



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 7, July 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

Investigation and Determination of Partial Discharge Source using Gaussian Naive Bayes (GNB) Technique

Kajol Chaudhari¹, Akhilesh Kumar Mishra², Priyanka Kothoke³, Yogesh Chaudhari⁴

Research Scholar, Shri J. J. T. University, Rajasthan, India¹

Professor, Shri J. J. T. University, Rajasthan, India²

Assistant Professor, MGM College of Engineering and Technology, Maharashtra, India³

Assistant Professor, School of Engineering and Technology, Navrachana University, Vadodara, India⁴

ABSTRACT: Partial Discharge (PD) patterns are an important tool for the diagnosis of High Voltage (HV) insulation systems. Human experts can discover possible insulation defects in various representations of the PD data. One of the most widely used representations is phase-resolved PD (PRPD) patterns. In order to ensure reliable operation of HV equipment, it is vital to relate the observable statistical characteristics of PDs to the properties of the defect and ultimately to determine the type of the defect. Naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes theorem with strong independence assumptions between the features. Right now considering the importance of GNB to evaluate its suitability in PD course of action. Model is built for this method in Google co-laboratory in Python and by giving input as statistical parameters, source of PD can be obtained on Graphical User Interface that whether it is void, surface or corona discharge.

KEYWORDS: Gaussian Naive Bayes, Google co-laboratory, Partial Discharge, Phase-Resolved.

I. INTRODUCTION

PD is a localized electrical discharge that partially bridges the insulation between conductors and which may or may not occur adjacent to a conductor [1]. In general, PDs are concerned with dielectric materials used, and partially bridging the electrodes between which the voltage is applied [2]. The insulation may consist of solid, liquid, or gaseous materials, or any combination of them. PD is the main reason for the electrical ageing and insulation breakdown of high voltage electrical apparatus. Different sources of PD give different effect on insulation performance. Therefore, PD classification is important in order to evaluate the harmfulness of the discharge [2].

PD classification aims at the recognition of discharges of unknown origin. For many years, the process was performed by investigating the pattern of the discharge using the well-known ellipse on an oscilloscope screen, which was observed crudely by eye. Nowadays, there has been extensive published research to identify PD sources by using intelligent technique like artificial neural networks, fuzzy logic and acoustic emission [2].

The recent upsurge of research on PD phenomena has been driven in part by development of new fast digital and computer-based techniques that can process and analyze signals derived from PD measurements. There seems to be an expectation that, with sufficiently sophisticated digital processing techniques, it should be possible not only to gain new insight into the physical and chemical basis of PD phenomena, but also to define PD 'patterns' that can be used for identifying the characteristics of the insulation 'defects' at which the observed PD occur [3]. One of the undoubted advantages of a computer-aided measuring system is the ability to process a large amount of information and to transform this information into an understandable output [4].

There are many types of patterns that can be used for PD source identification. If these differences can be presented in terms of statistical parameters, identification of the defect type from the observed PD pattern may be possible. As each defect has its own particular degradation mechanism, it is important to know the correlation between discharge patterns and the kind of defect [5]. Therefore, progress in the recognition of internal discharge and their correlation with the

kind of defect is becoming increasingly important in the quality control in insulating systems [6]. Research has been carried out in recognition of partial discharge sources using statistical techniques and neural network [7].

The three distinct classifications of PD pulse data patterns gathered from the digital PD detectors during the experiments are -- phase-resolved data, time-resolved data and data having neither phase nor time information.

1. The phase-resolved data consists of three-dimensional discharge epoch, ϕ charge transfer, q discharge rate, n patterns ($\phi \sim q$, $q \sim n$ and $\phi \sim n$ patterns) at some specific test voltage.
2. The time-resolved data constitute the individual discharge pulse magnitudes over some interval of time, i.e., $q \sim t$ data pattern.
3. The third classification of information comprises of variations in discharge pulse magnitudes against the amplitude of the test voltage V (for both increasing and decreasing levels), i.e., $q \sim V$ data.

Out of these three classifications, phase resolved data method is used in this work for statistical method whose pre-processed data is used as an input to this GNB Technique. Phase resolved data is used since it is drawing 2D patterns from raw data showing variations between either phase angle and charge, phase angle and number of pulses or number of pulses and charge and then comparison of these patterns amongst themselves will result in improved accuracy of the outputs.

At present, in any industry, just the source of PD is notified by using a buzzer. However, the buzzer indicates only the place of its occurrence and not the type of discharge, which is very essential for removal of insulation defects. The technique suggested in this research enable to know the type of the PD accurately so that the insulation defects can be removed from any high voltage equipment by specific techniques to avoid chances of supply failure.

The important parameters to characterize PDs are phase angle ϕ , PD charge magnitude q and PD number of pulses n . PD distribution patterns are composed of these three parameters. Statistical parameters are obtained for phase resolved patterns ($\phi \sim q$), ($\phi \sim n$) and ($n \sim q$).

PD pulses are grouped by their phase angle with respect to 50 (± 5) Hz sine wave. Consequently, the voltage cycle is divided into phase windows representing the phase angle axis (0 to 360°). If the observations are made for several voltage cycles, the statistical distribution of individual PD events can be determined in each phase window. The mean values of these statistical distributions results in two dimensional patterns of the observed PD patterns throughout the whole phase angle axis [8]. A two-dimensional (2-D) distribution $\phi \sim q$ and $\phi \sim n$ represents PD charge magnitude ' q ' and PD number of pulses ' n ' as a function of the phase angle ' ϕ ' [9]. The mean pulse height distribution $H_{qn}(\phi)$ is the average PD charge magnitude in each window as a function of the phase angle ϕ . The pulse count distribution $H_n(\phi)$ is the number of PD pulses in each window as a function of phase angle ϕ . These two quantity are further divided into two separate distributions of the negative and positive half cycle resulting in four different distributions to appear: for the positive half of the voltage cycle $H_{qn+}(\phi)$ and $H_{n+}(\phi)$ and for the negative half of the voltage cycle $H_{qn-}(\phi)$ and $H_{n-}(\phi)$ [9]. For a single defect, PD quantities can be described by the normal distribution. The distribution profiles of $H_{qn}(\phi)$ and $H_n(\phi)$ have been modelled by the moments of the normal distribution: skewness and kurtosis.

Skewness and Kurtosis are evaluated with respect to a reference normal distribution. Skewness is a measure of asymmetry or degree of tilt of the data with respect to normal distribution. If the distribution is symmetric, $Sk=0$; if it is asymmetric to the left, $Sk>0$; and if it is asymmetric to the right, $Sk<0$. Kurtosis is an indicator of sharpness of distribution. If the distribution has same sharpness as a normal distribution, then $Ku=0$. If it is sharper than normal, $Ku>0$, and if it is flatter, $Ku<0$ [10].

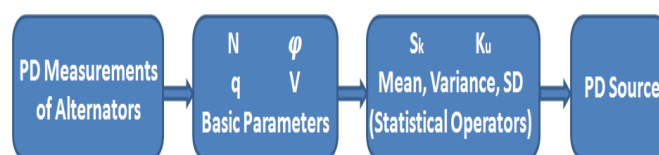


Fig.1. Block diagram of discharge analysis for (n-q)[15-17]

Where,

S.D = standard deviation

Sk = skewness

Ku = kurtosis

Statistical analysis is applied for the computation of several statistical operators. The definitions of most of these statistical operators are described below. The profile of all these discrete distribution functions can be put in a general function, i.e., $y_i=f(x_i)$ [15-17]. The statistical operators can be computed as follows:

Skewness and Kurtosis are evaluated with respect to a reference normal distribution. Skewness is a measure of asymmetry or degree of tilt of the data with respect to normal distribution. If the distribution is symmetric, $Sk=0$; if it is asymmetric to the left, $Sk>0$; and if it is asymmetric to the right, $Sk<0$. Kurtosis is an indicator of sharpness of distribution. If the distribution has same sharpness as a normal distribution, then $Ku=0$. If it is sharper than normal, $Ku>0$, and if it is flatter, $Ku<0$ [11][20].

In our study, we have tested various internal and external discharges like void, surface and corona using statistical parameters such as mean, standard deviation, variance, skewness and kurtosis for (n-q) (as explained above in Fig.1) as a initial step known as pre-processing and a rough draft is made about the results of PD source. Then it is given as an input in Google co-laboratory in Python for model building, GUI will display the type of discharge after entering the five values of statistical parameters as input.

II. GAUSSIAN NAÏVE BAYES CLASSIFIER

The group of classifiers which are build upon Bayes Theorem is called Naive Bayes classifiers, also called simple Bayes and independence Bayes. All the classifiers of this category share the common principle of classification. The reason behind calling it Naive Bayes is its assumption which says that all the attributes of a datasets have no correlation means each attribute is independent of other. Naive Bayes classifiers could be scaled easily. It requires linear parameters in the number of features in a classification problem. Training of Naive Bayes could be done using Maximum-likelihood. In these type of classifiers training is fast because iterative approximation is not done in Naïve Bayes classifiers. Naive Bayes is a simple technique for classification. It models a classifier which assigns a class labels to instances in the testing datasets. Each instance is represented by as a vector of feature values, where the class labels are drawn from training set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle. All Naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For some types of probability models, Naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for Naive Bayes models uses the method of maximum likelihood. Naive Bayes Revised Manuscript Received on December 22, 2018. Shikha Agarwal, Department of Computer Science, Central University of South Bihar, Gaya, India. Balmukund Jha, Department of Computer Science, Central University of South Bihar, Gaya, India. Tishu Kumar, Department of Computer Science, Central University of South Bihar, Gaya, India. Manish Kumar, Department of Computer Science, Central University of South Bihar, Gaya, India. Prabhat Ranjan, Department of Computer Science, Central University of South Bihar, Gaya, India. model could be used without accepting Bayesian probability or using any Bayesian methods. Despite their naive design and simple assumptions, Naive Bayes classifiers have worked quite well in many complex situations. An analysis of the Bayesian classification problem showed that there are sound theoretical reasons behind the apparently implausible efficacy of types of classifiers. A comprehensive comparison with other classification algorithms in 2006 showed that Bayes classification is outperformed by other approaches, such as boosted trees or random forests. An advantage of Naive Bayes is that it only requires a small number of training data to estimate the parameters necessary for classification. The fundamental property of Naive Bayes is that it works on discrete value. If attributes values are continuous it is recommended to use Gaussian Naive Bayes classifier[12].

Naive Bayes is a group of supervised machine learning techniques which are used for classification. The crux of this classification method is Bayes Theorem. It predicts membership probabilities for each class in the dataset such as the probability that given data point belongs to a particular class. The class with the highest membership probability is considered as the most likely class of data point. Classifier identifies in which category a new observation belongs, on the basis of a training set of data[13].

Gaussian Naïve Bayes: GNB is a supervised learning algorithm that is widely used for classification problems because of its simplicity and accurate results. It uses Bayes theorem as its framework and has strong independence assumptions between the independent variables. One important advantage of GNB is that it could estimate the parameters necessary for classification by training on a small training set[14].

The concept of image inpainting was first introduced by Bertamio et al. [1]. The method was inspired by the real inpainting process of artists. The image smoothness information interpolated by the image Laplacian is propagated along the isophote directions, which are estimated by the gradient of the image rotated by 90 degrees.

Exemplar Based method proposed by Criminisi et al. [2] used a best exemplar patch to propagate target patch including missing pixels. This technique uses an approach which combines structure propagation with texture synthesis and hence produced very good results. In [3], the authors decompose the image into sum of two functions and then reconstruct each function separately with structure and texture filling-in algorithms. Morphological technique is used to extract text from the images presented in [4]. In [5], the inpainting technique is combined with the techniques of finding text in images and a simple algorithm that links them. The technique is insensitive to noise, skew and text orientation. The authors in [6] have applied the CCL (connected component labelling) to detect the text and fast marching algorithm is used for Inpainting.

The work in this paper is divided in two stages. 1) Text- Detection 2) Inpainting. Text detection is done by applying morphological open-close and close-open filters and combines the images. Thereafter, gradient is applied to detect the edges followed by thresholding and morphological dilation, erosion operation. Then, connected component labelling is performed to label each object separately. Finally, the set of selection criteria is applied to filter out non text regions. After text detection, text inpainting is accomplished by using exemplar based Inpainting algorithm.

III. METHODOLOGY

1. This research uses google colab for simplifying as it is cheaper and easily accessible being Open source software and could be suitably modified for future scope.
2. Five parameters viz. mean, standard deviation, variance, skewness and kurtosis for both known (void, surface and corona) and unknown discharges (data1, data2 and data3) are given as an input for GNB method (as shown in Fig.2)
3. Initially, in Google Co- laboratory, common model is built. For model building, the obtained statistical parameters are used for both training and testing.
4. Secondly, the final processed file (merging all six known and unknown data) is created.
5. Thirdly, this built model of google colab is taken in python server for creating Graphical User Interface (GUI). GUI will display the type of discharge by entering the statistical parameters viz mean, standard deviation, variance, skewness and kurtosis[18][19] .

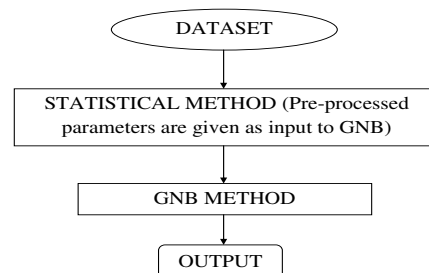


Fig.2. Flow of techniques used

IV. EXPERIMENTAL RESULTS

The output (type of discharge) for the above method viz GNB method will be shown by entering the input values (statistical parameters) as shown in figures below. Fig 3. below is showing the screenshot of window which will be used for entering input values and Fig 4. is used to select the method and Fig 5. shows the output.

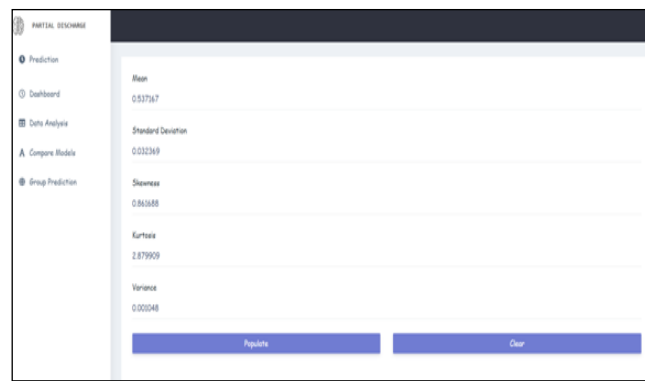
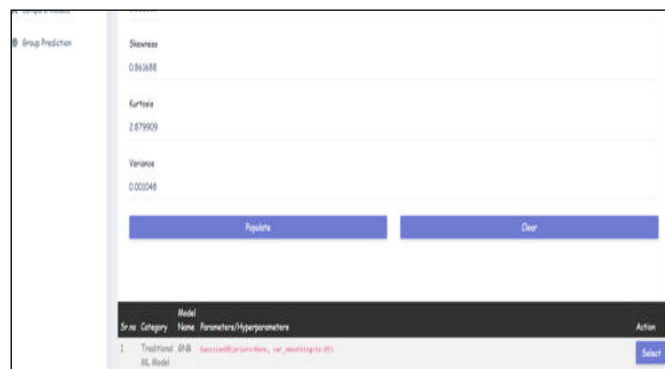
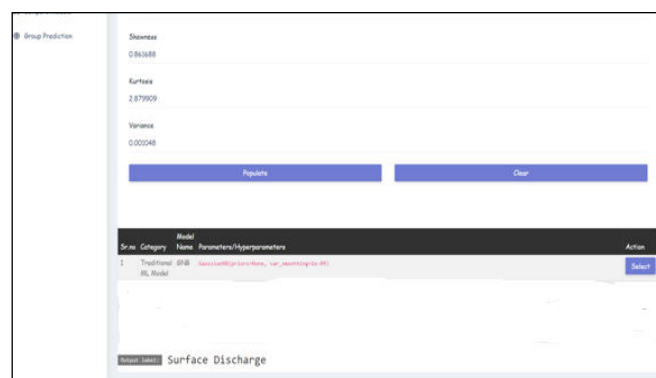


Fig.3. Figure showing the entered input values on GUI



S.no	Category	Model	Name	Parameters/Hyperparameters	Action
1	Traditional GNB	DecisionTreeClassifier	gnb	DecisionTreeClassifier	Select

Fig.4. Figure showing the icon to select the GNB Technique on GUI



S.no	Category	Model	Name	Parameters/Hyperparameters	Action
1	Traditional GNB	DecisionTreeClassifier	gnb	DecisionTreeClassifier	Select

Output: Surface Discharge

Fig.5. Figure showing the output as surface discharge on GUI

Detailed classification report:				
	precision	recall	f1-score	support
0	0.81	0.73	0.77	907
1	0.75	0.84	0.79	908
2	0.97	0.95	0.96	863
accuracy			0.84	2678
macro avg	0.84	0.84	0.84	2678
weighted avg	0.84	0.84	0.84	2678

Fig.6. Figure showing the detailed classification report i.e. precision, recall, f1-score and support of GNB Technique

In Fig.6, "0" stands for known discharge viz. corona discharge, "1" stands for surface discharge and "2" stands for void discharge. Comparative analysis of all the parameters is shown in Fig.7 below.

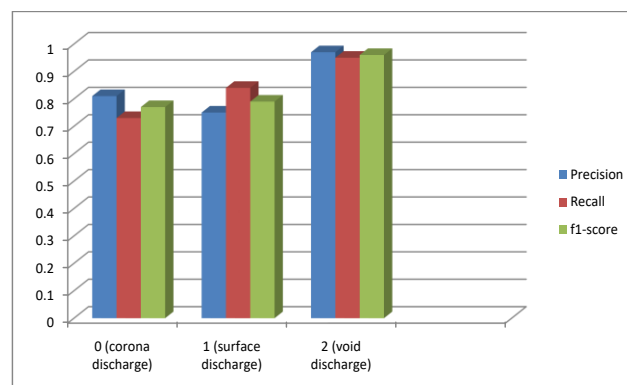


Fig.7. Comparative plot of Precision, Recall and f1-score for all three known discharges from GNB Method

Table I shows the hyper-parameters used in GNB Method. The hyper-parameters in any algorithm contributes greatly to the output of the model, therefore, determining the optimal (or near-optimal) combination of hyper-parameters would yield the best result. Hyper-parameters are defined as the properties of a model that the user can set the value to. They are different from parameters as parameters are changed internally by the model itself during training rather than set by the user before the training process.

Sr.No	Hyperparameters	Description	Value
1	priors	Prior probabilities of the classes.If specified the priors are not adjusted according to the data	None
2	var smoothing	Portion of the largest variance of all the features that is added to variances for calculation stability	1.00E-09

Table I. Table showing hyper-parameters used in GNB Method

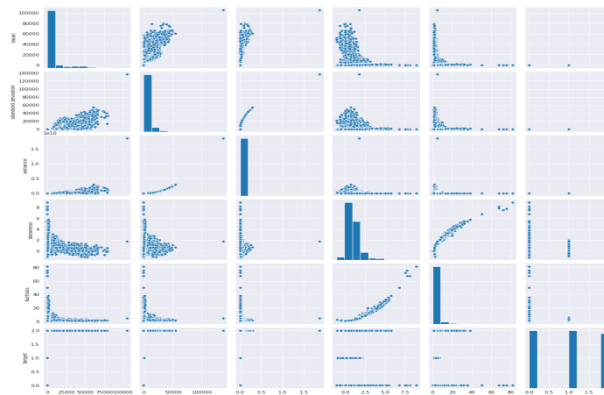


Fig.8. Attributes distributions and statistical analysis of the used dataset

Fig.8. shows the distributions and statistical analysis of the dataset used as we can see from the last column and last row figure that how we can reach to target.

```

▼ Naive Bayes

[ ] from sklearn.naive_bayes import GaussianNB

[ ] gnb = GaussianNB()
    gnb.fit(X_train, y_train)

    GaussianNB(priors=None, var_smoothing=1e-09)

[ ] print("Detailed classification report:")
    y_true, gnb_pred = y_test, gnb.predict(X_test)
    print(classification_report(y_true, gnb_pred))

    confusion = confusion_matrix(y_test, gnb_pred)
    print('Confusion Matrix:')
    print(confusion)
    
```

Fig.9. Steps for execution of GNB method on google co-laboratory window

Above figure shows the programming steps for execution of the GNB method, after uploading the pre-processed data obtained from statistical method, click on these 3 steps so that the program will get executed and output can be seen on GUI (as shown in Fig.3, Fig 4 and Fig.5)

IV. DISCUSSION AND CONCLUSION

Unknown data	Partial discharge Source
Data 1	Surface
Data 2	Surface
Data 3	Void

Table II. Results of GNB Method

It was seen that results of unknown data, data2 were not clearly confirming the source of discharge as void or surface using Statistical method in MATLAB software. But now, it can be finally concluded using GNB method using Python software that the data2 discharge is definitely a surface discharge. The accuracy from these method is 87.26%.

Created a common model and confirmed the results accurately for type of PD. Use of google co-laboratory gives accurate results. The advantage of google co-laboratory in Python is to reduce the error to zero faster.

Advantage of using GNB is to increase the speed of PD type recognition and avoid a lot of power supply failures in industry. GNB can be considered as the best method for detecting type of PD since it is giving accurate results.

Finally it can be concluded as --For detection of type of partial discharge will suggest to use GNB method.

REFERENCES

1. Statistical Pattern Analysis of Partial Discharge Measurements for Quality in High Assessment of Insulation Systems Voltage Electrical Machinery Birsan Yazici Symposium on Diagnostin far Electric Machines, Power Electronics and Drives Atlanta, Georgia, GA USA 24-26, August 2003.
2. Akimasa Hirata, Syou Nakata, and Zen-Ichiro Kawasaki, "Toward Automatic Classification of Partial Discharge Sources With Neural Networks", IEEE transactions on power delivery, vol. 21, no. 1, january 2006.
3. J. R Hyde and I. J. Kemp, "Partial Discharge Pattern Recognition and Associated Technologies", IEEE Colloquium on Monitoring Technologies for Plant Insulation, 1994.
4. S.H.Park, K. W.Leel, K.J.Lim. and S.H.Kang, "Classification of External and Internal PD Signals Generated in Molded Transformer by Neural Networks", Proc. of the 7th Int. Conf. on Properties and Applications of Dielectric Materials June 1-5 Nagoya P2-41, 2003.
5. Shan Ping', XuDake', Wang Guolil, Li Yanming, "Application of Neural Network with Genetic Algorithm to UHF PD Pattern Recognition in Transformers", Annual Report Conference on Electrical Insulation and Dielectric Phenomena, 2002.
6. E. Gulski and F. H. Kreuger, "Computer-aided recognition of Discharge Sources", IEEE Trans. on Electrical Insulation Vol. 27 No. 1, February, 2001.
7. S. Ghosh and N.K. Kishore, "Modelling Of Partial Discharge Inception And Extinction Voltages Of Sheet Samples Of Solid Insulating Materials Using an Artificial Network", Proc. Of IEEE Volume: 149, Issue: 2 2002.
8. Gagan Deep Meena, Dr. Girish Kumar Choudhary, Manoj Gupta, "Neural Network Based Recognition Of Partial Discharge Patterns", International Journal of Advanced Engineering Research and Studies. Vol. I/ Issue II/January-March, 2012/121-126.
9. V. M. Catterson, B. Sheng, "Deep Neural Networks for Understanding and Diagnosing Partial Discharge Data", Institute for Energy and Environment, University of Strathclyde, Glasgow, United Kingdom.
10. M. M. A. Salama, and R. Bartnikas, "Determination of Neural-Network Topology for Partial Discharge Pulse Pattern Recognition", IEEE Transactions On Neural Networks, Vol. 13, No. 2, March 2002.
11. Radio Location of Partial Discharge Sources: A Support Vector Regression Approach K. P. Bennett, and C. Campbell, "Support vector machines: hype or hallelujah," ACM SIGKDD Explorations Newsletter, vol. 2, no. 1, pp. 1-13, 2000.
12. M. Junshui, T. James and P. Simon, "Accurate On-line Support Vector Regression", Neural Computation, vol. 15, no. 11, pp. 2683-- 2703, 2003.
13. Shikha Agarwal, Balmukumd Jha, Tisu Kumar, Manish Kumar, Prabhat Ranjan, "Hybrid of Naive Bayes and Gaussian Naive Bayes for Classification: A Map Reduce Approach", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-6S3, April 2019.
14. Dana Bani-Hani, Pruthak Patel & Tasneem Alshaikh, "An Optimized Recursive General Regression Neural Network Oracle for the Prediction and Diagnosis of Diabetes", Global Journal of Computer Science and Technology: D Neural & Artificial Intelligence Volume 19 Issue 2 Version 1.0 Year 2019.
15. Namrata Bhosale, Priyanka Kothoke Amol Deshpande, Dr. Alice Cheeran, "Analysis of Partial Discharge using Phase-Resolved(ϕ -q) and (ϕ -n) Statistical Techniques", International Journal of Engineering Research and Technology, Vol. 2 (05), 2013, ISSN 2278-0181.
16. Priyanka Kothoke, Namrata Bhosale, Amol Deshpande, Dr. Alice Cheeran, "Analysis of Partial Discharge using Phase-Resolved (n-q) Statistical Techniques", International Journal of Engineering Research and Applications.
17. Yogesh R. Chaudhari, Namrata R. Bhosale, Priyanka M. Kothoke, "Composite Analysis of Phase Resolved PD Patterns using Statistical Techniques", International Journal of Modern Engineering Research (IJMER) Vol. 3, Issue. 4, pp-1947-1957, 2013.
18. . Priyanka Kothoke, Dr. Anupama Deshpande, Yogesh Chaudhari, "Investigation and Determination of PD source utilizing SVR and RF", International Conference on Inventive Computation Technologies (ICICT-2020), inclusion into IEEE Xplore and Scopus.



19. Priyanka Kothoke, Dr. Anupama Deshpande, Yogesh Chaudhari, “Examination and Determination of Partial Discharge Source utilizing SOM and BPM Techniques of ANN”, International Journal of Advanced Science and Technology, Vol. 29, No. 8s, (2020), pp. 3299-3306.
20. Priyanka Kothoke, Dr. Anupama Deshpande, Yogesh Chaudhari, “Analysis and Determination of Partial Discharge Type using Statistical Techniques and Back Propagation Method of Artificial Neural Network for Phase Resolved Data;”, International Journal of Engineering Research and Technology(IJERT),Vol. 08, Issue 08, August 2019.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details