Article

Fire Risk Assessment Models Using Statistical Machine Learning and Optimized Risk Indexing

Myoung-Young Choi [1] and Sunghae Jun [2],*

[1] Risk Management Center, Korean Fire Protection Association, Seoul 07328, Korea; mychoi@kfpa.or.kr
[2] Department of Big Data and Statistics, Cheongju University, Chungbuk 28503, Korea
* Correspondence: shjun@cju.ac.kr; Tel.: +82-10-7745-5677

check for updates

Abstract: It is very difficult for us to accurately predict occurrence of a fire. But, this is very important to protect human life and property. So, we study fire hazard prediction and evaluation methods to cope with fire risks. In this paper, we propose three models based on statistical machine learning and optimized risk indexing for fire risk assessment. We build logistic regression, deep neural networks (DNN) and fire risk indexing models, and verify performances between proposed and traditional models using real investigated data related to fire occurrence in Korea. In general, fire prediction models currently in use do not provide satisfactory levels of accuracy. The reason for this result is that the factors affecting fire occurrence are very diverse and frequency of fire occurrence is very sparse. To improve accuracy of fire occurrence, we first build logistic regression and DNN models. In addition, we construct a fire risk indexing model for a more improved model of fire prediction. To illustrate comparison results between our research models and current fire prediction model, we use real fire data investigated in Korea between 2011 to 2017. From the experimental results of this paper, we can confirm that accuracy of prediction by the proposed method is superior to the existing fire occurrence prediction model. Therefore, we expect the proposed model to contribute to evaluating the possibility of fire risk in buildings and factories in the field of fire insurance and to calculate the fire insurance premium.

Keywords: fire risk assessment and prediction; logistic regression analysis; deep neural networks; optimized fire risk indexing

1. Introduction

Fires have been devastating to human life and property. So, humans have been making various efforts to deal with the fires. One of them is to predict the possibility of fire. However, it is actually very difficult to predict the fire occurrence [1]. This is because there are many variables that affect fire, and the number of fire occurrences in the total data is very small. That is, the fire occurrence data set is very sparse because most values of the data set are zeros. To solve this problem of fire data set, we study novel models to predict fire occurrence in this paper. The fire risk evaluation is a popular approach for fire prediction [1]. Rishickesh, et al. (2019) studied a model to predict the forest fires using various machine learning algorithms such as logistic regression, support vector machines, random forest, boosting, etc. In addition, they showed that the experimental results from logistic regression and gradient boosting are better than other machine learning algorithms. In their experimental results, they showed the accuracy results of 0.6826 and 0.6838 as better prediction results than others. The first accuracy was the result of logistic regression with principal component analysis (PCA), and the second accuracy was the result of gradient boosting without PCA. The other machine learning algorithms except these two algorithms had lower accuracy results than logistic regression with PCA and gradient boosting without PCA. In general, the prediction accuracy of 0.6826 or 0.6838

is not satisfactory. So, we need to improve the performance of prediction models in fire occurrence. We consider the logistic regression as one of our candidate models for fire risk prediction. In our research, we apply the model to not forest but factory and building and try to increase the accuracy of fire prediction using new proposed methods. There are a number of approaches to fire risk evaluation including fire risk indexing for buildings, factories, forests, etc. [2–5]. The fire risk index is an index made up of variables that describe the fire hazards and prevention for evaluating fire hazards. Madaio, et al. (2016) developed the 'Firebird' framework for predicting fire risk and prioritizing fire inspections. The authors also used support vector machines and random forest for fire risk prediction, and built an interactive map for prioritization of fire inspections. From the experimental results of 'Firebird' system, we found that the performance of random forest is better than the support vector machines. In both studies conducted by Rishickesh et al. (2019) and Madaio et al. (2016), we can see that the accuracy of the fire prediction models was not satisfactorily high. So, we can confirm it is difficult to predict fire occurrence accurately. Watts (2016) introduced the fire risk indexing as another new method for predicting fire risk. In his research, the Fire risk indexing is a heuristic model based on the knowledge and experience of fire experts for fire safety [4]. The fire risk index consists of the factors (variables) representing the influences of fire risk. This index contributes to the quantification of fire risk. Therefore, we use the fire risk indexing to model the hazard and prevention of fire in our paper. Sakennaite and Vaidogas (2010) compared fire risk index with fire risk analysis. The authors assessed the fire safety by means of fire risk indexing [5]. They used various variables related to geometry and fire-specific data for building fire risk indexing and analysis [5]. Nikolopoulos et al. (2018) illustrated the model performance for prediction of post-fire debris flow occurrence. Using a contingence table, they evaluated the performances of three approaches which are rainfall thresholds, logistic regression and random forest [6]. They found random forest model has the best performance in the predictive models [6].

From previous research results related to fire occurrence prediction, we confirmed that the analytical methods that provide best prediction performance differ according to the detailed prediction fields related to fire occurrence. In our paper, we also propose a novel fire risk assessment model using fire risk indexing and analysis. We apply statistical machine learning and optimized risk indexing models for fire risk assessment. The Korea Fire Protection Association (KFPA) conducts fire safety inspections and is sponsored by fire insurers in Korea [7]. KFPA uses fire risk indexing method to evaluate fire risk of buildings known as KFPA Fire Risk Index (KFRI). In order to improve KFRI, KFPA has developed a lot of models including statistical machine learning and optimization of fire risk indexing. Therefore, we propose novel models to improve the performance of the KFRI for fire risk assessment and prediction. We organize this paper as follows. In Section 2, the statistical machine learning for fire risk prediction is introduced. We propose our statistical machine learning and optimized risk indexing models for fire risk assessment in Section 3. In Section 4, we illustrate the performance of our proposed models for fire prediction and evaluation using the real investigated data related to fire risk from the KFPA. Lastly, we show our conclusions and future works in Section 5.

2. Statistical Machine Learning

Statistics is defined as learning from data [8]. Machine learning is to make machines (computers) intelligent by learning from data [9]. So, statistical machine learning is to apply statistics to machine learning. In general, statistics uses the concept of inference with estimation and hypotheses testing [10]. In addition, statistics has a normality assumption for data [11]. Therefore, statistical machine learning leads to the improvement of the performance of existing machine learning by using inference and normality assumptions of statistics [12]. Regression is a representative method of statistics [13]. Also, deep learning is a popular algorithm of machine learning [14]. In this paper, we use logistic regression from the regression and deep neural networks from deep learning for statistical machine learning. We can consider so many methods for statistical machine learning. Most of them are for classification, prediction and clustering. The aim of this paper is to study, using a predictive model,

fire risk assessment. Therefore, we use statistical machine learning models for fire risk forecasting and assessment.

3. Fire Risk Assessment Models Using Statistical Machine Learning and Optimized Risk Indexing

In this paper, we construct a fire risk model using statistical machine learning and optimized risk indexing. The data related to fire risk consist of explanatory variables (X) affecting the occurrence of fire and response variable (Y), indicating the frequency of fire occurrence.

3.1. Statistical Machine Learning for Fire Occurrence Prediction

Since the response variable contains frequency data, we use count data analysis methods among statistical machine learning models [15]. In the discrete probability distributions, Poisson and binomial distributions can be applied to analyze the fire occurrence data. In the proposed study, we focused on building a model for the possibility of fire, so we built a predictive model based on the binomial distribution. A random variable Y is distributed to a binomial distribution with n and p when Y is represented as follows [16].

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y} \quad y = 0, 1, \ldots, n \tag{1}$$

where n is the number of Bernoulli trials, and p is the probability of success. The expectation (E(Y)) and variance (Var(Y)) of Y are np and np(1-p) respectively. Each y has a binary data value (1: occurred fire or 0: no occurred fire) representing whether a fire has occurred. So, we build a logistic regression model based on binary response variable (Y) to forecast fire risk. The logistic regression model is defined as the following model [17]:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{2}$$

where p is P(Y = 1), which is the probability of occurred fire, and $\log\left(\frac{p}{1-p}\right)$ is logit function of p. $(x_1, x_2, \ldots, x_k)$ are the explanatory variables that affect Y. In addition, error term ε is followed to normal distribution with mean=0 and variance $= \sigma^2$. From the logistic model of equation (2), we get the following model for predicting the probability of occurred fire (p):

$$p = \frac{1}{1 + \exp(-(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k))} \tag{3}$$

where b_i represents the estimate for regression parameter β_i under minimizing ε. We consider another method of statistical machine learning.

We apply deep neural networks (DNN) to fire risk assessment. $(y, x_1, x_2, \ldots, x_k)$ is the data set for DNN [18]. Like the logistic regression model, y is a response variable representing fire occurrence or not. Also, $(x_1, x_2, \ldots, x_k)$ are explanatory variables affecting Y. Our DNN consists of input, hidden and output layers. Input and hidden layers are related to X and Y respectively. We design the size and structure of hidden layers to improve model performance. Figure 1 shows the proposed DNN model for fire risk prediction.
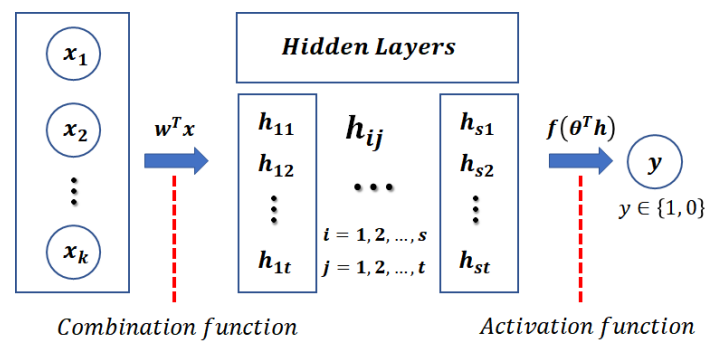
Figure 1. Proposed DNN model for fire risk prediction.

In Figure 1, w is connecting weight vector from input to hidden layers, and $w^T x$ is linear combination x with w. Also, we call $w^T x$ combination function. $\theta^T h$ is also a linear combination h with θ, where θ is a vector of connecting weights from hidden to output layer. $f(\cdot)$ is a function to transform $\theta^T h$ into the values between 0 to 1. This function is called as activation function, and we use logistic growth function as follows [9,12]:

$$f\left(\theta^T h\right) = \frac{1}{1 + e^{\theta^T h}} \tag{4}$$

We train the DNN model based on perceptron cost function that minimizes the difference between predicted value $\hat{y}$ and real value y.

3.2. New Fire Risk Indexing for Fire Risk Assessment

Lastly, we carry out an optimized risk indexing. In this paper, we use various variables (components) related to fire occurrence. Table 1 illustrates each component, module with components, and category with modules for the fire risk modeling.

Some components such as years, floors, size, and distance from the fire brigade are generated objectively but most other components are made rather subjectively. Therefore, we perform the proposed methods for fire risk assessment using the variables in Table 1. We illustrate our proposed procedure for fire risk assessment in Figure 2.
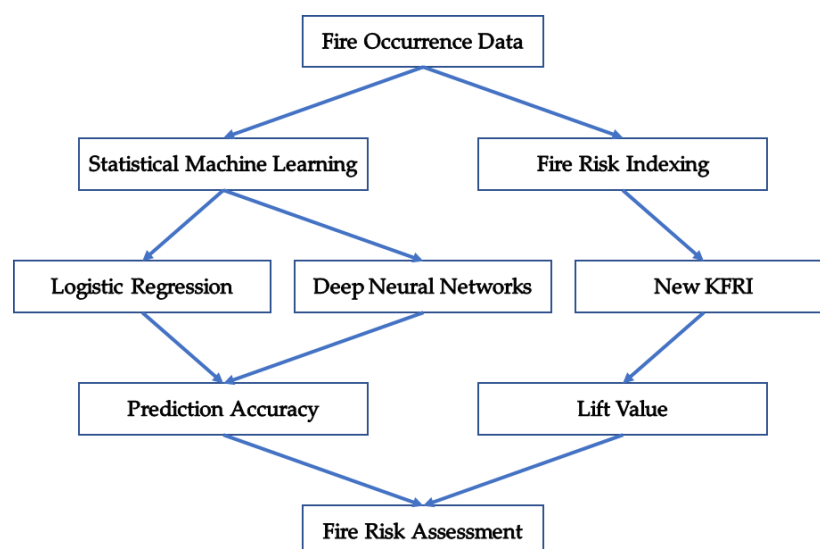

Figure 2. Proposed procedure for fire risk assessment.

To assess fire risk for each building and factory, we use fire occurrence data provided by the KFPA. This data set has a large number of variables related to fire occurrence. In this paper, we consider two approaches to fire risk assessment. First, we use statistical machine learning and DNN for fire occurrence prediction. Next, we build a risk index of fire occurrence, we call this index New KFRI (NKFRI). We use prediction accuracy and lift value as performance measures for statistical machine learning and fire risk indexing respectively. Using the results, we carry out fire risk assessment.

Table 1. Components, modules and categories of KFRI for manufacturing facilities.

Category	Module	Components
Hazards	Basic (B) $B = y \times e \times s \times g \times c \times d \times r \times q$	Years (y)
		Floors (e)
		Structure (s)
		Size (g)
		Fire load (q)
	Ignition (I) $I = i1 \times i2 \times i3 \times i4$	Fire facility (i1)
		Gas facility (i2)
		Hazardous material facility (i3)
		Electricity facility (i4)
	Process (P) $P = p0 \times p1 \times p2 \times p3 \times p4 \times p5 \times p6 \times p7$	Basic process (p0)
		Hazardous material (p1)
		Hot work (p2)
		Flammable gas (p3)
		High temperature or pressure (p4)
		Static electricity (p5)
		Dust (p6)
		High voltage (p7)
Counter-measures	Safety management (M)	Safety management (M)
	Firefighting equipment or system(F) $F = F1 \times F2 \times F3 \times F4 \times F5$	Fire extinguisher (F1)
		Standpipe (F2)
		Hydrant (F3)
		Sprinkler (F4)
		Total flooding gas system (F5)
	Fire alarm system (A) $A = A1 \times A2 \times A3$	Fire detection system (A1)
		Emergency alarm system (A2)
		Emergency notifying system (A3)
	Passive fire protection system (V) $V = V1 \times V2$	Fire compartment (V1)
		Evacuation system (V2)
	Smoke control system and auxiliary equipment required for fire brigade (S) $S = S1 \times S2$	Smoke control system (S1)
		Auxiliary equipment required for fire brigade (S2)

4. Experimental Results

To illustrate the validity of our research, we made experiments using real investigated data related to fire in Korea. The data set consists of 201,082 observations (buildings or factories) and 107 variables that cause and prevent fire [7] as follows: Identification number, Number of fires, Sum of property loss, Average of property loss, Sum of casualties, Average of casualties, Sum of damaged area,

Average of damaged area, Spread of fire, Inspection year, Inspection period, Property number, Name of property, Address1, Address2, Address3, Address4, Address5, Address6, Fiscal year of the report, Inspection date, Sort1, Sort2, Sort3, Estimated property value, Estimated value of machine, Estimated total value, Important level, KFRI 1, KFRI(before discount rate), Maximum Possible Loss, Estimated Maximum Loss, Year of construction 1, Building structure class, Number of buildings, Number of floor, Number of underground floor, Total area of buildings, Number of total floor, Building structure, Building size, Multiple use 1, Multiple use 2, Movement of occupants 1, Movement of occupants 2, Accommodation 1, Accommodation 2, Fire load, Fire facility, Gas facility, Hazardous material facility, Electricity facility, Basic process 1, Basic process 2, Hazardous material, Hot work, Flammable gas, High temperature or pressure, Static electricity, Dust, High voltage, Safety management 1, Safety management 2, Fire extinguisher 1, Protected ratio by Fire extinguisher, Fire extinguisher 2, Standpipe 1, Protected ratio by standpipe, Standpipe 2, Hydrant 1, Protected ratio by hydrant, Hydrant 2, Sprinkler 1, Protected ratio by sprinkler, Sprinkler 2, Total flooding gas system 1, Protected ratio by total flooding gas system, Total flooding gas system 2, Fire detection system 1, Protected ratio by fire detection system, Fire detection system 2, Emergency alarm system 1, Protected ratio by emergency alarm system, Emergency alarm system 2, Emergency notifying system 1, Emergency notifying system 2, Fire compartment 1, Fire compartment 2, Evacuation system 1, Evacuation system 2, Smoke control system 1, Smoke control system 2, Smoke control system 3, Auxiliary equipment required for fire brigade 1, Auxiliary equipment required for fire brigade 2, Public Fire Service 1, Public Fire Service 2, Public Fire Service 3, Public Fire Service 4, Additional insurance discount, Distance between buildings, Type of report, Year of construction 2, Year of construction 2, KFRI 2, KFRI 3, Percent rank of KFRI, KFRI 4, Percent rank of KFRI 4, Likelihood of KFRI 1, Percent rank of likelihood of KFRI, Likelihood of KFRI 4, Percent rank of likelihood of KFRI 4, Likelihood of KFRI 5, Percent rank of likelihood of KFRI 5, Likelihood of KFRI 6, and Percent rank of likelihood of KFRI 6.

We applied fire risk index of each building and factory for our proposed methods. First, we carried out logistic regression for fire risk forecasting. The response variable (Y) is the occurrence result of fire (1: occurred fire, 0: not occurred fire) and the explanatory variables are 106 variables except Y. To select the significant variables (features) explaining Y significantly, we used p-value (probability value) of each variable in the following hypothesis [10]:

$$H_0 : \beta_i = 0 \quad vs. \quad H_1 : \beta_i \neq 0 \tag{5}$$

where

β_i: model parameter of *ith* variable X_i

We reject the null hypothesis H_0 when p-value is less than 0.05 (95% confidence level). This means that X_i explains Y significantly. Using this feature selection process, we selected 18 variables from an entire 106 variables as follows: Year of construction ($< 2 \times 10^{-16}$), Number of total floor ($< 2 \times 10^{-16}$), Building structure ($< 2 \times 10^{-16}$), Building size ($< 2 \times 10^{\times}$), Fire load (1.70×10^{-12}), Fire facility (3.08×10^{-5}), Gas facility (8.71×10^{-10}), Hazardous material facility ($< 2 \times 10^{-16}$), Electricity facility (4.39×10^{-12}), Basic process ($< 2 \times 10^{-16}$), Hazardous material ($< 2 \times 10^{-16}$), Hot work (0.00089), Flammable gas (0.0138), High temperature or pressure (1.86×10^{-7}), Static electricity (0.000418), Dust ($< 2 \times 10^{-16}$), High voltage (1.15×10^{-12}), and Safety management (0.0429). The value in parenthesis for each variable is the p-value.

In this paper, we carried out the logistic regression analysis and DNN using the 18 explanatory variables and one response variable. We divided the entire data into training and test data sets. We extracted 70% randomly from our data for training and 30% for test. After various trials and errors, we finally designed our DNN architecture as follows: hidden layers, 3; hidden nodes in each hidden layer, 100; activation function, sigmoid; learning rate, 0.8; learning momentum, 0.5; epoch, 100; and weight decay, 0.000001. Also, we used R data language and its packages for fire data analysis [19]. Using the training data set, we constructed the forecasting models. To evaluate the performance

of structured models, we used the test data set. Table 2 shows the comparison result of prediction accuracy between the models.

Table 2. Prediction accuracy comparison.

Model	Accuracy
KFRI	0.5384
Logistic regression	0.5552
DNN	0.7514

KFRI is the predictive model currently used, and the accuracy result of the KFRI model is to predict whether a fire has occurred using the KFRI. Therefore, we found that fire prediction is a very difficult task. The prediction accuracy of logistic regression was higher than that of KFRI, but the performance improvement was not significant. Finally, we can see that the accuracy of fire prediction of the DNN model is significantly higher than that of KFRI or logistic regression.

For further performance improvement, we considered fire risk indexing. The fire risk indexing has been recommended and used as a rapid assessment to evaluate the fire risk of alternative concepts for large buildings. KFPA also uses this fire risk indexing method known as one of the KFRI models. In Table 1, we show the components (variables) for fire risk indexing and calculation formula of KFRI for manufacturing facilities. The hazard components of KFRI are placed in the numerator and the countermeasure components are placed in the denominator to reflect the risk of building. We calculate the KFRI by equation (6):

$$KFRI = \frac{B \times I \times P}{M \times F \times A \times V \times S \times G} \times 100 \tag{6}$$

where

B = Basic hazards or intrinsic hazards such as number of floor, structure, size, fire load, etc.

I = Ignition hazards due to fire, gas, electrical facilities, and hazardous material

P = Process hazards apply only for factories consist of basic process, hazardous material treating process, hot work process, etc.

M = Building safety management based on fire drill, education, hot work, smoking control, etc.

F = Fire protection equipment and system such as fire extinguisher, fire sprinkler, standpipe, etc.

A = Fire alarm system such as fire detecting system, notifying system, etc.

V = Fire compartment and evacuation system

S = Smoke control system and auxiliary equipment required for fire brigade.

G = Public fire service

The higher the risk, the higher the value of KFRI. Each component of KFRI is calculated combining its weight and inspection results by KFPA surveyors. The weights of each component for KFRI are originally decided by the analytical hierarchy process (AHP) with experienced KFPA field surveyors. After collecting enough inspection and fire incidents data for many years, we tried various approaches to improve its performance. New KFRI (NKFRI) was considered in order to compare the risks among factories with a total area of more than 3,000 m^2 and other specific buildings designated by the Korean law [20].

In the NKFRI model, we performed the fire risk indexing for factories. NKFRI more specifically introduced the concepts of likelihood and severity of fire by rearranging variables and then assigning optimal weight for each components based on its deviation. One of the main reasons to develop NKFRI is to make it easier to compare the risk among factories that have similar processes. In order to evaluate the relative standing of a value within a group, "PERCENTRANK" function of Excel 2016 was used.

$$NKFRI = Percentrank\ of\ likelihood\ index \times Percentrank\ of\ severity\ index \tag{7}$$

After calculating NKFRI, "PERCENTRANK" function is adapted again to finally compare the relative risk in a group. Table 3 shows the variable lists adapted when evaluating fire risk.

Table 3. Application range of variables.

Application Range	Variables
All	Years (y), Floors (e), Structure (s), Size (g), Fire load (q), Fire facility (i1), Gas facility (i2), Hazardous material facility (i3), Electricity facility (i4), Safety management(M), Fire extinguisher (F1), Standpipe (F2), Hydrant (F3), Sprinkler (F4), Total flooding gas system (F5), Fire detection system (A1), Emergency alarm system(A2), Emergency notifying system (A3), Fire compartment (V1), Evacuation system (V2), Smoke control system (S1), Auxiliary equipment required for fire brigade (S2)
General buildings only	Multi use (c), Movement of occupants (d), Accommodation (r)
Factories only	Basic process (p0), Hazardous material(p1), Hot work (p2), Flammable gas (p3), High temperature or pressure(p4), Static electricity (p5), Dust (p6), High voltage (p7)

To keep the independence of data, data was separated into model construction and model verification. The fire data during 2011–2017 was used to find out the optimized weights of each factor and then they were verified using the fire data during 2018–2019 data.

Table 4 shows the weight change for likelihood index of modules.

Table 4. Weight change for likelihood index.

Module	Components	KFRI Weights	NFKRI Weights
Basic (B) $B = y \times e \times s \times g \times c \times d \times r \times q$	Years (y)	Not used	6.99
	Floors (e)	1.25	2.42
	Structure (s)	1.25	4.70
	Size (g)	1.43	2.31
	Fire load (q)	2.5	2.28
Ignition (I) $I = i1 \times i2 \times i3 \times i4$	Fire facility (i1)	1.82	8.62
	Gas facility (i2)	1.44	9.98
	Hazardous material facility (i3)	1.33	9.98
	Electricity facility (i4)	1.41	9.98
Process (P) $P = p0 \times p1 \times p2 \times p3 \times p4 \times p5 \times p6 \times p7$	Basic process (p0)	1.15	9.93
	Hazardous material (p1)	1.03	1.18
	Hot work (p2)	1.03	1.39
	Flammable gas (p3)	1.03	1.29
	High temperature or pressure (p4)	1.03	1.44
	Static electricity (p5)	1.03	1.51
	Dust (p6)	1.03	1.96
	High voltage (p7)	1.03	9.99
Safety management (M)	Safety management (M)	4.25	1.15

The components and modules of likelihood and severity indexes were selected by KFPA experienced surveyors from the fire engineering point of view. In this paper, we tried to optimize the weights of each component for the likelihood index.

$$Likelihood\ index = \frac{B \times I \times P}{M} \tag{8}$$

Inspection data was combined with fire incidents data, and then classified into whether fire happened or not. In order to optimize the weights of each component, the concept of deviation based on the mean value of the likelihood index with or without fire incident was used. The deviation below was calculated as changing the weight of each component in the designated range, and then try to find optimized one that maximizes the deviation of likelihood index that consists of fire occurrence related components of KFRI using the tailor-made program:

$$Deviation = \frac{X - Y}{X} \times 100 \tag{9}$$

where

X = The mean value of fire frequency-related KFRI for buildings with fire
Y = The mean value of fire frequency-related KFRI for buildings without fire

The performance of optimized weights of each component was evaluated using the lift value defined below:

$$Lift\ Value = \frac{Top\ 10\%\ lift}{Baseline\ lift} \tag{10}$$

where

Baseline lift: Ratio of the number of fires included in the overall data before building the model.
Top 10% lift: Ratio of fires in the top 10% of the data sorted in descending order of the fire frequency-related KFRI. This value ranges from 0 to infinity. In addition, the larger this value, the better the performance of the model.

Table 5 illustrates the lift values of likelihood index of KFRI and proposed NKFRI models. In order to compare likelihood performance, identical components were adapted and for KFRI, origin weights decided by the AHP were used.

Table 5. Lift values of likelihood index of KFRI and NKFRI.

Likelihood Index of KFRI	Likelihood Index of NKFRI
1.5191	2.1421

The weights obtained using data from 2011–2017 were applied in 2018–2019, and then the likelihood index was recalculated respectively. When comparing the lift value using the first weights (KFRI) and the optimized weights (NKFRI), it was confirmed that the proposed NKFRI improved the lift value by 41.01%. So, we verified the improved performance of our proposed work, and this paper contributes to the fire risk assessment for various buildings and factories.

5. Discussion

The goal of this paper was to predict the occurrence of fire accurately. In this paper, we considered logistic regression, DNN, and optimized risk indexing for fire forecasting. From our experimental results, we found that the prediction accuracy of DNN is better than others such as KFRI or logistic regression. The accuracy of the prediction of the DNN model is 0.7514, which is an improved result compared to the prediction accuracy by the previous study of Rishickesh et al. (2019), 0.6838.

In addition, we also proposed an optimized risk indexing for fire prediction. We got 2.1421 as the lift value of this indexing. This result means that the accuracy of prediction is increased by 2.1421 times through the modeling based on the optimized risk indexing compared to before building the model. Therefore, using the proposed fire risk indexing with a lift value of 2.1421, we can predict fire occurrence efficiently for fire risk management.

6. Conclusions

In this paper, we proposed a model for fire risk assessment using statistical machine learning and optimized risk indexing. In general, to predict fire risk prediction is very difficult, this is because fires occur very rarely and the causes that affect fire occurrence are very diverse. So, we need more advanced models for fire risk prediction and assessment. Currently, the KFRI is widely used in practice to evaluate fire risk and then decide the discount rate of fire insurance premium in Korea. In this paper, we compared our proposed models with the KFRI and previous other researches. For the fire risk forecasting and management, we considered and constructed three models that are logistic regression, DNN and NKFRI (new fire risk index). The logistic regression analysis and DNN learning algorithm are based on the statistical machine learning. Also, the NKFRI is an optimized risk indexing model, an extended model of KFRI. In the experimental results, we found that the DNN and NKFRI provide more improved performance than traditional KFRI or previous research. However, it is difficult to actually apply machine learning to sensitive areas such as insurance premium decisions despite its better performance, because it is hard to explain the result clearly.

Therefore, in our future works, we will try to solve this problem preferentially. To increase the explanatory power of the model, we will apply probability theory and statistical inference to fire prediction models based on machine learning and fire risk indexing. Because of the characteristic of fire occurrence, it is very difficult to accurately predict the fire occurrence and risk. Although the accuracy of fire prediction by our proposed model has been improved compared to the existing research results, the accuracy of our fire prediction was not raised to the highest level. So, additional research strategies are needed to develop a model that can improve the performance of fire prediction. In this paper, we considered DNN with large hidden layers and nodes and got the prediction accuracy of 0.7514. We will consider convolutional neural networks with convolution operation for improving fire prediction accuracy larger than 0.7514. One of the advantages of using our models is it is easy to combine with other information. There are plenty of chances of more sources or information related to fire is open to public in this big data era. This process makes it easy to find out the optimal weights, whereas the AHP process requires a quantity of time and collaboration among engineers to get proper weights of each component. In addition, we will consider more advanced models such as Bayesian deep learning for improved fire risk prediction.

This paper contributes to the real domains related to fire risk assessment such as calculation and imposition of fire insurance premiums. Traditionally, the calculation of fire hazard ratings for buildings and factories relies on the subjective knowledge of a group of fire experts, but the results of our study have enabled this work to be quantifiable and objective. In our research, we focused on the fire risk forecasting and assessment for buildings and factories. We also expect that the application of this paper can be extended to fire risk related to forests and other domains.

Author Contributions: M.-Y.C. designed this research and collected the data set for the experiment. Also, M.-Y.C. developed proposed methodology. S.J. wrote this manuscript and made original draft. M.-Y.C. and S.J. analyzed the data to show the validity of this paper and performed all the research steps. All authors have read and agreed to the published version of the manuscript.

References

1. Rishickesh, R.; Shahina, A.; Nayeemulla Khan, A. Predicting Forest Fires using Supervised and Ensemble Machine Learning Algorithms. *Int. J. Recent Technol. Eng.* **2019**, *8*, 3697–3705.
2. Madaio, M.; Chen, S.T.; Haimson, O.L.; Zhang, W.; Cheng, X.; Hinds-Aldrich, M.; Chau, D.H.; Dilkina, B. Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta. In Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 185–194.
3. Watts, J.M. Index Approach to Quantifying Fire Risk. In Proceedings of the SFPE Symposium on Risk, Uncertainty, and Reliability in Fire Protection Engineering, Society of Fire Protection Engineers, Baltimore, MD, USA, 12–14 May 1999; pp. 39–45.
4. Watts, J.M. Fire Risk Indexing. *SFPE Handb. Fire Prot. Eng.* **2016**, *3*, 3158–3182.
5. Šakėnaitė, J.; Vaidogas, E. Fire Risk Indexing and Fire Risk Analysis: A Comparison of Pros and Cons. In Proceedings of the 10th International Conference Modern Building Materials, Structures and Techniques, Vilnius, Lithuania, 19–21 May 2010; pp. 1297–1305.
6. Nikolopoulos, E.I.; Destro, E.; Bhuiyan, M.A.E.; Borga, M.; Anagnostou, E.N. Evaluation of predictive models for post-fire debris flow occurrence in the western United States. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2331–2343. [CrossRef]
7. KFPA. Korea Fire Protection Association. Available online: https://www.kfpa.or.kr/eng/ (accessed on 10 January 2020).
8. Ross, S.M. *Introduction to Probability and Statistics for Engineers and Scientists*, 4th ed.; Elsevier: Seoul, Korea, 2012.
9. Theodoridis, S. *Machine Learning, A Bayesian and Optimization Perspective*; Elsevier: London, UK, 2015.
10. Hogg, R.V.; Tanis, E.A.; Zimmerman, D.L. *Probability and Statistical Inference*, 9th ed.; Pearson: Essex, UK, 2015.
11. Hogg, R.V.; McKean, J.M.; Craig, A.T. *Introduction to Mathematical Statistics*, 8th ed.; Pearson: Upper Saddle River, NJ, USA, 2018.
12. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.
13. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*, 3rd ed.; Morgan Kaufmann: Waltham, MA, USA, 2012.
14. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
15. Hilbe, J.M. *Modeling Count Data*; Cambridge University Press: New York, NY, USA, 2014.
16. Hilbe, J.M. *Negative Binomial Regression*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2011.
17. Cameron, A.C.; Trivedi, P.K. *Regression Analysis of Count Data*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2013.
18. Lesmeister, C. *Mastering Machine Learning with R*, 2nd ed.; Packt: Birmingham, UK, 2017.
19. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018. Available online: http://www.R-project.org/ (accessed on 1 August 2018).
20. Ministry of Government Legislation. Act on The Indemnification for Fire-Caused Loss and The Purchase of Insurance Policies. Available online: https://elaw.klri.re.kr/ (accessed on 21 May 2020).