STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–17.
Published online in International Academic Press (www.IAPress.org)

Database System and Model for Predicting Risk Level of Flood That Damages Rice Farming in Thailand

Wittawin Susutti [1], Pirun Dilokpatpongsa [1], Pawaton Kaemawichanurat [1], Teerapol Saleewong [1], Boonkong Dhakonlayodhin [2], Thidaporn Supapakorn [3], Wiboonsak Watthayu [1,*]

[1] *Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi*

[2] *Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok*

[3] *Department of Statistics, Faculty of Science, Kasetsart University*

Abstract Rice is always an important economic crop of Thailand as it is not only the staple in every family of the entire country but it also earns an extremely large income among all the Thai crop exports. However, Thai farmers are considered to be economically vulnerable and still have to face many difficulties, in particular, the flood problem. The flood problem has destroyed rice farming areas over the past decade until now. Risk and severity assessments mainly contribute to the government promptly subsidizing the farmers. In any case, the updated and reliable database systems are the main ingredients for developing the model of these assessments. In this paper, we develop a database system from a survey with 5,000 samples across the whole country. All the raw data has been managed to cleaned and prepared in order to develop a model that is used to predict risk level. The model achieves 87.24 percent accuracy with a significance level of 0.05. In addition, the model is able to select variables that have a statistically significant effect on the risk level forecast, and these variables could be used to improve the quality and data structure for developing a Web Application (WebApp). The WebApp of our research group for individual risk assessment of the rice farmers has been developed by JavaScript for the front end, while the back end is run by Python. The WebApp was evaluated satisfactions by over 370 farmers from three public hearings. The average satisfaction scores are over 4 to a maximum of 5 in all categories.

Keywords Crop Damage Area Assessment, Farmer Relief Fund, Disaster, Flood

AMS 2010 subject classifications 91G70, 86A32

DOI: 10.19139/soic-2310-5070-2595

1. Introduction and Literature Review

Rice is one of the economic crops of Thailand that, in addition to being a staple in every household, generates an immense export value as presented in Table 1 below concerning the export values of agricultural products during 2017-2021 (source from the Bank of Thailand [1]).

*Correspondence to: Wiboonsak Watthayu (Email: wiboonsak.wat@kmutt.ac.th). Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi. 126 Pracha Uthit Road, Bang Mod, Thung Khru, Bangkok, Thailand (10140).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

Table 1. Export Values of Agricultural Produces during 2017–2021 (Unit: Million Baht)

Agricultural Produce	2021	2020	2019	2018	2017
Rice	107,758.36	116,044.96	130,584.56	182,081.67	175,160.78
Rubber	175,977.88	108,900.20	128,490.41	147,343.37	204,556.41
Cassava	93,818.08	58,092.42	55,268.21	72,674.45	71,795.74
Fruits	179,895.21	120,257.35	106,249.75	79,331.87	73,505.08
Others	108,142.00	108,324.00	95,658.29	101,035.47	87,706.04
Export Value	**665,591.53**	**511,618.93**	**516,251.22**	**582,466.82**	**612,724.05**

However, most rice farmers in Thailand are considered low-income earners as a result of existing farming problems, including product pricing and natural disasters (floods and droughts in particular), which affect farming, yield, and price. They have to suffer the damage themselves and receive insufficient remedy from the government due to communication limitations and delays in assessing the damage.

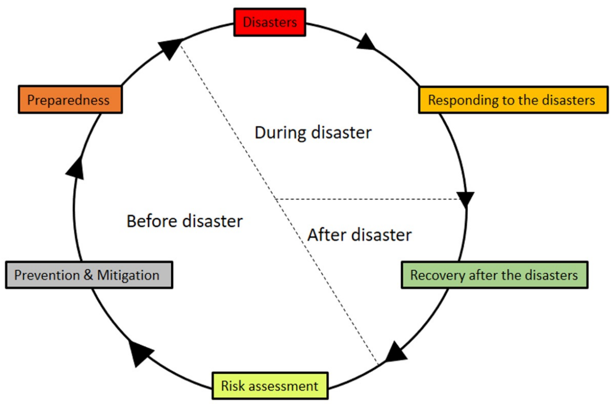

Figure 1. Disaster Management Plan
Source: Disaster Risk Mitigation towards Sustainability Development [14]

In the past, the management approach for disasters prioritized the rescue, lifesaving, house reparation, distribution of provisions, and compensation to the disaster victims to be able to recover financially. The response to the disasters and recovery after the disasters are considered passive measures. However, according to the experiences on disaster response and upon consideration to the damage and understanding of disaster cause, many countries around the world admitted that the effects caused by the disasters could be prevented and prepared in advance to mitigate the severity before an actual disaster occurs by following the principle of disaster risk management (DRM), covering from the preparation before, during and after the disasters. Three additional steps that can be done before the disaster occurs are Risk assessment, Prevention & Mitigation, and Preparedness as presented in Figure 1 (data from [14]).

The Risk assessment comprises 3 steps: risk identification - identify the nature and detail of the disaster, including the location and affected area, analyze the frequency and estimate the severity of the disaster; risk analysis - identify what may be risked, such as buildings, physical items, and society, by assessing the potential for responding to the disasters and estimating the damage; and risk estimation - arrange the frequency or probability, and severity of the disasters in order from low to high to estimate the risk. For example, a disaster with low probability and low severity results in low risk, while a disaster with high probability and high severity results in high risk. The statistical data or forecasting model may be used to support the estimation. Therefore, the disaster risk estimation is meant to identify the option to prevent, mitigate and prepare for the disasters.

Using the data from the risk estimation, Prevention & Mitigation are to determine the directions of the disaster responses, divided into 2 types of measure: structural measure, such as the technological and material development for building construction in the area where the earthquake frequently occurs, rapid tsunami warning development, irrigation system to response to flood and drought, mangrove forestation serving as windbreaker; and non-structural measure, such as the establishment of water user cooperative for the purpose of water management, strengthening the bargaining power to the members, and supporting and mentoring various groups of farmers, and the establishment of reserve fund to support the suffered farmers and reduce illegal loans.

Apart from the aforementioned measure, the disaster responses also include the policy formulation and enactment to mitigate the disasters, such as to prohibit road construction blocking the waterway that might cause the flood, to prohibit deforestation, which prevents the landslide, and to promote the casualty insurance.

Preparedness focuses on the preparation for response and management during disasters, including evacuation drills in the community, a stable telephone or internet communication system development, up-to-date first aid kits, life-saving equipment, and medical or rescue vehicles to respond to the disasters, such as boats or helicopters.

For related works on disaster management on farming, Matchaya et. al. [9] investigated the impact of erratic rainfall, dry spell, and water logging in Zambia. The data was obtained from the Zambia Statistical Agency from 2017-2018, which covered 13,512 farming households. The data includes the number of family members, distance from the house to the market, gender, and age of the family head. The aim was to create empirical evidence on how rainfall, dry spells, and waterlogging affected agricultural productivity in order to guide agricultural risk management strategies and the role of investment at both the household and policy level. The authors further employed the Cobb-Douglas production function to measure productivity. The productivity consists of Maize, Groundnut, Sweet Potato, Sunflower, Soybean, and Mixed bean. The results show dry spell has the most negative effect on Groundnut, followed by Maize, and the least effect on Sweet Potato. Rainfall also has a negative effect on Maize and Groundnut.

To improve real-time flash flood warning accuracy in the riverside area in Hubei Province in China, Tu et. al. [15] developed a dynamic early warning threshold chart depending on 4 soil moisture levels, which are dry, normal, wet, and saturated, to fulfill empirical early warning, the current one. The results from 2020 to 2022 showed that 396 early warnings were generated by the empirical threshold, while 456 were generated by the dynamic threshold. Compared to the current one, dynamic thresholds were relatively high in May, giving fewer early warnings, but were relatively low in June and July. By taking the total rainfall into account, dynamical early warnings were generated in wet years more often, while the empirical early warnings were generated more often in dry years.

In [5], Holyoak managed over 400 telephone surveys for farmer perception with farm dam management in four states in Australia that have weak, moderately strong, increasingly strong, and very strong policy environments. The four states were South Australia, New South Wales, Victoria, and Tasmania. Demographics of farms, farmers, and dams have been surveyed and analyzed by a chi-square test of independence. Farmer perception data were analyzed and compared by means, and were examined by ANOVA. The results showed significantly different perceptions of farmers among the 4 different strength policies. For example, farmers in weaker policy states have more spillway blocking behavior than the strongest policy state.

By door-to-door household survey, Dulawan et. al. [4] gained data to both qualitatively and quantitatively analyze what criteria that people in high flood risk areas still lived in their places. The results showed that the reasons to remain in flood-prone areas are home attachment, environmental acclimatization, convenience of basic need accessibility, economic concern, and flood safety perceive. The researchers pointed out two-pronged strategies. The first involves the active participation such as local officers and affected residents so that their roles meet the demand of the community, while the second involves policy makers or urban planners to develop a relocation action plan that meets the economic and environmental needs of affected families to encourage relocation.

Panrungsri and Sangiamkul [12] employed a Business Intelligence Model by constructing a data warehouse which was obtained from the Department of Disaster Prevention and Mitigation in order to develop a Multidimensional Model to predict the damage of flood and landslide in Phuket, Thailand. The author used Power BI for data visualization to deliver their assessment results.

As a new policy of agricultural and flood defense by the use of catchments on River Laver and River Skell in Northern Yorkshire has arisen to protect the flood problem, Posthumus et. al. [13] aimed to survey the perception of stakeholders in the rural area. The total sample size was 92, which consisted of local farmers, representatives from government and non-governmental organizations. The samples have been asked for insight and perception during the workshop by using the Food and Agriculture Risk Matrix (FARM) as a decision support matrix, as well as using "Post-it" stickers. As a result, the participants agreed that the target areas that are affected by the flood need to be specific, encourage farmers to see the advantage of solving this problem, and that suitable measures and funding are needed.

By taking into account the official data from 2017 to 2021, Li et. al. [7] have obtained the 8 most influential factors that cause flash floods in the whole country of China, which were precipitation, elevation, slope, landform, soil texture, vegetation, land use, and population density. The risk assessment is evaluated by the model called Flash Flood Potential Index, which is based on geographical and topographical factors. The process called ArcGIS was also employed to locate the places on the map that experienced flash floods. By these tools, the authors were able to rank the "explanatory power" of each factor and found that precipitation and soil texture have the most two highest explanatory power, while population density and land use only gently influenced the flash flood.

By the aforementioned disaster management plan, in particular the disaster risk management, prevention, and mitigation, requires accurate and updated data of the area, types of disaster, and farmers. Therefore, the purposes of this research, under the sponsorship of the Basic Research Fund from the Thailand Science Research and Innovation or TRSI, are:

1. Develop the farmer database system for managing the Farmer Relief Fund for those suffering the natural disasters.
2. Develop a statistical model to evaluate risk of flood for individual farmers in the whole country.
3. Develop a web application, linking to the model in 2, that farmers can evaluate their risk of flood and can report the crop damage to provide the preliminary remedy to the farmers after the disasters.

The result obtained from the first purpose will be potentially useful for making the casualty policy and establishing the Farmer Relief Fund, while the result obtained from the third purpose will remind farmers to be aware of their own risk and to assess the rice field damage in order to provide preliminary support after the disasters. Later, the affected farmers can directly contact a relevant agency, which will approve the preliminary financial support for their rapid recovery.

This research project is divided into 3 phases in order to achieve the set purposes.

Phase 1: Data Collection and Management-This phase consists of:

1. Data collection plan with 5,000 samples.
2. Data management - prepare and clean the data. In this phase, all the data was collected, prepared, and cleaned.
3. The structured data was available for developing the model and web application.
4. Identification of variables related to the risk classification. In Phase 1, only directly related variables perceivable by the farmers themselves were identified (excluding the information related to the agencies, such as risk level of each area, which would be identified in Phase 2).
5. A suitable model for this phase was created and used as an approach in Phase 2 for creating the main database system model that required all data from the samples.
6. Phase 1 was completed, and the result from this phase was also applied to the other 2 phases. We presented the result and the ongoing process at the first public hearing held in Sale Phum, Roi Et, on 27 February 2022, and at the 4th National Conference on Science, Technology, and Innovation 2022, organized by the Faculty of Science and Technology, Loei Rajabhat University.

Phase 2: Creation of Risk Level Forecast Model
Phase 1 led to the development of a complete and accurate model in terms of data, statistical approach, and risk

level identification (subject to each individual farmer from 0 to 1) that could be adapted to the group level as required. In addition, the suitable variables were identified for further modification of the data structure in Phase 3.
Phase 3: Web Application Creation
This phase will focus on the creation of a web application by applying the results from Phases 1-2 (suitable model and database system) to develop on a larger scale in the future (big data) and assessing the functionality of the system. The result of this phase was presented at another public hearing held in Bang Pahan, Phra Nakhon Si Ayutthaya, on 28 May, followed by Warin Chamrap, Ubon Ratchathani, on 21 June 2022. Totally, there were over 370 farmers who participated in our public hearings. The results from Phases 1-3 were also presented in the 26th Annual Meeting in Mathematics 2022 (AMM 2022), Thailand.

2. Methodology

This research surveyed the data of farmers and rice fields, including statistics from several agencies, to analyze the risk and assess the damage caused by floods. It was preceded by the following procedures.

2.1. Data Collection (Phase 1)

In this procedure, we collected the data from farmers from 77 provinces nationwide, using the provinces where farmers have encountered flooding as a criterion. The samples were then selected from these provinces. We sought permission to collect data and collaborated with relevant agencies to obtain the specified data for the areas. It is worth noting that the surveys faced the difficulty of COVID-19 measures, which banned gatherings of more than 30 people. However, there was a measure of relaxation duration when we could conduct fieldwork to collect data by in-person interviews with farmers in Nakhon Ratchasima who suffered from the flood. Apart from this in-person survey, we collaborated with people in agricultural offices in rural areas across the country to manage the survey for us. The period of data collection lasted for 3 months, from 1 November 2021 to 31 January 2022.

2.1.1. Research Tool

The tool for this research was a questionnaire, which was improved and developed from the research title "A Study of Guidelines for the Establishment of the Farmer Relief Fund Based upon the Principles of Crop Damage Assessment" [16]. It passed an evaluation by experts. The Index of Item–Objective Congruence (IOC) stood at 0.80–1.00, suggesting that the contents of the questionnaire were appropriate for obtaining essential data for this research.

As it is a study succeeding the research, the questionnaire consisted of 3 parts. (The full questionnaire (in Thai) can be accessed by the following link, https://drive.google.com/file/d/1zg7UNQT2PoEGwI4bUjBs-xNdEUwl8VOW/view?usp=sharing)

The questionnaire consists of 3 parts as follows:

- Part 1: Basic Data of Farmers
- Part 2: Data on Farming and Disaster-Related Problems
- Part 3: Opinions on the Establishment of the Farmer Relief Fund

Parts 1 and 2 were essential for the development of a web application to assess the crop damage so as to give preliminary support to the farmers after the disasters, as well as for the development of a farmer's database system.

2.1.2. Population

The population in this study was farmers who had faced floods in 77 provinces throughout the country. The data is obtained from Agricultural Research Development Agency during 2017-2022. The stratified random sampling was then conducted, using each region of the country as a stratum. The sample size was determined using the Yamane table with a 1.4% error. As a result, the overall sample size was 5,000 farmers. The calculation for the sample size of each region and province followed the principle of proportional allocation as shown in Table 2.

Table 2. Population Suffered from Flood in Each Region

	Region (Stratum)	Population and Sample		Percentage (%)
		N_i	n_i	
1	North	52,126	872	18.00
2	Northeast	131,392	2,198	44.00
3	West	18,651	312	6.00
4	Central	75,978	1,271	25.00
5	East	6,036	101	2.00
6	South	14,705	246	5.00
	Total Population (N)	**298,890**	**5,000**	**100.00**

2.1.3. Scope of the Study and Preliminary Assumptions on the Population

1. The population in this research was farmers who had encountered the flood only and were classified as vulnerable.
2. The main economic crop in this research was rice.
3. The natural disaster encountered by farmers was a flood.
4. The variables adopted by this study were as presented in Table 3.

This research considered only the variables that related directly to the farmers and were issues perceivable by the farmers themselves. It has yet to include information from relevant agencies, such as the risk level of each area.

4.1 Independent variables (R) comprised both quantitative and grouping (qualitative) variables, including:

4.1.1 Variables related to farmers and data on rice farming – for example, income and damage.

4.1.2 External variables perceivable by farmers – for example, problem conditions of disasters and disaster-prone seasons.

4.2 Dependent variables (Y) were 2 levels of risk.

Table 3. Variables in This Study

Variable	Meaning	Variable	Meaning
Y:	Risk Level Encountered by Farmers as a Result of the Flood	Level 1:	Low Risk Level
		Level 2:	High Risk Level
R1	Ratio of Income to Expense	**R5**	Total Damage
R2	Farmer Registration	**R6**	Total Number of Disasters
R3	BAAC Membership	**R7**	Number of Disasters in 3 Years
R4	Ratio of Income to Farming Expense	**INF**	Risk during Disaster-Prone Seasons

The theory of 2-group classification indicated 0.5 as a cutting score [3], whereas

$$P(\text{occurrence}) < 0.50 \text{ was classified as } Y = 1, \text{ meaning a low level of risk}$$

$$P(\text{occurrence}) \geq 0.50 \text{ was classified as } Y = 2, \text{ meaning a high level of risk}$$

The determination of the classification cutoff or the cutting score could be adjusted to suit each research. This research studies the fundamental factors that influence the risk confronted by each individual farmer, which could be different. Therefore, the average risk opportunity of each person had to be classified by the degree of risk for accurate classification. Importantly, if the risk was used as part of the calculation of the premium for the insured party, the classification must be accurate as well as just for all parties.

The fundamental risk levels adopted by this research rested upon "A Study of Guidelines for the Establishment of the Farmer Relief Fund Based upon the Principles of Crop Damage Assessment" [13]. It was found that the average risk level of the farmers was 0.60, close to the preliminary assessment of this research. Before deciding, we conducted a test and compared between the cutting scores of 0.50 and 0.60. It was found that the percentage of correct predictions stood at 85.33 and 87.24, respectively. Therefore, it was decided that the value under 0.60 would be considered the low level of risk, and the value at or above 0.60 would be considered the high level of risk.

2.2. Data Management (Phase 1)

To obtain the qualified data, the data management required a quality process, consisting of data preparation and data cleaning. The data collected from the farmers were in paper-based as well as Google Form formats. A great deal of variety of the data was prone to cause problems with its quality and structure, which would affect the development of the farmer's database system and the model for the development of a web application in the next stage. Therefore, the data preparation and data cleaning were necessary and consumed a considerable amount of time. The problems arising out of the data ranged from outliers, duplicate values, missing values, over-necessity, invalidity, to ambiguity. Importantly, the variety of the data necessitated structuring and standardizing for common use, including the development of a program that employed the machine learning technique in the next stage. The said data must be complete with regard to both its quality and its structure. In short, the data preparation and data cleaning procedures of this research were as follows:

1. Prepare a coding book to standardize the data utilization and preliminary data structuring.

2. Specify the definition of quality data, study problems of data collection, and find causes that lead to unqualified data.

3. Proceed to statistical data cleaning [8] in response to the problems with the R and the Excel VBA programs.

4. Structure the data as per the specification so that it is ready for common use.

5. Use the data to build the model for risk prediction and select the suitable variables.

6. Obtain the suitable model, which would also result in obtaining the appropriate variables for re-structuring the data.

7. Examine the data for developing the farmer database system and build a web application.

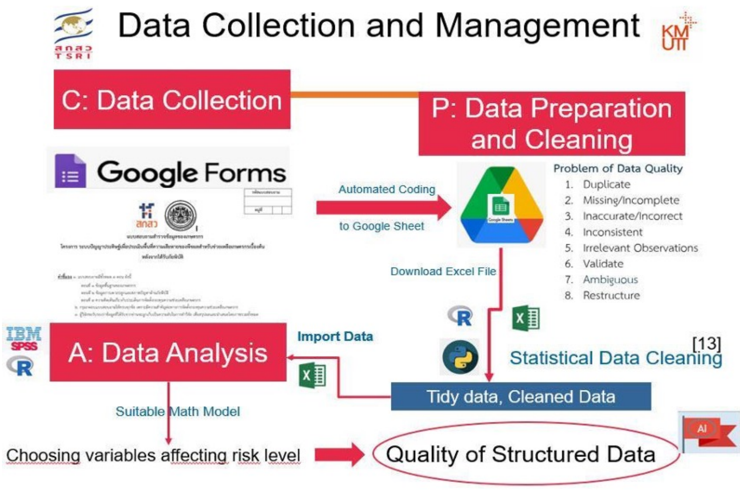

Figure 2. Process and Result of Data Management

2.3. Data Modelling for Risk Assessment (Phase 2)

The analysis of statistical data was divided into 2 parts as follows:

Part 1: Basic Data of Farmers – The data was analyzed by descriptive statistics, which involved finding the frequency and percentage, in order to examine the amount of data in each studied group, e.g., farmers' risk groups, growing seasons, etc.

Part 2: Farmer's Risk Assessment - This involved 2 methods of statistical analysis.

1. The chi-square test was conducted to examine the correlation between the variables and risk level of encountering flood, creating the model in the next statistical method.

2. The binary logistic regression analysis [6] was conducted for model creation and variable selection. The suitable model had to pass the significant evaluation in all of its procedures with a statistical significance value of 0.05. The accuracy of the model was considered from the percentage of correct predictions, which must stand at or above 85.

There are many statistical methods for classifying data into 2 groups, but the two most widespread methods are discriminant analysis and binary logistic regression analysis. There are many studies that compare the two methods. In most studies, it is found that the binary logistic regression analysis is more efficient for data classification [3, 11]. For this research, we therefore decided to adopt the binary logistic regression analysis. The other reasons that were factored into the decision were that the independent variables consisted of both quantitative and qualitative data, and that there were only 2 dichotomous variables. Moreover, the binary logistic regression analysis has fewer conditions than the normal regression analysis and the discriminant analysis.

2.4. Developing Web Application (Phase 3)

The Web Application (Webapp) has been designed and developed after obtaining an appropriate statistical model. The Webapp is designed to be easy and friendly for users. From the survey, the farmers prefer not to install any new or unfamiliar Webapp on their smartphones. So, we decided to link our browser to LINE OA as it is the most common application among all farmers in the same rural area for the whole country.

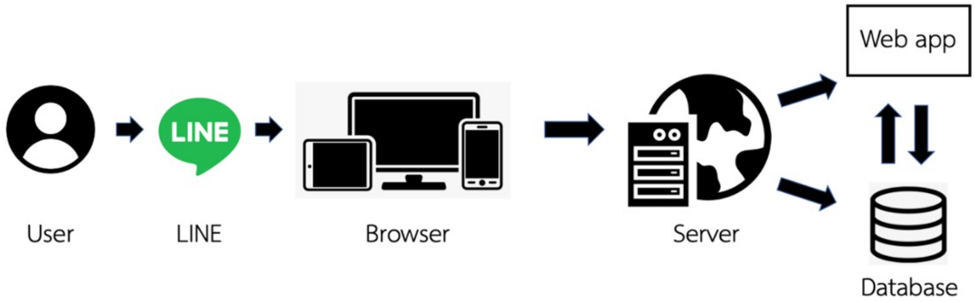

Figure 3. The linkage of the Web Application to users

Step of working process of web application are detailed as follows:

1. Users insert area information (tumbol, amphoe, province), income, monthly expense, farmer registration, BAAC membership, crop information and record of flood.
2. The system saves data into the database system.
3. The system calculates risk of flood from the inserted data.
4. The system saves risk level into the database system.
5. The system shows risk report to the user. The user can also recall all the record to see if needed.

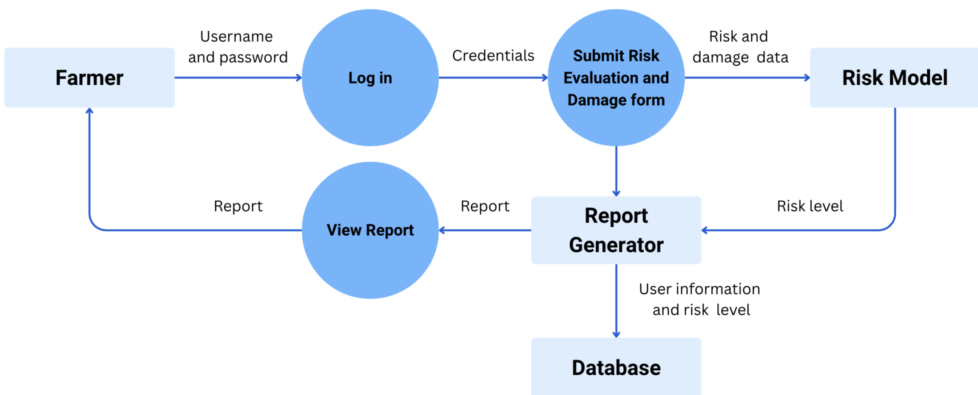

Figure 4. Data flow diagram of the system

The Web Application is developed in three parts as follows

1. Frontend
   This part is developed by HTML, CSS and JavaScript by using React as the library.
2. Backend
   This part is developed in Python by using Django as the framework.
3. Database
   The database is a Relational Database (RD) which consists of the following tables:
   - The table of user data - to collect personal data of the user when registering.

- The table of province data - to collect related data of all provinces in Thailand
- The table of amphoe data - to collect related data of all amphoe (districts) in every province.
- The table of tumbol data - to collect related data of all tumbol (sub-districts) in every amphoe.
- The table of pictures - to collect the pictures the farmers use to evaluate damage.
- The table of risk evaluation data - to collect related data, the user inserts into the Web Application for risk evaluation.
- The table of crop data for evaluation - to collect crop data to calculate the risk evaluation.

3. Results

Part 1: Essential Data

1. Sample size of the farmers suffered from flood (Table 4).
2. Risk level of the suffered farmers (Table 5).

Table 4. Sample Size of Farmers Suffered from Disasters in Each Region

Region	Frequency	Percentage
North	872	17.44
Northeast	2,198	43.96
West	312	6.24
Central	1,271	25.42
East	101	2.02
South	246	4.92
Total	**5,000**	**100.00**

Table 5. Risk Level of Suffered Farmers

Group	Risk Level	Number of Sample	
		Frequency	Percentage
1	Low	645	12.90
2	High	4,355	87.10
Total		**5,000**	**100.00**

Part 2: Selection of Variables and Risk Level Prediction Model

The Chi-Square test was employed to select the independent variables potentially influencing the risk level before creating the model to estimate the risk level of rice farmers suffering from the flood.

Step 1: Selection of Preliminary Independent Variables for Model Creation
1.1) The Chi-Square test was employed to verify the preliminary correlation between the independent variables potentially influencing the risk level. The hypotheses and results are presented below:

H_0 : No correlation between the variables
H_1 : Correlation between the variables

Table 6. Chi-Square Test of Dependent and Independent Variables

	Variables	Mean	Std Dev	Min	Max	χ^2	**p-value**	**C.C.**
R1	Ratio of Income to Expense	1.25	0.5	0.4	3.2	297.43	0.000*	0.337*
R2	Farmer Registration	0.88	0.33	0	1	0.89	0.346	0.013
R3	BAAC Membership	0.65	0.48	0	1	0.06	0.784	0.004
R4	Ratio of Income to Farming Expense	1.1	0.4	0.3	2.8	108.13	0.020	0.145*
R5	Total Damage	15200	8300	0	60000	958.24	0.000*	0.401*
R6	Total Number of Disasters	2.3	1.2	0	6	128.67	0.000*	0.158*
R7	Number of Disasters in 3 Years	1.5	0.9	0	4	155.63	0.000*	0.174*
INF	Risk during Disaster-Prone Seasons	0.72	0.45	0	1	219.67	0.000*	0.205*

* Reject H_0 = correlation between the variables.

It was found that 5 independent variables influenced the risk level of the farmers, given χ^2, the contingency coefficient (C.C.), and $p - value < 0.05$. The binary logistic regression analysis was then conducted to verify the results before creating the risk estimation model. However, these 5 variables still could not meet the objectives of this research. Therefore, the proper model had to be created to:

1. Select the desired variables and re-verify the results;
2. Estimate and classify the risk level;
3. Confirm the accuracy of the prediction;
4. Use the model to further develop the system in other modules.

1.2) Model Creation Using Binary Logistic Regression Analysis

In the binary logistic regression analysis, the backward stepwise selection was employed to select the variables that are presented, along with other statistical values used to verify the model, in Table 7

Table 7. Variables in Binary Logistic Regression (Backward Stepwise Selection: Wald)

Variable	**Coefficient: B**	**e^B = Exp(B)**	**S.E.**	**Wald**	**p-value***	**Independent Degree**
Constant	1.495	4.459	0.259	33.339	0.000	1
R1	-2.462	0.085	0.236	108.488	0.000	1
R4	-0.625	0.535	0.311	4.039	0.044	1
R5	1.130	3.097	0.109	108.443	0.000	1
R7	-0.609	0.544	0.142	18.373	0.000	1
INF	2.685	14.654	0.224	143.119	0.000	1

According to Table 7, five independent variables were selected as presented. The dependent variables were significantly predicted. The test used the Wald statistic (Column 5), given that the p-value < 0.05 for every variable. The logistic regression coefficients $(b_0, b_1, \ldots, b_n)$ (Column 2) reflected the influence of an additional independent variable, resulting in the variation of $\log(\text{Odds})$. Therefore, the logistic regression coefficients were converted to e^B (Column 3) for better interpretability. However, in this research, we focused on using $b_0, b_1, \ldots, b_n$ to predict the dependent variables as presented in the following equations:

$$P(\textit{Occurring Probability}) = \frac{1}{1 + e^{-W}} \tag{1}$$

whereas,

$$W = 1.5 - 2.46(R1) - 0.63(R4) + 1.13(R5) - 0.61(R7) + 2.69(\text{INF}) \tag{2}$$

To identify and classify the risk levels, $R1$, $R4$, $R5$, $R7$, and INF were required in Equation (2), resulting in W, which was used in Equation (1) to calculate the probability, varied from 0.00 to -1.00 (as presented in Figure 5). The results, indicating a high risk level, were later classified to calculate the percentage of correct predictions as presented in Table 9.

The logistic regression coefficients provide insights into how different factors influence farmers' flood risk classification. The negative coefficient for the ratio of income to expense (R1) indicates that farmers with stronger financial resilience are less likely to be classified as high risk, as higher income relative to expenses reduces vulnerability. Similarly, the ratio of income to farming expense (R4) also carries a negative coefficient, suggesting that more profitable farming operations buffer against disaster impacts. In contrast, the total reported damage (R5) has a strong positive effect, meaning that higher past damage substantially increases the likelihood of high-risk classification, underscoring the predictive power of previous loss experience. The number of disasters experienced in the past three years (R7) shows a negative coefficient, which is somewhat counterintuitive. This may reflect adaptive responses, government support, or possible biases in self-reported data, and thus warrants cautious interpretation. Finally, the variable indicating risk during disaster-prone seasons (INF) has the strongest positive effect, suggesting that seasonal vulnerability perception is a direct and significant predictor of high-risk status. Taken together, these results demonstrate that both economic resilience and experiential factors play important roles in shaping farmers' flood risk levels, while also revealing areas where data limitations may influence interpretation.

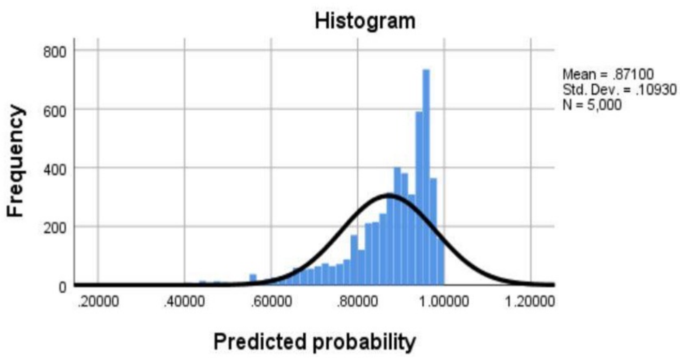

Figure 5. Predicted Probability of Risk Level

Step 2: Suitability Verification of Risk Level Assessment Model

The verification results for the risk level assessment logistic regression model, including key statistics from the omnibus tests of model coefficients and model summary, are presented in Table 8. The omnibus test examines whether the predictors included in the logistic regression model significantly improve the model's fit compared to the null model. Based on the $p - values < 0.001$, it was confirmed that the predictor variables significantly improved the model's fit. Also, the -2 Log Likelihood value indicated a reasonable fit of the model to the data. Two additional descriptive measures of fit, as shown in Table 8, were Cox and Snell's $R2$ and Nagelkerke's $R2$. These measures indicated that the model explained approximately 18.90% to 20.60% of the variance in the outcome variable, which is typical for logistic regression models in social and behavioral sciences. Thus, Equation (2) was suitable for prediction.

Table 8. Verification of Risk Level Assessment Model with Key Statistics

Omnibus Tests of Model Coefficients				Model Summary		
Step 4	**Chi-square**	**df**	**p-value**	**-2 Log likelihood**	**Cox & Snell R-Squared**	**Nagelkerke R-Squared**
Step	-0.134	1	0.71			
Block	466.61	5	0.00	3378.207	0.189 or 18.90%	0.266 or 26.60%
Model	466.61	5	0.00			

Step 3: Percentage of Correction Prediction of Risk Level Prediction Model

Focusing on the correct prediction of the suitable model, this research prioritized the percentage of correct prediction as presented in Table 9. The table provided a detailed breakdown of the performance of a risk level prediction model in terms of its ability to predict the low and high risk levels outcomes. The total number of predictions across both risk levels was 5,000 instances. Using the binary logistic model, the success rate is 6.36% and 99.20% respectively for the low and high risk levels. The overall percentage of correct predictions, according to Table 9, was 87.24%.

Table 9. Percentage of Correct Prediction of Risk Level Prediction Model

Observed Value: Y	**Predicted Value: Y**		**Correctness**
	Low Risk Level: 1	**High Risk Level: 2**	
Low Risk Level: 1	41	604	6.36%
High Risk Level: 2	34	4321	99.20%
Precision $= \frac{4321}{4321+604} \times 100 = 87.72\%$		**Percentage of Correct Prediction**	$\frac{41+4321}{5000} \times 100 = 87.24\%$
Recall $= \frac{4321}{4321+34} \times 100 = 99.22\%$		$F_1 = 2 \times \frac{0.8772 \times 0.9922}{0.8772+0.9922} \times 100 = 93.12\%$	

From the confusion matrix in Table 9, the positive class is the high-risk level, and the negative class is the low-risk level. The precision is 87.72%, i.e., out of all the predictions where the model flags someone as high-risk level, 87.72% are correct. The recall is 99.22%, that is, the model is able to detect 99.22% of all actual high-risk individuals, which is crucial in risk management or healthcare settings. The F1-score of 93.12% confirms that the model achieves a strong balance: it not only identifies high-risk cases accurately (high recall) but also does not falsely label too many low-risk cases as high-risk (decent precision). It is especially useful when both false positives and false negatives carry consequences.

Part 3: Web Application

The Web Application for farmers' risk assessment can be accessed from LINE OA, and some working instructions are presented as follows:

The Web Application consists of 2 main parts, which are:

3.1. User Part

This part consists of 5 functions as follows:

3.1.1. Registration for new users – To register, the users are required to insert all real data and create a password.

3.1.2. Signing in for registered users – Users can log in with their citizen ID and the password that was created when registering for the first time.

3.1.3. Risk Evaluation and Damage Submission – To evaluate the risk of flood, users need to insert their own farming information, tumbol, amphoe, province, income, monthly expense, Farmer Registration, BAAC Membership, crop information, and record of flood history together with a photo of the farm (if exists). Once finish inserting all the information, the Web Application will calculate risk into 3 levels based on Equation (1). Although in Phase 2, we divided the risk level into 2, which are "low" when the probability of Equation (1) is

0–0.6 and "high" when the probability is $0.61 - 1$. For a more detailed report on WebApp users, we divided the range $0 - 0.6$ of low risk from Phase 2 into two levels as detailed below:

- Low risk: probability $0 - -0.3$
- Medium risk: probability $0.31 - -0.6$
- High risk: probability $0.61 - -1$

In every evaluation, the system will save the results, which can be recalled for checking at any time. The pseudo code of this step is shown in appendix A.

3.1.4. Follow-up and Record Checking – Users are able to recall all the risk evaluation records that have been done earlier.

3.1.5. Changing Personal Data – User can change personal data, but is not able to change the record of risk evaluation.

3.2. Admin Part

This part consists of 5 functions as follows:

3.2.1. Logging in by admin - The Web Application Browser has specific icon at menu tab for the admin to log in. Admin is requested to insert the citizen ID as the account name and password when registering.

3.2.2. Listing all the users - Admin can see the details of the user from the screen. For convenience, the admin can find a record of the user by using the "search" function.

3.2.3. Recalling risk evaluation record - Admin is able to observe details of the risk evaluation. Admin is also able to "search" the evaluation details in each province.

3.2.4. Adding disaster type - Admin is able to add the type of disaster in this function and is able to search disasters by type from the search function.

3.2.5. Adding crop type - Admin is able to add the type of crops that are planted by the farmers in each rural area. The type of crop is able to be search from the "search" function too.

Part 4: Web Application by Public Hearings

The research team organized 3 public hearings which are at (1) Sale Phum, Roi Et on 27 February 2022, in (2) Bang Pahan, Phra Nakhon Si Ayutthaya on 28 May, followed by (3) Warin Chamrap, Ubon Ratchathani on 21 June 2022. Across these three public hearings, a total of 374 unique risk assessments were completed and used to test satisfaction. The satisfaction is divided into categories, all of which are measured by the satisfaction scores as follows:

- 1.00 – 1.80 Very unsatisfied
- 1.81 – 2.61 Unsatisfied
- 2.62 – 3.42 Neutral
- 3.43 – 4.23 Satisfied

- 4.24 – 5.00 Very satisfied

Table 10. The evaluation of Web application and public hearing management

List of satisfactions	P.H. (1)	P.H. (2)	P.H. (3)
Category 1: Evaluation of Web Application operation efficiency			
1. Reliability in real application	3.92	4.21	4.15
2. Understanding the purpose of the Web Application	3.61	4.13	3.97
3. Knowing the benefit of the Web Application	3.83	4.13	4.00
4. Easy to use	3.92	4.14	4.23
5. Speed of the Web Application operation	4.14	4.11	4.33
6. Nice to use and front end is beautiful	3.94	4.18	4.38
7. The Web Application has all the lists of working purposes	3.97	4.14	4.31
8. The Web Application is secured and stable	3.86	4.11	4.46
Overall satisfaction in Category 1	**4.00**	**4.12**	**4.51**
Category 2: Evaluation of public hearing satisfaction			
1. The topic of public hearing is interesting and approachable	4.03	4.27	4.13
2. The place is appropriate	4.03	4.09	4.23
3. The presentation is easy to understand	4.00	4.09	4.31
4. Media and material during the public hearing	4.00	4.14	4.21
5. Clearness of answering by the expert	4.17	4.05	4.46
6. Appropriate time	3.92	4.08	4.23
Overall satisfaction in Category 2	**4.03**	**4.13**	**4.44**

We have improved the Web Application and public hearing management every time by the comments of the participants, as evidenced by the increase in satisfaction scores.

4. Conclusions

This research has a key objective to develop a farmer database system and model for predicting risk level of flood that damages rice farming in Thailand, and to develop a Web Application for farmers to evaluate the risk of flood by themselves and yet further submit their farm damage results to the government via this channel. The authors would like to conclude our project due to the main objective as follows:

4.1. Developing Database System

The research group scoped the population to be the farmers in 77 provinces across Thailand who experienced flooding during 2017-2021. By Stratified Random Sampling and together with Yamane at an error 1.4%, we obtain a sample size at 5,000 to survey. The main statistical process required the data management (data preparation and cleaning) to properly structure the data. The coding book was also upgraded to a data dictionary for further big data development. Therefore, the processed data was in the form of tidy data in every aspect.

4.2. Statistical Modelling for Risk Evaluation

In this phase, 5 variables statistically significant to the risk estimation were selected for improving the database structure to be used for web application creation in the next phase. The selected variables are R1 Ratio of Income to Expense, R4 Ratio of Income to Farming Expense, R5 Total Damage, R7 Number of Disaster in 3 Years and INF Risk during Disaster-Prone Seasons. The suitable model was also created by logistic regression for risk level estimation. The model predicts 87.24% accuracy.

4.3. Limitations

Despite the promising results, this study has several limitations. First, the dataset exhibited a significant class imbalance, with 87.1% of farmers classified as high risk and only 12.9% as low risk. This imbalance led to reduced sensitivity in detecting the low-risk group (6.36%), which may affect the robustness of the model for minority cases. Future work should consider resampling techniques or alternative algorithms that better handle imbalanced data. Second, the model did not incorporate geospatial or environmental variables, such as elevation, slope, precipitation, land use, and soil texture, which are widely recognized as important predictors of flood risk. Integrating such data, possibly through GIS and weather forecasting systems, could enhance the model's precision and applicability across diverse regions. Finally, the study relied heavily on self-reported farmer data collected via questionnaires. While practical for nationwide coverage, self-reported data are prone to recall errors, reporting bias, and subjective assessments of damage and expenses. Cross-validation with official records or agency datasets would help reduce these biases and improve the reliability of the model.

4.4. Developing Web Application

By the satisfaction evaluation from the 3 public hearings involving over 374 farmers, we found that the average score of every list is more than 4, when 5 being very satisfied. This information shows that the Web Application will most likely be one of the important channels to contact the farmers in the future.

5. Suggestions

5.1. Statistical Methods for Risk Estimation

5.1.1. The prediction correctness could be higher by considering other statistical methods for better comparison, such as the multinomial logistic regression analysis. Furthermore, the risk level prediction should be identified specifically for each farmer in order to develop other related research, such as premium calculation, urgent support, etc.

5.1.2. The variables should be quantitative and qualitative for consistency purposes. The related information obtained from any agencies should be considered as well for a more correct prediction.

5.2. Suggestions for Research Implementation

5.2.1. Web Application Design and Development: This research will lead to the sustainable system development as a result of statistically processed reliable data. Therefore, the suitable model and data are available.

5.2.2. Further Design and Development: Focusing only on rice, the most important economic crop, and the rice farmers who suffered from the flood, the system should be developed to support other crops and disasters as well.

5.2.3. Policy on Data Security and Privacy: For practical implementation, it is essential to establish comprehensive policies and technical mechanisms to safeguard data security and protect user privacy. These should include, but are not limited to, encrypted data transmission via HTTPS, role-based access control, and secure data storage utilizing hashed personal identifiers.

5.2.4. System Scalability Considerations: When deployed in practice, the system must be designed to handle large volumes of data efficiently. This requires adopting scalable architecture, such as cloud-based infrastructure, load balancing, and distributed databases to ensure high availability, fault tolerance, and responsiveness as the system expands.

Appendix A

Function: Risk Level Calculation

Step 1: Collect Input Data

farming_information
income
monthly_expense
farming_income
farming_expense
total_damage
number_of_Disasters_in_3_Years
risk_during_disaster_prone_seasons

Step 2: Calculate Variables

$R1 = \min\left(\frac{\text{monthly\_expense}}{\text{income}}, 1\right)$

$R4 = \min\left(\frac{\text{farming\_income}}{\text{farming\_expense}}, 1\right)$

$R5 = \text{total\_damage}$

$R7 = \text{number\_of\_disasters\_in\_3\_Years}$

$INF = \text{risk\_during\_disaster\_prone\_seasons}$

Step 3: Calculate Probability (Equation 1)

$W = 1.5 - 2.46R1 - 0.63R4 + 1.13R5 - 0.61R7 + 2.69INF$

$\text{probability} = \frac{1}{1+e^{-W}}$

Step 4: Determine Risk Level

$$\text{risk\_level} = \begin{cases} \text{Low}, & 0 \leq \text{probability} \leq 0.3 \\ \text{Medium}, & 0.3 < \text{probability} \leq 0.6 \\ \text{High}, & 0.6 < \text{probability} \leq 1 \end{cases}$$

Return: $(\text{risk\_level}, \text{probability})$

Acknowledgement

This research has been sponsored by the Basic Research Fund 64000154 or TRSI. And we would like to express our appreciation to the Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi, for facilitating this research.

Conflicts of Interest:

The authors declare no conflicts of interest.

REFERENCES

1. Bank of Thailand, *Value and Amount of Goods Exported Categorized by Production Activities*, 2022. [Online]. Available: https://www.bot.or.th/App/BTWS_STAT/statistics/BOTWEBSTAT.aspx?reportID=747&language=th (in Thai) (accessed on March 5, 2022).
2. S. Chaudhuri and U. Dayal, *An overview of data warehousing and OLAP technology*, ACM SIGMOD Record, vol. 26, no. 1, pp. 65–74, 1997.
3. B. Demosthenes and P. Panagiotakos, *A comparison between Logistic Regression and Linear Discriminant Analysis for the Prediction of Categorical Health Outcomes*, International Journal of Statistical Sciences, vol. 5, pp. 73–84, 2006. ISSN 1683–5603.
4. J. M. T. Dulawan, Y. Imamura, H. Amaguchi, and M. Ohara, *Social Drivers of Flood Vulnerability: Understanding Household Perspectives and Persistence of Living in Flood Zones of Metro Manila, Philippines*, Water, vol. 16, no. 799, 2024.
5. J. L. T. Holyoak, *Water sharing risk in agriculture: Perceptions of farm fam management accountability in Australia*, Agricultural Water Management, vol. 145, pp. 123–133, 2014.
6. J. M. Hilbe, *Logistic Regression Models*, Chapman and Hall/CRC, 2009, 656 pages.
7. Q. Li, L. Zhao, Z. Zhang, Y. Wang, and M. Ma, *Comprehensive Risk Assessment Framework for Flash Floods in China*, Water, vol. 16, no. 616, 2024.
8. M. van der Loo and E. de Jonge, *Statistical Data Cleaning with Applications in R*, 1st Edition, John Wiley and Sons, 2018, 300 pages.
9. G. C. Matchaya, G. Tadesse, and A. N. Kuteya, *Rainfall and crop productivity in Zambia: Implication for agricultural water risk management*, Agricultural Water Management, vol. 269, 107648, 2022.
10. D. McFadden et al., *Quantitative methods for analyzing travel behavior of individuals: Some recent developments*, Institute of Transportation Studies, University of California, 1977, p. 307.
11. M. E. Montgomery, M. E. White, and S. W. Martin, *A comparison of discriminant analysis and logistic regression for the prediction of coliform mastitis in dairy cows*, Canadian Journal of Veterinary Research, vol. 51, no. 4, pp. 495–498, 1987.
12. T. Panrungsri and E. Sangiamkul, *Business Intelligence Model for Disaster Management: A Case Study in Phuket, Thailand*, In Proc. of the 14th ISCRAM Conference, May 2017, France, pp. 727–738.
13. H. Posthumus, C. J. M. Hewett, J. Morris, and P. F. Quinn, *Agricultural land use and flood risk management: Engaging with stakeholders in North Yorkshire*, Agricultural Water Management, vol. 95, no. 7, pp. 787–798, 2008.
14. Public Safety Canada, *Emergency Management Planning Guide 2010–2011*, 2011. [Online]. Available: https://www.publicsafety.gc.ca/index-en.aspx (accessed on March 4, 2022).
15. Y. Tu, Y. Zhao, L. Meng, W. Tang, W. Xu, J. Tian, G. Lyu, and N. Qiao, *Study on Dynamic Early Warning of Flash Floods in Hubei Province*, Water, vol. 15, no. 3153, 2024.
16. W. Wathayu and P. Varnakovida, *A Study of Guidelines for the Establishment of the Farmer Relief Fund Based upon the Principles of Crop Damage Assessment*, Research Report, Land Development Department (Public Organization), 2019. (in Thai)
17. W. Zhong, J. Fan, and C. Li, *A framework of applying BI to social security systems*, In *Proc. of the International Conference on Intelligent Computation Technology and Automation (ICICTA)*, October 2008, China, pp. 189–193.