CABLE FAULT DETECTION FOR DSL COMMUNICATION USING MACHINE LEARNING

NURUL BASHIRAH BINTI GHAZALI

A thesis submitted in
fulfillment of the requirement for the award of the
Degree of Master of Electrical Engineering

Faculty of Electrical and Electronic Engineering
Universiti Tun Hussein Onn Malaysia

NOVEMBER 2022

To,

Izzat Zayany Ramli

Akil Aidan Izzat Zayany

Ghazali Mat Isa

Noorul Huda Adnan

Che Salina Md Yusof

and the rest of the family.

This is for all of you. Thank you and I love you.

May all the support you have given me be the reason Allah grants Jannah for all of you.

ACKNOWLEDGEMENT

Foremost, I would like to acknowledge the Almighty God for His benevolence and for granting me wisdom and perseverance not only during the time of research and writing of this thesis, but indeed, throughout my life.

From the bottom of my heart, I would like to thank my parents, husband and family members who have been supporting me from the beginning of my study's journey. They have always been my inspiration and they are what keep me going.

I express, with heartfelt appreciation, my gratitude to my supervisor, Assoc. Prof. Dr. Fauziahanim Che Seman, and my co-supervisors, Assoc. Prof. Dr. Khairun Nidzam Ramli, Ts. Dr. Khalid Isa, Dr. Zuhairiah Zainal Abidin and Dr. Saizalmurisidi Md. Mustam, for their sincere and invaluable intellectual guidance extended to me throughout the years of my postgraduate studies.

Finally, special thanks to my beloved friends for always helping and encouraging me never to stop achieving my goals in life.

ABSTRACT

Very-High-Bit-Rate Digital Subscriber Line 2 (VDSL2) is a technology that is widely deployed all over the world, but its performance may be limited due to its nature, thus causing it to be more susceptible to signal degradation and faulty lines. Some of the VDSL2 line conditions are ideal, bridge tap, partial-short, short-wired, partial-open and open-wired faults. However, VDSL2 is still one of the technologies that users in sub-urban areas depend on to access the internet service. This thesis aimed to demonstrate the classification of VDSL2 line impairments and the regression of VDSL2 line-impairment locations using a machine learning algorithm. VDSL2 line-impairment classification and localisation detection may increase the technology's overall performance. The proposed system was developed using Python programming language following the stages for classification and regression using the Random Forest machine learning algorithm. The proposed algorithm was selected based on comparisons of several algorithms' performances in the WEKA software. The first stage involved data acquisition from the Multi-Service Access Node (MSAN) in the VDSL2 laboratory setup. The laboratory was set up to imitate the actual VDSL2 network. Due to certain VDSL2 line impairments that could cause the unavailability of some attributes, the acquired data needed to undergo the pre-processing stage. The training of the pre-processed data was evaluated using the percentage-split method. The classification of VDSL2 line impairments was assessed based on the percentage of accuracy and correctly classified instances. Using the proposed method, the developed system for the classification and localisation of VDSL2 line impairments was able to achieve 97.73% accuracy and a correlation coefficient of 0.9933 for the VDSL2 classification and localisation models, respectively, after using same dataset as that for training. Validation using the latest daily VDSL2 dataset was able to achieve 100% accuracy and a correlation coefficient of 0.36.

ABSTRAK

Talian Pelanggan Digital Kadar Bit Sangat Tinggi 2 (VDSL2) ialah teknologi yang digunakan secara meluas di seluruh dunia, tetapi prestasinya mungkin terhad disebabkan sifatnya dan lebih mudah terdedah kepada kemerosotan isyarat dan talian rosak. Talian VDSL2 boleh berlaku dalam beberapa keadaan, lebihan kabel yang tidak ditamatkan, litar pintas separa, litar pintas, litar terbuka separa dan litar terbuka. Namun begitu, VDSL2 masih menjadi salah satu teknologi yang penting kepada pengguna di kawasan pinggir bandar untuk menerima perkhidmatan internet. Tesis ini bertujuan untuk menunjukkan pengkelasan jenis kerosakan talian VDSL2 dan regresi lokasi kerosakan talian VDSL2 menggunakan algoritma pembelajaran mesin. Pengkelasan jenis kerosakan talian VDSL2 dan pengesanan kerosakan talian boleh meningkatkan prestasi keseluruhan teknologi. Sistem yang dicadangkan dibangunkan menggunakan bahasa pengaturcaraan Python mengikut peringkat pengkelasan dan regresi menggunakan algoritma pembelajaran mesin Random Forest. Algoritma yang dicadangkan dipilih berdasarkan perbandingan prestasi beberapa algoritma di dalam perisian WEKA. Peringkat pertama melibatkan perolehan data daripada Nod Akses Berbilang Perkhidmatan (MSAN) menerusi pengukuran dalam persediaan makmal VDSL2. Makmal telah disediakan untuk menyerupai rangkaian VDSL2 sebenar. Beberapa jenis kerosakan yang diperkenalkan menyebabkan variasi data yang berbeza, maka data yang diperoleh perlu menjalani peringkat pra-pemprosesan. Latihan data pra-proses dinilai menggunakan kaedah penilaian pecahan peratusan. Klasifikasi jenis kerosakan talian VDSL2 dinilai berdasarkan peratusan ketepatannya dan kerosakan yang diklasifikasikan dengan betul. Menerusi kaedah yang dicadangkan di atas, sistem yang dibangunkan untuk pengkelasan kerosakan talian VDSL2 adalah tepat 98% ketepatan. Manakala, bagi model regresi bagi pengesanan lokasi kerosakan, ketepatan model menghasilkan pkeali korelasi 0.9933 bagi validasi model menggunakan data yang sama seperti data latihan. Manakala pengesahan menggunakan data VDSL2 harian baharu mampu mencapai ketepatan 100% dan pekali korelasi 0.36.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

DSL	–	Digital Subscriber Line
VDSL2	–	Very-High-Bit-Rate Digital Subscriber Line 2
MSAN	–	Multi-Service Access Node
DT	–	Decision Tree
kNN	–	k-Nearest Neighbour
MLP	–	Multi-Layer Perceptron
NB	–	Naïve Bayes
RF	–	Random Forest
LR	–	Linear Regression
Ω	–	Ohm
V	–	Voltage
D	–	Aperture diameter

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of the study

The government of Malaysia is committed to providing a minimum internet speed of 20 megabits per second (Mbps) in 2020. The Communications and Multimedia minister has expressed the intention of achieving broadband connections that are "double the speed, half the price" for the people, which is also one of the government's new policies to improve the quality of broadband services in the country [1]. Telekom Malaysia (TM), being the largest government-linked company, has reduced broadband packages' prices to more than 30% and has introduced more affordable entry-level packages following the implementation of Mandatory Standard on Access Pricing (MSAP).

TM offers two key broadband services, which are Streamyx and Unifi. Those in landed properties or high-rise areas can contemplate a 100Mbps speed upgrade. In areas where Unifi infrastructure is not available yet, Streamyx users' broadband speed is upgradable up to 8 Mbps, which is the highest speed available for the service. The lowest high-speed internet package offered by TM is the 30Mbps internet package, the price of which has been reduced from RM139 to RM119 following the implementation of MSAP. Table 1.1 shows the prices and speed rates for Unifi internet packages that offer high speeds at lower prices [2]. It can be observed that the broadband speed has increased for each internet package price compared with those offered in year 2018.

Table 1.1: Unifi price and speed comparison [3].

TM Package			
Price (2018)	**Speed before MSAP**	**Price (2022)**	**Speed after MSAP**
RM129	10 Mbps	RM119	30 Mbps
RM139	30 Mbps	RM159	100 Mbps
RM329	100 Mbps	RM199	300 Mbps
		RM249	500 Mbps
		RM349	800 Mbps

Broadband technology is a technology with transmission speeds that are faster than the primary rates of Integrated Services Digital Network (ISDN) at 1.5 or 2.0 Mbps [4]. Some of the common broadband technology types are the Digital Subscriber Line (DSL) technology, cable modems, fibre optic cables and Wireless Local Area Network (WLAN). As of August 2018, Malaysia had accomplished 94.4% broadband penetration in populated areas of the country. Thus, the Communications and Multimedia Ministry's target of accomplishing 95% broadband coverage in populated areas throughout the country by 2020 is achievable. At the moment, the minimum speed offered by Telekom Malaysia is 30 Mbps. Even so, penetrating the remaining 5% of the populated region to achieve the minimum internet speed is challenging and costly due to various territories, populace densities and contrasting separations to existing broadband framework and systems [5]. 70% of Telekom Malaysia's transmission lines currently use copper cables. Users in urban areas would not have any complications or issues in achieving high internet speeds due to fibre infrastructures, which are mostly fully deployed in urban areas. However, users in sub-urban areas would have difficulties in achieving high-speed internet services, and thus there is a need for the integration of fibre optic and copper cables, which may aid in increasing internet speed in order to achieve the broadband penetration target. The integration of fibre optic and copper cables is possible with Digital Subscriber Line (DSL) technologies based on the International Telecommunication Union (ITU) regulation of standardisation. DSL is a network access technology that utilises copper wire pairs as the data transmission medium to or from users. DSL technology allows internet and telephone services to work over the same phone line, enabling customers to use both internet and voice simultaneously. DSL offers assigned point-to-point public network access. The connection of DSL technology is usually between the service provider's central office and the customer's site. DSL attracts a considerable amount of interest from service providers due to its ability to provide high-bandwidth transmission rates to various localities, as well as requiring rather small modifications of current network infrastructures.

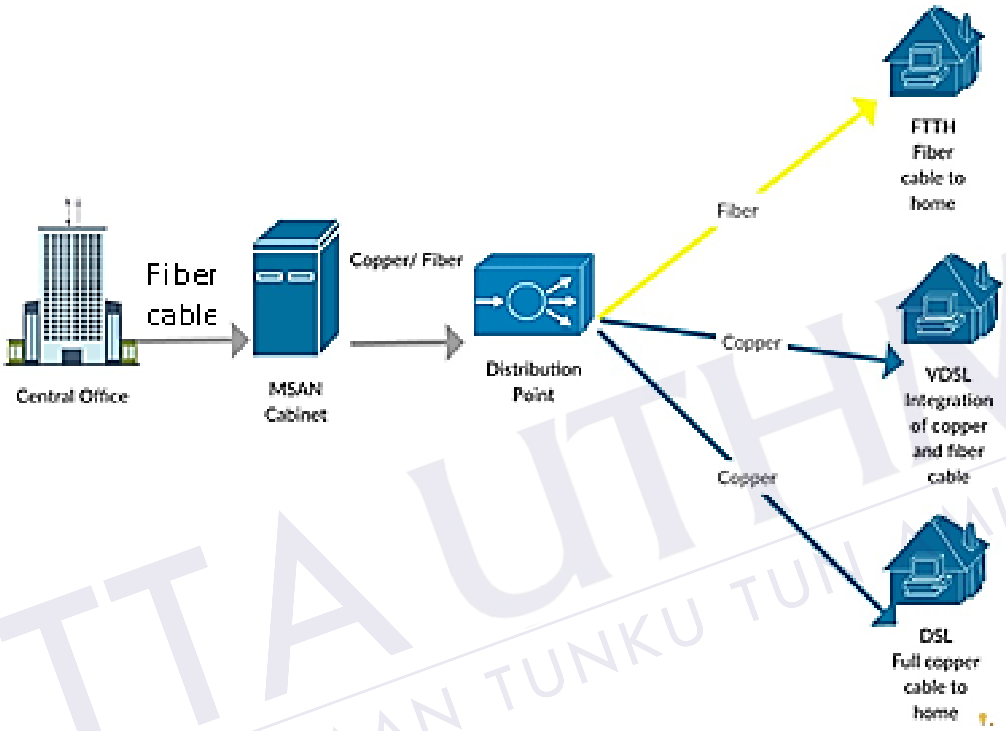

Figure 1.1: TM access network infrastructure.

Figure 1.1 shows the typical TM access network infrastructure. The infrastructure consists of the central office (HQ), Multi-Service Access Node (MSAN) cabinet, distribution point (DP) and customers' premises. The backbone of the infrastructure is basically fibre cables, starting from the central office to the MSAN cabinet. The connection between the MSAN cabinet to the distribution point is usually established using fibre cables in urban areas, but some sub-urban areas still are deployed with copper cables as the medium. The connection from the distribution point to customers' premises can be classified into three types of connection. The first type of connection is based on Fibre to the Home (FTTH), where fibres are laid from the MSAN cabinet to customers' premises. A pure fibre network can deliver 100 Mbps of data to homes and offices at distances up to 20 km between the optical line terminal (OLT) in the exchange and the optical network unit (fibre modem) at users' premises

[6]. The second type of connection is the integration of fibre and existing copper cables. The technology type that corporate with this integration is Very-High-Bit-Rate Digital Subscriber Line 2 (VDSL2), which offers speeds up to 50 Mbps for copper distances up to 1.5 km from the MSAN cabinet to the distribution point. The third type of connection of the infrastructure usually utilises full copper cable connection from the MSAN cabinet to customers' premises, which is usually used by the Asymmetrical Digital Subscriber Line (ADSL) technology and lower. ADSL technology supports up to 8Mbps internet speed for distances up to 5 km.

1.2 Problem statement

As highlighted in Section 1.1, DSL technology utilises copper cables to offer the necessary speeds required by the Malaysian government. Nonetheless, the characteristics of copper cables, which are based on twisting copper cables, cause electromagnetic interferences and are prone to a wide variety of environmental interferences. Thus, the cables are subject to impairments, which will limit the achievable information capacity and degrade the performance of the network. Furthermore, line impairments may introduce alterations in ideal capacitive and resistive values of the line. Some examples of copper line impairments induced by the factors highlighted above are open-wired, short-wired and bridge tap faults. Bridge tap faults, for instance, may result in increased attenuation and may produce impedance discontinuities in the line. Nevertheless, the copper cable faults detection still depend on on the manual detection by utilizing on-site commercial test gear even though there are an extensive amount of network data available. There are available Machine Learning algorithms that cater on the fault diagnosis application such as power line and CAT5 cable but these algorithms is not suitable to be applied as fault diagnosis in a complete DSL network architecture as mentioned above. Therefore, it is important to develop diagnostic tools to make use of the extensive amount of raw VDSL2 network data for the detection of faults and their localisation which may occurs between the MSAN cabinet and customer's house which is less than 1.2km.

1.3 Objectives

This research work embarks on the following objectives:

a) To perform data acquisition and data extraction of VDSL2 datasets from the MSAN server.

b) To develop the classification and regression models for the detections of VDSL2 faults and localisation by utilizing Random Forest machine learning algorithm.

c) To evaluate the accuracy performances of the classification and regression models of VDSL2 faults and localisation, respectively, using VDSL2 lab data.

1.4 Scopes of study

The project scope focused on major components, as listed below:

a) Random Forest machine learning models were designed to cater to VDSL2 copper access network technology.

b) This project focused on the development of VDSL2 technology's classification and regression models that targeted to assess fault diagnosis and fault localisation under real conditions that exist on a VDSL2 infrastructure.

c) The copper access network consisted of the MSAN, traffic generator, router, switch and copper cables. Data were extracted from the MSAN.

d) Useful data from MSAN were extracted during the data cleansing stage.

e) Pre-assessments using WEKA software were done on several machine learning algorithms for the classification and localisation models of the VDSL2 access network.

f) The machine learning algorithm with the best performance for the classification and localisation models was selected.

g) The selected machine learning algorithm was used to train the features of the data gathered to classify the types of faults, which were open, short, bridge tap, partial open and partial short.

h) The efficiency of the classification of faults was evaluated based on the data obtained in the UTHM lab.

1.5 Research contribution

The main contribution of this thesis :

a) Perform automatic data acquisition, extraction and processing of VDSL2 network using Python Programming Language.
b) Develop VDSL2 fault type classification model and fault location using machine learning method and was able to achieve more than 90% of accuracy under 5 minutes of computing time.
c) The VDSL2 fault type classification model and fault location machine learning model is able to be used on online or latest data.

1.6 Outline of the thesis

This thesis consists of five chapters. The first chapter of this thesis introduces the background of the study, the problem statement, the objective, the scope of study and the research contribution.

Next, Chapter 2 of the thesis presents the literature review of previous studies. It consists of the general background of the ITU standardisation in DSL technology, the performance parameters of the copper access network and VDSL2 faults and the general background on artificial intelligence and machine learning models' development. This chapter also presents 10 previous studies on classification and localisation machine learning models in several primary-service applications and in related telecommunication sectors.

In Chapter 3, the settings for the development of the classification model of VDSL2 fault types and the regression of the VDSL2 localisation model are explained. Besides that, all preparation processes, which were data acquisition, pre-assessments of related machine learning algorithms, development using Python programming language and evaluations, are presented in this chapter.

In Chapter 4, an example of the data acquired from the MSAN and an example of the cleansed data are presented. In addition, accuracy performances from pre-assessments done using the WEKA software on several machine learning algorithms for classification and regression and on the evaluation methods are presented. Based on the comparisons of pre-assessment results, one machine learning algorithm and one evaluation method were selected for each machine learning model. The selected algorithms and evaluation methods were applied in the development of the VDSL2 fault-type classification model and the VDSL2 fault-localisation model using Python programming language. Later on, the accuracy reports for the Python machine learning models were validated using several datasets, and their accuracies are presented in this chapter.

Lastly, Chapter 5 presents the conclusion and the limitations of the study. This chapter also includes recommendations for future research.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

This chapter presents the general background of International Telecommunication Union (ITU) standardisation in DSL technology, the performance parameters of the copper access network and VDSL2 faults and the general background on artificial intelligence and machine learning models' development. This chapter also presents 10 previous studies on classification and localisation machine learning models on several primary service applications and in related telecommunication sectors. Most importantly, this chapter serves as the guideline for the development of the VDSL2 fault-type classification model and the localisation regression model.

2.2 ITU standardisation in DSL technology

With the rapid increase in users' demand for resilience and ascendable communication services, broadband communication networks are confronting a growing challenge in accommodating diverse broadband services using a single communication architecture. The newest broadband services use fibre optic cables, which provide the fastest internet connection thus far. However, this type of internet service is still new, and its service areas are quite limited because the laying down of fibre optic cables takes a while to complete. However, wherever it is available, its cost not only competes with that of DSL, and it provides a much faster connection. DSL speeds are influenced by the distance between the subscribers and the local exchange and by the type of DSL technology [7].

As presented in Table 2.1, there are several types of DSL technology: R-ADSL (Rate-Adaptive Digital Subscriber Line), ADSL (Asymmetrical Digital Subscriber Line), ADSL Lite, HDSL (High-Bit-Rate Digital Subscriber Line), IDSL (ISDN Digital Subscriber Line), SDSL (Symmetric Digital Subscriber Line), VDSL (Very-High-Bit-Rate Digital Subscriber Line), VDSL2 (Very-High-Bit-Rate Digital Subscriber Line 2) and G.Fast [8].

Table 2.1: DSL technology types.

Service	**Maximum Distance to Central Office**	**Maximum Upload Speed**	**Maximum Download Speed**	**Notes**
ADSL	5.5 km	1500 kbps	9 Mbps	Asymmetrical
ADSL Lite	5.5 km	384 kbps	1.5 Mbps	No splitter is required
RADSL	5.5 km	384 kbps	8 Mbps	Rate adapts based on distance and quality
IDSL	10 km	144 kbps	144 kbps	DSL over ISDN (BRI)
SDSL	6.7 km	2.3 Mbps	2.3 Mbps	Targets T1 replacement. Symmetrical DSL service.
HDSL	5.5 km	1.54 Mbps	1.54 Mbps	Four-wire, similar to T1 service
VDSL	0.916 km	16 Mbps	52 Mbps	Few installations
VDSL2	1.5 km	100 Mbps	100 Mbps	Higher symmetrical speeds than VDSL
G.Fast	0.25 km	1 Gbps		

Figure 2.1 shows an overview of the family of DSL recommendations based on the International Telecommunication Union Telecommunication Standardization Sector's (ITU-T) standardisation, where ITU-T is the United Nations' specialised agency in the field of telecommunications [9].

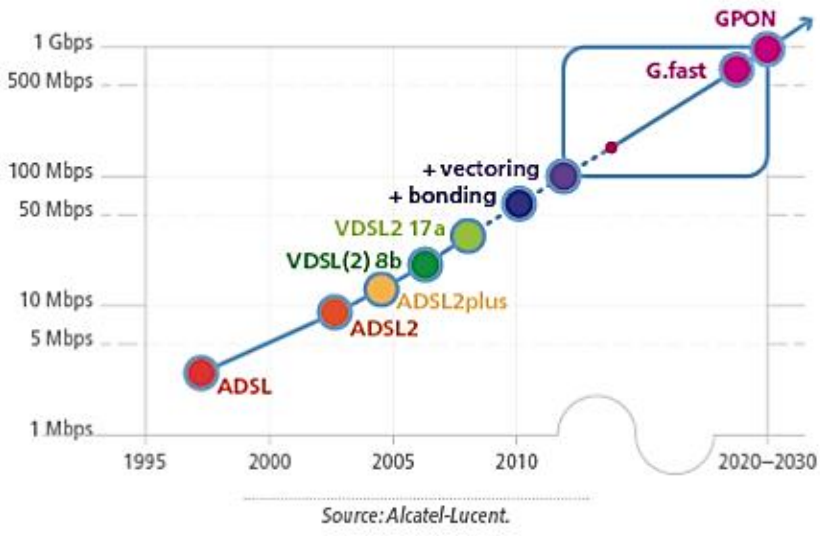

Figure 2.1: Evolution of xDSL technology [10].

The first Digital Subscriber Line (DSL) standard was approved in 1993. Also in the 1990s, ITU–T approved its first ITU–T ATM (Asynchronous Transfer Mode) Recommendations as a key layer technology for many ADSL implementations today. Throughout the late 1990s and the early years of the millennium, ITU–T continued upgrading its ground-breaking suite of xDSL standards. Although governments and operators in many industrialised countries are now looking to deploy or upgrade to fibre networks, it is important to note that globally, in 2014, xDSL still accounted for over half of all internet access lines in use worldwide. Technological innovation continues, with further advances continuing to help prolong the lifetime of existing infrastructures. ITU–T Study Group 15 approved the Very-High-Bit-Rate Digital Subscriber Line 2 (VDSL2) standard in 2005, enabling telecom operators to offer triple-play video, internet and voice services at speeds ten times faster than ADSL. Although many operators are extending fibres into their network, this can prove expensive; in addition, copper lines are not redundant today—indeed, the most exciting use of copper lines promises to be when used in conjunction with fibres cables. ITU–T approved its G.Fast ITU–T 9701 standard in December 2014, which continues the exciting work of getting the most out of legacy copper lines. G.Fast promises to deliver “fibre-like” speeds of up to 1 Gbps over short ranges [11].

ADSL and VDSL technologies are currently being utilised extensively as the dominant broadband technologies for residential and business customers. But at the

APPENDIX C

VDSL2 PSEUDO CODE FOR FAULT-TYPE CLASSIFICATION AND LOCALISATION

- Imports various python modules needed in this code.
- Read "UTHM-Data.xlsx " contents.
- Replace the word “Activating” and “Activated” to integer “0” and “1”.
- Access all cell of “Line SNR margin upstream(dB)” parameter in "UTHM-Data.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Access all cell of “Line SNR margin downstream(dB)” parameter in "UTHM-Data.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Access all cell of “Signal attenuation upstream(dB)” parameter in "UTHM-Data.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Access all cell of “Signal attenuation downstream(dB)” parameter in "UTHM-Data.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Access all cell of “'Maximum attainable rate upstream (Kbps)” parameter in "UTHM-Data.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Access all cell of “'Maximum attainable rate downstream (Kbps)” parameter in "UTHM-Data.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Access all cell of “'Actual line rate upstream (Kbps)” parameter in "UTHM-Data.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.

- Access all cell of “'Actual line rate downstream (Kbps)” parameter in "UTHM-Data.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Drop all data and columns of 'TIME_STAMP', 'LENGTH', 'Test_Type', 'Card_Type', 'TECHNOLOGY', 'PORT_NUMBER' from dataframe.
- Assign column “Test_Type” as first target or “y” variable.
- Assign column “LENGTH” as second target or “z” variable.
- Assign StandardScaler library to “scaler” variable.
- Transform all training parameters (X) using their respective mean and variance that has been calculated by the scaler variable.
- Split all training and target parameter into training and testing data.
- Train the first model, random forest classifier (rfc1), using the hyperparameter obtain from GridSearchCV function.
- Train the second model, random forest regressor (rfc2), using the hyperparameter obtain from GridSearchCV function.
- Training the random forest classifier (rfc1) model using training data features and target.
- Training the random forest regressor (rfc2) model using training data features and target.
- Assign the calculation of mean absolute error between z target for test data set and its predictions, to “mae2” variable.
- Output the value of “mae2” variable
- Output "===Fault Type Testing Confusion Matrix ===" text.
- Output the confusion matrix to evaluate the accuracy of a classification between the actual “y” target for test data set and its prediction.
- Output "=== Classification Report ===" text.
- Output text report showing the main classification metrics between the actual “y” target and its prediction.
- Output the coefficient of determination R^2 of the prediction of “y” target for training data set.
- Output the coefficient of determination R^2 of the prediction of “y” target for test data set.

- Output the coefficient of determination R^2 of the prediction of “z” target for train data set.
- Output the coefficient of determination R^2 of the prediction of “z” target for test data set.
- Output the mean absolute error between z target for test data set and its predictions, to “mae2” variable.
- Output "===scores localization===" text.
- Assign and output the value of mean squared error between z target for test data set and its predictions, to “MSE” variable.
- Assign and output the value of root mean squared error between z target for test data set and its predictions, to “RMSE” variable.
- Assign and output the value of mean absolute percentage error between z target for test data set and its predictions, to “mape” variable.
- Output the Cohen kappa score for “y” target of test data set and its prediction.
- Output the value of calculation of Pearson correlation coefficient and the p-value for testing non-correlation, between fault location parameter “LENGTH” and “Signal attenuation upstream(dB)” parameter.
- Assign a variable name “Pkl_Filename1” to rename “Pickle_RL_Model1.pkl”.
- Store random forest classifier (rfc1) model in “Pkl_Filename1.”
- Assign a variable name “Pkl_Filename2” to rename “Pickle_RL_Model2.pkl”.
- Store random forest regressor (rfc1) model in “Pkl_Filename2.”
- Open “Pkl_Filename1” file as assign the it to “Pickled_LR_Model1” variable.
- Read " indooroutdoorVDSL2autoprofile.xlsx " contents.
- Replace all cell with keywords “-“ to “-999.99” integer.
- Replace the word “Activating” and “Activated” to integer “0” and “1”.
- Access all cell of “Line SNR margin upstream(dB)” parameter in " indooroutdoorVDSL2autoprofile.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Access all cell of “Line SNR margin downstream(dB)” parameter in " indooroutdoorVDSL2autoprofile.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.

- Access all cell of “Signal attenuation upstream(dB)” parameter in " indooroutdoorVDSL2autoprofile.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Access all cell of “Signal attenuation downstream(dB)” parameter in " indooroutdoorVDSL2autoprofile.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Access all cell of “'Maximum attainable rate upstream (Kbps)” parameter in " indooroutdoorVDSL2autoprofile.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Access all cell of “'Maximum attainable rate downstream (Kbps)” parameter in " indooroutdoorVDSL2autoprofile.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Access all cell of “'Actual line rate upstream (Kbps)” parameter in " indooroutdoorVDSL2autoprofile.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Access all cell of “'Actual line rate downstream (Kbps)” parameter in " indooroutdoorVDSL2autoprofile.xlsx " data frame with a value of “0” in column “STATUS” and replace with “-999.99”.
- Set parameter in column 1 of “indooroutdoorVDSL2autoprofile.xlsx” data set as target and assign it to “target_cols1” variable.
- Drop all data and columns of 'TIME_STAMP', 'LENGTH', 'Test_Type', 'Card_Type', 'TECHNOLOGY', 'PORT_NUMBER' from data frame and assign it to “pred_cols1” variable.
- Assign StandardScaler library to “scaler” variable.
- Transform all parameters (pred_cols1) using their respective mean and variance that has been calculated by the scaler variable and assign all the values to “pred_cols2” variable.
- Predict VDSL2 fault type using “Pickled_LR_Model1” and assign the predicted value to “Ypredict1” value.
- Open “Pkl_Filename2” file as assign the it to “Pickled_LR_Model2” variable.
- Drop all data and columns of 'TIME_STAMP', 'LENGTH', 'Test_Type', 'Card_Type', 'TECHNOLOGY', 'PORT_NUMBER' from data frame and assign it to “pred_cols3” variable.

- Assign StandardScaler library to “scaler” variable.
- Transform all parameters (pred_cols3) using their respective mean and variance that has been calculated by the scaler variable and assign all the values to “pred_cols4” variable.
- Predict VDSL2 fault location using “Pickled_LR_Model2” and assign the predicted value to “Ypredict2” value.
- Computes and output the subset accuracy of multilabel classification between actual value of “y” target of " indooroutdoorVDSL2autoprofile.xlsx " data and its predicted value.
- Assign the calculation of mean absolute error between z target of " indooroutdoorVDSL2autoprofile.xlsx " data set and its predictions, to “mae2” variable.
- Assign and output the value of mean squared error between z target for test data set and its predictions, to “MSE” variable.
- Assign and output the value of root mean squared error between z target for " indooroutdoorVDSL2autoprofile.xlsx " data set and its predictions, to “RMSE” variable.
- Assign and output the value of mean absolute percentage error between z target for " indooroutdoorVDSL2autoprofile.xlsx " data set and its predictions, to “mape” variable.
- Output the confusion matrix to evaluate the accuracy of a classification between the actual “y” target for for " indooroutdoorVDSL2autoprofile.xlsx " data set and its prediction.
- Output text report showing the main classification metrics between the actual “y” target and its prediction.
- Compute and output the Cohen kappa score between “y” target of “indooroutdoorVDSL2autoprofile.xlsx “ data set and its prediction value.

APPENDIX I

VITA

The author was born in April 4, 1993, in Perak, Malaysia. She went to Sekolah Menengah Kebangsaan Bandar Tasik Puteri, Selangor, Malaysia for her secondary school. She pursued her degree at the University of Tun Hussein Onn, Johor, and graduated with the Bachelor of Electronic Engineering with Honours in 2019. He then enrolled at the University of Tun Hussein Onn, Johor, in 2019, where he was awarded the Master of Electrical Engineering in 2022. She authored three papers in areas of Machine Learning application in Digital Subscriber Line (DSL).