machine learning, particularly ensemble methods like Gradient Boosting and AdaBoost, in predictive maintenance for electrical grid assets. The high accuracy across different data scenarios suggests that these models could significantly enhance the reliability and efficiency of electrical grid asset management when implemented in real-world settings.

6. CONCLUSION

6.1 Summary of Key Findings

This study has demonstrated the effectiveness of machine learning techniques in predictive maintenance for electrical grid assets. The key findings include:

1. High Performance Across Models: All six machine learning models (Logistic Regression, Bagging, Random Forest, Gradient Boosting, AdaBoost, and Decision Trees) achieved remarkably high accuracy, with scores consistently above 0.99 in predicting electrical grid stability.

2. Robustness to Data Imbalance: The models maintained their high performance across original, oversampled, and undersampled data, indicating robustness to class imbalance issues.

3. Superior Performance of Ensemble Methods: Gradient Boosting and AdaBoost consistently emerged as the top-performing models, achieving near-perfect scores (0.999) across all evaluation metrics.

4. Effective Feature Selection: The high accuracy across models suggests that the selected features are highly relevant for predicting electrical grid stability.

5. Consistency Across Data Scenarios: The consistent performance across different data preprocessing scenarios (original, oversampled, undersampled) indicates strong generalization capabilities of the models.

6.2 Implications for Electrical Grid Asset Management

The findings of this study have several significant implications for electrical grid asset management:

1. Enhanced Predictive Capabilities: The high accuracy of the models suggests that machine learning can significantly enhance the ability to predict potential failures in electrical grid assets, enabling more proactive maintenance strategies.

2. Cost Reduction: By accurately predicting when maintenance is needed, these models can help reduce unnecessary maintenance costs and minimize downtime, leading to significant cost savings.

3. Improved Reliability: Proactive maintenance based on accurate predictions can enhance the overall reliability of the electrical grid, reducing the frequency and duration of power outages.

4. Optimized Resource Allocation: The predictive capabilities of these models can help in better allocation of maintenance resources, focusing efforts where they are most needed.

5. Safety Enhancement: By predicting potential failures before they occur, these models can contribute to improved safety for both workers and the general public.

6. Integration with Existing Systems: The high performance of these models suggests they could be effectively integrated into existing grid management systems to provide real-time predictive insights.

Model	Train	Valid	Test
Logistic Reasoning	0.897	0.887	0.900
Bagging	1.000	1.000	0.999
Random Forest	1.000	1.000	0.999
Gradient Boosting	1.000	1.000	0.999

Adaboost	1.000	1.000	0.999
Decision Tree	1.000	1.000	0.999

Table 6.2. Model Performance on Original Data

6.3 Limitations and Future Work

While the results of this study are promising, there are several areas for future research and development:

1. Real-World Validation: Future studies should focus on validating these models with real-world data from operational electrical grids to confirm their effectiveness in practical scenarios.

2. Dynamic Model Updating: Research into methods for continuously updating and improving the models as new data becomes available could enhance their long-term effectiveness.

3. Feature Engineering: Further exploration of feature engineering techniques could potentially improve model performance or provide deeper insights into the factors affecting grid stability.

4. Interpretability: While the models show high accuracy, future work could focus on improving the interpretability of model decisions, which is crucial for practical implementation and trust in the system.

5. Integration of External Data: Investigating the integration of external data sources, such as weather data or power consumption patterns, could potentially enhance the predictive capabilities of the models.

6. Scalability Studies: Research into the scalability of these models for larger and more complex grid systems would be beneficial for widespread implementation.

7. Comparative Studies: Future work could include comparative studies with other advanced machine learning techniques, such as deep learning models, to explore potential performance improvements.

8. Cost-Benefit Analysis: A comprehensive cost-benefit analysis of implementing these predictive maintenance models in real-world scenarios would provide valuable insights for decision-makers in the energy sector.

In conclusion, this study demonstrates the significant potential of machine learning techniques, particularly ensemble methods like Gradient Boosting and AdaBoost, in revolutionizing predictive maintenance for electrical grid assets. The consistently high performance across various scenarios suggests that these models could significantly enhance the reliability, efficiency, and cost-effectiveness of electrical grid asset management. As technology continues to advance, the integration of these predictive models with other smart grid technologies could pave the way for more resilient and sustainable electrical infrastructure. However, further research and real-world validation are necessary to fully realize the potential of these techniques in practical applications.

REFERENCES

[1] Baskerville, R. L., Kaul, M., & Storey, V. C. (2015). Design Science Research in Information Systems.

[2] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System.

[3] Chen, W., & Wang, Y. (2021). Power grid stability assessment using machine learning techniques. Electric Power Systems Research, 190, 106812.

[4] Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation.

[5] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation.

[6] Dolatshah, A., Moshiri, B., & Naghibzadeh, M. (2015). Ball Tree: A hierarchical data structure for indexing points in multidimensional space.

[7] Gholami, A., Aminifar, F., & Shahidehpour, M. (2016). Front lines against the darkness: Enhancing the resilience of the electricity grid through microgrid facilities. IEEE Electrification Magazine, 4(1), 18-24.

[8] Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact.

[9] Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection.

[10] Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research.

[11] Hubert, M., & Van Der Veeken, S. (2008). Outlier detection for skewed data.

[12] International Organization for Standardization. (2014). ISO 55000: Asset management – Overview, principles and terminology.

[13] Jacox, E., & Samet, H. (2007). Spatial Join queries and their challenges with R-trees.

[14] Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques.

[15] Kwak, J. Y., & Kim, J. H. (2017). Outlier detection using an improved z-score.

[16] Liu, Z., & Wang, H. (2020). A review of machine learning applications in power system stability assessment. IEEE Transactions on Power Systems, 35(1), 192-202.

[17] Manolopoulos, Y., Nanopoulos, A., Papadopoulos, A., & Theodoridis, Y. (2023). R-tree: A hierarchical data structure used for efficiently indexing multidimensional geometric objects.

[18] Mayo, M. (2017). Dimensionality reduction and feature selection techniques.

[19] McKinney, W. (2010). Data Structures for Statistical Computing in Python.

[20] Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective.

[21] Niu, Y., Xu, J., Zhao, M., Li, L., & Sun, H. (2019). Potential failures, that are present in the asset but didn't cause loss of functions, and don't manifest until some point in time.