ICIC Express Letters
Volume **17**, Number **1**, January **2023**

ICIC International ©2023 ISSN 1881-803X
pp. 1–11

PREDICTING KEY FACTORS IN AGRICULTURAL DATABASES FOR THAI FARMERS TO MAKE THE RIGHT DECISION

Chattrakul Sombattheera[1,*], Rapeeporn Chamchong[2]
and Phatthanaphong Chomphuwiset[2]

[1]Multiagent, Intelligent and Simulation Laboratory
Department of Information Technology, Faculty of Informatics
[2]POLAR Laboratory
Department of Computer Science, Faculty of Informatics
Mahasarakham University
Khamriang Sub-District, Kantarawichai District, Mahasarakham 44150, Thailand
{ rapeeporn.c; phatthanaphong.c }@msu.ac.th
*Corresponding author: chattrakul.s@msu.ac.th

Received May 2022; accepted July 2022

Abstract. *In recent years, Artificial Intelligence (AI) has been widely adopted to real world applications on several domains. The Department of Agriculture Extension (DOAE) of Thailand has courageously developed a large AI-based system, namely, Personalized Data (PD), in order to deliver suggestions to millions of Thai farmer in a convenient, affordable and timely fashion. While typical AI research expects experimental datasets and consumes a lot of time, this research investigates how such a system can deliver the required capabilities. There are three key factors, price, cost and yields of crops to be analyzed, in order to make it simple but meaningful to farmers. We found that artificial neural network with Multilayer Perceptron (MLP) and Random Forest (RF) models could effectively predict yield, cost, and price of crops. Adjusting parameters such as learning rate and the number of hidden nodes affect the accuracy of crop yield predictions. Smaller data sets required fewer hidden layers in model optimization. MLP models consistently produced more accurate yield predictions than RF models. MAPE (Mean Absolute Percentage Error) is used to measure both RF and MLP. It is found that MLP models produce accurate predictions. Similarly, RF is used to provide suggestions when time constraints are tight. We found that the accuracy is very high, and MAPE is around 5% in most cases. In some difficult scenarios where data is not adequate, the results are still good, and MAPE is around 20%.*
Keywords: Artificial neural network, Random forest regressor, Regression, Crop yield prediction

1. **Introduction.** The Department of Agriculture Extension (DOAE), Ministry of Agriculture, of Thailand is among the first government agencies in the country to adopt AI technologies to real world applications. DOAE has developed such a system, namely, Personalized Data (PD) system. The objective of the project is to bring about advances in AI to provide personalized information to help millions of farmers in the country make better decisions, including what crops to grow, where to sell, etc. The idea is to present simple, precise and succinct enough information to the farmers. DOAE has seven large databases, collecting data from real world environment over several decades.

These incur many challenges on various perspectives for the project, taking account of the norm of real world agriculture environment of the country. The first perspective is easiness for millions of farmers, whose majority are aged 40 or older. From easiness perspective, the challenge of the project is to make the information understandable, accessible and acceptable by farmers. The other perspective is time. While the computational

DOI: 10.24507/icicel.17.01.1

time for typical machine learning techniques is lengthy and the need for intensive computation is generally common, the computation time should be short, making the farmers feel that the application is relatively quick. The idea is that DOAE officers periodically analyze data for farmers in their respective areas by submitting requests via the system. The system must be able to execute rapidly and deliver reasonable accurate results. The last perspective is the correctness of the data. There could be a lot of incorrect and incomplete data. The new system should be able to work with existing system.

In order to make it easy for farmers, DOAE wants to simply focus on a few factors that make sense to farmers. DOAE names fifteen crops, including rice, corn, cassava, sugar cane, etc., in the nation's prime list. There are also a number highly prices crops, emerging in recent years. However, rice is generally the main crop that most farmers grow, spare a small amount of the yield for their own consumption and sell the rest of it as their main incomes. There are a few issues that the farmers need to know, including what plants they should grow for maximal yield, where to sell at the highest possible price and how much it might cost them. This relation can be put in the simplest mathematical form suitable to farmers, i.e., $\mathit{Profit} = (\mathit{Yield} \times \mathit{Price}) - \mathit{Cost}$. This looks very simple but can be very complex and tricky, taking account of the immense amount of data and other related issues.

This paper presents an attempt to tackle with the time and correctness issues – what would be the most appropriate algorithms that can cope with both issues. Instead of fancying advanced machine learning techniques, DOAE wants to use robust and fast AI techniques according to the above equation and then deliver the proper information to farmers.

The paper is structured as follows. We review related works, including available machine learning techniques and their usage in similar domains. We then discuss our real world environment. We present our methodology, including the architecture and major models of the system, datasets and cleansing data. We present the selected techniques used in the project. After that, we present our experiments and results. Lastly, we conclude the paper.

2. **Literature Review.** Having done preliminary research with respect to our constraints and purposes, we found that Multilayer Perceptron (MLP) [1], a fine-tuned technique of Artificial Neuron Network (ANN) [2], and Random Forest (RN) [1] are possible choices. In the following, we explore both techniques used in real world domains.

Throughout the years, multilayer perceptron has been widely adopted to agriculture. Wang et al. [3] predict sugarcane yield with highly accurate results. Ji et al. [8] deploy the ANN technique to predict rice yield in Fujian, China. Nosratabadi et al. [5] predict food production in Iran. Venkatesh and Thangaraj [6] help map suitable crops with soils. Al-Saif et al. [4] identify peach cultivars to help ensure high quality fruits. Pathane et al. [7] propose an ANN model to predict prices of agricultural commodities. Yuan and Ling [9] develop a software application to help farmers predict prices of crops. Silva et al. [10] develop an early warning system for facing the increase in prices of agricultural products. Jha and Sinha [11] use ANN to develop models for predicting prices of soybean in India.

Similarly, random forest has also been widely adopted to agriculture. The following works predict both yields and prices. Ramos et al. [12] predict maize yield production. Ok and Gungor [13] classify crop productions in Turkey. Jeong et al. [14] predict crop yield at regional and global scales. Mariano and Monica [15] map crop yield with spatial interpolation. Roell et al. [16] predict wheat yield in winter time. Ma et al. [18] predict prices for small and marginal farmers in India. Oktoviany et al. [19] predict prices of agricultural commodities. Brunda et al. [20] predict prices of multiple crops in India. Ghutake et al. [21] also use RF to predict crop prices in India with large datasets.

Although there are differences in detail about using MLP (or ANN) and RF in many countries, these works provide satisfactory results in their respective environments. We are to investigate their results in Thailand environment. Therefore, measuring the precision of forecasting is also important for evaluating the results. Among many techniques, Mean Absolute Percentage Error (MAPE) [17] is a popular technique for measuring the goodness-of-fit of ML results.

3. **Architecture, Internal and External Data Sources of the Personalized Data System.** Figure 1 depicts the architecture of this system. The system is composed of internal and external data sources, artificial intelligence system, executive system, configuration system an notification system. These components reside in application servers and mobile devices. The system is used by farmers and DOAE officers. Millions of farmers in Thailand are geographically categorized into seven layers. Officers are categorized into 4 levels, according to their roles and responsibilities. Internal data sources are seven major databases, including Farmer Registration (DB1), Digitized Farm Land (DB2), Current Crop (DB3), Disaster Victims (DB4), Large Farm (DB5), Efficiency Enhancement Centre (DB6) and Participating Farmer (DB7). These databases are running of different platforms, comprising a great deal of difficulty in composing a sensible system. The external data sources are public groups in faces, Talaad Thai (https://talaadthai.com, one of the largest e-market places of agricultural crops in Thailand), etc. Data collected from all sources are stored in DB9, the database of PD. Following MVC (Model View Control) design paradigm, the system uses Java, Python and React Native technologies plus M/L libraries to develop the system. There are hundreds of Java classes dealing with data model level, Java classes and Python for AI work in control level, and JSP (Java Server Pages) and React Native for view level. There are numerous intelligent agents, developed by Java, Python and shell scripts, coordinating components at different levels.

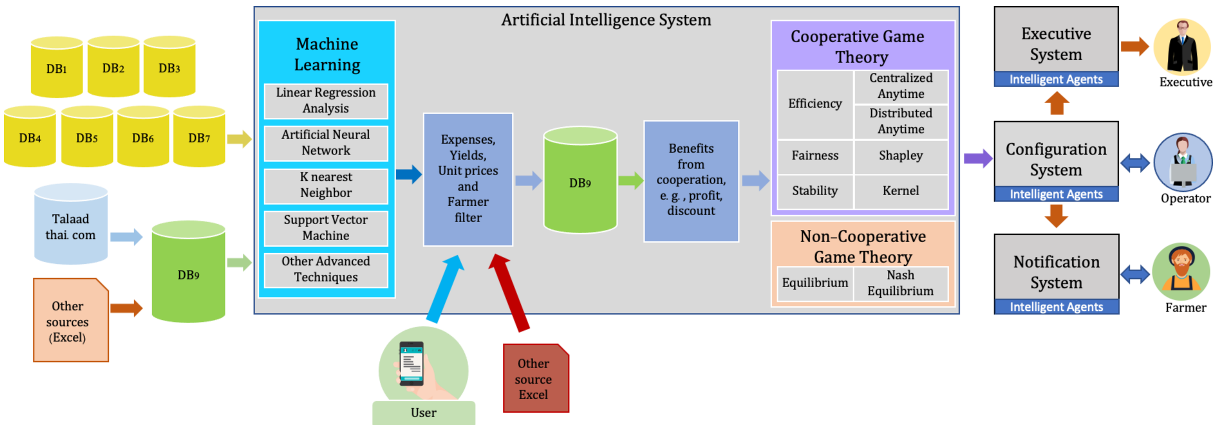

FIGURE 1. Architecture of personalized data system, developed and deployed at the Department of Agriculture Extension, Thailand

3.1. **Artificial intelligence system. 1) Machine Learning Module.** Machine learning [1] is used to forecast the three factors: yield, price and cost. There are two main sources of data used in the machine learning system. The internal sources are seven databases. External sources are 1) databases owned by outside organizations, and 2) excel files containing data. This paper is focused on the module, discussing how the analysis is done under real world environment. As shown in Figure 1, there could be many machine learning libraries added to this module. **2) Game Theoretic Module.** We use three concepts of cooperative game [22], efficiency, fairness, and stability, to help farmers make decisions. We use non-cooperative game [22] for strategic analysis.

3.2. **Configuration system: PO.** This web-based system allows operators at all levels around Thailand to configure how the artificial system and the notification system work. This system also allows operators to input data and monitor the results of their operations. At the first level, there are 8 menus available (altogether, there are 36 menus), such as News and Notification, Analyze, and Prepare Data. The most important one is Prepare Data menu, which allows for preparing data for ML analysis. DOAE officers can view histograms of data related to forecasting yields, prices and costs of crops. As previously mentioned, data can be incomplete and incorrect. Figure 2 shows a snapshot of the filter screen of rice cost. The overall distribution is shown in a bar graph. While unit cost of rice (x-axis) ranging from 0-4,100, the majority is within 0-1,500 Baht/Rai (0.395 Acre). The frequency range (y-axis) is 0-1,400. Other information includes mean (731.25), standard deviation (582.07), mode (800), median (700), minimum (50), and maximum (4,100). At the bottom of the screen, the user can use the sliding bars to specify the lower bound (left), and the upper bound (right) of the range.

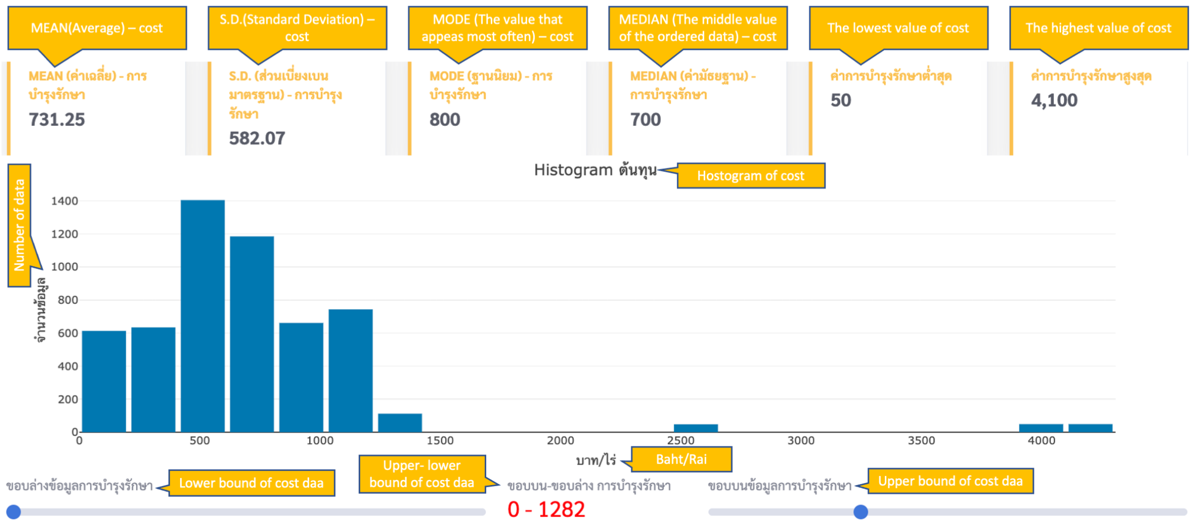

FIGURE 2. Filtering tool and histograms of DOAE application

There are also **Executive System** for presenting summarized information to executives at all levels, and **Notification System** for presenting analyzed data to farmers. However, we do not discuss about them here, because of limited space.

4. **Methodology.** Predicting factors for a specific purpose vary depending on unique circumstances. The study of precision of forecasting factors in agriculture [23] suggests that crop growth and yield models are based on a combination of soil, crop and climatic variables. In this project, we used different factors suggested by Office of Agricultural Economics of Thailand.

4.1. **Choosing attributes from real world data.** To make it acceptable to farmers, we only focus on three main factors: price, yield and cost. However, acquiring such data is an extremely tough task. DOAE is concerned with the security and privacy issues. We do our best to reveal as much as we can. As previously mentioned, DOAE databases were created so many decades ago when nobody back then would be able to imagine they would be used for machine learning purposes. Instead, the databases were created and extended to serve various purposes via different projects over the years. In DB1, there are almost one thousand tables of terabytes of data to look at. In DB2 and DB3, there are hundreds of them as well. In order to find specific pieces of information, we really need to look into hundreds of them and filter out the irrelevant ones. There are some limited documents but nothing really covers the whole picture of everything in the system. The most difficult part is that the existence of the information is not in one place. To find out about availability of water, for example, we need to look across three databases. There are several attributes used to predict yield, price and cost. They will be discussed in detail below.

4.2. **Dataset.** There are three kinds of prediction system on personalized data, which are 1) cost prediction, 2) yield prediction, and 3) sale-price prediction. Features in 1) cost prediction are water, disaster, soil, yield and sale-price, 2) yield prediction are water, disaster, soil, sale-price and cost, and 3) sale-price prediction are month, province, district, sub-district and export.

As this dataset is the real-world database, the main problem is the missing values and outliers. Therefore, any suitable model must be able to deal with these features. The significant amount of missing values and outliers for some key features made it particularly hard for prediction. In this study, missing values and outliers have to come up with strategies for solving this problem before users generate these models. The next subsection describes the pre-processing of cleansing data to cope with real-world database.

4.3. **Cleansing data.** Cleansing data is divided into two separate parts that are (i) outlier rejection and (ii) missing values for reliable prediction model. In our data, the outliers are consolidated mainly from human-errors, where the operators were not familiar with the conventional system and provided inaccurate data. In statistics, missing data, or missing values, occur when no data value is stored for the variable in an observation. Missing data are a common occurrence and can have a significant effect on the conclusions that can be drawn from the data. The goal of cleaning operations is to prevent problems caused by missing data that can arise when training a model. In this study, the record of data should be extracted from the dataset and replace the missing by using mean value of all data. In this study, we found that the outlier data sometimes has higher frequency value than the correct data. Moreover, there are a lot of missing data. If we used statistics for automatic decision, it will affect the correct data. For this reason of real-world database, visual application for user analysis should be developed to make a decision for selecting the range of correct dataset. To help officers filter out incorrect data, missing value and outlier, the system is equiped with visualized tools showing related attributes and histograms, as shown in Figure 2. After the cleansing data, the data normalization is to proceed by using Z-score. The overview picture of cleansing data for prediction is shown in Figure 3.

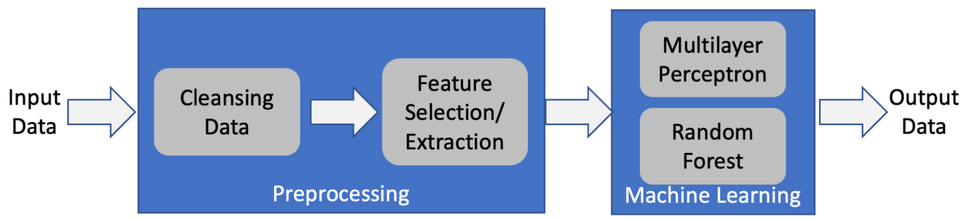

FIGURE 3. Overview of the prediction system on DOAE database

4.4. **Choosing right algorithms for real world data.** Knowing that there are incorrect and incomplete data in the databases, it would be better to start off with something reasonably solid. We have to choose algorithms that work well under this circumstance. There are two criteria to choose algorithms: i) time with acceptable quality, and ii) quality with acceptable time. We have primary experiments in order to choose merely two machine learning algorithms as a set of the system's original algorithms. Additional algorithms can be added on into the system's library at later stage. For the first criterion, Random Forest (RF) [24] suits our requirement well. It does not assume much of resources and take relative little execution time. It is suitable for quick results for a large amount of requests. For the second criterion, we allow more resources and time comsuption but pay more attention to quality. In this regard, Multilayer Perceptron (MLP) [1], a powerful empirical modelling approach and yet relatively simple compared with mechanistic models, is a good choice. The architectures of MLP and RF are shown in Figure 4. For the sake of completeness, we briefly review them below.

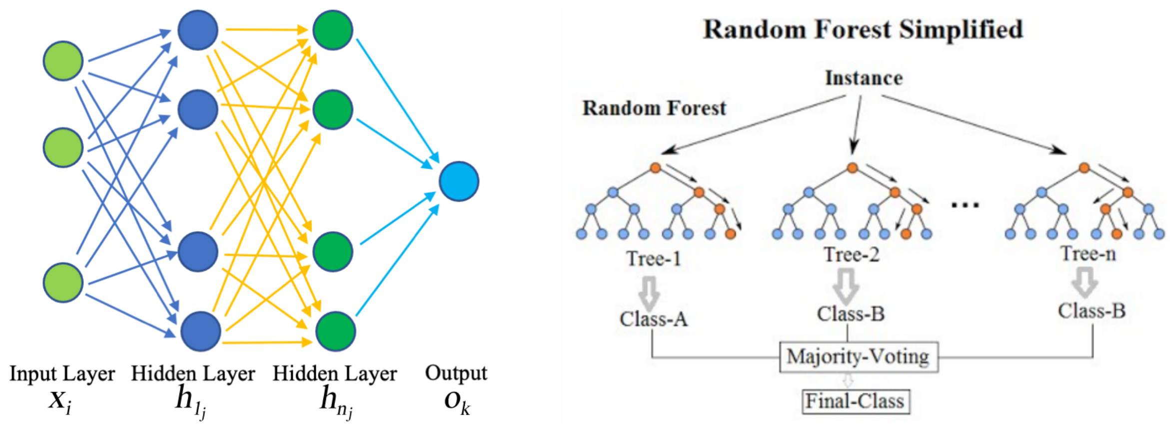

FIGURE 4. Multilayer perceptron (left) and random forest (right) architectures

4.4.1. *Multilayer perceptron.* Since MLP [1] is a type of ANN, MLP has layers of nodes. As shown in Figure 4 (left), x_i is the input nodes of input layer, h_{n_j} is the hidden nodes of layer n and o_k is the output nodes. To achieve better results, we simply have multiple layers of neuron cells. Weighted data are fed through input nodes, layered nodes and outputs in the end. The weighted sum of inputs at each node will be computed and pass on as output. The hidden node h_{1_j} at layer n that is replaced with $h_{1_j} = \sum_i w_{1_{ij}} x_i + b$ where h_{1_j} is hidden node j at layer 1, x_i is input node i, $w_{1_{ij}}$ is weight i at layer 1 and j, and b is bias. The hidden layer n is calculated by following equation: $h_{n_j} = \sum_i w_{n_{ij}} h_{(n-1)_i} + b$ where h_{n_j} is hidden node j at layer n, $h_{(n-1)_i}$ is input node i from prior hidden layer, $w_{n_{ij}}$ is weight at i and j at layer n, and b is bias. The non-linear activation function σ takes on the value h_{n_j}, controls the ranges of input values and generates the final output for the neuron.

4.4.2. *Random forest.* Among multiple benefits of RF, including high dimensionality, robust to outliers and non-linear data, handling of unbalanced data, low bias, moderate variance, RF is very fast, configurable to do parallelization. Random forest [24], as shown in Figure 4 (right), works by following bagging techniques. Let $X = \{x_1, \ldots, x_n\}$ be a training set. Let $Y = \{y_1, \ldots, y_n\}$ be a set of corresponding responses. Random forest repeatedly chooses (B times) a sample randomly, replaces with training set and inserts trees to the samples: Once training has been completed, unseen samples x' can be predicted by averaging each regression tree on x': $\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x')$. Bagging simply decorrelates trees, being correlated because many trees are trained by the same training set, by providing different training sets. In addition, the prediction uncertainty can be estimated as the predictions' standard deviation of all individual regression trees on x': $\sigma = \sqrt{\frac{\sum_{b=1}^{B}\left(f_b(x') - \hat{f}\right)^2}{B-1}}$.

4.5. **Measuring precision of results.** The measurement of this study is Mean Absolute Percentage Error (MAPE) [17] as $MAPE = \frac{\sum |Y_t - F_t|/Y_t}{n} \times 100$ where n is the size of sample, F_t is the value predicted by the model for time point t and Y_t is the value observed at

Table 1. Typical MAPE for prediction by Lewis [25]

MAPE	Interpretation
Less than 10%	Highly accurate prediction
10%-20%	Good accurate prediction
20%-50%	Reasonable accurate prediction
More than 50%	Inaccurate prediction

time point t. Table 1 contains typical MAPE values for industrial and business data and their interpretation.

5. **Experiment and Results.** The main objective of the experiment is to explore how suitable the selected algorithms are for real world usage. There are issues to be taken into account. We need to find an appropriate tradeoff between time required for computing and available computation time. The chosen algorithms must robustly provide good enough results in real world environment. There are fifteen plants that enlisted as Thailand prime crops. We choose to present only rice's because it is true main crop of the nation. We try to minimize the training time and see how accurate the results are.

5.1. **Rice yield from raw data.** Since the users are thousands of officers all over the country, it is possible that the data being fed into the system is the raw ones. Firstly, we examine what the results will be, should this undesirable incidents take place. Figure 5 shows loss of prediction model from raw data of rice yield (left) and comparison of actual and predicted data of MLP (center) and RF (right). Each dot is in light blue and is plotted according to the predicted and actual values. Dots appear darker when they are plotted on each other, showing higher frequency. The number of datasets is 3,786 and 20% test set. As obviously shown, the rice yield raw data possess extreme loss of prediction model of MLP, with the batch size of 128, 300 epochs, 5 hidden layers and 256 hidden notes. This is highlighted in the figure by zooming only from range 0-50. With RF, using 100 estimators and 0 random state, MAPE is reduced by 10 times. From the figure, there are roughly two sets of inaccurate results, where the actual values are around 0 and 200,000. Other than that actual and predicted values are reasonably accurate. (Note that the scale of actual data is from 0-600,000.) However, it is recommended that officers must be careful not to accidentally use raw data.

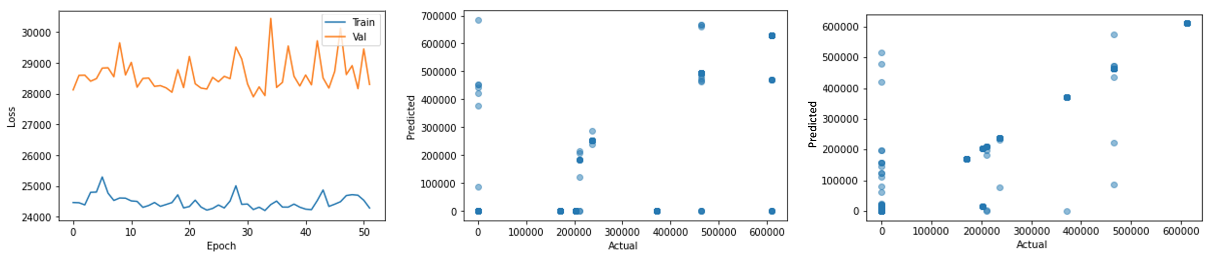

Figure 5. Results of rice yield from raw data: Loss of rice yield by MLP (left) and comparison of observed and predicted rice yields by MLP (center) and RF (right)

5.2. **Rice yield from clean data.** Given that data are cleaned by using filters provided in the system, we discuss experiments and results below. Figure 6 shows results of rice yield from clean data: loss of rice yield by MLP (left) and comparison of observed and predicted rice yields by MLP (center) and RF (right). The number of datasets is 2,788 and 20% test set. The features are water, disaster, quality of soil, cost, and sale-price. Having cleaned the data, the rice yield by MLP, with the batch size of 128, 300 epochs,

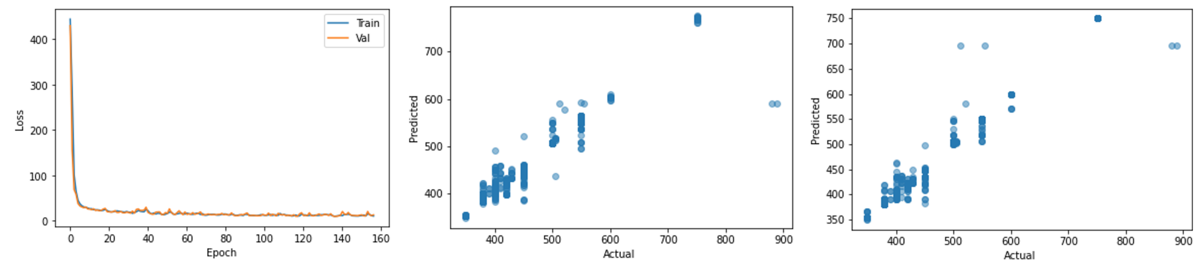

FIGURE 6. Results of rice yield from clean data: Loss of rice yield by MLP (left) and comparison of observed and predicted rice yields by MLP (center) and RF (right)

5 hidden layers and 256 hidden nodes, is reasonably good. The comparison of observed and predicted yield shows that there could be tiny errors. The loss of training of yield with saturation in 20 epochs and MAPE of test set is 2.37 percent. We only highlight the range of epoch between 0-160. With RF, having cleaned data also helps improve the quality of results significantly. The data range presented in the graph is 200-800. Most results fit reasonably well along the diagonal line between the actual and predicted values. MAPE is merely 1.76%.

5.3. **Rice price from clean data.** Another factor which is probably the most important one, based on farmer perception, is the price. Figure 7 shows results of rice price from clean data: loss of rice yield by MLP (left) and comparison of observed and predicted rice yields by MLP (center) and RF (right). The total number of datasets is 5,731 with 20% testing set. The features are month, province, district, sub-district and export. Similar to rice yields, we collect data of random sources in the database and examine the loss of prediction of MLP and receive satisfactory results, with the batch size of 128, 300 epochs, 5 hidden layers and 256 hidden notes. MAPE is also reasonably good, 22.3%. The loss of training of sale-price with saturation in 20 epochs and MAPE of test set is 10.74 percent. The comparison of observed and predicted price shows that there could still be some errors. With RF, we have merely two sets of data scattered off the diagonal line. Accordingly, the MAPE is very low, 1.7%.

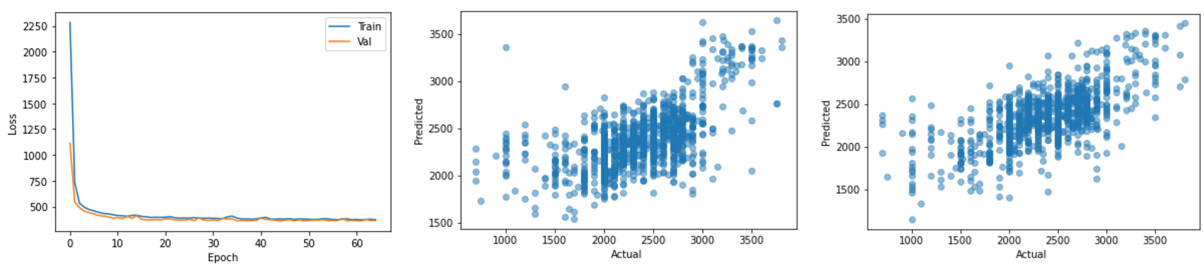

FIGURE 7. Results of rice price from clean data: Loss of rice yield by MLP (left) and comparison of observed and predicted rice yields by MLP (center) and RF (right)

5.4. **Rice cost from clean data.** The last factor is cost. Cost for growing rice has been steadily increasing over decades, in Thailand. Figure 8 shows results of rice cost from clean data: loss of rice yield by MLP (left) and comparison of observed and predicted rice yields by MLP (center) and RF (right). The total number of datasets is 2,714 with 20% testing set. The features are water, disaster, quality of soil, sale-price, and yield. Note that the use of fertilizer is considered as part of the quality of soil. With the batch size of 128, 300 epochs, 5 hidden layers and 256 hidden notes, MLP also receives quite satisfactory results. The loss of training of rice cost with saturation in 20 epochs and MAPE of test

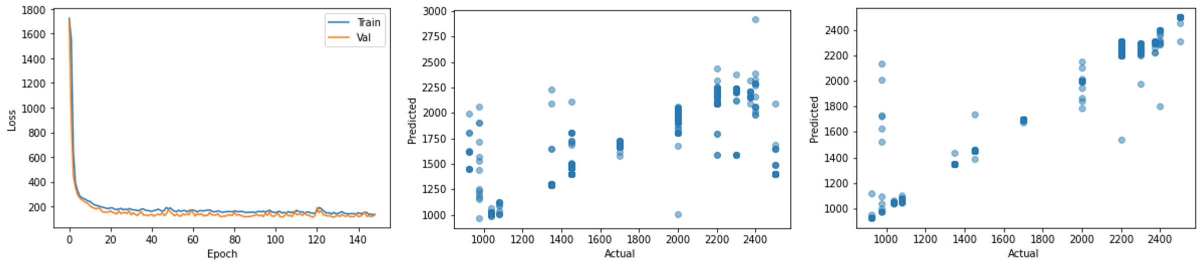

Figure 8. Results of rice cost from clean data: Loss of rice yield by MLP (left) and comparison of observed and predicted rice yields by MLP (center) and RF (right)

set is 4.75 percent. We selectively show the comparison of observed and predicted cost carried out with some errors, but the quality remains good according to MAPE. With RF, we selectively show the comparison of observed and predicted cost carried out with some errors. In general, we still receive good quality results.

However, we surprisingly receive 28.6% MAPE, (reasonable accurate, according to MAPE) in some cases. This happens because of several factors. The first one is how the operators filter the data. The second one is that the data that were collected over several decades from all over the country could be incomplete and incorrect. Note that we have to compromise with incomplete factors and find out the most efficient approach. There are pieces of information about soil quality in databases and there is also information about fertilizing, which is considered to be part of soil quality. We have to refer to a number of databases in order to represent the data of soil quality. In addition, there are several forms of data representing amount of water. These workarounds could affect the results.

6. **Conclusion.** As the underpinning workhorse of the personalized data system, MLP and RF are used to learn terabytes of data collected over several decades. The prediction of crops in real world database is a difficult task. This work presents a technique for agricultural prediction. The proposed technique applies MLP with cleansing data and feature selection/extraction is processed to resize before feeding to the network. Datasets were collected from Department of Agriculture Extension in Thailand. The experiments are conducted and the results demonstrate that the proposed technique provides promising results. Similarly, RF is also deployed, particularly to provide suggestions when time constraints are tight. Although considered a less accurate method, compared to MLP, RF works just fine in most cases. We choose three key factors, price, cost and yields in order to make it simple but meaningful to farmers. We found that the accuracy is very high, and MAPE is around 5% in most cases. In some difficult scenarios where data is not edequate, the results are still good, and MAPE is around 20%. In the future, more advanced learning techniques can be analyzed and installed in the system. The involved factors should also be taken into account because there will also be more external data sources connected to the system.

Acknowledgment. This research was supported by The Department of Agriculture Extension (DOAE), Ministry of Agriculture, Thailand. We are grateful to the cooperation and support of DOAE IT personnel. We are also thankful to Zyntelligent Co., Ltd., and Aditep Campira for their help.

REFERENCES

[1] E. Alpaydın, *Introduction to Machine Learning*, The MIT Press, 2013.

[2] Y. Sun, X. Chen, L. Jun, J. Zhao, Q. Hu, X. Fang and Y. Yan, Ship trajectory cleansing and prediction with historical AIS data using an ensemble ANN framework, *International Journal of Innovative Computing, Information and Control*, vol.17, no.2, pp.443-459, 2020.

[3] P. Wang, B. A. Hafshejani and D. Wang, An improved multilayer perceptron approach for detecting sugarcane yield production in IoT based smart agriculture, *Microprocessors and Microsystems*, vol.82, 103822, DOI: 10.1016/j.micpro.2021.103822, 2021.

[4] A. M. Al-Saif, M. Abdel-Sattar, A. Aboukarima and E. Dalia, Application of a multilayer perceptron artificial neural network for identification of peach cultivars based on physical characteristics, *PeerJ*, vol.9, no.11, e11529, DOI: 10.7717/peerj.11529, 2021.

[5] S. Nosratabadi, S. Ardabili, Z. Lakner, C. Mako and A. Mosavi, Prediction of food production using machine learning algorithms of multilayer perceptron and ANFIS, *Agriculture*, vol.11, no.5, 2021.

[6] E. T. Venkatesh and D. P. Thangaraj, Self-organizing map and multi-layer perceptron neural network based data mining to envisage agriculture cultivation, *Journal of Computer Science*, vol.4, pp.494-502, 2008.

[7] S. Pathane, U. Patil and N. Sidnal, Prediction of future market price for agricultural commodities, *International Journal of System and Software Engineering*, vol.3, no.1, pp.10-17, 2015.

[8] B. Ji, Y. Sun, S. Yang and J. Wan, Artificial neural networks for rice yield prediction in mountainous regions, *The Journal of Agricultural Science*, vol.145, no.3, pp.249-261, 2007.

[9] C. Z. Yuan and S. K. Ling, Long short-term memory model based agriculture commodity price prediction application, *Proc. of the 2020 2nd International Conference on Information Technology and Computer Communications*, New York, NY, USA, pp.43-49, 2020.

[10] J. Silva, M. Angulo, J. R. Borré, L. P. Ayarza, O. B. Lezama, Z. D. M. Galán and J. N. Beltran, An early warning method for agricultural products price spike based on artificial neural networks prediction, in *Computational Science – ICCS 2019. ICCS 2019. Lecture Notes in Computer Science*, J. M. F. Rodrigues, P. J. S. Cardoso, J. Monteiro et al. (eds.), Cham, Springer International Publishing, 2019.

[11] G. K. Jha and K. Sinha, Agricultural price forecasting using neural network model: An innovative information delivery system, *Agricultural Economics Research Review*, vol.26, no.26, pp.229-239, 2013.

[12] A. P. M. Ramos, L. P. Osco, D. E. G. Furuya, W. N. Goncalves, D. C. Santana and L. P. R. Teodoro, A random forest ranking approach to predict yield in maize with UAV-based vegetation spectral indices, *Computers and Electronics in Agriculture*, vol.178, 105791, DOI: 10.1016/j.compag.2020.105791, 2020.

[13] A. Ok and O. Gungor, Evaluation of random forest method for agricultural crop classification, *European Journal of Remote Sensing*, vol.45, no.1, pp.421-432, 2012.

[14] J. Jeong, J. Resop, N. Mueller, D. Fleisher, K. Yun, E. Butler et al., Random forests for global and regional crop yield predictions, *PLoS One*, vol.11, no.6, e0156571, DOI: 10.1371/journal.pone.0156571, 2016.

[15] C. Mariano and B. Monica, A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping, *Computers and Electronics in Agriculture*, vol.184, 106094, DOI: 10.1016/j.compag.2021.106094, 2021.

[16] Y. E. Roell, A. Beucher, P. G. Møller and M. B. Greve, Comparing a random forest based prediction of winter wheat yield to historical yield potential, *Agronomy*, vol.10, no.3, 395, DOI: 10.3390/agronomy10030395, 2020.

[17] J. Montano, A. Palmer, A. Sese and B. Cajal, Using the R-MAPE index as a resistant measure of forecast accuracy, *Psicothema*, vol.25, pp.500-506, 2013.

[18] W. Ma, K. Nowocin, N. Marathe and G. H. Chen, An interpretable produce price forecasting system for small and marginal farmers in India using collaborative filtering and adaptive nearest neighbors, *Proc. of the 10th International Conference on Information and Communication Technologies and Development (ICTD'19)*, New York, NY, USA, pp.1-11, 2019.

[19] P. Oktoviany, R. Knobloch and R. Korn, A machine learning-based price state prediction model for agricultural commodities using external factors, *Decisions in Economics and Finance*, vol.44, no.2, pp.1063-1085, 2021.

[20] S. Brunda, L. Nimish, S. Chiranthan and A. Khan, Crop price prediction using random forest and decision tree regression, *International Research Journal of Engineering and Technology (IRJET)*, vol.7, pp.235-237, 2020.

[21] I. Ghutake, R. Verma, R. Chaudhari and V. Amarsinh, An intelligent crop price prediction using suitable machine learning algorithm, *The 10th International Conference on Advances in Computing & Communications*, pp.1-4, 2021.

[22] J. V. Neumann and O. Morgenstern, *Theory of Games and Economic Behaviour*, Princeton University Press, Princeton, New Jersey, 1963.
[23] V. A. Bandel and E. A. Heger, *MASCAP: Maryland's Agronomic Soil Capability Assessment Program*, Cooperative Extension Service, University of Maryland, College Park, Maryland, USA, 1994.
[24] P. Sihag, S. M. Karimi and A. Angelaki, Random forest, M5P and regression analysis to estimate the field unsaturated hydraulic conductivity, *Applied Water Science*, vol.9, no.5, pp.1-9, 2019.
[25] C. D. Lewis, *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*, Butterworth Scientific, Boston, London, 1982.