International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

Analysis of Real Time Fault Data of Multi Terminal Transmission System using Python Learning Tools

Gyanesh Singh, A.Q. Ansari, Md.Abul Kalam

Abstract: ***Transmission line faults are common difficulties in today's world. Faults mainly depend on types of load and their nature. Transmission lines are divided in different zones and hence cannot predict easily the faults and their types. Several protective devices have incorporated in the past few years to classification of fault but none guaranteed the accuracy. In this paper author proposes an advanced machine learning algorithm to classification of faults and provided some future protection technique to minimize faults and reliability of supply to the consumer. In this python learning tools author compare the two algorithm namely using KNeighbors Classifier and Using Multinomial Logistic Regression. The data for experiments are obtained from BBMB Punjabi Bagh 220 KV substation New Delhi. The real experiments validates the improvements and future forecasting regarding faults in lines and analysis of results with their accuracy are elaborated in this paper. A discussion of real data and their implication in future conservation of energy.***

Index Terms: ***Power system real faults, Python, KNeighbors Classifier and Using Multinomial Logistic Regression algorithm. Python and MS Excel.Python libraries numpy, pandas, matplotlib, seaborn, IDEs Jupyter Notebook.***

I. INTRODUCTION

The data for each sample or event has been provided in a separate excel sheet, which is not suitable for tradition analysis so we need to gather data manually from the sheet in a single sheet for the analysis with features as columns and each event information as rows. There are several observations for each event so we need to choose what is relevant to our purpose and may act as a feature. The prominent looking features in the data are **Voltage, Load of Line** and **Frequency**. Features like **Line Trip on other End**, **Other Circuit, Date-Time, Weather** also seemed relevant. Some features like **Repairs Carried** and **Observation of the site** may or may not have been very much relevant but selected at first. Other information about data was mostly not available or missing. After we collected the data as in required way, we came to around 40 events of faults. We simply left the cells empty for the observation we didn't have any value for a feature [1-9].

II. PROPOSED LAYOUT OF POWER SYSTEM

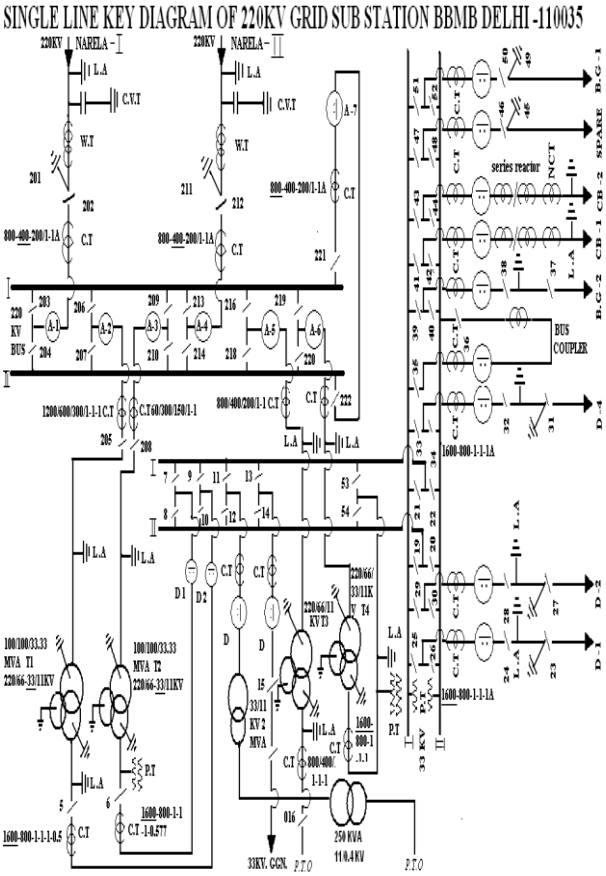


Fig.1. General layout of Power System

Revised Manuscript Received on 30 May 2019.
* Correspondence Author
Gyanesh Singh, Electrical & Electronics Engineering Department, IMS Engineering College Ghaziabad Uttar Pradesh India
A.Q. Ansari, Electrical Engineering Department Faculty of Engineering JMI New Delhi Jamia Millia Islamia New Delhi India
Md.Abul Kalam, Electrical Engineering Department JSSATE NOIDA Uttar Pradesh India.

Table 1 Real Data observed from 220 KV Grid Substation Punjabi Bagh New Delhi

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
date of trip	time of t	weather	other cir	line trip	load of li	voltage	frequency	other lin	date of restoration	time of restoration	observation	repair carried	tripping reason	type of fault
08-11-2017	05:39	cloudy	healthy	yes	71	237	49.87		08-11-2017	07:50		nil	bad weather	low
10-05-2017	15:56	cloudy	healthy	yes	137	220	49.98		10-05-2017	16:34		nil	transient fault	high
11-10-2017	12:35	clear	healthy	yes	124	229	49.94		11-10-2017	13:45		nil	foreign element	medium
18-06-2017	05:37	clear	healthy	yes	165	228	50.01		18-06-2017	06:17		nil	foreign element	medium
19-06-2017	06:30	rainy	healthy	yes	102	229	50.05		19-06-2017	07:10	disc puncture	feeder isolated	transient fault	medium
20-06-2017	17:04	cloudy	healthy	no	137	221	50.03		20-06-2017	17:10	relay burnt	relay removed	relay burn	medium
24-07-2017	11:32	cloudy	healthy	yes	173	220	49.98		24-07-2017	12:07		nil	transient fault	high
29-06-2017	13:51	cloudy	healthy	yes	100	227	50.03		29-06-2017	14:39		nil	foreign element	high
31-12-2017	07:38	foggy	healthy	yes	67	235	50.06		31-12-2017	09:26		nil	bad weather	low
11-06-2017	03:49	clear	healthy	no	146	232	50.01		11-06-2017	04:20		fuse	fuse failure	medium
11-06-2017	05:39	clear	healthy	no	135	233	49.98		11-06-2017	09:10		fuse	fuse failure	medium
10-06-2017	07:30	clear	healthy	no	122	231	50.03		10-06-2017	08:14		nil	fuse failure	medium
04-04-2014	07:40	clear		yes	140	230	50.06		04-04-2014	11:32			earth fault	medium
30-05-2014	16:52	cloudy	healthy	no	230	226	50		30-05-2014	18:11			earth fault	high
26-07-2014	18:22	cloudy	healthy	yes	155	229	49.91		26-07-2014	19:00			transient fault	high
27-06-2014	19:35	clear	healthy	no	215	223	49.85		27-06-2016	19:29			transient fault	high

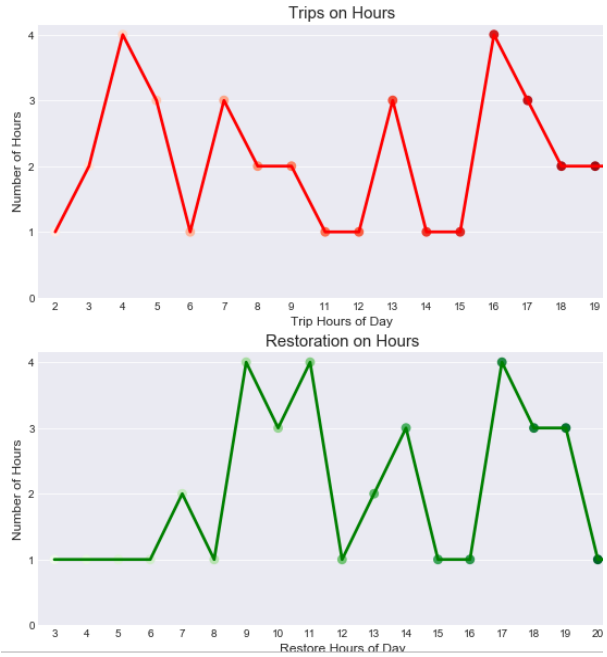


Fig.2. Trip of circuit breaker in particular time duration

voltage vs frequency

```
sns.jointplot(x = data['voltage'], y = data['frequency'], kind = 'reg', color = 'blue')
plt.show()
```

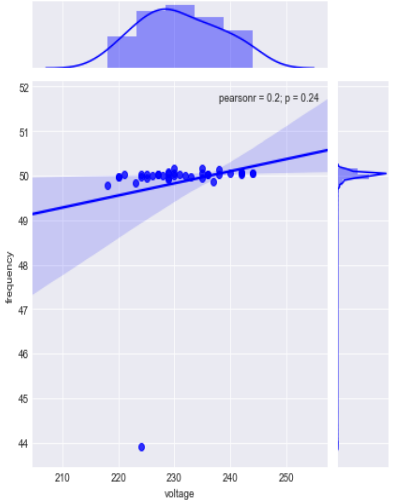


Fig.3. Frequency v/s voltage plot

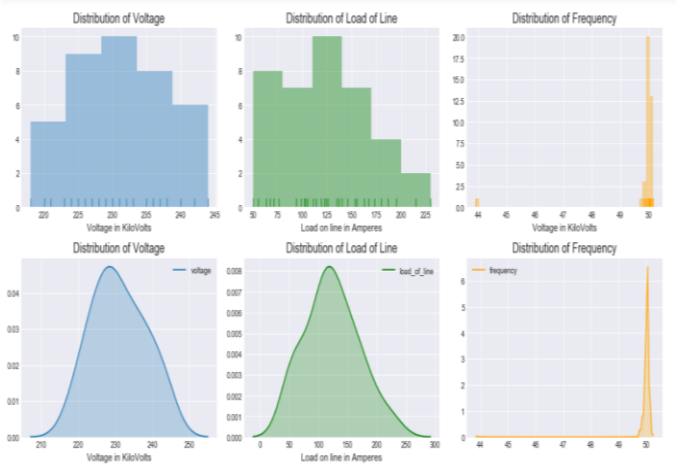


Fig.4. Distribution of voltage, frequency and load plot

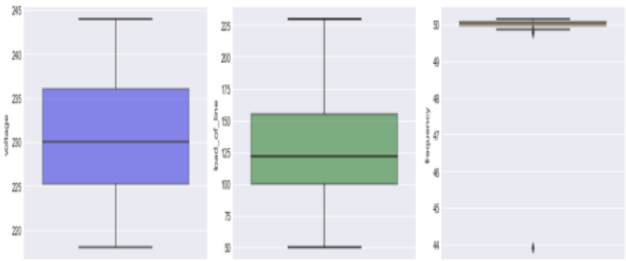


Fig.5. load v/s voltage plot

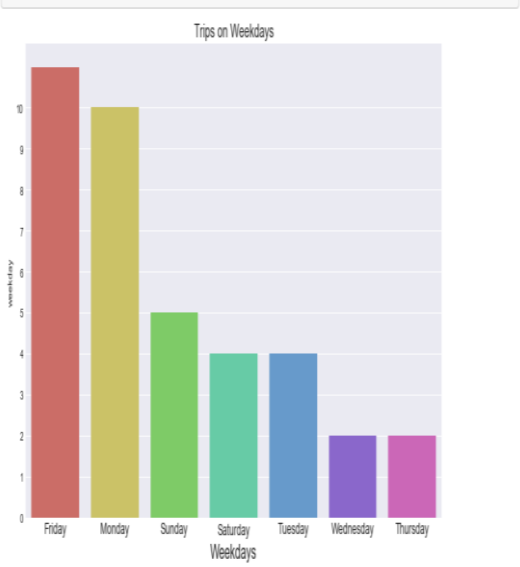


Fig.6. Tripping of week days plot

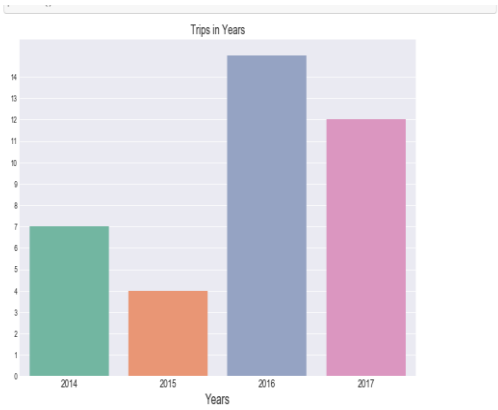


Fig.7. Tripping of years plot

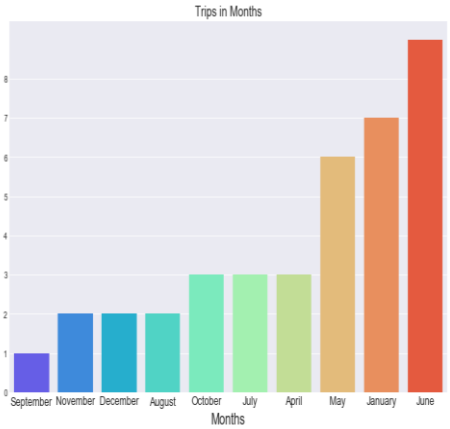


Fig.8. Tripping of months plot

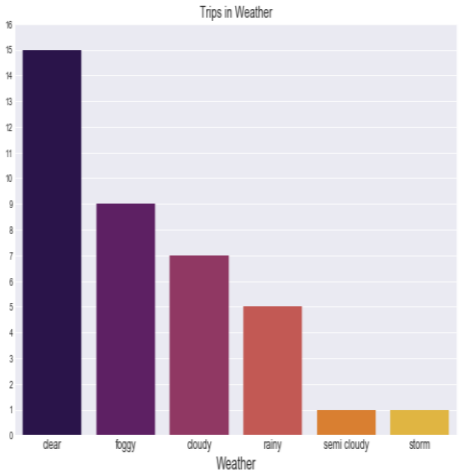


Fig.9. Tripping of weather plot

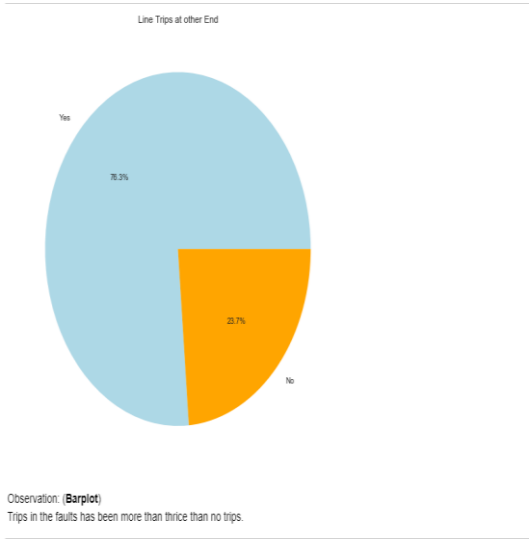


Fig.10. percentage of tripping plot

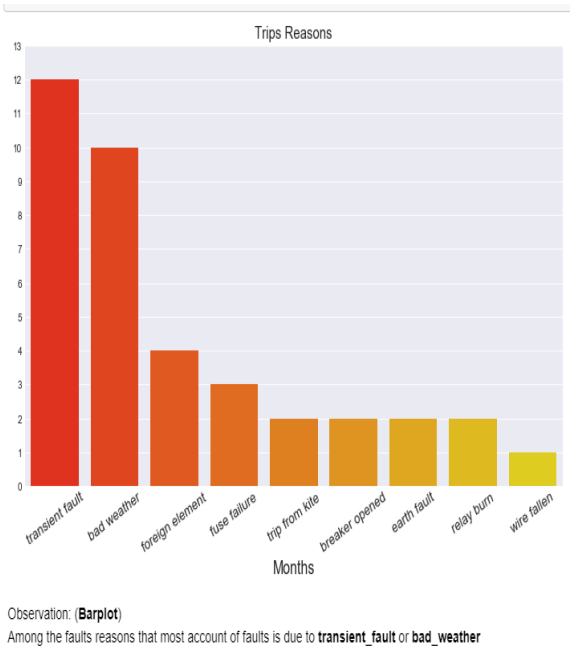
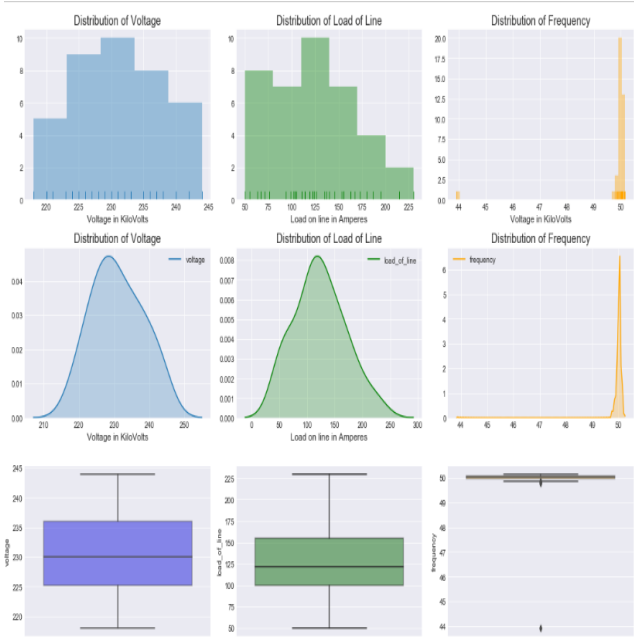


Fig.11. Tripping reason of months plot



Observation: (**Histogram, Kernel Plot and Box Plot**)

Fig.12. Histogram, kernel plot and box plot for load, frequency and voltage

Voltage vs load_of_line

```
sns.jointplot(x = data['load_of_line'], y = data['voltage'], kind = 'reg', color= 'g')
plt.show()
```

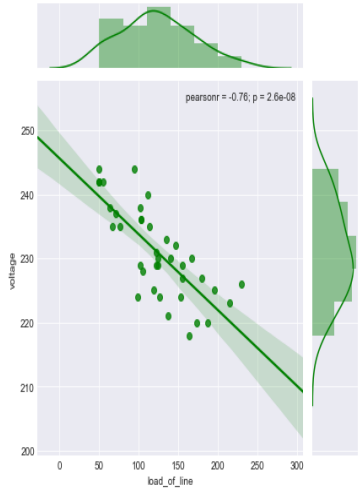


Fig.13. Voltage v/s load line plot in python toolsUsing KNeighbors Classifier

Table 2 Accuracy train test set

Accuracy on Train Set

```
print(classification_report(y_train, classifier.predict(X_train)))
```

	precision	recall	f1-score	support
-1	1.00	1.00	1.00	7
0	0.83	1.00	0.91	10
1	1.00	0.82	0.90	11
avg / total	0.94	0.93	0.93	28

So we got approximately 94% accuracy on the train set.

Table 3 Accuracy test set

Accuracy on Test Set

```
from sklearn.metrics import classification_report
cr = classification_report(y_test, y_test_pred)
print(cr)
```

	precision	recall	f1-score	support
-1	1.00	0.67	0.80	3
0	0.88	1.00	0.93	7
avg / total	0.91	0.90	0.89	10

So we got approximately 91% accuracy on the train set.

Table 4 Accuracy train test

Using Multinomial Logistic Regression

Accuracy on Train Test

```
print(classification_report(y_train, classifier.predict(X_train)))
```

	precision	recall	f1-score	support
-1	0.88	1.00	0.93	7
0	1.00	0.80	0.89	10
1	0.92	1.00	0.96	11
avg / total	0.94	0.93	0.93	28

So we got approximately 94% accuracy on the train set.

Table 5 Accuracy test set

Accuracy on Test Set

```
from sklearn.metrics import classification_report
cr = classification_report(y_test, y_test_pred)
print(cr)
```

	precision	recall	f1-score	support
-1	0.75	1.00	0.86	3
0	1.00	0.86	0.92	7
avg / total	0.93	0.90	0.90	10

So we got approximately 93% accuracy on the train set.

From the data we clearly observed that feature like other end only had one value for all events which was '**Healthy**', or was either missing so any such feature which has so high certainty and very less information gain doesn't improve our model so we drop it. We dive into analysis, first we looked for columns with missing values, **other_line_status** and

Exploratory Data Analysis (EDA)

We began with parsing the dates and time, **date_of_trip**, **date of restoration, time_of_trip,** and **time_of_restoration** we separated the months and years from the dates, and separated hours from the time, forming new features (columns) in the data. These months and days were initially parsed as numbers which were mapped to their respective names. Then we plotted the charts such as bar charts or count plots for these features and made some observations:

- We found that most fault were on Fridays and Mondays.
- Years were 2016 and 2017
- Most frequent months of faults were January, May and June.

Then we plotted the **weather** for the faults.

- The most faults were during clear and foggy weather

When checked for **line trips** at other end we found

- There were no line trips 24% and line trips 76% of the times, more than thrice than no trips.

Then we looked for the most frequent reasons observed for the faults and concluded

Most faults were transient_faults or caused by bad_weather.

Next, we moved on to our prime three features **Voltage, load_of_line** and **frequency.** The histograms, box distributions and kernel plots provided that **Voltage** and **load_of_line** have bell-shaped curve (follow normal distribution) and have satisfactory spread, no skewness of data or like that. However the frequency had very sharp peaked distribution and less variance[10-14]. So ,we inferred from it that it may be important now to scale our data features before modelling in order to avoid any dominance of features over other. Then we plotted these prime features against each other in pairs and out of three one very important observation was made from the Voltage vs load_of_lineplot which showed a pearson correlation of -0.76 between them. Such negative correlation is considered to be a strong negative correlation which implies that higher one feature (either **observation** had almost no data values so we dropped them right away. There were significant missing values in **repairs** and very few missing ones in **tripping_reason** so we can fill the missing values in **tripping_reason**with themost frequent reason for tripping. We filled in the missing values in the **repairs** column as nil, which implies no repairs were carried. voltage or load_of_line) leads to lower value of other. Now we know that features which are strongly correlated are linearly dependent, that is information about one can be obtained by other. So, we can't involve any feature which provides information that can be given by other in order to avoid any multi collinearity, so we drop **voltage** as a feature for our model but keep **load_of_line.**

The next we visualized are the tripping hours to see any patterns in faults during the span of 24 hours.

- The faults showed two higher peaks during the very early morning and dusk time. But there are not very much data to ensure that as a pattern. So we rule it out as well.

Further visualization included **Repair** types for the faults.

- There mostly no repairs provided other than that there were at most 2 counts for any other repair value. So, this feature could also be considered to be victim of less data.

Predictive Modelling

The final features worth the model were three – **line_trip, load_of_line** and **frequency.** Next, we mapped the labelled target variable **type_of_fault**from string values (low, medium and high) to -1, 0 and 1. To make them suitable for model calculations. We also mapped the **line_trip** values from (yes and no) to 1 and 0.

We used Standardization Equation to scale our data

$$z=\frac{x-\mu}{\sigma} \qquad \textbf{(i)}$$

μ= Mean

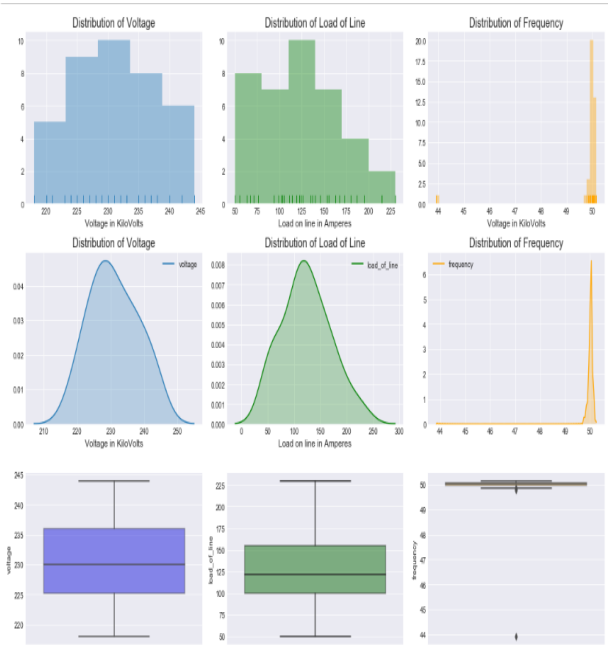
σ=Standard Deviation

We finally came up with two predictive models one working on the algorithm **KNeighbors Classifier** and **Multinomial Logistic Regression**. Both provided a very good accuracy on the training data, expected. However we only have only few observations so we cannot test it on any validation set or splitting training set into two parts. Final built model is able to predict the type of fault for any single observation as well, asking for three values, **line_trip, load_of_line** and **frequency.** Most of our features have not been very fruitful but that were mainly due to lack of sufficient data, in presence of enough data this analysis may be very productive for insightful purposes. Note, this model is able to predict the type of fault provided that there is a fault occurred definitely but can't predict if there will be a fault or not as the data only contains faults' observations only.[15-22]

III. CONCLUSION

We compared two algorithm namely Using KNeighbors Classifier and Using Multinomial Logistic Regression and we got good accuracy in terms of training and testing of data by multinomial logistic regression.

APPENDIX



Observation: (**Histogram, Kernel Plot and Box Plot**)

ACKNOWLEDGMENT

I acknowledge Mr. S.K. Pandey Sr. Engineer transmission and their team for providing me real time fault data and information of transformer data for calculation, classification and regression analysis.

REFERENCES

1. Adrian Stetco, Fateme Dinmohammadi, Xingyu Zhao, Valentin Robu, David Flynn, Mike Barnes, John Keane, Goran Nenadic Machine learning methods for wind turbine condition monitoring: A review Renewable Energy, Volume 133, 2019, pp. 620-635.
2. Hongshan Zhao, Huihai Liu, Wenjing Hu, Xihui YanAnomaly detection and fault analysis of wind turbine components based on deep learning network Renewable Energy, Volume 127, 2018, pp. 825-834.
3. Alberto Diez-Olivan, Javier Del Ser, Diego Galar, BasilioSierra
4. Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0 Information Fusion, Volume 50, 2019, pp. 92-111.
5. PhongB. Dao, Wieslaw . Staszewski, Tomasz Barszcz, Tadeusz Uhl Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data Renewable Energy, Volume 116, Part B, 2018, pp. 107-122.
6. Georg Helbing, Matthias Ritter Deep Learning for fault detection in wind turbines Renewable and Sustainable Energy Reviews, Volume 98, 2018, pp. 189-198.
7. Estefania Artigao, Sergio Martín-Martínez, AndrésHonrubia-Escribano, Emilio Gómez-Lázaro Wind turbine reliability: A comprehensive review towards effective condition monitoring development Applied Energy, Volume 228, 2018, pp. 1569-1583.
8. S. S. S. Rawat, V. A. Polavarapu, V. Kumar, E. Aruna, V. Sumathi, "Anomaly detection in smart grid using rough set theory and K cross validation", Proc. Int. Conf. Circuits Power Comput. Technol., pp. 479-483, Mar. 2014.
9. Y. Cui, J. Shi, Z. Wang, "Power system fault reasoning and diagnosis based on the improved temporal constraint network", IEEE Trans. Power Del., vol. 31, no. 3, pp. 946-954, Jun. 2016.
10. Y. Tao, J. Zheng, T. Wang, Y. Hu, "A state and fault prediction method based on RBF neural networks", Proc. IEEE Workshop Adv. Robot. Social Impacts, pp. 221-225, Jul. 2016.
11. E. Rakhshani, I. Sariri, K. Rouzbehi, "Application of data mining on fault detection and prediction in boiler of power plant using artificial neural network", Proc. Int. Conf. Power Eng. Energy Elect. Drives, pp. 473-478, Mar. 2009.
12. M. Mahdi, V. M. I. Genc, "Artificial neural network based algorithm for early prediction of transient stability using wide area measurements", Proc. 5th Int. Istanbul Smart Grid Cities Congr. Fair, pp. 17-21, Apr. 2017.
13. L. Mou, P. Ghamisi, X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification", IEEE Trans. Geosci. Remote Sens., vol. 55, no. 7, pp. 3639-3655, Jul. 2017.
14. H. Zuo, H. Fan, E. Blasch, H. Ling, "Combining convolutional and recurrent neural networks for human skin detection", IEEE Signal Process. Lett., vol. 24, no. 3, pp. 289-293, Mar. 2017.
15. T. Nakashika, T. Takiguchi, Y. Ariki, "Voice conversion using RNN pre-trained by recurrent temporal restricted Boltzmann machines", IEEE/ACM Trans. Audio Speech Language Process., vol. 23, no. 3, pp. 580-587, Mar. 2015.
16. S. Leglaive, R. Hennequin, R. Badeau, "Singing voice detection with deep recurrent neural networks", Proc. IEEE Int. Conf. Acoust. Speech Signal Process., pp. 121-125, Apr. 2015.
17. V.-T. Tran, K.-H. Nguyen, D.-H. Bui, "A vietnamese language model based on recurrent neural network", Proc. 8th Int. Conf. Knowl. Syst. Eng., pp. 274-278, Oct. 2016.
18. C. Xu, G. Wang, X. Liu, D. Guo, T.-Y. Liu, "Health status assessment and failure prediction for hard drives with recurrent neural networks", IEEE Trans. Comput., vol. 65, no. 11, pp. 3502-3508, Nov. 2016.
19. T. de Bruin, K. Verbert, R. Babuska, "Railway track circuit fault diagnosis using recurrent neural networks", IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 3, pp. 523-533, Mar. 2017.
20. M. Yuan, Y. Wu, L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network", Proc. IEEE Int. Conf. Aircraft Utility Syst., pp. 135-140, Oct. 2016.
21. Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast", IET Intell. Transp. Syst., vol. 11, no. 2, pp. 68-75, Jan. 2017.
22. S. U. Jan, Y. D. Lee, J. Shin, I. Koo, "Sensor fault classification based on support vector machine and statistical time-domainfeatures", IEEE Access, vol. 5, pp. 8682-8690, May 2017.
23. Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery", IEEE Access, vol. 5, pp. 15066-15079, Jul. 2017.
24. Y. Wang, M. Liu, Z. Bao, "Deep learning neural network for power system fault diagnosis", Proc. 35th Chin. Control Conf., pp. 6678-6683, Jul. 2016.

AUTHORS PROFILE

Gyanesh Singh Assistant professor IMS Engineering College, completed BE degree in Electrical Engineering in 2002 from AKGEC GZB, Master in Electrical Engineering from NITTTR, Chandigarh in 2011, Persuing Phd in Electrical Engineering From Jamia Millia Islamia Central University new delhi. Research area in power system fault, protection of transmission lines multiterminal. More than 7 papers in National and International Journal, International conferences and national conferences. 17 years of teaching experience in the field of electrical engineering, Teaching undergraduate students in subject like Electrical engineering, Measurement and Instrumentation, Network analysis and synthesis, Electrical Machinery, Control System, Power electronics, Power system ,Electric Drives etc.

Prof. A. Q. Ansari is an astute performer with proven expertise in devising and effectuating programmes and policies aimed at ensuring smooth operations and execution of academic, research and administration tasks within the institute premises, possesses potential for distinctive achievement through strategy, innovation, implementation and control.

He is a stickler for quality, team builder to the core and a natural motivator with perseverance and integrity. He commands excellent communication skills that have been honed through interacting with people at national and international levels. He is popular among his students for excellent content delivery, clarity of presentation, and extraordinary punctuality and discipline.

His research contributions are in the areas of Mobile Adhoc Networks, Multimodal Biometrics, Networks-on-Chip, and Fuzzy Logic and its Variants. His mathematical formulations for the "Fuzzification of Intuitionistic Fuzzy Sets" have been widely appreciated. He has undertaken several prestigious research and consultancy assignments. Having completed three R & D Major Research Projects, he has got hands on experience of providing consultancy to one of the leading system integration companies of India, ONNYX Electronics, dealing in installation and maintenance of Traffic Signals in major cities of the country.

Prof. Ansari has successfully guided eleven PhDs, twelve M. Tech. dissertations, and several B. Tech., B.E., M.C.A, and B.C.A. projects and has produced excellent results with proven records through two patent applications, 44 Peer Reviewed International Journal papers, one book each written and edited, 5 Book Chapters, and 77 Peer Reviewed International Conference papers. He also worked as Guest Editor of the International Journal of Embedded Systems, special issue on Emerging Trends in On-Chip Communications, published by InderScience, Switzerland.

Prof. Ansari has many firsts to his credit in Jamia: He is the first person to have applied for a patent in the name of Jamia Millia Islamia in the year 2011. His two patents are under consideration these days in the Indian Patent Office. He was the one who established the IEEE Students' Branch in Jamia in the year 1992 and worked as its Founder Counselor for seven long years. He also established the Students' Chapters of IETE and ISTE in Jamia in the years 1990 and 1994 respectively.

Prof. Ansari is a Senior Member of IEEE, Fellow and Chartered Engineer of the Institution of Engineers and IETE. He is the recipient of the Rajarambapu Patil National Award for Promising Engineering Teacher for Creative Work Done in Technical Education for the year 2011 awarded by the Indian Society for Technical Education, New Delhi, which is a testimony to the fact that he is a person with excellent experience in Policy Planning in Education, Curriculum Development, Human Resource Management, and Innovative Design of Motivational Techniques.

Dr. Md. Abul Kalam is Associate Prof. with the Electrical Engg. Deptt. JSSATE Noida(U.P). He has received his BE Degree in Electrical Engg. from Jamia Millia Islamia, New Delhi in 1999 and M.Tech. From UPTU, Lucknow in 2007. He received Ph.D degree in 2014 in Power system. from Jamia Millia Islamia, New Delhi. His area of Interest are Power System, Intelligent Techniques based Protection of Transmission System and Energy. He has published/presented many research papers in National/International Journals/conferences. Faults can occur in any components of power system including generator, transformer, buses and transmission lines. The transmission line is one of the main component of power system and being exposed to the environmental conditions, the possibility of experiencing faults on the transmission system is generally higher than that on other components, therefore the detection of faults on transmission lines is the primary concern, necessitating high speed fault clearance and improved transient stability. The time needed to determine the fault point along the line affect the quality of the power delivery. Therefore, to maintain the efficient and reliable operation of power systems, it is extremely important that the transmission line faults need to and located in a reliable and accurate manner as fast as possible.

The transmission line being geographically wide distributed systems and exposed to the atmosphere there is uncertainty about the fault events and the line parameters. Therefore, in such situation, where there is ambiguity and uncertainty about the fault events and the line parameters, it is difficult to deal with the transmission lines protection problems through strict mathematical approaches (deterministic approaches) effectively.

Keeping in view of uncertainty in the fault events and the line parameters, in the present research work, application of intelligent techniques such as fuzzy logic, artificial neural network (ANN) and generalized neural network (GNN) along with signal processing tools like discrete Fourier and wavelet transforms are proposed for fault detection, faulty phase identification and for fault location estimation of transmission system for protective relaying purposes.

In the proposed research work, wavelet transform is used as a features extractor of transient signals (fault current signals) of the faulted transmission system because of its ability to extract time frequency information from the transient signal to detect and to identify the faulty phase and to locate the fault of a given transmission system. The task of faults classification of the transmission line is carried out extracting the features of the fault current signals using Discrete Fourier transform (DFT) at the fundamental frequency phasors of the fault current signals. The obtained features are applied as input parameters to the fuzzy logic system and output as types of fault.

An artificial neural network (ANN) model which provides computationally efficient way of determining an empirical, possibly non linear relationship between a number of inputs and output is developed in the present work and used for fault location estimation of the transmission system. To overcome some of the problems of ANN and to improve its training and testing performance, the simple neuron is modified and a generalized neural network (GNN) model is developed, to estimate the fault location of the given transmission system.

In this study five methods such as conventional method based fault location estimation wavelet transform based fault detection and faulty phase identification, Fuzzy logic and discrete Fourier transform based faults type classification, wavelet-ANN and wavelet-GNN model based fault location estimation of the three phase transmission system have been investigated and it is found that the distinctive characteristics features of wavelet signals detect the fault and identify the faulty phase successfully, and the fuzzy logic approach is capable to identify all the phase to phase and ground faults and the generalized neural network (GNN) is found more efficient and the better technique to estimate fault location of the transmission line as compared to the conventional mathematical model and the artificial neural network (ANN) for a wide variation of fault operating conditions and the obtained results found quite satisfactory. Moreover, GNN requires less data to analyze, therefore the time taken for identifying the location of the fault is quite less using GNN model.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication