A Hybrid Forecasting Model of Cassava Price Based on Artificial Neural Network with Support Vector Machine Technique

Korawat Polyiam and Pudsadee Boonrawd
Department of Information Technology, Faculty of Information Technology
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
e-mail: koorawat@hotmail.com , pudsadee.b@it.kmutnb.ac.th

Abstract— **Thailand is the world's largest exporter of cassava. The cassava prices fluctuate because of many factors such as the production cost, economic condition, and price intervention. Therefore, this research aims to propose a forecasting model of cassava price based on the 11-year data (from 2005 to 2015) obtained from the Thai Tapioca Starch Association and Office of Agricultural Economics. Various techniques were applied for the forecast such as Artificial Neural Network, Support Vector Machine, k-Nearest Neighbor and Hybrid Technique. The statistics used to determine the effectiveness of this model were Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) and Mean Squared Error (MSE). The results of this research showed that Hybrid Technique demonstrated the lowest value of error followed by Artificial Neural Network, k-Nearest Neighbor and Support Vector Machine, respectively. Therefore, it could be concluded that using the Hybrid Technique to forecast the price of cassava was better than other techniques and generated the predicted price closest to the actual price.**

Keywords-Forecasting Model; Artificial Neural Network; k-Nearest Neighbor; Support Vector Machine; Hybrid Technique;

I. INTRODUCTION

Cassava has been ranked the fourth highest grossing crop of Thailand, next to sugarcane, rubber and rice [1]. Thailand is also the largest exporting country of cassava in the world [2]. In 2015, Thailand had an export value of 172,778 million baht [1]. This crop was exported in the forms of chips, starch, pellets and ethanol. Thailand can grow cassava in every region of the country throughout the year. However, the prices of cassava fluctuate due to many factors such as business cycle, natural seasons, changes in price intervention policy and the global economic network [3] [4]. Cassava is a drought -tolerant crop, which can grow in every kind of soil, even soil with low fertility. It can be harvested after twelve-months of growth.

Cassava can be categorized according to the content of hydrocyanic acid into two types; sweet type and bitter type. Sweet type contains low content of hydrocyanic acid and is suitable for human consumption. This type is commonly used for cooking desserts but only at the household industry level because of the limited market. The bitter type has a high amount of hydrocyanic acid. Farmers mostly grow this type because it is rich in starch. The bitter type is not suitable for human consumption but widely used in the processing industry for export [5] [6]. The criteria to determine the cassava price is the percentage of starch content in the cassava roots. The fluctuation of prices and the increase of production cost are the main problems. If the production cost or cassava price can be forecasted, it may bring about the benefits from planning and management, which should lead to more business advantages.

The objective of this study is to propose a forecasting model for cassava price by applying the Artificial Neural Network, k-Nearest Neighbor and Support Vector Machine techniques to create an effective forecasting model for use. This model could be further developed to be a tool for analysis and planning of production cost management. This research is divided into five mains topics: Introduction, Related Work, Research Methodology, Results and Conclusion.

II. RELATED WORK

A. Machine Learning

Machine Learning is the use of a computer to find the mathematic model from the input data. This model can be used for prediction, explanation and visualization of the data [7]. Machine Learning is divided into two types; Supervised Learning and Unsupervised Learning as shown in Figure 1.

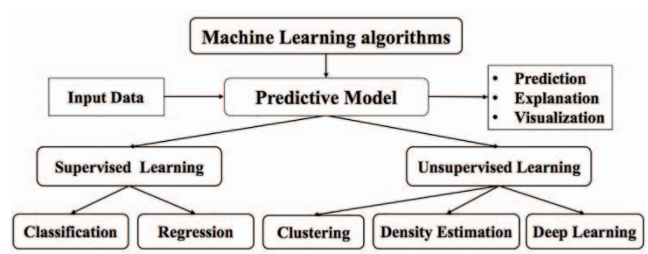


Figure 1. Machine Learning Framework

1. Supervised Learning

Supervised Learning is learning by adopting the input data, which are classified into learning and test datasets, to validate or determine the target output. If the target output does not match with the actual data, then weight adjustment will be conducted to make it equal to the output data [7] [8] such as classification and regression. The principles of classification and regression are as follows.

Classification is used to classify the data with different characteristics. It is an analysis of the discrete data and appropriate for creating a model to predict the data. In classification, the output is definite value and there are many techniques such as Artificial Neural Network, k-Nearest Neighbor and Support Vector Machine.

Regression is the statistical process which uses regression equation for the prediction. It is an analysis of the continuous data and the output can be definite or numerical value.

2. Unsupervised Learning

Unsupervised Learning is learning by using the classification of input data according to their characteristics without the validation of output data. In unsupervised learning, the output is represented in a cluster of data and there are many methods such as Clustering, Density Estimation and Deep Learning [8].

B. Forecasting Techniques

From the literature review, there are many techniques used for data forecasting. For example, data mining is the process of data analyzing and summarizing. This technique can be used for forecasting, planning or making a decision. Researchers have used mathematic methods to predict agricultural products such as the ARIMA technique. The exponential smoothing technique can accurately forecast a short-term outcome, but errors occurred for the long-term outcome. There has also been researches related to cassava, for example, the forecast of cassava starch-export prices [9], the forecast of demand and supply management on cassava products [10] and the planning of price transmission of cassava products [11]. The Artificial Neural Network technique has been used for various kinds of predictions, for example, the forecast of electricity prices, the forecast of stock prices [12] [13] [14] and the forecast of oil price [15]. This technique is flexible and able to reduce the limitation of mathematic methods which were described previously, therefore, the comparison between this technique and other techniques such as Support Vector Machine [14] [15] and k-Nearest Neighbor had been conducted in many researches [16]. The results indicated that the Artificial Neural Network technique demonstrated greater accuracy than the other techniques mentioned previously [17]. It could also effectively forecast the data that were numeric values and had the reliability for both short-term and long-term outcome [7] [8] [12] [15]. In addition, if a Hybrid Technique in which there are more than one technique used to the model will be more effective forecast [18] [19] [20].

III. RESEARCH METHODOLOGY

The concept of Cross-Industry Standard Process for Data Mining (CRISP-DM), which is the learning process from data mining, was used in this study as the reference. This learning process can be divided into six phases as shown in Figure 2.

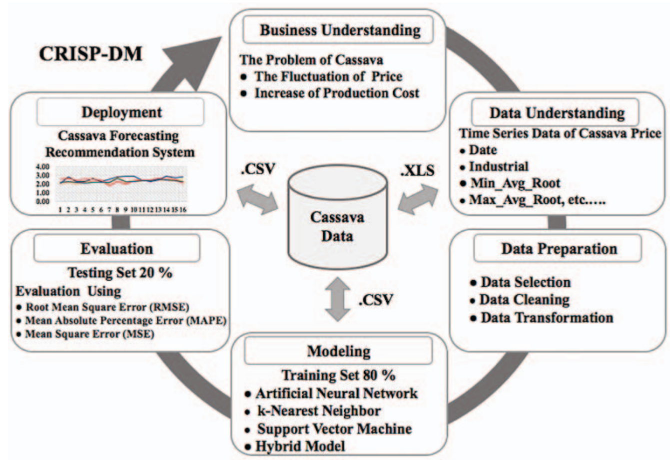


Figure 2. Research Framework

A. Business Understanding

This phase is focused on understanding the problem and finding the solution. It was demonstrated that the fluctuation of cassava root prices resulted in an inconsistency of the prices. If there is a prediction or forecast on this, it will appropriately help the price management and production cost planning. Therefore, this study developed a forecasting model of cassava price using Hybrid Technique.

B. Data Understanding

The data from January 2005 to December 2015, totally 11 years, from the Thai Tapioca Starch Association and Office of Agricultural Economics recorded weekly were used in this study. These data came from factories located in nine areas as follows: the east of Nakhon Ratchasima and nearby area; the west of Nakhon Ratchasima and nearby area; Kalasin province and nearby area; the upper area of Northeastern region; Sakaeo, Chanthaburi and Chachoengsao; Chonburi and Rayong; the lower part of the Northern region; the southern part of Ratchaburi and Kanchanaburi; and Uthai Thani.

C. Data Preparation

This phase is aimed at preparing and manipulating the data into an appropriate format to create a forecast model. This can be divided into four procedures as follows:

a) Data Selection

The data in this research comprised of the recorded date of the price, the location of the purchasing factory, the minimum and maximum prices of cassava roots, the average price of cassava roots, the minimum and maximum prices of mixed cassava, the average price of mixed cassava, the minimum percentage of starch content, the maximum percentage of starch content, average percentage of starch content and price at the average percentage of starch content. Only the data from cassava containing the twenty-five percent of starch content was used in this research because it is the criterion of setting the price of cassava roots.

b) Data Cleaning

Incorrect data or missing value was substituted by a weekly average price according to equation (1) shown below.

$$\sum_{i=1}^{n} \frac{x_i}{n} \tag{1}$$

where x_i is the observational value at number i
n is the total number of observational values

c) Data Transformation

The data were transformed from spread-sheet into the comma separate value (CSV) format in order to be used for developing the forecasting model.

D. Modeling

Rapid Miner Studio version 7, the data mining software, was used to develop the model for forecasting the prices of cassava roots. The techniques used for modeling were Artificial Neural Network, k-Nearest Neighbor, Support Vector Machine and Hybrid Technique. The split test was used in this research to separate the data into two sets which were x Set and Testing Set. Eighty percent of data from the Training Set and Twenty percent of data from the Testing Set were used for modeling. The details are described below.

a) Forecasting Model Using Artificial Neural Network

To develop the model, Multi-layer Artificial Neural Network is applied to analyze complex data. There were three layers as follows; Input Layer, Hidden Layer and Output Layer. Each layer had an input and calculated the sum of the input data (p_i) and weight (w_i).

Parameters were set for model development. The range of Learning Rate and Momentum were set at 0 to 1. The Training Cycle was set at 300, 500, 800, 1000 and 1500. The function is expressed in equation (2) as shown below.

$$n = \sum_{i=1}^{z} x_i w_i + b \tag{2}$$

where n is the sum of the function
x_i is the data at number i
w_i is the weight of neuron number i
z is the total neuron number of input layer
b is the propensity value
i is the number from i to z

b) Forecasting Model Using k-Nearest Neighbor

The model was developed by using k-Nearest Neighbor which is one of the techniques for data mining. In this research k-Nearest Neighbor was used for classification. The distance between each attribute and the training set was determined from each dataset, called Euclidean Distance. The parameters were set as described below. The K value was 1 to 10. Measure types were assigned as mixed measures, nominal measures and divergences. This assumption is expressed as equation (3).

$$D_{\text{Euclidean}} = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\ldots+(x_L-y_L)^2} \tag{3}$$

where x1 is attribute number 1 of data at point number 1
y1 is attribute number 1 of data at point number 2
which both x and y have the same number of attribute equal to L

c) Forecasting Model Using Support Vector Machine

Support Vector Machine was applied by using Kernel Function for the modeling [12]. The variable for decision-making is called Property and the changing variable used for determining various dimensions is called Feature. The process to select the most appropriate feature is called Feature Selection. The number of feature set is used to explain one particular situation [12] [13] such as the pattern of prediction called Vector.

Radial Basis Kernel Function was used in this research to develop the model. The parameters were assigned. Kernel gamma ranged from 1 to 50. C value ranged from 1 to 50 which is a constant value to identify the error in data classification. The structure of equation is shown in equation (4).

$$k(x_i,x_i) = \exp\left(-\gamma \left\|x_i,x_j\right\|^2\right) \tag{4}$$

d) Forecasting Model Using Hybrid Technique

From the results of previous experiments, the modeling method using Hybrid Technique and tested with many combinations of techniques was developed. As a result, the Artificial Neural Network was used in combination with the Support Vector Machine. The parameters set in Artificial Neural Network was Learning Rate in the range of 0 to 1, Momentum in the range of 0 to 1 and Training Cycles at 300, 500, 800, 1000 and 1500. In Support Vector Machine, Kernel gamma was set at 1 to 50 and C value was also set at 1 to 50. With these assumptions, the sum function was expressed as shown in equation (5).

$$Y_t = N_t + L_t + \varepsilon_t \tag{5}$$

where Y_t is the data at time t
N_t is the non-linear data at time t
L_t is the linear data at time t
ε_t is the random error at time t

which the Hybrid Model was shown in Figure 3.

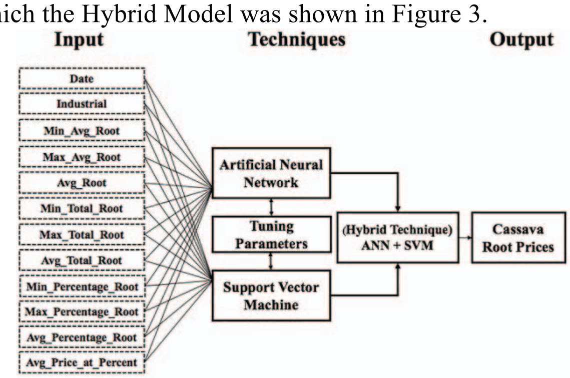


Figure 3. Conceptual model of Hybrid Model

E. Model Evaluation

This phase is aimed at evaluating the effectiveness of the developed cassava price forecasting model. Twenty percent of the data from the test set was used. The effectiveness of the model was evaluated by three methods; Mean Absolute

Percentage Error (MAPE), Root Mean Squared Error (RMSE) and Mean Squared Error (MSE). Mean Absolute Percentage Error (MAPE) was used to measure the error of prediction compared to the actual value, the formula is shown in equation (6).

$$\text{MAPE} = \frac{1}{n}\sum\left(\frac{|\text{Actual-Forecast}|}{|\text{Actual}|}\right)*100 \qquad (6)$$

Root Mean Squared Error (RMSE) is used to measure the deviation of the predicted value from the actual value, the formula is shown in equation (7).

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(\text{Actual-Forecast})^2}{n}} \qquad (7)$$

Mean Squared Error (MSE) uses the same principle of calculation with the statistical variance. It can be calculated by summing up the square of deviation at any time and then calculating the average of this value. The formula is shown in equation (8).

$$\text{MSE} = \frac{\sum_{t=1}^{n}(\text{Actual-Forecast})^2}{n} \qquad (8)$$

where Actual is the actual value of the data
Forecast is the predicted value of the data
n is the total number of the data

The principle of all three methods in measuring the deviation of the predicted value from the actual value are the same. If MAPE, RMSE and MSE are low, it indicates that the model can forecast close to the actual value.

F. Model Deployment

After the four forecasting models were developed, the most effective forecasting model was taken to use in the development of the Recommendation System which was designed as the information system in the form of a web application. This web application would support the production planning and decision-making related to cassava.

IIII. RESEULT

A. Results of Forecasting Model Effectiveness Evaluation

Evaluation results of the forecasting model effectiveness with the four models using Artificial Neural Network, k-Nearest Neighbor, Support Vector Machine and Hybrid Technique were shown in Table 1.

TABLE I. RESULTS OF FORECASTING MODEL EFFECTIVENESS EVALUATION

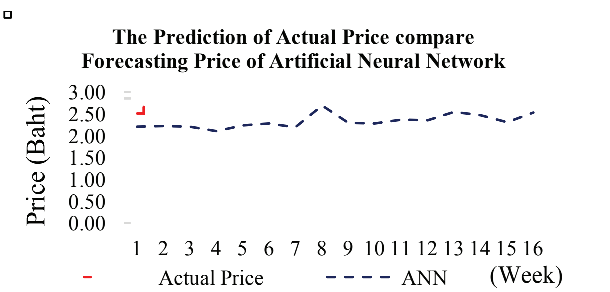
Techniques	Error Criteria		
	MAPE	**RMSE**	**MSE**
Artificial Neural Network (ANN)	2.720	0.006	0.009
k-Nearest Neighbor (k-NN)	10.584	0.129	0.073
Support Vector Machine (SVM)	13.683	0.522	1.143
Hybrid Technique	0.790	0.004	0.003

In Table 1, the forecasting model using Hybrid Technique demonstrated the lowest error values. The MAPE, RMSE and MSE were equal to 0.790, 0.004 and 0.003, respectively. The followings were Artificial Neural Network, k-Nearest Neighbor, Support Vector Machine.

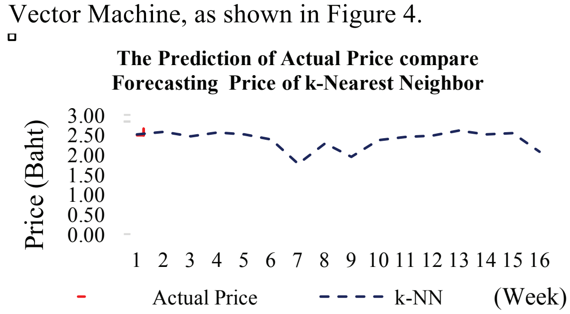
The parameters of each technique to develop the model were assigned as described below. In the Artificial Neural Network, Learning Rate was set at 0.2, Momentum was set at 0.2 and the Training Cycles was 800. In k-Nearest Neighbor, K value was equal to 9 and the Measure Types was assigned as Mixed Measures. In the Support Vector Machine, Radial Kernel Function was used. Kernel gamma was equal to 1 and C value was equal to 14

B. Results of Cassava Price Forecast

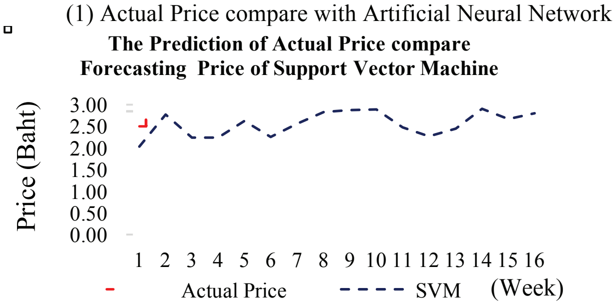
Four developed models were used to forecast the cassava price for the next sixteen weeks. The results indicated that Hybrid Technique was the most accurate followed by the Artificial Neural Network, k-Nearest Neighbor and Support Vector Machine, as shown in Figure 4.



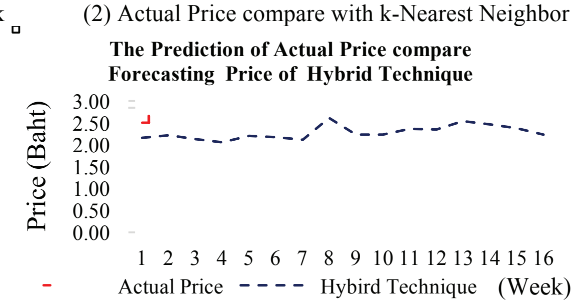
(1) Actual Price compare with Artificial Neural Network



(2) Actual Price compare with k-Nearest Neighbor



(3) Actual Price compare with Support Vector Machine



(4) Actual Price compare with Hybrid Technique

Figure 4. Results of Predictive Cassava Price

IIV. CONCLUSION

The forecasting models of cassava price using Artificial Neural Network, Support Vector Machine, k-Nearest Neighbor and Hybrid Technique were developed in this research. The effectiveness of these models was evaluated by Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) and Mean Squared Error (MSE). It can be concluded that the model using Hybrid Technique had the lowest value of error when compared to other techniques. The MAPE, RMSE and MSE were equal to 0.790, 0.004 and 0.003. This means that the developed model could generate the predicted price closest to the actual price for both short-term and long-term durations. For a future research, more dataset should be added to the model which should develop the process and parameters of each technique. The factors that may result in an overfitting should also be studied. The process used can also be improved in order to obtain a more accurate and effective model.

ACKNOWLEDGMENT

The authors wish to thank the Thai Tapioca Starch Association, Office of Agricultural Economics for providing data of this research. This research was funded by the Graduate College, Faculty of Information Technology, King Mongkut's University of Technology North Bangkok.

REFERENCES

[1] Office of Agricultural Economics. "Prospect for Agricultural of Thailand in ASEAN ." Ministry of Agriculture and Cooperatives. [Online Journal] March 2015. [Searched on 1st July 2016]. From http://www.oae.go.th/download/download_journal/2558/yearbook 57.pdf

[2] Office of Agricultural Economics. "Thailand Agricultural statistic year 2015." Ministry of Agriculture and Cooperatives. [Online Journal] March 2016. [Searched on 27th July 2016]. From http://www.oae.go.th/download/download_journal/ 2559/yearbook58.pdf

[3] Research Strategies and Cassava Industry Development in Thailand (2012-2016) and Cassava Research and Development Program under Research Strategies and Development, National Science and Technology Development Agency. Second Phase 2011-2016 (Strategic Planning Alliance II: SPA II). National Science and Technology Development Agency, Pathumthani: National Science and Technology Development Agency, 2011.

[4] TTSA. Analysis to Thai tapioca market. Technical report, Thai Tapioca Starch Association (TTSA), 2010.

[5] B. Adamolekun, "Neurological disorders associated with cassava diet: a review of putative etiological mechanisms. Metabolic Brain Disorder", Vol.26, pp.79–85, 2011.

[6] K. O. Falade, J. O. Akingbala, "Utilization of cassava for food. Food Reviews International", Vol.27, pp.51–83, 2010.

[7] A. Alizadeh, K. Mafinezhad, "Monthly Brent oil price forecasting using artificial neural network and a crisis index," Electronics and Information Engineering (ICEIE), 2010 International Conference On. Vol. 2 IEEE, 2010.

[8] J. Patel, S. Shah, P. Thakkar, K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques Expert Systems with Applications," Vol. 42, pp. 259–268, 2015.

[9] P. Warut, H. Van-Nam, S. Songsak, "ARIMA Versus Artificial Neural Network for Thailand's Cassava Starch Export Forecasting", Causal Inference in Econometrics, Vol. 622, December, pp. 255-277, 2015.

[10] N. Choosuk, A. Kengpol, "An Application of Forecasting Models for the supply and Demand Management of Cassava Product," KMUTNB Int J. Appl. Sci. Technol. Vol. 9, No.3, pp.175–187, 2016.

[11] I. Ojiako, , C. Ezedinma, G. Asumugha, and N.M. Nkang, "Spatial Integration and Price Transmission in Selected Cassava Products Markets in Nigeria: A Case of Lafun, World Applied Sciences Journal. Vol. 18, No.9, pp.1209–1219, 2012.

[12] Y. Kara, M. Acar Boyacioglu, O. K. Baykan, "Predicting direction of stock price index movement using artificial neural network and Support Vector Machine," The sample of the Istanbul stock exchange. Expert systems with Applications, Vol. 38, pp.5311–5319, 2011.

[13] P. Jyothi, V.D. Mytri., "A Prediction Model for Population Dynamics of Cotton Pest (Thrips tabaci Linde) using Multilayer-Perceptron Neural Network", International Journal Computer Applications, Vol. 67, No. 4, pp.19-26, 2013.

[14] N. Cristianini, J. Shawe-Taylor, "An Introduction to Support Vector Machine and Other Kernel-based Learning Methods," Cambridge University Press, 2000.

[15] C. Jinxing, W. Jianzhou, "Short-term electricity prices forecasting based on support vector regression and Auto-regressive integrated moving average modeling," Energy Conversion and Management, Vol 51, pp 1911-1917, 2010.

[16] K. Alkhatib, H. Najadat, I. Hmeidi, and MK. Ali-Shatnawi, "Stock price prediction using k-Nearest Neighbor algorithm" International journal of business humanities and technology, pp.32–44, 2013.

[17] Y. Shynkevich1, T. M. McGinnity, S. Coleman and A. Belatreche, "Predicting Stock Price Movements Based on Different Categoriesm of News Articles," IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (IEEE CIFEr), 2015.

[18] J. L. Zhang, Y. J. Zhang, L. Zhang, "A novel hybrid method for crude oil price forecasting," Energy Economic, Vol. 49, pp.649-659, 2015.

[19] L. Hu, G. Taylor, "A novel hybrid technique for short-term electricity price forecasting in UK electricity markets," J. Int. Council Elect. Eng., vol. 4, no. 2, pp.114–120, 2014.

[20] G. F. Lin, M. C. Wu, "A hybrid neural network model for typhoon-rainfall forecasting," Journal of Hydrology, vol. 375, no. 3, pp.450–458, 2009.