FOREST FIRE MANAGEMENT AND RISK PREDICTION SYSTEM IN CHIANGMAI PROVINCE: A COMPARATIVE STUDY OF SUPPORT VECTOR MACHINE AND NEURAL NETWORK

A-SAN SOMBOONYING

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)
FACULTY OF GRADUATE STUDIES
2012**

Thesis
entitled

FOREST FIRE MANAGEMENT AND RISK PREDICTION SYSTEM IN CHIANGMAI PROVINCE: A COMPARATIVE STUDY OF SUPPORT VECTOR MACHINE AND NEURAL NETWORK

………………..…………..…..………
Mr. A-san Somboonying
Candidate

………………………..…..…..………
Lect. Mingmanas Sivaraksa, Ph.D.
Major advisor

………………………..…..…..………
Lect.Suratose Tritilanunt, Ph.D.
Co-advisor

………………………..…..…..………
Asst.Prof. Worasit Choochaiwattana, Ph.D.
Co-advisor

……………………………..…..…..………
Prof. Banchong Mahaisavariya,
M.D., Dip Thai Board of Orthopedics
Dean
Faculty of Graduate Studies
Mahidol University

……………………………..…..…..………
Lect. Supaporn Kiattisin, Ph.D.
Program Director
Master of Science Program in
Technology of Information System
Management
Faculty of Engineering
Mahidol University

Thesis
entitled

FOREST FIRE MANAGEMENT AND RISK PREDICTION SYSTEM IN CHIANGMAI PROVINCE: A COMPARATIVE STUDY OF SUPPORT VECTOR MACHINE AND NEURAL NETWORK

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science
(Technology of Information System Management)
on
June 7, 2012

……………….…………..…..………
Mr. A-san Somboonying
Candidate

……………………..…..………………
Lect.Waranyu Wongseree, Ph.D.
Chair

……………………..…..……………...…
Asst.Prof. Worasit Choochaiwattana, Ph.D.
Member

……………………..…..…………..……
Lect. Mingmanas Sivaraksa, Ph.D.
Member

……………………..…..……………...…
Mr. Theerapat Prayurasiddhi, Ph.D.
Member

……………………..…..………………..
Lect. Suratose Tritilanunt, Ph.D.
Member

……………………..…..……………...…
Prof. Banchong Mahaisavariya, M.D., Dip Thai Board of Orthopedics
Dean
Faculty of Graduate Studies
Mahidol University

……………………..…..……………...…
Lect. Worawit Israngkul, M.S.
Dean
Faculty of Engineering
Mahidol University

ACKNOWLEDGEMENTS

The success of this thesis can be attributed to the extensive support and assistance from my major advisor, Lect. Mingmanas Sivaraksa, and my co–advisor, Suratose Tritilanunt Lect. and Asst.Prof. Worasit Choochaiwattana. I deeply thank them for their valuable advice and guidance in this research.

I would like to acknowledge Thai Meteorological Department for providing the meteorological data and Bureau of Forest Fire prevention and controls, Department of National Parks, Wildlife and Plant Conservation Ministry of Natural Resources and Environment for providing the forest fire data.

This thesis is partially supported by Graduate Studies of Mahidol University Alumni Association and University Industry Research Collaboration Program (U – IRC). I would like to thank for their support and scholarship. And I would like to thank Lect. Kanat Poolsawasd, Dr. Theerapat Prayurasiddhi and the expert from Department of National Parks, Wildlife and Plant Conservation (DNP.) for many advices in this thesis.

My thanks also go to all the lectures of Technology Information System Management and Computer engineering in Faculty of Engineering, Mahidol University for their valuable advice.

Finally, I am grateful to my family for their financials support and love.

A-san Somboonying

FOREST FIRE MANAGEMENT AND RISK PREDICTION SYSTEM IN CHIANGMAI PROVINCE: A COMPARATIVE STUDY OF SUPPORT VECTOR MACHINE AND NEURAL NETWOK

A-SAN SOMBOONYING 5136039 EGTI/M

M.Sc. (TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)

THESIS ADVISORY COMMITTEE: MINGMANAS SIVARAKSA, Ph.D., SURATOSE TRITILANUNT, Ph.D., WORASIT CHOOCHAIWATTANA, Ph.D.

ABSTRACT

Forest fires destroy many forest areas in Thailand every year. In order to reduce the risks and hazards of forest fires, the first part of this study was to design and develop a web based application for forest fire management. The system was developed using the PHP language and MySQL as the database. The new system makes the working process more convenient and reduces duplication. For the evaluation process of the forest fire system, it was done by the users who work in the Bureau of Forest Fire Prevention and Control. The results of the evaluation were that the levels of the functional requirement test and security test were fine, the functional test and usability test were excellent. When the first part is finished, the forest fire data in the new system will be used for predicting forest fires by using data mining techniques and meteorological data. The data mining techniques used include Multilayer Perceptron (MLP) and Support Vector Machine (SVM). In this study, we incorporate Synthetic Minority Over-sampling Technique (SMOTE) to improve the performance of the model. Without SMOTE, it is very difficult to correctly identify the occurrence of forest fires given the high number of no fire occurrences. Our results show that SVM model produced better results than the MLP model. Therefore, SVM and metrological inputs are the best way to predict the risk of forest fire occurrences.

KEY WORDS: FOREST FIRE / IMBALANCE DATA / DATA MINING / SMOTE

79 pages

ระบบจัดการและพยากรณ์ไฟป่าในจังหวัดเชียงใหม่ กรณีศึกษาเปรียบเทียบระหว่างซัพพอร์ตเวกเตอร์แมชชีนและโครงข่ายประสาทเทียม

FOREST FIRE MANAGEMENT AND RISK PREDICTION SYSTEM IN CHIANGMAI PROVINCE: A COMPARATIVE STUDY OF SUPPORT VECTOR MACHINE AND NEURAL NETWOK

อาสาพ สมบูรณ์ยิ่ง 5136039 EGTI/M

วท.ม. (เทคโนโลยีการจัดการระบบสารสนเทศ)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์ : มิ่งมานัส ศิวรักษ์, Ph.D., สุรทศ ไตรติลานันท์, Ph.D., วรสิทธิ์ ชูชัยวัฒนา, Ph.D.

บทคัดย่อ

พื้นที่ป่าของประเทศไทยได้ถูกทำลายจากไฟป่าทุกปี เพื่อลดความเสี่ยงและอันตรายจากไฟป่า ในส่วนแรกของการศึกษาจะทำการออกแบบและพัฒนาระบบจัดการข้อมูลไฟป่าในรูปแบบของเวบ โดยพัฒนาโดยใช้ภาษา PHP และใช้ MySQL เป็นฐานข้อมูล ซึ่งในระบบที่พัฒนาขึ้นใหม่นี้จะทำให้การทำงานมีความสะดวกมากขึ้น ลดการทำงานที่ซ้ำซ้อนลง ในการประเมินระบบจัดการไฟป่านั้น จะทำโดยผู้ใช้งานที่ทำงานในสำนักควบคุมและป้องกันไฟป่า โดยผลของการประเมินมีดังนี้ คะแนนของ functional requirement test และ security test อยู่ในระดับดี และคะแนนของ functional test และ usability test อยู่ในระดับดีมาก เมื่อพัฒนาส่วนแรกของการศึกษาเสร็จจะนำข้อมูลไฟป่าในระบบที่สร้างขึ้นใหม่มาพยากรณ์การเกิดไฟป่าโดยใช้เทคนิคเหมืองข้อมูลร่วมกับข้อมูลทางอุตุนิยมวิทยา โดยเทคนิคเหมืองข้อมูลที่ใช้ประกอบด้วย Multilayer Perceptron (MLP) and Support Vector Machine (SVM) ในการศึกษานี้จะนำเทคนิค Synthetic Minority Over-sampling Technique (SMOTE) มาเพิ่มประสิทธิภาพของโมเดล เนื่องจากถ้าไม่ใช้ SMOTE การทำนายจะทำได้ยากเนื่องจากมีคลาสการไม่เกิดไฟป่ามากกว่าคลาสอื่น โดยผลการทดสอบพบกว่าผลการทำนายของ SVM ดีกว่า MLP ดังนั้นวิธีที่ดีที่สุดในการทำนายการเกิดไฟป่าคือการใช้เทคนิค SVM ร่วมกับข้อมูลทางอุตุนิยมวิทยา 6 ตัว

79 หน้า

CONTENTS

CONTENTS (cont.)

LIST OF TABLES

LIST OF TABLES (cont.)

LIST OF FIGURES

LIST OF FIGURES (cont.)

CHAPTER I
INTRODUCTION

1.1 Background and Problem Statement

The area of forest in Thailand has decreased dramatically from 273,632 square kilometers (53.33 % of area of Thailand) in 1961 to 172,176 square kilometers (33.56 % of area of Thailand) in 2009. The major cause of the decreasing forest area is increasing population. As a result, people burn forests for agriculture, tourism industry such as tourism accommodations, golf courses or development infrastructure of the government including the construction of the road and dam which are one of the destruction of forest over wide areas. The Forest encroachment will change the high density forest with high moisture to the low density forest (eg. Mixed Deciduous Forest**,** Dry Dipterocarp Forest, Savanna and grove wood) which easily create forest fire. Therefore, forest fire is one of the problems that affect the decreasing area of forest in Thailand. The graph in Figure 1.1 illustrates that forest fire in Thailand has occurred every year especially in the dry season from November to May estimably.

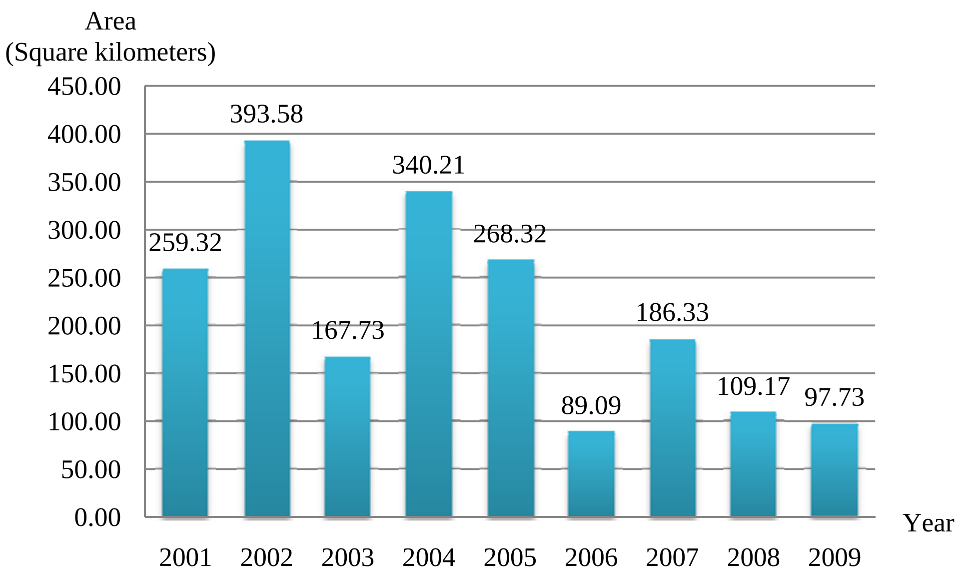


Figure 1.1 The graph shows the area of forest fire between 2001 - 2009 year.

In addition, in year 2001 to 2009, over 3,500 Square kilometers of forest area have been destroyed by forest fire. Therefore, the forest fire can burn a forest in a wide area and quickly. When the forest fire occurred, the first thing that be destroyed is the plants, animals living in the area. The next impact is changing the natural environment and damaged ecological balance of forest including soil and forests which difficult to recover and cause human suffering. Moreover, forest fire not only causes specific damage to the area where fire occurred only but it will extend damage to the environment and ecosystems of the world [1].

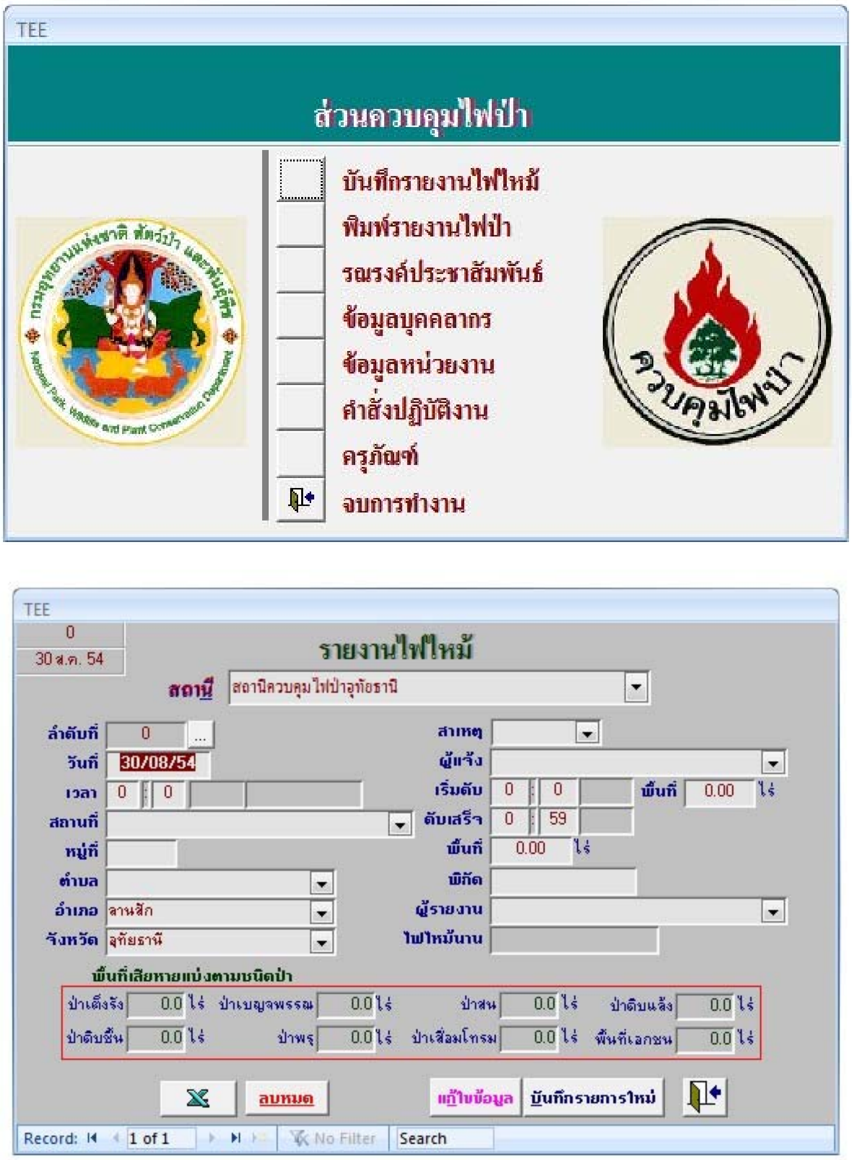


Figure 1.2 Shows current system of forest fire system management.

To manage and control forest fire in protected areas, Thailand has established Bureau of Forest Fire Prevention and Controls in Department of National Parks, Wildlife and Plant Conservation, Ministry of Natural Resources and Environment. The duties of Bureau of Forest Fire Prevention and Control are to manage and solve forest fire problems for conservation and restoration forest resources. The current fire system management of Bureau of Forest Fire Prevention and Control is only reporting when forest fire occurred, the people who see the forest fire will call to Forest Fire control station which located in all regions of Thailand. After that, the Forest Fire control station will send the staff to extinguish the fire and record the information such as time of forest occurred, location of forest fire and cause of forest fire (see Figure 1.2). The information is recorded in Access database which is not well designed so the information in this database can be inconsistent and duplicate. Moreover, the Access database is a stand-alone database which cannot report the incident of fire to the agency in Bangkok conveniently because the agencies in the region will report by sending the Access files to the agency in Bangkok. Because of this, it makes the delay in working process. If the system is web-based application, fire reporting will work more effectively. Because computer and the internet technology have developed rapidly and become tools to change the ways of working process. A development of the database system in web based application would make the working process more conveniently. In order to reduce risks and hazards of forest fires, the Bureau of Forest Fire prevention and controls uses the forest fire weather index (FWI) to predict the risk of forest fire area(more detail in chapter 2). However, FWI use many factors in the calculation and the result of predicting is inaccuracy so if creating a model that uses fewer factors and having high accuracy of prediction, the control and prevention of forest fire will be better.

Table 1.1 Show the frequency of forest fires occurred and damaged areas that cause by forest fires between 1 October 2010 to 17 August 2011.

Province	Frequency of forest fires (time)	Damaged areas (Square kilometers)
Chiangmai	369	3.64224
Mae Hong Son	176	1.312
Lamphun	141	1.3488
Chaiyaphum	143	5.336
Udonthani	136	1.9072

Studying environmental factors that influence the occurrence of forest fire including Biological, Physical and basic information are necessary. The study area of this research is Chiangmai province where in year 2010-2011 has the most frequency of forest fire (Table 1.1). Forest fire information will be transferred in a new and better designed database. In addition, use the forest fire information and weather information for comparing with two data mining technique to analysis risk of forest fire. For the purpose of fire control plans can be very effective.

1.2 Objective

1.2.1 To design and develop a forest fire system for creating reports and showing forest fire places on a google map.

1.2.2 To apply and compare artificial intelligence techniques for predicting the fire risk level of a day using fire data and meteorological data.

1.3 Scope of the study

1.3.1 Area of work

Focus on the Bureau of Forest Fire Prevention and Control, Department of National Parks, Wildlife and Plant Conservation, Ministry of Natural Resources and Environment in Chiangmai province. The available data sets are composed of:

1.3.1.1 Forest fire data from year 2001 to 2009 of Chiang Mai province.

1.3.1.2 Climate data including maximum temperature data, minimum temperature data, mean temperature data, humidity data, speed data and rainfall data of Chiangmai province from year 2001 to 2009.

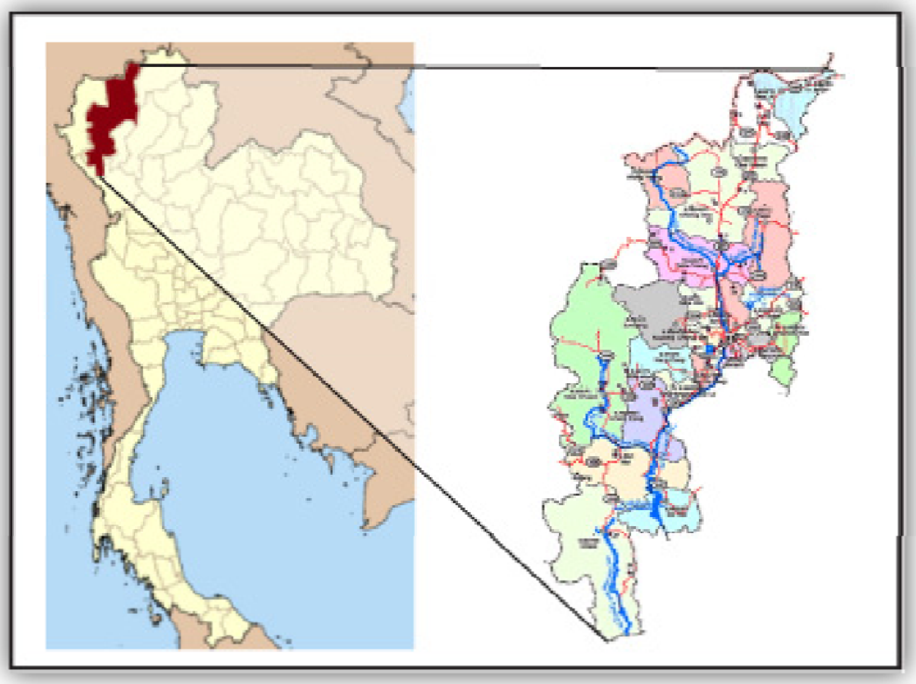


Figure 1.3 Shows map of Chiangmai province.

1.3.2 Area of solution

1.3.2.1 Developing a new system by using a web-based application and creating a new database for using data mining techniques to predict forest fire in the future.

1.3.2.2 Apply multi-layer perceptron and support vector machines for predicting risk forest fire.

CHAPTER II
LITERATURE REVIEW

The web-based fire forest system a case study for National Park in Chiangmai province was created as a website that can present the information about forest fire in Chiangmai province to help the staff get information and created the model to predict the forest fire. The review of related literatures is hereinafter.

2.1 Related theory

2.1.1 Definition of forest fire

There are various definitions that have been suggested for forest fire incidents.

Brown and Davis [2] define forest fire as “The fire that burns infinitely and spread freely, with the burning of natural material such as lumber, grass, weed and tree.”

U.S. Forest Service [3] defines forest fire as “The fire that burn natural material in forest and spread infinitely, uncontrolled. The natural materials are grass, leaf, stick.”

Apinan [4] defines forest fire as “The fire which caused by anything in the forest, prairie, and forestry plantations and spreading freely, uncontrolled.”

2.1.2 The cause of forest fire

The cause of forest can be separated into two main reasons including:

2.1.2.1 Natural

The natural forest fire was coming from many reason such as eruption, strict fiction or lightning but the mainly cause are following:

- Lightning - It can be divided into two kinds such as wet or blue lightning and dry or red lightning
- Wet of blue lightning - Wet of blue lightning is lightning when the rain begins to fall. The color of the lightening is blue. Although the fire will be burning, it will be burning in specific area and it does not spread into another area.
- Dry or red lightning - Dry or red lightning occurs without any rain. Lightning has a red light that occurs from the lightning cloud. The cloud is moving by itself in the own zone on an annual basis. If lightening on the fuel, the fuel will flare up.
- Friction - The friction will occur in dense and dry forests. When the tree will cause friction, it will make heat and flame will be spread. However in Thailand, it does not have a fire that caused by this because climate of Thailand has high humidity.

2.1.2.2 Human

Forest fire will cause by the people who do activities in the forest including

- Find the wild goods are the major reason to occur forest fire. The villagers burn forests in order to facilitate and to find the wild goods such as mushroom and firewood.
- Burn for remove weed , prepare for the planting area
- Hunting
- Agricultural
- Camping

2.1.3 Learning Algorithm

2.1.3.1 Multilayer Perceptron

Multilayer Perceptron (MLP) is a feed-forward artificial neural network model(ANN). It is supervised learning and using back propagation. MLP composes of nodes in layers which are an input layer, hidden layer and an output layer, as shown in Figure 2.1.

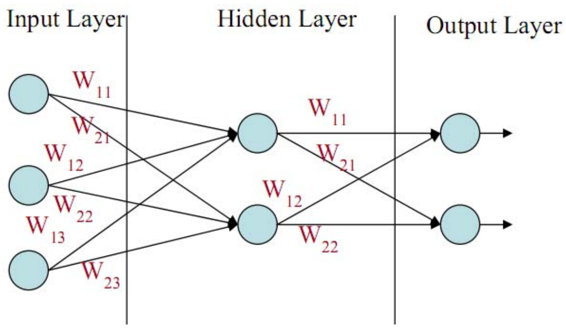
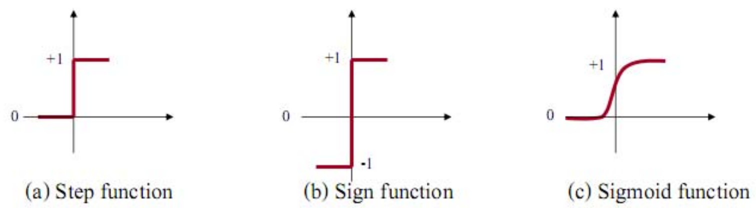


Figure 2.1 Layers of Multilayer Perceptron[5]

In each hidden layer, there is an activation function which calculates output when it receives the output from the previous layer. The number of hidden layer can be more than one and the activation function of each layer can be different type, as shown in Figure 2.2.



$$step\ (g(x)) = \begin{cases} 1, if\ g(x) > 0 \\ 0, otherwise \end{cases}$$

$$sigmoid\,(g(x)) = \frac{1}{1+e^{-x}}$$

$$sign(g(x)) = \begin{cases} +1, if\ g(x) \geq 0 \\ -1, otherwise \end{cases}$$

Figure 2.2 Activation Functions[5]

After one round of MLP process (Epoch), the weight will be adjusted and it is repeated until the error is satisfied or the number of Epoch reaches defined value.

ANNs model [6] composes of number elements, called neurons, as shown in Figure 2.1. Its mathematical model is

$$o_j = f_j \sum_k (w_{jk} \, x_k) , \tag{2.1}$$

where o_j is the output of a neuron,
f_j is a transfer function or activation function,
w_{jk} is an adjustable weight,
x_k is the input of a neuron.

The model of each node is shown in Figure 2.3

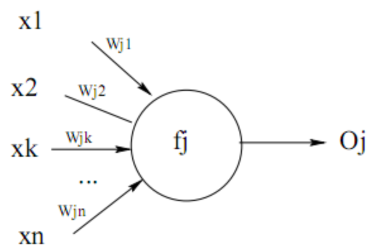


Figure 2.3 Model of each node in ANN[6]

MLP learning algorithm is shown in Figure 2.4. In MLP algorithm, the learning rate will affect the speed of learning. It will apply a smaller or larger proportion of the current adjustment to the previous weight. The higher the rate is set, the faster the network will learn, but if there is large variability in the input the network will not learn very well if at all.

Step 1: Set all weights to random numbers, between -1 and +1 to initialize the network.
Step 2: Calculate g(x) and output.
Step 3: Compare the output with the target.
Step 4: Propagate the error backwards.

(a) Adjust the output layer weights using the following formula

$$w_{ho} = w_{ho} + \Delta w_{ho}$$

$$, \quad \Delta w_{ho} = \eta \delta_o o_h$$

where w_{ho} is the weight of connecting hidden unit h with output unit o
η is the learning rate
o_h is the output at hidden unit h
δ_o is given by the following formula

$$\delta_o = o_o \times (1 - o_o) \times (t_o - o_o)$$

where o_o is the output at hidden unit h
t_o is the target for the node

(b) Adjust the input layer weights using the following formula

$$w_{ih} = w_{ih} + \Delta w_{ih}$$

$$, \quad \Delta w_{ih} = \eta \delta_h o_i$$

where w_{ih} is the weight of connecting node i of input layer with node h of hidden layer
$\eta$is the learning rate
o_i is the input of node i of input layer
δ_h is given by the following formula

$$\delta_h = o_h \times (1 - o_h) \times \sum_o (w_{ho} \times \delta_o)$$

Step5: Calculate the error from the difference between the target and the output. For example formula:

$$E = \frac{\sqrt{\sum_{n=1}^{p} (t_o - o_o)^2}}{p}$$

Step6: Repeat from step 2 for each pattern in the training set to complete one epoch.
Step7: Repeat from step 2 for the number of epochs, or until the error changes to desired value

Figure 2.4 MLP Learning Algorithm

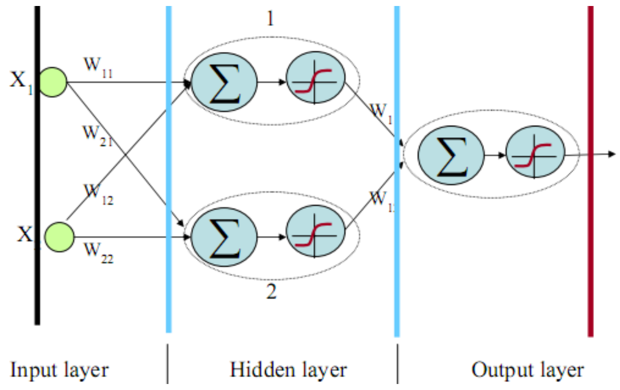


Figure 2.5 An Example of Multi-layer Perceptron[5]

2.1.3.2 Support Vector Machines

Support vector machines (SVM) are a useful technique for data classification, presented by Vapnik[7]. In this technique high generalization performance can be achieved with high quality and speed are guaranteed. In support vector machines, a predictor variable is called an "attribute", and a transformed attribute that is used to define the hyperplane is called a "feature". The task of choosing the most suitable representation is known as "feature selection". A set of features that describes one case (i.e., a row of predictor values) is called a "vector". So the objective of SVM modeling is to find the optimal hyperplane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. The vectors near the hyperplane are the "support vectors".

To construct an optimal hyperplane, SVM will iterate training algorithm which is used to minimize an error function. According to the form of the error function, SVM models for classifying can be separated into two distinct groups:

- Classification SVM Type 1 (also known as C-SVM classification)
- Classification SVM Type 2 (also known as nu-SVM classification)

Classification SVM Type 1 - The original SVM is known as C-SVM classification. For this type, SVM will separate a group of information and find W from equation 2.2 that regarded as the solution to minimize of the error function

$$\phi(w,\xi) = \frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i \qquad (2.2)$$

subject to the constraints:

$$y_i(w^T\phi(x_i) + b) \ \geq 1 - \xi_i \ and \ \xi_i \geq 0, i = 1,..,N$$

where C is the capacity constant,

w is the vector of coefficients

b is a constant

ξ_i is parameters for handling nonseparable data (inputs).

i is the N training cases.

$y \in \pm 1$ is the class labels and x_i is the independent variables.

The kernel (ϕ) is used to transform data from the input (independent) to the feature space. $C\sum_{i=1}^{N}\xi_i$is the sum of data error for analytic result. So the SVM performance is affected by these parameters: C – a trade-off between the model complexity and the amount up to which deviations larger than and γ – the parameter of the kernel.

Classification SVM Type 2 - In contrast to Classification SVM Type 1, the Classification SVM Type 2 will replaces C with a different parameter $v \in [0,1]$ that serves as an upper bound on the fraction of margin errors and a lower bound on the fraction of support vectors.

$$\phi(w,\xi) = \frac{1}{2}w^T w - v_\rho + \frac{1}{N}\sum_{i=1}^{N}\xi_i \qquad (2.3)$$

subject to the constraints:

$$y_i(w^T\phi(x_i) + b) \geq \rho - \xi_i \; and \; \xi_i \geq 0, i = 1,..,N \; and \; \rho \geq 0$$

The common used kernel functions were as follows:

- Linear kernel

$$K(x_i, x_j) = x_i^T x_j \tag{2.4}$$

- Gaussian radial basis function

$$K(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) \tag{2.5}$$

- Polynomial kernel

$$K(x_i, x_j) = \left(x_i^T x_j + D\right)^p \tag{2.6}$$

Figure 2.6 shows the resulting SVM classifier with the Gaussian radial basis function kernel, by $\sigma = 1.75$, have been used. Dotted lines mark the margin and circled points the support vectors.

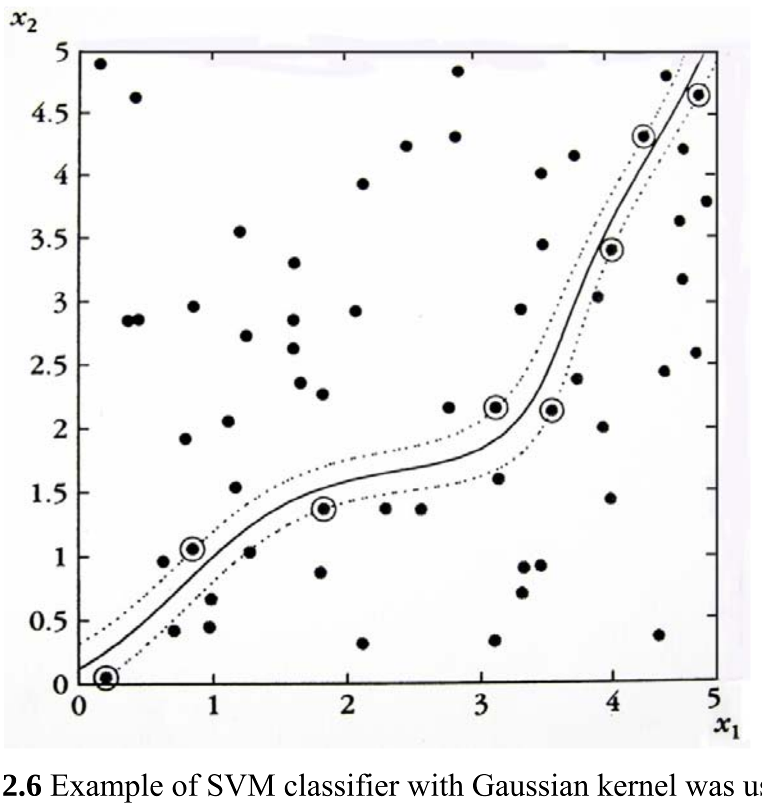


Figure 2.6 Example of SVM classifier with Gaussian kernel was used.[8]

However, using a hyperplane to separate the feature vectors into two groups works well when there are only two target categories. In the case which the target variable has more than two categories, there are two techniques to solve these include one against many and one against one. One against many will construct one SVM per class, which is trained to distinguish samples of one class from samples of all remaining classes. In one against one, the *c(c-1)/2* are constructed where *c* is the number of categories and trained to distinguish the samples of one class from the samples of another class.

2.1.4 Prediction Tool

Weka (Waikato Environment for Knowledge Analysis) is a popular program of machine learning software written in Java. It is developed by the University of Waikato, Hamilton, New Zealand. WEKA is open sources software available under the GUN General Public License [9]. The Weka GUI Chooser is shown in Figure 2.7. Weka composes of tools for data mining problems; regression, classification, clustering, association rules, and visualization.

Figure 2.7 Weka GUI Chooser

In this study, using Weka as a tool to predict the risk of forest fire occurred and the algorithms which used were Multi-layer Perceptron and Support Vector Machines. The data that input to Weka come from Forest Fire prevention and control station which located in Chiangmai Province.

2.1.5 Imbalance data sets[10]

Imbalanced data sets can occur in many domains, such as medical, information technology, biology, and finance. With imbalanced datasets, the conventional way of maximizing overall performance will often fail to learn anything useful about the minority class, because of the dominating effect of the majority class. Consider a problem where 99% of the data belongs to one class, and only 1% is rare class examples. A learner can probably achieve 99% accuracy with ease, but still fail to correctly classify any rare examples.

2.1.6 Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE (Synthetic Minority Over-sampling Technique) is an over-sampling approach proposed and designed in Nitesh V. Chawla and el.[11] SMOTE will generate synthetic examples in a less application-specific manner, by operating in "feature space" rather than "data space". The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the *k* minority class nearest neighbors. Depending upon the amount of over-sampling required, neighbors from the *k* nearest neighbors are randomly chosen. Our implementation currently uses five nearest neighbors. For instance, if the amount of over-sampling needed is 200%, only two neighbors from the five nearest neighbors are chosen and one sample is generated in the direction of each. Synthetic samples are generated in the following way: take the difference between the feature vector (sample) under consideration and its nearest neighbor; multiply this difference by a random number between 0 and 1, and add it to the feature vector under consideration. This causes the selection of a random point along the line segment between two specific features (Fig 2.8). This approach effectively forces the decision region of the minority class to become more general.

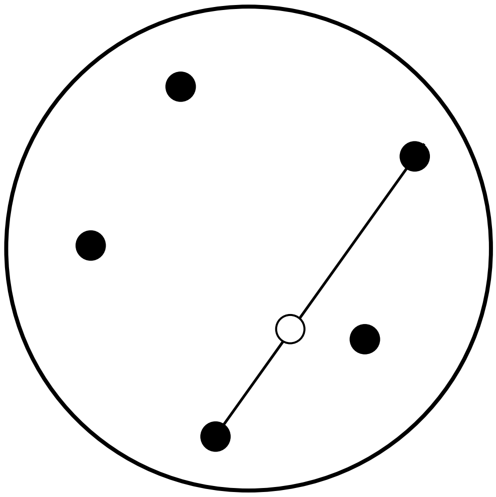


Figure 2.8 Over-Sampling with SMOTE. The minority class is oversampled by taking each minority class sample and introducing synthetic examples (white circle) along the line segments joining any/all of the *k* (default=5) minority class nearest neighbors (black circles).[12]

2.1.7 General information of Chiangmai Province

2.1.7.1 Location and Boundary

Chiangmai province is in the northern part of Thailand. It covers an area of approximately 20,107.057 km² or 12,566,910 rais divided to forest 69.92% (8,787,656 rais), agriculture 12.82% (1,611,971 rais) residential and others 17.26% (2,167,971 rais)

The boundary of the province is connected to;

North Shan State, Myanmar

South Amphur Samngao, Tak Province

East Chiang Rai, Lamphun, Lampana

West Mae Hongson

2.1.7.2 Geography

The most part (78%) of Chiangmai's land is covered by mountains and forests. The mountain ranges generally run in a north – south alignment through the province connecting to the mountain range in Yunnan, China and Shan State, Myanmar.

2.1.7.3 Weather

The average temperature is 26°C Chiangmai consists of 3 seasons.

1) Hot Season, the weather is very hot in the midday and become cool at night, low humidity, the average high is about 30°C.

2) Rainy Season, the average rainfall is 1,270 millimeters, relative humidity is 65.6%.

3) Cold Season, normally is not too cold, sometime foggy, the average temperature is 13.94°C.

Chiangmai has several kind of forest such as hilly evergreen forest, dry evergreen forest, mixed deciduous forest, dry dipterocarp forest, mixed dry dipterocarp and hilly dipterocarp forest, deciduous dipterocarp forest, etc. The forest area combined from natural forest, forest garden and natural restoration forest by 8,787,656 rais (by law) or 69.62% of Chiangmai area, divided to 25 Conserved Forests, 13 National parks, 1 Wildlife Sanctuary. Most Forest Fire Prevention and Control stations in Chinagmai are located in National parks

Forest Fire Prevention and Control stations in Chinagmai including:

1) Chiangmai Forest Fire Prevention and Control
2) Maeon Forest Fire Prevention and Control
3) Huaynamdang Forest Fire Prevention and control
4) Maeping Forest Fire Prevention and control
5) Maekuang Forest Fire Prevention and control
6) Doi Inthanon Forest Fire Prevention and Control
7) Doi Booluang Forest Fire Prevention and Control
8) Omkoi Forest Fire Prevention and Control
9) Chaiprakarn Forest Fire Prevention and Control
10) The Royal Development Project’s Forest Fire Prevention and Control
11) Samerng Forest Fire Prevention and Control
12) Ban Lek Ni Pa Yai Forest Fire Prevention and Control

To efficiently manage and control forest fire, many researchers were developing the fire forest prediction models which were following:

2.2 Related literatures

2.2.1 A prediction model for forest fire-burnt area based on meteorological factor [13].

The purpose of this study was to study and analyze the occurrence patterns of forest fire in Heilongjiang Province, China by using meteorological factor base on statistic analysis theory to build the model in prediction fire-burnt area. The meteorological factors that used in this study were the average wind speed, the relative humidity and the mean temperature. This average precision of the model was 63.3%.

2.2.2 Application of Apriori Algorithm to the Data Mining of Wildfire [14].

The purpose of this study was to learn the rules for predicting the forest fire in Fangshan Forestry Bureau in the south of Beijing, China from 1986 to 2007 by using the data, that is the meteorological and forest fire data. The daily meteorological data was including temperature, moisture, wing speed and rainfall. The daily forest fire data was including the fire location, the burn area, and the time finding the fire, the time putting out the fire, the number of the people, vehicles and coordinate. Before in put this data to Apriori algorithm, the data preparation was needed.

The data preparation was following:

1. The input data was divided into two kinds, one was that the fire number was zero, and the other was not
2. The dataset, which number is zero, will be divided further, from which the maximum and the minimum will be got and the difference value is divided by 3, so all the data are divided into 3 groups according to the quotient, and each group will be set a value as high, medium or low
3. The output data was divided into four kinds according to its fire number, that is often, sometime, rare and none

When complete the data preparation, the data was input to Apriori algorithm. The results of Apriori algorithm were the rules between the wildfire number and the weather for example, when moisture=low and rainfall=low, it had a chance to occur forest fire. (Table 2.1)

Table 2.1 Rules between the wildfire number and the weather factors

Rule number	Antecedent	Consequent (Fire Broken?)
1.	Rainfall=low	Y
2.	Moisture=low	Y
3.	Moisture=low and rainfall=low	Y
4.	Air temperature=mid	Y
5.	Air temperature=mid and rainfall=low	Y
6.	Wind speed=mid	Y
7.	Wind speed=mid rainfall=low	Y
8.	Moisture=low wind speed=mid	Y
9.	Moisture=low wind speed=mid	Y
10.	rainfall=low	Y
11.	Wind speed=low	Y
12.	Wind speed=low rainfall=low	Y
13.	Temperature=high	Y
14.	Temperature=mid moisture=low	Y
15.	Temperature=mid moisture=low rainfall=low	Y
	Temperature=high rainfall=low	

The conclusion was the first factor which affected the wildfire was temperature and there was no fire occurring while the temperature is low. When the temperature is middle, the fire can occur when the moisture is low. However, if the temperature is high, the fire will occur out only if the rainfall is low. The temperature value of the most records is middle. Moisture is a constrained factors but not a crucial one. There is no fire when the moisture is not low.

2.2.3 A data mining approach to predict forest fires using meteorological data [15].

The purpose of this study was to find the best configuration to predict forest fire by using meteorological data. The meteorological data was used because it includes in the forest fire weather index (FWI). The forest fire weather index (FWI) is the Canadian system for rating danger with six components including:

1. Fine Fuel Moisture Code (FFMC) denotes moisture content surface litter and influences ignition and fire spread.
2. Duff Moisture Code (DMC) the moisture content of shallow.
3. Drought Code (DC) the moisture content of deep organic layers and deep organic layers, which affect fire intensity.
4. Initial Spread Index (ISI) was the same with DC.
5. Buildup Index (BUI) represents the amount of available fuel.
6. FWI index is an indicator of fire intensity.

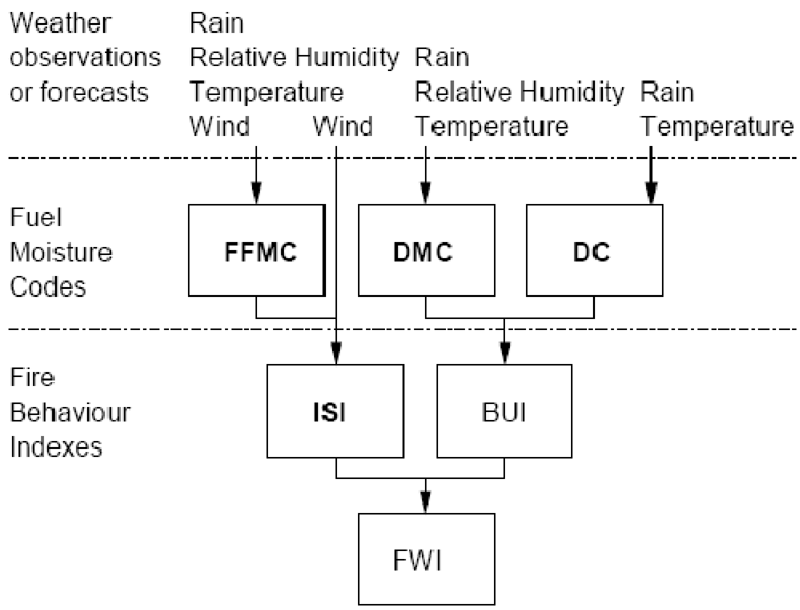


Figure 2.9 The Fire Weather Index structure (adapted from [16])

The other data used in the experiments was collected from January 2000 to December 2003 from the Montesinho natural park located in Tr´as-os-Montes northeast region of Portugal supra-Mediterranean with an average annual temperature within the range 8 to 12◦C. and the database are separated into two sources. The first database was collected by the inspector who was responsible for the Montesinho fire occurrences including forest fire occurred, time, date, spatial location within a 9×9 grid (Figure 2.8), the type of vegetation involved, the six components of the FWI system and the total burned area. The second database was collected by the Braganca Polytechnic Institute, containing several weather observations (e.g. temperature, relative humidity, rain and wind). The two databases were stored in tens of individual spreadsheets (Table 2.2).

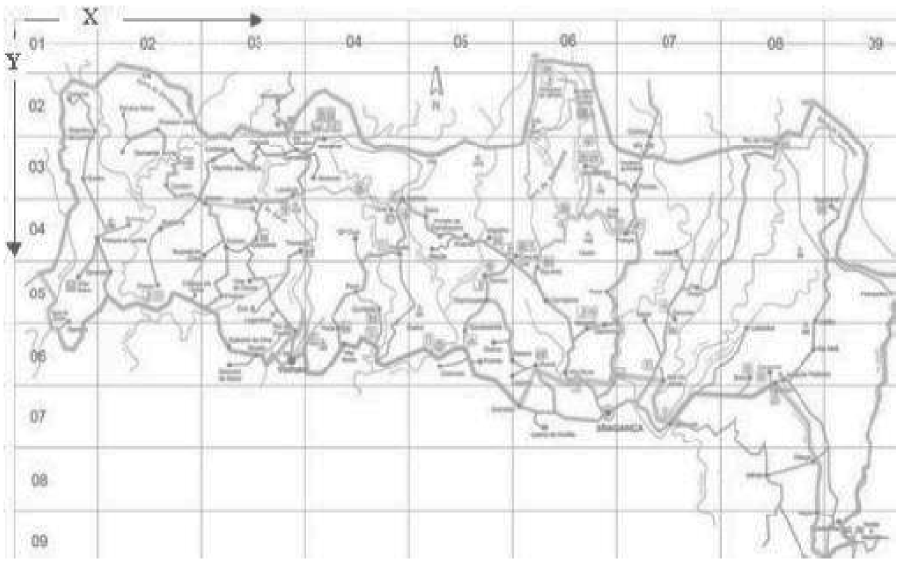


Figure 2.10 The map of the Montesinho natural park with divided into 9×9 grid

The data mining used in the experiments including: Multiple regressing, Decision tree, Random forest, Neural network, Support vector machine. This study was conducted using conducted using the RMiner [17], an open source library for the R statistical environment [18] that facilitates the use of DM techniques in classification and regression tasks. In particular, the RMiner uses the random Forest (RF algorithm by L. Breiman and A. Cutler), nnet (for the NN) and kernlab (LIBSVM tool [19]) packages.

Table 2.2 The preprocessed dataset attributes

Attribute	**Description**
X	x-axis coordinate (from 1 to 9)
Y	y-axis coordinate (from 1 to 9)
month	Month of the year (January to December)
day	Day of the week (Monday to Sunday)
FFMC	FFMC code
DMC	DMC code
DC	DC code
ISI	ISI index
temp	Outside temperature (in ◦C)
RH	Outside relative humidity (in %)
wind	Outside wind speed (in km/h)
rain	Outside rain (in mm/m2)
area	Total burned area (in ha)

Finally, the experimental results show the best configuration to predict the forest fire area were a support vector machine with four meteorological inputs including: temperature, relative humidity, rain and wind.

2.2.4 Artificial Intelligence for Forest Fire Prediction: A Comparative Study [20].

The purpose of this study was to analyze the forest fire prediction methods based on artificial intelligence techniques. Two forest fire risk prediction algorithms, based on support vector machines and artificial neural networks. The daily number of forest fires that used in this study provided by the Lebanese Ministry of Environment and weather parameter provided by the Lebanese Agricultural Research Institute (LARI), covering the Lebanese territory and spans nine years between 2000 and 2008. The weather parameters providing includes:

- The minimal temperature of the day, **Tmin**
- The maximal temperature of the day, **Tmax**
- The average humidity of the day

- The solar radiation over the day
- The average wind speed over the day
- The cumulative precipitation level starting in October 1 up to the specific day

Although **Tmin** might not be useful for indicating the hazard of forest fires, but the difference between **Tmax** and **Tmin** gives an additional indicator about the air moisture. Indeed when the air moisture is high, the difference between **Tmax** and **Tmin** over the day is low. In order to fuse the above features and make fire predictions, ANN and SVM are used. These learning mechanisms are introduced next and their performances are compared.

Table 2.3 Output scale of forest fire risk.

	Scale 1	Scale 2	Scale 3	Scale 4
June	0	1-3	4-7	>7
July	0	1-4	5-8	>8
August	0	1-3	4-15	>15
September	0	1-4	5-16	>16
October	0	1-7	7-11	>11

The proposed method introduces a fire risk index on a scale of 1 to 4, where 1 corresponds to the lowest fire risk and 4 to the highest fire risk. This index is based on the number of fires that occurred on a specific day can be used to estimate the range of number of fires that could happen on that day. Class 1 always corresponds to a no fire day. As for class 2, it corresponds to any number of fires that falls between the first quartile and the third quartile of the number of fire distribution that is given by the historical data. Class 3 corresponds to an increase of risk by 10% from the third quartile and class 4 corresponds to any risk greater than 10%. Table 2.3 shows the corresponding number of fires per month and per scale that corresponds to the boundaries between the different classes. The performance of the architecture is evaluated by computing the average error of the number of fires predicted. Denote by Ni the number of days of true class *i* and by d_{ij} the total number of days of true class *i* predicted to be of class *j*. Denote also by $q_{min,c}$ the lower boundary of class *c* and by $q_{max,c}$ the upper boundary for that same class. Using the class boundaries defined in

Table 2.3 it is possible to state that if a true class of a day is 1 and the decision made by the algorithm is also 1 then there is no fire error. If the decision was *j* then there is an error equal to $q_{min,j}$.

Table 2.4 The forest fire prediction results.

	June				July				August				September				October			
Scale	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
SVM	0.5	1.0	1.0	1.1	1.1	0.8	1.2	1.1	0.6	0.6	0.4	2.5	1.7	0.6	1.4	1.0	2.3	1.1	1.8	5.5
ANN	0.5	0.5	1.7	2.1	0.8	0.1	1.0	1.1	0.2	0.7	1.2	0.9	1.7	0.6	0.7	0.2	1.4	1	0.4	4.8

Finally, the summary of results for both algorithms is presented in Table 2.4. This table shows the average error in the number of fires predicted per class. This error is computed as the average per class of prediction deviation from the closest error class boundary. For example, for the month of June, a point belonging to class 2 mistakenly being predicted as class 4 will contribute a four-fire error. The four-fire error is the difference between the closest class bounds of 3 fires (for class 2) and 7 fires (for class 4). Comparing the two methods, except for the five highlighted cases, the average error for ANN is less than the average error for SVM. In general, both algorithms are giving acceptable results for all months. Note that the system is only being tested for the Lebanese fire season months of June, July, August, September and October in year 2009.

2.2.5 Efficient forest fire occurrence prediction for developing countries using two weather parameters [21].

The purpose of this study was to predict the forest fire occurrence by reducing the number of monitored features, and eliminating the need for weather prediction mechanisms and comparing with two artificial intelligence based methods, artificial neural networks (ANN) and support vector machines (SVM). The reason is to reduce errors due to inaccuracies in weather prediction. The challenge is to choose a limited number of easily measurable features in the aim of reducing the cost of the system and deployment and maintenance. At the same time, the chosen features must have a high correlation with the risk of fire occurrence. The daily number of forest fires that used in this study provided by the Lebanese Ministry of Environment and weather parameter provided by the Lebanese Agricultural Research Institute (LARI), covers the Lebanese territory and spans the nine years between 2000 and 2008. Typical weather parameters used as features are following:

- The average humidity of the day
- The cumulative precipitation level starting in October 1 up to the specific day

In order to fuse the above features and to make fire predictions, SVM and ANN are used. The proposed method introduces a fire risk index on a scale of 1 to 4, where 1 corresponds to the lowest fire risk and 4 to the highest fire risk. This index is based on the number of fires that occurred on a specific day can be used to estimate the range of number of fires that could happen on that day. Class 1 always corresponds to a no fire day. As for class 2, it corresponds to any number of fires that falls between the first quartile and the third quartile of the number of fire distribution that is given by the historical data. Class 3 corresponds to an increase of risk by 10% from the third quartile and class 4 corresponds to any risk greater than 10%.

Table 2.5 Prediction results for the five-month fire season.

Reduced features					Full features				
(a) *June*									
Class	1	2	3	4	Class	1	2	3	4
SVM E_i	**0.38**	**0.83**	1.78	**1**	SVM E_i	0.51	1.0	**1.07**	1.16
ANN E_i	0.86	**0.45**	**0.85**	**0.83**	ANN E_i	**0.51**	0.56	1.78	2.16
SVM Es_i	**0.3**	0.61	1.07	**0.54**	SVM Es_i	0.40	**0.57**	**0.78**	0.55
ANN Es_i	0.70	**0.38**	**0.71**	**0.5**	ANN Es_i	**0.51**	0.43	1.07	1
(b) *July*									
Class	1	2	3	4	Class	1	2	3	4
SVM E_i	**0.45**	**0.57**	1.44	**0.47**	SVM E_i	1.18	0.87	**1.27**	1.17
ANN E_i	**0.66**	0.32	**0.94**	**0.52**	ANN E_i	0.80	**0.16**	1.0	1.11
SVM Es_i	**0.45**	**0.46**	**0.77**	**0.29**	SVM Es_i	0.68	0.60	0.77	0.64
ANN Es_i	**0.60**	0.26	**0.76**	**0.35**	ANN Es_i	0.68	**0.16**	0.82	0.58
(c) *August*									
Class	1	2	3	4	Class	1	2	3	4
SVM E_i	**0.17**	0.86	1.4	**0.22**	SVM E_i	0.61	**0.61**	**0.42**	2.52
ANN E_i	**0.18**	0.81	1.66	**0.14**	ANN E_i	0.27	**0.77**	**1.25**	0.90
SVM Es_i	**0.17**	0.66	1.0	**0.22**	SVM Es_i	0.46	**0.45**	**0.42**	0.76
ANN Es_i	**0.18**	**0.60**	1.16	**0.14**	ANN Es_i	0.27	0.77	**0.91**	0.42
(d) *September*									
Class	1	2	3	4	Class	1	2	3	4
SVM E_i	**0.51**	**0.45**	**1.0**	**0.16**	SVM E_i	1.68	0.59	1.41	1.03
ANN E_i	**0.80**	**0.34**	0.92	0.31	ANN E_i	1.73	0.63	**0.78**	**0.21**
SVM Es_i	**0.43**	**0.45**	**0.80**	**0.16**	SVM Es_i	0.56	0.56	0.93	0.22
ANN Es_i	**0.5**	**0.34**	**0.71**	0.31	ANN Es_i	0.91	0.43	0.78	**0.21**
(e) *October*									
Class	1	2	3	4	Class	1	2	3	4
SVM E_i	**0.93**	1.0	**1.2**	**3.09**	SVM E_i	1.21	**0.97**	2.4	5.63
ANN E_i	1.43	0.91	**0.45**	**1.76**	ANN E_i	**1.23**	**0.44**	1.54	4.38
SVM Es_i	**0.41**	0.64	**0.9**	**1.04**	SVM Es_i	0.57	**0.61**	0.2	1.54
ANN Es_i	0.66	0.58	**0.45**	**0.9**	ANN Es_i	**0.63**	**0.36**	1.0	1.52

The experiments are shown in Table 2.5. The second row corresponds to the errors over the four classes using the SVM prediction algorithm and the third row presents the fire errors over the four classes using the ANN algorithm. The forth row corresponds to the average class error taken over all the days of class i for SVM and the fifth row is that same average class error for ANN. For example, in the reduced feature results for June, the average error on the number of fires over all days of class 1 is 0.38 fires, while the same average error for ANN is 0.86. The corresponding scale error is 0.3 for SVM and 0.7 for ANN. It is worth noting that the highlighted cells in the reduced features tables show the months and classes where ANN outperforms SVM. It is shown that the average error for ANN is in 12 cases out of 20 better than the average error of SVM. The comparison of the performance of the two algorithms over all months and over all classes for the reduced features shows that ANN is better than SVM on average by 0.17 fires and 0.04 classes. Note that bold face is used to compare between reduced features and full features cases indicating which is better. It is visible that the reduced features algorithm for SVM and for ANN outperforms the corresponding algorithm trained with the full features in most of the cases. Finally, note that the system is only tested for the Lebanese fire season months of June, July, August, September and October.

In the end, the fire/no fire scenario was tested. This scenario was a binary classification between days where no fires had occurred, and days where any number of fires had occurred.SVM and ANN were used in this test which using the K-fold cross-validation technique. The Gaussian kernel function was used for SVM and a 4–4– 1 architecture was used for ANN with the tangent sigmoid transfer function. The accuracy reported in Table 2.6.

Table 2.6 Fire/no fire scenario accuracy

Month	*SVM* (%)	*ANN* (%)
June	93.02	91.2
July	92.11	90.1
August	94.21	90.2
September	94.12	90.1
October	93.5	89.1

In summary, the support vector machines and artificial neural networks techniques have proved to achieve very low error on the number of fires predicted. It also showed that ANN outperforms SVM on average by 0.17 fires, while SVM outperforms ANN in the binary classification of fire/no fire scenario. The operational use of the developed algorithms for the prediction of forest fire occurrence in more fire seasons and in different study areas might be worth investigating in the future.

2.2.6 The analysis of the influential factors on quantitative fluctuation of student case study : Wangsapung industrial and community college [22].

The purpose of this study was to develop the analysis factor that influence the increasing and decreasing of the number of students by using Muti-layer Perceptron with Weka program. The data that used in this study is coming from the database of Wangsapug college students between 2004 to 2008 including: gender, age, status of parents, incoming of family, occupation of parents, major and levels of study. And then transforms the data into same pattern and use ratio 60:40, 70:30 and 80:20 of training set and testing set. In this study, use Weka for analysis. The result demonstrated that the priority of the factors is less significant to the age, tumbon , level , sex , major , status , income and occupant.

2.2.7 Learning Classifiers from Imbalanced, Only Positive and Unlabeled Data Sets [12].

The purpose of this study was investigated several re-sampling techniques in improving the learning from the imbalanced data. These include over-sampling with SMOTE(OS_SMOTE), Oversampling by duplicating minority examples(OSbD), random undersampling(US). Then three data mining algorithm (Decision Tree(DT), Naïve Bayes(NB), and Neural Network(NN)) were trained on the rebalanced training sets and used to classify the test set. The results showed the re-sampling techniques significantly improved the accuracy of the minority class on the test set specifically SMOTE techniques but the Naïve Bayes classifier with three re-sampling techniques did not improve the accuracy. The effect of resampling techniques on imbalanced data was shown in Fig 2.10

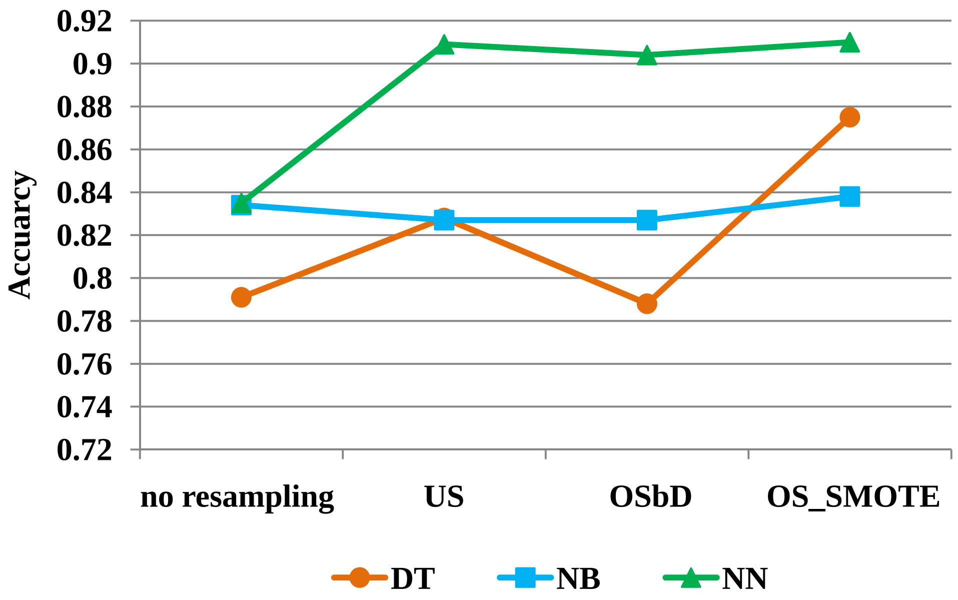


Figure 2.11 Effect of resampling techniques on imbalanced data

2.2.8 The summary

To effectively manage and control forest fire, there are many forest fire prediction models that use the forest fire information and weather information with data mining technique. The briefly summary from related literature are following:

QuZhi-Lin and Hu Hai-Qing[13] proposed a model to analyze the occurrence patterns of forest fire in Heilongjiang Province, China by using meteorological factors based on statistical analysis theory to build a model in prediction fire-burnt area. The meteorological factors that used in this study are the average wind speed, the relative humidity and the mean temperature. This average precision of the model is 63.3%.

HU Lin et al.[14] presented the rules to predict forest fire in Fangshan Forestry Bureau in the south of Beijing, China from 1986 to 2007 by the meteorological and forest fire data with Apriori Algorithm. The daily meteorological data are temperature, moisture, wind speed and rainfall. The daily forest fire data are the fire location, the burnt area, and the time finding the fire, the time putting out the fire, the number of the people, vehicles and coordinate. In conclusion, the result of this model found that the primary factor which affects the forest fire is temperature. There is no fire occurring while the temperature is low.

Cortez Paulo and MoraisAnibal[15] use temperature, relative humidity, rain and wind as the meteorological data and use Multiple Regressing, Decision Tree, Random Forest, Neural Network, Support Vector Machine as algorithm for predicting the forest fire occurrence. The results of study show the best configuration to predict the forest fire area are a support vector machine.

George Sakr and Elhajj Imad[20] adopted Neural Network and Support Vector Machine to predict forest fire by used six meteorological data included the minimal temperature, the maximal temperature**,** the average humidity, the solar radiation, the average wind speed and the cumulative precipitation level. When comparing the two algorithms, the average error of Neural Network is less than the average error of Support Vector Machine.

George Sakret al. [21] discovered the way to reduce the meteorological data factor by using Support Vector Machine and Neural Network. The result shows average humidity and cumulative precipitation level can use to predict forest fire for reducing cost.

Moreover, in a forest fire problem, the issue of imbalance data normally occurred. Fire occurrence does not happen that frequently, therefore the number of no fire occurrence class is enormously higher than other classes. In these cases, standard classifiers algorithm will tend to be overwhelmed by the large classes and ignore the small ones. Particularly, they tend to produce high predictive accuracy over the majority class, but poor predictive accuracy over the minority class given a higher overall accuracy. The solutions to solve these problems are re-sampling data such as over-sampling and under-sampling. There are many researches to solve imbalance data. For example, Yetian Chen investigated several re-sampling techniques in improving the learning from the imbalanced data [12]. These include random under-sampling (US), oversampling by duplicating minority examples (OSbD), oversampling with SMOTE (OS_SMOTE). The result shows the re-sampling techniques significantly improved the accuracy of the minority class on the test set specifically on SMOTE techniques.

Finally, in this study the researchers used the maximum temperature data, minimum temperature data, mean temperature data, humidity data, speed data and rainfall data of Chiang Mai province for the meteorological data which agreed with other publications. The program used for analysis was Weka. SMOTE was used to solve the imbalance data set which other publications did not consider. Multi-layer perceptron and Support vector machine was the algorithm for predicting the forest fire occurred.

CHAPTER III
MATERIALS AND METHODOLOGY

3.1 Research methodology

The objective of this research was to analysis and design a web application of the Bureau of Forest Fire Prevention and Control, Department of National Parks, Wildlife and Plant Conservation Ministry of Natural Resources and Environment. Another objective was to create a system that could predict the risk of forest fire system. The research methodology of this study was demonstrated in Figure 3.1 and Figure 3.2.

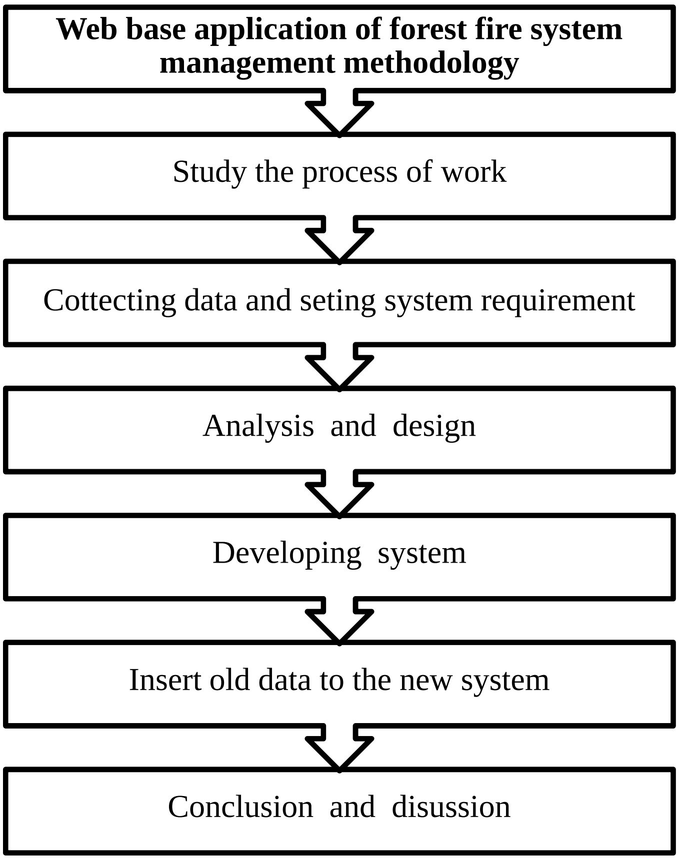


Figure 3.1 The developing process of a web base application for forest fire management.

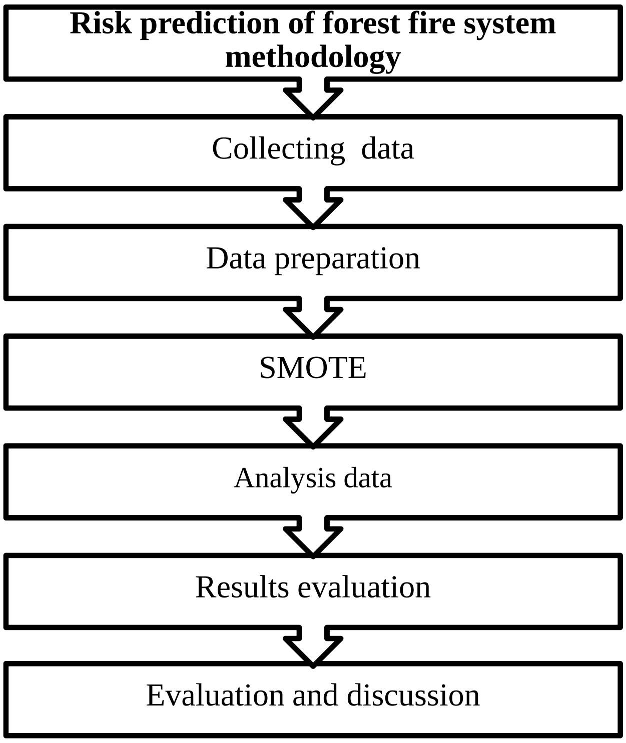


Figure 3.2 The developing process to predict the risk of forest fire system

3.2 Web base application of forest fire management methodology

3.2.1. Study of the process of work of the Bureau of Forest Fire Prevention and Control

In this study, the researcher will discover the process of work of the Bureau of Forest Fire Prevention and Control in Bangkok and interview the users to identify the problems of original system and determine the area of research.

3.2.2. Data Collection and determining system requirements

3.2.2.1. Collect the basic information that stored in original database including the area of forest fire data and causes of forest fire.

3.2.2.2. Study documents, reports and forms of the Bureau of Forest Fire Prevention and Control for determining the needs of users.

3.2.2.3. Interview the users of system for determining the requirements of users and studying the process of work of the Bureau of Forest Fire Prevention and Control.

3.2.3. Analysis and design

3.2.3.1. Analyze and develop by using user requirements.

3.2.3.2. Create entity relationship(ER diagram).

3.2.3.3. Create a database from entity relationship.

3.2.3.4. Design a user interface by using Adobe Dreamweaver CS3 program , Adobe flash and Adobe Illustrator CS3.

3.2.4. Developing the user interface

The data that was analyzed in step 3.2.3 are used in developing the system by using a PHP language for software development and using Mysql for database.

There were six pages of user interface which composed of:

1. **Login page** (Figure 3.3)

Users who want to use this system must login before use.

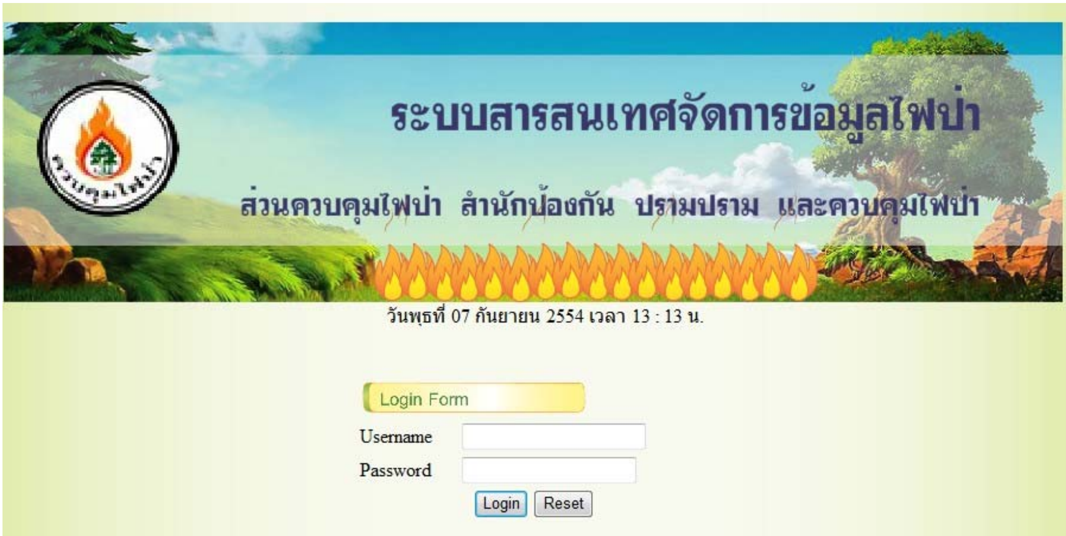


Figure 3.3 Show the login page.

2. Manage information (Figure 3.4**)**

In this page, users can add, edit, and delete the information of fire and the information of station. For example can add the information of fire by click the adding button or when users want to see more information of fire, user can click on the time that show on the table to get more information.

รหัสสถานี	วัน	เวลาที่แจ้ง	คนแจ้ง	คนรายงาน	เวลาเริ่มดับ	เวลาดับเสร็จ	พื้นที่ดับ	Update	Delete
31104	31 ม.ค. 2553	19:00:00	พบ ดวงแก้ว	สุพล	19:30:00	20:00:00	3.00	Update	Delete
31106	31 ม.ค. 2553	15:00:00	ลาด ตระเวน	เทพ สร้อยเรือน	15:40:00	17:30:00	3.00	Update	Delete
31100	31 ม.ค. 2553	13:25:00	อาทิตย์	อดิศักดิ์	14:00:00	14:44:00	2.00	Update	Delete
31104	31 ม.ค. 2553	11:20:00	สมบูรณ์ บัวดอกตูม	ประจวบ	11:30:00	12:40:00	5.00	Update	Delete
31101	30 ม.ค. 2553	17:51:00	นายเหลา นางเมาะ	ภัทรกุล	19:56:00	20:08:00	6.00	Update	Delete
31105	30 ม.ค. 2553	16:30:00	สุชาติ	ติกา	16:50:00	17:10:00	1.00	Update	Delete
31106	29 ม.ค. 2553	15:11:00	ปราโมทย์	ผัด	15:25:00	17:00:00	2.00	Update	Delete
31104	29 ม.ค. 2553	14:30:00	จำนง	ชำนาญ	14:59:00	15:29:00	4.00	Update	Delete
31104	29 ม.ค. 2553	13:34:00	กมล	กมล	13:44:00	13:56:00	2.00	Update	Delete

Figure 3.4 Show the manage information page.

3. Searching (Figure 3.5)

In this page, users can search the information of fire by using date of fire occurred and search the information of station by using the name of the station.

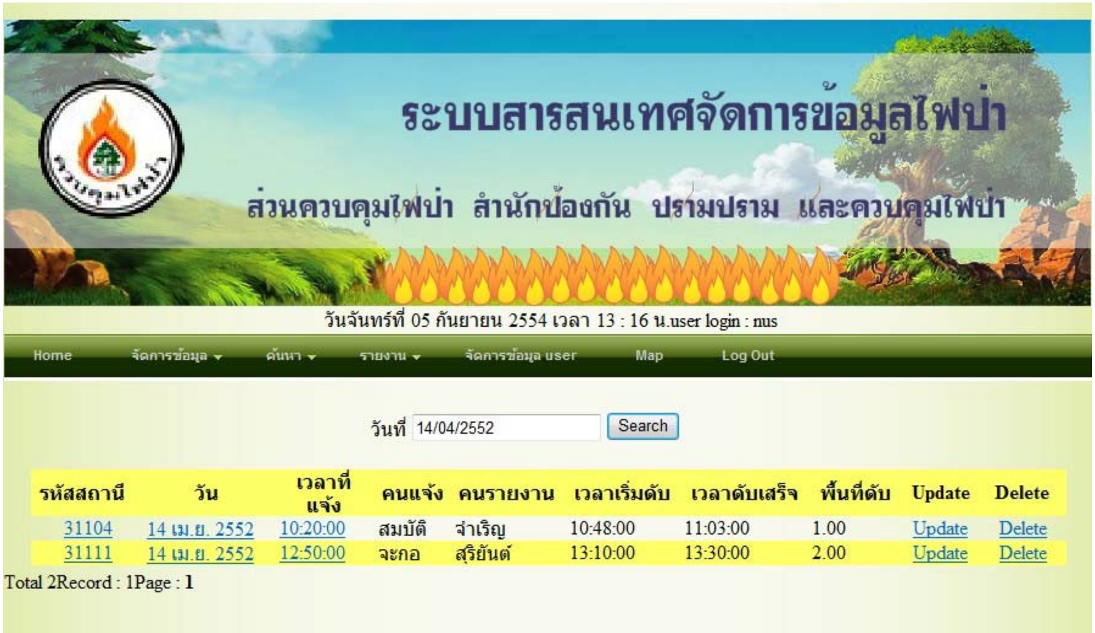


Figure 3.5 Show the searching page.

4. Reports (Figure 3.6)

In this page, user can create the reports which include

- The report of forest fire which in ff1 form briefly (ฟป.1 อย่างย่อ)
- The report of forest fire which in ff2 form (ฟป.2)
- The report of forest fire which in ff3 form (ฟป.3)
- The report of forest fire by station.
- The report of forest fire by Tumbon.
- The report of forest fire by Ampur.
- The report of forest fire by province.
- The report of forest fire by part of Thailand.
- The report of forest fire by month.
- The report of forest fire by province and type of forests.

- The report of forest fire by part of Thailand and type of forests.

The example of the report was shown in Figure 3.7

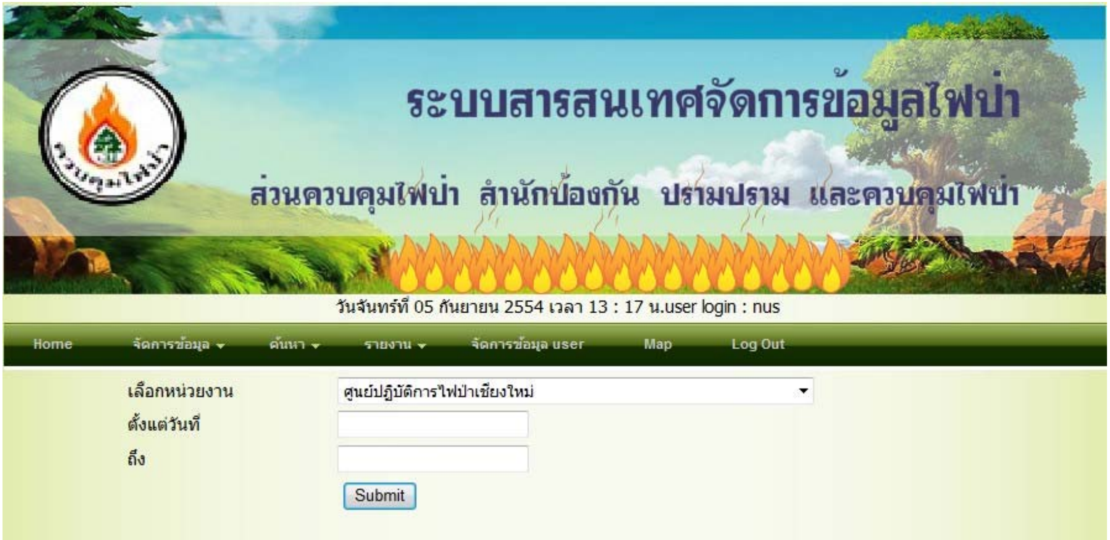


Figure 3.6 Show the reporting page.

รายงานการเกิดไฟป่าแยกอำเภอ

ตั้งแต่วันที่ 01 เดือน มกราคม 2544 ถึงวันที่ 01 เดือน มกราคม 2553

ศูนย์ปฏิบัติการไฟป่าเชียงใหม่เมืองเชียงใหม่จังหวัดเชียงใหม่

อำเภอ	จังหวัด	จำนวนครั้ง	จำนวนไร่
เมืองเชียงใหม่	เชียงใหม่	965	3884.13
เชียงดาว	เชียงใหม่	1544	9949.50
จอมทอง	เชียงใหม่	2706	25574.75
ดอยสะเก็ด	เชียงใหม่	1923	7710.75
ฝาง	เชียงใหม่	111	801.00
หางดง	เชียงใหม่	521	2280.50
ฮอด	เชียงใหม่	1342	6968.75
แม่แจ่ม	เชียงใหม่	1031	8698.00
แม่ริม	เชียงใหม่	509	2404.50
แม่แตง	เชียงใหม่	502	4719.50
อมก๋อย	เชียงใหม่	830	5321.50
พร้าว	เชียงใหม่	2	41.00
สะเมิง	เชียงใหม่	1393	10401.50
สันกำแพง	เชียงใหม่	187	723.00
สันป่าตอง	เชียงใหม่	6	53.00
สันทราย	เชียงใหม่	9	54.00
แม่อาย	เชียงใหม่	509	4186.75
ดอยเต่า	เชียงใหม่	118	732.00
เวียงแหง	เชียงใหม่	7	112.00
ไชยปราการ	เชียงใหม่	569	3552.25
แม่วาง	เชียงใหม่	65	1322.00
กิ่งอำเภอแม่ออน	เชียงใหม่	1698	5681.75
กิ่งอำเภอดอยหล่อ	เชียงใหม่	2	16.00
รวม		16597	105811.13

Figure 3.7 Show the example of the report of forest fire by Districts.

5. Manage users (Figure 3.8)

In this page, user's admin can add, edit, and delete the information of user from this system.

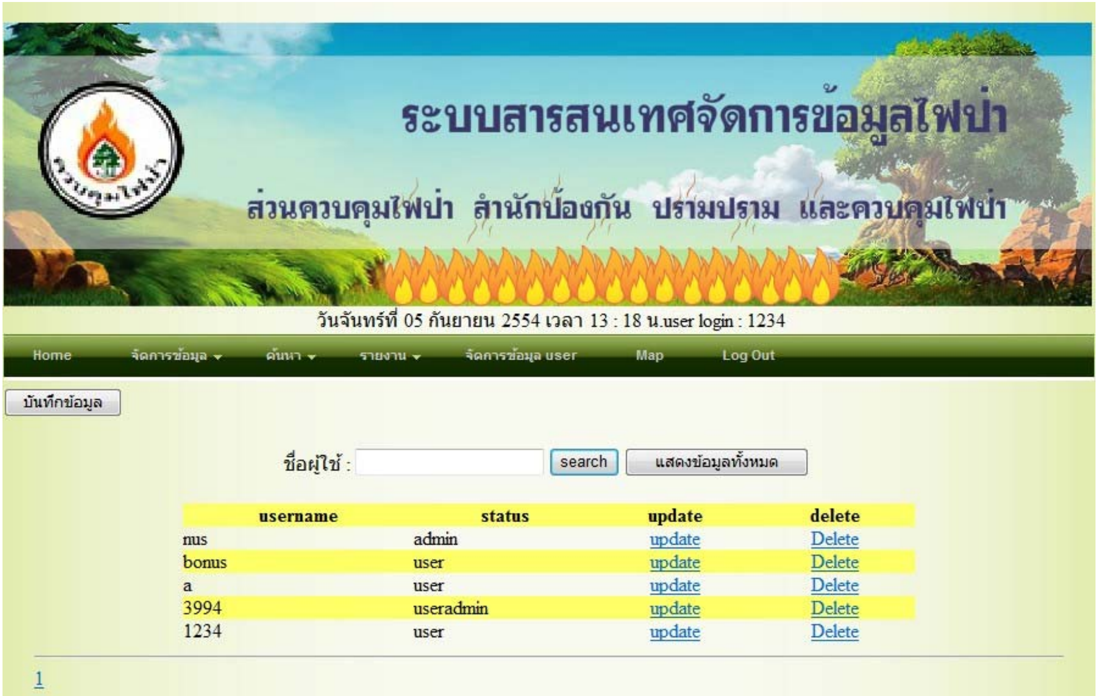


Figure 3.8 Show the user management page.

6. Map (Figure 3.9)

In this page, users can search the location of forest fire by using date of fire occurred which showed on the google map. The locations of forest fire on google map were created by.

1. Object of the Google Map was created by set the size of the map page with 800 * 600 pixels and set the center of google map location in the northern of Thailand by set the latitude – longitude coronate at 18.78856,98.98234.

2. The locations of forest fire were selected from database which set in the latitude – longitude coronate formats. Then, the location of the forest fire were generated to Marker in XML files.

3. Using AJAX to retrieve a PHP page using the XML in step 2 to activate JavaScript. Then, the locations of forest fire will appear in google map.

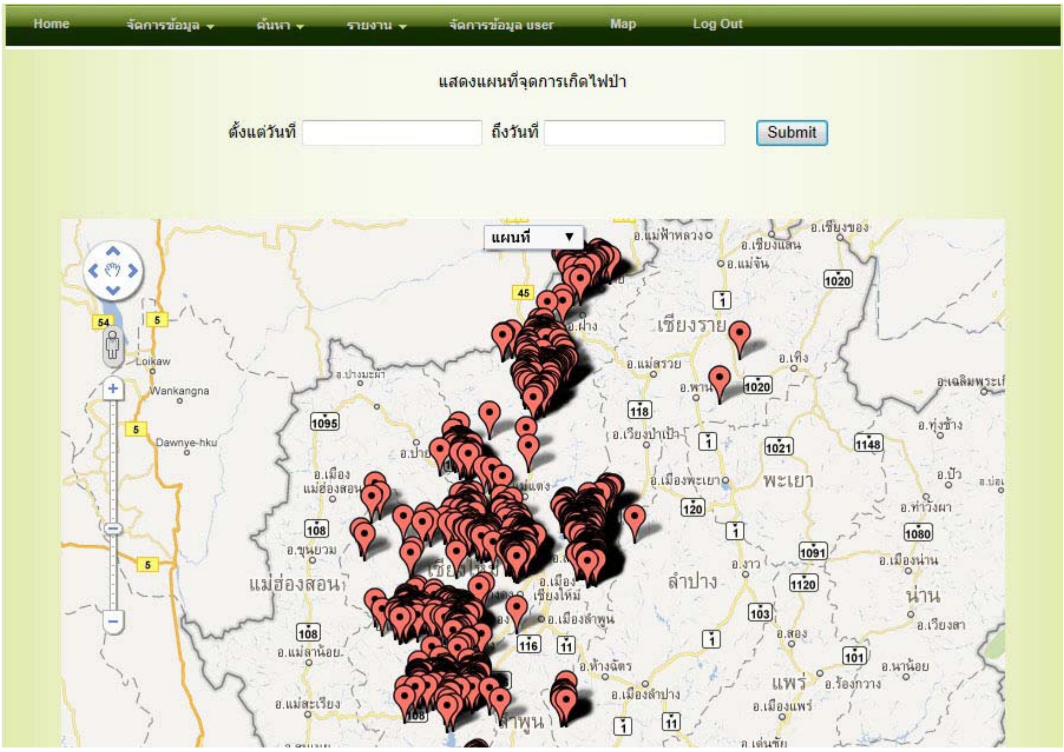


Figure 3.9 Show the map of forest fire occurred page.

The database in the new system will be normalized. The normalization was the process of efficiently organizing data in a database. The objectives of the normalization process were eliminating redundant data (for example, storing the same data in more than one table) and ensuring data dependencies make sense (only storing related data in a table). Both of these can reduce the amount of space and database will be faster. The new database was composed with five tables mainly. The main tables include Fire_data table, Station_datatable and Placefire_data table (Figure 3.10) and the descriptions of each table was shown in Table 3.1 , Table 3.2 and Table 3.3 respectively

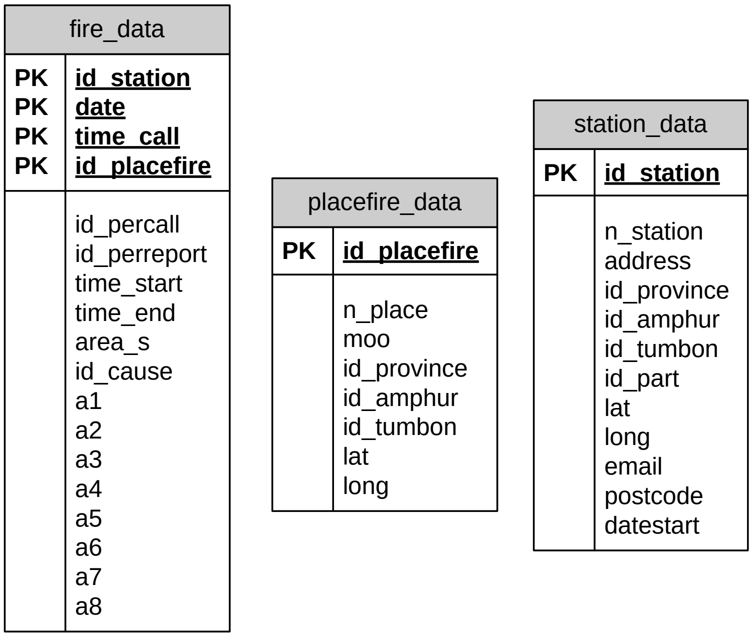


Figure 3.10 Show the main tables in the system.

- Fire_data table - record the information of forest fire occurred which had primary key included.

Table 3.1 Show the descriptions of Fire_data table

Primary key	Description
Id_station	Code of forest fire station
Date	Date of fire forest occurred
Time_call	Time of fire forest occurred
Id_placefire	Code of fire forest location

- Station_data table - record the information of Forest Fire Prevention and Control station which had primary key included.

Table 3.2 Show the descriptions of Station_data table

Primary key	**Description**
Id_station	Code of forest fire station

- Placefire_data table - record the information of forest fire location which had primary key included

Table 3.3 Show the descriptions of Placefire_data table

Primary key	**Description**
Id_placefire	Code of fire forest location

3.2.5. Convert the data from the old system to the new system

When the system development was complete, the next step was to insert the data from the old database to the new database. The wear in the following steps:

First step was converting the UTM coordinate data into latitude-longitude by using the UTMConversions which was the spreadsheet converting program. There were three sheets in this spreadsheet included: (Figure 3.11)

Main page – Displays the datum selections and conversion windows. Users could convert the latitude-longitude to UTM coordinate or convert the UTM coordinate to the latitude-longitude.

Batch Convert UTM to Lat-Long – Users could convert the UTM coordinate to the latitude-longitude coordinate in large quantities of data.

Batch Convert Lat Long to UTM – Users could convert the latitude-longitude to UTM coordinate in large quantities of data.

In this studied, the researcher used the Batch Convert UTM to Lat-Long sheet for converting data. Before converting data, preparing data by copying the data in Microsoft access file to Microsoft excel and use function text to columns for separating the UTM coordinate into two columns because in the old database, the UTM coordinate was inserted in one column. After that, inserting the data to UTMConversions and put UTM coordinate in northing and easting columns. Then, latitude and longitude wear given in dd.dddd format in columns AB and AC. When finish converting, insert the data into MySQL database in fire_data table.

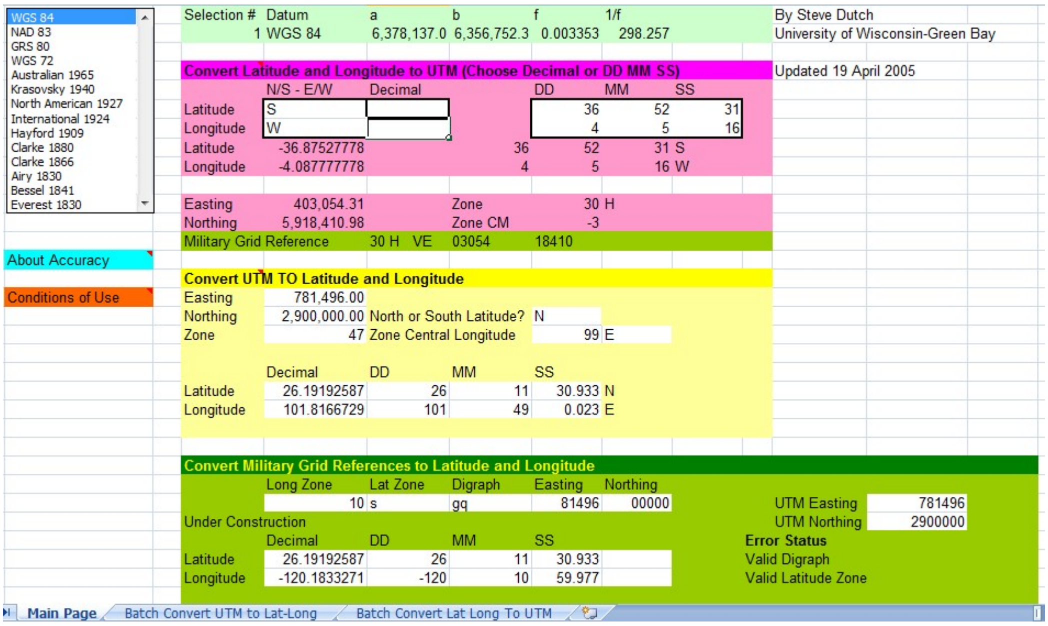


Figure 3.11 The UTMConversions program.

The next steps were to convert some of fire location data including provinces, amphurs, tumbons into a codename in Figure 3.4by copy the data in Microsoft access file to Microsoft excel and created the converting function in Microsoft visual basic. The code name was related with province table, amphur table and tumbon table. The objective of this converting was eliminating redundant data. Then, insert the new data that converted into placefire_data table in MySQL database. The placefire_data table was shown in Figure 3.12.

Location	Moo	District	Amphur	Province
ทิศเหนือเขื่อนแม่กวง	1	ลวงเหนือ	ดอยสะเก็ด	เชียงใหม่
ห้วยแม่ลอง	1	หางดง	ฮอด	เชียงใหม่
ดอยเหล็ก	3	แม่เหียะ	เมือง	เชียงใหม่
บ้านวังธาร ดงป่าก่อ	8	ลวงเหนือ	ดอยสะเก็ด	เชียงใหม่
หลังเทคนิคบนสันเขา	1	สหกรณ์	แม่ออน	เชียงใหม่

n_place	moo	id_province	id_amphur	id_tumbon
ทิศเหนือเขื่อนแม่กวง	1	171	17104	1710409
ห้วยแม่ลอง	1	171	17107	1710701
ดอยเหล็ก	3	171	17101	1710101
บ้านวังธาร ดงป่าก่อ	8	171	17104	1710409
หลังเทคนิคบนสันเขา	1	171	17123	1712306

Figure 3.12 Placefire_data table.

Finally, the data in Microsoft Access (the database in old system) would be converted to MySQL database. The data that added to new database had 16000 records approximately which start from 1 January 2001 to 31 January 2010.

3.2.6. Evaluation and discussion.

Evaluate the forest fire management system by using questionnaires as a tool to ask and sample users who use this system. The result of system evaluation by measuring user’s satisfaction level which were five levels of evaluation; poor (1.00 – 1.49), fair (1.50 – 2.49), good (2.50 – 3.49), fine (3.50 – 4.49) and excellent (4.50 – 5.00).The results of the questionnaire were discussed for leaning the advantages and disadvantage of this system for development in the future.

3.3 Risk prediction of forest fire system methodology

3.3.1. Collection data

The data used in this study is secondary data which collected information from relevant agencies including:

3.3.1.1 The daily forest fire data from year 2001 to 2010 from the Office of Protected Area Administrative 16 in Chiangmai province which composing 12 stations following : Chiangmai, Maeon, Huaynamdang, Maeping, Maekuang, Inthanon, Booluang, Banleknaiparyai, Damri, Omkoi, Samerng and Chaiprakarn.

3.3.1.2 The daily climate data from Thai Meteorological Department which including maximum temperature data(^{o}C), minimum temperature data(^{o}C), mean temperature data(^{o}C), humidity data(%), wind speed data(knots) and rainfall data(mm.) of Chiangmai province from year 2001 to 2010.

3.3.2. Data preparation

3.3.2.1. Merge the daily climate data including maximum temperature data, minimum temperature data, mean temperature data, humidity data, wind speed data and rainfall data of Chiang Mai province from year 2001 to 2010 into one table because the daily climate data come from different file source.

3.3.2.2. Counting the number of forest fire occurred and forest fire station from database (Figure 3.13) by using SQL statement. Then , convert the number of forest fire occurred into five numbers include

(5) Very high - had the number of forest fire occurred > 21 per days.

(4) High - had the number of forest fire occurred between 14 and 21 per day.

(3) Normal - had the number of forest fire occurred between 7 and 14 per days.

(2) Low - had the number of forest fire occurred between 14 and 7 per days.

(1) None - had not the number of forest fire occurred.

Table 3.4 The result of level the number of forest fire occurred

station	date	fire
31102	3/3/2001	2
31102	5/2/2001	1
31102	9/2/2001	3
31102	12/2/2001	1
31102	14/2/2001	1
31102	20/2/2001	1
31102	21/2/2001	1
31102	24/2/2001	1
31102	25/2/2001	1

3.3.2.3 Merge the fire data in 3.3.2.2 and climate data in 3.3.2.1 into one table by using SQL statement.

3.3.2.4 Replace missing data. A few missing data was found in wind speed data so the researcher will replace missing data manually. The average between previous wind speed and next wind speed will be used to replace the missing data.

3.3.2.5 In the experiment, the training data set was data from year 2001 to 2008 and the testing data sets were data from year 2009 to 2010. This experiment would be only used for the fire season months of January, February, March, April and December. Since fire occurrence is only found in these months and to remove abundance of unnecessary data. All the data sets were converted to .arff format to be used by Weka. Then, the preprocessed data sets would be resampling with SMOTE. The preprocessed data sets attributes are shown in Table 3.5. The preprocessed data sets in this experiment consisted of 8 attributes. The attributes including mintemp, rate, rh, meantemp, wind, max, place and rate. The instances of this experiment were 14520 composing:

- None class - 9369 instances
- Low class - 2844 instances
- Medium class - 1321 instances
- High class - 568 instances
- Very high class - 418 instances

Table 3.5 The preprocessed data sets attributes.

Attribute	Description
Min	minimum temperature data(^{o}C)
Rain	rainfall data(mm.)
RH	humidity data (%)
Meantemp	mean temperature data(^{o}C)
Wind	wind speed data(knots)
Max	maximum temperature data(^{o}C),
Place	The forest fire occurrence area.
Rate	The rate of forest fire occurrence area.

The examples of data sets of this experiment are shown in Table 3.6

Table 3.6 The example of data sets of this experiment

min	rain	rh	mean temp	wind	max	place	rate
16.2	0	72	22.5	6	28.9	chiangmai	None
15.1	0	72	22	6	29	chiangmai	None
15	0	73	22.1	4	29.2	chiangmai	Low
15.9	0	73	22.8	7	29.7	chiangmai	None
16	0	74	22.7	11	29.5	chiangmai	None
15.6	0	72	22.8	7	30	chiangmai	None

The first row in Table 3.6 demonstrates the daily data which composing the minimal temperature is 16.2 ^{o}C, the rain fall is 0 (no raining), the humidity is 72%, the mean temperature is 22 ^{o}C, the wind speed is 6 knots, the maximum temperature is 28.9 ^{o}C, the place where fire occurrence is in Mueang District, Chiangmai province and the rate of forest fire occurrence is none (no fire occurrence).

3.3.3 Solving the imbalanced data sets problem

In the experiment, the data set is imbalance data because there are the none class a lot larger than other classes (the none class is 64.52 % of total instances).

Therefore, the SMOTE technique was used to re-sampling data sets before analysis with data mining algorithm. The SMOTE technique was embedded in the Weka :weka.fiters.supervised.instance.SMOTE.

3.3.4 Data analysis

After preparing data in steps 3.3.2, select the data in weka program by using .arff file. The data for training model was data from year 2001 to 2008. The input data were maximum temperature data, minimum temperature data, mean temperature data, humidity data, speed data, rainfall data and forest fire station. The output data was rate of forest fire occurred. Then, the data was analysis by using multi-layer perceptron (MLP) and support vector machine (SVM) for training. Cross-validation measurement models with 10 folds would be used. The information was divided into ten parts by one part to store information for using in the experiment and the remaining nine parts were used in testing the model. When finished, it would be used the part that has not been tested in the next test. Then, repeating these steps until all the data is tested. Finally, the complete model was gotten.

The importance parameters of MLP algorithm in this study are hidden layer = 7, learning rate = 0.3 and momentum = 0.2. The importance parameters of SVM algorithm in this study are degree = 3, gamma = 0 and kernel is radial basis function (Gaussian kernel). The parameters that can adjust of MLP algorithm and SVM algorithm were shown in Figure 3.13 and 3.14 respectively.

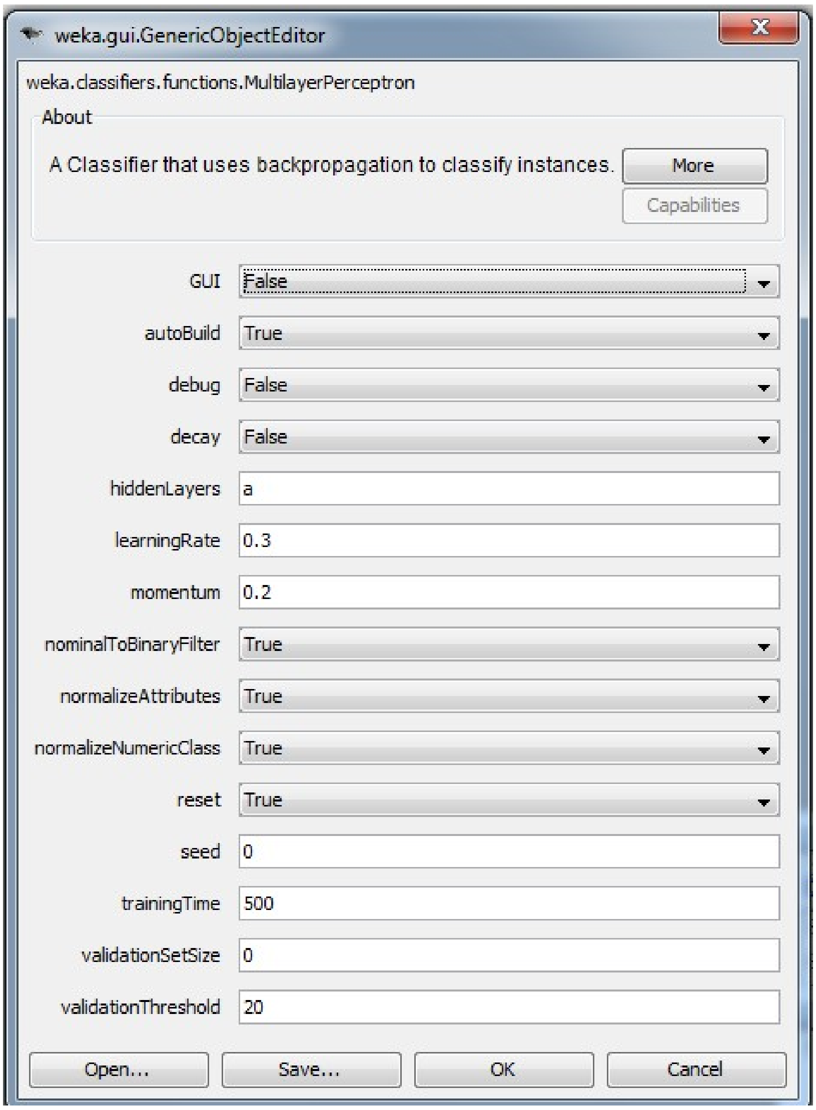


Figure 3.13 The parameters of MLP algorithm. (the hidden layers are calculated from (number of attributes + number of classed) / 2 so in this study the hidden layers are seven.)

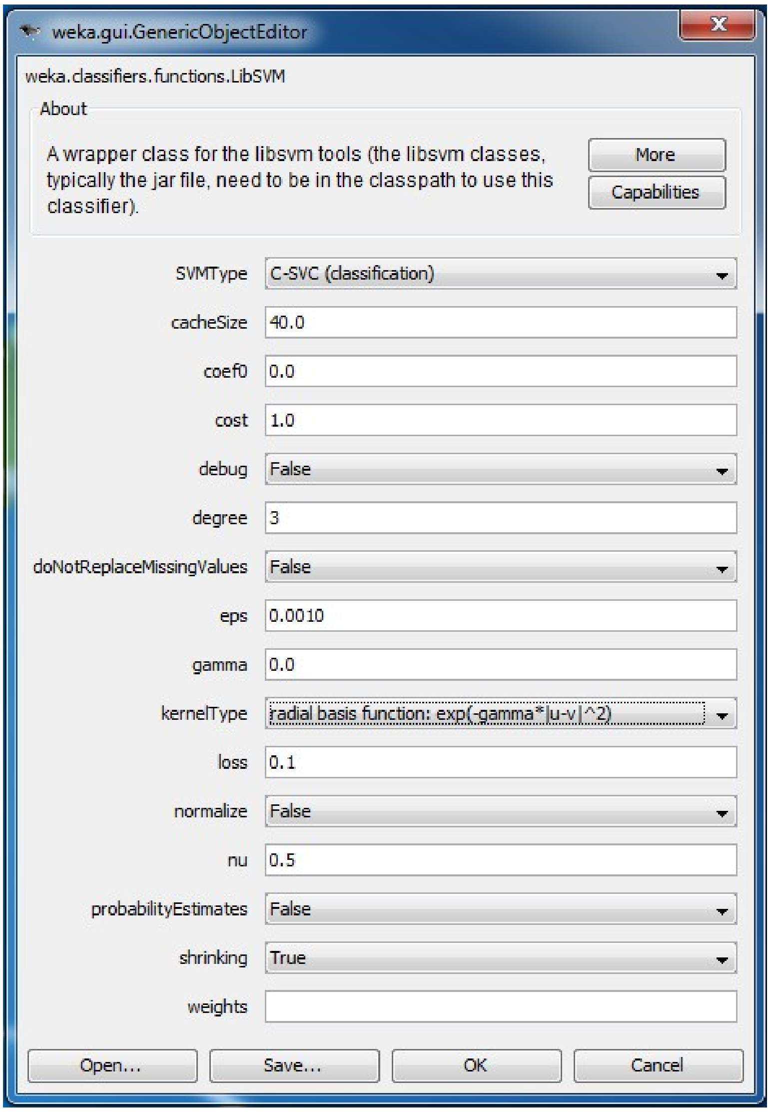


Figure 3.14 The parameters of SVM algorithm.

3.3.5 Results evaluation

To evaluate the performance of experiments, the testing set would be tested by the data from 2008 to 2010. Finally, the results of data sets with no re-sampling and the data sets with re-sampling by SMOTE will be compared.

3.3.6 Conclusion and discussion

The best predicting results the risk of fire is found, the performance of multi-layer perceptron and supporting vector machines will be compared. The results of each experiment will be presented.

3.4 Materials and Development Tools

3.4.1 Hardware

Computer Name:	:	Lenovo y450
Processor:	:	Intel Core 2 Duo P8700 / 2.53 GHz
RAM	:	DDRIII 8 GB
Hard Drive:	:	500 GB

3.4.2 Software

DBMS	:	My SQL
System development language	:	PHP
Application development	:	Adobe Photoshop, Adobe Illustrator Adobe Dreamweaver, Weka
Operating System	:	Microsoft Window 7
Web browser	:	Microsoft Internet Explorer version 9

3.5 Schedule of research

Table 3.7 Schedule of research

Schedule of research	2011-2012			
	Dec.	Jan.	Feb.	Mar.
1. Planning and collecting the information				
2. Analysis the information.				
3. Developing web base application of forest fire management system.				
4. Developing risk prediction of forest fire system.				
5. Analyze the risk of forest fire.				
6. Evaluation of result the risk of forest fire				
7. Testing and documenting.				

CHAPTER IV
RESULTS AND DISCUSSION

The result of forest fire management and risk prediction system in Chiangmai province: a comparative study of support vector machine and neural network are divided into two parts. The first part is the result and discussion of forest fire management system and the second part is the result and discussion of forest fire risk prediction. The data used in this study come from year 2001-2010.

4.1 The result of forest fire management system

The forest fire management system was developed by using a PHP language for coding the application and using MySQL as a database. After the development of the system was completed, system testing was done for debugging error, and system evaluation was done for system performance. The users from Bureau of Forest Fire prevention and controls and the Director of Information and Technology Office were the target group in the evaluation as they are most likely to adopt this application, thus they is well suited for evaluating the importance of presented perspectives. The questionnaire was aimed to investigate the quality and the usability of the system by questions from 5 users. The result of system evaluation by measuring users satisfaction level was divided into four parts; functional requirement test, functional test, usability test and security test. Each part divided into 5 levels as described in Table 4.1 to 4.4 the level ranges 1.00 – 1.49 is poor, 1.50 – 2.49 is fair, 2.50 – 3.49 is good, 3.50 – 4.49 is fine and 4.50 – 5.00 is excellent.

Table 4.1 The results of questionnaire for functional requirement test

Functional requirement test	Result
1. Ability to present information	4.6
2. Ability to search	4.2
3. Ability to manage the database	4.4
4. Ability to categorize data	4.2
Mean	**4.35**

Table 4.1 shows the results of questionnaire for functional requirement test found that the overall satisfactions of functional requirement test from users in the system are fine(4.35) which is divided to;

Ability to present information from users in the system are excellent (4.6).

Ability to search from users in the system are fine (4.2).

Ability to manage the database from users in the system are fine (4.4).

Ability to categorize data from users in the system are fine (4.2).

Table 4.2 The results of questionnaire for functional test

Functional Test	Result
1. Accuracy of the information presentation	4.8
2. Accuracy in searching	4.6
3. Accuracy in the database	4.2
4. Accuracy in classifying	4.4
Mean	**4.5**

Table 4.2 shows the results of questionnaire for functional test found that the overall satisfactions of functional test from users in the system are excellent (4.5) which is divided to;

Accuracy of the information presentation from users in the system are excellent (4.8).

Accuracy in searching from users in the system are excellent (4.6).

Accuracy in the database from users in the system are fine (4.2).

Accuracy in classifying from expert from users in the system are fine (4.4).

Table 4.3 The results of questionnaire for usability test

Usability Test	**Result**
1. Suitability of information presented	4.6
2. Clarity of the text display	4.2
3.Ease of system use	4.6
4. Ability of the system in overview	4.6
5. System satisfaction	4.6
Mean	**4.52**

Table 4.3 shows the results of questionnaire for usability test found that overall satisfaction of usability test form users in the system are excellent (4.52) which is divided to;

Suitability of information presented from expert from users in the system are excellent (4.6).

Clarity of the text display from users in the system are fine (4.2).

Ease of system use from users in the system are excellent (4.6).

Ability of the system in overview from users in the system are excellent (4.6).

System satisfactions from users in the system are excellent (4.6).

Table 4.4 The results of questionnaire for security test

Security Test	**Result**
1. Permission to access	4.2
2. Ability to access for editing data	4.2
3. Permission to update system administrator	4.4
4. Ability to categorize data	4.2
Mean	**4.25**

Table 4.4 shows the results of questionnaire for security test found that overall satisfaction security test from users in the system are fine (4.25) which is divided to;

Permission to access from expert from users in the system are fine (4.2).
Ability to access for editing data from users in the system are fine (4.2).
Permission to update system administrator from users in the system are fine (4.4).
Ability to categorize data from users in the system are fine (4.2).

4.2 Discussion the forest fire management System

The new forest fire system makes the working process more conveniently because it is a web-based application. The users in this system do not record the data in Access database which is not well designed so the information in this database can be inconsistent and duplicate. Moreover, the forest fire occurrence on the google map are improve efficiently on the forest fire prevent and control planning because the staff can see the forest fire in overview, easy to make a decision.

From the system evaluation, the score from users is quite high in overall. The result of functional requirement test and security test from users in the system are fine. The result of functional test and usability test are excellent.

4.3 The result of forest fire risk prediction

After finish forest fire management system, the data in this system will be used for predicting forest fire which the results are following.

Because of classification with the class imbalance problem such as our problem with has more 'no fire' data compared to 'fire occurrence data', accuracy is no longer a proper measure even though accuracy is more common than other measurement. Since the minority class has very little impact on accuracy as compared to the prevalent class [23]. In Table 4.6, the most accuracy is 81.09% but most data are predicted to be ‘none’ and ‘low’ fire occurrence. A few more is in a ‘medium’ class. However, almost none is predicted to be ‘high/very high’ class. Given that it is very unlikely to correctly predict the fire occurrence. Therefore, this experiment would use the average recall to measure performance of the model.

Table 4.5 A confusion matrix for a two-class classification.

	Predicted as Positive	Predicted as Negative
Actual Positive class	True Positive(TP)	False Negative(FN)
Actual Negative class	False Positive(FP)	True Negative(TN)

Table 4.5 illustrates a confusion matrix of a two-class problem. The first column of the table is the actual class label of the examples, and the first row presents their predicted class label. TP and TN denote the number of positive and negative examples that are classified correctly, while FN and FP denoted the number of misclassified positive and negative examples respectively.

Overall accuracy is defined as the proportion of the all correct prediction against all the instances. Meanwhile, the recall is defined as the proportion of the correct prediction of this class against all the instances of this class.

The overall accuracy and recall are calculated by using a confusion matrix in Table 4.5 and equation (4.1), (4.2) respectively.

$$\text{Overall accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{4.1}$$

$$\text{Recall} = \frac{y}{z} \text{ or } \frac{TP}{TP+FN} \tag{4.2}$$

Where y is the number of correct prediction in each class.

z is the overall number of prediction in each class.

Table 4.6 The confusion matrix of SVM

Actual \ Predicted	None	Low	Medium	High	Very high	Recall	Accuracy
None	1551	56	1	0	0	0.97	81.09 %
Low	164	195	9	0	0	0.53	
Medium	41	65	24	0	0	0.19	
High	10	38	8	1	0	0.02	
Very high	4	13	4	0	0	0	
Average						0.34	

Table 4.6 shows the confusion matrix of SVM model. The overall accuracy of this model is 81.09 %. This model predicts excellent in none class which predict correctly 1551 instances out of 1608 total none class instances (96.5%). In low class, this model can predict fairly which predict correctly 195 instances out of 368 total low class instance (53%). On the other hand, this model can predict poorly in medium, high and very high class. The model predict correctly 24 out of 130 (18.5%), 1 out of 57 (1.8%) in medium and high class respectively. Moreover, this model never predict correctly in very high class. The recall of none class is the highest while the very high class has the lowest recall (no recall that means this model never predict correctly in very high class). The average recall of this model is 0.34.

Table 4.7 The confusion matrix of SVM-SMOTE

Actual \ Predicted	None	Low	Medium	High	Very high	Recall	Accuracy
None	**1282**	182	78	42	24	0.80	75.60 %
Low	45	**222**	47	41	13	0.60	
Medium	7	13	**88**	13	9	0.68	
High	1	3	5	**44**	4	0.77	
Very high	0	3	1	2	**15**	0.71	
Average						**0.71**	

Table 4.7 shows the confusion matrix of SVM using the SMOTE model for training. The overall accuracy of this model is 75.60 %. This model predicts well in every class which predict correctly 1282 instances out of 1608 total none class instances (79.7%) in none class. In low class, this model can predict correctly 222 instances out of 368 total low class instance (60.3%). In medium class, this model can predict correctly 88 instances out of 130 total medium class instance (67.7%). In high class, this model can predict correctly 44 instances out of 57 total high class instance (77.2%). Finally, in very high class, this model can predict correctly 15 instances out of 21 total very high class instance (71.4%). The recall of none class is the highest and the low class has the lowest recall. The average recall of this model is 0.71.

Table 4.8 The confusion matrix of MLP

Predicted / **Actual**	**None**	**Low**	**Medium**	**High**	**Very high**	**Recall**	**Accuracy**
None	**1489**	112	7	0	0	0.93	76.14 %
Low	185	**157**	26	0	0	0.43	
Medium	32	81	**17**	0	0	0.13	
High	8	31	18	**0**	0	0	
Very high	1	2	18	0	**0**	0	
Average						**0.30**	

Table 4.8 shows the confusion matrix of MLP model. The overall accuracy of this model is 76.14 %. Similar to the SVM model, MLP predicts excellent in none class which predict correctly 1489 instances out of 1608 total none class instances (93%). In low class, this model can predict correctly 157 instances out of 368 total low class instance (43%). On the other hand, this model can predict poorly in medium, high and very high class. The model predict correctly 17 out of 130 (13%) in medium class. In high and very high class, this model cannot predict correctly. The recall of none class was the highest and the high and very high class have no recall. The average recall of this model is 0.30.

Table 4.9 The confusion matrix of MLP-SMOTE

Predicted / Actual	None	Low	Medium	High	Very high	Recall	Accuracy
None	**1032**	420	13	55	88	0.64	56.14 %
Low	40	**161**	21	59	87	0.44	
Medium	6	35	**7**	29	53	0.05	
High	1	9	3	**9**	35	0.16	
Very high	0	2	0	2	**17**	0.81	
Average						**0.42**	

Table 4.9 shows the confusion matrix of MLP-SMOTE model. The overall accuracy of this model is 56.14 %. This model predict correctly 1032 instances out of 1608 total none class instances (64%) in none class. In low class, this model can predict correctly 161 instances out of 368 total low class instance (44%). In medium class, this model can predict correctly 7 instances out of 130 total medium class instance (5%).In high class, this model can predict correctly 9 instances out of 57 total high class instance (16%). Finally, in very high class, this model can predict correctly 17 instances out of 21 total very high class instance (81%). The recall of very high class is the highest and the high class has the lowest recall. The average recall of this model is 0.42

4.4 The discussion of forest fire risk prediction

The data in this study is imbalance data. Since data is none fire occurrence class is larger than other classes (64.5% out of overall instances). To solve this problem, SMOTE is used for improving performance of the model. Moreover, the average recall is used for evaluating the performance of the experiments. The recall is more appropriate because if the accuracy is used the correct prediction on 'none' class will have higher significance than other classes. As a result, even though very few 'high', 'very high' can be correctly predicted, the accuracy of 'none' still weigh the overall accuracy up. Moreover, because of the data in this study is imbalance data which none class instances is 64.5% of overall instances so the model will aim to predict highly accuracy in none class but the other classes can predict with poor accuracy. Therefore, the objectives of the study have to predict fire occurrence better than predict the no occurrence of fire. The SMOTE technique is used to balance this problem. The comparison between the model which using SMOTE and do not use are shown in Table 4.10 and 4.11.

Table 4.10 The comparative confusion matrix between SVM and SVM-SMOTE

Predicted / Actual	None		Low		Medium		High		Very high		Recall		Accuracy	
	SVM	SVM - SMOTE	SVM	SVM - SMOTE	SVM	SVM - SMOTE	SVM	SVM - SMOTE	SVM	SVM - SMOTE	SVM	SVM - SMOTE	SVM	SVM - SMOTE
None	**1551**	**1282**	56	182	1	78	0	42	0	24	0.97	0.80	81.09 %	75.60 %
Low	164	45	**195**	**222**	9	47	0	41	0	13	0.53	0.60		
Medium	41	7	65	13	**24**	**88**	0	13	0	9	0.19	0.68		
High	10	1	38	3	8	5	**1**	**44**	0	4	0.02	0.77		
Very high	4	0	13	3	4	1	0	2	**0**	**15**	0	0.71		
Average											**0.34**	**0.71**		

When using SMOTE to solve imbalance problem, the SVM model can predict better than when SMOTE is not used. Although, the overall accuracy is decrease from 81.09% to 75.60% but the average recall has increased from 0.34 to 0.71. Table 4.10 shows the comparative result between SVM and SVM-SMOTE. SMOTE improve the result in every classes expect none class which recall decrease from 0.97 to 0.80. In low class, the recall slightly improves from 0.53 to 0.60. In medium, high and very high class, the recall significant improves from 0.16 to 0.68, 0.02 to 0.77 and 0 to 0.71 respectively.

Table 4.11 The comparative confusion matrix between MLP and MLP-SMOTE

Predicted / Actual	None		Low		Medium		High		Very high		Recall		Accuracy	
	MLP	MLP-SMOTE	MLP	MLP-SMOTE	MLP	MLP-SMOTE	MLP	MLP-SMOTE	MLP	MLP-SMOTE	MLP	MLP-SMOTE	MLP	MLP-SMOTE
None	**1489**	**1032**	112	420	7	13	0	55	0	88	0.93	0.64		
Low	185	40	**157**	**161**	26	21	0	59	0	87	0.43	0.44		
Medium	32	6	81	35	**17**	**7**	0	29	0	53	0.13	0.05	76.14 %	56.14 %
High	8	1	31	9	18	3	**0**	**9**	0	35	0	0.16		
Very high	1	0	2	2	18	0	0	2	**0**	**17**	0	0.81		
Average											0.30	0.42		

When using SMOTE to solve imbalance problem, the MLP model can predict better than when does not using SMOTE. Although, the overall accuracy is drop from 76.14% to 56.14% but the average recall is increase from 0.30 to 0.42. Table 4.11 shows the comparative result between MLP and MLP-SMOTE. SMOTE improve the result in every classes expect none class which recall decrease from 0.97 to 0.80 and medium class which recall decrease from 0.13 to 0.05. In low class, the recall slightly improves from 0.43 to 0.44. In high and very high class, the recall improves from 0 to 0.16 and 0 to 0.81 respectively.

Table 4.12 The prediction results of SVM and SVM with SMOTE

Class / Technique	recall: none	low	medium	high	very high	Avg. recall	Overall accuracy
SVM	0.97	0.53	0.19	0.02	0	0.34	81.09%
SVM with SMOTE	0.80	0.60	0.68	0.77	0.71	0.71	75.60%
MLP	0.93	0.43	0.13	0.00	0.00	0.30	76.14%
MLP with SMOTE	0.64	0.44	0.05	0.16	0.81	0.42	56.14 %

The result in Table 4.12 shows the recall of each class, SVM and MLP model can predict with high accuracy in the none class but the other classes can predict with poor accuracy. The average recall is 0.34 and 0.30 respectively. When uses SMOTE with SVM and MLP, the result significantly improves in SVM with SMOTE. The average recall improves from 0.34 to 0.71. On the other hand, the results improve slightly in MLP with SMOTE. The average recall has improved from 0.30 to 0.42. SVM with SMOTE improves the recall a lot more significantly.

CHAPTER V
CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

This study has an objective to design and develop a forest fire system for creating reports and showing forest fire places on a google map and to compare the data mining algorithm for predicting forest fire occurrence based on six metrological data.

During the designed and development phase of a forest fire system, the Bureau of Forest Fire prevention and control process were intensively studied and interview the users to identify the problems. Then, the system was developed by PHP language, and the interface designed by Macromedia Dreamweaver CS. MySQL was the database that use in this system. The new system was a web based application which user in the region could report the forest fire occurrences directly to agency in Bangkok. Moreover, The new system can create reports which including : the report of forest fire 1 briefly, forest fire 2, forest fire 3, by station, by Tumbon, by Ampur, by province, by part of Thailand, by month, by province and type of forests and by part of Thailand and type of forests. The map of forest fire occurrences on google map was newly function in new system which user could see the forest fire occurrence spots on the google map.

Therefore, the new system made the working process more conveniently, reduced the duplicate working. For the evaluation process of the forest fire system, it was done by the users who work in the Bureau of Forest Fire prevention and control. After using the system, the users gave the optimal average scores for evaluation which the level of functional requirement test and security test were fine (4.35 and 4.25 out of 5 scores respectively), the functional test, usability test were excellent (4.5 and 4.52 out of 5 scores respectively).

In the forest fire risk prediction, used the forest fire data from year 2001 to 2010 which were composed of 12 forest fire prevention and control stations;

Chiangmai, Maeon, Huaynamdang, Maeping, Maekuang, Inthanon, Booluang, Banleknaiparyai, Damri, Omkoi, Samerng and Chaiprakarn where was recorded every day. Moreover, maximum temperature, minimum temperature, mean temperature, humidity data, wind speed data and rainfall were also used for the prediction. The data in this study were imbalance because fire occurrence is obviously not as often occurrence as no fire occurrence. The imbalance data problem was a significant factor, very few publications had mentioned about this problem. To solve this problem, SMOTE was used in this study. The result demonstrates that SMOTE can improve the performance of the model to better prediction of tendency of high risks of fire . However, the comparison study between the results of two algorithms showed that SVM with SMOTE outperformed MLP with SMOTE. Thus, the best way to predict forest fire occurrence was the support vector machine with SMOTE.

5.2 Recommendations for forest fire system.

The following recommendations were recommendations for improvement in further study is below:

1. The forest fire occurrence on google map should have more information such as the station areas.
2. The report system should customized a newly report for add into the system.

5.3 Recommendations for forest fire prediction.

The result of forest fire prediction cannot show on the map and do not have user interface for predicting. To ease of use, in the future should have user interface which users can see the risk of forest fire occurrences on the map.

Support vector machine (SVM) is a machine learning technique. The SVM algorithm are the parameters which can be adjusted, for example, C, ε, gamma. So the experiments should be performed using parameters adjustment to find the more accuracy predicting

Multilayer perceptron is a neural network algorithm. To use parameters adjustment can be advantage too, for example, the adjustment of the number of hidden layers, the number of nodes in hidden layers, the learning rate and the training time.

Moreover, a better method to improve the prediction performance may try the new techniques suggested, such as RBF Neural Network.

REFERENCES

1 กรกนก วชิโรภาสนันท์. การกำหนดพื้นที่เสี่ยงภัยต่อการเกิดไฟป่า บริเวณเขตรักษาพันธุ์สัตว์ป่าห้วยขาแข้ง อุทัยธานี. (วิทยานิพนธ์). นครปฐม : มหาวิทยาลัยมหิดล; 2542.

2 Brown Arthur Allen and K.P. David. Forest Fire: Control and Use. New York : Mc Graw – Hill; 1973.

3 ชนะชัย เลิศสุชาตวณิช. ดัชนีไฟเพื่อการจัดการสิ่งแวดล้อมในพื้นที่ป่าเต็งรัง กรณีศึกษา : อุทยานแห่งชาติสุเทพ-ปุย จังหวัดเชียงใหม่. (วิทยานิพนธ์). นครปฐม : มหาวิทยาลัยมหิดล; 2538.

4 อภินันท์ ปลอดเปลี่ยว และคณะ. แนวทางการควบคุมไฟป่าในประเทศไทย. กรุงเทพฯ : ฝ่ายวิชาการและแผนงาน. สำนักงานช่วยเหลือผู้ประสบภัยธรรมชาติ.กรมป่าไม้; 2536.

5 Noriega Leonardo. Multilayer perceptron tutorial [serial on the Internet]. School of Computing Staffordshire University; 2005 [cited 2011 Jun 3]. Available from: http://www.cs.sun.ac.za/~kroon/courses/machine_learning/lecture5/mlp.pdf

6 Hong Chen, Canizares C A, Singh A. ANN-based Short-Term Load Forecasting in Electricity Markets. Power engineering society winter meeting 2001 IEEE; 2001.

7 Vapnik Vladimir. Statistical learning theory. New York: John Wiley and Sons Inc; 1998.

8 Theodoridis Sergios, Koutroumbas Konstantinos. Pattern Recognition.4th ed. San Diego: Elsevier; 2009.

9 Weka [homepage on the Internet].University of Waikato, Hamilton, New Zealand, [cited 2010 Jan 20]. Available from: http://www.cs.waikato.ac.nz/ml/weka/

10 Cheng G. Weng and Josiah Poon. A New Evaluation Measure for Imbalanced Datasets. Seventh Australasian Data Mining Conference; 2008.

11 Nitesh V. Chawla et. al. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research; 2002.

12 Yetian Chen. Learning Classifiers from Imbalanced, Only Positive and Unlabeled Data Sets. The 14th international conference on Knowledge discovery and data mining; 2008.

13 Zhi-Lin Qu and Hai-Qing Hu. A prediction model for forest fire-burnt area based on meteorological factors [serial on the Internet]. [cited 2011 Feb 10]; Available from: http://en.cnki.com.cn/Article_en/CJFDTOTALYYSB200712009. htm

14 HU Lin, ZHOU Goumin and QIU Yun. Application of Apriori Algorithm to the Data Mining of the Wildfire. 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery; 2009.

15 Cortez Paulo , Morais Anibal. A data mining approach to predict forest fires using meteorological data [serial on the Internet]. [cited 2009 Nov 29]; Available from: http://www.dsi.uminho.pt/~pcortez/

16 Stephen Taylor and Martin Alexander. Science, technology, and human factors in fire danger rating: the Canadian experience. International Journal of Wild and Fire; 2006.

17 Davenport Mark. The 2*v*-SVM: A Cost-Sensitive Extension of the *v*-SVM. (Dissertation). Rice University; 2005.

18 R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2006.

19 Hsu Chih-Wei, Chang Chih-Chung, Lin Chih-Jen. A Practical Guide to Support Vector Classification. [homepage on the Internet]. [cited 2010 Dec 18]. Available from: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

20 George Sakr, Elhajj Imad. Artificial Intelligence for Forest Fire Prediction: A Comparative Study. I International Conference on Forest Fire Research D. X. Viegas (Ed.); 2010.

21 George Sakr,Elhajj Imad, Mitri George. Efficient forest fire occurrence prediction for developing countries using two weather parameters. Engineering Applications of Artificial Intelligence; 2011.

22 สุมาลัย นุชิต. การวิเคราะห์ปัจจัยที่มีอิทธิพลต่อการเพิ่มลดของจำนวนนักศึกษา กรณีศึกษา : วิทยาลัยการอาชีพวังสะพุง .(วิทยานิพนธ์). ขอนแก่น : มหาวิทยาลัยขอนแก่น; 2553.

23 Qiong Gu et al. Data mining on imbalanced data sets. 2008 International Conference on Advanced Computer Theory and Engineering;2008.

APPENDICES

APPENDIX A

Table Name : amphore

Description : to store Amphur information.

Primary Key : Id_amphore

Foreign Key : Id_province

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
1	Id_amphur	PK	text	10	Code of Amphur.
2	n_amphore		text	50	Name of Amphur.
3	Id_province	FK	text	10	Code of Province.

Table Name : cause

Description : to store cause of fire occurrence

Primary Key : Id_cause

Foreign Key : -

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
1	Id_cause	PK	Text	10	Code of cause of fire occurrence.
2	n_cause		Text	50	cause of fire occurrence.

Table Name : fire_data

Description : to store of fire data

Primary Key : id_station
Date
Time_call
id_placefire

Foreign Key : id_station
id_placefire
id_percall
id_pererport

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
1	id_station	PK,FK	Text	10	Code of station.
2	date	PK	Date/time	-	Days which get the information of fire occurrence.
3	time_call	PK	Date/time	-	Time which get the information of fire occurrence.
4	id_placefire	FK	Text	10	Code of place of fire occurrence.
5	id_percall	FK	Text	10	Code of person who inform the fire occurrence.
6	id_perreport	FK	Text	10	Code of person who record the fire occurrence.
7	Time_start		Date/time	-	Time to start extinguish fire.
8	Time_end		Date/time	-	Time when finishing extinguish fire .
9	Area_s		int	20	Total area of extinguish fire.
10	Id_cause	PK	Text	10	Code of cause of fire occurrence.

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
11	A1		Float	11	Dry Dipterocarp Forest.
12	A2		Float	11	Mixed Deciduous Forest.
13	A3		Float	11	Coniferous Forest.
14	A4		Float	11	Dry Evergreen Forest.
15	A5		Float	11	Tropical Rain Forest.
16	A6		Float	11	Peat Swamp Forest.
17	A7		Float	11	Forest deterioration.
18	A8		Float	11	Private area.

Table Name : member

Description : to store username and password of this system

Primary Key : userid

Foreign Key : id_status

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
1	userid	PK	Text	10	Code of person.
2	username		Text	10	Name of username.
3	password		Text	30	Name of password.
4	Id_status	FK	Text	10	Code of status of member.

Table Name : part

Description : to store part of Thailand

Primary Key : id_part

Foreign Key : -

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
1	Id_part	PK	Text	10	Code of part of Thailand.
2	n_part		Text	50	Name of part of Thailand.

Table Name : personal

Description : to store person information

Primary Key : id_person

Foreign Key : id_title

id_station

id_province

id_amphur

id_tumbon

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
1	Id_person	PK	Text	10	Code of person.
2	Id_title	FK	Text	10	Code of title.
3	F_name		Text	20	First name of person.
4	S_name		Text	20	Last name of person.
5	Address		Text	50	Name of address.
6	Id_province	FK	Text	10	Code of province.
7	Id_amphur	FK	Text	10	Code of Amphur.
8	Id_tumbon	FK	Text	10	Code of Tumbon.
9	Postcode		Text	50	Postcode.
10	Email		Text	50	Email.
11	tel		Text	10	Telephone number.
12	Id_station	FK	Text	10	Code of station.

Table Name : placefire_data

Description : to store place of fire occurrence information.

Primary Key : id_placefire

Foreign Key : id_province
id_amphur
id_tumbon

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
1	Id_placefire	PK	Text	10	Code of place of fire occurrence.
2	N_place		Text	30	Name of place of fire occurrence.
3	Moo		Text	10	Moo.
4	Id_province	FK	Text	10	Code of province.
5	Id_amphur	FK	Text	10	Code of Amphur.
6	Id_tumbon	FK	Text	10	Code of Tumbon.
7	Lat		Text	10	Latitude.
8	long		Text	10	Longitude.

Table Name : province

Description : to store province information

Primary Key : id_province

Foreign Key : id_part

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
1	Id_province	PK	Text	10	Code of province.
2	N_province		Text	50	Name of province.
3	Id_part	FK	Text	10	Code of part of Thailand.

Table Name : station_data

Description : to store station data.

Primary Key : id_station

Foreign Key : id_province

id_amphur

id_tumbon

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
1	Id_station	PK	Text	10	Code of station.
2	N_station		Text	50	Name of station.
3	N_stationpast		Text	50	Name of station in the past.5
3	Address		Text	50	Name of address.
4	Id_province	FK	Text	10	Code of province.
5	Id_amphur	FK	Text	10	Code of Amphur.
6	Id_tumbon	FK	Text	10	Code of Tumbon.
8	Postcode		Text	50	Postcode.
9	Email		Text	50	Email.
10	tel		Text	10	Telephone number.
11	cover		int	10	Area that station cover.
12	Lat		Text	10	Latitude.
13	Long		Text	10	Longitude.

Table Name : status

Description : to store status of member

Primary Key : id_status

Foreign Key : -

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
1	Id_status	PK	Text	10	Code of status.
2	n_status		Text	50	Name of status.

Table Name : title

Description : to store title name.

Primary Key : id_title

Foreign Key : -

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
1	Id_title	PK	Text	10	Code of title.
2	n_title		Text	50	Name of title.

Table Name : tumbon

Description : to store Tumbon information

Primary Key : id_tumbon

Foreign Key : -

No.	NAME	KEY	TYPE	LENGTH	DESCRIPTION
1	Id_tumbon	PK	Text	10	Code of Tumbon.
2	n_tumbon		Text	50	Name of tumbon.
3	Id_amphur		Text	10	Code of Amphur.

APPENDIX B

แบบสอบถาม

ความพึงพอใจต่อการใช้ระบบจัดการไฟป่า กรณีศึกษาจังหวัดเชียงใหม่

คำชี้แจง: แบบสอบถามฉบับนี้จัดทำเพื่อประเมินประสิทธิภาพของระบบระบบจัดการไฟป่า กรณีศึกษาจังหวัดเชียงใหม่

หลังจากที่ท่านได้ทดลองใช้ระบบแล้ว แบบสอบถามฉบับนี้ประกอบด้วย 3 ตอน คือ

ตอนที่ 1: ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม

ตอนที่ 2: แบบสอบถามสำหรับประเมินประสิทธิภาพของระบบ

ตอนที่ 3: ความคิดเห็นเพิ่มเติม

ตอนที่ 1: ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม

คำชี้แจง: โปรดทำเครื่องหมาย ✓ ลงในช่องว่างที่ตรงกับข้อมูลส่วนตัวของท่าน

1. เพศ	☐ ชาย	☐ หญิง
2. ระดับการศึกษา	☐ อนุปริญญา	☐ ปริญญาตรี
	☐ ปริญญาโท	☐ ปริญญาเอก
3. อาชีพ	☐ นักศึกษา	☐ พนักงาน
	☐ ข้าราชการ	☐ อื่นๆ (ระบุ)
		....

ตอนที่ 2: แบบสอบถามสำหรับประเมินประสิทธิภาพของระบบ

คำชี้แจง: โปรดทำเครื่องหมาย ✓ ลงในช่องที่ท่านเห็นว่าเหมาะสมที่สุด

รายการประเมิน	**ระดับความคิดเห็น**				
	ดีมาก (5)	**ดี (4)**	**ปานกลาง (3)**	**น้อย (2)**	**น้อยมาก (1)**
การประเมินระบบด้าน Functional Requirement Test					
1. ความสามารถในการนำเสนอข้อมูล					
2. ความสามรถในการสืบค้นข้อมูล					
3. ความสามารถในการจัดการฐานข้อมูล					
4. ความสามารถในการจัดหมวดหมู่ข้อมูล					
การประเมินระบบด้าน Functional Test					
1. ความถูกต้องในการนำเสนอข้อมูล					
2. ความถูกต้องในการสืบค้นข้อมูล					
3. ความถูกต้องในการจัดการฐานข้อมูล					
4. ความถูกต้องในการจัดหมวดหมู่ข้อมูล					
การประเมินระบบด้าน Usability Test					
1. ความเหมาะสมของข้อมูลที่นำเสนอ					
2. ความชัดเจนของข้อความที่แสดงผล					
3. ความง่ายในการใช้งานระบบ					
4. ความสามารถของระบบในภาพรวม					
5. ความพึงพอใจของผู้ใช้งานระบบ					

ตอนที่ 2: แบบสอบถามสำหรับประเมินประสิทธิภาพของระบบ (ต่อ)

คำชี้แจง: โปรดทำเครื่องหมาย ✓ ลงในช่องที่ท่านเห็นว่าเหมาะสมที่สุด

รายการประเมิน	ระดับความคิดเห็น				
	ดีมาก (5)	**ดี (4)**	**ปานกลาง (3)**	**น้อย (2)**	**น้อยมาก (1)**
การประเมินระบบด้าน Security Test					
1. การกำหนดสิทธิในการเข้าถึง					
2. ความสามารถในการเข้าถึงข้อมูลเพื่อทำการแก้ไขข้อมูล					
3. การกำหนดสิทธิในการแก้ไขข้อมูลของผู้ดูแลระบบ					
4. ความสามารถในการจัดหมวดหมู่ข้อมูล					

ตอนที่ 3: ความคิดเห็นเพิ่มเติม

คำชี้แจง: โปรดกรอกข้อความลงในที่ว่าง

ความคิดเห็นเพิ่มเติม

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

BIOGRAPHY

NAME	Mr. A-san Somboonying
DATE OF BIRTH	12 July 1984
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTIONS ATTENDED	Kasetsart University, 2003-2007 Bachelor of Science (Packaging Technology) Mahidol University, 2008-2012 Master of Science (Technology of Information System Management)
RESEARCH GRANTS	Data mining, Database, Web application
HOME ADDRESS	22 moo 9 Theparak Road, Bangpla, Bangplee, Samutprakarn 10540 Tel 087-012-3675 E-mail : asan_nus@hotmail.com
PUBLICATION / PRESENTATION	NCIT 2012 (National Conference on Information Technology 2012)
SCHOLARSHIP	- This thesis is partially supported by Graduate Studies of Mahidol University Alumni Association - University Industry Research Collaboration Program (U – IRC)