Classification of the Supply Voltage conditions of a Three-Phase Induction Motor with Machine Learning techniques

Marcelo E. da Silva * **Manoel E. N. Oliveira** * **Felipe B. M. Nunes** *
Luiz Antonio G. Santos * **Arthur M. do Nascimento** *
Edênio Z. G. Costa *

* *Electrical Engineering, Sobral Campus, Federal University of Ceara, CE, (e-mail: marcelo.estevao@alu.ufc.br, manoeleric59@ieee.org, felipebrunomn@gmail.com, luiz.santos@ieee.org, arthurmnphb@gmail.com, edeniogalvao@alu.ufc.br).*

Abstract: Three-Phase Induction Motors (TPIM) is a fundamental part, as they are the main responsible for carrying out the mechanical work process in the industry. It is estimated that they are responsible for consuming more than half of all energy destined for the industrial sector. Thus, any failure of operation in motors of this type is reflected in energy, economic and environmental losses. Among the most common failures is the unbalance of the supply voltages, which can cause total loss of the machine depending on the magnitude of the unbalance. This article addresses a comparative analysis between the Machine Learning K-Nearest Neighbors (KNN), Random Forest (RF), Suport Vector Machine (SVM), Principal Component Analysis (PCA) and Multilayer Perceptron Neural Network (MLP) techniques applied to the classification of unbalanced supply voltages of a three-phase induction motor. For this, a database was used with mechanical and electrical variables related to the balanced and unbalanced operation of the motor, divided into classes of different levels of unbalance according to the National Electrical Manufactores Association (NEMA).

Keywords: Fault Diagnostic; Machine Learning; Pattern Recognition; Three-Phase Induction Motor; Unbalanced Supply Voltages.

1. INTRODUCTION

The Three-Phase Induction Motor (TPIM) is a rotating electrical machine designed to perform the process of converting electrical energy into mechanical energy, being widely used in the industrial sector. In this process of energy conversion, TPIM uses rotating magnetic fields.

It is estimated that TPIMs are responsible for consuming more than 70% of all energy destined for industrial plants and for more than 40% of a country's total energy consumption (Nascimento et al., 2020). The widespread use of these machines in the industry is justified not only by their robustness, but also by their simplicity and low maintenance cost (Santos et al., 2015).

In this way, it is evident that any failures in machines of this type will reflect in energy losses, due to the drop in performance; economic, by spending wasted energy; and environmental, due to the natural resources used in the generation of unused energy. As a consequence of unbalanced voltages, a TPIM may exhibit sequence current and pulsating torque sequence, which causes an increase in the winding temperature and motor losses (Adekitan and Abdulkareem, 2019). In addition, the efficiency of the motor is also impaired due to increased losses and harmonics of current, also reducing the life of the machine (Adekitan and Abdulkareem, 2019).

In general, the voltage unbalance can be found at any level, with the possibility of exceptions for levels below 2%. In most cases, the greatest voltage unbalance appears in the electrical installations of the final consumer. It is estimated that 98% of concessionaires have less than 3% unbalance, with more than half between 0 and 1% unbalance (Fernando Mantilla, 2008).

In Adekitan and Abdulkareem (2019), a comparison was made between the performance of the Tree Emsemble (TE), Decision Tree (DT), Random Forest (RF) and Support Vector Machine (SVM) techniques applied in the classification of under voltage (2-10%), rated voltage and over voltage (2-10%)

In Bazan et al. (2019), an unbalanced voltage diagnosis was carried out, along with load torque variations and short-circuit levels, in the first stage C4.5 DT was used and in the second stage, it used Multilayer Perpectron Neural Networks (MLP).

In addition, in Sawitri et al. (2013), the detection of unbalanced supply voltages was performed using SVM with the extraction of features from the Wavelet transform and the Principal Component Analysis (PCA) algorithm.

Therefore, the purpose of this work is to perform a comparative analysis between the techniques of K-Nearest Neighbors (KNN), Random Forest (RF), Suport Vector

Machine (SVM), Principal Component Analysis (PCA) and Multilayer Perceptron Neural Network (MLP) in order to find out which technique has the highest hit rate for the problem of unbalanced power supply voltages in TPIM. For this, a stratified database strategy is used for a more detailed analysis with subsequent reduction of dimensionality with PCA and based on the importance of each feature for the model.

The differential of this work is the application of the techniques mentioned in a database different from the one used by previous works, obtained in Adekitan et al. (2019). In addition, the use of Machine Learning techniques does not require the use of traditional features for the calculation of unbalance of supply voltages, being therefore, an alternative for cases in which the features of traditional methods are unavailable.

The structure of this work is divided into four more sections. Section 2 presents the theoretical basis for the problem of unbalanced supply voltages. Section 3 describes the methodology adopted. In section 4, the results obtained in section 3 are shown. And finally, in section 5 is the conclusion of the work and indications of future work.

2. THEORETICAL BACKGROUND

2.1 Voltage Unbalance

The balanced voltage is characterized by having the same magnitude value and a difference of 120° between the phases, when this condition is not ture the voltage is characterized as unbalanced. When there is an unbalance in the supply voltage, a TPIM can present some problems such as decreased performance and reduced machine life (Alham et al., 2020).

The most frequent causes of unbalanced voltages in a TPIM occur due to unstable power, single-phase loads distributed in the same energy system unevenly, an open circuit in the primary distribution system and atmospheric discharges in distribution circuits (Araújo et al., 2020). Unbalances considered small can reflect a great unbalance in the current of the TPIM, which can result in an increase in temperature and, thus, compromise the isolation of the TPIM (Alham et al., 2020).

2.2 Measuring the Voltage Unbalance

It is possible to measure the level of unbalanced supply voltages through three definitions: the definition of the National Electrical Manufacturer Association (NEMA), the definition of the Institute of Electrical and Electronics Engineers (IEEE) and the definition of International Electrotechnical Commission (IEC) (Refaat and Abu-Rub, 2015).

The NEMA definition says that the line voltage unbalance rate (LVUR) is the maximum deviation from the average line voltage in relation to the average value of the line voltages. This definition takes into account only the magnitudes or modules of the voltages, in this case, the phase angles are not considered (Refaat and Abu-Rub, 2015)

$$LVUR(\%) = \frac{Max\ volt\ deviation\ from\ avg\ line\ volt}{Average\ Line\ Voltage} \times 100 \quad (1)$$

The definition given by the IEEE says that a rate of phase voltage unbalance (PVUR) is given by the maximum deviation from the average phase voltage related to the average value of the voltages of the three phases (Refaat and Abu-Rub, 2015).

$$PVUR(\%) = \frac{Max\ volt\ deviation\ from\ avg\ phase\ volt}{Average\ Phase\ Voltage} \times 100 \quad (2)$$

The definition given by the IEC says that the voltage unbalance factor (VUF) is the ratio between the negative sequence voltage component and the positive sequence voltage component (Refaat and Abu-Rub, 2015).

$$VUF(\%) = \frac{Negative\ Sequence\ Voltage\ Component}{Positive\ Sequence\ Voltage\ Component} \times 100 \quad (3)$$

Normally, the effect of voltage unbalance in a TPIM according to NEMA is reflected in the behavior of the negative sequence voltage, in this case, the rotation of the TPIM occurs in the opposite direction to that of the voltage unbalance. Therefore, positive and negative sequence voltages can be used to analyze the behavior of a TPIM in an unbalanced condition (Sawitri et al., 2013).

Thus, the voltage of the positive (V_{sp}) and negative (V_{sn}) sequence components are obtained from the unbalance of each phase (V_{ab}, V_{bc} and V_{ca}). For the balanced condition, we have (Sawitri et al., 2013):

$$V_{sp} = \frac{V_{ab} + a' \times V_{bc} + a'' \times V_{ca}}{3} \quad (4)$$

$$V_{sn} = \frac{V_{ab} + a'' \times V_{bc} + a' \times V_{ca}}{3} \quad (5)$$

Where:

$$a' = 1\angle 120^\circ \quad and \quad a'' = 1\angle 240^\circ \quad (6)$$

2.3 Machine Learning Techniques

The techniques of KNN, SVM, RF, MLP and PCA are well known and used in the classification of phenomena of one or more classes. Thus, the PCA technique can be used to reduce the dimensionality of the data and the criterion for excluding features due to their importance for the classification can give greater accuracy to the problem to be solved.

The KNN is an algorithm that implements the classification of unknown samples based on the distance between the unknown samples and the samples associated with the problem classes. For this, it is necessary to define the number of k neighbors, the metric of the distance calculation and the decision rule (Soares et al., 2020).

In addition, SVM aims to find the hyperparameter of separation between the classes of the problem, in order to find the maximum distance between the closest samples (Soares et al., 2020).

Furthermore, RF can be described as a combination of tree predictors so that each tree is dependent on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Breiman, 2001).

On the other hand, MLP is a neural network used to solve non-linearly separable problems. For this, the network uses one or more intermediate layers of neurons and an output layer normally fully connected (Faceli et al., 2011).

In addition, the PCA transforms high-dimensional data into a low-dimensional subspace component and a noise component. This decomposition is quite useful for data compression and noise elimination tasks, thus being an important step for many data processing tasks (Minka, 2000).

Finally, feature importance applied in DT is based on the average reduction of impurity. Feature importances are calculated as the mean and standard deviation of accumulation of the reduction of impurities within each tree (Cassidy and Deviney Jr, 2014).

3. METHODOLOGY

A methodology used in this procecdure follow a steps sequential of according with a specific database characteristics . How data contains one unique dataframe with all six conditions of unbalance, three steps was necessary for identification of failures on the TPIM.

In the first stage, database was apportion in five new dataframes contains the five levels of unbalance in five dataframes, where each dataframe contais each condition of unbalance concatened with balance condition (0%). The intention of this step is to apply in each new base several classification techniques to select the four best algorithms for the next stage.

In the second stage, the dataframes used before, are concatened resulting on unique dataframe with a six labels, in the other words, the new dataframe is a multiclass database.

In the last stage, some hypotheses are applied to try to improve the performance of the best algorithms. The hypotheses are the reduction of size by two techniques explained later.

3.1 Dataset

In this work was used a dataset that contains several scenarios of voltage change of a Three-Phase Induction Motor (TPIM), to obtain the variations of the motor operational parameters for 6 unbalancing conditions (0%, 1%, 2%, 3%, 4% e 5%). The conditions mentioned before vary independently to obtain the 5% unbalance according to the National Electrical Manufacturers Association (NEMA) (Pillay and Manyage, 2001).

In the experimental procedure to obtain the data, was used one TPIM of 415 V with the specifications, $Xm = 7.9636\Omega$, $Xs = 0.3965\Omega$, $Rr = 0.2775\Omega$ and $Rs = 0.2412\Omega$. The data presented here, in terms of slip refers to the interval where, $-1 \leq \text{slip} \leq 2$, thus ensuring all conditions for slip acordding NEMA . With this the specific data base the obtains electrical and mechanical data of the motor how, rotor current, stator current, winding copper losses, real input power, reactive input power the apparent power, and air gap power, torque and electromechanical power (Adekitan et al., 2019).

3.2 Stage 1

As the database used here initially consisted of a single CSV file containing the motor unbalance data on each sheet (0 %, 1 %, 2 %, 3 %, 4 %, 5 %), it was necessary to subdivide the database into 5 new files each containing an unbalance condition (1 %, 2 %, 3 %, 4%, 5 %) concatened with the efficiency condition (0 % unbalance) and add a label column for each condition , 0 for balancing and 1 for unbalance. Thus, the initial problem was divided into 5 new binary variable problems. The first stage of the method can be seen in the Figure 1.

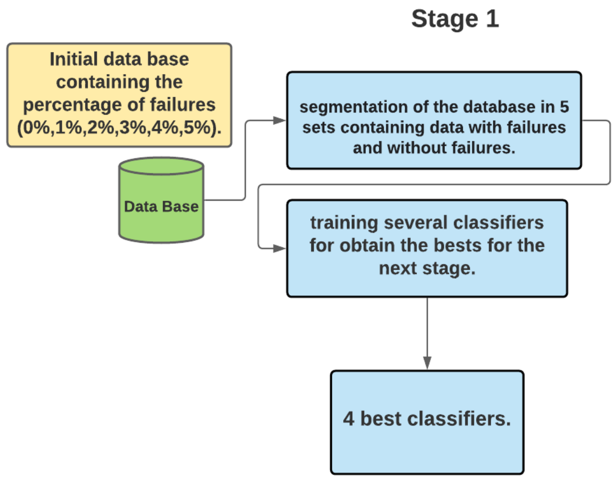

Figure 1. Methodology of Stage 1

As showed in the previous figure, the intention of this step is the application of several classification techniques in machine learning using the Python Scikit Learning library, to obtain the 4 best classifiers. This step acts as a selection filter for the best models. This method is based on the model proposed in (Tanwani et al., 2009),where by applying several algorithms in a generic way in a database, it is possible to obtain the tendency of which may be the best algorithm for the next stage of the methodology. The classifying algorithms were selected from some evaluation metrics evidenced in the next section.

3.3 Evaluation metrics

For to evaluate the classifiers, was used the some metrics based on confusion matrix for each dataframe. They were, accuracy, precision and recall.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$precision = \frac{TP}{TP + FP} \tag{8}$$

$$recall = \frac{TP}{TP + FN} \tag{9}$$

where, TP, TN, FN and FP stand for True-Positive, True Negative, False Negative and False Positive. In addition for the F1-Score was used:

$$F1\text{-}Score = \frac{2(precision)(recall)}{(precision) + (recall)} \tag{10}$$

3.4 Stage 2

With the procedure of the previous stage, a select group of classifiers algorithms was obtained . However, the purpose of this stage is to join the previously segmented database to a single dataframe containing all the unbalance data. Thus, the four best algorithms from the previous stage will be used to address the new multiclass problem, with the aim of analyzing the performance of the fault identification task in TPIM. The Figure 2, show the diagram of the second stage.

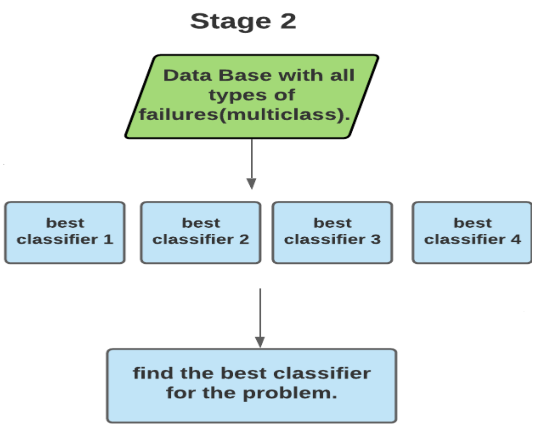

Figure 2. Methodology of Stage 2

3.5 Stage 3

In the last stage, as the methodology addressed so far is sequential, the results obtained in the second stage were used to finally select the best classification algorithm, evaluating again based on the evaluation metrics of stage 1 and stage 2.

In addition, the hypothesis that irrelevant or related data existed led to the application of two dimensionality reduction techniques: Principal Component Analysis (PCA) and Feature Importance.

The figure 3, denotes the flowchart of the last stage of the methodology.

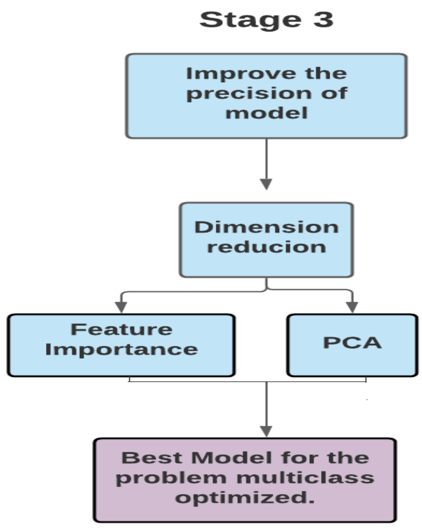

Figure 3. Methodology of Stage 3

4. RESULTS

In this section, is presented at results obtained according to the theory proposed in the methodology stages. Thus, as each stage has a specific objective, they will be divided into 3 sections for results.

4.1 Results for First Stage

The proposal of the first stage, was to select the best classifiers in the segmented databases in binary problems, containing the unbalance of the engine from 1 % to 5 %. Thus, the following tables present the 4 best algorithms for each unbalance dataframe.

Table 1. Results for dataframe with 1% unbalance

Classifier	Precision	Recall	F1-Score	Accuracy
Decision Tree	94,5 %	98,2 %	96,3 %	99,6%
Quadratic SVM	93,8 %	95,6 %	94,7 %	99,5%
KNN	92,1 %	90,0 %	91,1 %	98,3%
Bagged Trees	97,5 %	99,3 %	98,4 %	99,9%

Table 2. Results for dataframe with 2% unbalance

Classifier	Precision	Recall	F1-Score	Accuracy
Decision Tree	96,5 %	98,9 %	97,4 %	99,6%
Quadratic SVM	95,8 %	94,7 %	94,7 %	99,6%
KNN	92,1 %	95,1 %	92,6 %	99,2%
Bagged Trees	98,5 %	99,1 %	98,8 %	99,6%

Table 3. Results for dataframe with 3% unbalance

Classifier	Precision	Recall	F1-Score	Accuracy
Decision Tree	96,5 %	97,9 %	93,4 %	99,4%
Quadratic SVM	94,8 %	95,7 %	92,7 %	99,7%
KNN	92,1 %	95,1 %	92,6 %	99,6%
Bagged Trees	98,5 %	98,1 %	98,1 %	99,8%

Table 4. Results for dataframe with 4% unbalance

Classifier	Precision	Recall	F1-Score	Accuracy
Decision Tree	96,5 %	97,9 %	93,4 %	99,6 %
Quadratic SVM	94,8 %	95,7 %	92,7 %	99,6 %
Subspace Discriminant	95,2 %	95,1 %	94,6 %	98,7 %
Bagged Trees	98,6 %	98,1 %	98,2 %	99,8 %

Table 5. Results for dataframe with 5% unbalance

Classifier	Precision	Recall	F1-Score	Accuracy
Decision Tree	96,2 %	96,9 %	95,4 %	99,6 %
Quadratic SVM	96,8 %	95,8 %	92,8 %	99,6 %
Subspace Discriminant	94,2 %	95,2 %	94,1 %	98,2 %
Bagged Trees	98,6 %	98,1 %	98,2 %	99,8 %

As expected, good results were obtained in view of the initial data provision taking into account only one decimal place (1 %, 2 %, 3 %, 4 % and 5 %) concatenated with the effective case (0 % unbalance).

4.2 Results for the Second stage

As in the methodology, the second stage aims to apply the 4 bests classification techniques in the base containing all types of unbalance and select the best hyperparameters for them. Thus, the following tables shows the best results and the worst results for: Decision Tree, Quadratic SVM, KNN and Bagged Trees.

Table 6. Multiclass problem with all unbalances of TPIM. (Best results)

Classifier	Precision	Recall	F1-Score	Accuracy
Decision Tree (Best result)	92,2 %	94,9 %	91,4 %	95,1 %
Quadratic SVM (Best result)	94,8 %	95,8 %	93,8 %	96,6 %
KNN (Best result)	89,2 %	82,2 %	85,1 %	89,2 %
Bagged Trees (Best result)	98,6 %	95,1 %	97,2 %	98,8 %

Table 7. Multiclass problem with all unbalances of TPIM. (Worse results)

Classifier	Precision	Recall	F1-Score	Accuracy
Decision Tree (Worse result)	87,2 %	91,9 %	89,4 %	91,1 %
Quadratic SVM (Worse result)	90,8 %	94,8 %	93,8 %	93,6 %
KNN (Worse result)	82,2 %	81,2 %	83,1 %	84,2 %
Bagged Trees (Worse result)	94,6 %	91,1 %	93,2 %	94,8 %

With this sequence of tests, after the variation of the hyperparameters, the best classifier was obtained in the multiclass problem. The best classifier for this database was Random Forest with the "bootstrap" agregation method. The best hyperparameters are: 'bootstrap' aggregation method, a type of classifier like the "Decision Tree" with 30 learners. The following figures show the confusion matrix's of the best and worst case respectively.

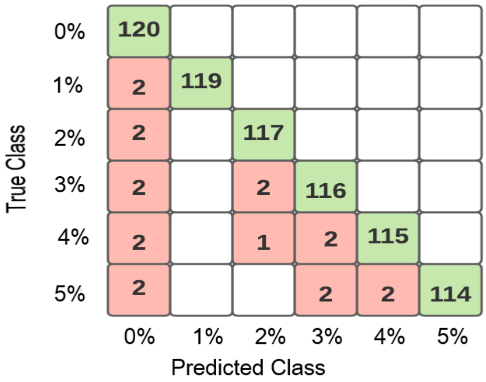

Figure 4. Confusion Matrix for the Bagged Trees in your best case

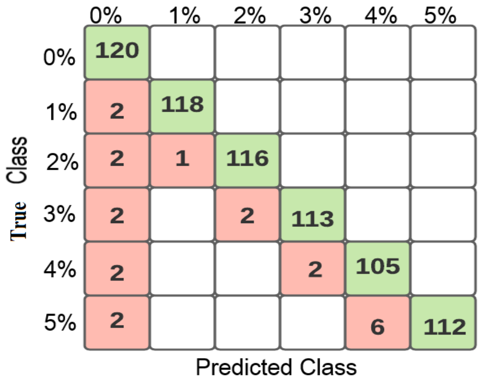

Figure 5. Confusion Matrix for the Bagged Trees in your worse case

4.3 Results for last Stage

According to stage 3, the hypothesis of improving the model by reducing the size of the data was applied using PCA and Feature Importance.

4.4 PCA results

The Figure 6, presents the data of accuracy and precision with reduction by PCA

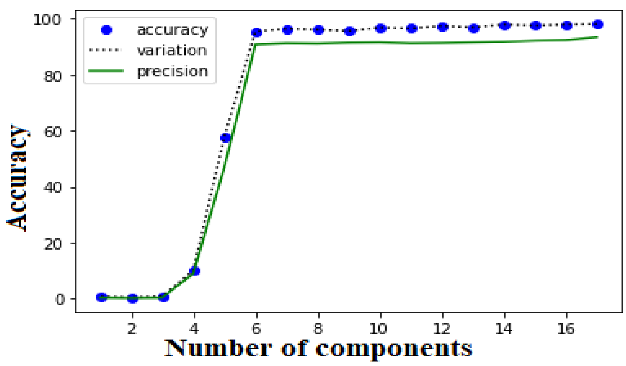

Figure 6. Reducion with PCA, results

As showed in the graph, reducing the dimension of data is advantageous until the point of drop of the graph, 6 components. However, with the previous results, the accuracy and precision decreases, however, there is a gain in training time.

4.5 *Feature Importance results*

The Figure 7 demonstrates the results obtained in the accuracy and precision with the dimension reduction embroidering using Feature Importance.

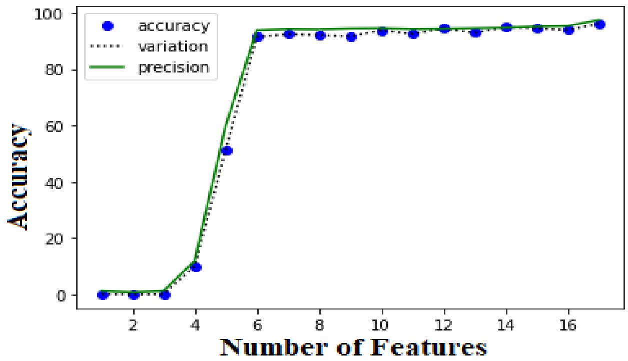

Figure 7. Reducion with Feature Importance, results

In the figure it can be seen that with the use of the Feature Importance function, there is a drop in the accuracy value. However, there is an increase in the model's precision value. Thus, with the objective of the two techniques approached, it is to improve the model, that is, to make it more efficient in a short time, the best option for reducing the dimension for this database, with the voltage unbalance data, is reduction with Feature Importance.

5. CONCLUSION

In this work presented the application/selection of machine learning techniques, applied in the detection of failures in TPIM. In addition, the method applied in the work constitutes a theorical background framework for future work, that will analyze the similar database, as in the detection of failures analytically it is necessary data on the electric current in the sectors of the motor.

The results showed that in the analysis of the unbalance of a TPIM under voltage unbalance conditions, the Random Forest algorithm with bagged tree algorithm with the possibility of improving the precision and the training time with the reduction of dimensionality with the PCA and Feature Importance techniques. In addition, it was defined that reducing the dimension using the Feature Importance function is more advantageous than using the reduction by the PCA that is commonly used in the area of computational intelligence.

REFERENCES

Adekitan, A.I. and Abdulkareem, A. (2019). Prediction of the voltage status of a three-phase induction motor using data mining algorithms. *SN Applied Sciences*, 1.

Adekitan, A.I., Samuel, I., and Amuta, E. (2019). Dataset on the performance of a three phase induction motor under balanced and unbalanced supply voltage conditions. *Data in brief*, 24.

Alham, N.R., Utomo, R.M., Asfani, D.A., Muslimin, and Dewantara, B.Y. (2020). Detection of unbalanced voltage supply and static eccentricity on three-phase induction motor using discrete wavelet transform. *2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 269–274.

Araújo, A.C.S., Lamim Filho, P.C.M., Baccarini, L.M.R., and Batista, F.B. (2020). Detecção de desequilíbrio de tensão e de desalinhamento no motor de indução através da análise do fluxo de dispersão. *Congresso Brasileiro de Automática*, 2.

Bazan, G.H., Scalassara, P.R., Endo, F., Goedtel, A., Palácios, R.H.C., and Godoy, W.F. (2019). Stator short-circuit diagnosis in induction motors using mutual information and intelligent systems. *IEEE Transactions on Industrial Electronics*, 66.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.

Cassidy, A.P. and Deviney Jr, F.A. (2014). Calculating feature importance in data streams with concept drift using online random forest. *2014 IEEE International Conference on Big Data (Big Data)*.

Faceli, K., Lorena, A.C., Gama, J., and Carvalho, A.C.P.L.F. (2011). *Inteligência Artificial: Uma Abordagem de Aprendizagem de Máquina*. LTC - Livros Técnicos e Científicos Editora Ltda., Rio de Janeiro.

Fernando Mantilla, L. (2008). Analysis of the voltage phasors characteristics for motor unbalanced supplies under constant voltage level. *Electrical Engineering*, 90.

Minka, T. (2000). Automatic choice of dimensionality for pca. *Advances in Neural Information Processing Systems*, 13.

Nascimento, A.d.A.R., Moreira, J.d.S., and Zaparoli, I.O. (2020). Detecção de falhas de rolamentos em motores de indução utilizando análise do envelope de corrente. *Congresso Brasileiro de Automática*, 2.

Pillay, P. and Manyage, M. (2001). Definitions of voltage unbalance. *IEEE Power Engineering Review*, 21(5), 50–51.

Refaat, S.S. and Abu-Rub, H. (2015). Ann-based diagnosis of incipient stator winding turn faults for three-phase induction motors in the presence of unbalanced supply voltage. *IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society*, 5328–5334.

Santos, V.S., Felipe, P.R.V., Sarduy, J.R.G., Lemozy, N.A., Jurado, A., and Quispe, E.C. (2015). Procedure for determining induction motor efficiency working under distorted grid voltages. *IEEE Transactions on Energy Conversion*, 30.

Sawitri, D.R., Asfani, D.A., Purnomo, M.H., Purnama, K.E., and Ashari, M. (2013). Early detection of unbalance voltage in three phase induction motor based on svm. *IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED)*, 573–578.

Soares, L.O., Pinto, L.A., and Lobão, D.A. (2020). Aplicaçao de machine learning e descritores estatısticos para diagnostico de falhas em rolamentos de motores eletricos industriais. *Congresso Brasileiro de Automática*, 2.

Tanwani, A.K., Afridi, J., Shafiq, M.Z., and Farooq, M. (2009). Guidelines to select machine learning scheme for classification of biomedical datasets. In *European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, 128–139. Springer.

ISSN: 2175-8905

DOI: 10.20906/sbai.v1i1.2645