Identification of Key Parameters Influencing Load Progression during Startup in Combined Cycle Power Plants (CCPP): Implications for Smart Grids

Mogana Vadhna Suntrakumar
Faculty of Information and Communication Technology (FTMK)
Universiti Teknikal Malaysia Melaka (UTeM)
Melaka, Malaysia
moganavadhna@gmail.com

Sharifah Sakinah Binti Syed Ahmad
Faculty of Information and Communication Technology (FTMK)
Universiti Teknikal Malaysia Melaka (UTeM)
Melaka, Malaysia
sakinah@utem.edu.my

Noor Fazillah Abd Yusof
Faculty of Information and Communication Technology (FTMK)
Universiti Teknikal Malaysia Melaka (UTeM)
Melaka, Malaysia
elle@utem.edu.my

Mohamad Lutfi Samsudin
Department of Engineering
Malakoff Power Berhad
Kuala Lumpur, Malaysia
ridzhuan.othman@malakoff.commy

2024 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia) | 979-8-3503-5229-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICPSASIA61913.2024.10761422

***Abstract*—This paper investigates integrating advanced analytics techniques into smart grids to enhance energy system efficiency and reliability. It focuses on identifying key operational parameters influencing load progression during startup processes in combined cycle power plants (CCPPs). Using a dataset categorized into different startup classes (Hot, Warm 1, Warm 2, Cold), it employs feature selection techniques such as PCC, LASSO regularization, and RF-RFE to identify influential parameters. Predictive modeling evaluates each technique's efficiency in capturing system dynamics, revealing variations in predictive performance across startup classes. The study highlights crucial variables contributing significantly to load progression and discusses implications for smart grid applications like predictive maintenance and load prediction. It advances understanding of CCPP operation within smart grids, providing insights for enhancing energy system intelligence and resilience.**

***Keywords*—Smart Grid, CCPP, energy, power, predictive modeling, feature selection, key operational parameters.**

I. INTRODUCTION

Traditional mechanical grids have long been the backbone of global energy distribution systems, facilitating the reliable supply of electricity to consumers. However, these grids face significant challenges, including limited flexibility, vulnerability to disruptions, and inefficiencies in energy utilization [1, 2]. As the demand for electricity continues to rise and the landscape of energy production evolves, the need for more adaptable, resilient, and efficient energy infrastructure becomes increasingly apparent.

Enter the smart grid—a transformative approach to energy distribution that leverages advanced technologies, big data analytics, and real-time communication to revolutionize how electricity is generated, transmitted, and consumed. At its core, a smart grid represents a paradigm shift from a passive, one-way energy flow to an interconnected [3], dynamic network capable of real-time monitoring, control, and optimization. By integrating digital communication and sensing capabilities into the traditional grid infrastructure, smart grids empower utilities, grid operators, and consumers to make data-driven decisions and respond proactively to changing energy demands and environmental conditions.

The significance of smart grids lies in their ability to address many of the shortcomings of conventional grids while unlocking a myriad of advantages. Foremost among these advantages is enhanced reliability and resilience [4], achieved through the implementation of self-healing capabilities and automated fault detection and isolation mechanisms. Additionally, smart grids offer greater flexibility and efficiency in energy management, enabling optimal utilization of renewable energy sources, demand response programs, and energy storage technologies.

In this context, big data analytics emerges as a cornerstone of smart grid intelligence, enabling the processing, analysis, and interpretation of vast amounts of data generated by sensors, meters, and other grid devices [5]. By harnessing the power of big data, smart grids can intelligently predict and respond to changes in energy demand, optimize grid operations, and mitigate potential disruption. This is particularly significant as big data analytics enables the collection, processing, and analysis of massive quantities of data from heterogeneous modules within the grid. Through enhanced energy regulation, predictive maintenance, and integration of renewable energy sources, big data analytics enhances the efficiency, reliability, and resilience of smart grid systems, ultimately contributing to the advancement of modern electrical power infrastructure [6].

In parallel, the identification and understanding of key operational parameters within energy systems, such as those in combined cycle power plants (CCPPs), play a vital role in realizing the full potential of smart grid technologies. With the increasing complexity of energy networks and the integration of renewable energy sources, the need for precise parameter identification becomes even more pronounced.

A combined cycle power plant (CCPP) integrates heat engines to efficiently convert heat into mechanical energy, yielding benefits such as increased efficiency and reduced fuel costs [7]. Recognized for their quick start capability and lower specific emissions, CCPPs are pivotal in meeting the rising electricity demand while minimizing environmental impact [8]. Their frequent start-up and shut-down cycles, coupled with the need for intricate control logic to manage seasonal demand variations, underscore the complexity of CCPP operation [9].

979-8-3503-5229-0/24/$31.00 ©2024 IEEE

Amidst these operational intricacies, the advent of big data and advanced analytics offers opportunities to optimize CCPP performance . By leveraging vast operational data and employing artificial intelligence (AI)-driven analytics, insights into plant dynamics and opportunities for efficiency enhancements are unlocked . However, selecting relevant features from the wealth of data remains a challenge . To address this challenge, our study primarily focuses on the identification of key operational parameters influencing load progression during start-up processes in CCPPs. While the effectiveness of three feature selection techniques (filtering using Pearson Correlation Coefficient (PCC), wrapper-based selection with Random Forest-based Recursive Feature Elimination (RF-RFE), and embedded selection through LASSO regularization) is compared, it's crucial to emphasize that our primary objective is to identify these key parameters. This emphasis stems from the recognition that dealing with hundreds of parameters would be computationally expensive [10] and impractical for real-world implementation. By focusing on identifying these key parameters, our research contributes significantly to advancing feature selection methodologies in CCPP optimization, thereby shaping future advancements in this critical domain and providing valuable insights to optimize smart grid operations further.

II. Methodology

A. Data Understanding

The load progression data utilized in this study was collected from a Combined Cycle Power Plant (CCPP) located in Malaysia. The dataset contains records of load progression during startup processes, recorded at regular intervals. It includes data categorized by startup class (HOT, WARM1, WARM2, COLD) and covers the period from 2017 to 2021. A detailed breakdown of the dataset is presented in Table 1, illustrating the distribution of data across different startup classes and years. Comprising 240 Excel files, each containing 14,401 rows and 92 columns, the dataset offers a wide range of information. The variables (columns) are grouped into categories such as load (Generator Watts), turbine speed, control and feedback, pressure, temperature, steam flow, loading rate, and steam conductivity, collectively representing factors influencing load progression. Table 2 provides an overview of the number of variables within each category. Data types in the dataset include integers, floats, datetime values, and Boolean indicators, reflecting the diverse parameters and attributes captured. This detailed overview of the dataset forms the basis for subsequent analysis and exploration.

TABLE I. DISTRIBUTION OF DATA ACROSS DIFFERENT STARTUP CLASSES AND YEARS.

Year	2017	2018	2019	2020	2021	TOTAL
HOT	7	27	25	4	0	**63**
WARM1	4	13	11	8	2	**38**
WARM2	7	23	20	14	8	**72**
COLD	2	11	8	28	18	**67**
TOTAL	**20**	**74**	**64**	**54**	**28**	**240**

TABLE II. NUMBER OF VARIABLES PRESENT IN EACH CATEGORY.

Variable Category	Count of Variables
GENERATOR WATTS (LOAD)	3
TURBINE SPEED	2
CONTROL AND FEEDBACK	39
PRESSURE	8
TEMPERATURE	28
STEAM FLOW	2
LOADING RATE	3
STEAM CONDUCTIVITY	3

In addition to the load progression data, understanding the operational aspects of the power plant is crucial for contextualizing the dataset. The power plant under study features key elements including a cutting-edge GE 9FA+e gas turbine, a GE D10 steam turbine, and a GE 390H hydrogen-cooled generator, integrated into a single-shaft configuration for optimal efficiency and sustainability, with a capacity of 350MW. Furthermore, the startup sequences at the power plant are meticulously orchestrated steps designed to initiate power generation reliably. These sequences include initiating the gas turbine, purging potentially hazardous gases, firing the combustion chamber, warming up the turbine, reaching full speed without load, synchronizing with the electrical grid, activating temperature control, coupling with the steam turbine if applicable, enabling inlet pressure control, enabling back pressure systems, deactivating temperature matching, marking sequence completion, gradually loading to minimum capacity, and loading to designated baseload or declared available capacity. The datasets used in this study encompass parameters recorded at each of these stages, providing comprehensive insights into the operational dynamics of the power plant.

B. Data Preprocessing

The data preprocessing process begins with extracting relevant data from Excel files obtained from the CCPP in Malaysia. Unnecessary rows are eliminated, and columns are renamed for clarity and consistency. Missing values are handled using appropriate imputation techniques such as filling with zero values or linear interpolation. Outliers are flagged using Z-Score and retained for valuable insights. Feature engineering includes adding new columns like "Reference" and "Class" and removing unnecessary ones, resulting in 67 variables. Individual Data Frames are merged into a unified CSV file named 'COMBINED_DATA.csv', containing 3,312,227 rows, ready for analysis.

C. Data Selection for Analysis

The data selection process was crucial for refining the analysis to concentrate on the most relevant startup procedures. By carefully choosing the top 10 startup processes with the shortest total startup time (LST) and the top 10 with the longest total startup time (LGT) from each startup class (HOT, WARM1, WARM2, COLD), a total of 80 startup process files were meticulously curated. This methodological approach played a pivotal role in identifying startup durations across different classes, providing insight into load progression dynamics. Highlighting the 10 files with the shortest total startup times emphasizes the efficiency of specific startup procedures, while acknowledging the 10 files with the longest total startup times illuminates potential complexities within particular classes. Subsequently consolidating the selected startup processes into two datasets, COMBINED_LST.csv and COMBINED_LGT.csv, further refined the analysis, allowing for a focused examination of influential variables impacting load progression. By separately analysing the two datasets for each class, we can explore the variables (parameters) that significantly influence both favourable and less favourable scenarios, thereby offering valuable insights for predictive modelling of load progression.

D. Feature Selection

In this section, three advanced feature selection methods are employed: PCC filtering, RF-RFE wrapper-based selection, and LASSO regularization. These techniques refine LSTM model development by identifying influential features to enhance predictive accuracy. The goal is to forecast load progression (Generator Watts) during startup processes. Each technique's application will be discussed, highlighting their effectiveness in identifying key features and providing insights into factors influencing load progression in CCPP power plants.

1) PCC

The Pearson Correlation Coefficient (PCC) formula (1) provides a statistical measure of the linear relationship between two continuous variables [11]. It gauges how much two variables change together, relative to their means [12]. A positive coefficient indicates a positive linear relationship, while a negative coefficient signifies an inverse relationship. The magnitude of the coefficient indicates the strength of the relationship, with values closer to +1 or -1 representing stronger correlations [13]. In feature selection, highly correlated features may introduce redundancy, impacting model accuracy. In this project, the selection criteria for using PCC involved assessing the correlation of each variable with the target variable, Generator Watts. A threshold of 0.7 was set to select the variables, indicating that only variables with a PCC magnitude greater than or equal to +0.7/-0.7 were considered relevant for inclusion in the analysis. This threshold value ensured that only highly correlated features, demonstrating a strong linear relationship with the target variable, were retained for further investigation.

$$\rho(x, y) = \frac{cov(x,y)}{\sigma(x) \times \sigma(y)} \tag{1}$$

2) RF-RFE.

RF-RFE, or Random Forest based-Recursive Feature Elimination, is a wrapper-based feature selection method widely used in machine learning to pinpoint the most crucial features for model performance. Unlike filter methods such as PCC, which evaluate individual feature correlations with the target variable independently, RF-RFE assesses feature importance collectively by iteratively removing the least significant ones from the dataset. Initially, the model is trained on the entire feature set, ranking features based on their importance scores relative to the target variable, Generator Watts. Subsequently, the least important features are pruned, and the model is retrained on the reduced set [15]. This process repeats until the desired number of features is reached or the model's performance stabilizes. RF-RFE is valuable for high-dimensional datasets, where feature count surpasses sample count, as it identifies a subset that optimizes predictive performance while minimizing computational complexity [16]. Leveraging Random Forest Regressor, a Random Forest variant known for its ensemble learning capabilities, RFE iteratively ranks and eliminates less impactful features, ensuring retention of the most relevant ones for predicting load progression during startup.

3) LASSO

LASSO, or Least Absolute Shrinkage and Selection Operator, proposed by Tibshirani, is a widely used feature selection technique in machine learning and regression analysis. It introduces regularization into the model fitting process to prevent overfitting and promote sparsity in the resulting feature set. In LASSO, an objective function includes a penalty term that penalizes the absolute magnitude of feature coefficients, in addition to the standard loss function. This penalty encourages smaller coefficient values and can drive some coefficients to exactly zero, effectively removing corresponding features from the model. When features exhibit high correlation, LASSO selects only one from the group and shrinks the rest to zero. The formula for the LASSO objective function is presented by formula (2). Here n is the number of observations, p is the number of features, yi is the target variable for observation i, xij is the value of feature j for observation i, $\beta 0$ is the intercept term, βj is the coefficient for feature j, and λ is the regularization parameter. The regularization parameter (λ) controls the strength of the regularization penalty.

$$\min\beta \left\{ \frac{1}{2n} \sum_{i=1}^{n} (yi - \beta 0 - \sum_{j=1}^{p} \beta j xij)^2 + \lambda \sum_{j=1}^{p} \mid \beta j \mid \right\} \tag{2}$$

- **Feature Selection Criteria**

The process of identifying influential features involved applying all three feature selection methods—PCC, RF-RFE, and LASSO—individually to both the COMBINED_LST and COMBINED_LGT datasets for each startup class (HOT, WARM 1, WARM 2, COLD). By conducting separate analyses for each class, distinct sets of influential features were obtained, indicating that each class has a unique selection of influential variables for all techniques. The most influential variables were identified as those consistently selected across both LST and LGT scenarios, indicating their significant influence under various startup conditions.

E. Prediction Model

The prediction model utilized in this study employs Long Short-Term Memory (LSTM) neural networks, a type of recurrent neural network (RNN) renowned for its capacity to learn and retain long sequences of data. Specifically, for each of the four startup classes (HOT, WARM1, WARM2, COLD), a distinct LSTM model is trained using sets of influential variables selected through the PCC, RFE, and LASSO feature selection techniques. This process yields a total of 12 models, each tailored to the characteristics of its respective startup class. These models are trained and evaluated individually on sequences of influential variables unique to each startup class, enabling a thorough examination of feature selection's impact on predictive accuracy across various operational scenarios in the CCPP power plant. Each LSTM model is composed of three layers, each comprising 50 units, with dropout regularization layers incorporated to mitigate overfitting. Performance evaluation metrics such as root mean squared error (RMSE) is computed to gauge the effectiveness of each model, providing valuable insights into load progression during startup operations in the CCPP power plant.

F. Evaluation Metrics

- **Root Mean Squared Error (RMSE)**

Root Mean Squared Error (RMSE) is the square root of the average squared differences between predicted and actual values. It offers an interpretable measure of prediction error in the same units as the target variable, providing insight into the magnitude of errors made by the model in predicting the target variable. Formula (3) presents the calculation for RMSE. In the equation, y_i represents the actual value, $\hat{y}_i$ denotes the predicted value, and n is the total number of samples.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (3)$$

III. RESULT AND DISCUSSION

This section presents the outcomes of applying various feature selection techniques to predict load progression across different startup classes (HOT, WARM1, WARM2, COLD) in Combined Cycle Power Plants (CCPPs), with a particular emphasis on selected feature categories. The selected features encompass a diverse range of variables, including pressure, temperature, turbine speed, control and feedback parameters, and steam flow. Rather than solely assessing the efficiency of the feature selection techniques, this study evaluates the performance of each technique based on its ability to predict load progression dynamics during startup operations in CCPPs using the selected features. Table III displays the critical operational parameter categories identified by each feature selection technique, predominantly across all startup classes. Each technique, namely PCC, RF-RFE, and LASSO, is assessed using the Root Mean Squared Error (RMSE) metric. Detailed results are visualized in Figure 1, providing a comprehensive overview of the RMSE performance of each technique across all startup classes. Subsequent sections will analyze the results for each class, shedding light on the effectiveness of the selected features in predicting load progression during startup operations in CCPPs.

TABLE III. IDENTIFIED CRITICAL OPERATIONAL PARAMETER OF CCPP

PCC	RF-RFE	LASSO
Pressure	Turbine Speed	Turbine Speed
Steam Flow	Temperature	Control and Feedback
Control and Feedback	Pressure	Temperature
Temperature	Steam Flow	Pressure
	Control and Feedback	Steam Flow

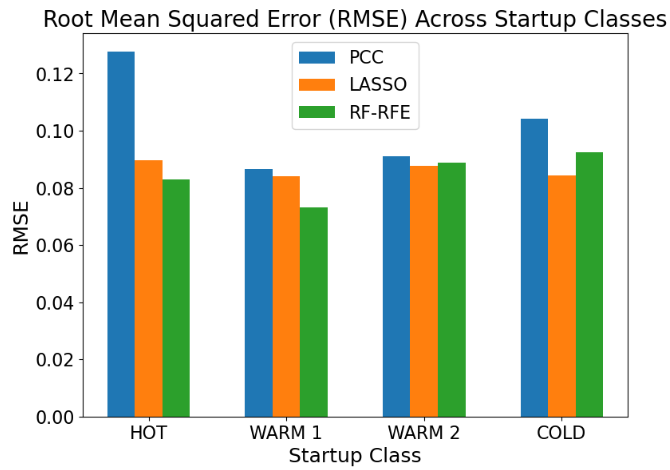

Fig. 1. RMSE value comparison of different feature selection techniques for all the startup classes.

- **Hot Start-up Class**

PCC selected pressure, steam flow, control & feedback, and temperature as important variables. RF-RFE and LASSO also identified these variables but additionally included turbine speed in their feature set. This extension led to improved predictive performance, as indicated by lower RMSE values compared to PCC.

- **Warm 1 Start-up class**

PCC initially selected pressure, steam flow, control & feedback, and temperature. RF-RFE and LASSO, like in the HOT class, selected the same set of variables and augmented them with turbine speed. Consequently, RF-RFE and LASSO achieved lower RMSE values compared to PCC, indicating enhanced predictive capabilities.

- **Warm 2 Start-up class**

Similar to the HOT and WARM 1 classes, PCC selected pressure, steam flow, control & feedback, and temperature. RF-RFE and LASSO also identified these variables and included turbine speed, resulting in an improved predictive performance with lower RMSE values compared to PCC.

- **Cold Start-up class**

In the COLD class, PCC selected pressure, steam flow, and temperature as significant variables. RF-RFE and LASSO, like in the other classes, selected the same set of variables and included turbine speed. This led to improved predictive accuracy, as evidenced by lower RMSE values compared to PCC.

A. Identification of Critical Operational Parameter Categories for Load Progression Prediction

In the analysis of load progression prediction during startup operations in Combined Cycle Power Plants (CCPPs), the identification of critical variable categories plays a pivotal role in understanding the underlying dynamics and optimizing plant performance. Among the original pool of variables, including generator watts (load), turbine speed, control and feedback, pressure, temperature, steam flow, loading rate, and steam conductivity, only specific variable categories emerged as crucial for predicting load progression. The visual representation in Table 3 highlights the relative importance of these selected variable categories across the three feature selection techniques. The color gradient within the table denotes the rank of these categories, ranging from the brightest green (indicating the highest importance) to the lightest green (indicating diminishing importance) from top to bottom. This gradient provides valuable insights into the relative significance of each variable category, facilitating a deeper understanding of their impact on load progression dynamics. The following sections detail the importance of these selected variable categories and suggest optimization strategies to enhance predictive accuracy and operational efficiency.

Pressure is a fundamental variable in CCPPs, influencing steam generation, turbine efficiency, and overall power output. This variable was selected by all three feature selection techniques: PCC, RF-RFE, and LASSO. To optimize the inclusion of pressure in predictive models, sophisticated control algorithms and sensors can be employed to continuously monitor and adjust pressure levels in real-time. This approach minimizes deviations and ensures smooth load progression.

Temperature directly affects the thermal conditions within CCPPs and significantly impacts power generation dynamics. Like pressure, temperature was selected by PCC, RF-RFE, and LASSO. Implementing advanced temperature control systems, such as proportional-integral-derivative (PID), along with utilizing predictive analytics to anticipate temperature changes, can help adjust plant operations accordingly. This ensures safe operating conditions and efficient energy conversion.

Steam flow regulation is critical for supplying the necessary steam to turbines for power generation. Variations in steam flow can directly influence turbine speed and power output, making it essential for accurate load progression prediction. This variable was identified by PCC, RF-RFE, and LASSO. Maintaining precise control over steam valves and flow rates is crucial for consistent turbine performance. Advanced control strategies, such as model predictive control (MPC), can enhance steam flow regulation and improve load progression prediction accuracy.

Control and feedback mechanisms play a vital role in regulating plant components and optimizing energy conversion processes. These mechanisms were selected by PCC, RF-RFE, and LASSO. Fine-tuning control algorithms and feedback mechanisms to respond effectively to changing operating conditions can significantly improve load progression prediction accuracy. Utilizing advanced control strategies, such as adaptive control and fuzzy logic control, can further enhance system responsiveness and adaptability.

Turbine speed directly influences power generation and system efficiency in CCPPs. This variable provides valuable insights into load progression dynamics and allows for anticipating changes in turbine performance. It was selected by RF-RFE and LASSO, highlighting its importance. Implementing advanced speed control algorithms, such as PID control and utilizing condition monitoring systems to detect potential turbine issues can prevent disruptions to load progression, ensuring stable turbine operation within the desired range.

By incorporating these variable categories into predictive models and optimizing their control and monitoring strategies, CCPP operators can improve load progression prediction accuracy, enhance operational efficiency, and ensure consistent power generation.

B. Comparative Performance of Feature Selection Techniques Across Startup Classes

The selected variables by each feature selection technique—Pearson Correlation Coefficient (PCC), Recursive Feature Elimination with Random Forest (RF-RFE), and LASSO regularization—play a crucial role in predicting load progression during startup operations in Combined Cycle Power Plants (CCPPs). The variables identified by each technique provide insights into the key factors influencing load progression dynamics and contribute to the overall predictive model's accuracy.

In the case of **PCC**, the selected variables include pressure, steam flow, control and feedback, and temperature. However, it's notable that turbine speed, a significant predictor in startup operations, was not selected by PCC across all classes. The inherent constraint of filter feature selection techniques, such as PCC, lies in their reliance on individual correlations between variables, without considering their interactions. This limitation can hinder their predictive performance, especially in scenarios where variables interact synergistically to influence system behavior.

On the other hand, **RF-RFE** demonstrated superior performance by selecting turbine speed as one of the most influential variables across all startup classes. Alongside pressure, steam flow, control & feedback, and temperature, turbine speed played a pivotal role in predicting load progression during startup. Despite its indirect relationship with generator watts, this selection underscores the importance of turbine speed in predicting load progression during startup. RF-RFE's ability to iteratively evaluate feature importance based on model performance allowed it to capture the collective influence of variables and identify key predictors contributing to model accuracy.

Similarly, **LASSO** regularization also identified turbine speed as an influential variable, alongside control and feedback, pressure, steam flow and temperature, across all startup classes. The regularization effect of LASSO, which penalizes coefficients of less relevant features, effectively promotes the inclusion of variables contributing to predictive accuracy while shrinking coefficients of less influential features. This emphasizes turbine speed's significance in predicting load progression during startup operations, despite its indirect relationship with the target variable.

Overall, while PCC exhibited lower performance compared to RF-RFE and LASSO in all classes, RF-RFE outperformed others in the Hot and Warm 1 classes, whereas LASSO excelled in the Warm 2 and Cold classes. The consistent selection of turbine speed by RF-RFE and LASSO highlights its importance in capturing the underlying dynamics of startup processes, further emphasizing the need for comprehensive feature selection techniques to enhance predictive accuracy in CCPP operations.

IV. CONCLUSION

This study underscores the importance of advanced machine learning techniques and feature selection methodologies in improving smart grid intelligence, particularly in predicting load progression during CCPP startup operations. Leveraging LSTM neural networks and various feature selection techniques, including filter, wrapper, and embedded methods, our analysis aimed to enhance predictive accuracy and model performance, optimizing energy generation processes within smart grids. The project focused on identifying key operational parameters that significantly impact load progression prediction. This focused feature selection is crucial for improving model efficiency, enhancing predictive accuracy, and ensuring reliable energy system operations by prioritizing influential variables. Among the original pool of variables—generator watts (load), turbine speed, control and feedback, pressure, temperature, steam flow, loading rate, and steam conductivity—specific variable categories emerged as crucial for accurate prediction: Turbine Speed, Control and Feedback, Pressure, Temperature, and Steam Flow. These findings are pivotal in understanding the underlying dynamics and optimizing plant performance. Variations in the effectiveness of feature selection techniques across different startup classes were observed, highlighting the importance of tailored methodologies for specific operational contexts. For instance, superior performance in capturing startup dynamics in Hot and Warm 1 classes was demonstrated by RF-RFE, while excellence in Warm 2 and Cold classes was shown by LASSO. The omission of Turbine Speed by the PCC method, despite its selection by RF-RFE and LASSO, underscores limitations in traditional filter-based approaches and the need for sophisticated machine learning techniques capable of capturing complex interactions and non-linear relationships among variables. Moving forward, future research could explore integrating additional variables and advanced modeling techniques, such as ensemble learning and hybrid approaches, to further enhance smart grid intelligence. By improving model robustness and generalization capabilities, these advancements promise more accurate real-time predictions and informed decision-making in grid operations. Ultimately, this study contributes to the ongoing evolution of smart grids, offering insights that can optimize power generation, improve load distribution efficiency, and advance sustainable energy practices on a broader scale.

ACKNOWLEDGMENT

This work was supported by Malakoff Energy Sdn Bhd under the Industrial Grant Vote No: INDUSTRI/MALAKOFF/FTMK/2023/I00077. The authors express their appreciation for this support.

REFERENCES

[1] H. Farhangi, "The Path of the Smart Grid," IEEE Power and Energy Magazine, vol. 8, no. 1, pp. 18 - 28, 2010.

[2] A. M. COLAK and O. KAPLAN, " A Review on the Efficiency Increment in a Power System Using Smart Grid Technologies," in 9th IEEE International Conference on Smart Grid, Setubal, Portugal., 2021.

[3] A. Shobol1, M. H. Ali2, M. Wadi3 and M. R. TÜR, " Overview of Big Data in Smart Grid," in 8th International Conference on Renewable Energy Research and Applications, Brasov, Romania, 2019.

[4] X. Zhou, K. Li, Y. Ma, Gao and Zhiqiang, "Research Review on Big Data of the Smart Grid," in Proceedings of 2018 IEEE International Conference on Mechatronics and Automation, Changchun, China, 2018.

[5] L. Hou, Y. Zhang, Y. Yu, Y. Shi and K. Liang, " Overview of data mining and visual analytics towards big data in smart grid," in 2016 International Conference on Identification, Information and Knowledge in the Internet of Things, Beijing, China, 2016.

[6] S. Mishra, T. Sivani, T. Swain, V. Sharma and A. Alkhayyat, "Evaluating Energy Consumption Patterns in a Smart Grid with Data Analytics Models," in 2023 Second International Conference On Smart Technologies For Smart Nation (SmartTechCon), Singapore, 2023.

[7] R. A. Joy, "An Interpretable Catboost Model to Predict the Power of Combined Cycle Power Plants," in 2021 International Conference on Information Technology (ICIT), Amman, Jordan., 2021.

[8] P. Kaewprapha, P. Prempaneerach, V. Singh, T. Tinikul and N. Intarangsi, "Machine Learning Approaches for Estimating the Efficiency of Combined Cycle Power Plant," in 2022 International Electrical Engineering Congress (iEECON), Khon Kaen, Thailand, 2022.

[9] A. Kole, "A review and study on advanced control and automation functions and future control for a modern combined cycle power plant," in 2016 International Conference on Intelligent Control Power and Instrumentation (ICICPI), Kolkata, India, 2016.

[10] A. Bagheri and I. Y. Gu, "Big Data from Smart Grids," in IEEE PES Innovative Smart Grid Technologies Conference Europe, New York, 2017.

[11] Priyanka Saha, S. P. (2020). A Correlation - Sequential Forward Selection Based Feature Selection Method for Healthcare Data Analysis. IEEE International Conference on Computing, Power and Communication Technologies (GUCON). New Delhi, India.

[12] Y. Zheng, Y. Li, G. Wang, Y. Chen, Q. Xu, J. Fan and X. Cui, "A Novel Hybrid Algorithm for Feature Selection Based on Whale Optimization Algorithm," IEEE Access, vol. 7, pp. 14908 - 14923, 2018.

[13] B. W. Maulana, S. B. Utomo, K. Anam and I. Bilfaqih, " Analysis of Movement Patterns Based on Electroencephalograph Signals Using Pearson Correlation Coefficient," in 2023 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, Indonesia, 2023.

[14] B. W. Maulana, S. B. Utomo, K. Anam and I. Bilfaqih, " Analysis of Movement Patterns Based on Electroencephalograph Signals Using Pearson Correlation Coefficient," in 2023 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, Indonesia, 2023.

[15] M. G. Ismail, M. A. E. Ghany and M. A.-M. Salem, "Enhanced Recursive Feature Elimination for IoT Intrusion Detection Systems," in 2022 International Conference on Microelectronics (ICM), Casablanca, Morocco, 2022.

[16] R.K.Jeyauthmigha and R.C.Suganthe, "Recursive Feature Elimination and Clustering Technique for Network Anomaly Detection," in IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India, 2018.