Power System Technology
ISSN:1000-3673

Received: 06-06-2024 *Revised: 15-07-2024* *Accepted: 28-08-2024*

Development of suitable Data Mining Algorithm for Power System Fault Detection using Machine Learning Approach

Mahesh Yenagimath1, Dr. Shekhappa Ankaliki2, Dr. Girish V3
1Department of Electrical and Electronics Engineering, Hirasugar Institute of Technology, Nidasoshi, Karnataka, India.
2Department of Electrical and Electronics Engineering, S.D.M College of Engineering and Technology, Dharwad, Karnataka, India.
3Executive Engineer (EI), ALDC, HESCOM Corporate office, HESCOM, Navanagar,Hubli, Karnataka, India
Corresponding Author Email: mahesh28mar@gmail.com

Abstract: - A Modern power system (PS) is a complex interconnected network which includes four different parts such as generation, transmission, distribution, and loads. Since many advanced measurements and protection instruments are equipped with modern power system, a huge amount of data has been collected with larger dimensionality and quantity. To utilize the long previous data efficiently for post fault analysis, large capacity memory devices are required at PDC level. Accurate prediction. Past literarature reports indicates that the traditional analytical methods which are presently used in practice are not accurate and also not providing enough speed. Therefore, Machine Learning algorithms are used to create emulation of power system. These algorithms are used to make fast and accurate decisions. Here an attempt is made to implement various Machine Learning algorithms such as K-Nearest Neighbour (KNN), Decision Tree (DT), Support Vector Machine (SVM) on power system fault data collected by PMU from Indian power system for protection. Here an attempt is made to check the performance of different Machine learning algorithms using different normalization techniques such as standard scaler and Minmax scaler which helps to extract the useful information for improving situational awareness in power system. .
Keywords: Machine Learning, Phasor Data Concentrator, Data Mining.

1. Introduction

In 20th century, one of the greatest achievements done in power system is its automation. But when we consider large amount of losses and interuuptions which areoccurying currently on power system, its reliability needs to be enhanced considerably [1]. Local measurements are playing a vital role in various controlling strategies of power system and Overall status of the power system mainly depends on dynamic bbehaviour of the region [2]. The present SCADA-based monitoring system measures 2-5 samples per second which is use of PMU, which measures all the parameters synchronously in faster rate i.e 25 samples per second [4]. But analysis of such a large amount of collected data will help to improve the reliability aspect of the power systemin a larger way. In addition to that, the new challenge in infront of the current electric grid is managinag effective storage system of this large accumulated data by PMU[5]. The synchrophasor technology was firstly introduced in the year 2000 in the

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

power system by keeping an objective to improve the energy quality of the power system. This new PMU technology monitors voltage and current parameters with much higher sampling rate as compared to traditional previous technologies used in power system[6]. This advanced synchrophasor technology will help engineers and researchers to analyse dynamic behavior or events of the power system in better way. In addition to that, overall reliability of the power system can be improved by establishing Wide area measurement and control systems(WAMCS) in vary efficient manner[7]. The only choice to improve the overall efficiency and control operation of the power system is to adapt synchrophasor technology [8].

2. Machine Learning in Power System

Machine learning(ML) is an application of AI that makes the systems to learn and these algorithms try to learn hidden patterns from training data and based on knowledge gained by rigourous training process it makes the accurate predictions[9]. A well defined function refers the inputs and generates output as shown in Figure 1[10].

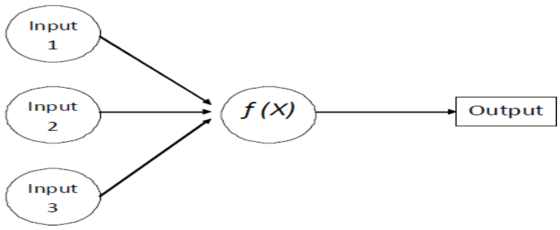

Figure 1: Mapping of input to output by machine learning

For accurate prediction and forecasting, machine learning algorithms mainly depends on historical data which contains mathematical calculations and simulations of respective power system. By processing synchrophasor data fast event detection is possible in power system[11]. Especially fast event detection in power system requires real time monitoring. As we know fault detection is one application of machine learning technique which is used in power system, which determines the system failure based on collected PMU data with all parameters of the power system[12]. In real time application, machine learning algorithms are most suitable because of its fast decision making capability.

After going through reported progress in power system, two major deficiencies are still exist in fast and accurate fault detection. One of the most important deficiency among that is collection of large online data from the monitoring system. Here an attempt is made to implement various Machine Learning algorithms such as K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Logistic Regression (LR) on power system fault data collected by PMU[13].

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 *Revised: 15-07-2024* *Accepted: 28-08-2024*

2.1 Logistic Regression (LR)

This is kind of supevised learning algorithm which is basically used to predict events (like yes or no) occurring. This type of algorithm is basically used to get solution for classification problems[14]. Here basically parameters are calculated. The Logistic regression algorithm tries to predict probability of any power system event or class that is basically dependent on many factors As the output of this algorithm always lies between 0 and 1 because of that it is most commonly used for classification of different kind of events or faults in power system.

2.2 Support Vector Machine (SVM)

For better classification, regression and outlier detection activities this algorithm is most suitable[15]. This algorithm is basically used for classification problems by creating best decision boundry that can separate n-dimensional space into classes. When we add new data point to check the result then it will categorize that new data point into correct category, hence these extreme cases or vectors or points that are used to make best decision boundry are called as support vectors [16].

Figure 2 shows the data segregation process made by SVM algortithn for making the best decision. Two classify two different types of dots with blue and red colour, the SVM algorithm will try to learn the best features of two different objects separately and while making the decision to classify the objects it will consider the extreme points those are called as support vectors into consideration.

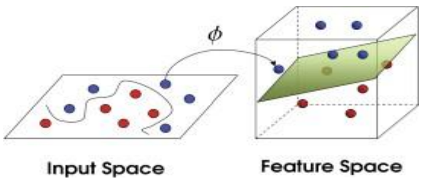

Figure. 2 SVM for Classification

SVM algorithm maps the input vectors into extreme points in high dimensional feature space through different nonlinear methods[17]. Because of these all features SVM algorithm is suitable for making the best decision in many applications like face recognition, handwriting detection, classification etc.

2.3 K-Nearest Neighbor Algorithm (K-NN)

It belongs to Supervised Learning technique and will store all available data having different cases for learning. It will try to identify the similarity between new data point with old class, so it will classify the new data based on most similar data among the available catagories [18]. If there are two categories of data then it will classify the new data by comparing similar features of both categories then it will take the decision considering most similar feature. The accuracy of this algorithm is very high when data is not normal with unknown distributions.

Received: 06-06-2024 *Revised: 15-07-2024* *Accepted: 28-08-2024*

Here Euclidean distance concept is used to identify K-closest points among the various data points [19].

2.4 Decision Tree (DT)

This is one of the typical data mining algorithm which is having highest performance among the available algorithms and whose compuatational efficiency is also more. This algorithm requires a set of attributes as input predictors which are categorical variables or numerical variables of n number whose measurements are real numbers and its target value may be discrete or continuous [20]. The data set is split into two subsets because idea here is to apply the rules to each subset to make the prediction as pure as possible [21]. As this algorithm is having an ability to process as well as analyses the time series signal efficiently this can be used as one of the best tool for dignosing power system faults especially in distribution lines.

3. Implementation of Machine Learning Algorithms for Fault Prediction

In this section, the implementation of various Machine Learning algorithms to predict the power system faults is discussed in detail. The simulation is carried out with Google Colab platform using python programming.

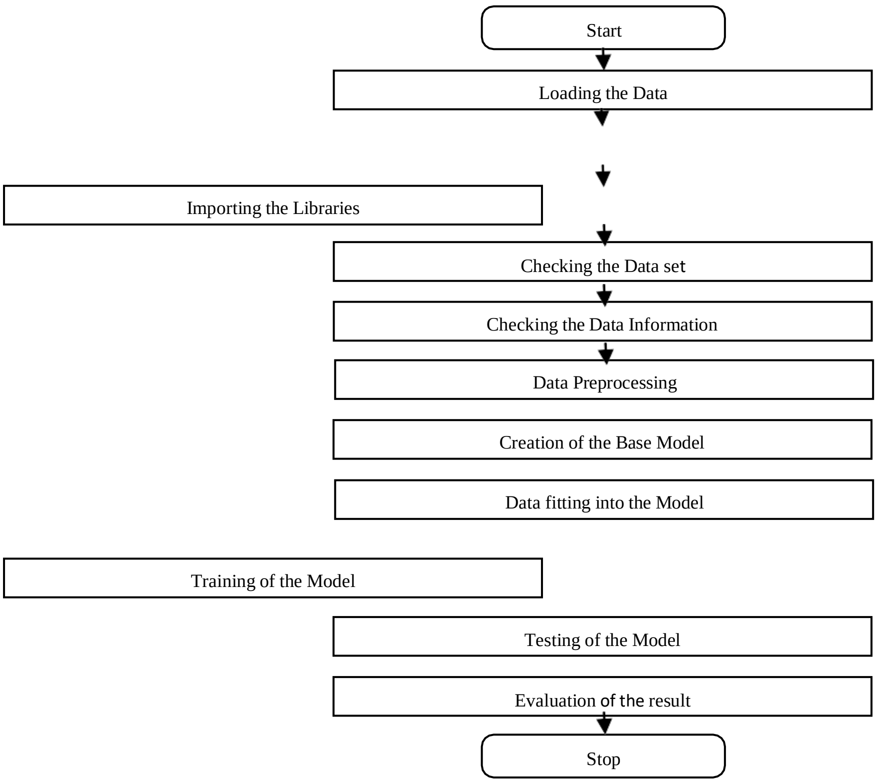

Figure 3. Details steps of implementing Machine Learning algorithm

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

Initially, we collected the fault data measured by PMU from the Kaggle website. After successful implementation, we tried to implement the algorithms using offline synchrophasor data collected from Indian power system. The detailed steps of implementation are as shown in figure 3 with the flowchart. The collected data is not suitable for implementing machine learning algorithms hence, we applied suitable data pre-processing technique to make the collected data suitable for Machine Learning Algorithm. Here around 800000 measurements of the time duration around 20 milliseconds of three phase power scheme which operates at 50 Hz frequency forone complete cycle is taken for analysis.

The proposed algorithms are applied on power system fault data to predict the fault. The simulation was carried out using Python language in Google Colab. The available synchrophasor data is used training all the proposed algorithms then after successful training, it is tested. Here 80% of available synchrophasor data is randomly selected for training purpose and remaining 20% data is used for testing purpose.

3.1 Implementation of Logistic Regression Algorithm

Step 1: Loading Data: As shown in Figure 4 the collected fault data is stored in google drive, which is imported to Google Colab using the following procedure. The file name is PMUdataSet.csv.

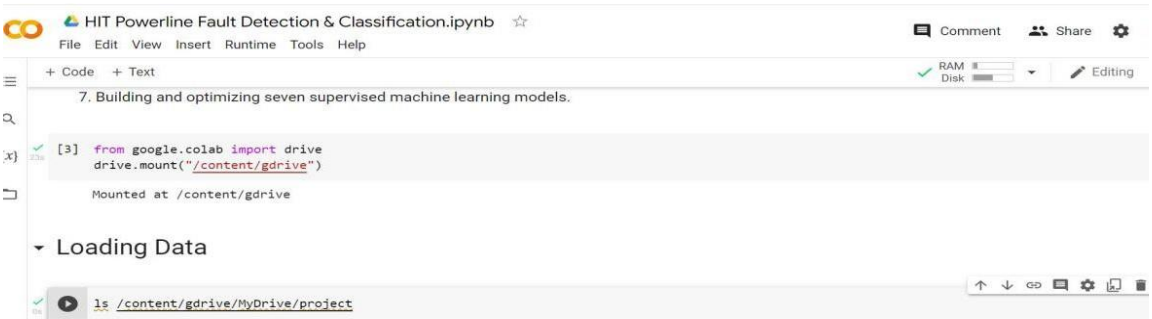

Figure. 4 Loading data

Step 2: Importing Libraries: After loading the data, it is necessary to import the libraries like Pandas and Numpy which are helpful for data processing application. A panda is an open-source python package, which is specially utilized for reading data from the file and the module Numpy works with the numerical data.

Step 3: Checking the datasets: It will check the available parameters in loaded data set. Here Ia, Ib, Ic, are the current parameters and Va, Vb, Vc are the voltage parameters.

Step 4: Checking Data Information: It checks the size of the dataset what we have loaded. All the 6 variables present in the dataset Ia, Ib, Ic, Va, Vb, Vc are the floating type and output is in int type and all are in 64 bit wide.

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

Step 5: Data Preprocessing: Our machine learning algorithm requires data in a particular format but whatever raw data collected from the power system is not in the form which is suitable for the machine learning algorithm [22]. Hence data pre-processing is required.

There are different types of data normalization techniques.

1. Standardization: Here each input variables are separately scales down by subtracting it by mean and after that dividing it by standard deviation.
2. Min-Max Normalization: In this technique each input variables is separately scaled down in the range of zero(0) to one(1) [23].

Step 6: Model Creation: After importing the necessary library functions the pre-processed data is downloaded into the model. This data having different features gives the information to create the model.

Step 7: Evaluating the Model:

Evaluation of Model is done using Confusion matrix which is as shown in figure 5. To evaluate the performance of particular classifier is done with the help of this matrix. This shows summary of correct and incorrect predictions in tabular form which is very much helpful to understand the performance of particular classifier.

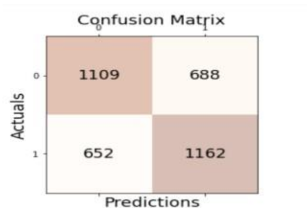

Figure 5. Confusion Matrix of Logistic Regression ModelStep 8: Result of the Logistic

Regression Model

The accuracy of proposed Logistic Regression Model is as shown in figure 6.

	precision	recall	f1-score	support
0	0.63	0.62	0.62	1797
1	0.63	0.64	0.63	1814
accuracy			0.63	3611

Figure 6. Result of the Logistic Regression Model

3.2 Implementation of KNN AlgorithmModel Creation of KNN Algorithm

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

In this section, the attempt is made to create a Model using KNN algorithm after importingthe necessary library functions using pre-processed data.

Result of the KNN model

The accuracy of proposed KNN Model is as shown in figure 7.

	precision	recall	f1-score	support
0	1.00	0.87	0.93	1797
1	0.89	1.00	0.94	1814
accuracy			0.94	3611

Figure 7. Result of the KNN model

Confusion Matrix of KNN Model

Evaluation of the KNN Model is as shown in figure 8.

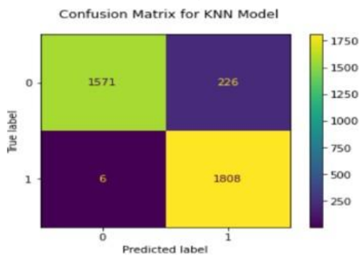

Figure 8. Confusion Matrix of KNN Model

Result Comparison of Logostic Regression and KNN Algorithms

After implementing both Logistic regression and KNN algorithms on same PMU data which is collected from Indian power system we concluded that KNN algorithm is showing more accuracy as compared to Logistic regression algorithm which as shown in Table 1.

Table 1 : Result Comparison of LR and KNN algorithm

Logistic Regression algorithm		Precision	Recall	F1 Score	Support	Accuracy
	0	0.63	0.62	0.62	1797	0.63
	1	0.63	0.64	0.63	1814	
KNN Algorithm	0	1.00	0.87	0.93	1797	0.94
	1	0.89	1.00	0.94	1814	

3.3 Implementation of SVM Model creation using Standardscaler normalization

In this section, the attempt is made to create a Model using SVM algorithm after importing the necessary library functions using pre-processed data using Standardscaler normalization .

Power System Technology
ISSN:1000-3673

Received: 06-06-2024 *Revised: 15-07-2024* *Accepted: 28-08-2024*

Evaluation of the SVM Model is as shown in figure 9.

	precision	recall	f1-score	support
0	0.88	1.00	0.94	1821
1	0.00	0.00	0.00	240
accuracy			0.88	2061
macro avg	0.44	0.50	0.47	2061
weighted avg	0.78	0.88	0.83	2061

Accuracy: 88.36

Figure 9. Result of SVM using Standardscaler normalization

3.4 Implementation of SVM Model creation using Minmaxscaler normalization

In this section, the attempt is made to create a Model using SVM algorithm after importing the necessary library functions using pre-processed data using Minmaxscaler normalization . Evaluation of the SVM Model is as shown in figure 10.

	precision	recall	f1-score	support
0	0.88	1.00	0.94	1821
1	0.00	0.00	0.00	240
accuracy			0.88	2061
macro avg	0.44	0.50	0.47	2061
weighted avg	0.78	0.88	0.83	2061

Accuracy: 88.36

Figure 10. Result of SVM using Minmaxscaler normalization

3.5 Implementation of KNN Model creation using Standardscaler normalization

Figure 11 shows the evaluation of the KNN Model using Standardscaler normalization.

	precision	recall	f1-score	support
0	1.00	0.99	1.00	1821
1	0.96	0.99	0.97	240
accuracy			0.99	2061
macro avg	0.98	0.99	0.98	2061
weighted avg	0.99	0.99	0.99	2061

Accuracy: 99.37

Figure 11. Result of KNN using Minmaxscaler normalization

3.6 Implementation of KNN Model creation using Minmaxscaler normalization

Figure 12 shows the evaluation of the KNN Model using Minmaxscaler normalization.

Power System Technology
ISSN:1000-3673

Received: 06-06-2024 *Revised: 15-07-2024* *Accepted: 28-08-2024*

	precision	recall	f1-score	support
0	0.98	0.98	0.98	1821
1	0.88	0.86	0.87	240
accuracy			0.97	2061
macro avg	0.93	0.92	0.93	2061
weighted avg	0.97	0.97	0.97	2061

Accuracy: 96.99

Figure 12. Result of KNN using Minmaxscaler normalization

3.7 Implementation of Decision Tree Model creation using Standardscaler normalization

Figure 13 shows the evaluation of the Decision Tree Model using Standardscaler normalization.

	precision	recall	f1-score	support
0	0.98	0.98	0.98	1821
1	0.88	0.86	0.87	240
accuracy			0.97	2061
macro avg	0.93	0.92	0.93	2061
weighted avg	0.97	0.97	0.97	2061

Accuracy: 96.99

Figure 13. Result of Decision Tree using Standardscaler normalization

3.8 Implementation of Decision Tree Model creation using Minmaxscaler normalization

Figure 14 shows the evaluation of the Decision Tree Model using Minmaxscaler normalization.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1821
1	0.97	0.97	0.97	240
accuracy			0.99	2061
macro avg	0.99	0.99	0.99	2061
weighted avg	0.99	0.99	0.99	2061

Accuracy: 99.42

Figure 14. Result of Decision Tree using Minmaxscaler normalization

Here an attempt is made to observe change in accuracy while implementing KNN,SVM,LR and DT algorithms. Two different normalization techniques such as standard scaler and Minmax scaler are used on same data to evaluate the same model. Table 2 shows the accuracy of two different normalization techniques.

Received: 06-06-2024 *Revised: 15-07-2024* *Accepted: 28-08-2024*

Table 2: Comparison results for the proposed methods

Name of the Algorithm	Type of Normalization	Accuracy
Support Vector Machine	Standard Scaler	88.36%
	MinMax Scaler	88.36%
K-Nearest Neighbour	Standard Scaler	99.37%
	MinMax Scaler	96.99%
Decision Tree	Standard Scaler	99.13%
	MinMax Scaler	99.42%

4. Conclusion

This paper highlights about the implementation of different Machine Learning algorithms on synchrophasor data which is collected by Indian power system which are specially used for improving situational awareness. Firstly, the simulation was carried out using Python language in Google Colab. The accuracy of the KNN algorithm better as compared to accuracy of the Logistic Regression algorithm. The paper also focuses on implementation of two different normalization techniques such as Standard scaler and Minmax Scaler on three different models for evaluating the result. But the interesting factor is, for KNN algorithm accuracy is better for standard scaler normalization and for Decision tree algorithm accuracy is more for Minmax scaler normalization. Mainly this paper highlights on different choices of software implementation for power system fault detection.

References

Texbooks:

[1] D. o. Energy, "The Smart Grid: An Introduction," D. o. Energy, Ed., ed.Journal:

[2] Anjan Bose, "Power system stability: new opportunities for control”, pp 315–330, https://link, springer com/chapter/10.1007/978-1-4612-0037-6_16.

[3] R. Avila-Rosales, M. J. Rice, J. Giri, et al. "Recent experience with a hybrid SCADA/PMU on-line state estimator" in Power & Energy Society General Meeting, 2009. PES '09. IEEE, 2009, pp. 1-8.

[4] L. Xie, Y. Chen, and P. R. Kumar, “Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis,” IEEE Trans.Power Syst.,vol.29,no.6,pp.2784–2794,Nov. 2014.

[5] Kedar V. Khandeparkar, Nitesh Pandit, A.M. Kulkarni et.al, “Design of a Phasor Data Concentrator for Wide Area Measurement System”,https://www.researchgate.net/publication /317664931.

[6] Philip Top and John Breneman, “Compressing Phasor Measurement Data”, DOI: 10.1109/NAPS.2013.6666959, https://ieeexplore.ieee.org/document/6666959.

[7] Michael Brown; Milan Biswal; Sukumar Brahma; Satish J Ranade et al. (2016), “Characterizing and quantifying noise in PMU data” In: 2016 IEEE Power and Energy

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

Society General Meeting (PESGM), IEEE. https://doi.org/10.1109/PESGM.2016.7741972.

[8] Sayali N. Muneshwar, R. P. Hasabet, Parag Kose et al. (2014), "A New Adaptive PMU based Protection Scheme for Interconnected Transmission Network System" 2014, International Conference on Circuits, Power and Computing Technologies[ICCPCT-2014]. IEEE.https:// doi.org/10.1109/ ICCPCT.2014.7054922.

[9] L. Wehenkel, "Machine learning approaches to power-system security assessment,"IEEE Expert, vol. 12, pp. 60-72, 1997.

[10] N. J. Nilsson, "Introduction to machine learning", 1996. An early draft of aproposed textbook.

[11] Senlin Zhang,Yixing Wang, Meiqin Liu; Zhejing Bao et al (2017), "Data-based line trip fault prediction in power systems using LSTM networks and SVM", https://ieeexplore.ieee.org /document/8233109.

[12] B. H. Chowdhury and K. Wang, "Fault Classification using Kohonen Feature Mapping," 1996, pp. 194-198,https://ieeexplore.ieee.org/document/501067/authors#authors.

[13] N Hatziargyriou, "Machine learning applications to powe rsystems",pp.308-317,2001,https:// link.springer.com/chapter/10.1007/3-540-44673-7_20.

[14] Le Xu, Mo Yuen Chow, "Power Distribution Systems Fault Cause Identification UsingLogistic Regression and Artificial Neural Network", 1-59975-028-7/05/$20.00©2005 ISAP,file:///C:/ Users / Om / Downloads / Power distribution systems fault cause_identificat.pdf.

[15] Jakkula, V.Tutorial on support vector machine (svm). School of EECS, WashingtonState University (2006).

[16] N. Cristianini, J. Shawe-Taylor, "An Introduction to Support Vector Machines andOther Kernel-based Learning Methods", ambridge university press, 2000.

[17] Ladjici Ahmed Amine, et al.(2014), "Power System Applications of Support Vector Machine in Classification and Regression", 3rd International Conference on Electrical Engineering, May 19-21 2009, Algiers Algeria,https://www.researchgate.net/publication/262284775.

[18] Aida Asadi Majd et.al,Haidar Samet,Teymoor Chanbari, "K-NN based Fault Detection and Classification Methods for Power Transmission Systems",DOI 10.1186/s41601-017-0063-z, https:// pcmp.springeropen.com / track/pdf/10.1186/s41601-017-0063-z.pdf.

[19] Kashvi Taunk; Sanjukta De; Srishti Verma; Aleena Swetapadma,et al."A BriefReview of Nearest Neighbor Algorithm for Learning and Classification" https:// ieeexplore.ieee.org/ document/9065747.

[20] Chengxi Liu, Zakir Hussain Rather, Z. Chen, et.al (2013), "An overview of DecisionTree Applied to Power Systems", Manuscript received July 11, 2013, revised September 11, 2013, https: //www.researchgate.net/publication/ 274464463.

[21] Luigi Vanfretti, V. S. Narasimham Arava, Member, "Decision Tree-Based Classification of Multiple Operating Conditions for Power System Voltage Stability Assessment",https://www.sciencedirect.com/science/article/pii/S0142061519334775.

[22] Giuseppe Ciaburro et al(2022), "Machine fault detection methods based on machine learningalgorithms: A review", https://www.aimspress.com/aimspress-data/mbe/2022/11/PDF/mbe-19-11-534.pdf.

Received: 06-06-2024 *Revised: 15-07-2024* *Accepted: 28-08-2024*

[23] Oyeniyi Akeem Alimi,Khmaies Ouahada, Adnan M.Abu-Mahfouz et.al(2020),“.AReview of Machine Learning Approaches to Power System Security and Stability”,DOI 10.1109/ACCESS.2020.3003568, IEEE Access.
[24] Mahesh Yenagimath, Shekhappa Giriyappa Ankaliki, “Lossless Compression of Synchrophasor Data in Wide Area Monitoring (WAM) of Power System”, © November 2021| IJIRT | Volume 8 Issue 6 | ISSN: 2349-6002, ijirt 153319international journal of innovative research in Technology.
[25] Mahesh Yanagimath, Shekhappa G. Ankaliki, “Review on Synchrophasor basedData Mining Techniques and Tools”, in i-manager’s Journal on Power Systems Engineering, Vol. 7, No. 4, November 2019 - January 2020.