

An Integrated System of Geographic Information and Water Quality: Lamtakong River

Manoch Thinsungnoen

Program of Computer Science, Nakhon Ratchasima Rajabhat University (NRRU), Thailand

Email: manoch.t@gmail.com

Tippaya Thinsungnoen and Sunsanee Liangpanit

NRRU, Program of Information Technology, Management Information System, Thailand

Email: {tippayasot, sunsaneel}@hotmail.com

Pasapitch Chujai

Electrical Technology Education Department, King Mongkut's University of Technology Thonburi, Thailand

Email: pasapitchchujai@gmail.com

Abstract—A cause of the river being shallow, and the water quality being rotten from land-used were changed. Therefore, the study of a relationship between water quality and land-used that usefully for a water pollution management. The purposes of this research were 1) to develop a model of factors that influence water quality using j48 algorithm, 2) to develop a Geographic Information System for Retrieve and display data quality and Water Pollution (GIS_IRWP), and 3) to evaluate the effectiveness of GIS_IRWP. The results of this research found that: The first, the decision tree model provided an accuracy 78.21%. The second, the GIS_IRWP was the information system of spatial data (included: district boundary lines, river transport, Lamtakong River lines, land-used and agricultural areas) integrated with water quality data. The GIS_IRWP can retrieved the information that presented in spatial form. The last, an evaluation the effectiveness of the system by both of experts and users found that the system had effective in high-level means at 4.07 and 4.48 respectively.

Index Terms—decision tree model, J48 algorithm, geographic information system, water quality

I. INTRODUCTION

The Lamtakong is an important river in Nakhon Ratchasima province that has a total length is about 224 kms, and areas about 3,050 km². The Lamtakong River has the Lamtakong dam is a large headwaters source. Now Lamtakong River is shallow, and the water quality is quite poor because changing used forest areas transform to resort, golf court and livestock. In addition, community wastewater drainage to river without healing that cause increased the crisis of quality in the Lamtakong River. This problem leads many agencies are looking for solutions to the crisis of water shortage, restore water resources for sustainable develop natural resources and

environment in the Lamtakong watershed. The solutions are using of integrated natural resources to maximize the benefits, including prevention and remediation of pollution in the Lamtakong watershed area. The information technology and data mining have been advances rapidly. Some technologies integrated in several a research or a study, and applied in bioinformatics, agricultural, industry, environmental management and others [1]-[3]. A Geographic Information System (GIS) Technology is process of spatial data using computer systems that defined descriptive data (attribute data) and an information such as address relative location of spatial data such as houses, roads, rivers and others in the database. These technologies are effective if we are applied in the management or decision making [4]. GIS technologies can be found in research widely [5]-[8]. Next technology is a "knowledge discovery in database (KDD)" or data mining. Data mining is technology that explores a large data to finding a relationship, pattern or trend from the dataset [3]. A popular technique in data mining such as decision tree [9] that classified as supervised learning in machine learning. The structure of the decision tree model can help users understand easily to decision condition, work is not complicated, and fast processing [10], [11]. Thus, decision tree techniques it found in several studies, and can be used to support a decision making across multiple tasks. [12]-[24], for example, in [16] finding the factors that related to diabetes mellitus based-on a combination of techniques included association rule and decision tree. In the research [13] development of a decision support system for analyzing the risk of two chronic diseases in cases of diabetes mellitus and hypertension based-on the decision tree technique. In addition can be found some research that integrates both of GIS and decision tree, such as the research on the land-cover, and land-used data collected [17], [18]. The study to develop a classification module for hotspot occurrences in a GIS, and classification module was made using the decision tree [19].

Manuscript received April 24, 2019; revised September 26, 2019.

Therefore, this paper proposes an integrated system of geographic information and water quality of the Lamtakong River in Nakhon Ratchasima province. The GIS_IRWP was developed in web-based application using PHP language, and a database management system was using MySQL. An application was using the maps of Nakhon Ratchasima province. This research integrated all of water quality data, spatial data, land-used and water quality model in form decision tree model. The decision tree was constructed that using the best model derived from c4.5 (j48) algorithms.

II. LITERATURE REVIEW

The contents of this section consist of five topics as the following.

A. Information System

The information system section can be combines to two part such as.

Information is the result of raw data processing by collecting data from various sources. The process, whether it is grouping data, sorting data, calculation and summary. Then, bring the information to make a proper report, which creates benefits for human life, such as in the areas of daily life, news, academic and business knowledge [20].

Information System is a system of storage, data processing, and information presenting, with people and information technology in operation to get information that is suitable for each task. The basic activities of information system have three types: Inputting, Processing and Outputting. The work starts from changing the raw data that it comes through the input step through processing step will be information that comes out of the output step. Output will return to Input for further evaluation [21], [22].

B. Geographic Information System

The Geographic Information System (GIS) is an information system designed to collect, store, analyze geographic data including data retrieval and information display, in other words, is Geographic Information Systems (GIS) are both database systems, and operating systems to analyze those data. GIS has ability to manage spatial data, and use for a decision making in the form of numerical maps, feature information [4]. The GIS was considered a tool used to analyze spatial data by the characteristic data in the area.

The GIS has five components that interoperate with the computer system for transforming the spatial data into useful information [23]. The five components of a GIS are hardware, software, people, procedure, and data.

A feature of data in GIS to consist of three parts as including: spatial data, attribute data, and time [4]. The spatial data can be referenced to geo-address (Geo-referenced data) on land. The attribute data is an attribute of geo-data often called non-spatial data because it does not show any information related to the geo-address. The time is important because the Geo-data often reference to any point in any period. Thus, the knowing time that

collection the Geo-data being the most important in proper usage Geo-data.

C. Water Pollution Sources

The Ministry of Resources and Environment issued the Ministerial Regulations to prescribe rules, procedures and forms for collecting statistics and data. Preparation of detailed records and a summary report of the wastewater treatment system 2012 in the enforcement of the Ministerial Regulations has been enforced since August 2, 2012, as announced in the Government Gazette, volume 129, episode 39 A, on the 4th of May 2012, according to the Ministerial Regulations above. The Ministerial Regulations have determined the source of pollution that controlled release wastewater into public water sources or to the environment outside the area sets by section 69 of the National Environmental Quality Act and the Environmental Protection Act in 1992, which has a wastewater treatment system as belonging of self, according to section 70 that divided into 10 sources [24]. The source of pollution as including: industrial factories and industrial estates, some types of buildings and sizes, allocated land, swine farming, fishing pier (fish pier, and fish raft business), fuel service stations, coastal aquaculture ponds, brackish aquaculture ponds, freshwater aquaculture ponds, and community wastewater treatment system.

D. Water Quality Index (WQI)

The meaning of water quality has a wildly that will be used differently in each place and each country. In addition, it depends on the purpose of using water together, such as for consumption, consumption in the community, used in industrial are used in agriculture, and relax etc. In determining the quality of water, it is considered the nature of water, physical and chemical [25]. Thus, the water quality requirements are different, depends on how much water is used. The water quality index mentioned is an index that indicates the condition of the river in general.

The general water quality index (WQI) mentioned has a score of 0 to 100, with 91–100 points as the water quality is very good, 71–90 points is good, 61–70 points is fair, 31–60 points is poor, and 0–30 points as the water quality is in a very poor. These scores are usually caused by a combination of 9 indices including: positive potential of the hydrogen ions (pH), dissolved oxygen (DO), total solid (TS), fecal coliform bacteria (FCB), nitrate (NO_3^-), phosphate (PO_4^{3-}), Turbidity, Temperature, and dirtiness in the form of organic matter (Biological Oxygen Demand: BOD) together as a single score using the equation as follows:

$$WQI = [(pH)(DO)(TS)(FCB)(NO_3^-)(PO_4^{3-})(Turbid)(Temp)(BOD))]^{1/9} \quad (1)$$

Later, the factors of WQI were reduced to five factors included DO, FCB, NH_3 , BOD, and TCB, which compute for WQI then transforms WQI to water quality level into five levels as follows: very good, good, fair, poor, and very poor.

E. Decision Tree Algorithms

There are many a decision tree algorithms can use developing the decision tree model, i.e. ID3, CART, C4.5 (J48), LADTree, BFTree and NBTree. A model was created by decision tree technique is a classification scheme which generates a tree and a set of rules, representing the model of different classes. The decision tree induction has a learning and classification steps are simple and fast.

J48 algorithm in weka application or call name C4.5, this algorithm is an extension of ID3, presented by E. B. Hunt, J. Marin, and P. T. Stone. Quinlan [11]. A basic decision tree algorithm is summarized as below.

Step 1: he leaf is labelled with the same class if the instances belong to the same class.

Step 2: For every attribute, the potential information will be calculated and the gain in information will be taken from the test on the attribute.

Step 3: Finally the best attribute will be selected based on the current selection parameter.

III. OBJECTIVES OF RESEARCH

The Objectives of this research as following:

To develop a model of factors that influence water quality based on the monitoring of water quality of the Lamtakong River using j48 algorithm.

To develop the Geographic Information System for Retrieve and display data quality and Water Pollution (GIS_IRWP)

To evaluate the effectiveness of GIS_IRWP.

IV. RESEARCH METHODOLOGY

The contents of the research methodology section consist: dataset and study area, system analysis and design, database design, and decision tree model as the following.

A. Dataset and Study Area

The dataset and study area used in this research included as follows:

Water quality: This research collected data on water quality from all of 15 monitoring stations that distributed along with the Lamtakong River line in Nakhon Ratchasima province between the years 2008-2013 by environment office region 11 [26] as show in Fig. 1, and a description shows in Table I. The dataset of water quality used for analysis were the value of factors inside the river consists of: depth, temperature air, temperature water, PH, turbidity (Turbid), total solids (TS), nitrate-nitrogen (NO₃), total phosphorus (TP), dissolved oxygen (DO), biochemical oxygen demand (BOD), total coli form bacteria (TCB), fecal coli form bacteria (FCB), ammonia-nitrogen (NH₃), and water quality index (WQI), which WQI can be computed by equation (1).

Study Area: All of dataset as water quality and land-used were collected from the study area that along the Lamtakong River line. The study area has boundary cover the area of 6 districts, namely Chaloem phra Kiat, Mueang, Kham Thale So, Sung Noen, Sikhio, and Pak Chong as show in Fig. 1.

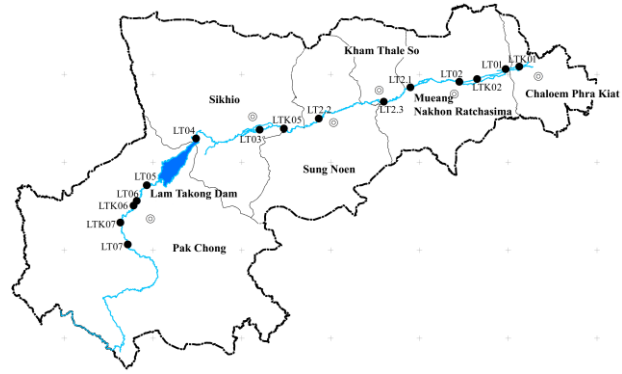


Figure 1. The study area on Lamtakong river in Nakhon Ratchasima province.

TABLE I. THE MONITORING STATION IN NAKHON RATCHASIMA PROVINCE

Code.	Address
LT01	Gate of Lamtakong River, Phanao sub-district, Mueang district.
LT02	Bridges of Wat Samakee community, Nai Mueang sub-district, Mueang district.
LT2.1	MakhamTao dam, BanMai sub-district, Mueang district. (LTK03)
LT2.2	TakhongLong bridge, Sung Noen sub-district, Sung Noen district.
LT2.3	KhamThaleSo bridge, KhamThaleSo sub-district, KhamThaleSo district. (LTK04)
LT03	BanKutChanuan bridge, Mittraphap sub-district, Sikhio district.
LT04	Lamtakong pump-point, Khlong Phai sub-district, Sikhio district.
LT05	Department of Quartermaster of Army Division bridge, Pak Chong district.
LT06	Nong Sarai bridge, Pak Chong District.
LT07	Ban BuKruead, KhanongPhra sub-district, Pak Chong district.
LTK01	Kanphom dam, PhraBuddha, ChaloemPhraKiat district.
LTK02	BanThaKraSang bridge, HuaThale sub-district, Mueang district.
LTK05	BungLamyai bridge, Sikhio Subdistrict, Sikhio District
LTK06	ThaManao bridge, Nong Sarai sub-district, Pak Chong district.
LTK07	Chantarasathit bridge, Pak Chong sub-district, Pak Chong district.

Land-used: The pollution sources that used in this research assume that were land-used. The land-used might be important reason leads water source became shallow, and the water quality being rotten. This research we collected the land-used data from an environment office region 11 [26], included: Residences, Industrials and Commercials, Government institutions, Education institutions, Hospitals, Petrol stations, Hotels, Golf Courses, Livestock, Other Farms, Rice fields, Cassavas, Corns, Canes.

Maps: An application was using the maps of Nakhon Ratchasima province in 1:50,000 scales, consist of Administrative boundaries of Nakhon Ratchasima province (district and province), map of sloping river, the Lamtakong River line in Nakhon Ratchasima area, and other relevant maps. A description of the maps shows in the Table II.

TABLE II. LIST OF DATA LAYERS, AND DATA TYPES

Data	Data Layers	Data Types
1. Administrative boundaries	1.1 Provincial boundary 1.2 District boundaries 1.3 Sub-district boundaries 1.4 District location	1.1 Polygon 1.2 Polygon 1.3 Polygon 1.4 Point
2. Transportation route	2.1 Road 2.2 Railway	2.1 Line 2.2 Line
3. River	3.1 Lamtakong River	3.1 Line
4. Land-used	4.1 Residences 4.2 Industrials and Commercials 4.3 Government institutions 4.4 Education institutions 4.5 Hospitals 4.6 Petrol stations 4.7 Hotels 4.8 Golf courses 4.9 Livestock Farm 4.10 Other Farms 4.11 Rice fields 4.12 Cassavas 4.13 Corns 4.14 Canes	4.1 Point 4.2 Point 4.3 Point 4.4 Point 4.5 Point 4.6 Point 4.7 Point 4.8 Point 4.9 Point 4.10 Point 4.11 Polygon 4.12 Polygon 4.13 Polygon 4.14 Polygon
5. Monitoring stations	5.1 Monitoring Stations	5.1 Point
6. Water quality	6.1 Water quality in 2008 6.2 Water quality in 2009 6.3 Water quality in 2010 6.4 Water quality in 2011 6.5 Water quality in 2012 6.6 Water quality in 2013	6.1 Point 6.2 Point 6.3 Point 6.4 Point 6.5 Point 6.6 Point

B. The Framework of GIS_IRWP

This research was designing a system from data that collected, and divided to five modules consist of: Data Management module, Retrieval and Report module, Prediction module, User Interface module (website), and Database module, all of them can shows as Fig. 2.

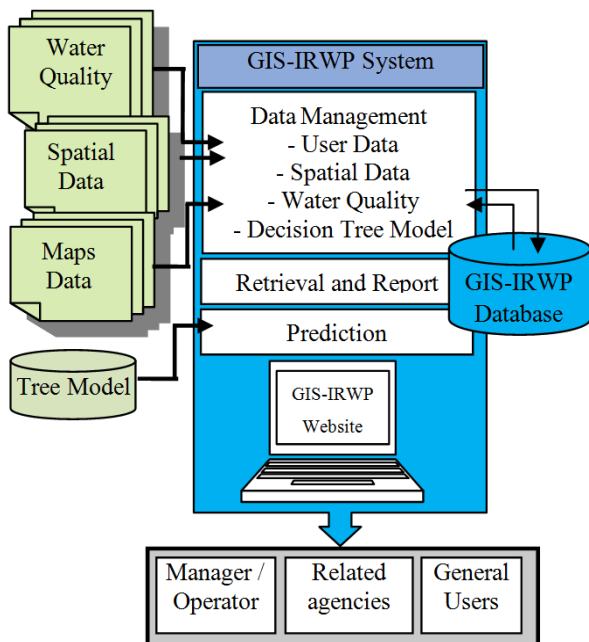


Figure 2. The framework of GIS_IRWP.

C. System Analysis and Design

In this section we describe for a system analysis and design of GIS_IRWP. The design separates into four

sections as consist: data flow design, functional design, user interface design, and report design as follows.

Data flow design: In Fig. 3 shows the context diagram for GIS_IRWP system. The elements are described as follows: an input data was included data of users, water quality, spatial data, land-used data, and decision tree model, which an information that derived from the system were according to input data. For the external entities that associated with the system includes general user, and administrator.

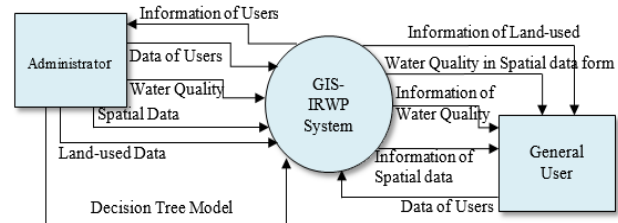


Figure 3. The context diagram of GIS_IRWP.

Functional design: The GIS_IRWP was designed for four functions as follows.

Data management process was designed for the administrator that can be inputting, modifying, and deleting of data that have authorization.

The spatial retrieval process was designed that will be displayed as a map by showing the layer of data or can was combined. The data that shows consist of general information, land-used, agricultural area, and water quality during the year 2008-2013.

The attribute retrieval process that attributes, data retrieval will be displayed in the data table format. The process can be retrieved items as follows: the quality measurement stations (according to sub-district and district codes), the water quality data (according to the date range, station code, and water quality level), and the type of pollution source (according to the code and the name of the source of pollution).

User authorization process was designed for all users to operating them work according to access rights was assigned.

User interface design: This work was designed the user interface for two types of user include administrator and general user.

An administrator interface was designed to be able to access the system via the home screen, login screen, retrieve-report screen, and the admin screen. Which the first three screens, access rights like general users. The admin screen is different, require administrators to manage data Initial configuration user data, master data and tree model.

The General user interface was designed to be able to access the same system as the first three screens of administrator. Which on the screen of retrieve-report can be used as well for both general users and level supervisor.

Reporting design: The design of part of the retrieval-report, there are details for designing as follows:

General information was designed to display the boundary line of the district consisting of sub-district

boundaries, district boundaries and provincial boundaries, transportation route, the Lamtakhong River line, Show detailed information about the sub-district boundary data.

The *station of measurement* was designed to display the location of all 15 stations. The measurement station can display the details of the sub-district boundary data.

The *land-used* was designed to display the location of the area of use, and agricultural areas, which can display the details of the sub-district data, the location and accountability of the stations.

The *representing of density for land-used* was assigned to five levels. This work was using the color to represent for level of density as follows: green instead of the lowest level where the density is 0-50, yellow instead of the low level where the density is 51-100, orange instead of the medium level where the density is 101-300, red instead of the high level where the density is 301-600, dark red instead of the highest level where the density is more than 600, and black instead of none of data.

The *water quality* was assigned to six levels as follows: dark green instead of the very good level where the WQI is 91-100, green instead of the good level where the WQI is 71-90, orange instead of the fair level where the WQI is 61-70, brown instead of the poor level where the WQI is 31-60, red instead of the very poor level where the WQI is 0-30, and black instead of none of data.

D. Database Design

In this research, the database was designing using MySQL. The design of the table for storing data were five entities include: *tbStationWater* for stores the measurement station, *tbWaterQuality* for stores water quality data, *tbAgriculture* for stores the area of agriculture, *tbPollSrc* for stores the data of land-used, and *tbPollutionType* for stores a types of water pollution, which a relationship of the entities as shows in Fig. 4.

E. Decision Tree Model

The decision tree model in this research was created to support a management of water pollution. The model has found a relationship between water quality and Land-used. Which, the factors of modelling has used six factors including: Biological Oxygen Demand (BOD), Total Coli form Bacteria (TCB), Dissolved Oxygen (DO), Positive potential of the Hydrogen ions (pH), Ammonia-nitrogen (NH₃) and Industrials and Commercials (Trade).

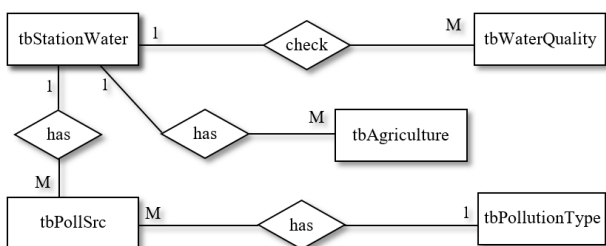


Figure 4. The entity relationship diagram of GIS_IRWP database.

This research was creating a decision tree model using J48 algorithm that can transform to the decision rules, and can use for predicting situate of water quality, the model shows as follows.

```

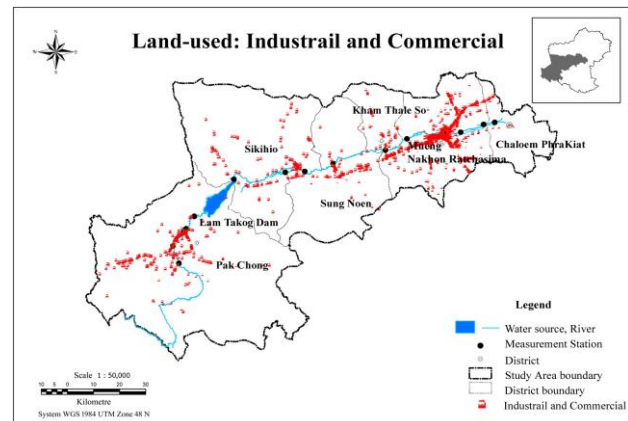
BOD <= 2
| TCB <= 9000
| | DO <= 4.1: poor (3.0/1.0)
| | DO > 4.1
| | | TCB <= 300
| | | | BOD <= 1: very_good (7.0)
| | | | BOD > 1
| | | | pH <= 8.2: good (14.0)
| | | | pH > 8.2: very_good (3.0/1.0)
| | | TCB > 300: good (35.0/1.0)
| | TCB > 9000: fair (5.0/1.0)
BOD > 2
| NH3 <= 8.13
| | DO <= 5.5
| | | BOD <= 3.8
| | | | TCB <= 5000
| | | | trade1 <= 1: poor (3.0)
| | | | trade1 > 1: fair (13.0/1.0)
| | | | TCB > 5000: poor (16.0/1.0)
| | | BOD > 3.8: poor (35.0/1.0)
| | DO > 5.5
| | | TCB <= 300
| | | | trade1 <= 24: good (3.0)
| | | | trade1 > 24: fair (2.0)
| | | TCB > 300: fair (14.0/1.0)
| | NH3 > 8.13: very poor (3.0)
  
```

V. RESEARCH RESULTS

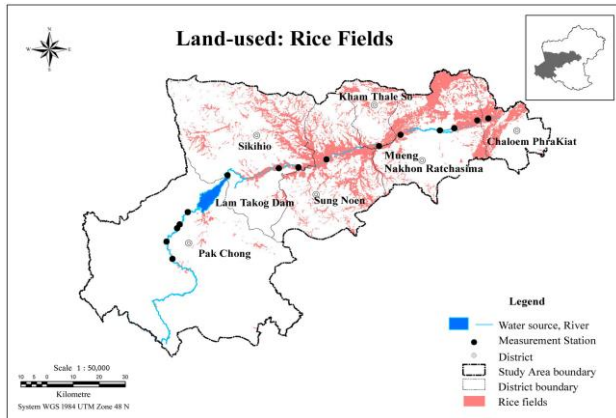
The research results can separating to four parts include the map development results, the decision tree modelling results, the GIS_IRWP development results, and the system evaluation results as follows.

A. The Map Development Results

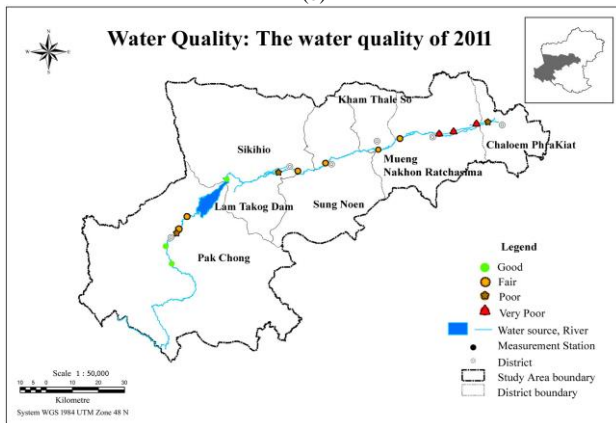
The map developed that was used for the GIS_IRWP system is due to the stacking of the data layer. The development of the map was using the data layers, and data types that both of the polygon, point and line types. Then, we were integrating a decision tree model derived from j48 algorithm to presenting the station zone according to water quality, an example of the map results as shows in Fig. 5(a) presented the points of the industrial and commercials, in Fig. 5(b) presented the polygon of rice fields, and the water quality in 2011 was presenting in Fig. 5(c).



(a)



(b)



(c)

Figure 5. The results of the map developed.

B. The Decision Tree Modelling Results

The decision tree was modelling with J48 algorithm that using data set are 156 instances. The results of modelling has shown that correctly classified instances are 122 instances, and incorrectly are 34 instances, an accuracy as 78.21%, RMSE as 0.29, recall as 0.92, precision as 0.93, and f-measure as 0.92.

After creating the tree model, this research was applying the model for the module of predicting that received factors based-on the tree model. User inputs the six factors into predict screen included BOD, TCB, DO, pH, NH3 and number of the trade around the area, then the system will predict the water quality at any level, including the probability of previewing the water quality forecast from the model as shown in Fig. 6.

Water quality forecasting in Nakhon Ratchasima province

Input the factors of water quality

Biological Oxygen Demand (BOD)	4.2
Total Coli form Bacteria (TCB)	100000
Dissolved Oxygen (DO)	4.8
Positive potential of the Hydrogen ions (pH)	4
Ammonia-nitrogen (NH3)	0.2
Industrials and Commercials (Trade)	10

Predicts Cancel

The results of predicting a water quality

The water quality as level: เลือนโทรม (Poor)

The probability: 0.971

Figure 6. The water quality forecasting based-on decision tree model.

C. The GIS_IRWP Development Results

The development results of GIS_IRWP have a detail separated to two section included section of management data, and section of retrieval and report as follow:

Section of management data is the section for administrator has a role for managing all data included: news, user, station of water quality, pollution type, pollution source, agriculture area, and water quality. The managing process consists of add, edit, and delete that show as Fig. 7.

Section of retrieval and report is the section for all users, both administrator and general user to use the system for searching data on land-used, and water quality that presenting in spatial data, and table. The system can show as Fig. 8.

Figure 7. The section of management data.

Figure 8. The section of retrieval and report.

D. The System Evaluation Results

In evaluating the effectiveness of GIS_IRWP. The system was tested in three aspects: Functionality and Reliability, Meet the requirements, and Usability. The test results are as follows.

The system evaluation of user amount 30 people consist of general user as 23, the policy level user as 3, and operator as 4. An evaluating was found that the system has the functionality and reliability, meet the requirements at the highest level, mean of 4.45, and 4.62 respectively. The system has a usability at high level, mean as 4.38, and the summary, the performance of GIS_IRWP at high level, and mean as 4.48.

The system evaluation of expert amount 5 people consist of an expert of software application as 2, an expert of system development as 1, an expert of geographic information system as 1, and an expert of environment as 1. An evaluating was found that the system has the functionality and reliability, meet the requirements, and a usability at the high level, mean as 3.95, 4.07, and 4.20 respectively. In summary, the GIS_IRWP has the performance at the high level, mean as 4.07.

VI. CONCLUSION AND DISCUSSIONS

This research focuses on developing an information system based on the integration of spatial data, water quality data, and decision tree model, which aiming to enable the system to be a tool for personal in relevant local agencies to utilize the information system, by concluding and discussing the results as follows:

A. Conclusion

A conclusion of the research can be separated to three parts as follows:

Database development: The development of database of GIS_IRWP was using the MySQL tool that can be storing data into five tables including station table, pollution type, pollution source, agriculture area, and a water quality table.

Decision tree modelling: The decision tree model was constructed using the best model derived from j48 algorithms. The decision tree model has six factors included BOD, TCB, DO, pH, NH₃ and number of the trade. The model has an accuracy at 78.21%.

GIS_IRWP development: The GIS_IRWP was developing on web application using PHP language. An application was using the maps of Nakhon Ratchasima province in 1:50,000 scales. The system was integrated all of water quality data, spatial data, land-used, and water quality model in form decision tree model. The GIS_IRWP have main ability 3 functions consist of data management, retrieval and report, and prediction.

B. Discussions

The results of all the development can be discussed as follows: Based on the results of the map development, considering the severity, utilization area (source of pollution) and water quality level of the 15 stations during the years 2008-2013, we found interesting information as follows:

Density of land-used: When considering the level of density of the area used in the surrounding area, each station has the following details: The highest density is the station LT02, and the lowest are LTK01, LTK02, LT2.1, LT03, LT05, LT06, and LT07 shows as Fig. 9.

Water quality and Density of land-used: Considering the relevance between density levels of land-used and water quality levels, we found interesting information as follows:

The area where the density of pollution sources is at the lowest to low level, we found that the water quality level will be of fair quality to a good level clearly visible

in LT07, LTK07, LTK06, LT06, LT05, LT03, LTK05 and LT2.1 stations. Where the pollution source is at the highest level found that the water quality level would be somewhat deteriorated in LT02 station.

An interesting final issue that should be further studied is the area where the density of pollution sources is medium to very high, such as LT2.3, LT2.2, LT04 stations, the quality level has not poor, as it should be. But in the station such as LTK01, LT01 and LTK02, the level of density of land-used is the low to the lowest, the water quality level is not as good as it should be, more details in Fig. 10.

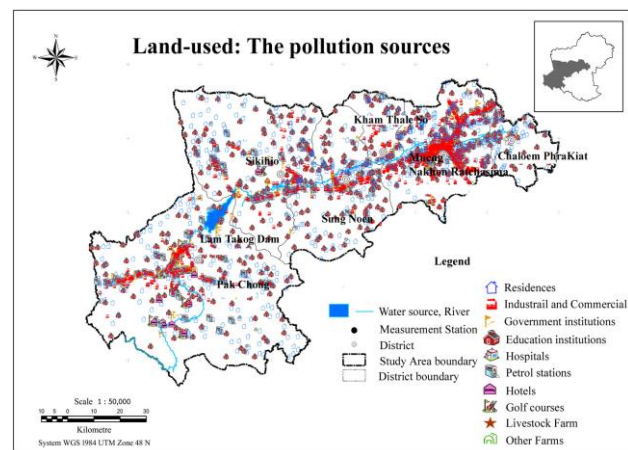


Figure 9. The map showing density level of pollution source.

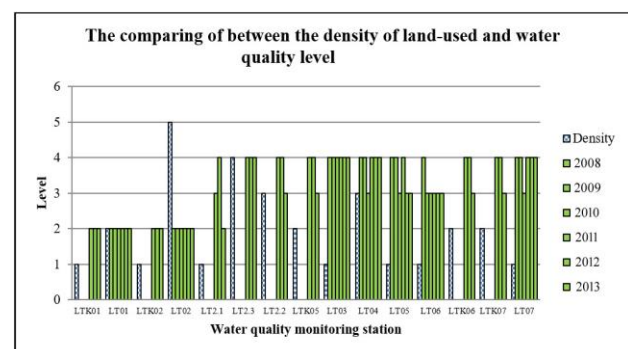


Figure 10. Comparison between density of land-used and water quality level.

ACKNOWLEDGMENT

This research was supported by Research and Development Institute have been funded by Nakhon Ratchasima Rajabhat University (NRRU).

REFERENCES

- [1] K. M. Raval, "Data mining techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 10, 2012.
- [2] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, Elsevier, 2011.
- [3] W. Lee, S. J. Stolfo, P. K. Chan, and W. Fan, *Real Time Data Minin-based Intrusion Detection*, NY: Columbia University, 2001.
- [4] S. Klindao, *Geographic Information System: Basic Principles*, Bangkok: Thammasat University, 1999.

- [5] W. Phaengsang, "Geographic information system for determining risk areas for air pollution in Bangkok," master of arts thesis, Arts Council Chulalongkorn University, 2004.
- [6] P. C. Lee, Y. Wang, T. P. Lo, and D. Long, "An integrated system framework of building information modelling and geographical information system for utility tunnel maintenance management," *Tunnelling and Underground Space Technology*, vol. 79, pp. 263-273, 2018.
- [7] E. Lepuschitz, "Geographic information systems in mountain risk and disaster management," *Applied Geography*, vol. 63, pp. 212-219, 2015.
- [8] E. Higgins, M. Taylor, H. Francis, M. Jones, and D. Appleton, "The evolution of geographical information systems for fire prevention support," *Fire Safety Journal*, vol. 69, pp. 117-125, 2014.
- [9] R. J. Roiger and M. W. Geatz, *Data Mining a Tutorial – Based Primer*, Addison Wesley, 2003.
- [10] A. K. Pujari, *Data Mining Techniques*, Universities Press, 2001.
- [11] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, Elsevier, 2011.
- [12] S. A. Mulay, P. R. Devale, and G. V. Garje, "Intrusion detection system using support vector machine and decision tree," *International Journal of Computer Applications*, vol. 3, no. 3, pp. 40-43, 2010.
- [13] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Applied Mathematics and Computation*, vol. 187, no. 2, pp. 1017-1026, 2007.
- [14] M. Saimurugan, K. I. Ramachandran, V. Sugumaran, and N. R. Sakthivel, "Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3819-3826, 2011.
- [15] M. Zorman, G. Masuda, P. Kokol, R. Yamanoto, and B. Stiglic, "Mining diabetes database with decision trees and association rules," in *Proc. the 15th IEEE Symposium Computer-based Medical Systems*, 2002, pp. 134-139.
- [16] N. N. Jongkasikit, "A decision support system for analyzing the risk of two chronic diseases: Diabetes mellitus and hypertension," *Industrial Technology Lampang Rajabhat University Journal*, vol. 9, no. 2, 2016.
- [17] M. A. Friedl and C. E. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote Sensing of Environment*, vol. 61, no. 3, pp. 399-409, 1997.
- [18] M. Punia, P. K. Joshi, and M. C. Porwal, "Decision tree classification of land use land cover for Delhi, India using IRS-P6 AWiFS data," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5577-5583, 2011.
- [19] K. Amri and I. S. Sitanggang, "A geographic information system for hotspot occurrences classification in Riau province Indonesia," *Procedia Environmental Sciences*, vol. 24, pp. 127-131, 2015.
- [20] P. Phanichakul, *Information Technology*, Bangkok: KTP Comp & Consulting, 2008.
- [21] S. Kiranandana, *Information Technology Statistics: Data in Information Systems*, Bangkok: Chulalongkorn University Printing House, 1998.
- [22] K. C. Laudon and J. P. Laudon, *Essentials of Management Information Systems: Organization and Technology in the Enterprise*, 4th ed., Upper Saddle River, NJ: Rentice Hall, 2011.
- [23] G. H. Castle, *Profiting from a Geographic Information System*, 3rd ed., Colorado, 1993.
- [24] Water Quality Management Bureau, "Ministerial regulations prescribing rules, procedures and forms for collecting statistics and data Preparation of detailed records and the summary report of the wastewater treatment system," Pollution Control Department, 2012.
- [25] P. Panichphan, P. Ruenwongsa, S. Sawangphol, and J. Tolim. (2012). Water pollution. [Online]. Available: http://www.il.mahidol.ac.th/edia/ecology/chapter3/chapter3_water8.htm
- [26] Office of the Environment Region 11. "Data Files," Regional Environment Office 11 (Nakhon Ratchasima), Ministry of Natural Resources and Environment, 2014.



University (NRRU), Thailand. His research of interest includes information system development, communication systems and networks.



She worked as a programmer at Saim-multisoft Company, as a system analyst at Alpha office automation Company, and as a system analyst at Turbosoft Company. She is currently a lecturer at the Program of Information Technology, Faculty of Science and Technology, Nakhon Ratchasima Rajabhat University. Her research of interest includes data mining, deep learning, artificial intelligence, software development and software engineering. She has published several papers about data mining topics, such as "The Clustering Validity with Silhouette and Sum of Squared Errors" (Kitakyushu: ICIAE2015, 2015), "A Deep Learning of Time Series for Efficient Analysis" (Nagoya: IJFCC, 2017), "Deep Autoencoder Networks Optimized with Genetic Algorithm for Efficient ECGs Clustering" (Singapore: IJMLC, 2018). Dr. Tippaya is a member of ResearchGate that the professional network for scientists and researchers.



Science and Technology, Nakhon Ratchasima Rajabhat University. Her research of interest includes Data mining, software development, project management and software quality management.



She is currently a lecturer at the Electrical Technology Education Department, Faculty of Industrial Education and Technology, King Mongkut's University of Technology Thonburi, Thailand. Her current research includes ontology, recommendation system, time series, and imbalanced data classification.

Manoch Thinsungnoen was born in Nakhon Ratchasima of Thailand in 1978. He received him bachelor degree in computer engineering from Suranaree University of Technology (SUT), Thailand, in 2001, master degree in information technology from Khon Kaen University (KKU), Thailand, in 2009.

He is currently a lecturers at the Program of Computer Science, Faculty of Science and Technology, Nakhon Ratchasima Rajabhat

Tippaya Thinsungnoen was born in Roi-ed of Thailand in 1976. She received her bachelor degree in computer science from Nakhon Ratchasima Rajabhat Institute, Thailand, in 1999, master degree and doctoral degree in computer engineering from Suranaree University of Technology (SUT), Thailand, in 2007, and the Ph.D. degree in computer engineering from Suranaree University of Technology, Thailand, in 2007.

Sunsanee Liangpanit was born in Nakhon Ratchasima of Thailand in 1976. She received her bachelor degree in computer science from Mahasarakham University (MSU), Thailand, in 1999, master degree in information technology from King Mongkut's University of Technology North Bangkok (KMITNB), in 2004, Thailand.

She is currently a lecturer at the Program of Management Information System, Faculty of

Pasapitch Chujai was born in Nakhon Si Thammarat of Thailand in 1975. She received her bachelor degree in computer science from Ramkhamhaeng University, Thailand, in 2000, master degree in computer and information technology from King Mongkut's University of Technology Thonburi, Thailand, in 2004, and the Ph.D. degree in computer engineering, Suranaree University of Technology, Thailand, in 2015.