PREDICTION MODEL FOR THE DISTRIBUTION TRANSFORMER FAILURE USING CORRELATION OF WEATHER DATA

A Thesis

by

EUN HUI KO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Mladen Kezunovic
Committee Members,	Chanan Singh
	Shankar P. Bhattacharyya
	Ivan Damnjanovic
Head of Department,	Miroslav Begovic

August 2019

Major Subject: Electrical Engineering

ABSTRACT

Distribution Transformer (DT) is an integral component of a distribution network. Electric utilities have invested interest in reducing DTs failure rates. This paper presents a method for prediction of probability of DT failure by analyzing a correlation between weather data and historical DT failure data. Logistic regression prediction model is used in order to predict DT failure, and to extract the correlation between weather parameters and DT failure rates. Accuracy of prediction is reliable, which is presented using evaluation metrics. This method not only has a vital significance for the maintenance of DTs, but also improves the economic efficiency and reliability of distribution network operation.

DEDICATION

I dedicate this work to my family.

ACKNOWLEDGEMENTS

I would like to sincerely thank my academic advisor Dr. Kezunovic for his guidance throughout my education at Texas A&M University. Through his patient and thoughtful mentorship, he helped me make the best decisions regarding my thesis and coursework. I owe my successful academic experience to him.

Thanks go to my committee chair Dr. Mladen Kezunovic and my committee member Dr. Chanan Singh, Dr. Shankar P. Bhattacharyya, and Dr. Ivan, Damnjanovic for their support throughout my research.

I am also thankful to my coworker Dr. Tatjana Dokic and graduate students in Dr. Kezunovic's lab in electrical and computer engineering department for being supportive and making my experience a great one.

Finally, thanks to my family for their encouragement.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor Mladen Kezunovic [advisor], and Professor Chanan, Singh, Professor Shankar P. Bhattacharyya, of the Department of Electrical & Computer Engineering, and Professor Ivan, Damnjanovic of the Department of Civil Engineering.

All work for the thesis was completed by the student, under the advisement of Professor Mladen Kezunovic of the Department of Electrical & Computer Engineering.

Funding Sources

This research was funded by Korea Electricity Power Cooperation (KEPCO) in Korea.

NOMENCLATURE

DT	Distribution Transformer
AT	Average Temperature
HT	Highest Temperature
RH	Relative Humidity
MWS	Maximum Wind Speed
WG	Wind Gust
PT	Precipitation
AUC	Area Under the Receiver Operating Characteristic Curve
ROC	Receiver Operating Characteristics
LV	Low Voltage
KMA	Korea Meteorological Administration
tp	true positive
fp	false positive

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Firstly, the importance of distribution (DT) and the causes of DT failure is addressed. The portion of weather factors as a main cause of DT failure is pointed out. After we look over the impact of DT failure dividing quantitative losses and qualitative losses, then the research objective is considered.

1.1 Motivation

The Distribution Transformer (DT) is a vital link in the chain of power apparatus supplying electric power to the customers [1]. DT failures have a major financial impact on both utility company and customers. The utility companies invest large amount of money annually for maintenance of distribution transformers [2, 3]. Utility customers may experience loss of power due to transformer failures, which also translates to a potential economic impact for the customers [1]. Likelihood and impact of transformer failure rates is regularly assessed and DTs may be replaced according to set of criteria during the evaluation [1]. However, there is still a limited capability to mitigate DT failures caused by weather [4, 5]. A practical utility example given in Table 1 [6] illustrates that the rate of DT failures caused by weather is still high and presents the second highest cause after aging, also affected by weather.

Table 1.1: The causes of DT failure from JeonllaNambo of South Korea (2011-2018) [6]

Cause	The number	Rate [%]
Aging	337	27.2
Weather	333	26.9
Corrosion	155	12.5
Animal contact	145	11.7
Lack of maintenance	92	7.4

Table 1.1: (continued) The causes of DT failure from JeonllaNambo of South Korea [6]

Cause	The number	Rate [%]
Overloading	87	7.0
Foreign objects	25	2.0
Tree	18	1.5
People error	16	1.3
Manufacture error	15	1.2
Flooding	13	1.0
Installment error	3	0.2
Fire	1	0.1
Total	1,240	100

1.2 The impact of DT failure

When we talk about impact of losses due to outages occurrence, we can divide them into quantitative losses and qualitative losses. They matter significantly for both utilities and customers.

The quantitative losses include workforce time engagement and economical loss. Once failure occurs, employees in utility company respond to the complains of customers who experienced outage, by providing maintenance or replacing equipment. This process can take from one day up to a week depending on the kind of an outage. The employees may have to delay their own scheduled work while dealing with a sudden and unexpected outage problem and have additional urgency taking care of customers who experience outage. Sometimes even the mental stress caused by communicating with angry customers may have a large impact on employee's work efficiency [7]. In addition, utility company faces an economic loss by not selling electricity during the time of an outage, and maintenance cost imposed by replacement of failed equipment. [8]

When it comes to qualitative losses, utility company may lose the trust by the customers about reliability of their service [9]. Especially when it comes to customers who run their own commercial business, utility can lose customer's trust even for a short outage, which may create substantial economic loss for customers [10]. As a result, utility company may have high probability of losing future customers because customers who experience a sudden outage can transfer to another utility company if available.

1.3 Research Objective

The utility companies have many challenges to prevent DT failure. There are various causes of DT failure, which are aging, weather, corrosion, animal contact, out of maintenance, overloading, etc. (see Table1). The utilities study these causes and implement diagnosis to try to decrease these factors [11]. In some cases, it is not obvious what combination of the causes has the higher impacts on DT failures, so researching the cause-effect relationship between weather parameters and failures requires further efforts [12, 13]. Our research identifies weather predictors causing DT failure, and analyzes how they correlate by presenting probability prediction. If utility companies embrace such methods in the future, they can be utilized as an essential risk-based decision-making tool, which is proactive.

1.4 Conclusions

The impact of DT failure is extensive, from economic loss to the loss of customer's reliability. Therefore, it is essential for the utility companies to try to decrease impact of DT failure. The predictive methods of DT failures caused by weather have not been addressed extensively in the past. This is because weather is not precisely predictable and there is limitation in being able to prevent DT failure due to strength and volatility of nature. Thus, if we analyze the correlation between DT failure rates and weather parameters, and prepare for future weather based on weather forecast, failure could be reduced by advanced implementation of mitigation methods such as timely positioning field crews or timely replacement or repair of transformers

2. SURVEY OF PRIOR WORK

In this section, we take a look at different maintenance practice and published work by others. The difference and limits between them and my research are analyzed. Based on studying the prior work, I present the research topic that I will be focusing on for my research.

2.1 Different maintenance practice

The causes of transformer failure are separated into fundamental and derived causes [12, 14]. The fundamental causes are failures that occur in the secondary interior components of the transformer due to mechanical, electrical or thermal stress such as winding, bushing, tap, core, tank, protection system, and cooling system [14, 15, 16]. On the other hand, derived causes are deduced from primary causes causing failure, which are aging, weather, overloading, corrosion, lack of maintenance, animal contact, installment error, and etc. Our research considers also some extra causes of DT failure. In summary, many factors are affecting DT failure comprise aging, weather, overloading, corrosion, out of maintenance, animal contact, installment error, people's error, etc. in JeonllaNamdo (see Table 1). Among the causes, aging, corrosion and overloading constitute 47%, which combined is the largest reason (Table 1). While we sampled one region of South Korea, the causes of DT failure in other regions are representative of the entire country. In order to decrease the failure rate caused by aging, corrosion and overloading, utility company implements different methods including internal inspection. It comprises of visual examination such as checking oil leakage and internal pressure [15], electric measurements for electrical insulation resistance tests [14], short circuit test [15, 17], leakage current measurement [15, 18], and gas chemical tests such as Dissolved Gas Analysis (DGA) [19,20], Furan Analysis [21] , Health Index [22, 23] and infrared thermography [19, 24].

It is relatively easy to figure out faults of DT by visual inspection and electric measurement methods because it takes short time to perform tests and the visibility of impacts is clear. However, to identify DT fault by visual inspection and electric measurements may not be sufficient. In contrast, DGA is currently the most accurate method to verify the internal status of insulation of gas and oil-paper transformer to determine a need for maintenance [25]. But, it takes a lot of time to carry on the procedure, and it is costly to implement for all transformers. In addition, extracting oil from on-site pole mounted transformer for DGA is a dangerous work, and also not convenient from the operation standpoint because DT has to be de-energized. To solve these problems, Furan analysis is introduced. It takes very short time because the result of test comes out on site instantly, but the approach been still studied, and not standardized until now [21]. The Health Index represents a practical tool that combines the results of operating observations, field inspections, as well as site and laboratory testing providing the overall health of the asset. The Health Index can effectively be employed but it needs justification for a capital plan because it needs various diagnosis tools to implement the analysis. Furthermore, infrared thermography is relatively easy method, but there is limit that it represents temperature only at a certain time, and it primarily can be used for verifying aging of connection components like bushing and connection lines.

As it may be noted, none of these standard practices in the electric utility consider causes of DT failure inflicted by weather.

2.2 Published work by others

While there are a quite a few papers that study the causes of DT failure, very limited research is done on prediction of weather-related distribution transformer failures. One of them [26] analyzes the root causes of failure, overvoltage and overloading, as well as the electrical surge affecting the core and winding failure. It considers fundamental causes of DT failure, and does not address weather parameters as causes. The other study [27] shows the main causes (Tree, animal, contamination, nature disaster, and human) according to geographical regions. This paper focuses a classification of causes to impact of regional failure rates of Thailand. According to study about DT failure root causes in India [28], they classify utility side and manufacturer sides. Utility side includes prolonged overloading, improper LT and HT protection, single phase loading, unbalanced loading, and etc. while there are poor insulation covering on conductor, improper joints and connection, and incomplete drying, and etc. as causes of manufacturer side. It mentions extreme weather as a one of external causes, however, does not consider the weather parameters as a principal cause. Literature explains how the causes of DT failure, represented through a diagnosis values represent current status of a DT. The insight into weather parameters related causes is largely missing. Some studies analyzed lightning impact, but without the insight into complete weather parameters [29, 30, 31]. Their focus was on finding the causes of mechanical external failure. Such analysis focuses on lightning-induced voltages and correlation between low voltage surge arresters and grounding resistance [32].

When it comes to studying prediction of DT failure, one of the references creates an algorithm for predicting the probability of failure applying extreme value theory [33]. This research of prediction generates a derivation from expected valued using an extreme distribution, however, it does not provide decision-making tool. In this study, they extract history data and current status of

DT using sensor measurement are considered. It utilizes the degree of polymerization value in order to identify overall health of the insulation paper. Another prediction method, AMI (Advanced Metering Infrastructure) reading [34] is used to collect voltage data of DT for the limited period. It then analyzes each prediction using various approaches, which are DNN (Deep Neural Networks) features using digital signal processing techniques, and gradient-boosting machines [35]. This research uses AMI reading as a collection of historical data, but the information about DT health is very limited utilizing voltage. The data represents only an overvoltage, which is not enough to verify DT failure.

2. 3 Conclusions

As explained, utility companies implement various diagnosis and measurement techniques according to standard and process. However, DT failure still occurs often, which tells that we need to research other approach. Our research provides such an approach by using logistic regression to analyze weather-related failures. The main advantage is that it can be used as an input to the replacement decision-making process using correlation of weather data, hence alleviating the limits of existing maintenance met

3. PROBLEM FORMULATION

For problem formulation, we introduce the hypothesis that the weather parameters are expected to have high impact on DT failure. Then the hypothesis premise is that the correlation between weather data and DT failure can be represented with the probability prediction by using logistic regression. Eventually, we present the evaluation of the results by analyzing the AUC (Area Under Curve) and coefficient values.

3.1 Research Hypothesis

The causes of DT failure are various and weather impact accounts for about 27%. Power utility company primarily implements diagnosis against causes of aging and overloading of DT. However, there is a limitation of insight for failure caused by weather. Utility company personnel know that there is correlation between weather and DT failure (Table 1), [36]. But they do not know about the correlation between specific parameter of weather and failure. The parameters of weather data include lightning, average temperature, highest temperature, relative humidity, maximum wind speed, wind gust, and precipitation. Highest temperature, precipitation, and lightning are expected to have higher impact on DT failures according to empirical data. The hypothesis premise is that the correlation between weather data and DT failure rates can be solved by logistic regression. Furthermore, in terms of high temperature, the hypothesis which the probability of failure can increase with high temperature by 86°F or 89.6°F is assumed. Therefore, the degree of high temperature is classified into 86°F or below, 86°F - 89.6°F, and 89.6°F or above.

3.2 Proposed approach

While the utility companies are aware that there is a correlation between weather and DT failure according to their field experiences, typically it is not known which weather parameter is most relevant among different parameters, and such correlation is not quantified. In order to verify the correlation between weather parameters and DT failure rates, it is necessary to collect historical weather and DT failure data. The information of DT failure and historical weather data is processed by using logistic regression. The goal of the logistic regression implemented in this study is prediction of future failures. We calculate coefficients of correlation between parameters and failure and apply to the data test set next.

Through this method, the degree of correlation between parameters of weather data and DT failure is measured and the probability of each failure is calculated.

3.3 Methodology

Extracting DT failure data is the first step. We will illustrate the process by using an example. In our case we are considering data from South Korea associated with Korea Electric Power Company. The area where the largest number of failures occurred the most in South Korea is selected and events are sorted by the dates of failure and areas. The next step is matching the areas of failure and areas where weather stations are located. The weather stations are selected at one of the closest locations. The last step is extracting the historical weather data according to the times and locations of DT failures. The weather station provides weather data about Lightning [0/1] (LI), Average Temperature [°F] (AT), Highest Temperature [°F] (HT), Relative Humidity [%] (RH), Maximum Wind Speed [m/s] (MWS), Wind Gust [m/s] (WG), and Precipitation (mm) (PT).

The information of DT failure history and associated weather data is analyzed by using logistic regression. The AUC and coefficient values are used for the results evaluation.

3.4 Experimental validation

During testing, we calculated the accuracy of prediction and consider maintenance of DT according to result of prediction. Logistic regression is a recognized method for gaining probability from relationship between dependent variables and independent variable. The DT failure data is extracted from real utility company for the most recent five years. There are 22 cities, 24 branches and 16 Korea Meteorological Administration (KMA) weather stations in JeonllaNamdo. The historical weather data is extracted from closest KMA location where failure occurs. The region of one branch covers on average 225.28mi^2. 237 historical DT failures are extracted and weather data is also collected according to the dates of these failures. For implementation of logistic regression, 148 weather data sets for dates when there was no outage are extracted. The total number of weather data samples is 385, which are sum of 237 dates with failure and 148 dates without failure. 90% of the total weather data is used for training set and remaining 10% of the data is used for testing set. The possible event of failure is one out of two; Y=1 and Y=0, which means failure occurrence and no failure respectively. The result is presented by probability of each failure and the accuracy of probability is described by using Receiver Operating Characteristics (ROC) graph.

3.5 Conclusions

The weather has an impact on DT failure and it is necessary to figure out the degree of correlation between the parameters of weather and DT failure. The reliable DT failure data and historical weather data are extracted and processed. In order to validate the correlation, logistic regression is used and analyzed. The result values of logistic regression are the probability of failure events. The accuracy of probability is computed by using training set and test set of failure events. Through the decision-making process based on the predictive failure analysis the utility company can develop efficient maintenance strategies for DT.

4. DATA SOURCES

In this section, we look over how historical DT failure data and weather data is addressed. Especially, we study from what source we select and why we choose the specific data. Thus, we elaborate the process of preparing the data for the study and make it methodological.

4.1 Historical outages

We studied step down transformers (22.9KV-220V) used in the distribution sectors in South Korea. The 237 DT failure events we selected comprise the data set. The various predictor variables are used for our study for modeling purpose. We also applied preprocessing steps to extract useful features and prepare data for prediction algorithm. We collected data for modeling outage events used for prediction and analysis starting from year 2011 up to year 2018. Before applying the logistic regression, let us go through our adjusted modeling data in simple descriptive statistical measures. The historical outages are extracted for five causes: lighting, tree contact, snow, rain, and dust. We select weather related parameters among causes of DT failure that utility company classifies. The distribution of locations of DT failures is demonstrated by mapping with ArcGIS [37] as shown in Fig 1.

Figure 4.1: Mapping of DT failure in JeonllaNamdo

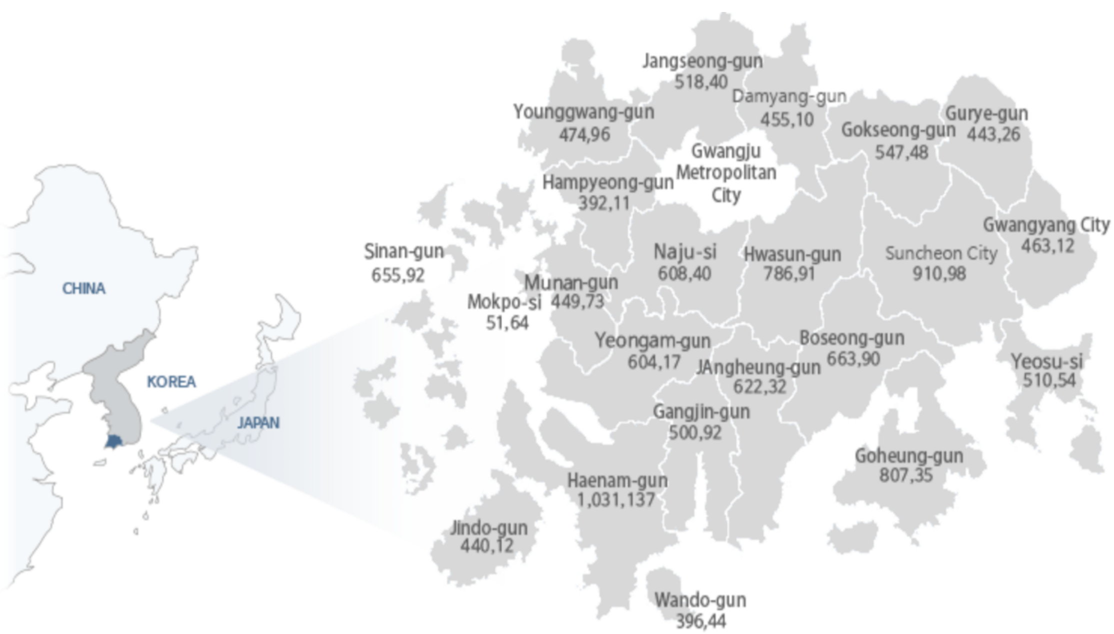


Figure 4.2: JeonllaNamdo area [38]

The region is JeonllaNam-do in South Korea and the size of the region is 4,729mi² with the popularity of 190 million and 842,668 households. It consists of 22 cities as described in Fig. 2, which account for 12.3% of the entire land area. The JeonllaNam-do area ranks first in the number of DTs and poles, route length and the length of low voltage lines in Korea as shown in Fig 3 and Fig 4. There are 1,447,320 poles and 242,577 DTs in JonllaNamdo area (see Table 2), which accounts for the largest portion in the country. Furthermore, the area has more countryside and seaside coverage, more than in other states, which make it is more vulnerable to weather impacts.

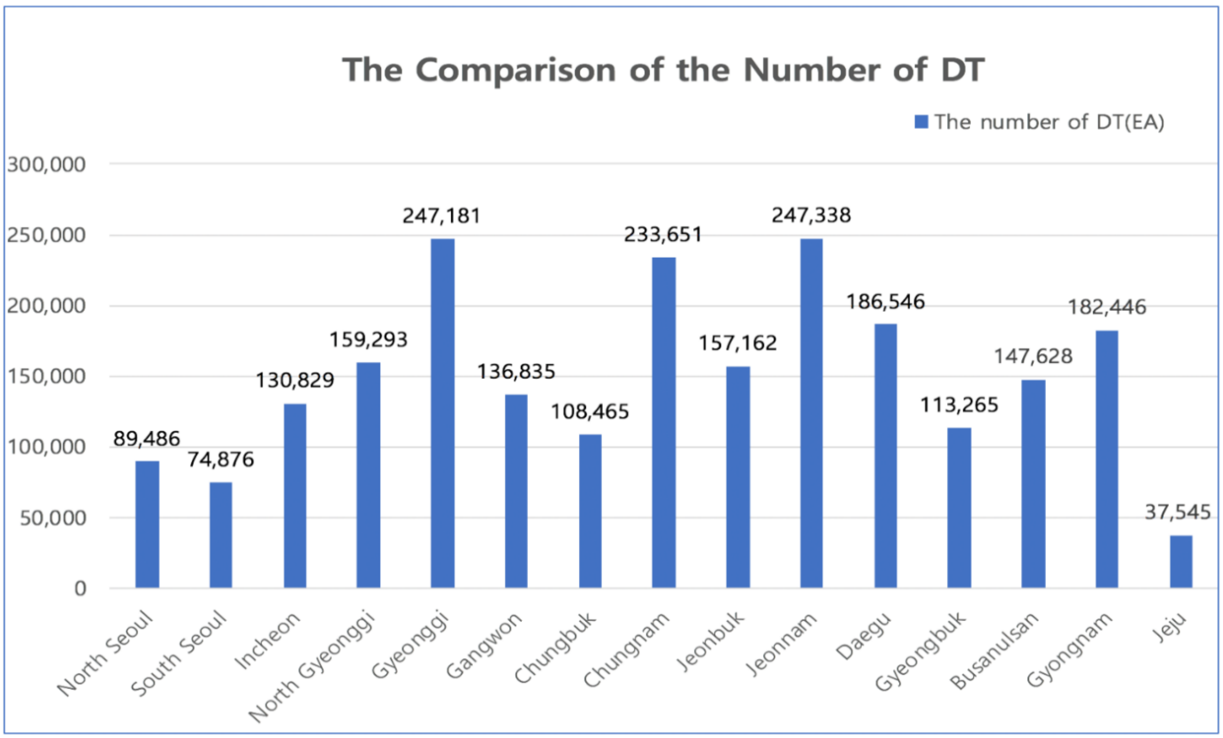


Figure 4.3: Distribution facilities in Korea [38]

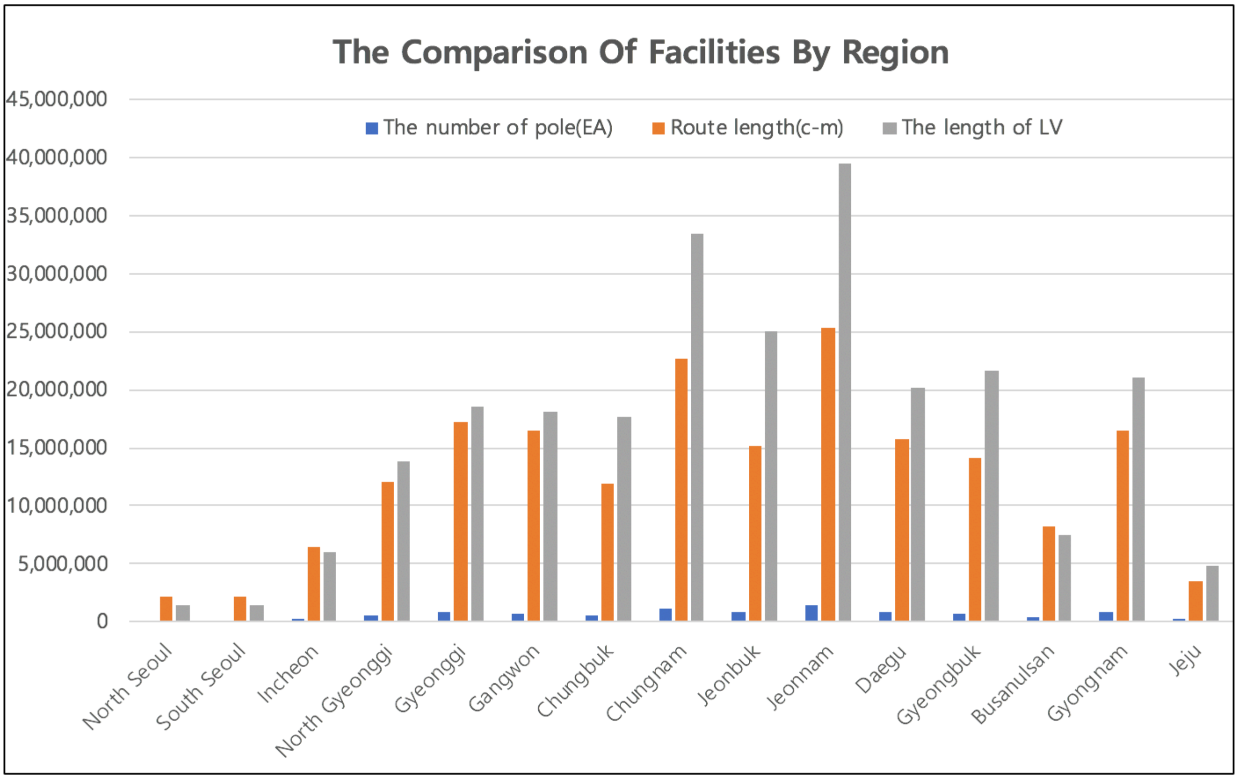


Figure 4.4: The comparison of Distribution facilities by region [38]

Table 4.2: Distribution facilities in JeonllaNamdo [6]

Pole	Distribution line		Ground line (m)
	Route length (c-m)	Total length (m)	
1,447,329	64,706,019	180,300,997	23,442,293

Transformer			Protective Device		
Bank	Number	Capacity (kVA)	Breaker	Equipment	COS
103,793	242,577	9,251,963	11,868	1,412	71,933

From 1/1/2011 to 11/2/2018 the number of total outages was 1,025, where failures caused by weather account for 237, which represents 24% of the total. The causes of all outages include lightning, three contact, snow, aging, overload, bird contact, people fault, installation fault, manufacture fault, corrosion, fire, etc.

The data used for the logistic regression is extracted from outages caused by weather. Aging and overload are related to temperature. Since those are not direct causes, they are excluded.

The dates which have outages caused by weather are denoted as Y=1 and the dates which don't have any outages are denoted as Y=0, and historical weather data are extracted for those dates.

JeonllaNamdo consists of 22 cities (see Table 3), and the size of cities averagely 225.28 mi^2.

Table 4.3: Size of a city in JeonllaNamdo [38]

No.	City	Size (mi^2)
1	Kwangju	193.51
2	Mokpo	19.94
3	Yeosoo	197.12
4	Sooncheon	351.73
5	Naju	234.90
6	Kwangyang	178.81
7	Damyang	175.72
8	Koksung	211.38
9	Gurae	171.14
10	Goheong	311.72
11	Bosung	256.33
12	Hwasoon	303.83
13	Jangheong	240.28
14	Kangjin	193.41
15	Haenam	398.21
16	Youngam	233.27
17	Mooan	173.64

Table 4.3: (continued) Size of a city in JeonllaNamdo [38]

No.	City	Size (mi^2)
18	Hampyeong	151.39
19	Youngkwang	183.38
20	Jangsung	200.16
22	Wando	153.07
22	Jindo	169.93
23	Sinan	253.25

4.2 Historical weather data

The region is located in the southwest part of South Korea, thus has many areas, which spread along long seaside as shown in Fig 1 and in Fig 2. This is the reason it has a strong oceanic climate. The climate of the region has different properties of weather between seaside cities vs islands and inland mountainous areas. The average of highest temperature from 2011 to 2015 is 94.46°F as values given in Fig 5 suggest, and the average temperature in August is 78.62°F as described in Fig 6. Additionally, there is a lot of precipitation in the region in April and August with an average of 159.6 mm as shown in Fig 7.

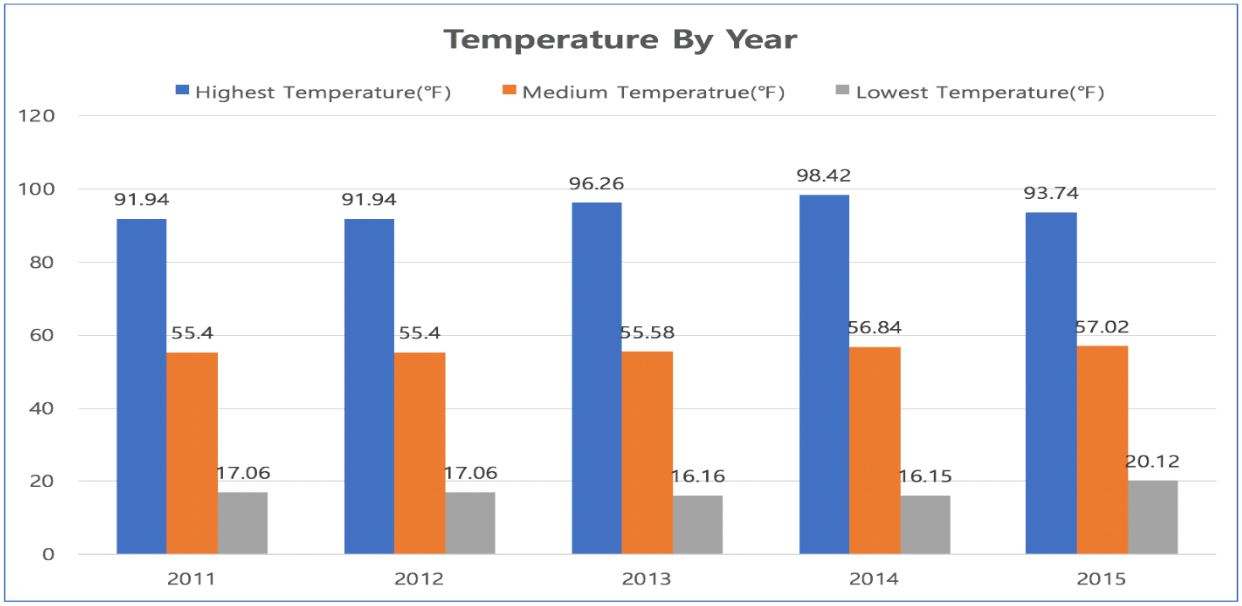


Figure 4.5: Temperature by year in JeonllaNamdo [38]

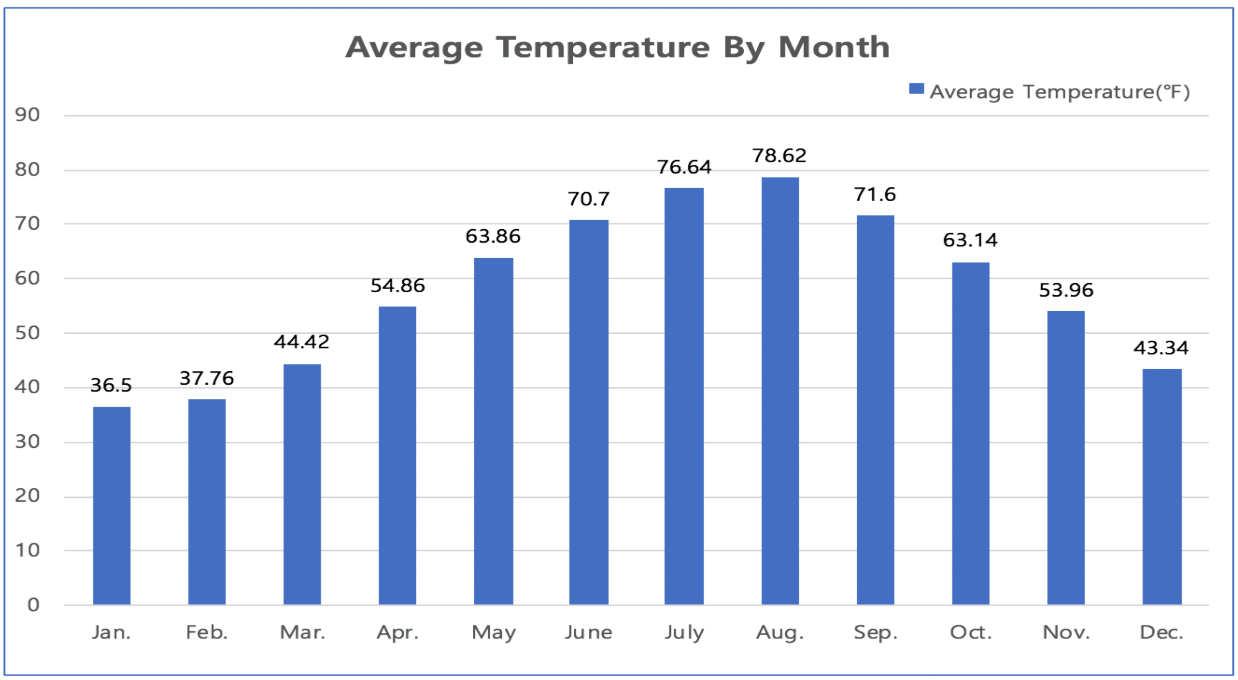


Figure 4.6: Average Temperature by month in JeonllaNamdo [38]

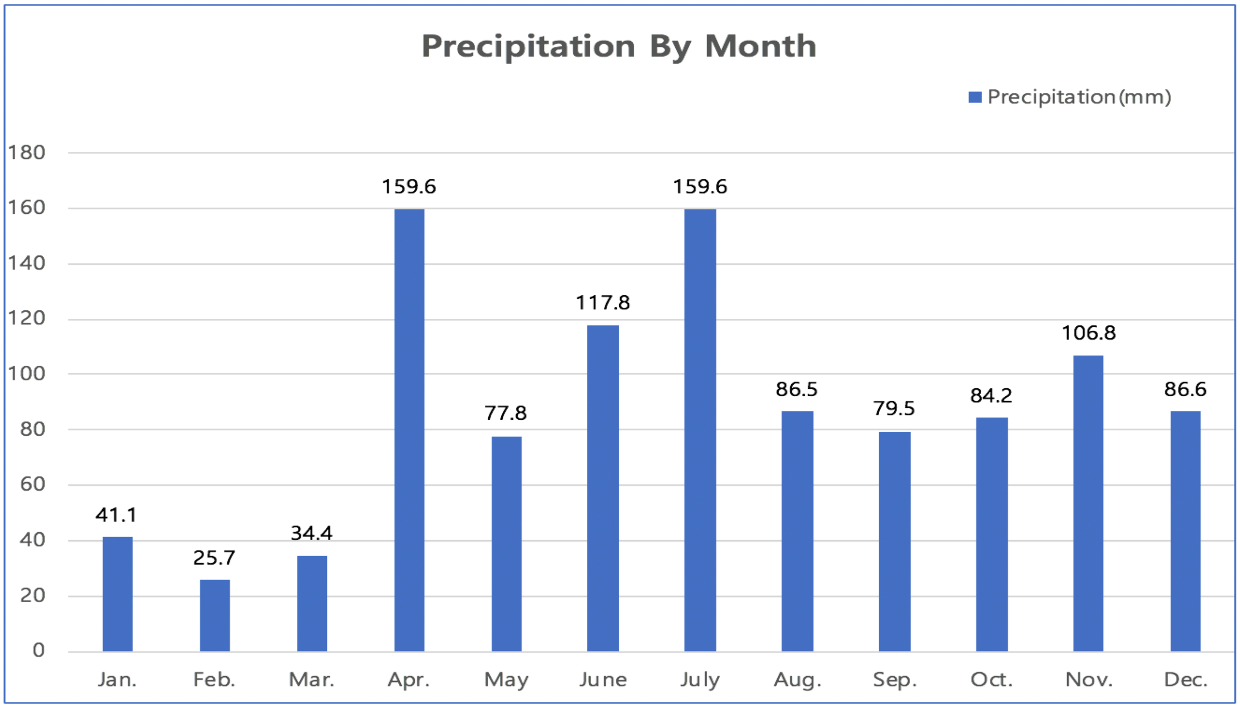


Figure 4.7: Precipitation by month in JeonllaNamdo [38]

The historical weather data is extracted from Korea Meteorological Administration (KMA). Branches of utility company in JeonllaNamdo are 24, and 13 branches coincide with locations of KMA (see Table 4). In 'consistency' section of table 4, 'Consistent' means a branch and a KMA station are located in the same city. There are 16 KMA stations in JeonllaNamdo [39] area, and the branches which don't have KMA stations are selected nearest to the existing KMA locations.

Table 4.4: Distribution Branches and KMA station locations [38, 39]

No.	Branches	KMA place	Consistency
1	Kangjin	Kangjin	Consistent
2	Koheoung	Koheoung	Consistent
3	Koksung	Sooncheon	
4	kwangsan	Kwangju	

Table 4.4: (continued) Distribution Branches and KMA station locations [38, 39]

No.	Branches	KMA place	Consistency
5	Kwangyang	Kwangyang	Consistent
6	Kwangju	Kwangju	Consistent
7	Naju	Kwangju	
8	Damyang	Kwangju	
9	Mokpo	Mokpo	Consistent
10	Mooan	Mokpo	
11	Bosung	Sooncheon	
12	Seokwangju	Kwangju	
13	Sooncheon	Sooncheon	Consistent
14	Sinan	Jindo	
15	Yeosoo	Yeosoo	Consistent
16	Youngkwang	Youngkwang	Consistent
17	Youngam	Kwangju	
18	Wando	Wando	Consistent
19	Jangsung	Kochang	
20	Jangheoung	Jangheoung	Consistent
21	Jindo	Jindo	Consistent
22	Hampeung	Youngkwang	
23	Haenam	Haenam	Consistent
24	Hwasoon	Kwangju	

KMA stations provide weather factors as follows: average temperature [˚F], the highest temperature [˚F], the lowest temperature [˚F], relative humidity [%], cloudiness, irradiation (MJ/m^2), sunshine [hour], average wind speed [m/s], maximum wind speed [m/s], wind gust [m/s], snow depth [cm], precipitation [mm], lightning [with/without], and evaporation [mm]. We selected average temperature [˚F], the highest temperature [˚F], relative humidity [%], maximum wind speed [m/s], wind gust [m/s], snow depth [cm], precipitation [mm], lightning [with/without] as weather variables. We did not choose cloudiness, irradiation [MJ/m^2], sunshine [hour], and evaporation [mm], the lowest temperature as parameters since these are not related to causes of DT failure based on papers that analyze causes of DT failure [22, 15, 16, 27]. Snow depth variable is selected, however, there was no date when there was snow among selected dates. Hence, snow depth is excluded from parameters of weather data.

The weather parameters that are taken into account are shown in Table 5.

Table 4.5: Parameters of weather data

Lightning [0/1] (LI)	Average Temperature [˚F] (AT)	Highest Temperature [˚F] (HT)	Relative Humidity [%] (RH)
Maximum Wind Speed [m/s] (MWS)		Wind Gust [m/s] (WG)	Precipitation [mm] (PT)

4.3 Economic evaluation of the failure impact

When failure occurs, the utility utilizes lots of time to restore service and labor of employees to repair equipment. Most of employees spend one or two days dealing with work orders about the customer's complains and repairing or replacing DT. In addition, the utility cannot sell electricity during the time of failure. It usually takes four hours to replace old DT and the loss of electricity causes loss of revenue from 20 to 150 customers connected to the transformer during this time. In the long term, customers may lose faith in supply of electricity, and there is a high chance for utility to lose the customers in the future. From the customer's point of view, they experience sudden outage, without any warning. If they run business directly attending a service for customers, they might not be able to provide such service to their customers and as a result they experience an economic loss. Most of customers who experience outage don't receive compensation and they consume the economic burden caused by the outage leaving them unhappy about the experience.

4.4 Conclusions

The region where DT failures occur is selected by criteria such as area that failures occur the most, the number of distribution facilities such as pole mounted DT, and length of distribution lines. In addition, the reason that this specific area is chosen is the number of historical weather events such as storm, rain, lighting etc. The historical weather data is extracted from Korea Meteorological Administration stations. They provide the weather data on their website without cost. The KMA stations closest to the locations where DT failures occurred are selected.

5. PREDICTION MODEL

The theoretical foundation, which it is the most important part of this research, is related to the prediction model. We study the reasons why we consider logistic regression as a prediction model for my research, and take a look at specific theory and mathematical foundations. Finally, it is demonstrated that logistic regression is the most appropriate method to represent correlation between weather parameters and DT failure.

5.1 Prediction Model

The goal of this research is to estimate the probability of DT failure caused by weather. For this purpose, logistic regression, a probabilistic nonlinear discriminant classifier, is considered as the method. Logistic regression is the fitting regression analysis to execute in the case that the dependent variable is binary. Logistic regression is useful for explanation of the relationship between one dependent binary variable and one or more independent predictors. In addition, the more familiar method of maximum likelihood is adopted, because it has better statistical characteristics. Maximum likelihood is a very broad method that is used to fit the non-linear models that we research. Finally, the condition that DT failure caused by weather takes binary values (i.e., failure/no failure) satisfies selection of logistic regression. Since there is a limit of modeling by a linear function due to complexity of the correlation between weather parameters and DT failure, logistic regression model is able to represent the correlation between DT failure and weather parameters through a non-linear function. In addition to such modeling advantage, logistic regression model can also inherit the advantage of classical linear regression in terms of coefficients interpretation. [40]

5.2 **Logistic Regression**

Logistic regression is used for modeling a binary response (i.e., failure/no failure) in many applications [41, 42]. This model estimates the probability of the response occurring P(X) = Pr (Y=1 | X) through a nonlinear function of explanatory variables X. In this study, it is natural that the response variable Y is a DT failure, i.e., 1 (failure) and 0 (no failure), and weather predictors like Lightning (LI), Average Temperature (AT), Highest Temperature (HT), Relative Humidity (RH), Maximum Wind Speed (MWS), Wind Gust (WG), and Precipitation (PT) are available for modeling logistic regression. Specifically, X is n × (p+1) design matrix where n is the number of observations and p is the number of weather predictors. Naturally, the number of coefficients is eight by seven predictors and an intercept. The corresponding coefficients β of predictors designate the effect of the weather predictors on the probability of DT failure in Eq (1).

$$\boldsymbol{\beta} = [\beta_0, \dots, \beta_7]^T \tag{1}$$

In Eq (1), β_0 is an intercept that when every weather variable is set to zero. $\beta_1 \sim \beta_7$ are the effects of the corresponding weather variables on the probability of DT failure: LT, AT, HT, RH, MWS, WG, and PT respectively.

The basic intuition behind using maximum likelihood to fit a logistic regression model is as following Eq (3): we seek estimates such that the predicted probability of failure for each individual DT is most likely to agree with its observed failure. This intuition can be formalized using the mathematical Equations (2) called a likelihood function.

$$\ell(\boldsymbol{\beta}) = \prod_{i:y_i=1} p(x_i) \quad \prod_{i':y_{i'}} (1 - p(x_{i'}\,) \tag{2}$$

$$\widehat{\boldsymbol{\beta}} = \max_{\beta} \ell(\beta) \tag{3}$$

The estimates $\widehat{\beta}_0, \dots, \widehat{\beta}_7$ are chosen to maximize this likelihood function.

We model $\hat{p}(\boldsymbol{x})$ using a function that gives outputs between 0 and 1 for all values of x. In logistic regression, we use the logistic function with discrete values.

Once the coefficients have been estimated by using Eq (3), the probability of failure is given by

$$\hat{p}(\boldsymbol{x}) = \frac{e^{x^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}}}}{1+e^{x^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}}}} \tag{4}$$

5.3 Conclusions

Logistic regression is used as a way to build a probabilistic classifier in this research. Each probability of DT failure is computed according to historical DT failure data and weather data. We use maximum likelihood function to fit a logistic regression model, and the coefficients of parameters of weather data are calculated by using logistic regression model.

6. RESULTS EVALUATION

Lastly, we discuss the experimental setup and model evaluation. In terms of evaluation metrics, AUC significance and coefficient analysis are used for analysis of the results. Every probability of DT failure event is calculated and we expect the results are analyzed appropriately by using evaluation metrics.

6.1 Data Processing

The causes of transformer failure are various such as aging, weather, animal contact, overloading, corrosion, out for maintenance, installment fault and people caused fault. The extracted failures for this study are the ones caused by weather. The date of failure is used to collect weather data during the failure occurrence. The weather data for the date with no outages is also collected because it is necessary for calculating the area under ROC curve. When selecting dates of no outage, not only outages caused by weather but also outages caused by other factors are excluded. The dates of no outage are selected with the same time interval AS used between the dates of outages. In addition, KMA data is chosen from the closest station from a branch of the utility company where DT failure occurs.

Every weather variable except for lighting is a continuous variable. Lightning variable is classified into 0 (without lighting) and 1 (with lightning).

6.2 Experimental Setup

All DT failures have their own failure number and date, and the failure data spans from 2011 to 2018. The weather data is correlated through the date and location of failure [39]. The historical DT failure data is divided into the training and testing sets. Training and testing sets are used for model validation. Out of entire data set some portion of the population for training is selected to build prediction model, then prediction model can be validated by getting the predictions on the remaining observations called testing set. We select the proportion 90:10 for training and testing sets. The total number of DT failure is 237, and 90% of the total is selected for the training set. The remaining 10% of the data is used for the testing sets for model estimates. There is total of 148 of no failure cases used.

The probability of the occurrence of 385 weather data sets is represented as a result of logistic regression. When a probability of DT failure is lower than 0.5, we show Y=0, and when a probability is above 0.5, we assign Y=1. The degree of high temperature (HT) is classified into two temperature thresholds such as 86°F and 89.6°F in order to make interpretation of HT coefficient precisely. The basic loading of distribution transformer for a normal life expectancy is continuous loading under a constant 30°C (86°F) ambient temperature as discussed in [43].

6.3 Model Evaluation and Effect of Predictors

To evaluate logistic regression, Receiver Operating Characteristics (ROC) [44] graphs are useful for organizing classifiers and visualizing their performance. The Area Under Curve (AUC) [45] is used, and the estimated coefficients quantify the effect of weather on the probability of failure. AUC is the most popular metric for visualizing the performance and analyzing the coefficient serves as intuitive interpretation of the effects of predictors.

6.3.1 AUC significance

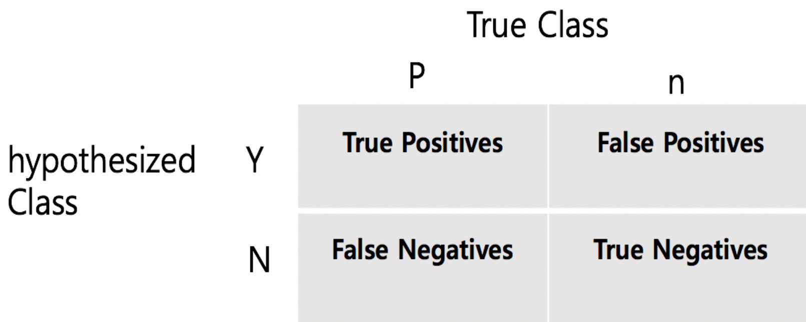


Figure 6.8: Confusion matrix and common performance metrics

To distinguish between the actual class and the predicted class, we use the labels Y, N for the class predictions produced by a model as shown in Fig 8 [23]. There are four possible cases. If the prediction is failure when real value is failure, it is true positive, and if the prediction is failure when real value is no failure, it is false negative. On the other hand, if the prediction is no failure when real value is no failure, it is true negative, and if the prediction is no failure when real value is failure, it is false positive.

The true positive rate of a classifier is estimated as

$$\text{tp rate} \approx \frac{Ture\ positive}{Total\ positives(ture\ positive + flase\ positive)}$$

The false positive (fp) rate of the classifier is estimated as

$$\text{fp rate} \approx \frac{False\ positive}{Total\ negatives(false\ negative + true\ negative)}$$

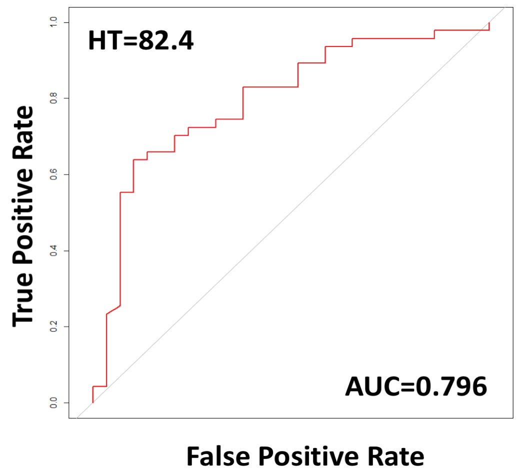


Figure 6.9: An example of ROC Curve

ROC (Receiver Operating Characteristics) graphs [45] are two-dimensional graphs in which tp rate is plotted on the Y axis fp rate is plotted on the X axis as shown in Fig 9. An ideal ROC curve will approach the top left corner, so the larger the AUC the better the performance. This is because if the false positive value shown on X-axis is smaller, and the true positive value on Y-axis is larger, the larger the Area Under Curve (AUC), which means that the accuracy of the model is high as shown in Fig 9.

6.3.2 Coefficient analysis

Coefficients show how the corresponding predictors have an impact on an outcome by describing the magnitude. Seven weather predictors are shown in X-axis and magnitude of the corresponding coefficient values are represented in Y-axis as shown in Fig 10. It shows which weather parameter among variables such as LI, AT, HT, RH, MWS and PT has the largest effect on DT failure.

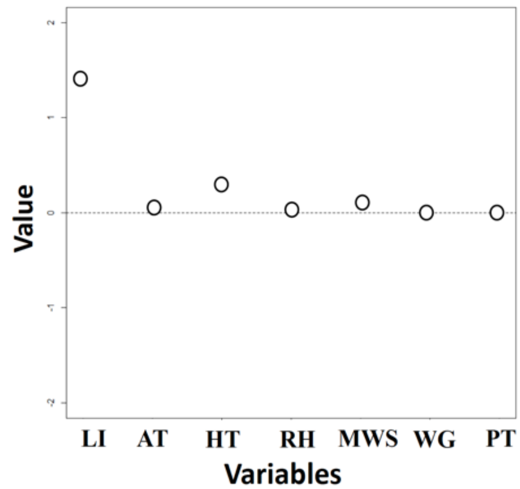


Figure 6.10: An example of coefficient value graph

6.4 Results

The partial results from predicting probability of DT failure can be represented as given in Tables 6, 7, and 8. The prediction results by logistic regression are summarized in the Table 9.

Table 6.6: Probability of DT failure (HT: 86°F or below)

LI	AT	HT	RH	MWS	WG	PT	Failure (Y=1/0)	Logistic Prob.	Logistic Pred.
1	71.96	low	192	4.7	13.7	180	0	0.961671	1
1	70.7	low	138	7.1	12.8	17.2	0	0.874356	1
1	34.34	low	163	12.4	18.8	0.2	0	0.936093	1
1	78.98	low	222	5.8	10.9	15.9	0	0.941851	1
0	40.1	low	152	14.2	21.7	1	0	0.743072	1
0	68.36	low	135	7.3	13.8	22.3	0	0.550675	1
0	48.38	low	121	6.1	8.7	0	0	0.425061	0
1	81.14	high	128	5.7	9.3	13	0	0.867763	1
0	73.04	high	107	4	7	2.5	0	0.396732	0
0	76.28	high	145	4.5	8.1	50.5	0	0.595443	1

Table 6.7: Probability of DT failure (HT: 86°F - 89.6°F)

LI	AT	HT	RH	MWS	WG	PT	Failure (Y=1/0)	Logistic Prob.	Logistic Pred.
1	75.92	low	168	4.2	5.7	0	0	0.828777	1
1	53.24	low	149	6.1	8.5	0	0	0.79653	1
0	31.46	low	154	4.5	13.7	1.3	0	0.405346	0
1	74.66	low	219	9	12.4	12.4	0	0.922829	1
1	81.14	high	128	5.7	9.3	13	0	0.861096	1
0	75.38	low	127	4.3	6.9	0	00	0.446147	0
1	71.96	low	192	4.7	13.7	180	0	0.982402	1
1	77.54	low	209	7.7	10.4	52.6	0	0.94545	1
1	67.28	low	220	11.9	24	92	0	0.972105	1
1	72.5	low	113	11.5	16.9	45	0	0.909012	1

Table 6.8: Probability of DT failure (HT: 89.6°F or above)

LI	AT	HT	RH	MWS	WG	PT	Failure (Y=1/0)	Logistic Prob.	Logistic Pred.
0	54.86	low	115	13.3	17.4	48.1	0	0.761106	1
1	70.7	low	138	7.1	12.8	17.2	0	0.854838	1
1	77.36	low	216	14.5	20.8	57.2	0	0.979367	1
1	60.98	low	195	6.2	10.7	12.6	0	0.868762	1
0	56.66	low	212	5.8	21.6	21.6	0	0.667786	1
1	63.14	low	162	5.2	0	0	0	0.782837	1
0	68.36	low	135	7.3	22.4	22.4	0	0.621972	1
1	71.06	low	66	8.4	0.1	0.1	0	0.724282	1
0	67.64	low	170	4.6	10	10	0	0.537834	1
1	76.1	low	203	5.9	86.1	86.1	0	0.970742	1

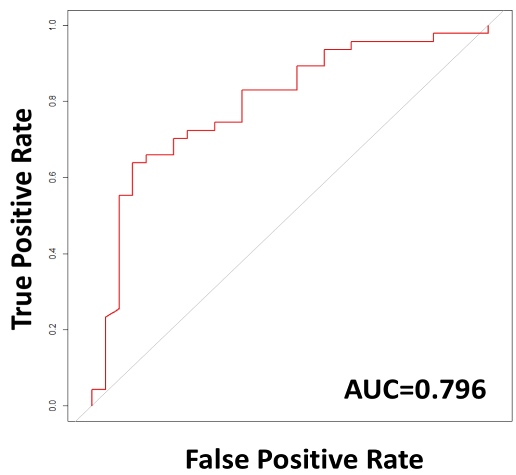
The prediction of probability according to an event is presented in Table 9.

Table 6.9: Event Vs. Prediction of failure

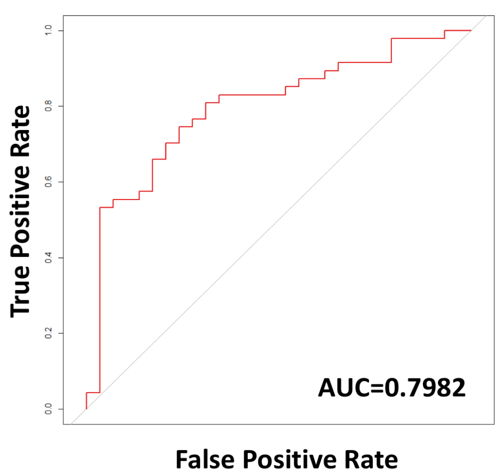
Degree of HT	Failure(Y/N)	Prediction	
		Y=0	Y=1
86°F or below	Y=0	113	47
	Y=1	35	190
86°F - 89.6°F	Y=0	112	47
	Y=1	36	190
89.6°F or above	Y=0	111	54
	Y=1	37	183

For the case of 86°F or below, 113 cases are predicted as no failure (i.e., Y=0), and 47 cases show prediction that there will be a failure (i.e., Y=1). For the cases of 86°F-89.6°F, and 89.6°F or above, 112 events and 111 cases are predicted as no failure respectively. On the other hand, 190 cases in 86°F-89.6°F and 183 cases in 89.6°F or above have prediction of failure (see Table 9). The accuracy of prediction the condition of 86°F or below is 0.789 by calculating (113+190)/384 in. In the cases of 86°F - 89.6°F, and 89.6°F or above, the accuracy of prediction is 0.786 ((112+190)/384) and 0.766 ((111+183)/384) respectfully. For 86°F or below, the accuracy of prediction of probability is the highest.

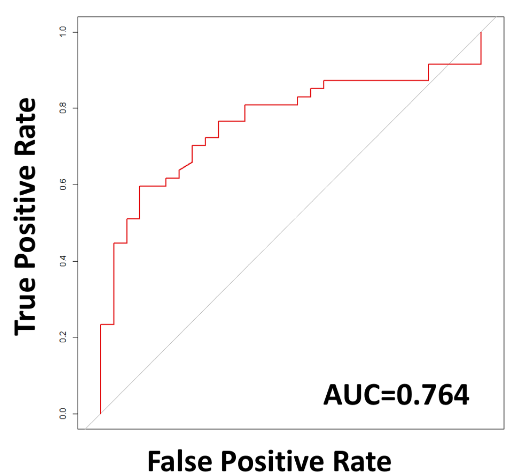
The AUC is 0.796, 0.798 and 0.764 in HT 86°F or below, 86°F - 89.6°F, and 89.6°F or above respectively as shown in Fig 10, which means the result can be considered as very good. The AUC of HT 86°F and below have the highest values (see in Fig 11, a)). HT is divided by three groups (0/1) according to the temperature. We assume that probability of DT failure would increase by HT. Three section of HTs such as 86°F or below, 86°F-89.6°F, and 89.6°F or above are selected. For example, for a certain temperature below 86°F, the probability is 0 and if the temperature is above 86°F, the value is 1.



a) HT: 86°F or below



b) HT: 86°F - 89.6°F



c) HT: 89.6°F or above

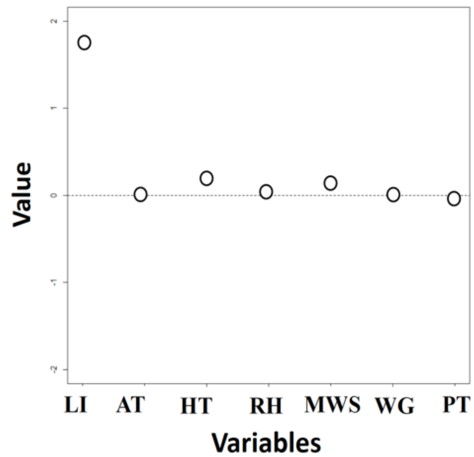
Figure 6.11: ROC for classification by logistic regression

Table 6.10: Coefficient values

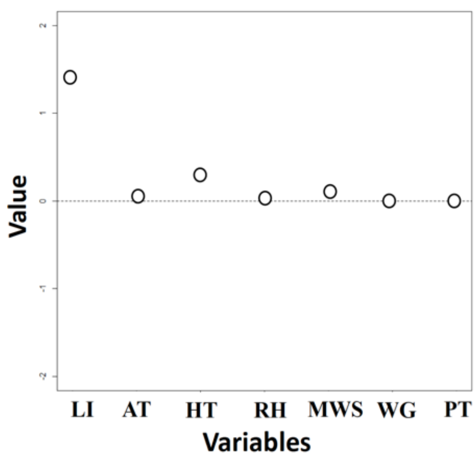
Degree of HT	LI	AT	HT
86°F or below	1.7455	0.003565	0.21771
86°F - 89.6°F	1.4329	0.007634	0.30125
89.6°F or above	1.3649	0.010839	0.78165

RH	MWS	PT
0.01176	0.13367	0.005994
0.008571	0.0823	0.01289
0.007447	0.060257	0.01922

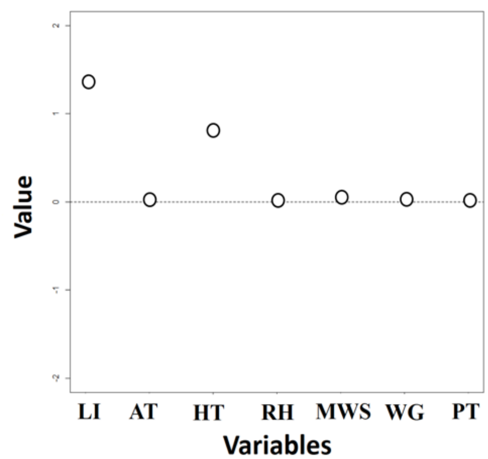
The positive coefficient of the predictor indicates that the predictor increases the probability of DT failure, while the negative coefficient makes the probability of DT failure decrease. That is, predictor with the positive coefficient is the most relevant contributor to the failure, and predictor with the negative coefficient is the least relevant contributor to the failure. The positive coefficient is likely to have an effect on failure, and on the other hand, negative coefficient is less likely to have an influence on failure. The coefficients for DT failure as shown in Table 10 and in Fig 12, which are LI, AT, HT, RH, MWS and PT, have all positive values. Lightning is the most influential coefficient, since it is the biggest one and high temperature is the second.



a) HT: 86°F or below



b) HT: 86°F - 98.6°F



c) HT: 89.6°F or above

Figure 6.12: Coefficients value estimate

We realize that lightning and higher temperature, especially 89.6°F or above have the biggest effect on DT failure as shown Fig 12 c). For the coefficient value of HT, there is no huge differences between 86°F or below and 86 °F - 89.6°F from 0.21771 to 0.30125 in Table 10. However, the value increases for 89.6°F or above as 0.78165. The Average Temperature has low correlation but it is positive. In the case Average Temperature is positive, HT coefficient turns into the important factor, which is in turn associated with higher probability of DT failure. On the other hand, Relative Humidity, Maximum Wind Speed, and Precipitation have a low positive correlation.

6.5 Conclusion

The logistic regression model is used to be calculated the probability for different types of weather caused failures. In addition, the prediction model shows relatively high-level accuracy, where the average AUC is 0.78. Every parameter has positive value, and Lightning and HT is the most important factor that affects the DT failure. HT has more effect on failure when it has high temperature such as 89.6°F or above.

7. RESEARCH SUMMARY AND FUTURE WORK

In this last section, research summary is discussed. We summarize the discovery associated with results, and then we look further what will need to be studied more in the future research.

7.1 **Research summary**

The Thesis describes the use of a logistic regression prediction model for estimation of probability of distribution transformers failure by using correlation of weather parameters. The main findings of our study are:

- The variety of weather predictors causing DT failure are identified as being Lightning [0/1], Average Temperature [°F], Highest Temperature [°F], Relative Humidity [%], Maximum Wind Speed [m/s], Wind Gust [m/s], and Precipitation [mm].
- The prediction model shows high-level accuracy, where the average AUC is 0.78. Each AUC is 0.796 in case of HT 86°F or below, 0.798 in condition of HT between 86°F - 89.6°F, and 0.764 in HT 89.6°F or above.
- The coefficient values estimate show that all weather predictors, LI, AT, HT, RH, MWS and PT, with three HT conditions have positive value indicating all predictors have positive effect on failure. Lightning and HT being the most important factor which affect the DT failure. HT has more effect on failure when it reaches high temperature such as 89.6°F or above.
- The hypothesis which the probability of failure can increase with temperature by 86°F or 89.6°F is proven through categorizing the degree of high temperature, and this work quantifies the risk of DT failure by coefficient analysis. There is a substantial increase in risk.

• This work provides more reliable model than existing physical-based model that are not considered in physical-based model.

➢ Utility Benefits

• The approach is a fundamental step to predicting not only DT failure but also other outages of other components in the power network.

• The logistic regression can be applied for other variables such as Health Index, historical maintenance work as well as weather variable. Therefore, the prediction of failure can be more accurate, and it covers every cause of DT failure by addressing comprehensive causes.

• The risk analysis of each DT enables when all DT data of distribution grids are used as well as DT failure data. Through this work, utility company is able to utilize proactive maintenance in preparation for lighting and heat wave as a strategy of choice and concentration.

• Utilities can create more precise criteria by defining future weather data as one of the maintenance criteria. By using the maintenance criteria, DT failure rate can be declined, and losses of utility company and customers will decrease.

7.2 Future work

The parameters of weather data are correlated when considering impact on DT failures. For example, Average Temperature is correlated to High Temperature and there is correlation between Maximum Wind Speed and Wind Gust and Relative Humidity and Precipitation respectively. In order to verify more detailed correlation between variables, further correlation study is needed. For analysis of correlation between variables, the least absolute shrinkage and selection operator (LASSO) logistic regression [46, 47] can be used. In statistics and machine learning, the LASSO is a penalized regression method that performs variable selection in order to enhance the prediction accuracy and interpretability of the statistical model it produces. LASSO was originally formulated for least squares models. This simple case reveals a substantial insight into the behavior of the estimator, including its relationship to other method such as a ridge regression and best subset selection and the connections between LASSO coefficient estimates and so-called soft thresholding. It also reveals that (like standard linear regression) the coefficient estimates need not be unique if covariates are collinear.

7.3 Conclusion

The hypothesis that the probability of failure can increase with high temperature, and a range of temperature can be specified is proven through categorizing the degree of high temperature. This research quantifies the risk of DT failure over weather variables, and provides more reliable model than existing physical-based model by considering various variables. My proactive maintenance, DT failure can be reduced, and losses of utility and customers will decrease.

REFERENCES

[1] Singh, Ranjana, A. S. Zadgaonkar, and Amarjit Singh. "Premature failure of distribution transformer-a case study." International Journal of Scientific &Engineering Research 5.6 (2014).

[2] Suwnansri, Thanapong. "Asset management of power transformer: Optimization of operation and maintenance costs." 2014 International Electrical Engineering Congress (iEECON). IEEE, 2014.

[3] Jahromi, A., et al. "An approach to power transformer asset management using health index." IEEE Electrical Insulation Magazine 25.2 (2009): 20-34.

[4] Treinish, L., et al. "On-going utilization and evaluation of a coupled weather and outage prediction service for electric distribution operations." Proc. 4th Conf. Weather Climate New Energy Economy. 2011.

[5] Li, Gengfeng, et al. "Risk analysis for distribution systems in the northeast US under wind storms." IEEE Transactions on Power Systems 29.2 (2014): 889-898.

[6] EPSIS, Korea Electric Power Statistics Information system, https://epsis.kpx.or.kr/

[7] Boles, James S., and Barry J. Babin. "On the front lines: Stress, conflict, and the customer service provider." Journal of Business Research 37.1 (1996): 41-50.

[8] Rose, Adam, Gbadebo Oladosu, and Shu-Yi Liao. "Business interruption impacts of a terrorist attack on the electric power system of Los Angeles: customer resilience to a total blackout." Risk Analysis: An International Journal 27.3 (2007): 513-531.

[9] Wacker, Garry, and Roy Billinton. "Customer cost of electric service interruptions." Proceedings of the IEEE 77.6 (1989): 919-930.

[10] El-Bassiouny, A., M. El-Shimy, and R. Hammouda. "Impact of power transformer failures on customer interruptions costs using customer damage functions." 2017 Nineteenth International Middle East Power Systems Conference (MEPCON). IEEE, 2017.

[11] Zhang, Xiang, and Ernst Gockenbach. "Asset-management of transformers based on condition monitoring and standard diagnosis." IEEE Electrical Insulation Magazine 24.4 (2008): 26-40.

[12] Eaton, "Blackout Tracker United States Annual Report 2013, "Eaton, 2018. [Online] Available:file:///Users/eunhuiko/Downloads/BROCHURE-Blackout-Tracker-2018-MZ153043EN-LR-v6.pdf

[13] Abi-Samra, N. C., K. R. Forsten, and R. Entriken. "Sample Effects of extreme weather on power systems and components, part I: Sample effects on distribution systems." IEEE PES General Meeting. IEEE, 2010.

[14] Rexhepi, Vezir. "An analysis of power transformer outages and reliability monitoring." Energy Procedia 141 (2017): 418-422.

[15] Jan, Shayan Tariq, Raheel Afzal, and Akif Zia Khan. "Transformer failures, causes & impact." International conference data mining, civil and mechanical engineering. 2015

[16] Ndungu, C., et al. "A Study of the root causes of high failure rate of distribution transformer-a case study." International Journal Engineering and Science 2 (2017): 14-18.

[17] Kumbhar, Ganesh B., and S. V. Kulkarni. "Analysis of short-circuit performance of split-winding transformer using coupled field-circuit approach." IEEE Transactions on power delivery 22.2 (2007): 936-943.

[18] Paillat, T., et al. "Flow electrification in transformers: relevance of ECT and leakage current measurements." IEEE transactions on dielectrics and electrical insulation 10.1 (2003): 7-14.

[19] C.Rajotte et al., Guide for Transformer Maintenance, CIGRE, 2011

[20] Muhamad, N. A., et al. "Comparative study and analysis of DGA methods for transformer mineral oil." 2007 IEEE Lausanne Power Tech. IEEE, 2007.

[21] Udayakanthi, Munasinghe Vidana Pathirana Geetha. "Identification of causes of distribution transformer failures and introduction of measures to minimize failures." (2015).

[22] Jahromi, A., et al. "An approach to power transformer asset management using health index." IEEE Electrical Insulation Magazine 25.2 (2009): 20-34.

[23] Haema, Juthathip, and Rattanakorn Phadungthin. "Condition assessment of the health index for power transformer." 2012 Power Engineering and Automation Conference. IEEE, 2012.

[24] Utami, N. Y., et al. "Evaluation condition of transformer based on infrared thermography results." 2009 IEEE 9th International Conference on the Properties and Applications of Dielectric Materials. IEEE, 2009.

[25] Yanming, Tu, and Qian Zheng. "DGA based insulation diagnosis of power transformer via ANN." Proceedings of the 6th International Conference on Properties and Applications of Dielectric Materials (Cat. No. 00CH36347). Vol. 1. IEEE, 2000.

[26] Cheim, Luiz, et al. "Furan analysis for liquid power transformers." IEEE Electrical Insulation Magazine 28.2 (2012): 8-21.

[27] Marín, OJ Soto, et al. "Causes of failure of Distribution Transformers in the East Zone of Caldas."

[28] Tippachon, W., et al. "Failure Analysis of Power Distribution Systems in Thailand." 2006, International Conference on Power System Technology. IEEE, 2006.

[29] T&D world, Predicting Distribution Transformer Failures, https://www.tdworld.com/asset-managementservice/predicting-distribution-transformer-failures

[30] Piantini, Alexandre, et al. "Lightning-caused transformer failures in distribution systems." 2014 International Conference on Lightning Protection (ICLP). IEEE, 2014.

[31] Al-Arainy, A. A., N. H. Malik, and M. I. Qureshi. "A study of failure of pole mounted distribution transformers." 2012 International Conference on High Voltage Engineering and Application. IEEE, 2012.

[32] Puri, J. L., et al. "Lightning induced failures in distributed transformers." IEEE Transactions on Power Delivery 3.4 (1988): 1784-1801.

[33] Pandit, Narasimha, and R. L. Chakrasali. "Distribution transformer failure in India root causes and remedies." 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). IEEE, 2017.

[34] Gupta, Manju, Sushma Gupta, and Tripta Thakur. "Design and impact of wireless AMI for distribution utilities." 2013 Annual IEEE India Conference (INDICON). IEEE, 2013.

[35] Osorio, Christian, and Nandan Sawant. "Transformer Lifetime Prediction." Intelligent Energy Projects, ee292k edition: 5.

[36] De Carvalho, T. O., et al. "Analysis of lightning-caused distribution transformer failures." 2011 International Symposium on Lightning Protection. IEEE, 2011

[37] ArcGIS, Esri.2017 [Online] Available: http://www.arcgis.com

[38] JeollaNamdo, JeollaNam Provincial Government, http://www.jeonnam.go.kr/contentsView.do?menuId=jeonnam0600000000

[39] Korea Meteorological Administration, KMA, http://www. kma.go.kr/eng/index.jsp

[40] K. Murphy, "Logistic regression," Machine Learning: A probabilistic Perspective, Chapter 8, pp. 245-279, 2012

[41] Adwere-Boamah, Joseph, and Shirley Hufstedler. "Predicting Social Trust with Binary Logistic Regression." Research in Higher Education Journal 27 (2015).

[42] Soule, Patrick. "Predicting Student Success: A Logistic Regression Analysis of Data From Multiple SIU-C Courses." (2017).

[43] Biçen, Yunus, et al. "An assessment on aging model of IEEE/IEC standards for natural. and mineral oil-immersed transformer." 2011 IEEE International Conference on Dielectric Liquids. IEEE, 2011.

[44] Fawcett, Tom. "An introduction to ROC. analysis." Pattern recognition letters 27.8 (2006): 861-874.

[45] Gareth James, Daniela Witten, Trevor Hastie Robert Tibshiran, “An Introduction to Statistical Learning”, 2013

[46] Austin, P., et al. "Regression Shrinkage and Selection via the Lasso." (2015): 586-593.

[47] Tibshirani, Robert. "Regression shrinkage and selection via the lasso." Journal of the Royal Statistical Society: Series B (Methodological) 58.1 (1996): 267-288.