

The use if machine learning techniques to classify power transmission line fault types and location

K. Chandrakala

Department of Computer Science, Sri Venkateswara University, Tirupati

Abstract—Transmission lines are very important component of the electric power system. Therefore, it is necessary to predict and detect transmission line's fault types and locations to enhance the power system protection scheme and increase its reliability. This paper investigates the use of four powerful machine learning classifiers to detect and predict fault types and locations over a 750KV, 600km long power transmission line. Bagging, Boosting, radial basis functions and naïve Bayesian classifiers were utilized for locating and detecting faults in a power transmission line. Findings exhibits that using machine learning technique could be feasible for such task and may represent a great opportunity to increase the power system protection and efficiency.

I. INTRODUCTION

In recent days, Trustworthy and efficient algorithms and techniques that can provide a correct and accurate analysis of faults on overhead power transmission lines have become necessary in order to implement a modern transmission line protection system. Transmission line protection systems usually operate by identifying the fault and separate only the faulty zone. There are four primary fault types: 1- 3-Phase (Triple line with or without ground) 2- Phase to phase 3- Phase to Ground 4- Double Phase to Ground Three-phase and three-phase to ground faults are similar in terms of electrical quantities (current and short-circuit voltage), which explains the fact that some researchers limit themselves to enumerating 10 types of electrical faults instead of 11 faults. However, in practice, it is important to differentiate them because the short-circuit current is different in both cases and will thus cause different damage to some extent. Restoration or putting into service of an electrical power transmission line after a permanent fault occurrence can only be carried out after maintenance of the electrical power zone which is at fault. The search for faults can be difficult especially over a long transmission line. Thus, it is important to detect and to locate the fault or to evaluate it within minimum error and time. Also quick fault detection can help protect equipment by allowing the disconnection of the power transmission line before any damage occurs. Hence increasing cost savings and power transmission system efficiency and reliability. Due to its simplicity and effectiveness, recent research began to introduce machine learning techniques to fault type and location detection and prediction. Techniques such as multilayer perceptron, random forests, support vector machines were used to locate and predict faults on transmission lines. However, none have introduced the powerful machine learning ensemble methods of Bagging, and Boosting, nor the state of the art the naïve Bayesian classifier and the radial basis function classifier. In addition, a few research was conducted on a fairly long power transmission lines (length > 240km), however none experimented on a length reached 600km. In recent years, machine learning algorithms (ML) has been successfully applied in many engineering fields which include computer systems, vision, finance, hospital and medicine, transportation, telecommunications, heuristic classification, aviation, gaming, data mining, speech recognition, and heavy industry. The use of such powerful machine learning techniques in fault prediction could result on enhancing the protection procedures for the power transmission system. In addition it will reduce the time needed to clear the faults, especially for a long transmission line, hence increasing the overall power system reliability and efficiency. There are two main contributions for this paper. The first one is to apply four powerful and popular machine learning techniques of Bagging, Boosting, naïve Bayesian classifier and radial basis function to classify and detect 11 different fault types and locations. The second one is to experiment on a high voltage (750 kV) and a long power transmission line (600 km) with 11 different fault types.

1.1 Artificial Intelligence:

Artificial intelligence (AI) is the ability of a computer program or a machine to think and learn. It is also a field of study which tries to make computers "smart". As machines become increasingly capable, mental facilities once thought to require intelligence are removed from the definition. AI is an area of computer sciences that emphasizes the creation of intelligent machines that work and reacts like humans. Some of the activities computers with artificial intelligence are designed for include: Face recognition, Learning, Planning, Decision making etc., Artificial intelligence is the use of computer science programming

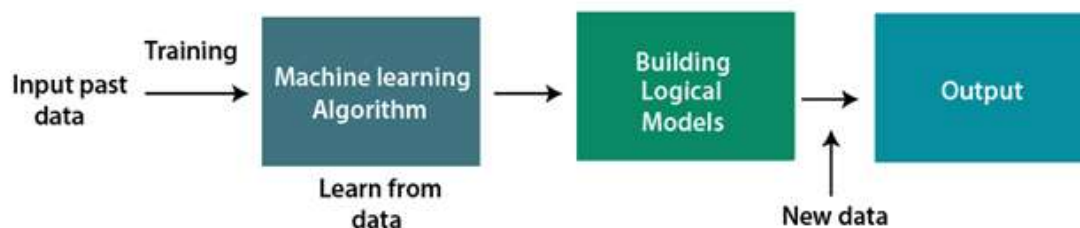
to imitate human thought and action by analysing data and surroundings, solving or anticipating problems and learning or self-teaching to adapt to a variety of tasks.

1.2 Machine Learning

Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for building mathematical models and making predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, Facebook auto-tagging, recommender system, and many more. Machine Learning is said as a subset of artificial intelligence that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own. The term machine learning was first introduced by Arthur Samuel in 1959. We can define it in a summarized way as: "Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed".

A Machine Learning system learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it. The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately.

Suppose we have a complex problem, where we need to perform some predictions, so instead of writing a code for it, we just need to feed the data to generic algorithms, and with the help of these algorithms, machine builds the logic as per the data and predict the output. Machine learning has changed our way of thinking about the problem. The below block diagram explains the working of Machine Learning algorithm:



1.2.1 Features of Machine Learning:

- Machine learning uses data to detect various patterns in a given dataset.
- It can learn from past data and improve automatically.
- It is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with the huge amount of the data.

1.2.2 Classification of Machine Learning

At a broad level, machine learning can be classified into three types:

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Supervised Learning

Supervised learning is a type of machine learning method in which we provide sample labeled data to the machine learning system in order to train it, and on that basis, it predicts the output.

The system creates a model using labeled data to understand the datasets and learn about each data, once the training and processing are done then we test the model by providing a sample data to check whether it is predicting the exact output or not.

The goal of supervised learning is to map input data with the output data. The supervised learning is based on supervision, and it is the same as when a student learns things in the supervision of the teacher. The example of supervised learning is **spam filtering**.

Supervised learning can be grouped further in two categories of algorithms:

- **Classification**
- **Regression**

Unsupervised Learning

Unsupervised learning is a learning method in which a machine learns without any supervision. The training is provided to the machine with the set of data that has not been labeled, classified, or categorized, and the algorithm needs to act on that data without any supervision. The goal of unsupervised learning is to restructure the input data into new features or a group of objects with similar patterns.

In unsupervised learning, we don't have a predetermined result. The machine tries to find useful insights from the huge amount of data.

It can be further classified into two categories of algorithms:

- **Clustering**
- **Association**

1.3 Naïve Bayes:

Naïve Bayes' Classifier (NBC) The naïve Bayes' classifier is considered an easy method, with simple approach, to learn and classify probabilistic knowledge. It is founded by implementing Bayes' theorem with solid (naïve) independence assumptions, particularly, independent feature model. It is easy to construct a Naive Bayes model. In addition, it is used efficiently for very large data sets. Along with simplicity, Naive Bayes is known to over-perform several powerful classification methods. There are several cases where the naïve Bayesian classifier does not perform well for certain datasets. This is due to the fact that attributes are dealt with as if they are entirely independent; the addition of redundant ones slants the learning process.

1.4 Bagging Classifier

Bagging Classifier The use of Bagging or Bootstrap Aggregation classifier was offered by Breiman in 1996 [15]. Bagging technique is widely used to enhance the classification accuracy outcomes by merging the findings of the trained classifiers using randomly generated training sets. Bagging is considered an ensemble or a multiclassifier algorithm technique which superior to a single classifier. Bootstrap creates samples of the training set by using sampling with replacement. Bagging technique is known for its outstanding performance as proven in many of the recent research. Bagging technique aims to enhance stability and accuracy of machine learning algorithms by reducing variance and avoiding over-fitting. As Bagging technique resamples the training set with replacement, certain instance will be represented couple of times while other instances will not be represented at all.

1.5 Boosting Classifier

Boosting classifier Boosting classifier was presented by Schapire and Freund. Boosting main approach is to introduce a series of classifiers within the data set. The hypothesis behind Boosting technique is simply referred to the process of converting a weak learner into a strong one. The training set used for each classifier within the generated series is selected depending on the performance of the previous generated classifier(s) in the series. Most Boosting algorithms contain iteratively learning weak classifiers with respect to a distribution and adding them to the last strong classifier. Usually, in Boosting technique, the cases which were incorrectly classified by the prior classifiers in the series are selected than cases that were accurately classified. As a result, Boosting tries to build new enhanced classifiers that can predict cases in high prediction accuracy.

II. LITERATURE REVIEW

[1] **TITLE:** An Ensemble Machine Learning Based Fault Classification Method for Faults during Power Swing.

AUTHOR NAME: Dinesh Patil

DESCRIPTION: In this paper, a novel method for phase selection of faults during power swing is presented. The proposed method operates with current signals from one end of the protected transmission line. The most popular and powerful ensemble machine learning technique Random Forest is used for fault classification. The magnitude of the three-phase incremental quantities of currents is used as input to random forest for classifying the fault. The incremental quantity of the current signal is computed using a moving average filter. The proposed method is validated with 9216 cases covering varying swing frequency, inception angle, fault type, fault resistance, source to line impedance ratio and fault distance. Achieved fault classification within 10msec with an accuracy of 99.8%.

[2] TITLE: Extreme learning machine-based fault classification in a series compensated transmission line.

AUTHOR NAME: Papia Ray

DESCRIPTION: This project presents an accurate hybrid method for fault classification in a series compensated transmission line. The proposed scheme uses discrete wavelet transform in combination with extreme learning machine for fault classification. Instantaneous value of current signal is measured from the relaying end of the transmission line for one cycle duration from the inception of fault. Discrete wavelet transform is used to decompose the signal and extract certain features from it. The feature set is then normalized and best features are selected from the total feature set by forward feature selection method. Selected features are then fed as an input to the extreme learning machine for fault classification. To evaluate the feasibility of the proposed technique, it is tested on a 400 kV, 300 km series compensated transmission line for all the ten type of fault using MATLAB/ SIMULINK. A wide range of simulation condition is taken to generate the train and test pattern. Simulation result indicates that the proposed approach is robust, fast in learning and classifies the fault very accurately.

[3] TITLE: <https://ieeexplore.ieee.org/abstract/document/1017309/> Transmission Line Fault Classification and Location Using Multi-Layer Perceptron Artificial Neural Network.

AUTHOR NAME: A. K. Onaolapo

DESCRIPTION: Most transmission lines are overhead, spanning possibly thousands of kilometers and are exposed to different climatic conditions. and therefore prone to faults. This paper develops accurate, fast and reliable algorithms that can classify and locate faults on transmission lines. These reduce outage time. Multi-Layer Perceptron Artificial Neural Networks (MLP-ANN) are used to implement the algorithms because of their ability to learn during a wide range of conditions. The ANN inputs are the voltage and current signals. These were measured at one terminal of the line. The test system is the 29-bus Great Britain transmission network. The algorithm sensitivity is investigated with varying fault impedance and inception angle, and different locations on the transmission line.

[4] TITLE: Fault Classification and Detection in Transmission Lines Using ANN

AUTHOR NAME: ShreyaUpadhyay

DESCRIPTION: The ANN is a potent tool for the classification and detection of the faults on the transmission line due to pattern recognition ability. This paper centers on progressing an ANN to classify and detect a fault on power transmission line. The study employs feed-forward ANN with back-propagation algorithm in evolving the fault detector & classifier. Simulation and modelling of transmission lines was done in MATLAB using sim power systems Toolbox. The instant currents and voltage numeric are taken after training the fault classifier & detector. Simulation results have been delivered to determine the effectiveness of the technologically advanced intellectual detector & classifier for fault. The performance of the classifier & detector is assessed by the Mean Square Error (MSE) and confusion matrix. The detector attained a tolerable MSE of $8.5571e-7$ and accuracy of 100%. The classifier attained a tolerable MSE of 0.63035 and accuracy of 88%, presenting that the performance of the technologically advanced intellectual classifier detector is acceptable.

[5] TITLE: Fault Classification and Location Identification on Electrical Transmission Network Based on Machine Learning Methods.

AUTHOR NAME: Vidya Venkatesh.

DESCRIPTION: Power transmission network is the most important link in the country's energy system as they carry large amounts of power at high voltages from generators to substations. Modern power system is a complex network and requires high-speed, precise, and reliable protective system. Faults in power system are unavoidable and overhead transmission line faults are generally higher compare to other major components. They not only affect the reliability of the system but also cause

widespread impact on the end users. Additionally, the complexity of protecting transmission line configurations increases with as the configurations get more complex. Therefore, prediction of faults (type and location) with high accuracy increases the operational stability and reliability of the power system and helps to avoid huge power failure. Furthermore, proper operation of the protective relays requires the correct determination of the fault type as quickly as possible (e.g., reclosing relays). With advent of smart grid, digital technology is implemented allowing deployment of sensors along the transmission lines which can collect live fault data as they contain useful information which can be used for analyzing disturbances that occur in transmission lines. In this thesis, application of machine learning algorithms for fault classification and location identification on the transmission line has been explored. They have ability to “learn” from the data without explicitly programmed and can independently adapt when exposed to new data. The work presented makes following contributions:

- 1) Two different architectures are proposed which adapts to any N-terminal in the transmission line.
- 2) The models proposed do not require large dataset or high sampling frequency. Additionally, they can be trained quickly and generalize well to the problem.
- 3) The first architecture is based off decision trees for its simplicity, easy visualization which have not been used earlier. Fault location method uses traveling wave-based approach for location of faults. The method is tested with performance better than expected accuracy and fault location error is less than $\pm 1\%$.
- 4) The second architecture uses single support vector machine to classify ten types of shunt faults and Regression model for fault location which eliminates manual work. The architecture was tested on real data and has proven to be better than first architecture. The regression model has fault location error less than $\pm 1\%$ for both three and two terminals.
- 5) Both the architectures are tested on real fault data which gives substantial evidence of its application.

III. PROBLEM STATEMENT

- In recent years, machine learning algorithms (ML) has been successfully applied in many engineering fields which include computer systems, vision, finance, hospital and medicine, transportation, telecommunications, heuristic classification, aviation, gaming, data mining, speech recognition, and heavy industry.
- The use of such powerful machine learning techniques in fault prediction could result on enhancing the protection procedures for the power transmission system.
- In addition it will reduce the time needed to clear the faults, especially for a long transmission line, hence increasing the overall power system reliability and efficiency.

Disadvantage

- It will take too much resource for the processing.
- Take too much processing time.
- Accuracy is less for the prediction in testing audio.

IV. DEVELOPMENT PROCESS

4.1 Requirement Analysis and Specifications

The requirement engineering process consists of feasibility study, requirements elicitation and analysis, requirements specification, requirements validation and requirements management. Requirements elicitation and analysis is an iterative process that can be represented as a spiral of activities, namely requirements discovery, requirements classification and organisation, requirement negotiation and requirements documentation.

4.2 Input Requirement and Output Requirements

Input Design

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?

- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

Objectives

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus, the objective of input design is to create an input layout that is easy to follow

Output Design

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

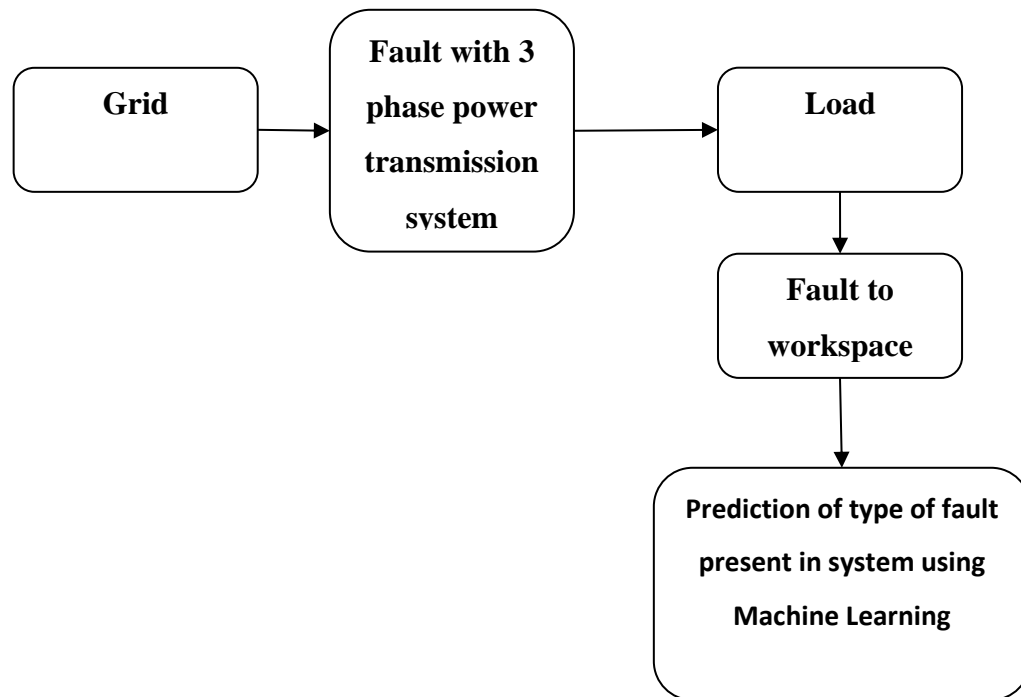
- ❖ Convey information about past activities, current status or projections of the
- ❖ Future.
- ❖ Signal important events, opportunities, problems, or warnings.
- ❖ Trigger an action.
- ❖ Confirm an action.

V. PROPOSED METHOD

- There are two main contributions for this paper. The first one is to apply four powerful and popular machine learning techniques of Bagging, Boosting, naïve Bayesian classifier and radial basis function to classify and detect 11 different fault types and locations.
- The second one is to experiment on a high voltage (750 kV) and a long power transmission line (600 km) with 11 different fault types.

Advantages

- An improved version of Machine learning parameters was proposed in this proposed method to improve the accuracy of classification.
- Based on the experimental results, our method was superior to the techniques described in the conventional methods.



MODULES USED

- 1. Datasets Collection.
- 2. Pre-processing.
- 3. Model Implementation.
- 4. Classification.

1. Datasets Collection:

A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question.

A Collection of data is called datasets. Here, the datasets collection is audio file.

2. Preprocessing:

Pre-processing routines prepare the data for analysis. Before we start the actual processing, the data has to be pre-processed to remove the detector effects. Preprocessing is the most important aspect of data processing. Hence, data filtering, data ordering, data editing and noise modeling play an important role in any data preprocessing.

A Pre-Processing is one of the techniques is used to reduce the noise in the file format of the system. Smoothing and de-trending are processes for removing noise and linear trends from data, while scaling changes the bounds of the data. Grouping and binning methods are techniques that identify relationships among the data variables.

3. Model Implementation:

Machine learning

1. Naïve Bayes' Classifier (NBC) The naïve Bayes' classifier is considered an easy method, with simple approach, to learn and classify probabilistic knowledge. It is founded by implementing Bayes' theorem with solid (naïve)

independence assumptions, particularly, independent feature model. It is easy to construct a Naive Bayes model. In addition, it is used efficiently for very large data sets. Along with simplicity, Naive Bayes is known to over-perform several powerful classification methods. There are several cases where the naïve Bayesian classifier does not perform well for certain datasets. This is due to the fact that attributes are dealt with as if they are entirely independent; the addition of redundant ones slants the learning process.

2. **Bagging Classifier** The use of Bagging or Bootstrap Aggregation classifier was offered by Breiman in 1996 [15]. Bagging technique is widely used to enhance the classification accuracy outcomes by merging the findings of the trained classifiers using randomly generated training sets. Bagging is considered an ensemble or a multiclassifier algorithm technique which superior to a single classifier. Bootstrap creates samples of the training set by using sampling with replacement. Bagging technique is known for its outstanding performance as proven in many of the recent research. Bagging technique aims to enhance stability and accuracy of machine learning algorithms by reducing variance and avoiding over-fitting. As Bagging technique resamples the training set with replacement, certain instance will be represented couple of times while other instances will not be represented at all.
3. **Boosting classifier** Boosting classifier was presented by Schapire and Freund. Boosting main approach is to introduce a series of classifiers within the data set. The hypothesis behind Boosting technique is simply referred to the process of converting a weak learner into a strong one. The training set used for each classifier within the generated series is selected depending on the performance of the previous generated classifier(s) in the series. Most Boosting algorithms contain iteratively learning weak classifiers with respect to a distribution and adding them to the last strong classifier. Usually, in Boosting technique, the cases which were incorrectly classified by the prior classifiers in the series are selected than cases that were accurately classified. As a result, Boosting tries to build new enhanced classifiers that can predict cases in high prediction accuracy.

4. Classification:

Classification neural networks used for feature categorization are very similar to fault-diagnosis networks, except that they only allow one output response for any input pattern, instead of allowing multiple faults to occur for a given set of operating conditions. Classification is a process related to categorization, the process in which ideas and objects are recognized, differentiated and understood the system. Classification is a term used both about the process to classify the condition of the fault wire. A classification process can be obtained to training the network of the system. Here, classification techniques can be obtained by using Machine Learning model to classify the result of the model. This project finally predicts the type of fault present wiring system.

VI. FUTURE ENHANCEMENT

In our future work, we can implement the concept of colon cancer detection disease using image processing in deep learning techniques. In this project, we are processing in software. In future, we will implement in hardware side.

VII. CONCLUSION

In this project the use of four powerful and popular machine learning techniques (Bagging, Boosting, RBF, and NBC) for identifying and locating faults over a 600 Km power transmission line using was introduced and investigated. 11 faults were recognized to be detected, predicted and located. The experiment results suggest that the RBF, the Bagging and the NBC techniques can be used for fault type prediction as they achieved high prediction accuracy, while Boosting did not achieve enough accuracy to be considered for such application. For fault location prediction experiment, the attained prediction results were not as accurate as in the case of fault type prediction, however, the RBF, the NBC and the Bagging techniques achieved the highest prediction accuracy. From the experimental results it can be concluded that machine learning techniques could be used for identifying the transmission lines faults. However results for location prediction accuracy may need to be improved in order to locate the fault precisely. This may be achieved by introducing other machine learning techniques for such application or increase the data size.

REFERENCES

- [1] Z. M. Radojević, C. H. Kim, M. Popov, G. Preston, and V. Terzija, "New Approach for Fault Location on Transmission Lines Not Requiring Line Parameters", International Conference on Power Systems Transients IPST, Kyoto Japan, 2009.

- [2] J. Izykowski, "Power System faults", Wrocław University of Technology, ISBN 978-83-62098-80-4, Published by PRINTPAP Łódź, 2011.
- [3] M. Mirzaei, M.Z. Ab Kadir, E. Moazami, H. Hizam, "Review of Fault Location Methods for Distribution Power System", Australian Journal of Basic and Applied Sciences , 3(3): 2670-2676, ISSN 1991-8178, 2009.
- [4] M. Farshad, J. Sadeh, "Fault Locating in High Voltage Transmission Lines Based on Harmonic Components of One-End Voltage Using Random Forests", Iranian Journal of Electrical & Electronic Engineering, Vol. 9, No. 3, Sep. 2013.
- [5] R. Malhotra, and A. Jain, "Fault Prediction Using Statistical and Machine Learning Methods for Improving Software Quality", Journal of Information Processing Systems, Vol.8, No.2, pp. 241-262, June 2012.
- [6] A. N Hasan, and B. Twala, "Improving Single Classifiers Prediction Accuracy for Underground Water Pump Station in a Gold Mine Using Ensemble Techniques", IEEE EUROCON, Salamanca, Spain, 2015.
- [7] M. Tayeb, "Faults Detection in Power Systems Using Artificial Neural Network", American journal of Engineering Research AJER, vol.02(06): pp.69-75, 2013.
- [8] I Mohamed, M. Hassan, and MA Moustafa, "Distance Relay Protection for Short and Long Transmission line", Proceedings of International Conference on Modelling, Identification & Control (ICMIC), pp.204-211, 2013.
- [9] T. G. Tan, P. Anthony, J. Teo, and J. H. Ong, "Neural network ensembles for video game ai using evolutionary multiobjective optimization," in Hybrid Intelligent Systems (HIS), 2011 11th International Conference on. IEEE, pp. 605-610, 2011.
- [10] D. Asfani, I. M. Negara, and M. Rajavalens, "Sensitive Detection for High Impedance Fault on Transmission Line using Wavelet and Naive Bayes", JAVA Journal of Electrical and Electronics Engineering, Vol. 12, No.2, pp 29-33, October 2014.