Open Access | Rapid and quality publishing

International Journal of Computer Science Research (IJCSR)
Survey Article | Volume 3, Issue 1 | Pages 01-10 | ISSN XXXX-XXXX

A Comparative Analysis of Support Vector Machine and Decision Tree Algorithm for Predicting Fault in Technical Data Schedule Uninterruptible Power Supply System for Ghana Gas Limited

Isaac M. Doe[1], J. K. Annan[2] and B. Odoi[3]

[1]*University of Mines and Technology, UMaT, Tarkwa, Ghana*
[2]*Department of Electrical and Electronic, UMaT, Tarkwa, Ghana*
[3]*Department of Mathematical Sciences, UMaT, Tarkwa, Ghana, bodoi@umat.edu.gh*

*__Correspondence:__ *bodoi@umat.edu.gh*

ABSTRACT- Power supply systems can have problems, and Ghana Gas Limited is not an exception. Ghana Gas Limited uses an intricate Uninterruptible Power Supply (UPS) system called the Technical Data Schedule (TDS), which is made up of several parts such as electromechanical components, PCB boards, and electrolytic capacitors.The majority of components have technical lifespans that are governed by usage, operational environment, and working conditions, such as electrical stress, working hours, and working cycles. Most of the time, these errors affect the integrity and power supply after manufacture. The issue is that it takes longer for the professionals who operate on this machine to recognize these flaws, which makes it difficult for them to predict errors quickly or anticipate the likelihood of faults happening in the system components at an early stage for effective corrective action to be performed. Support vector machines (SVM) and decision trees were used in this study to anticipate faults for technical data scheduling of uninterruptible power supply systems for Ghana Gas Limited in an efficient manner. Based on a comparative analysis using these two techniques, faults in Ghana Gas Limited's power supply system were predicted using a four-hour daily interval dataset on TDS UPS recordings, including input voltage, battery voltage, battery current, and alarm, spanning from August 2017 to October 2023. The findings depicted that the support vector machine was more efficient in detecting the fault locations in the power supply system with an accuracy of 96.80%, recall of 99.80%, precision of 100 %, F1-score of 93.15%.The results from the error metrics also validate the measures in assessing the predictive ability of the model with MAE of 0.42%, MSE of 1.18%, RMSE of 4.45%, R2 of 99.97%, RMSLE of 0.036%, and MAPE of 0.21%.

Keywords: Power Supply System; Support Vector Machine; Decision Tree Algorithm; Precision; Accuracy; Error Metrics

ARTICLE INFORMATION

Author(s): Isaac M. Doe, J. K. Annan and B. Odo;

Received: 17/01/2024; **Accepted**: 25/01/2024; **Published**: 30/03/2024;
e-ISSN: XXXX-XXXX;
Paper Id: IJCSR-030106;
Citation: 10.37391/IJCSR.030106

crossref member
CROSSREF.ORG
THE CITATION LINKING BACKBONE

Publisher's Note: FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

The majority of engineers now find that their ability to distribute an uninterruptible electrical power supply and safeguard their operations is limited by how sophisticated they are at handling power supply outages [1].Larger industries utilize uninterruptible power supplies (UPS) as an emergency power source to make sure business activities continue as usual in the event that the primary power supply fails. When it comes to providing almost instantaneous safety from input power failures, a UPS is different from a backup or supplemental generator since it uses electricity produced by batteries, ultracapacitors, or flywheels [2]. A variety of scenarios need the use of uninterruptible power supply (UPS) to provide consistent and well-regulated AC voltages for critical workloads, including computer servers, medical equipment, air traffic control systems, and communication networks [3]. It has been noted in numerous deployments that when system load increases over time, an upgraded UPS with a larger capacity is needed [4]. As power grids grow and loads rise, the primary goal of a UPS is to ensure the stability and dependability of the electrical supply. Furthermore, as the system serves as a precaution or protection against any unanticipated power outages that the companies may suffer, it must be reliable, secure, and less prone to malfunctions [5] and [6]. However, overloading, overvoltage, power fluctuations, and outdated components can result in malfunctions in UPS systems, which can cost the industry a lot of money in terms of lost productivity and system replacement [4]. Fault prediction is the process of tracking and evaluating historical data to identify the presence of a failure in the power system, so that actions may be taken to prevent accidents and assure system recovery [7]. As a result, it is essential to implement fault prediction models that are efficient and accurate in determining when faults may occur [1]. Fault prediction is an essential technology and maintenance security technique that is more advanced than fault diagnosis, which is frequently carried out after issues have happened [8]. Making wise decisions to avoid errors and minimize their negative consequences is made easier with the help of fault prediction. In order to reduce the frequency and duration of

Survey Article | Volume 3, Issue 1 | Pages 01-10 | ISSN XXXX-XXXX

power outages, utility workers can find and eliminate persistent defects with the use of high-accuracy fault prediction in power systems [9].

Previously, UPS system maintenance was reactive, starting when a problem was found. Recent years have seen an increase in the popularity of preventive maintenance, which involves changing out components to increase system reliability [10]. The simplest kind of preventive maintenance is to target consumable parts before they are expected to reach the end of their useful life. Fans, AC and DC capacitors, and of course the battery are all consumable parts of UPS systems [11]. Preventative maintenance is the failure prediction based on prediction models created from gathered data or log files in more sophisticated systems [12]. Making the right decisions to prevent power system problems and estimating their likelihood can both be aided by the analysis of historical data [13]. In general, utilizing electrical measurement data to its fullest potential will increase the accuracy of failure prediction and guarantee the stability and dependability of the UPS system.

Studies utilizing artificial intelligence (AI) and machine learning have been developed in the past several years to predict faults. Right now, it's a worthwhile and pressing topic [13]. [14] provided a more accurate prediction strategy for optimized Artificial Neural Networks (ANN) based on multilayer evolutionary algorithms in an effort to improve the fault forecasting model's accuracy. [15] employed ANN to obtain likelihoods of success for five fault prediction techniques ranging from 87% to 100% using 33 data sets. [16] looked at the use of convolutional neural networks (CNNs) to forecast refrigerant charge failures. Two classification and regression predictive models were suggested in order to predict the quantitative refrigerant in both cooling and heating applications. In summary, the recommended tasks were finished with a 3.1% mistake rate and 99% accuracy. In order to predict power converter failures, [17] created a fault prediction model utilizing Markov Chains Analysis based on data collected from several UPS installations.

Furthermore, classification is an important part of the fault prediction process. The Support Vector Machine (SVM) is a hyperplane-configured discriminant classifier. SVM-based applications have been shown to be feasible in [18] and [19].
This study aims to estimate the failure rate of the Technical Data Schedule (TDS) UPS systems by using Markov Chains Analysis, Support Vector Machine (SVM), and Decision Tree (DT) Algorithms to create prediction models that the Ghana Gas engineering team would use. The Ghana National Gas Company, also known as Ghana Gas, was established with the responsibility of creating, acquiring, and overseeing the natural gas infrastructure required for the processing, transportation, and marketing of gas to satisfy the nation's needs for both household and commercial electricity. Most importantly, a consistent and dependable power source is essential to Ghana Gas's operations and activities. Predicting TDS UPS failures is therefore necessary to enhance power supply performance and lower the company's total operational expenditure (OPEX).

The TDS UPS employed by Ghana Gas is a complicated system consisting of a number of parts, including PCB boards, electromechanical components (such as relays and fans), and electrolytic capacitors. The majority of components' lifespans are determined by their technical attributes and are influenced by their usage, operational environment, and working conditions, that is, working hours, working cycles and electrical stress [20]. Currently, preventative maintenance is conducted over a predetermined time period without considering the level of stress experienced or the overall health of the UPS system. For instance, fans are typically replaced every five years without considering whether the UPS was, perhaps, in a clean room or in a harsher, dustier environment [21]. Costly on-site maintenance is carried out irrespective of the device's status and may therefore be too late or too early. The latter situation could result in the servicing of a healthy component, thereby increasing the company's financial costs and decreasing the UPS systems' reliability [22].

A model for accurate fault prediction and forecasting will help to improve the level of UPS system reliability and reduce power quality disturbances as well as equipment damages [22]. Furthermore, the prediction model will enable Ghana Gas to better manage their engineering resources by forecasting failures and enabling precautionary actions to be performed in order to reduce operating expenses caused by unnecessary component replacement and additional charges. An important step in understanding the reliability of the UPS system as a whole is determining the significance of various UPS parameters [23]. Hence, by monitoring UPS parameters such as output voltage, output current, frequency, power factor, working hours, active and reactive powers, a run-time lifespan calculation of the components could be conducted to determine its health state and predict potential UPS failures. Obviously, to conduct such a task, an intelligent data methodology has to be employed.

This study emphasis is on utilizing machine learning algorithms for fault evaluation and prediction in the TDS UPS system. The research discusses fault prediction models with a particular emphasis on UPS operations. Particularly, operational data from UPS installations has been recorded. In order to create the failure prediction models, data is then processed using Support Vector Machines (SVM), and Decision Tree (DT) algorithms.

2. MATERIALS AND METHODS

2.1 Data collection

The data for the study were obtained by TDS UPS (recordings include input voltage, battery voltage, battery current, and alarm), which records an observation every four hours. The data was gathered between August 2017 and October 2023. The Python programming language was used to do the analysis.

2.2 Support Vector Machine (SVM)

Given a set of training data $\{(x_1, y_1), \ldots, (x_1, y_N)\}$, where $x_i \in R^D$ are the input vectors and $y_i \in \{-1,1\}$ are the corresponding class labels, an SVM seeks to construct a hyperplane that separates the data with the maximum margin of separability

Survey Article | Volume 3, Issue 1 | Pages 01-10 | ISSN XXXX-XXXX

[24]. N is the number of observations, and D is the dimension of the input vectors. The decision function can be written as

$$f(x) = sign\{\sum_{j=1}^{N^{sv}} \alpha_j y_j^{sv} (\Phi(x) \cdot \Phi(x_j^{sv})) + b\} \quad (1)$$

Where x_j^{sv} are the support vectors, is $\Phi(x)$ a nonlinear vector function that maps the input vector onto a higher dimensional feature space [25], y_j^{sv} is the label corresponding to the th support vector, N^{sv} is the number of support vectors, b is a bias term, and α_j are the Lagrangian multipliers, The inner product is $(\Phi(x) \cdot \Phi(x_j^{sv}))$ called the kernel function.s

2.3 Decision Tree (DT):

Decision Tree builds models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. Information gain is one metric used for segmentation. In order to split the data at the most informative features, we define an objective function;

$$IG(S, A) = \sum_{v \in V(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

Where the range of attribute A is $V(A)$, and S_v is a subset of set S equal to the attribute value of attribute v. Entropy is employed to measure a dataset's impurity or randomness. The value of entropy always lies between 0 and 1.

$$Entropy(S) = \sum_{i=1}^{c} P_i \log \log 2^{P_I} \quad (3)$$

Where P_i is the ratio of the sample number of the subset and the $i - th$ attribute value.

2.4 Evaluation Metrics

This study compares many machine learning techniques in an attempt to forecast the location of a wire drawing process problem. Therefore, six (6) well-known assessment measures that are frequently utilized in fault prediction applications were used to gauge how well these algorithms performed. Since a model's effectiveness cannot be determined by a single metric, these evaluation metrics were selected. Below is a discussion of various evaluation metrics [26]–[33].

2.4.1 Accuracy

The accuracy is determined by the fraction of correctly predicted classes, as (4). Therefore, the ratio of accurately predicted events to all predicted events is known as accuracy.fraction of classes that are predicted correctly determines the accuracy as shown (4). Thus, accuracy is the proportion of correctly predicted events to all predicted events.

$$\text{Accuracy} = \frac{TP + FN}{TP + FN + TN + FP} \quad (4)$$

where, in order of representation, FN, TP, FP, and TN stand for false negatives, true positives, false positives, and true negatives.

2.4.2 Recall

It shows how well a classification model classifies the positive class when the actual result is positive. As a result, it is better to have a recall score that is closest to 1. By (5), its equation is provided.illustrates how effective a classification model is at categorizing the positive class when the true outcome is positive. Consequently, a recall score that is closest to 1 is preferable. Its equation is given by (5).

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

2.4.3 Precision

The ability of a classification model to classify a positive class is evaluated. Therefore, of all the projected classes of defects, it displays the proportion of faults that were accurately predicted. Nearly perfect outcomes are desirable. (6) displays its equation.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

2.4.4 F1-score

It displays the equality of the consonant mean (P) and sensitivity (R), as illustrated in (7). It is utilized as a trade-off between the classifiers and basically combines recall and precision into one parameter.

$$\text{F1-score} = \frac{2 \times P \times R}{P + R} \in \{0,1\} \quad (7)$$

2.4.5 Kappa

K The likelihood of agreement is measured using kappa in relation to what would be predicted if the assessments were independent. The values in the range are [-1, 1], where 1 denotes complete agreement, 0 denotes independence, and negative denotes worse agreement. (8) gives its equation.

$$\text{Kappa} = \left\{ \frac{P_{(s)} - P_{(y)}}{(1 - P_{(y)}} \right\} \in \{-1,1\} \quad (8)$$

where P(y) represents the expected agreement and P(s) represents the predicted agreement.

3. ERROR METRICS

The fitness and accuracy of the various models' predictions were assessed using the error measures. As will be covered below, six indices were used to evaluate the faults. In general, the model performs better the lower the value of these error metrics. On the other hand, a greater R^2 value is preferred [31]–[33].

3.1 Root Mean Square Error

The Root Mean Square Error (RMSE) measures how anticipated mistakes vary among actual data sets. To put it another way, the RMSE clarifies the degree to which the projected values of an estimated model agree with the observed data points. The following is the formula:

International Journal of
Computer Science Research (IJCSR)
Survey Article | Volume 3, Issue 1 | Pages 01-10 | ISSN XXXX-XXXX

$$RMSE_{forecast} = \sqrt{\sum_{i=1}^{n} \left(\frac{Y_t - \hat{Y}_t}{n} \right)^2} \quad (9)$$

where $\hat{Y}_t$ indicates the prediction, Y_t the real data sets, and n the size of the sample.

3.2 Mean Absolute Percentage Error

A forecast's error is measured as a percentage magnitude using the Mean Absolute Percentage Error (MAPE). The following is how it is expressed and used to assess forecast accuracy:

$$MAPE_{forecast} = \left(\frac{1}{n} \sum \frac{|Y_t - \hat{Y}_t|}{|Y_t|} \right) \times 100\% \quad (10)$$

3.3 Mean Absolute Error

For predicting performance, the most basic metric is the Mean Absolute Error (MAE). As demonstrated in (11), MAE indicates the size of the expected error from the mean estimation.

$$MAE = \left(\frac{1}{n} \sum |Y_t - \hat{Y}_t| \right) \quad (11)$$

3.4 Mean Squared Error

The variance between the actual and anticipated values is calculated as the mean squared. Its formula is displayed in (12).

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_t - \hat{Y}_t)^2 \quad (12)$$

3.5 Root Mean Squared Logarithmic Error

The ratio between the actual and expected values is calculated by taking the log of the predicted and actual values. An underestimation is used instead of RMSE if it is deemed to be worse than an overestimation. It's articulated by (13).

$$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\log(Y_t + 1) - \log(\hat{Y}_t + 1)^2} \quad (13)$$

3.6 R^2 Value

The degree of agreement between the expected and actual values is shown by the R^2 value, also known as the coefficient of determination. As such, it denotes the percentage of variation accounted for by a model, as shown in (14).

$$R^2 = 1 - \frac{\frac{1}{n} \sum_{i=1}^{n} (Y_t - \hat{Y}_t)^2}{\frac{1}{n} \sum_{i=1}^{n} (Y_t - \bar{Y}_t)^2} \quad (14)$$

The UPS data includes four-hour time periods for each of the equipment's four shifts. In all, 3912-time frames were created over the course of three years. The objective is to generate statistical features for each window in order to better define each daily signal and to reduce the data's dimensionality. To further describe the signal and understand its evolution over time, eight different characteristics were extracted from each data frame.

Maximum, Minimum, Mean, Standard Deviation, Root Mean Square (RMS), Skewness, and Kurtosis Mean Absolute Deviation (MAD) were features generated. These eight variables were chosen to limit the study's resources and see whether they are adequate to achieve an identification. As none of the variables need frequency analysis, the features are all conducted in the time domain. Furthermore, according to [37], time domain statistical resources provide a high performance to characterize trends and changes.

Table 1: Extracted Statistical Features for the Input Voltage Attribute

	Max	Min	Mean	Std	RMS	Skewness	Kurtosis	MAD
04/01/2021	423	418	419.8333	1.94079	419.8371	0.46611	-1.52589	1.5
05/01/2021	427	409	421.5	7.148426	421.5505	-0.75421	-1.28666	5.666667
06/01/2021	422	416	418.8333	2.041241	418.8375	0.185073	-1.38893	1.5
07/01/2021	439	420	428.6667	7.763161	428.7252	0.066101	-1.9978	6.666667
08/01/2021	440	417	428.6667	9.884178	428.7711	0.014249	-2.07055	9.333333
09/01/2021	441	430	436	4.289522	436.0176	-0.15204	-1.92683	3.666667
10/01/2021	442	430	436.6667	4.082483	436.6826	-0.32932	-1.30493	3
11/01/2021	429	421	426.25	3.593976	426.2614	-0.63615	-1.76142	2.625
12/01/2021	431	422	425.5	4.1833	425.5171	0.293685	-2.03381	3.5

04/02/2021	427	421	423.3333	2.65832	423.3403	0.181385	-2.04845	2.333333

Table 2: Extracted Statistical Features for the Battery Voltage Attribute

	Max	Min	Mean	Std	RMS	Skewness	Kurtosis	MAD
04/01/2021	255	251	253.5	1.974842	253.5064	-0.45443	-1.97682	1.666667
05/01/2021	255	254	254.6667	0.516398	254.6671	-0.53791	-1.95833	0.444444
06/01/2021	255	250	253.8333	1.94079	253.8395	-1.17541	-0.39816	1.277778
07/01/2021	254	252	253.5	0.83666	253.5012	-0.85373	-1.17177	0.666667
08/01/2021	255	235	253	5.688905	253.0586	-2.61252	5.3613	3
09/01/2021	255	250	253.1667	2.136976	253.1742	-0.48579	-1.83196	1.777778
10/01/2021	256	254	254.3333	0.816497	254.3344	1.360828	-0.08333	0.555556
11/01/2021	254	251	252.5	1.290994	252.5025	0	-2.0775	1
12/01/2021	254	250	252.3333	1.861899	252.3391	-0.09181	-2.18023	1.666667
04/02/2021	255	251	253.1667	1.722401	253.1715	-0.4059	-1.92401	1.444444

Tables 1 and 2 show the first ten statistical features extracted from daily input and battery voltage values recorded by TDS UPS, which are Maximum, Minimum, Mean, Standard Deviation, Root Mean Square (RMS), Skewness and Kurtosis, and Mean Absolute Deviation (MAD). *Figure 1* shows a graphical representation of the features generated for each variable.

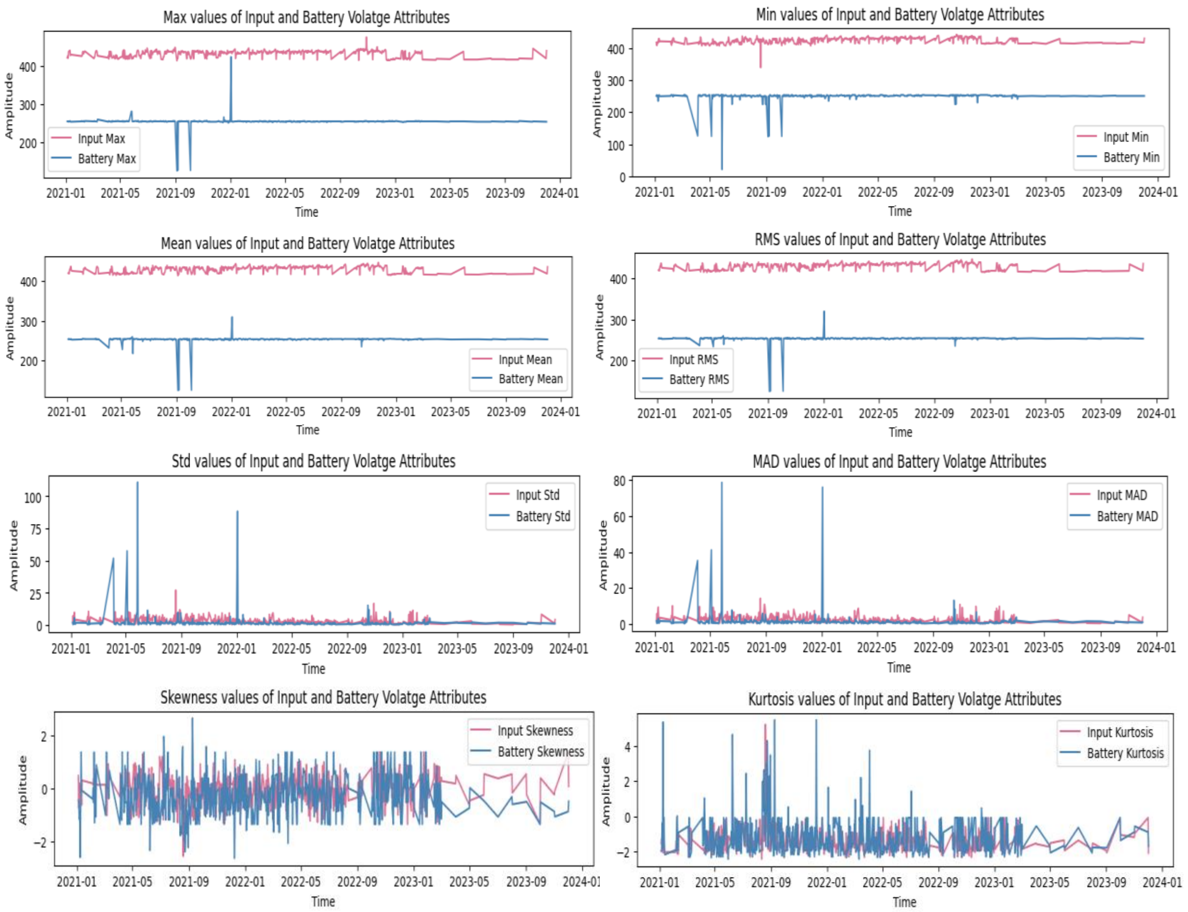

Figure 1. All features for both the Input and Battery Variables

International Journal of Computer Science Research (IJCSR)
Survey Article | Volume 3, Issue 1 | Pages 01-10 | ISSN XXXX-XXXX

4. ML MODELS DEVELOPED FAULT PREDICTION

Machine learning methods were used to process the input voltage and battery voltage statistical features. For the classification task including fault identification, the most popular and appropriate classifiers were used. In the modelling section, the machine learning models will be trained on the training subset and its performance will be tested against the unknown testing subset, resulting in confusion matrices and learning curves.

4.1 SVM Model for TDS UPS Fault Prediction

The SVM algorithms were designed with the following features: Mean, RMS, Maximum, and Minimum for input voltage data, and Minimum, Mean, RMS, and Kurtosis for battery voltage data. To boost model performance, the model's hyper-parameter parameters were fine-tuned. The radial basis function kernel was employed since it possessed the lowest margin of error and the most iterations (1,000,000). The model's performance is measured using multiple evaluation metrics such as Accuracy, F1-score, Recall, and Precision.

4.1.1 Model Evaluation

Table 3 and 4 illustrate the performance measures derived from the obtained SVM fault classification for the input voltage and battery voltage features, respectively. *Table 3 and 4* also demonstrate the performance for the training dataset and the validation or testing dataset, respectively.

Table 3: SVM Evaluation Metrics for the Input Voltage Attribute

Model	Training			
	Accuracy	F1-Score	Recall	Precision
SVM - Radial Kernel	0.9577	0.9211	0.8611	1.0000
	Testing			
	0.9624	0.9315	0.8718	1.0000

Table 4: SVM Evaluation Metrics for the Battery Voltage Attribute

Model	Training			
	Accuracy	F1-Score	Recall	Precision
SVM - Radial Kernel	0.9675	0.9140	0.8738	0.9732
	Testing			
	0.9474	0.8679	0.8214	0.9200

Table 3 shows that the radial basis function (rbf) kernel was calibrated and produced the smallest error margins having Accuracy, F1-Score, Recall, and Precision values of 0.9624, 0.9315, 0.8718, and 1.0000 for the input voltage dataset. Furthermore, the testing data Accuracy, F1-Score, Recall, and Precision scores for the battery voltage dataset presented in *table 4*. were slightly lower than those for the input voltage dataset.

4.1.2 Confusion Matrix

The confusion matrices for the input voltage and battery voltage datasets are shown in *figure 2 and 4.10*. The matrices show how many correct and wrong predictions the model produced for each class. The absolute numbers of classification success rates are recorded on the diagonals of the confusion matrices, while the misclassified samples are on the parts of the matrix based on the distribution of classification errors. For each confusion matrix, the predicted label represents the predicted value of the provided sample by the trained SVM algorithm, whereas the true label represents the desired value of that sample.

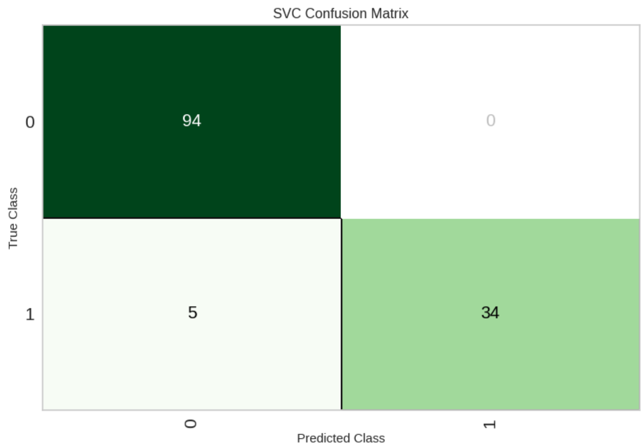

Figure 2: SVM Confusion Matrix for the Input Voltage Testing Data

In addition, *figure 3* shows that the SVM model accurately identified 103 fault free data points (True Positive) and misclassified 2 fault free data points as faults (False Negative) for the Battery Voltage dataset. Similarly, 23 True Negative means that the model accurately identified 23 fault data points while incorrectly classifying 5 fault data points as No fault data points (False Positive). Overall, the model appears to be performing reasonably well, however it generates quite a few false negatives for various classes.

4.1.3 Cross Validation

Cross validation is a technique for evaluating a machine learning model's performance by training it on multiple subsets of data and evaluating it on the remaining data. The models were validated and tested using the k-fold cross-validation approach. In this investigation, ten folds or a k value of ten were used. Using this cross-validation approach, the dataset is randomly divided into test and training data and then divided into k groups. The model is validated on one of the groups, then training is done on the remaining groups.

FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Computer Science Research (IJCSR)
Survey Article | Volume 3, Issue 1 | Pages 01-10 | ISSN XXXX-XXXX

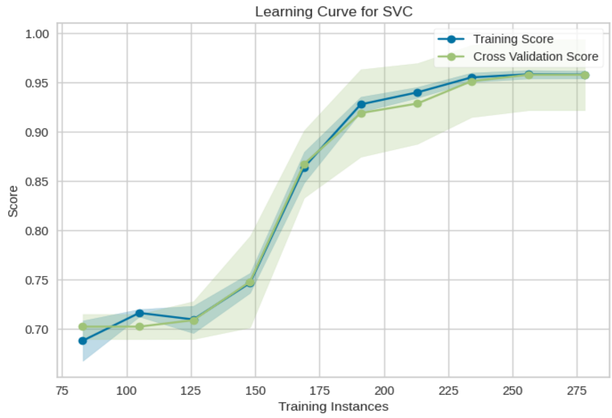

Figure 4: Plot of SVM Learning Curve for the Input Voltage Attribute

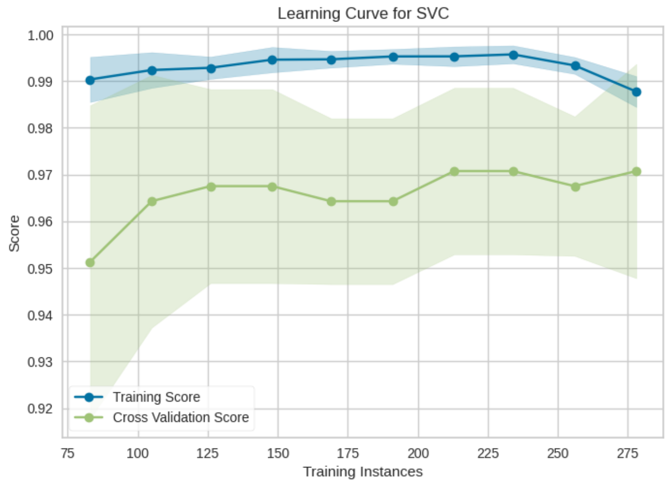

Figure 5: Plot of SVM Learning Curve for the Battery Voltage Attribute

Figure 4 and 12 compare a model's performance on training and testing data over a range of training instances. Figure 4 shows that the model is well-fitting, as indicated by training and validation scores that increase to a point of stability with a small difference between the two final score values. Figure 5 shows that the training score is exceptionally high regardless of training instances, and the cross-validation score grows over time. There is also a considerable variance between the training and testing scores, indicating that the SVM model is overfitting the data.

4.2 Decision Tree Model for TDS UPS Fault Prediction

The DT classification was carried out in order to categorize the data into binary targets and to construct a classification capable of correctly distinguishing between fault free and faulty data points. The criterion is the parameter that determines how the impurity of a split will be measured. The Gini impurity was the criterion parameter used. Also, "min_samples_split" parameter, which is the minimum number of samples required to split an internal node, was set to 20.

4.2.1 Model Evaluation

The classification performance results for the Decision Tree algorithm are summarized in *tables 5 and 6*. *Table 5* shows that the testing Accuracy, F1-Score, Recall, and Precision values for the input voltage dataset are 0.9774, 0.9620, 0.9744, and 0.9500, respectively. In addition, the battery voltage dataset has Accuracy, F1-Score, Recall, and Precision values of 0.9699, 0.9286, 0.9286, and 0.9286 in *table 6*.

Table 5: DT Evaluation Metrics for the Input Voltage Attribute

Model	Training			
	Accuracy	F1-Score	Recall	Precision
Decision Tree Classifier	0.9706	0.9486	0.9444	0.9633
	Testing			
	0.9774	0.9620	0.9744	0.9500

Table 6: DT Evaluation Metrics for the Battery Voltage Attribute

Model	Training			
	Accuracy	F1-Score	Recall	Precision
Decision Tree Classifier	0.9352	0.8175	0.7476	0.9399
	Testing			
	0.9699	0.9286	0.9286	0.9286

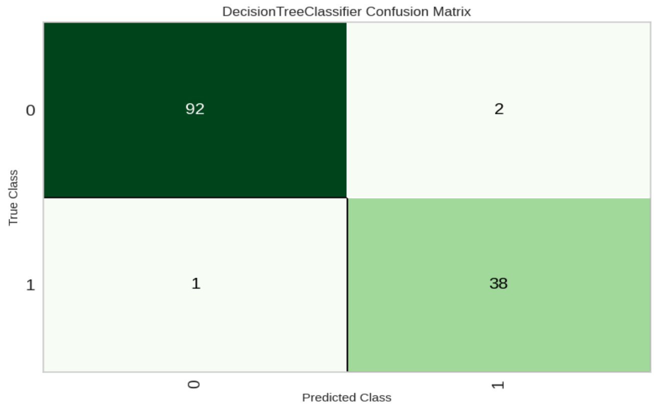

Figure 6: DT Confusion Matrix for the Input Voltage Attribute

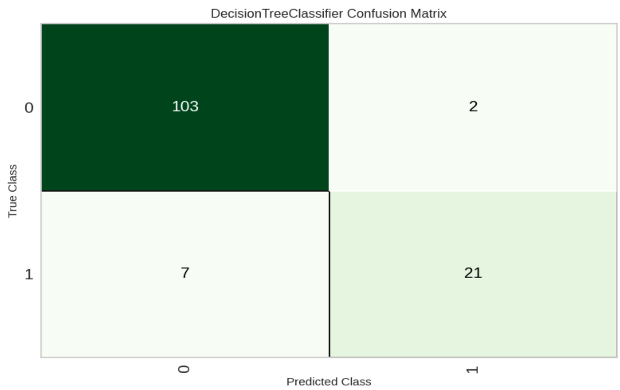

Figure 7: DT Confusion Matrix for the Battery Voltage Attribute

Figure 6 shows that the DT model made 92 correct predictions and 2 wrong predictions for the "No Failure" class. In addition, the model made 38 correct predictions and 1 incorrect prediction for the "Fault" class. Similarly, for the "No Failure" class, 102 data points were correctly classified and 2 were incorrectly classified in *figure 7* data points in the "Fault" class were accurately classified as faults, while 7 were wrongly labelled as No fault.

4.2.2 Cross Validation

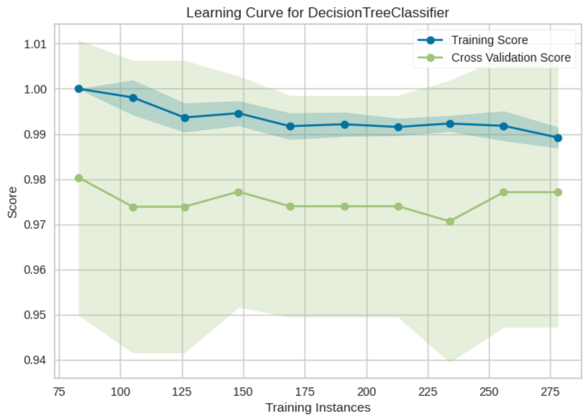

Figure 8: Plot of DT Learning Curve for the Input Voltage Attribute

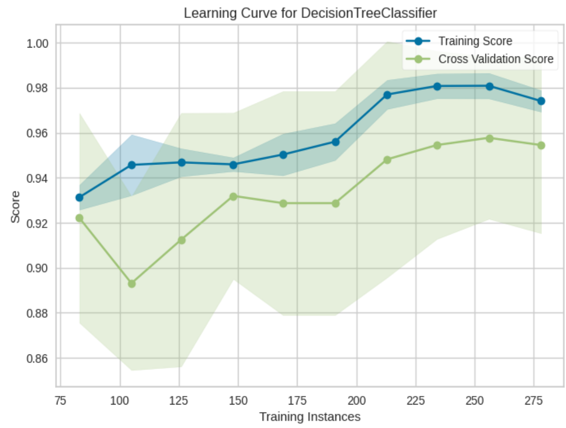

Figure 9: Plot of DT Learning Curve for the Battery Voltage Attribute

The learning curve in *figure 8* indicates substantial test variability and a high score up to around 280 instances, but after this threshold, the model begins to converge on an F1 score of around 0.98. Because the training and test results have not yet converged, this model could benefit from further training data. Finally, this model suffers mostly from error caused by variance (the test data scores are more variable than the training data), suggesting that the model may be overfitting.

Figure 9 also shows that the model has a very low training score at first, which steadily increases as more training examples are added. Both the training and testing scores begin to fall at 250 samples, showing that adding more examples will aid in the model's convergence and stability. The cross-validation graphs reveal the consistency and stability of the model's performance.

Table 7a Summary of comparison between ML models

Models	Input data			
	Accuracy	F1-Score	Recall	Precision
SVM	0.9624	0.9315	0.8718	1.0000
DT	0.9774	0.9620	0.9744	0.9500
	Battery Data			
SVM	0.9474	0.8679	0.8214	0.9200
DT	0.9699	0.9286	0.9286	0.9286

Table 7b Error Metrics between SVM and DT models

Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE
Input data						
SVM	0.42%	1.18%	4.45%	99.97%	0.036%	0.21%
DT	0.55%	2.25%	12.25%	98.23%	0.028%	0.31%
Battery data						
SVM	0.52%	2.28%	5.26%	90.57%	0.038%	0.51%
DT	1.24%	1.84%	6.07%	90.22%	0.13%	0.63%

The dependability of any developed model is determined by its capacity to generalize adequately based on test outcomes. As a result, a quantitative comparison of the proposed ML models was performed, as shown in *table 7*. Individual models were found to be capable of predicting faults in the TDS UPS. DT outperformed the SVM model in terms of Accuracy, F1-Score, Recall, and Precision.

In *table 7*, the DT-based prediction model had an Accuracy = 0.9774, F1-Score = 0.9620, Recall = 0.9744, and Precision = 0.9500 for the Input Voltage data, compared to SVM, which had an Accuracy = 0.9624, F1-Score = 0.9315, Recall = 0.8718, and Precision = 0.9500. The DT model's Accuracy (0.9774) rating indicates that the model is extremely dependable in forecasting failures in the TDS UPS system. Similarly, for the battery voltage data, DT demonstrated superior performance to the SVM model. The importance of input voltage and battery voltage-based variables to fault prediction was also discovered, with input voltage features better predictors for fault prediction.

5. CONCLUSION

This study introduced the notion of fault prediction and demonstrated how businesses may use it to enhance their maintenance cycles. The test result demonstrates that, with an accuracy of 97.74% for input voltage features and 96.99% for battery voltage features, the proposed decision tree's fault classification accuracy is superior to that of SVM. SVM and Decision tree were employed as trained classification models, and the recorded TDS UPS data from Ghana Gas was used in the ML modelling in order to create the prediction models, two variables (input voltage and battery voltage) were evaluated using data obtained from the TDS UPS during a three-year and three-month period. Furthermore, the eight statistical features were derived from both the input voltage and the battery voltage in order to further characterize the data. In pattern classification, the capabilities of machine learning algorithms were used. In addition, the major input parameters used by the models were the Mean, Min Max RMS, and Skewness. After modelling, each algorithm's performance for fault classification was examined. To assess the efficacy and capacities of the constructed models, four performance metrics were used: Accuracy, F-1 Score, Recall, and Precision. While the model performed well overall, it was discovered that it was unable to predict all classes with the same level of accuracy using battery voltage data. This implies that there might be opportunities to enhance the model's functionality.

6. RECOMMENDATIONS

The proposed method's robustness and accuracy demonstrated its potential for protecting UPS systems in major power companies. It provides companies additional support for equipment reliability decision-making, allowing them to be more competitive in the market. It is necessary to conduct additional study in order to forecast equipment failure times as well as to conduct real-time online calibration monitoring as data is being gathered. Also, the model enables an analysis of equipment data records, allowing the detection of faults without prior knowledge of the equipment's status. Furthermore, the model is generalizable to any number of UPS systems. However, to improve the model in the future, it advised to implement an online algorithm to diagnose and prognosis the equipment during an operation. An artificial neural network class like the Multi-Layer Perceptron (MLP) might be utilized to enhance the performance of the HMM.

REFERENCES

[1] Nan, B., Chen, L., Rodrigo, N.D., Borodin, O., Piao, N., Xia, J., Pollard, T., Hou, S., Zhang, J., Ji, X. and Xu, J., 2022. Enhancing Li+ transport in NMC811|| graphite lithium-ion batteries at low temperatures by using low-polarity-solvent electrolytes. Angewandte Chemie International Edition, 61(35), p.e202205967.

[2] Lukovic, M., Lukovic, V., Belca, I., Kasalica, B., Stanimirovic, I. and Vicic, M., 2016. LED-based Vis-NIR spectrally tunable light source-the optimization algorithm. Journal of the European Optical Society-Rapid Publications, 12, pp. 1-12.

[3] Guerrero, J.M., Matas, J., de Vicuna, L.G., Castilla, M. and Miret, J., 2007. Decentralized control for parallel operation of distributed generation inverters using resistive output impedance. IEEE Transactions on industrial electronics, 54(2), pp. 994-1004.

[4] Low, K.S. and Cao, R., 2008. Model predictive control of parallel-connected inverters for uninterruptible power supplies. IEEE Transactions on Industrial Electronics, 55(8), pp. 2884-2893.

[5] Martin, F. and Aguado, J.A., 2003. Wavelet-based ANN approach for transmission line protection. IEEE Transactions on power delivery, 18(4), pp.1572-1574.

[6] Sharafi, A., Jafarian, P. and Sanaye-Pasand, M., 2010, May. A combined algorithm for high speed transmission-line protection based on traveling-wave. In 2010 9th International Conference on Environment and Electrical Engineering, pp. 132-135).

[7] Park T, Efros AA, Zhang R, Zhu JY. Contrastive learning for unpaired image-to-image translation. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Springer International Publishing, Part IX 16 2020, pp. 319-345.

[8] Katipamula, S. and Brambley, M.R., 2005. Methods for fault detection, diagnostics, and prognostics for building systems—A review, part II. Hvac&R Research, 11(2), pp.169-187.

[9] Awadallah, M.A. and Morcos, M.M., 2003. Application of AI tools in fault diagnosis of electrical machines and drives-an overview. IEEE Transactions on energy conversion, 18(2), pp. 245-251.

[10] Giuntini, L. and Brioschi, M., 2015, March. Markov chain analysis for failure prediction of power converters. In 2015 IEEE International Conference on Industrial Technology (ICIT), pp. 1336-1341). IEEE.

[11] Eusgeld I, Fraikin F, Rohr M, Salfner F, Wappler U. Software reliability. Dependability Metrics: Advanced Lectures. 2008:104-25.

[12] Salfner, F. and Tschirpke, S., 2008. Error Log Processing for Accurate Failure Prediction. WASL, 8, pp.4.

[13] Zhang, X., Zhou, X., Lin, M. and Sun, J., 2018. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6848-6856).

[14] Tao L, Cinquanta E, Chiappe D, Grazianetti C, Fanciulli M, Dubey M, Molle A, Akinwande D. Silicene field-effect transistors operating at room temperature. Nature nanotechnology. 2015 Mar;10(3): pp. 227-31.

[15] Guardado, J.L., Naredo, J.L., Moreno, P. and Fuerte, C.R., 2001. A comparative study of neural network efficiency in power transformers diagnosis using dissolved gas analysis. IEEE Transactions on Power delivery, 16(4), pp. 643-647.

[16] Eon, C., Breadsell, J., Morrison, G. and Byrne, J., 2019. Shifting home energy consumption through a holistic understanding of the home system of practice. Decarbonising the built environment: Charting the transition, pp. 431-447.

[17] Giuntini, L. and Brioschi, M., 2015, March. Markov chain analysis for failure prediction of power converters. In 2015 IEEE International Conference on Industrial Technology (ICIT), pp. 1336-1341, IEEE.

[18] Jan, S.U., Lee, Y.D., Shin, J. and Koo, I., 2017. Sensor fault classification based on support vector machine and statistical time-domain features. IEEE Access, 5, pp. 8682-8690.

[19] Bigdeli, M., Vakilian, M. and Rahimpour, E., 2012. Transformer winding faults classification based on transfer function analysis by support vector machine. IET electric power applications, 6(5), pp. 268-276.

[20] Linzen, D., Buller, S., Karden, E. and De Doncker, R.W., 2005. Analysis and evaluation of charge-balancing circuits on performance, reliability, and lifetime of supercapacitor systems. IEEE transactions on industry applications, 41(5), pp. 1135-1141.

[21] Litchfield, N.J., Villamor, P., Dissen, R.J.V., Nicol, A., Barnes, P.M., A. Barrell, D.J., Pettinga, J.R., Langridge, R.M., Little, T.A., Mountjoy, J.J. and Ries, W.F., 2018. Surface rupture of multiple crustal faults in the 2016 M w 7.8 Kaikōura, New Zealand, Earthquake. Bulletin of the Seismological Society of America, 108(3B), pp. 1496-1520.

[22] Lind, H. and Muyingo, H., 2012. Building maintenance strategies: planning under uncertainty, Property Management, 30(1), pp. 14-28.

[23] Aamir, M., Kalwar, K.A. and Mekhilef, S., 2016. Uninterruptible power supply (UPS) system. Renewable and sustainable energy reviews, 58, pp. 1395-1410.

[24] Cherkassky V, Mulier FM. Learning from data: concepts, theory, and methods. John Wiley & Sons; 2007 Sep 10.

[25] Suykens JA. Nonlinear modelling and support vector machines. InIMTC 2001. proceedings of the 18th IEEE instrumentation and measurement technology conference. Rediscovering measurement in the age of informatics (Cat. No. 01CH 37188) 2001 May 21 (Vol. 1, pp. 287-294). IEEE.

[26] Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de-Pison FJ, Antonanzas-Torres F. Review of photovoltaic power forecasting. Sol Energy. 2016;136: pp. 78-111.

[27] Sharma G, Tripathi V, Mahajan M, Srivastava AK. Comparative analysis of supervised models for diamond price prediction. In: Proceedings of the Conflu 2021 11th Int Conf Cloud Comput Data Sci Eng. Published online 2021: pp. 1019-1022.

[28] Fontana F A, Mäntylä MV, Zanoni M, Marino A. Comparing and experimenting machine learning techniques for code smell detection. Empir Softw Eng Vol. 2016;21: pp. 1143–1191.

[29] Muhammad IA , Fabio P, Lin S, Qing W. Machine learning techniques for code smell detection: A systematic literature review and meta-analysis. Inf Softw Technol. 2019;108 (1), pp. 15-138.

[30] Ofosu A. R, Odoi B, Asamoah M, “Electricity consumption forecast for Tarkwa using autoregressive integrated moving average and adaptive neuro fuzzy inference system”, Serbian Electr Eng. 2021;18(1): pp. 75-94.

[31] Twumasi-Ankrah, S., Odoi, B., Adoma Pels, W. and Gyamfi, E.H., 2019. Efficiency of imputation techniques in univariate time series.

[32] Ofosu R. A, Zhu H, Odoi B. A Hybrid Prediction Fault Location Model for Copper Wire Manufacturing Process. Acta Polytechnica Hungarica. 2024; 21(6).

[33] Odoi B., Boahen A. D. and Brew L. (2024), Predicting Nitrous Oxide Emissions in Ghana using Long Short-term Memory and Gated Recurrent Neural Network, eNergetics 2023, 9 th Virtual International Conference on Science, Technology and Management in Energy, Conference, 81

[34] Rathakrishnan V, Bt. Beddu S, Ahmed AN. Predicting compressive strength of high-performance concrete with high volume ground granulated blast-furnace slag replacement using a boosting machine, learning algorithms. Sci Rep. 2022;12(1): pp. 1-16.

[35] Liu Q, Wang X, Huang X, Yin X. Prediction model of rock mass class using classification and regression tree integrated AdaBoost algorithm based on TBM driving data. Tunn Undergr Sp Technol. 2020; 106: pp. 1-15.

[36] Nguyen H, Vu T, Vo TP, Thai HT. Efficient machine learning models for prediction of concrete strengths. Constr Build Mater. 2021; 266: pp.1-17.

[37] Saucedo-Dorantes J.J, Arellano-Espitia F, Delgado-Prieto M, Osornio-Rios RA. Diagnosis methodology based on deep feature learning for fault identification in metallic, hybrid and ceramic bearings. Sensors. 2021 Aug 30;21(17):5832.