ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 8, Issue 2, February 2021

Power Quality Event Classification using Machine Learning Techniques

[1] Challa Sairam Reddy, [2] Divya R, [3] Dr. Manjula G Nair

[1][2][3] Department of Electrical and Electronics Engineering, Amrita vishwa vidyapeetham, Amritapuri, India

Abstract: The rapid transformation of the conventional electric power grid to the sophisticated smart grid involves many challenging factors like power quality and reliability issues. Smart grid is nothing but introduction of information and communication technology to the present existing grid, so the process of realizing smart grids involves various kinds of devices which results in degradation of power quality. As power quality is the promising factor for utilities as well as the end users for proper functioning of equipment, it should be maintained high and with in specified limits. Thus detection and classification of different power quality events and their causes must be known prior to take appropriate mitigating actions. The task of classification involves storing of huge amount of data for analysis which became feasible with increased databases. The power quality events like sag, swell, interruption and unbalance are classified with data mining algorithms like SVM, KNN, Random forest with supervised machine learning techniques using python software.

Keywords: Power quality, Machine Learning, classification algorithms

1. INTRODUCTION

Power quality (PQ) is the concerning factor which deals with the set of electrical standards required by any electrical appliances to function in its intended manufacturing efficiency and life expectancy. The quality of power can be considered as high when its constituents like voltage and current are purely sinusoidal wave forms at power frequency, also the magnitude of voltage at its reference level. These are the primary considerations of power quality, deviation in which effects the equipment connected to power system in terms of its efficiency and life expectancy

One of the fundamental reasons for degradation of power quality is the restructuring of power industry and increased usage of semi conducting materials in almost all types of industrial and consumer equipment, which made power quality a concerning factor and measures to mitigate power quality disturbances. The term power quality can be used to address both voltage and current qualities as they work hand in hand, disturbance in one quantity effects the other. Inspite of the disturbance may be for a fraction of second but the affects are huge resulting in increased losses, decreased life expectancy and down time in case of manufacturing units. Hence power quality disturbance monitoring and providing necessary mitigating actions is a concerning task for both the utilities and the users. Monitoring and classification of different events require analysing huge data which was bit challenging in past days but due to the advancements in data analysis with machine learning techniques and improved databases the task became much simpler. In this paper the event monitoring and classification are done by training the data mining algorithms with machine learning techniques using python software.

This paper deals with the following: section 2 system modelling with a block diagram, section 3 different power quality disturbances with their typical figures, section 4 Data mining algorithms for classification, section 5 simulink circuit model with different disturbances, section 6 results and discussions, section 7 conclusions of the work.

2. SYSTEM MODELLING

The block diagram shown in Fig. 1 depicts the order in which the paper is organized. Different power quality events like voltage sag, swell, interruption, unbalance are modelled in Matlab simulink. The rms voltage waveform obtained at point of common coupling (PCC) is sampled with data logging feature of simulink and the sampled data is stored in the form of an excel file. The data set thus obtained is pre-processed to make it ready for training machine learning models for classification. The pre-processed data set is then used for training the algorithms and the accuracy of the trained models are checked using test data set.

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

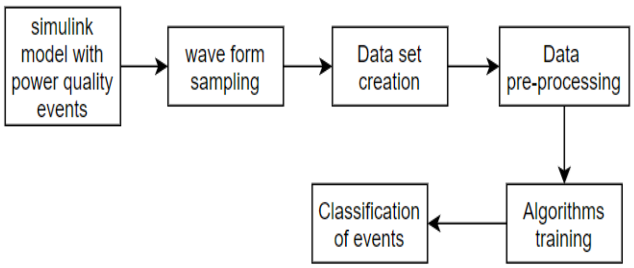

Fig. 1 block diagram

3. POWER QUALITY DISTRUBANCES

PQ issues in a power system include different types of electric disturbances, such as voltage sag, swell, interruption, Unbalance, harmonics etc. There are different types of PQ disturbances and also several ways to define and designate them. Definitions of the different disturbance types as well as their usual causes and unfortunate consequences of them are described below:

3.1 Voltage Sag:

It is known (ANSI std. 1100-1992) as the reduction in the magnitude of AC voltage (0.1-0.9 pu) for a period of half a cycle to a few seconds at the device frequency.

The key causes for the voltage sag are heavy load energization, switching in broad induction motors, unsymmetrical and symmetrical faults, shifting of enormous loads from one source to another, interaction with animals or interference with the tree branches.

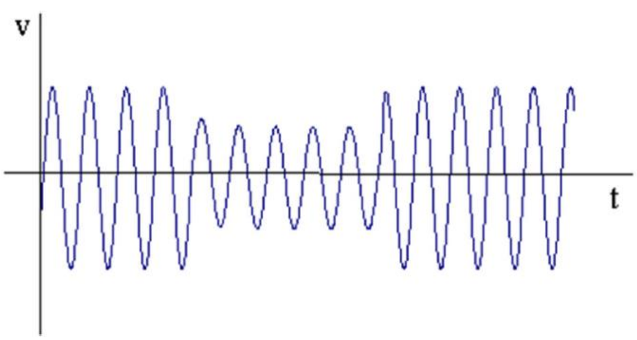

Fig. 2 Voltage Sag

3.2 Voltage Swell:

It is characterized as the increase in the magnitude of AC voltage (1.1-1.8 pu) at the system frequency, for duration from half a cycle to a few seconds.

Main reasons of voltage swells include energizing of capacitor banks, shutdown of large loads, unbalanced faults, transients and power frequency surges.

It causes problems with equipment that require constant steady-state voltage.

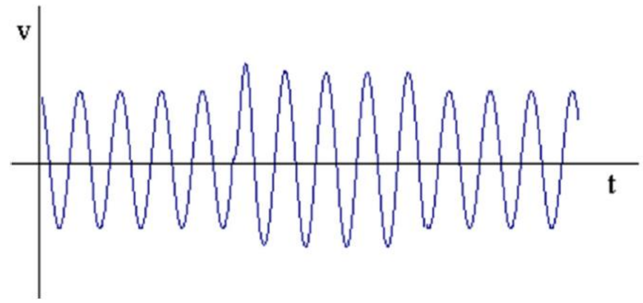

Fig. 3 Voltage Swell

3.3 Voltage Interruption:

Usually, it is the cumulative AC voltage loss (< 0.1 pu) for a few seconds to as long as one minute. As a consequence of a momentary short circuit, this may happen. This occurrence can be very momentary or may be repetitive for a brief period of time often.

Planned interruptions in the power system are typically triggered by construction or maintenance. Temporary interruptions are typically triggered by faults and are typically events that are erratic and spontaneous.

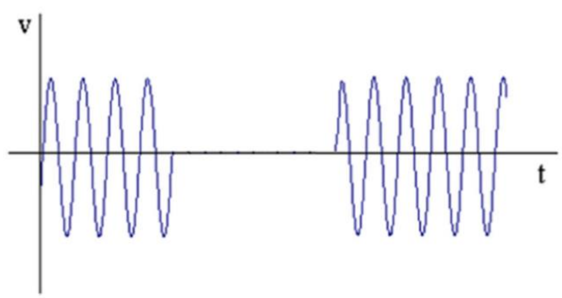

Fig. 4 Voltage Interruption

3.4 Voltage Unbalance:

It is defined as in a three-phase system in which the three voltage magnitudes or the phase angle variations between them are not identical.

Unbalanced faults, 1-ph loads on the 3-ph circuit and damaged fuses in 1-ph of a 3-ph are the key causes of unbalance.

Negative series currents flow, motor performance decreases, vibration and noise are the consequences due to unbalance.

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

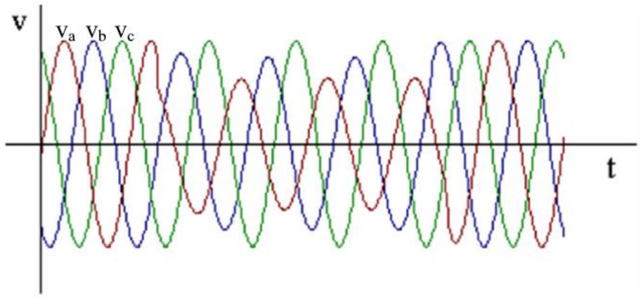

Fig. 5 Voltage Unbalance

4. DATA MINING

The process of data mining involves discovering the concealed patterns in the large quantity of information. Its a multi-disciplinary field which consists of machine learning, statistics and database. Data mining process contains knowledge extraction from data set and transforms it into structured data for later usage. Data mining is that the analysis step of the "knowledge discovery in databases" process or KDD. The knowledge discovery process involves the following steps: (i) Data Cleaning, (ii) Data Integration, (iii) Data Selection, (iv) Data Transformation, (v) Data mining, (vi)Pattern Evaluation and (vii) knowledge presentation. Apart from the raw analysis step, it also involves database and data management aspects, data pre-processing, model and inference considerations, metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. The process of data mining differs from classical statistical methods in that only model estimation focuses on solutions from statistical methods, whereas data mining techniques concentrate on both model creation and efficiency. Another essential distinction is that statistical approaches struggle to evaluate data containing a combination of numeric and qualitative types with missing values or data. Instead, data mining techniques can evaluate and interact intelligently with documents such as, missing values, without repetitive manual manipulation, as well as a combination of qualitative and quantitative results.

It is possible to consider data mining as a sub-set of data processing. Finding essential patterns and laws is the exploration and review of enormous knowledge. Data mining may also be a systematic and successive process in a large data collection to find and uncover hidden trends and data. In addition, it is used to construct models of machine learning which are further used in artificial intelligence. The first phase in the process of data mining is data collection from field equipment. When more information is obtained, the accuracy of the training model would be more precise. The data collected may be either big, medium, small, semi-organized or unstructured, as well as organized.

The real challenge of data processing is to semi-automatically or automatically analyse vast volumes of knowledge to retrieve previously unknown information. Data mining technology is an effective tool to deal with massive data, power quality data mining mainly includes data clustering analysis, correlation analysis and forecast analysis, and etc... The discovered patterns are then visualized in accordance with the input data and now this is pre processed data. this cleaned data can be used to train a machine learning model in order to fulfill our application like Classification, Clustering, Predictive analytics etc. Data pre processing is primarily to ensure data accuracy, completeness and consistency. In addition to the raw analysis phase, it also includes elements of data management and data visualization.

4.1 Algorithms for Data Mining

Different application based data mining algorithms are available, with SVM, KNN and Random Forest classifier being the commonly used classification algorithms.

4.1.1 Support Vector Machine (SVM):

Support vector machine can be used for both classification and regression tasks. It introduces the concept of decision planes for its classification task by outlining decision boundaries. When data of different classes are grouped to gather, SVM can classify them by using hyper plane or decision plane.

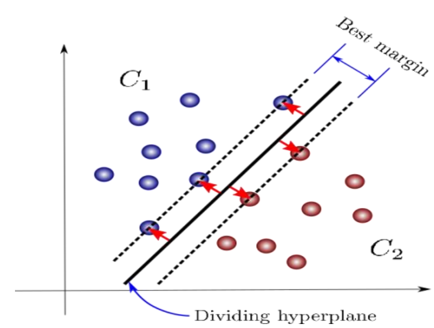

Fig. 6 SVM Classifier

This hyper plane can be created by using different support vectors, which are the linear separators between different data classes. Maximizing the margin of hyper plane results in the classification accuracy among different

International Journal of Engineering Research in Computer Science and Engineering
(IJERCSE)
Vol 8, Issue 2, February 2021

classes.

4.1.2 **K- Nearest Neighbour (KNN):**

K Nearest Neighbors is the simplest classification algorithm which classifies the target variable based on the no. of its nearest neighbors regardless of labels. By defining the value of K we can select no. of nearest neighbors should be present in order to classify the target variable. Each training vector defines a region in the space. KNN is also a non-parametric, lazy learning algorithm, which classifies the new target variables supported by the measure of distance or similarity functions. In general, for a 2 class problem, the odd value of K is selected.

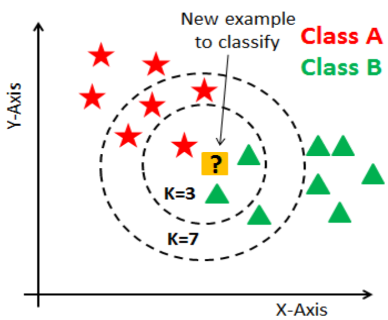

Fig. 7 Decision Tree Classifier

Non-parametric [illegible] the model structure calculated from the data set, for the underlying data distribution. It is also referred to as a lazy learning algorithm because no training data points for model generation are needed. In the testing process, all training data is used, which makes training quicker and the testing phase slower and more expensive. KNN is a simple algorithm that stores all the cases available and classifies the new data or cases on the basis of a measure of similarity.

4.1.3 Random Forest:

An ensemble learning is a group of models consisting of multiple independently trained supervised learning models, so the findings are combined in different ways to realize the ultimate prediction. The performance of the ensemble has greater predictive power than the independent outcomes of each of its constituent learning algorithms. The general principle is that the aggregation of learning models increases the precision of classification.

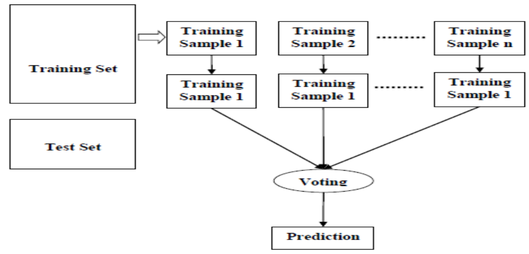

Fig. 8 Random Forest Classifier

This algorithm develops a lot of decision trees based on random selection of data as well as random selection of variables. It describes the class of dependent variable based on many decision trees. Most of the trees can provide correct prediction of class for most part of data. The final classification output is considered by majority voting from different models.

5. SIMULINK MODEL FOR GENERATING POWER QUALITY EVENTS DATA

The circuit shown in the Fig. 9 is modelled in MATLAB simulink. The circuit consists of a 33KV voltage source which is stepped down to 11KV and connected to a distribution line through the circuit breaker. The line is terminated with a distribution transformer 11/0.433KV which supplies a load of 190KW, 140KVAR.

Initially it is simulated without any disturbances and the rms voltage at point of common coupling (PCC) is recorded, this is considered as the actual rms value for the simulation. It is then simulated to obtain data for different events of power quality, such as voltage sags, swells, interruptions, and problems of imbalance. In order to show the difference between NO power quality event and power quality event the total simulation time is considered such that all the events that we want to classify exists within single run of simulation.

Voltage sags are generated for a given time period by a balanced three-phase ground fault with fault impedance. Voltage swells are created by connecting the three-phase capacitor bank to the grid. By opening the three phase circuit breaker, voltage interference is added, thus disconnecting the supply. Unbalance is created by an unbalanced fault in three phases. All events are generated with the simulink function block at their specified time intervals. The rms voltage is measured at PCC.

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 2, February 2021

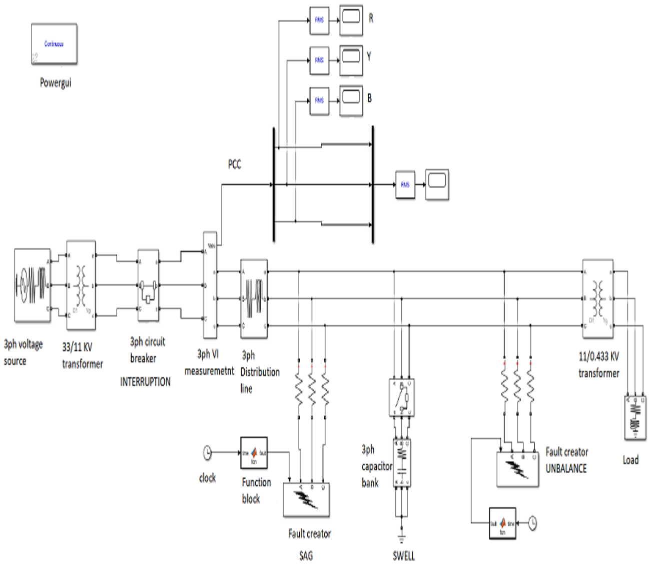

Fig. 9 Simulation circuit diagram of system

The rms voltage waveform is sampled at a sampling frequency of 6KHz throughout its simulation time. The sampled data is collected from the MATLAB SIMULINK shown in Fig. 8 and data set is stored in form of an excel file. A total of 64269 samples are obtained by sampling the obtained rms waveform at PCC throughout its simulation time.

6. RESULTS AND DISCUSSIONS

The obtained data set is thus pre-processed in the form of an excel file and the class attributes are formulated to define sag, swell, interruption, unbalance and no power quality. As data set formed is by sampling rms voltage, the python code in 'jupyter notebook' is written in such a way that the raw data collected with all the events mixed is diversified based on the voltage magnitude.

The total samples obtained are 64269, of which 123 samples contain SAG, 3540 samples contain SWELL, 1771 samples contain INTERRUPTION, 5900 samples contain UNBALANCE, but there are no power quality issues with the remaining 52935. Attributes used here are voltage of 1 phase rms along with class attributes that are 'NOPQ' for no issue with power quality,' SAG' for voltage sag,' SWELL' for voltage swell,' INTERRUPTION' for voltage interruption,' UNBALANCE' for voltage imbalance.

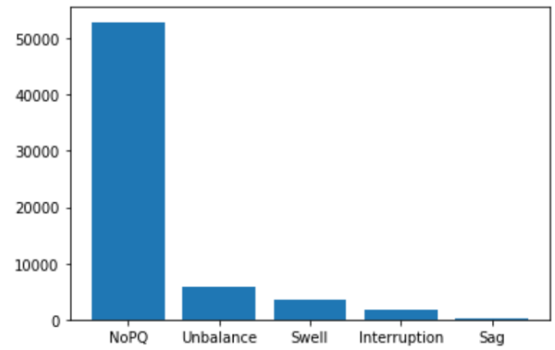

Fig. 10 pre-processed dataset Output

Fig. 10 displays the pre-processed python output, showing the number of attributes and the number of samples in each power quality problem classification case. The attributes and their corresponding number of samples are visualized in the form of a bar graph using the python visualization method Matplotlib.

The pre-processed data set is used for checking the accuracy of different classification algorithms. Using sklearn tool of python the pre-processed data set is splitted into train data set and test data set. Train data is used to train the data mining algorithms for classification, i.e. SVM, KNN, Random Forest. The algorithms are evaluated after training on the basis of the test data set.

The test data set size considered is 0.25 i.e 75% of the pre-processed data set is used for training the model and remaining 25% data is used for testing the accuracy of different classification algorithms. The value counts are taken with the training data set i.e to determine how many class attributes are present in 75% of pre-processed set.

To test its efficiency, a Confusion matrix is generated for all the algorithms. Maybe a tabular description of the amount of right and incorrect predictions made by a classifier is an uncertainty matrix. It can be used through the measurement of performance metrics such as accuracy , precision , recall, and F1-score to assess the performance of a classification model.

sno	algorithm	event tested	no of samples	classification	
				correct	incorrect
1	SVM	sag	27	26	1
		swell	892	892	0
		interruption	430	430	0
		unbalance	1507	23	1484
		no PQ	13212	13212	0
		overall	16068	14583	1485
2	KNN	sag	27	26	1
		swell	892	892	0
		interruption	430	430	0
		unbalance	1507	1507	0
		no PQ	13212	13211	1
		overall	16068	16066	2
3	Random Forest	sag	27	26	1
		swell	892	892	0
		interruption	430	430	0
		unbalance	1507	1505	2
		no PQ	13212	13210	2
		overall	16068	16063	5

Table.1 Comparison of data mining algorithms with confusion matrix results

Table.1 shows the results obtained after evaluating the algorithms using the evaluating collection. The overall accuracy of the SVM algorithm is 90.75 percent, 99.99 percent for KNN and 99.96 percent for Random Forest in the classification of power quality problems, as shown in Table 2.

Feature	SVM	KNN (K=2)	Random Forest
Accuracy(%)	90.75	99.99	99.96

Table.2 Performance Comparison

7. CONCLUSION

As power quality is becoming the utmost important factor today, identifying and mitigating should be very fast and accurate to protect the equipment. This paper presents different data mining algorithms like SVM, KNN, Random Forest use Python tools for classification of various power quality issues such as voltage sag , swell, interruption, unbalance. With the information set obtained from the sampled rms voltage waveforms generated in Matlab simulink, the algorithms are trained and checked. The accuracy depends on the type of data collected and amount of data used for training the model and the training time. So the best performing algorithm can be decided based on the trade-off between accuracy and training time. This work can be extended to accurately classify different hormonics present in the power system and hence necessary mitigation methods can be implemented to safeguard the system.

8. REFERENCES

[1] Power Quality Standards IEC 61000-4-30
[2] ANN Based Solar Power Forecasting in a Smart Microgrid System for Power Flow Management; 2019 Innovations in Power and Advanced Computing Technologies (i-PACT).
[3] D. R. Nair, Devi S, M. G. Nair and K. Ilango, ”Tariff based fuzzy logic controller for active power sharing between microgrid to grid with improved power quality,” 2016 International Conference on Energy Efficient Technologies for Sustainability (ICEETS), Nagercoil, pp. 406-409, 2016.
[4] Hardware implementation of power sharing and power quality improvement for grid integration of microgrid; 2017 International Conference on Technological Advancements in Power and Energy (TAP Energy).
[5] Lekshmi R. Chandran, Karuppasamy, D. Ilango, and Dr. Manjula G. Nair, “Voltage quality enhancement in distribution systems using dynamic voltage restorer with adaptive fuzzy pi controller”, vol. 8, pp. 3036-3041, 2019.
[6] Power quality improvement using STATCOM with renewable energy sources, K Ilango, A Bhargav, A Trivikram, PS Kavya, G Mounika, MG Nair 2012 IEEE 5th India International Conference on Power Electronics (IICPE).
[7] Fault Classification and Diagnosis of Industrial Application Motor Drives using Soft Computing Techniques, GS Ayyappan, K Venugopal, RM Raja, R Poonthalir, K Ilango; 2019 4th International Conference on Recent Trends on Electronics.
[8] Vinit Kumar, Manbir Kaur,“Power Quality Event Generation in MATLAB/Simulink Environment,”International Journal of Advanced Research in Engineering and Applied Sciences.
[9] MATLAB/Simulink Based Modeling and Simulation of Power Quality Disturbances,”978-1-4799-4848-2/14/$31.00 ©2014 IEEE”
[10] Dr.A jaya Laxmi,S Asha Kiranmai “DATA EXTRACTION FOR CLASSIFICATION AND CHARACTERIZATION OF POWER QUALITY PROBLEMS,” International Journal of Advanced Research in Engineering and

Applied Sciences .

[11] Abdullah Asuhaimi Mohd zin,"MATLAB/Simulink Based Modeling and Simulation of Power Quality Disturbances", IEEE Conference paper.

[12] Pandit, N., & Chakrasali, R. L. (2017). Power quality issues in Indian power distribution utilities and feasible solutions. International Journal of Scientific & Technology Research, 6(5), 23–25.

[13] Sharmila, M., Sundarabalan, C. K., & Selvi, K. (2017). Power quality disturbances classification using data mining technique. International Journal of Research in Science & Engineering, 3(3), 77–90.

[14] Akinola, S., & Oyabugbe, O. (2015). Accuracies and training times of data mining classification algorithms: An empirical comparative study. Journal of Software Engineering and Applications, 8, 470–477

[15] Upadhyaya, S., & Mohanty, S. (2016). Localization and classification of power quality disturbances using maximal overlap discrete wavelet transform and data mining based classifiers. IFAC-Papers OnLine, 49(1), 437–442.

[16] Jeya Sheela, Y., & Krishnaveni, S. H. (2017). A comparative analysis of various classification trees (pp. 1346–1352). Kollam: IEEE International Conference on Circuit, Power and Computing Technologies (ICCPCT).

[17] Pandey, P., & Prabhakar, R. (2016). An analysis of machine learning techniques (J48 & AdaBoost)-for classification (pp. 1–6). Delhi: IEEE 1st India International Conference on Information Processing (IICIP).

[18] Hamsagayathri, P., & Sampath, P. (2017). Performance analysis of breast cancer classification using decision tree classifiers. Int J Curr Pharm Res, 9(2), 19–25.

[19] Suresh, K., & Chandrashekhar, T. (2012). Characteristics analysis of voltage sag in distribution system using RMS voltage method. ACEEE Int J on Electrical and Power Engineering, 3(1), 55–61.

[20] Soman, K. P., Diwakar, S., & Ajay, V. (2006). Data mining: Theory and practice. Delhi: PHI Learning Pvt. Ltd..

[21] Dinakaran, S., & Thangaiah, D. R. J. (2013). Role of attribute selection in classification algorithms. International Journal of Scientific & Engineering Research, 4(6), 67–71.

[22] Udemy : Data Analytics using Python