Machine Learning Application on Area Yield Rice Insurance Development Process

Krish Sethanand[1]
Technopreneurship and Innovation Management Program, Chulalongkorn University.
krish.sethanand@gmail.com
Associate Professor Thitivadee Chaiyawat, Ph.D.[2]
Chulalongkorn Business School, Chulalongkorn University.
thitivadee@cbs.chula.ac.th
Chupun Gowanit, Ph.D.[3]
Technopreneurship and Innovation Management Program, Graduate School, Chulalongkorn University.
Chupun@gmail.com

Abstract

Crop insurance in Thailand is quite limited to rice and maize and is designed as indemnity-based insurance which claim compensation is made under losses from pests and all natural risks such as flood, drought, dry spell, windstorm, frost, hailstorm and fire. However, economies of scale and efficiency are being argued due to high administrative costs since damage assessments are conducted on the use of manpower. In addition, indemnity payment of crop insurance in Thailand is criticized since farmers receive claim payments only based on official declarations of disaster areas. Pay-out is made to rice and maize farmers for total losses as inspected by local government in declared calamity areas, thus adding basis risk to the claim position. Ineffective and shortfall of assessment contributes to farmers' misperception in agricultural insurance and leads to adverse selection and moral hazard problems.

Area yield crop insurance has possessed a variety of prominent benefits and preserved effort incentives for insurers as no individual farmer can influence the probability of an insurance pay-off. Area yield insurance is also immune to problems of adverse selection. This paper therefore aims to develop yield agricultural insurance, at a provincial level, focusing on rice insurance scheme. This is because rice is the preeminent economic crop in Thai agriculture. However, major damage to rice crop is natural disasters resulting from flood or drought. The production risks are from both high frequency rate and a wide range of cultivation areas.

Crop yield prediction is an essential task for successful area yield agricultural insurance. However, efficiency of crop yield prediction is one of the challenging problems in precision agriculture. Machine learning is a practical approach that can provide better yield prediction. Therefore, this study uses machine learning technique in predicting growing season rice production yield at a provincial level. Three machine learning algorithms - polynomial regression, multivariate polynomial regression, and random forest – are adopted in this analysis. This study uses 10-fold cross validation to analyze the most suitable algorithm for this existing dataset.

Primary data are collected from 107 farmers in Ubon Ratchathani province cultivated area via an in-depth interview from rice-plant experts. Farmers in Ubon Ratchathani plant Jasmine105 rice which acclimates to moderate water stress. Three risk factors which are flood depth (cm), flood duration (weeks), and life cycle of the rice during the flood are suspected to have an impact on growing season rice yield.

The result of the study reveals that all risks have a negative correlation to the rice production especially flood depth and flood duration. This is because the higher water level and the longer flooding time length the rice paddy is ruined. The result of machine learning shows that every degree (2^{nd}, 3^{rd}, and 4^{th}) of polynomial regression gives an acceptable fitting with R-Square around 75-90%. The 2^{nd} and the 3^{rd} degree of multivariate polynomial regressions show a better fit compared with single polynomial regressions. However, the 4^{th} degree of multivariate polynomial regression shows a lower precision. This may explain the induction of multi-collinearity among multiple features. The 3^{rd} degree multivariate polynomial regression appears to be the suitable algorithm to avoid underfitting and overfitting. Anyhow, random forest is the most efficient tool and exhibits enhanced prediction accuracy with a very low error - 0.04 of MAE and 0.005 of MSE. This may be because random forest combines the output of multiple

decision tree to generate the final prediction which can avoid overfitting. Therefore, area yield index crop insurance as proposed by this study can be a solution to enhance the economic viability and sustainability of the Thai rice industry.

1. Introduction

Natural disaster, especially floods, affects rice production in Thailand and costs government budget for farmers loss compensation. During 2005-2014, 1-5 million hectares of cultivation area were damaged. Thai government has established 1,000-4,000 million baht each year for natural disaster relief to help farmers. The incident of great flood in Thailand happens in 2011 caused the damage of cultivation areas up to 9,300,000 rai and the release of financial relief was up to 30 billion baht.

Agriculture has been exposed to crucial risks – uncertainties in weather and climate sensitivity, plant diseases, and pests – affecting productivity. Varity approaches has been adopted to manage agricultural risks, e.g., crop diversification, off-farm activities (Salazar César et al., 2019), and agricultural insurance. In Thailand, there are a number of crop, livestock, and aquaculture insurance programs – a type of risk management tool – issuing in a format of indemnity-based agricultural insurance for rice, maize, livestock, and prawn. Crop insurance in Thailand is quite limited to rice and maize and is designed as indemnity-based insurance which claim compensation is made under losses from pests and all natural risks such as flood, drought, dry spell, windstorm, frost, hailstorm and fire. However, economies of scale and efficiency are being argued due to high administrative costs since damage assessments are conducted on the use of manpower. Second concern about this current agriculture scheme is a time-consuming claim payment process where crop damage claim settlement and payment may take more than 3 months.

In addition, indemnity payment of crop insurance in Thailand is criticized since farmers receive claim payments only based on local official declarations of disaster areas. Pay-out is made to rice and maize farmers for total losses as inspected by local government in declared calamity areas, thus adding basis risk to the claim position. Ineffectiveness and shortfall of assessment may contribute to farmers' misperception in agricultural insurance and can lead to adverse selection and moral hazard problems. Rice is Thailand's leading harvested crop ahead of the other Thai agricultural products. Rice farming is prevalence across the entire country and Thailand harvest areas of rice is about 11.19 million hectares (Napasintuwong, 2019). In addition, rice is the most country's important crop. However, in terms of the harvested yield Thailand has a lower average rice yield compared with other large producing rice countries such as Vietnam, Philippines, Myanmar, and Cambodia (Global Yield Gap Atlas, 2021). The contrast between growing area and production arises from several components. One of the crucial threats is attributable to natural disasters which impose an impact on production risks creating both high frequency rate and a wide range of cultivation areas. An indemnity-based rice insurance scheme in Thailand covers the largest insured farmland area over other agriculture types. However, only half of rice planted areas are in the rice insurance program while the rest of the farmers are not interested in participating although Thai government subsidizes insurance premium. This is due to basis risk, adverse selection, and moral hazard affecting not a high-level rice insurance penetration rate and causing some portions of farmers arguing to participate in this current rice insurance program. To improve a resounding successful agricultural insurance program in Thailand under weather and climate uncertainty more effective and efficient agricultural management practice through higher accuracy and fast speed in risk and loss assessment should be implemented.

Technology, e.g., satellite imagery, drone, or farmer photography has been utilized on agricultural insurance in this present day to assess damage of cultivated areas. Data collection and processing technology such as machine learning - a branch of Artificial Intelligence (AI) - is a practical approach that can process a large amount of data (big data), analyze the input data (disaster data), and contribute the output data (damage or indemnity payment). Therefore, this study uses machine learning technique in predicting growing season rice production yield. Machine learning method has been efficiently used for predicting automatized damage outcome. This should help to improve higher accuracy and better performance in yield prediction and additionally should help to reduce administrative cost and time of loss assessment process.

2. Literature Review

2.1 Agricultural insurance model

There are various models of agricultural insurance. Each model has a specific method of damage assessment and indemnity evaluation. Model selection on development of agricultural insurance for each crop and area has factors to be considered to ensure that the model has a reliability of indemnity payment scheme for farmers and produce business potential for the insurers as well as is consistent with the government regulation (Georgievich, 2019). Insured under a traditional crop insurance typically receives an indemnity from insurer when the crops are damaged by drought, hail or frost. Index-based model has increasingly used in many areas of agricultural insurance. Weather index insurance – a type of indexed-based insurance - uses the weather condition for damage assessment. Indemnity is based on realizations of a specific weather parameter measured over a prespecified period at a particular weather station. The insurance can be structured to protect against index realizations that are either too high or too low which is expected to cause crop losses (World Bank, 2011). Another type of index-based insurance - yield index insurance - uses crop production for indemnity payment while farmers will receive the same compensation per area size (Wen Chen et al., 2017). Index-based insurance benefits insurer on insurance administration management (Choudhury et al., 2016). The compensation evaluation of the index-based model is more viable than indemnity-based insurance as index-based insurance takes into account several levels of damage factors to match crops and perils. This is suitable for a single peril crop insurance model which indemnity is paid when the particular index meets the specified condition (Biffis et al., 2017).

Index-based model has been implemented on agricultural insurance in several countries. Specified indexes are used for assessing the damage caused by natural disasters or determining the indemnity payment. Basis risk is significantly reduced, and the model is suitable for the weather conditions that change over time. Index-based model also benefits crop insurance to be improved in terms of premium management and compensation funds (Collier B. et al., 2009). Index-based model requires historical data to evaluate the damage assessment therefore data collection is needed. Index-based model usage can be described as follows:

- Area yield insurance uses the level of crop production as a measure of indemnity payment evaluation. The insured yield is established as a percentage of the average yield for the area. An indemnity is paid if the realized yield for the area is less than the insured yield regardless of the actual yield on a policyholder’s farm. This area yield insurance contract allows to significantly reduce moral hazard and adverse selection. However, this type of index insurance requires historical area yield data (World Bank, 2011).
- Weather index insurance uses climate condition for damage assessment. Data recorded by weather stations are evaluated for the specified damage level.
- Remote sensing data from the satellite identify plant health indexes or take the photo of farmland to determine the damage level or disaster condition.

2.2 Technology for data collection and analytic

Agricultural insurance models are improved according to the development of cultivation (Wenjun Zhu et al., 2019). Index-based models are designed to use data processing technology implementation instead of manpower. One major limitation is that index-based insurance contracts are subject to basis risk that means some residual risk is left with the insured farmer. Basis risk arises because the index used is not perfectly correlated with yield losses on each farm. Therefore, special attention should be paid to quality and sampling data and the design of the index. Digital Technologies can help overcome this basis risk issue by improving yield prediction. Therefore, digital technologies can facilitate and help to increase related parties’ satisfaction and improve demand for agricultural insurance (Ceballos et al., 2019).

Enrico Biffis and Erik Chavez (2017) studied satellite image utilization in conjunction with machine learning to identify shortages of rainfall and overheating of farmland to design agricultural insurance for drought comparing of the cultivation period from the beginning, middle, and harvesting periods. Climate conditions in the area that changed over time can be assembled as big data for helpful analytics. Benefits of implementing technologies to agricultural insurance based on big data are increasing damage and indemnity assessment efficacy, data collection and processing proficiency, reducing cost of data handling and manpower, as well as helping insurers to effectively perform risk assessment and loss ratio management (Swiss Re, 2019). In present day, data mining technique has been extensively used in various agricultural data utilization, including crop-health, weather, and yield data (Mangani and Kousalya, 2019). While deep learning utilizes data to implement an innovative numerical model for loss assessment and indemnity payment prediction. Regression analysis is effectively utilized in real-time applications for insurance payout with multi-features input (Mangani and Kousalya, 2020).

Machine learning implementation on agricultural insurance has input data prepared to predict continuous numerical values in the case that output is area yield or indemnity payout level. Training on the labelled data opts to supervise learning (Mahesh, 2019). Regression is defined to predict continuous values. There are several algorithms to deal with a variety of variable properties in the dataset and solve the problem by various forms of regression (Sinha, 2013). Simple linear regression and polynomial regression solve the problem with a one-to-one relationship between independent and dependent variables. Multiple linear regression, multivariate polynomial regression, support vector regression, decision tree, and random forest can handle more complex problem that has many independent factors to describe dependent variable (Ray, 2019). Model with multiple independent variables, e.g., multiple linear regression gives smaller prediction error and has better fitting compared to simple linear regression method. Using R-square figure shows that polynomial regression has noticeably better curve fitting than one of linear regression (Maulud, 2020). Support vector regression, generalized from support vector machine concepts, is an effective tool for continuous value prediction (Awad and Khanna, 2015). It has advantages on high dimensional of features (Chris and Kaufman, 2000). Random forest, a collaborative of decision tree (Prajwala, 2015), has been extensively used for classification, prediction, and regression. Random forest provides efficient learning with accuracy and avoids overfitting (Liu et. al., 2012).

3. Data and Methodology

3.1 Data

Primary data are collected from 107 farmers in Ubon Ratchathani province cultivated area via an in-depth interview from rice-plant experts. Thailand has a diversity of cultivation area. Each location has a different landscape and grows a variety of rice species. Therefore, this study selects Ubon Ratchathani province, lying in lower northeastern Thailand, where farmers plant Jasmine105 rice which acclimates to moderate water stress. Three risk factors which are flood depth (average water level exceeding normal) (cm), flood duration (flooding length of time) (weeks), and life cycle of the rice during the flood are suspected to have an impact on growing season rice yield.

This study pursues a semi-structured questionnaire to assess the value of 14 expert opinions in agricultural insurance sector in the context of area yield and risk in growing rice. The questionnaire covers the content of rice characteristics, crop modelling, risk management, peril, risk impact on rice cultivation, and agricultural insurance management. This paper recognizes the importance of a focus group interview and discussion method of data collection. Therefore, focus group meeting has been commenced with more than 100 participants to collect more relevant information.

3.2 Methodology

Agricultural insurance development process consists of several steps including crop identification, model selection, business model analysis, model testing and implementation. (RAJ Roberts (FAO), 2005, Blair et al. (ASEAN), 2017, GlobalAgRisk, 2009, Scheuing and Johnson, 1989). Service innovation development process can be applied to manage groups of development steps with internal processes and tools. The paper proposes the steps of rice insurance model development with clear procedures that could be appropriately applied to every type of agricultural insurance model with various types of disasters, crops, cultivated area, etc. Four processes are identifying, designing, developing, and implementing (IDDI). Details are as follows:

Step 1 Identifying

This step defines the specific components of the agricultural insurance model. The essential details of the insurance model to be created are as follows: 1) insured crop(s): single crop or multiple crops 2) disaster type(s): single peril or multi perils 3) cultivation area (each location faces different types of disasters, crops, and other related information) 4) model selection and specified index of the selected model 5) technologies and tools to be utilized in the model development 6) related parties in the insurance model scheme.

Step 2 Designing

Second step is designing structure of an insurance model with the elements defined in step 1 to create a flow diagram to represent input, process, and output to assess final model.

Step 3 Developing

Third step is developing a prototype of the insurance model generated in step 2 to implement technical or practical tools. This may be in the form of software or a policy that is consistent with the process of the insurance model.

Step 4 Implementing

Final step is implementing a new insurance model derived from the practical tool in step 3, collecting result and evaluating model efficiency to enable further development or improvement in any part of step 1, 2, or 3 until the finalization of the model as shown in Figure 1.

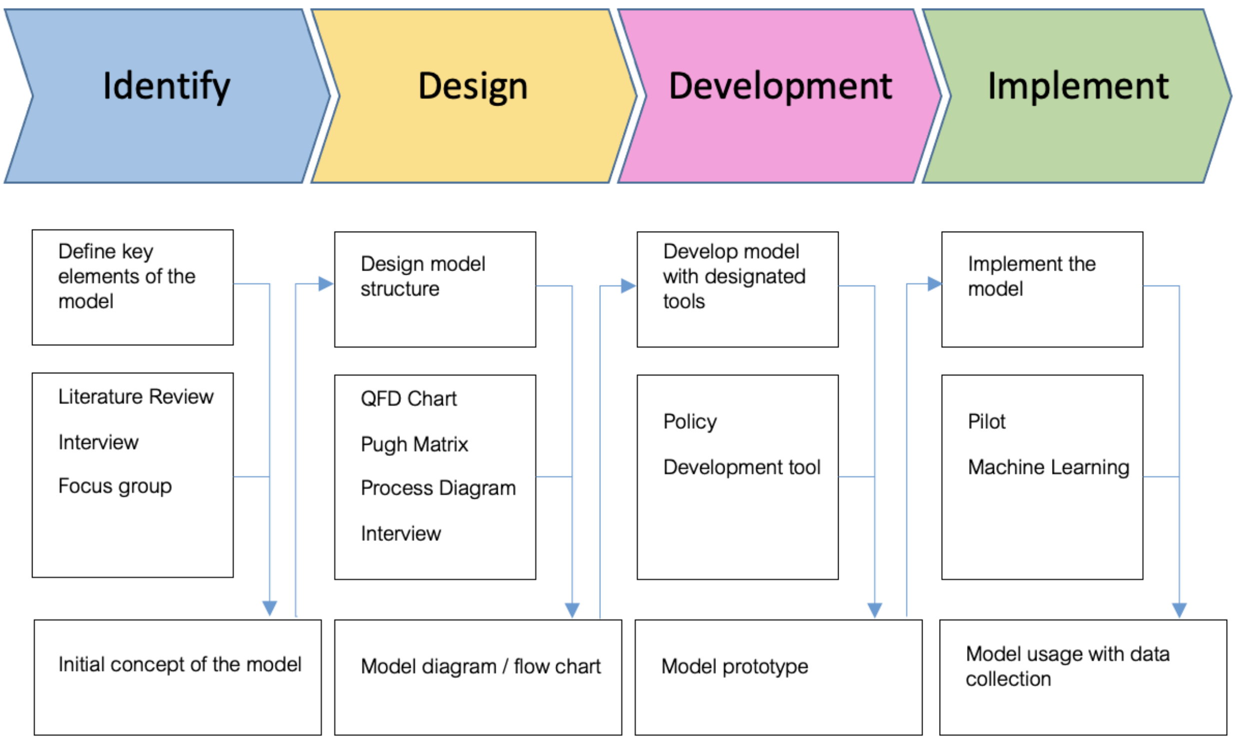

Figure 1: Identifying, Designing, Developing, and Implementing (IDDI) Process for Area Yield Crop Insurance

This paper puts a spotlight on the second step of rice insurance model development.

Model algorithm evaluation

This study uses machine learning technique in predicting growing season rice production yield at a provincial level (which has not typically played a role in traditional Thai crop insurance). Three machine learning algorithms - polynomial regression, multivariate polynomial regression, and random forest – are adopted in this analysis. Polynomial regression is selected to avoid underfitting from high data variation and to capture non-linear relationships of the data. The model selects the factor with the highest correlation to rice production as an independent variable to see how well the single feature method performs. The second algorithm is multivariate polynomial regression relying on multiple factors which have an effect on rice production. Each factor results in rice production with different levels of correlation. Any classification factors will be scaled to numerical independent variables. Random Forest – machine learning technique – is used in this analysis to avoid overfitting which may happen to other decision tree-based algorithms. Support vector machine is neglected in this paper since it is commonly used on the dataset with the number of dimensions is much greater than the number of examples. Machine learning and data analysis was done via Python programming language. K-fold cross validation is implemented to analyze the most suitable algorithm for this existing dataset. Cross-validation divides the data set into k groups, one group of which is used as validation data set each time, and the remaining groups (k-1) are used as training data set, so that k models can be obtained then the average score of k models can be used as the result parameter. The optimal split ratio is dependently on the data, the task, and the model. The smaller split ratio (higher number of k groups) tends to be optimal as the number of data growth (Goutte and Larsen, 1998). This study uses 10-fold cross validation.

4. Result & Discussion

Most insurance encounters 2 problems. First issue is adverse selection, where the farmers have inequality information of yields and crop risks compared to the insurers. This causes most of the farmers with high-risk farming areas to have a willingness to apply for insurance, resulting in a high loss ratio. The second matter is moral hazard, which results in the growth of payout ratio (Miranda, 1991). The area yield approach observes the range of yield in the area rather than the specific yields of each farmer. Thus, this could reduce the moral hazard as well as the cost of damage assessment (M. Boyd et al., 2011). Area yield index crop insurance is interesting to be investigate since insurance payments are based on an index rather than loss adjustments calculated for each farm that is insured, operating and administrative costs are significantly lower than those of other types of agricultural insurance (Barnett et al., 2008). Therefore, it is vital to explore in-depth investigation of the evaluation and prediction of an indemnity payout.

Data collected from farmers show 3 risk factors that influence the level of growing season rice yield: flood depth (the average water level exceeding normality) (cm), flood duration (flooding length of time) (weeks), and the life cycle of the rice during the flood. Although rice plants require large amount of water during growth, flooding stress from high flood depth and long duration results in severe loss of the crop. The crop damage and production loss would be more severe on account of unpredictable changes in climatic conditions causing floods (Vergara et al., 2014). Therefore, average water level and flooding time length are in concern since the exceeding to the proper water level will cause a low or no rice production. The life cycle of the rice is classified as before and after the “booting stage” and is suspected to have an impact on rice production. The total number of 107 data are collected as shown in Table 1 to predict production magnitude at a provincial level.

Table 1: Example of Data Exhibits Water Level, Flood Duration, Rice State, and Rice Production

1	Type	Normal Water Level	Actual Production	Normal Production	Actual Water Level	Water Rate	Time Span	Rice State	Production Rate
2	มะลิ105	50	332.5	350	80	1.6	0.5	0	0.95
3	มะลิ105	50	315	350	100	2	0.5	0	0.9
4	มะลิ105	50	35	350	200	4	1	1	0.1
5	มะลิ105	50	52.5	350	200	4	1	1	0.15
6	มะลิ105	50	0	360	300	6	6	1	0
7	มะลิ105	50	108	360	150	3	2	1	0.3
8	มะลิ105	50	0	350	150	3	12	1	0
9	มะลิ105	50	0	350	200	4	12	1	0
10	มะลิ105	50	52.5	350	80	1.6	8	1	0.15
11	มะลิ105	50	0	350	200	4	2	0	0
12	มะลิ105	50	0	350	200	4	2	1	0
13	มะลิ105	50	0	350	200	4	4	0	0
14	มะลิ105	50	0	350	200	4	4	1	0
15	มะลิ105	50	70	350	150	3	1	0	0.2

Correlation Test is analyzed to measure a relationship between each pair of variables. This would give an insight how strong each factor would affect the rate of rice production with flood stress compared to normal production without floods. The correlation test is shown in Table 2 and Figure 2.

Table 2: Correlation Among Actual Water Level, Flood Duration, Rice State, and Rice Production Rate

	Actual Water Level	Time Span	Rice State	Production Rate
Actual Water Level	1.000000	0.356293	0.116471	-0.633464
Time Span	0.356293	1.000000	0.168450	-0.568716
Rice State	0.116471	0.168450	1.000000	-0.273262
Production Rate	-0.633464	-0.568716	-0.273262	1.000000

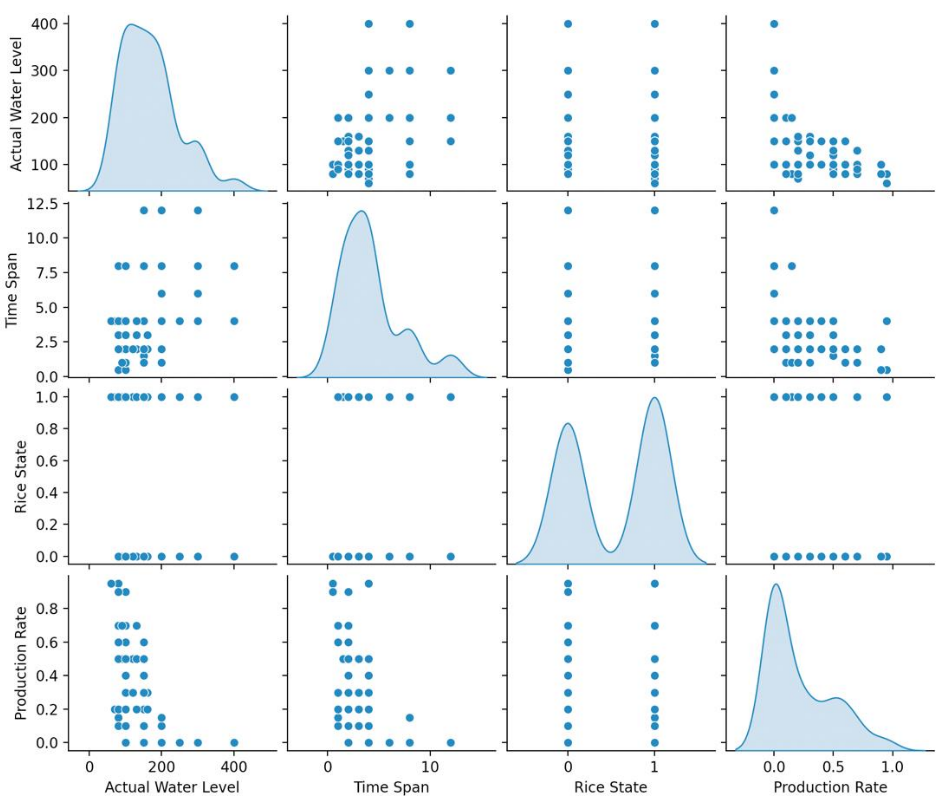

Figure 2: Correlation Among Actual Water Level, Flood Duration, Rice State, and Rice Production Rate

Figure 2 and Table 2 show correlation direction and coefficient among the relationship of actual water level with flood stress, flood duration, rice state, and rice production rate. The result of 0.356 correlation coefficient reveals a positive correlation between flood depth and duration convincing that the higher water level the longer flooding length of time. Rice state has a low positive correlation to the other two risk factors that intuitively means that rice state is not related to natural disasters. All factors especially flood depth and flood duration show a negative correlation to the rice production. The higher water level and the longer period the rice paddy under water the more severe rice plant would be damaged, thus lessening rice production. This is confirmed by correlation coefficients which are -0.633 and -0.569, respectively. Rice state data classified as before and after the booting state are scaled to 0 and 1. Rice state and rice production has a slightly negative correlation. This indicates that rice lifecycle after the booting state is more vulnerable to floods.

Figure 3 and Table 3 shows the result of polynomial regression where independent variable is flood depth due to the highest correlation with rice production. Polynomial regression had been tested on 2nd, 3rd, and 4th degree to analyze the view of a regression fitting and predicting rice production. To validate the model this paper splits data into train dataset and test dataset with a ratio of 9 to 1.

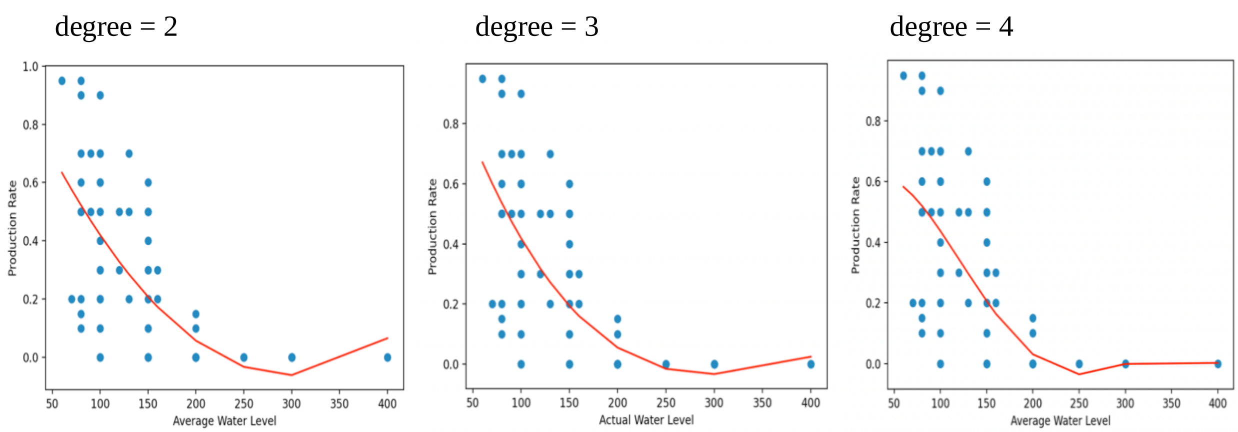

Figure 3: Polynomial Regression to Predict Rice Production Rate

Table 3: Polynomial Regression to Predict Rice Production Rate

Single Feature : Polynomial Regression	Single Feature : Polynomial Regression	Single Feature : Polynomial Regression
R-Square: 0.7317887399839189	R-Square: 0.7549412149987111	R-Square: 0.7463942128951213
MAE: 0.09642893569658297	MAE: 0.09127044092787293	MAE: 0.08441591541094108
MSE: 0.012590412866870583	MSE: 0.011503585940556367	MSE: 0.011904800584757945

The result shows that every degree of polynomial regression give an acceptable fitting with R-square around 75 percent. The 2nd degree probably is underfitting when water level is rising. The 3rd degree shows the best in fitting. However, the 4th degree may reasonably be the most appropriate curve for rice yield forecasting since it can correctly predict total rice production loss at high water depth with mean absolute error (MAE) of 0.084 and mean square error (MSE) of 0.011.

Multivariate polynomial algorithm is also used in this study to investigate the impact of all risk factors on the production loss. Multivariate polynomial algorithm includes 3 factors as independent variables to predict rice yield at 2nd, 3rd, and 4th degree. The results are presented in Table 4 as shown below.

Table 4: Multivariate Polynomial Regression to Predict Rice Production Rate

degree = 2

```
Multiple Features : Polynomial Regression
R-Square: 0.8744483710419328
MAE:  0.06066553153169368
MSE:  0.00589366324365142
```

degree = 3

```
Multiple Features : Polynomial Regression
R-Square: 0.8659119299263764
MAE:  0.06263399291122825
MSE:  0.00629438213238167
```

degree = 4

```
Multiple Features : Polynomial Regression
R-Square: 0.6737412804942561
MAE:  0.09391680241335347
MSE:  0.015315285345393599
```

The 4th degree of multivariate polynomial regression predicting correction drops compared to the model of 2nd and 3rd degree as seen by a reduction in R-square to 67 percent. This may explain the induction of multi-collinearity among multiple features or ill-conditioned. The 2nd and the 3rd degree of multivariate polynomial regression models show similar results, and it seems to be that it gives a better prediction compared to single polynomial regression. The 3rd degree multivariate polynomial regression appears to be a suitable algorithm to avoid underfitting and overfitting.

Table 5 shows the predicting result of rice production using random forest. Random forest shows very high accuracy and low error in prediction with MAE of 0.04 and MSE of 0.005. Random Forest combines the output of multiple decision tree to generate the final prediction to avoid overfitting.

Table 5: Random Forest to Predict Rice Production Rate

```
Multiple Features : Random Forest
R-Square: 0.8933732017097347
MAE:  0.0408079725829726
MSE:  0.005005291027179397
```

However, to confirm that random forest is more suitable algorithm over multivariate polynomial regression K-fold cross validation was done on 3rd degree multivariate polynomial regression and random forest to verify the performance of the algorithm in various splitting pattern of train and test dataset. The dataset was split into 10 folds (k=10). The average and the standard deviation of scores with 10 validations for both algorithms are shown in Figure 4. Results shown in Figure 4 indicates that values of R-square mean and standard deviation of multivariate polynomial regression and random forest are not much different. Multivariate polynomial regression gives slightly higher mean and lower standard deviation of R-square.

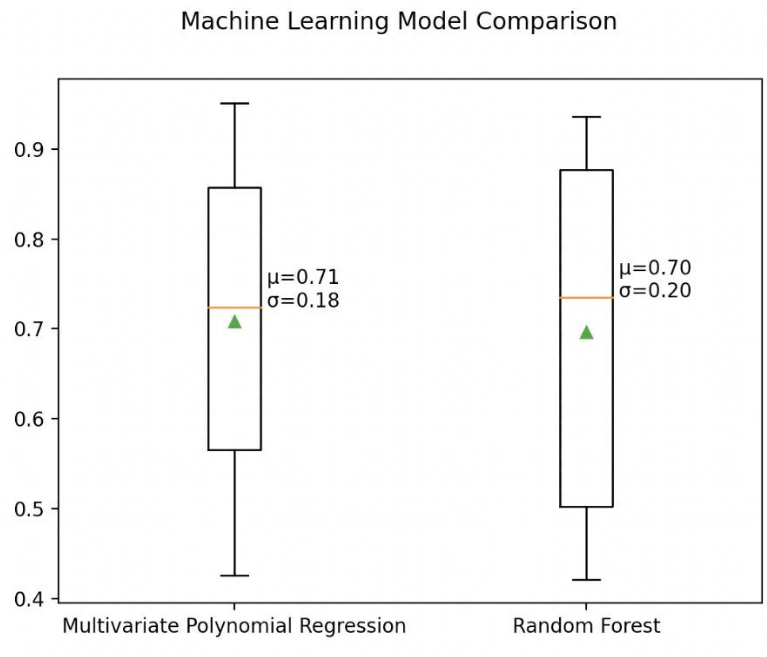

Figure 4: Average and Standard Deviation of R-Square with 10-Fold Validations for 3rd Degree Multivariate Polynomial Regression and Random Forest

5. Conclusion

Rice insurance development can be done through the IDDI process. The practice needs data collection to verify the scheme of the rice insurance consists of specific crop, peril, suitable insurance model, technology utilization, and coverage area. The rice insurance designed in this paper has been applied using machine learning to generate the model for rice production, at a provincial level, under flood stress. Random forest shows a very high accuracy and low error in predicting rice yield. However, in validation testing 3rd degree multivariate polynomial regression shows a similarity of R-square mean and error compared to those of random forest. To be able to correctly anticipate rice production loss would enhance a success in adopting area yield crop insurance in Thailand.

6. Reference

Askar Choudhury, James Jones, Adolph Okine, & Raquiba (Lena) Choudhury. (2016). “Drought-Triggered Index Insurance Using Cluster Analysis of Rainfall Affected by Climate Change.” *Journal of Insurance Issues*, 39(2), 169–186.

Awad, Mariette, and Rahul Khanna. *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. Springer nature, 2015.

Barnett, Barry J., Christopher B. Barrett, and Jerry R. Skees. "Poverty traps and index-based risk transfer products." *World Development* 36, no. 10 (2008): 1766-1785.

Boyd, Milton, Jeffrey Pai, Zhang Qiao, and Wang Ke. "Crop insurance principles and risk implications for China." *Human and ecological risk assessment* 17, no. 3 (2011): 554-565.

Chen, Wen, Roman Hohl, and Lee Kong Tiong. "Rainfall index insurance for corn farmers in Shandong based on high-resolution weather and yield data." *Agricultural Finance Review* (2017).

Collier, Benjamin, Jerry Skees, and Barry Barnett. "Weather index insurance and climate change: Opportunities and challenges in lower income countries." *The Geneva Papers on Risk and Insurance-Issues and Practice* 34, no. 3 (2009): 401-424.

Drucker, Harris, Chris JC Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. "Support vector regression machines." *Advances in neural information processing systems* 9 (1997): 155-161.

Enrico Biffis and Erik Chavez. "Satellite Data and Machine Learning for Weather Risk Management and Food Security." *Risk Analysis: An International Journal*, vol. 37, no. 8, Aug. 2017, pp. 1508–1521

Francisco Ceballos et al. "The Feasibility of Picture-Based Insurance (PBI): Smartphone Pictures for Affordable Crop Insurance." *Development Engineering*, vol. 4, Jan. 2019.

Georgievich, Kiselev Valeriy. "Statistical Modeling in Agricultural Insurance." In *2019 Twelfth International Conference" Management of large-scale system development"(MLSD)*, pp. 1-4. IEEE, 2019.

GlobalAgRisk. "Designing Agricultural Index Insurance in Developing Countries :" *A GlobalAgRisk Market Development Model Handbook for Policy and Decision Makers*. Lexington, KY: GlobalAgRisk, 2009.

Global Yield Gap Atlas. "Rice production in the five Southeast Asian countries, Vietnam, Thailand, Myanmar, Philippines, and Cambodia.", 2021

Goutte, Cyril, and Jan Larsen. "Optimal cross-validation split ratio: Experimental investigation." In *International Conference on Artificial Neural Networks*, pp. 681-686. Springer, London, 1998.

Laura Johnson Blair et al. "10 Phases in Developing a National Crop Insurance Program: Guide Overview" *Thirty Nine Meeting of the ASEAN Ministers on Agriculture and Forestry (39th AMAF)*, 28 September 2017, Chiang Mai, Thailand

Liu, Yanli, Yourong Wang, and Jian Zhang. "New machine learning algorithm: Random forest." In *International Conference on Information Computing and Applications*, pp. 246-252. Springer, Berlin, Heidelberg, 2012.

Mahesh, Batta. "Machine Learning Algorithms-A Review." *International Journal of Science and Research (IJSR).[Internet]* 9 (2020): 381-386.

Mangani, K. P., and R. Kousalya. "Designing weather based crop insurance payout estimation based on agro-meteorological data using machine learning techniques." *International Journal of International Journal of Recent Technology and Engineering (IJRTE)* 8, no. 3 (2019).

Mangani, K. P., and R. Kousalya. "Machine learning model for estimating agricultural crop insurance payout based on air temperature, rainfall, and relative humidity." *Agricultural Science & Technology (1313-8820)* 12, no. 2 (2020).

Maulud, Dastan, and Adnan M. Abdulazeez. "A Review on Linear Regression Comprehensive in Machine Learning." *Journal of Applied Science and Technology Trends* 1, no. 4 (2020): 140-147.

Miranda, Mario J. "Area-yield crop insurance reconsidered." *American Journal of Agricultural Economics* 73, no. 2 (1991): 233-242.

Napasintuwong, Orachos. “Rice Economy of Thailand.” *ARE Working Papers 284119*, Kasetsart University - Department of Agricultural and Resource Economics.No. 2193-2019-803. 2019.

Prajwala, T. R. "A comparative study on decision tree and random forest using R tool." *International journal of advanced research in computer and communication engineering* 4, no. 1 (2015): 196-199.

R.A.J. Roberts, "Insurance of crops in developing countries" *FAO Agricultural Services Bulletin 159*, Food and Agriculture Organization of the United Nations, Rome, 2005

Ray, Susmita. "A quick review of machine learning algorithms." In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, pp. 35-39. IEEE, 2019.

Salazar, César, Marcela Jaime, Cristián Pinto, and Andrés Acuña. "Interaction between crop insurance and technology adoption decisions: The case of wheat farmers in Chile." *Australian Journal of Agricultural and Resource Economics* 63, no. 3 (2019): 593-619.

Scheuing, Eberhard E., and Eugene M. Johnson. "A proposed model for new service development." *Journal of Services marketing* (1989).

Sinha, Priyanka. "Multivariate polynomial regression in data mining: methodology, problems and solutions." *International Journal of Scientific and Engineering Research* 4, no. 12 (2013): 962-965.

Swiss Re Institute, “Advanced analytics: unlocking new frontiers in P&C insurance”, *Swiss Re sigma No 4/2019*.

Vergara, Georgina V., Yudhistira Nugraha, Manuel Q. Esguerra, David J. Mackill, and Abdelbagi M. Ismail. "Variation in tolerance of rice to long-term stagnant flooding that submerges most of the shoot will aid in breeding tolerant cultivars." *AoB Plants* 6 (2014).

World Bank. “WEATHER INDEX INSURANCE FOR AGRICULTURE: Guidance for Development Practitioners”, November 2011

Zhu, Wenjun, Lysa Porth, and Ken Seng Tan. "A credibility-based yield forecasting model for crop reinsurance pricing and weather risk management." *Agricultural Finance Review* (2019).