Support Vector Machine Method for SF6 Gas Quality Classification

Zuhri Arieffasa S[1,2], Anton Sujarwo[1,2], Yandika Restu W[1,2], Ahmad Mushawir[1,3], Zainuddin[1,2] & Fahdzi Muttaqien[1]

[1]Institut Teknologi Bandung
[2]PT PLN (Persero)
[3]PT PLN Nusantara Power

Email: z.arieffasa@gmail.com[1,2], mail.anton.sujarwo@gmail.com[1,2], yandika.wara@gmail.com[1,2], ahmad.mushawir@gmail.com[1,3], zains2itbsk@gmail.com[1,2] & fahdzi@itb.ac.id[1]

Abstract. To meet the need for high reliability, increased load demand, and optimal maintenance strategy improvements, a machine learning-based method is introduced as an assessment approach for gas-insulated switchgear (GIS). Using historical data from different time periods, indicators representing the health level of GIS in terms of the quality of *Sulphur Hexafluoride* (SF6) gas are extracted using the support vector machine (SVM) algorithm. The data imbalance in the sample is normalized using the Synthetic Minority Oversampling Technique (SMOTE) to improve the accuracy of the training data. By combining inspection results and online monitoring data, the operational condition panorama of each GIS compartment is classified as the model output. Based on data obtained from three hundred and twenty-two GIS compartments in the Surabaya area, the method's effectiveness is demonstrated through sample test results.

Keywords: *gas-insulated-switchgear; SF6; machine learning; SVM*

1 Introduction

Based on data obtained from PLN, eight out of twenty-eight GIS (Gas Insulated Switchgear) installations have undergone dismantling (overhaul) due to anomalies over the past ten years. Although GIS systems are known for their high reliability and require minimal maintenance, these systems can deteriorate over time due to various factors such as environmental conditions, operational stress, and aging. Therefore, they must be regularly maintained and inspected at intervals specified and recommended by the manufacturer (Panmala, et al., 2023). In research conducted by Purnomoadi (2020), following energy sector deregulation in the 1990s, which was also triggered by aging infrastructure and increased demand from regulators and customers, many utility networks adopted asset management in the hope of generating more revenue, achieving better credit ratings, and stabilizing stock prices. GIS was selected as the subject of this research for the following reasons:

1. GIS is a critical asset in the transmission network.

2. It has fewer references for decision-making efforts compared to power transformers.

3. In tropical conditions, GIS failures are twice the rate reported in the 3rd CIGRE survey in 2007.

Given these studies and conditions, this research will discuss the extraction of SF6 gas quality parameters in GIS through various attributes processed using the SVM algorithm, resulting in output in the form of anomaly classifications. The research data used consists of measurement results of SF6 gas quality from 3,391 compartments spread across 26 different GIS locations. The data was collected over the course of one year. The assessment process, which employs the scoring method, takes six to eight months to complete. As a result, the anomaly classification outcomes can only be determined in the following year.

As an alternative option to obtain anomaly classification results without waiting for six to eight months, this study can produce classification results within minutes with an accuracy rate of over 90%. By utilizing machine learning algorithms, this research offers a comprehensive approach to generate a structured anomaly classification dataset for GIS.

2 Gas Insulated Switchgear

Gas-Insulated Switchgear (GIS) is an integrated combination of electrical equipment housed within a sealed enclosure. This structure consists of several electrical components insulated by sulfur hexafluoride gas, or SF6 (Wang et al., 2017). These electrical components include circuit breakers, current transformers, disconnect switches, and earthing switches (Cho H et al., 2006).

Although GIS has advantages over conventional substations, it has a complex structure in terms of manufacturing, mobilization, and assembly, which can lead to defects or even internal damage. This internal damage can result in partial discharge (PD), potentially degrading insulation quality. Partial discharge is an electron jump that indicates an abnormality within the insulation of GIS compartment. Partial discharge can only be detected by sensors when the GIS is in an active or energized state. If insulation quality continues to deteriorate and is not promptly addressed, it may lead to damage to the GIS structure or even widespread disruption in the electrical distribution system (Zeng et al., 2018:103). To maintain quality and detect anomalies in a GIS compartment, regular maintenance is necessary by conducting a condition assessment of the SF6 gas used as the insulation medium (Kurniawati et al., 2022).

3 Supervised Learning

Supervised learning is an approach to building AI. It is called "supervised" because, in this approach, machine learning is trained to recognize patterns between input data and output labels. Furthermore, machine learning is also trained to identify the underlying relationships connecting the input data with the output labels. Supervised learning is a subset of machine learning that involves a set of M inputs (xi) and outputs (yi). (xi) is an M by N matrix where M is the number of data points and N is the number of input features or attributes. (xi) represents a vector of response features. When the response feature (yi) is numeric, it indicates a regression problem. Alternatively, it could indicate a classification or pattern recognition issue. Examples of supervised machine learning algorithms include artificial neural networks (ANN), support vector machines (SVM), random forest (RF), extreme random forest (ERF), gradient boosting (GB), adaptive gradient boost (AGB), multi-linear regression (MLR), logistic regression (LR), and KNN (Belyadi & Haghighat, 2021:109).

4 Support Vector Machine

A type of machine learning that can be used in anomaly detection and diagnosis is the support vector machine (SVM). Support Vector Machine (SVM) is a machine learning algorithm used for data modeling and classification. SVM is effective in handling classification and regression problems by constructing a hyperplane that optimally separates two data classes. The design of SVM maximizes the margin, which is the distance between the hyperplane and the nearest points of each class, thereby providing stability and good generalization for new data. SVM maps input vectors into higher-dimensional feature spaces. SVM classification is designed for binary classification, aiming to separate a set of vectors into two different classes, with support vectors being the sample data points closest to the decision boundary. SVM also provides user-definable parameters that allow the user to adjust the balance between the number of misclassified samples and the width of the decision boundary (Sonajharia, et al., 2023).

SVM is known for its ability to handle high-dimensional data and its advantage in dealing with imbalanced data. Its flexibility and reliability make SVM a popular algorithm in various machine learning and pattern recognition applications. In this study, SVM with an RBF kernel is used due to the radial structure of data correlations observed in attribute plots. Therefore, RBF is considered effective and suitable for use with this dataset.

5 Methodology

The methodology used in this research includes preprocessing, training, and testing data. The methodology contains the following steps in developing a condition monitoring system for SF6 gas quality using the support vector machine (SVM) algorithm.

Data Collection
Collecting historical data is the primary step taken. This data consists of SF6 gas quality parameters such as purity, moisture content, dew point, SO2 content, and compartment circuits.

Sensor Integration
Special sensors designed to detect SF6 gas quality parameters will be integrated with GIS assets.

Data Normalization
Data Normalization involves managing the dataset to ensure consistency and accurate comparison between the parameters used.

Data Analysis Using SVM
Training with SVM will be conducted on historical data that has undergone normalization. This algorithm is used with the aim of producing outputs that identify anomalies in the data, indicating issues or deteriorating conditions.

Model Prediction
The predictive model is developed by leveraging detected anomaly patterns from two monitoring sessions. This model will enable the estimation of future changes in GIS conditions.

Model Evaluation
The predictive modeling will be evaluated and validated to determine the quality of the predictions, ensuring the accuracy and reliability of the prediction results.

By using SVM as the processing and analysis algorithm, this research aims to create an intelligent anomaly identification system and an accurate predictive model to improve GIS reliability, thereby enhancing operational efficiency in electricity distribution.

5.1 Preprocessing Data

This process is carried out to normalize the data, making it ready for analysis. At this stage, it can also be determined whether the data contains any missing values. In this study, the data types used are already in integer or float format, making the data fairly balanced. After the data is collected, the normalization process begins by identifying the data type in each column. Since the data processing can only be performed on numerical data types, the next step is to convert all data to numerical types. Then, it is necessary to ensure that every cell is completely filled. If there are any empty cells, they need to be deleted or replaced with random data to avoid obstructing the data processing.

Generally, the dataset represents the latest inspection and measurement results on the quality of SF6 gas functioning in dielectric subsections. The dataset comprises 8 attributes and 322 rows of data, described as follows.

Table 1 Normalized Dataset

	PURITY	MOISTURE_CONTENT	DEW_POINT	SO2	PRESSURE_GAUGE	CB_COMP	DS_COMP	ANOMALY
0	99.98	107.24	-47.58	0.0	3.69	0.0	1.0	0.0
1	99.29	373.73	-30.16	0.0	3.64	0.0	1.0	0.0
2	99.20	93.32	-42.79	0.0	3.54	0.0	1.0	0.0
3	99.93	85.54	-43.53	0.0	6.44	1.0	0.0	0.0
4	98.65	60.44	-46.47	0.0	6.28	1.0	0.0	0.0

The parameters used in this data include purity, moisture content, dew point, SO_2 (sulfur dioxide), pressure gauge, cb_comp (circuit breaker compartment), ds_comp (disconnecting switch compartment), and anomaly index which will be identified on the testing dataset.

Once the data composition is ready, the next step is to identify the number of instances in each class of the target variable. This is done to ensure that the machine learning model can comprehensively learn the data structure. In this case, an imbalance is found between the number of anomaly and non-anomaly data points. Python provides the random_over_sampling library, which can be used to create a new database with equal composition in each class. To address data imbalance, the application of SMOTE (Synthetic Minority Over-Sampling) can be used. SMOTE is a technique in machine learning to handle class imbalance issues, where the data in one class is significantly less than in the other classes. This helps to prevent the model form being biased towards the majority class. SMOTE enhances a model's ability to handle class imbalance by improving sensitivity to the minority class. This often results in better recall and balanced evaluation metrics. However, it is essential to use SMOTE to avoid overfitting or amplifying noise in the data.

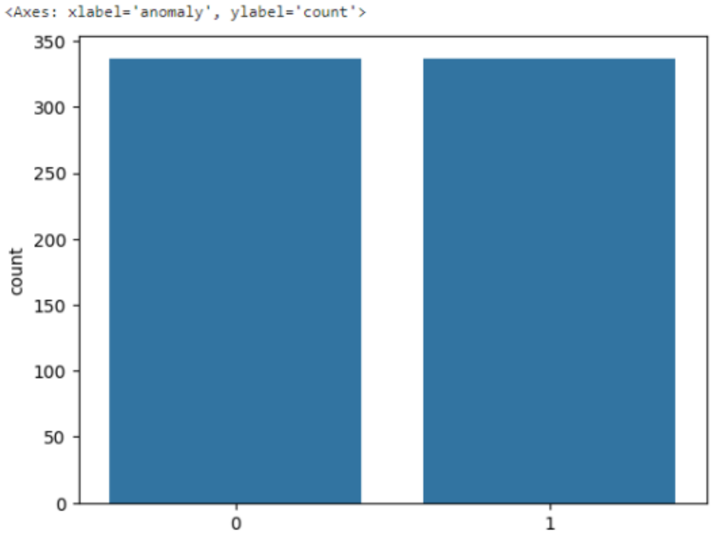

Figure 1 SMOTE Diagram

After addressing the data imbalance, the next step is to select the parameters most influential to the target. In this modeling, there are seven parameters, namely purity, moisture content, dew point, SO2, pressure gauge, CB compartment (cb_comp), and DS compartment (ds_comp), along with one target: anomaly. Purity, moisture content, dew point, and SO2 represent the quality conditions of SF6 gas, which can be determined through test results. Pressure gauge refers to the SF6 gas pressure used to isolate electron jumps in the GIS compartments. Meanwhile, cb_comp and ds_comp are parameters to identify the influence of switching compartment functions on changes in SF6 gas quality. The selection of these parameters can be seen through the Heatmap Correlation Table, as shown in Figure 2 below.

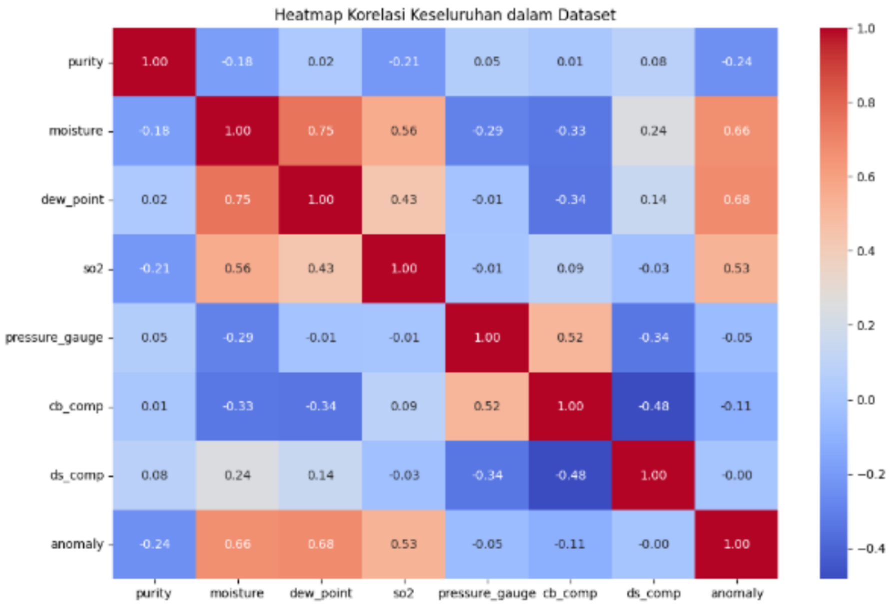

Figure 2 Heatmap Correlation Diagram

Based on the results shown in the heatmap correlation diagram in Figure 2, there are several attributes with relatively high correlations compared to others. These attributes are moisture, dew point, and SO_2 content. These attributes are then used in the testing data to improve the quality of the SVM in learning patterns and modeling the condition of SF6 gas quality.

5.2 Training and Testing Data

There were nine trials with differences in the percentage of the comparison between training data and testing data. The modelling result is summarized in the following table.

Table 2 Classification Result

NUM	TRAIN (%)	TEST (%)	ACCURACY	PRECISION	RECALL	SPECIFICITY	F1_SCORE
1st	90	10	0,9804	1,0	0,9524	1,0	0,9756
2nd	80	20	0,9802	1,0	0,9412	1,0	0,9697
3rd	70	30	0,9671	1,0	0,9000	1,0	0,9474
4th	60	40	0,9455	1,0	0,8514	1,0	0,9197
5th	50	50	0,9365	1,0	0,8202	1,0	0,9012
6th	40	60	0,9241	1,0	0,7788	1,0	0,8757
7th	30	70	0,9263	1,0	0,7778	1,0	0,8750
8th	20	80	0,9307	1,0	0,7910	1,0	0,8833
9th	10	90	0,9427	1,0	0,8255	1,0	0,9044

Accuracy: This is the proportion of correctly classified instances out of the total instances. A higher accuracy means the model is generally performing well. However, it can sometimes be misleading if there's an imbalance in classes.

Precision: Precision is the proportion of true positives out of all predicted positives. Since precision is 1.0 for all entries, it indicates that there are no false positives in these models — when the model predicts a positive, it's always correct.

Recall: Recall (or sensitivity) is the proportion of true positives out of all actual positives. This metric varies significantly across different splits. A higher recall means the model is successfully capturing the majority of the actual positives, so higher recall is better.

Specificity: Specificity is the proportion of true negatives out of all actual negatives. All models have a specificity of 1.0, indicating there are no false negatives in these models, which is excellent for negative class predictions.

F1 Score: The F1 Score is the harmonic mean of precision and recall. It gives a single metric that balances precision and recall. A higher F1 score indicates better overall performance in terms of positive class prediction.

Looking at Accuracy, Recall, and F1 Score together, the model with a 90% training and 10% testing split (the 1st row) achieves the highest accuracy (0.9804) and the highest F1 score (0.9756). However, since precision and specificity are perfect (1.0) for all models, the difference mainly lies in recall and F1 score. Here, the 90-10 split also has the highest recall (0.9524), making it the best performing model in terms of balanced metrics.

Based on the Table 2, the first, second, and the third model have slightly different value of accuracy, recall, and F1 Score. Each of them has more than 90% accurate and precise. To choose the best model among three of them are by looking for the learning curve of each models. Learning curve figures out how overfit or underfit the model is. If the training error is low but the validation error is high, the model is likely overfitting, meaning it performs well on the training data but poorly on unseen data. Meanwhile, if both training and validation errors are high and converge, the model is likely underfitting, meaning it's too simple to capture the underlying patterns in the data. The confussion matrix and learning curve of the considered model is illustrated on the figure 3 and 4 below.

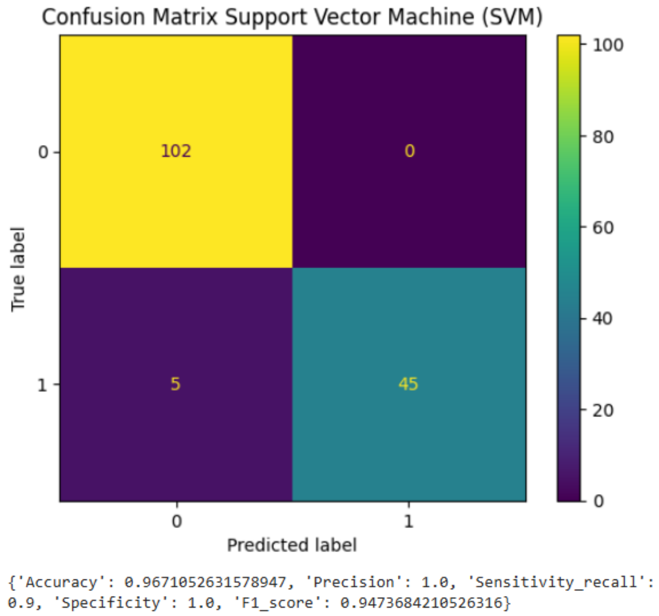

Figure 3 Confussion Matrix (Train:Test = 70:30)

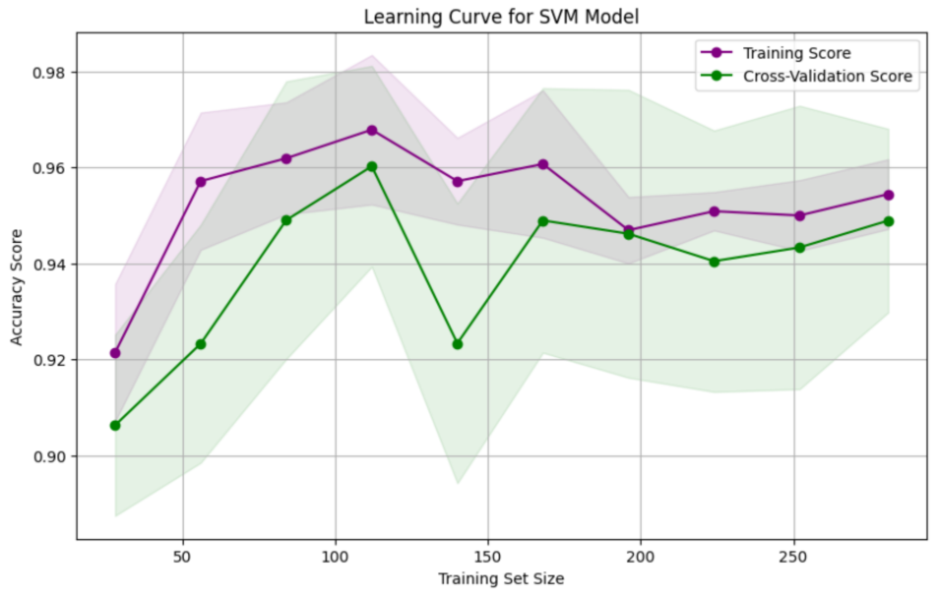

Figure 4 Cross-Validation Curve (Train:Test = 70:30)

Based on the learning curve illustrated in Figure 4, the model with a composition of 70% training data and 30% testing data shows that the testing and training curves do not intersect. The training set curve is positioned above the testing set curve. Although there are certain iterations where error points are almost overlapping, overall, the learning curve above demonstrates that the model is neither underfitting nor overfitting. This differentiates it from the learning curve in models with 90:10 and 80:20 training-to-testing compositions. In these two compositions, the training and testing curves show intersection and convergence. Therefore, the learning curve with a composition of 70% training and 30% testing is the best choice for this classification model.

6 Conclusion

The evaluation results of the SVM model performance show very high accuracy, with accuracy, precision, specificity, and F1 Score reaching 98.02%, 100%, 100%, and 96.97%, respectively. Based on these percentages, as demonstrated by the learning curve, the use of SVM with an RBF kernel in this study has proven to be effective. With a training and testing composition of 70% and 30%, the model created is neither overfitting nor underfitting. The modeling results indicate that SVM can identify GIS compartment anomalies very effectively. The identification of SF6 gas conditions based on quality testing attributes using machine learning methods has been successfully implemented. As part of GIS condition assessment, the use of this model provides benefits in terms of time and a more effective method. Furthermore, this SVM model can be further developed for application in GIS assessment from both electrical and environmental aspects.

References

[1] Belyadi, H., & Haghighat, A. (2021). *Machine Learning Guide for Oil and Gas Using Python. In Machine Learning Guide for Oil and Gas Using Python (p. iii)*. Elsevier. https://doi.org/10.1016/b978-0-12-821929-4.01001-5

[2] Cho, H., Lee, S., & Lim, S. (2006). *Dynamic analysis on the closing resistors of Gas Insulated Switchgea*r. Journal of Mechanical Science and Technology (KSME Int. J), 20(10), 1607–1613.

[3] Geng, S., & Wang, X. (2020). *Research on data-driven method for circuit breaker condition assessment based on back propagation neural network*. Computers and Electrical Engineering, 86. https://doi.org/10.1016/j.compeleceng.2020.106732

[4] Ibrahim, V. M., Abdul-Malek, Z., & Muhamad, N. A. (2017*). Status review on gas insulated switchgear partial discharge diagnostic technique for preventive maintenance.* Indonesian Journal of Electrical Engineering and Computer Science, 7(1), 9–17. https://doi.org/10.11591/ijeecs.v7.i1.pp9-17

[5] Kurniawati, A. H., Mumtaza Ghalya, A., Gayatri, N. C., Januar, D., & Kusnadi. (2022). *SF6 Gas Parameter for Condition Assessment in GIS*. Proceedings of the National Seminar on Electrical Engineering, 7.

[6] Panmala, N., Suwanasri, T., & Suwanasri, C. (2023). *A Fuzzy Logic Approach to Health Index Determination for a Gas-Insulated Switchgear.* Energies, 16(18). https://doi.org/10.3390/en16186605

[7] Purnomoadi, A. P., Mor, A. R., & Smit, J. J. (2020). *Health index and risk assessment models for Gas Insulated Switchgear (GIS) operating under tropical conditions*. International Journal of Electrical Power and Energy Systems, 117. https://doi.org/10.1016/j.ijepes.2019.105681

[8] Srinuntawong, W., Srangtook, W., Kerdmanee, S., Suwanasri, T., Suwansari, C., Fuangpian, P., Kumpalavalee, S., & Somsak, T. (2019). *Data Warehouse and Asset Management Intelligence Architecture for Condition Assessment of Major Equipment in Power Plant*. IEEE PES GTD Asia, 532–537.

[9] Wang, Q., He, Z., Lin, S., & Li, Z. (2017). *Failure Modeling and Maintenance Decision for GIS Equipment Subject to Degradation and Shocks*. IEEE Transactions on Power Delivery, 32(2), 1079–1088. https://doi.org/10.1109/TPWRD.2017.2655010

[10] Zeng, F., Tang, J., Zhang, X., Zhou, S., & Pan, C. (2018). *Typical Internal Defects of Gas-Insulated Switchgear and Partial Discharge Characteristics. In Simulation and Modelling of Electrical Insulation Weaknesses in Electrical Equipment.* InTech. https://doi.org/10.5772/intechopen.79090.