Local Outlier Factor based Data Mining Model for Three Phase Transmission Lines Faults Identification

Tan Yong Sing
Faculty of Engineering and Computer Technology, AIMST University
Bedong, Kedah, Malaysia

Emran bin Siraj Syahrel
Faculty of Engineering and Computer Technology, AIMST University
Bedong, Kedah, Malaysia

Pratap Nair Marimuthu
Faculty of Engineering and Computer Technology, AIMST University
Bedong, Kedah, Malaysia

Raman Raguraman
Faculty of Engineering and Computer Technology, AIMST University
Bedong, Kedah, Malaysia

K. Nithiyananthan, PhD
Department of Electrical and Electronics Engineering
Karpagam College of Engineering
Coimbatore, India

ABSTRACT
The main objective of this paper is to design and developed a model of power system transmission lines fault identification using Local Outlier Factor (LOF) technique based on data mining. 9 bus power system and 30 bus power systems overhead transmission line fault of are simulated as a test cases using PowerWorld software and the LOF computations using Matlab software. The proposed model consists of two main steps; the first step is used to determine single line-to-ground (SLG) fault and 3 phase fault. The second step was used to determine line-to-line (LL) fault and double line-to-ground (DLG) fault. The second step is required uninterrupted phase voltage in order to distinguish between LL fault and DLG faults. The proposed model successfully identified asymmetrical faults and symmetrical faults that particular single bus fault.

General Terms
Local Outlier Factor, Data Mining Model

Keywords
Power system, asymmetrical faults, symmetrical fault, Transmission lines

1. INTRODUCTION
Transmission lines in power system required constant delivery of electrical power to industries and the end consumers. The numbers of overhead transmission lines keep on increasing as the power system network become more distributed which is results in transmission lines fault identification and become more complicated. Due to huge amount of data collected from the data acquisition system, it may slow down the transmission lines restoration process. This may cause huge impact on economics of the particular areas and safety of end consumers. Hence, by utilizing the overflow data or information acquired from the database, transmission lines fault identification can be done easily, expeditiously and accurately with Data Mining methods. Data Mining is one of the processes in Knowledge Discovery in Database (KDD). Useful information and hidden knowledge are discovered from the raw data. These techniques will help to make decision for the operators when power system bus is at fault and can identify the type of fault that had occurred. Based on the knowledge or results acquired from data mining, it will help to reduce the human errors in making decision in the

future. Failure of transmission lines will cause financial losses to companies, safety of workers and unreliability from customers. Hence, restoration of transmission lines failure must be done expeditiously. Overhead transmission lines fault can be categorized in two categories, Symmetrical fault and Asymmetrical fault as shown in Figure 1. Three phase fault is a symmetrical fault where all three phases are affected equally it also may called as balanced fault. Asymmetrical unlike Symmetrical fault the faults are not affecting all the three phases equally. There are three types of Asymmetrical faults such as Single Line-to-Ground fault, Line-to-Line fault and Double Line-to-Ground fault.
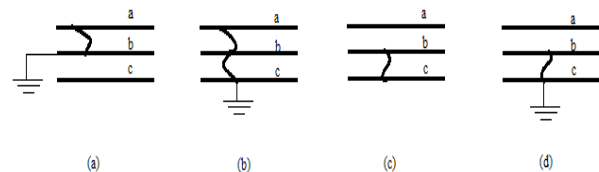

Figure 1: Overhead transmission lines illustration (a) Single line-to-ground fault (b) 3 phase fault (c) line-to-line fault (d) double line-to-ground fault

There are several numbers of methods used for transmission lines fault identification. Early models used impedance-based which utilize pure fault and pre-fault data to detect transmission lines fault location [1]. Travelling waves which is used for high frequency electromagnetic waves [2]. Travelling waves model have a lot of noises which will affect the fault detection and Discrete Wavelet Transform (DWT) is used to eliminate the distorted signals. [3] include actual data acquisition features into DWT by using Global Positioning System(GPS). However, DWT models unable to detect line-to-line fault and single line-to-ground fault. The mother wavelet in DWT must be chosen properly. Non-conventional models applied to power system fault detection such as Artificial Intelligent techniques. [4] had proposed Artificial Neural Network (ANN) on power system transmission lines fault and classified the fault type. [5] proposed to detect single line-to-ground fault using ANN. [6] used Fuzzy Logic model to detect single line-to-ground fault and double line-to-ground fault. It further improved by [7] to include detection of line-to-line fault by using Fuzzy-Neuro approach. Increase in transmission lines fault types identification had proposed by [8] based on Fuzzy logic model. However, it is time

consuming to detect transmission lines fault due to large network of power system.

A better solutions or enhancement of non-conventional models is KDD techniques. Many researchers and experts are interested in applying KDD models such as data mining techniques into field of power systems. [9] had proposed IEEE 9 bus transmission lines fault detection using Cluster Analysis based on Data Mining techniques. However, it was only tested for single line-to-ground fault and unable to detect various types of transmission lines faults. The proposed model use LOF which is one of the density-based methods to determine symmetrical and asymmetrical transmission lines fault for 9 bus and 30 bus power system. This paper is organized as follows, Section 2 briefly discussed on LOF as density-based model, Section 3 will discussed on the proposed model using LOF techniques to identify the bus at fault and the types of fault that occur, Section 4 presents the results of simulations. Finally, Section 5 will be the conclusion remarks and future works. Section 6 is the acknowledgement for this research projects.

2. LOCAL OUTLIER FACTOR BASED OUTLIER ANALYSIS MODEL

An Outlier Analysis consists of multiple of methods. Density based method is one of the method in Outlier Analysis. This method use complex mechanisms to model the outlier value of the data points. LOF is one of the Density-based methods in Outlier Analysis. LOF algorithm is able to measure the local density of the particular point and the local densities of the neighbor of that point. Thus, the degree of outlier of a data point is normally measure with the ratio of density of the point to the averaged densities of nearest neighbors of the point. This method has a powerful modeling capability, but it required a complicated computation. The density based method is first proposed by [10]. [10] given an example which shows that density-based method is having more advantage than distance-based methods to detect outlier. The sample data set is shown in Figure 2. The sample data set consists of an outlier point p, two clusters with different density D1 and D2. The distance based method fail to detect point p due to the density at cluster D1 and density of point p neighbor are almost similar. Whereas, density based method will measure the relative densities of each points and it neighbor density which make it able to detect point p as outlier point.
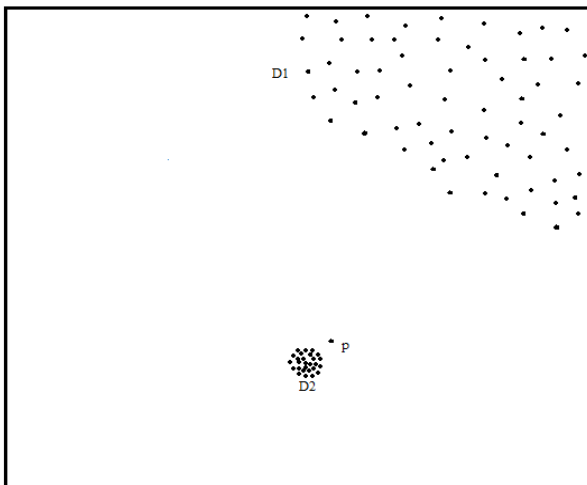

Figure 2: Illustration of Sample Data Set

This density-based formulation quantifies the outlying degree of points using LOF. Given parameter K, LOF of a point p is defined as:

$$LOF_K(p) = \frac{\sum_{o \in K(p)} \frac{lrd_{K(o)}}{lrd_{K(p)}}}{|N_K(p)|} \qquad (1)$$

Where $|N_K(p)|$ is the points falling into the K-neighborhood of p and $lrd_{K(p)}$ denotes the local reachability density of point p that is defined as the inverse of the average reachability distance based on the K nearest neighbors of p. The local reachability density can be express as:

$$lrd_K(p) = 1 / \frac{\sum_{o \in K(p)} reach_{dist_K}(p,o)}{|N_K(p)|} \qquad (2)$$

where the reachability distance of point p can be represent as:

$$reach_{dist_K}(p,o) = \max(K_{distance(o)}, dist(p,o)) \qquad (3)$$

LOF of a data point reflects the density difference between the density of the data point and those densities of the data point neighbor. The neighborhood is defined by the distance to the Kth nearest neighbor. The local outlier factor is a mean value of the ratio of the density distribution estimate in the neighborhood of the object analyzed to the distribution densities of its neighbors [10]. This mean that the lesser the density of point p and the higher the densities of neighbor of data point p, the higher the value of LOF(p), which indicates that p has a higher degree of being an outlier.

2.1 Application of Local Outlier Factor on 3 Phase Transmission Lines Fault Identification

Overhead transmission lines fault becoming harder to detect as the power system network became distributed. Due to the increased in data or information collected to the database many of the conventional models discussed in Section 1 unable to perform very well due to the complexity of today power system networks and many noise data available. The proposed model of fault identification flow chart was shown in the Figure 3. The proposed model utilized the three phase bus voltages (Va, Vb and Vc) at fault from PowerWorld simulation of 9 bus and 30 bus power system. The data collected was transformed into values 0 and about 1 with Cosine Similarity method and subjected to LOF computation for each phase voltages individually with Matlab.If there were non-empty LOF value outputs after evaluation, the raw data undergone data transformation at data transformation stage followed by LOF computation of each phase voltages.

Therefore, there were two processes occurred in the proposed model to obtain both types of symmetrical and asymmetrical transmission lines fault. The first process identify three phase fault, single line-to-ground fault and double line-to-ground fault(which was not simulated in this paper), whereas, the second process determine line-to-line fault and double line-to-ground fault. All the identification of bus at fault and type of fault were compared with the simulated data in order to make sure the result obtain was consistent.

2.2 Fault Simulation of 9 BUS and 30 BUS Power System

The 9 bus and 30 bus symmetrical and asymmetrical faults were simulated usingPowerWorld version 18. The 9 bus power system consists of 9 buses, 3 generators and 9 transmission lines and the 30 bus power system consists of 6 generators, 30 buses and 41 transmission lines. The one line diagram for 9 bus power systems are show in Figure 4.

The symmetrical fault and asymmetrical fault were simulated

at particular bus. Each types of fault were simulated at different buses. Table 1 shows the bus at specific type of fault for both 9 bus power system and 30 bus power system.

2.3 Cosine Similarity and Data Transformation Model

In order to perform LOF computations, the proximity matrix have to be calculated. The method used for calculating the proximity matrix is Cosine Similarity method. This method will represent the data similarities in term of angles. Cosine Similarity model focused on the minimal angle between two vectors. Generally, it can be mathematical represented as:

$$\cos\theta = \frac{a^T b}{\|a\| \, \|b\|} \tag{4}$$

Where a and b are vector

Based on Cauchy-Schwarz inequality [11] stated that for all vectors x and y of an inner product space it is true that

$$\left|a^T b\right| \leq \|a\| \cdot \|b\| \tag{5}$$

$$-1 \leq \cos\theta \leq 1 \tag{6}$$

Therefore, from equation (6) the Cosine Similarity values are within 1 and -1 range where 1 is 0° and -1 is 180°.

Data transformation model only occur in the second step when there are non-deterministic outputs in the first step. The data transformation model equations are as follows:

$$V_{ab} = V_{ai} - V_{bi} \tag{7}$$

$$V_{ac} = V_{ai} - V_{ci} \tag{8}$$

$$V_{bc} = V_{bi} - V_{ci} \tag{9}$$

Where a, b and c denote phase A, phase B and phase C respectively and i denote the bus number. By using the above equations (7), (8) and (9), the LL fault or DLG fault occur in particular bus can be identify.
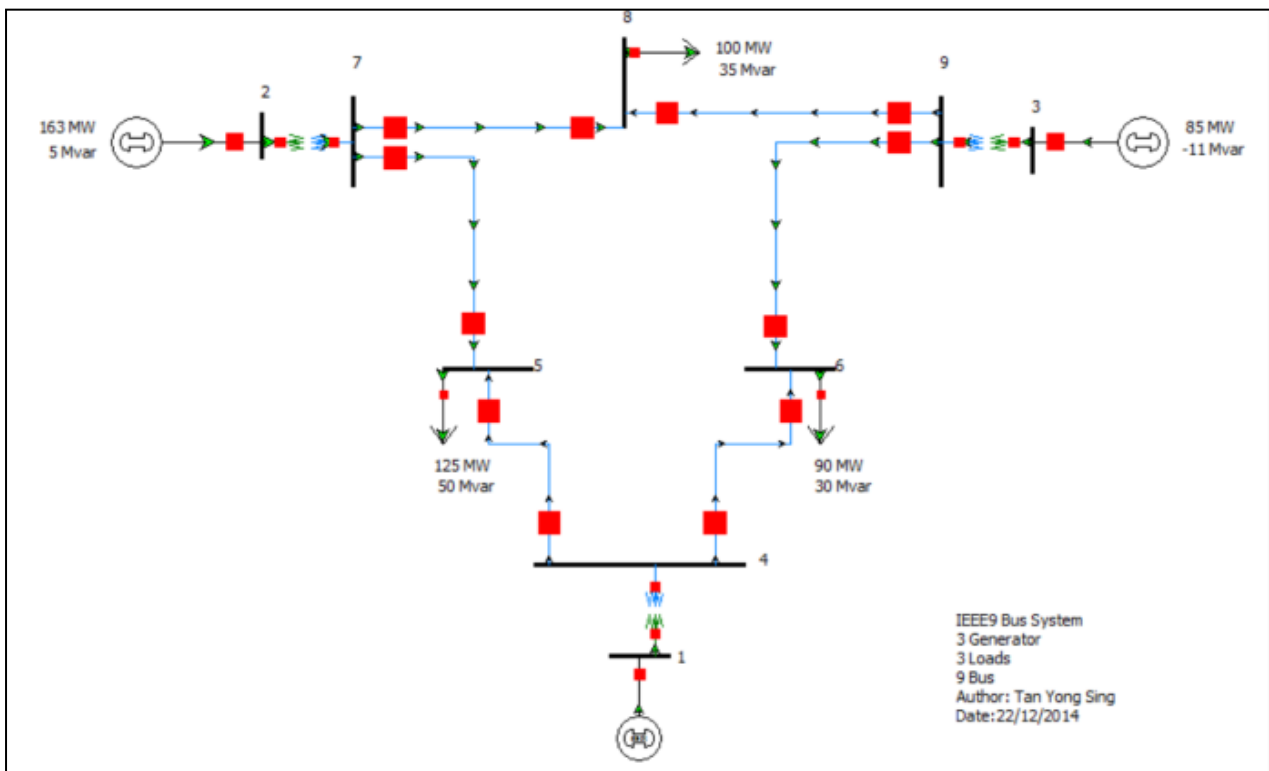

Figure 4: One-line diagram of 9 bus system

Table 1: Simulated faults at particular bus

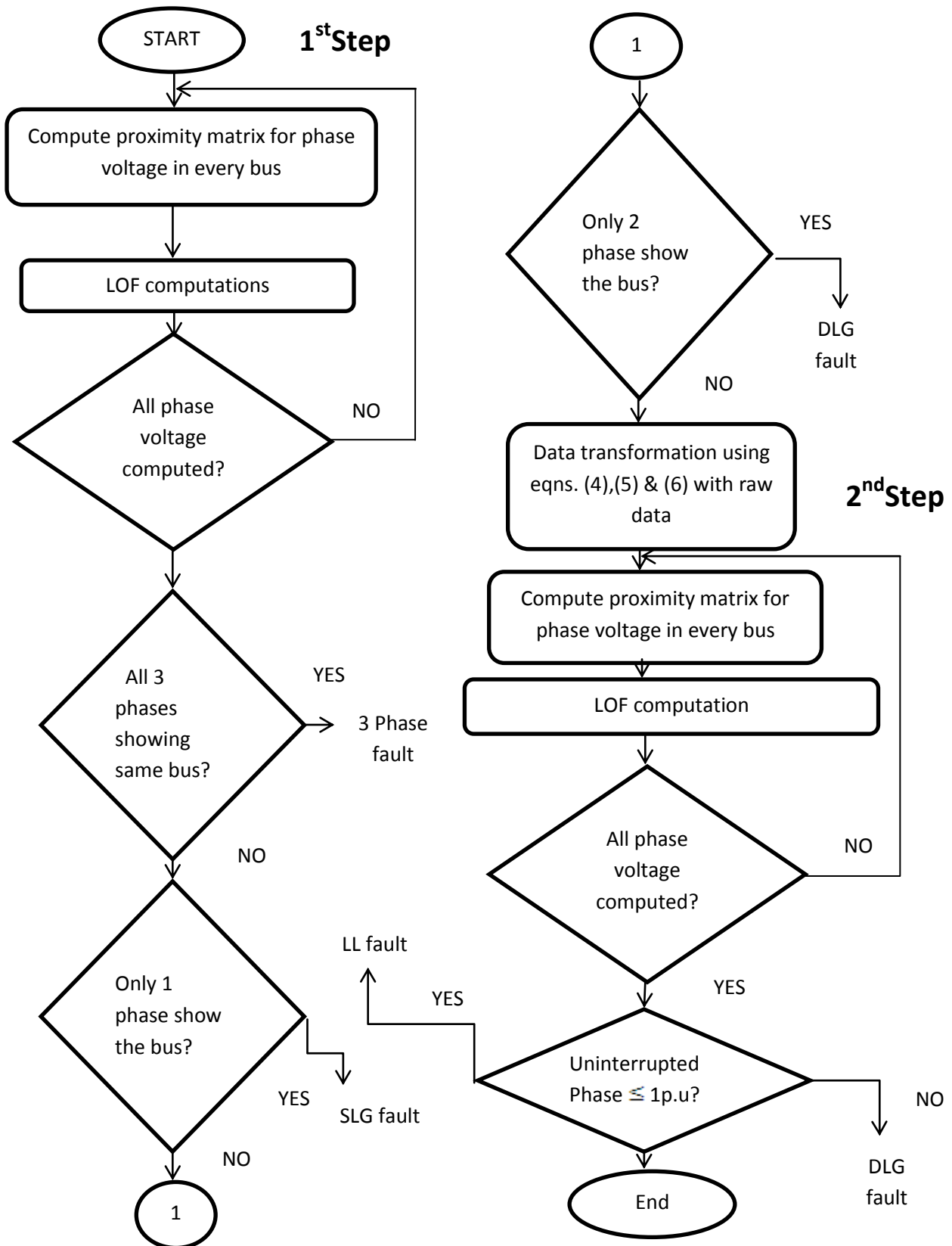| Types of faults	Bus at Fault	
	9 Bus Power System	**30 Bus Power System**
3 Phase	Bus 8	Bus 2
SLG	Bus 5	Bus 7
LL	Bus 7	Bus 11
DLG	Bus 4	Bus 18

Figure 3: Proposed model flow chart

LOF Three Phase Transmission Lines Fault Identification

After the proximity matrices are computed in first step or second step, LOF computation is carried out. The algorithm of LOF computation is shown in Figure 5 as below:
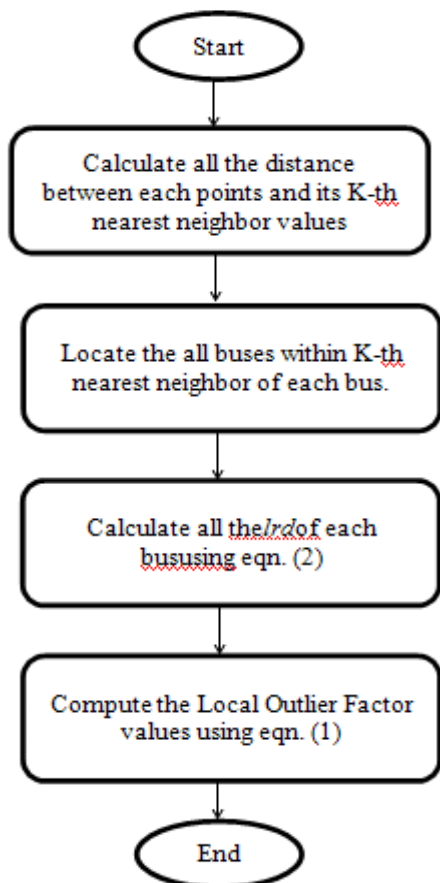
Figure 5: LOF algorithm flow chart

The K-th value used for 9 bus power system and 30 bus power systems are different as shown in Table 2.

Table 2: K-th values assignment for 9 bus and 30 bus power system

Power System	K-th Values
9 Bus	7
30 Bus	20

The K-th value as shown in Table 2 is for single bus fault test case. The K-th values may varies if more than single bus faults at a time.

3. RESULTS

Figure 6 shows the result in the form of LOF vs. Bus scatter plot of each phase except for Figure 6(c) and Figure 6(d). LL fault and DLG fault undergone second step as mention in the Section 3. The missing bubble head which having an empty LOF value in the scatter plot representing the bus is at fault. From the evaluation of the three scatter plot graphs SLG fault, 3 Phase fault, LL fault and DLG fault can be determined. Due to the proximity matrix is computed using Cosine Similarity method in the first and second process, the computed LOF values for unaffected phases in each bus was constant at value 49. If all three scatter plot show a constant values of 49, it is term as non-deterministic result. Thus, the proposed model successfully identifies both specific single bus fault and types of fault which the bus experienced.

The obtained results for 30 bus power system are almost similar to results for 9 bus power system. The Figure 7 shows that the computed LOF value was 400. The empty LOF values in the phases scatter plot represent the particular bus at fault. By comparing the three scatter plot of each case, both symmetrical and asymmetrical faults can be identified. Thus, the proposed model successfully identifies both specific single bus fault and types of fault which the bus experienced for 9 bus and 30 bus power system.
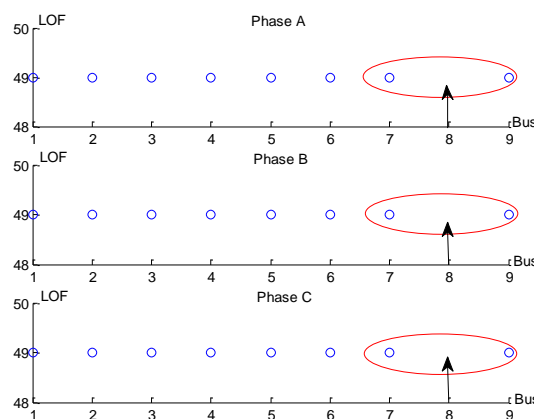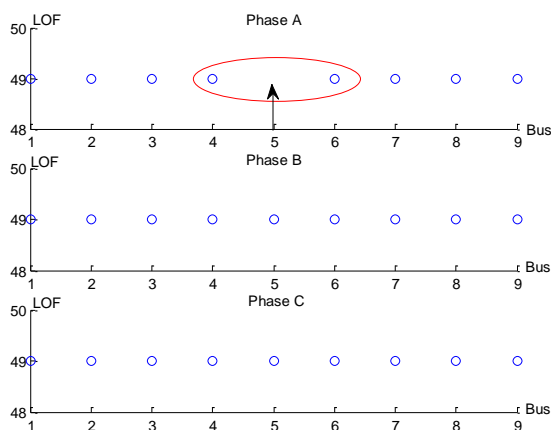
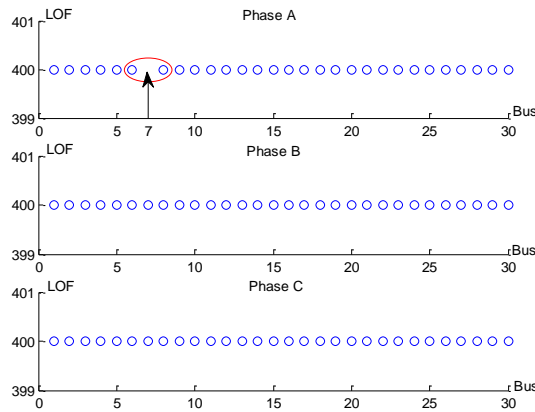

(a) Bus5 SLG Fault (b) Bus 8 3 Phase Fault
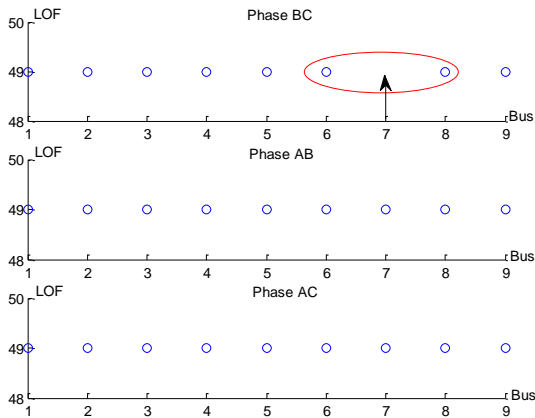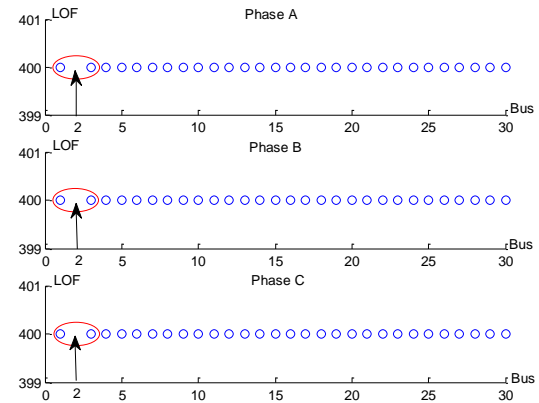
(c)Bus 7 LL fault with Va ≤ 1 at Bus 7
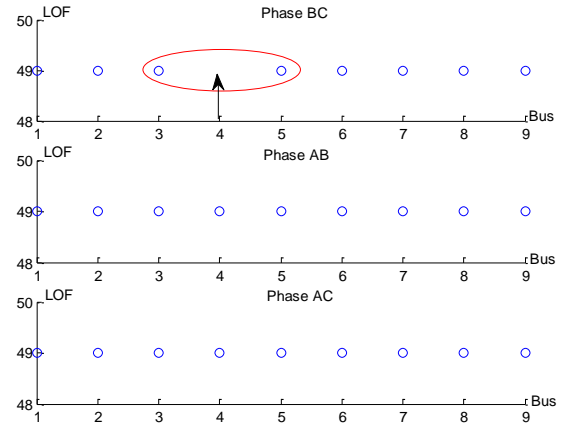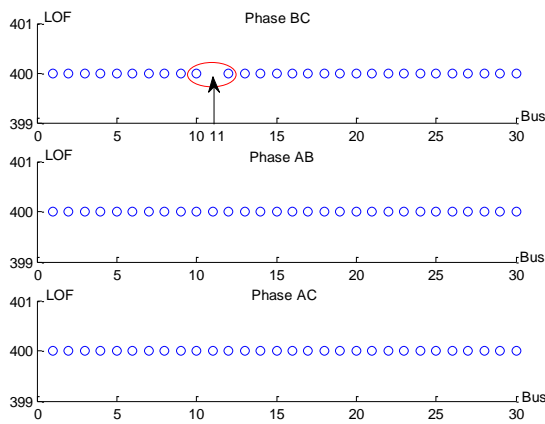
(d)Bus 4 DLG fault with Va> 1 at Bus 4

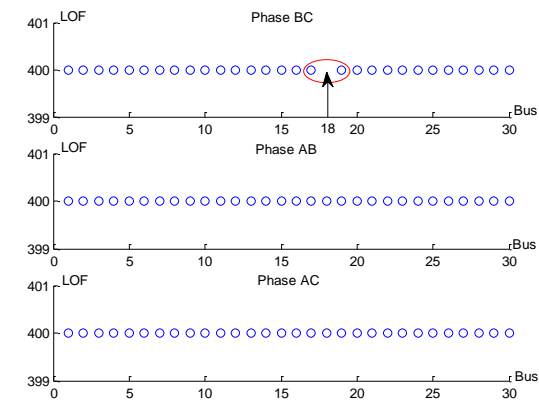**Figure 6: 9 bus power system fault identification scatter plot**

(a) Bus7 SLG Fault

(b) Bus 2 3 phase fault

(c) Bus 11 LL fault with Va ≤ 1 at Bus 11

(d) Bus 18 DLG fault with Va> 1 at Bus 18

Figure 7: 30 bus power system fault identification scatter plot

4. CONCLUSION

A simple and innovative proposed model has successfully detected a single bus fault and identified different types of transmission lines faults for both 9 bus power system and 30 bus power systems. As observed in Table 2, the K-th values increases as the number of bus in the power system increases for single bus fault case. The K-th value selected is more than 50% of the total number of buses in the power system. Future

work may include multiple faults simultaneously and inline fault detection. This shows that there is huge potential by using data mining technique in various power systems applications in planning and operations such as power system protection, voltage stability, contingency analysis, economic load dispatch etc. Therefore, in today world of ICT there are lot of information been presented in order to optimize the efficiency in solving complex problem in power systems through non-conventional methods. Hence it has been proved

that data Mining technique is of the best non-conventional methods that gaining popularity and has best implementation ability in big data power systems applications.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] T.Takagi, Y. Yamakoshi, R. Kondow, M. Yamaura, T. Matsushima, "Development of a New Type Fault Locator Using the One-Terminal Voltage and Current Data." IEEE Power EngRev. PER-2(8), 59-60 (1982).

[2] J.Rohrig, "Location of Faulty Places by Measuring with Cathode Ray Oscilloscope".ElecticitatZeitschrift, 19th Feb, 1931, pp 241-242.

[3] Amir Tabatabaei, Mohammad-Reza Mosavi, AbdolrezaRahmati, "Fault Location Techniques in Power System based on Traveling Wave using Wavelet Analysis and GPS Timing", Electrical Review, ISSN 0033-2097, 2012.

[4] Kulicke B, Dalstein T. "Neural Network Approach To Fault Classification For High Speed Protective Relaying", IEEE Transactions on Power Delivery, vol. 4, 1995, pp. 1002-1009.

[5] Aravinda Surya. V, EbhaKoley, Anamika Yadav and A.S.Thoke, "Artificial Neural Network Based Fault Locator for Single Line to Ground Fault in Double Circuit Transmission Line", DOI: 10.7763/IPEDR, vol. 75, 2014

[6] S.Sangiovanni, A.Ferrero, E. Zapitelli, "A Fuzzy Set Approach To Fault Type Identification In Digital Relaying", IEEE Trans. Power Delivery, vol.10, pp. 169-175, January 1995.

[7] W.W.L. Keerthipala and H.Wang, "Fuzzy Neuro Approach To Fault Classification For Transmission Line Protection", IEEE Trans. Power Delivery, vol.13, pp.1093-1104, October 1998.

[8] J.Vital Reddy and B.Das, "Fuzzy-Logic Based Fault Classification Scheme For Digital Distance Protection", IEEE Trans. Power Delivery, vol.20, pp. 609-616, April 2005.

[9] Yagang Zhang, Jing Ma, Jinfang Zhang, Zenping Wang, "Application of Data Mining Theory in Electrical Engineering", SciRes,DOI:10.4236/eng.2009.13025.

[10] M. Breuning, H-P.Kriegel, R. Ng, and J. Sander. "LOF: Identifying Density-Based Local Outliers". In Proc. of 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD'00), Dallas, Texas, pp 93-104, 2000.

[11] J.M. Aldaz, S.Barza, M.Fuji and M.S. Moslehian. "Advances in Operator Cauchy-Schwarz Inequalities and Their Reverse, Ann. Func.Anal. 6 (2015), no.3, p275-295.