The Selection of Variables for Cabbages Price Fluctuation Prediction by Factor Analysis

Chih-Chao Hsu
Dept of Statistics
National Cheng Chi University
Taipei, Taiwan R.O.C.
witty27818a@gmail.com

*Mao-Yuan Pai
General Research Service Center
National Pingtung University of Science and Technology
Pingtung, Taiwan R.O.C.
mypai@g4e.npust.edu.tw

Shu-Yi Liaw
Dept of Business Administration
National Pingtung University of Science and Technology
Pingtung, Taiwan R.O.C
syliaw@mail.npust.edu.tw

Tz-Heng Hsu
Dept. of Computer Science and Information Engineering
Southern Taiwan University of Science and Technology
Tainan, Taiwan R.O.C
hsuth@mail.stust.edu.tw

Zhi-Hao Wang
Dept. of System Design Chen-Fon-Wei Security Co.,Ltd.
Kaohsiung, Taiwan R.O.C.
zhwang0401@homail.com

***Abstract* — The variables affecting the fluctuations of cabbage prices are plentiful and fast-growing, which makes it hard for farmers to collect and analyze the data from the cabbage wholesale markets. To improve this situation, the study developed a method to predict the increase or decrease of cabbage prices. This study is mainly concerned with whether the accuracy of the predictions regarding the fluctuations of cabbage prices is improved if this study condensed the variables that cause the fluctuations into fewer factors. This study gathered meteorological, oil prices and historical transaction data for cabbages in Taipei City. By means of factor analysis, this study selected the variables that affect the price changes and extracted the common factors behind these variables. Based on the results, logistic regression analysis models were constructed, predicting the fluctuations of cabbage prices. In the end, 8 factors were extracted. It turned out that the predictive model for fluctuations of cabbages prices using the dataset of factor scores had better performance on the three indicators, namely accuracy, specificity and precision, than the predictive model built on the original dataset in testing data. This led to the model based on the modified dataset with reference to the results of factor analysis having better generalization ability. Hence, this study suggests the collection of information about the 8 factors by the farmers instead of dozens of variables influencing the price changes of cabbage to achieve better performance on price fluctuation predictions.**

Keywords—factor analysis, logistic regression analysis, price prediction, variable selection

I. INTRODUCTION

Cabbage is regarded as the favorite vegetable of many people due to its known health benefits and the various ways it can be prepared for consumption. Higher prices decrease the willingness to purchase cabbage whereas lower prices tend to increase the demand for cabbages. Therefore, price level is a key variable affecting consumers' choice on whether to purchase cabbage or not. Farmers tend to cultivate more cabbage when the price level is expected to be high in order to maximize profit. Consequently, this behavior often causes an imbalance between supply and demand which results in greater losses for these farmers. Two reasons are leading to the problem. Firstly, there are numerous variables that might cause cabbage price to rise or drop, but farmers are usually unaware of these variables. Secondly, the variables influencing the price are not static. The availability of information will continuously increase as time goes by. For example, there are countless transaction data for cabbage generated on a daily basis. The rapidly growing information is enormous and beyond an ordinary farmer's control. This situation accentuates the lack of abilities for these farmers to gather, process and analyze the data from cabbage markets.

To improve the abilities of the farmers to gather, process and analyze the transaction data, it is suggested to develop a systematic method for predicting the increase or decrease of cabbage prices, which would help the farmers formulate their planting strategies. Many variables might influence predictions of cabbage prices. In the past, it was hardly possible to gather these variables due to underdeveloped online information gathering techniques. Studies have shown that weather conditions [1][2], oil prices [2] and historical transaction data [3][5] are considered to have an influence on the fluctuation of cabbage prices Hence, this study gathered the following variables (about weather conditions, oil prices and historical transaction data) hoping to find out the common factors behind these variables through factor analysis. These crucial factors were then used to predict the fluctuations of cabbage prices. Nowadays, the techniques of web crawling are well-developed and thus it's rather easy to gather information online. It is necessary to identify the factors behind the variables affecting the cabbage prices' fluctuations by means of factor analysis.

Hence, hoping to improve the accuracy of the predictions regarding the fluctuations of cabbage prices, this study is mainly concerned with whether the accuracy of the predictions would be influenced if the price-related variables were concentrated in fewer factors. In order to identify the variables that trigger the fluctuations of cabbage prices, this study gathered data related to weather conditions, oil prices and historical transaction data of cabbage in Taipei City. This study hopes to extract the common factors behind the variables influencing the price fluctuation for cabbages by means of factor analysis, and use these results to build a logistic regression model, which is the prediction model commonly used in binary classification problems. Moreover, this study would like to compare the predictive performances of the model built on the data modified by the results of factor analysis with that of the model built on the original dataset, seeing whether using the results from factor analysis to build the price fluctuation prediction model for cabbage would help to increase the accuracy of the prediction. As for why this study chose to analyze the data of Taipei City, being the capital of R.O.C., it has the largest population among the cities in Taiwan, and thus makes the data rather representative. Additionally, cabbages are easy to cultivate, and have a high nutritive value as well as a crunchy texture, which make them widely planted and bought.

II. Related Works

A. *Variables Affecting Cabbages' Price fluctuation*

This study utilized web crawlers to collect the data related to the variables influencing cabbage prices at all the wholesale markets in Taiwan. From the literature reviewed, these price-rising-or-falling-influencing variables come from the following categories: the price of cabbage, the supply [3][5], and cost variables such as transportation costs [2]. In addition, many studies have found that meteorological variables such as temperature and precipitation [1][2] also play important roles in affecting the wholesale market for cabbages. From an economic point of view, consumers' expectations also contribute to changes in the price for cabbages. For example, an increase in purchases of food ingredients can be observed before special holidays [4].

Cabbages are common food ingredients in the wholesale markets in Taiwan. When natural disasters occur, the retail price of cabbages will immediately and drastically increase, as shown in Fig. 1. For example, in Taiwan, from June to September is the typhoon season, and the price of cabbage usually rises sharply. Conversely, when there is an excess of production, such as from October to May, the market price of cabbage will be lower than the cost of production. Therefore, cabbages' data are used in this study as the main analytical materials. In this study, historical transaction data of cabbage were collected at the No.1 wholesale market for vegetables and fruits in Taiwan, specifically highest price, lowest price, median price, average price and volume. Also, by the support of web crawlers, this study got the historical price data of 92, 95, 98-octane unleaded gasoline, diesel, kerosene, two-stroke oil, fishing boat fuel A and B, low sulfur boiler fuel oil 0.5% and 1%, low sulfur fuel oil and special low sulfur fuel oil. Finally, this study collected 25 meteorological variables, listed as follows: evapotranspiration, UV index, maximum temperature, minimum temperature, average temperature, dew point temperature, maximum pressure, minimum pressure, average pressure, sea-level pressure, solar insolation duration, solar insolation rate, global solar radiation, minimum relative humidity, average relative humidity, rainfall duration, total cloud cover, visibility, total rainfall, one-hour maximum rainfall, ten-minute maximum rainfall, average wind speed, average wind direction, peak gust wind speed and peak gust wind direction.

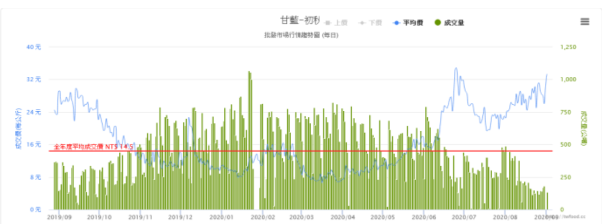

Fig. 1. Historical price and volume of cabbages

B. *The Selection of Analytical Methods*

In order to improve the accuracy of the prediction regarding fluctuations of cabbage prices, this study aimed to investigate whether the accuracy of the prediction is affected by condensing the variables that are mentioned in section 2-A into fewer factors. Therefore, two experimental procedures were proposed in this study, which are described as follows:

- Experiment 1: Many current studies use regression analysis for price forecasting [2][9]. However, this study focused on predicting the fluctuations of cabbage's price, which is a binary variable and therefore suitable for using logistic regression analysis model. The experimental procedure is shown in Fig. 2: First, this study collected the variables that affect cabbage prices by web crawlers, and then this study predicted the fluctuations of cabbage prices using logistic regression analysis.

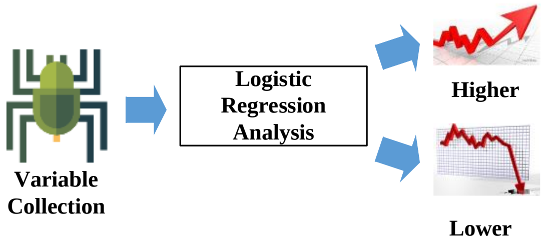

Fig. 2. Procedure of the experiment 1

- Experiment 2: Previous studies reviewed indicated that condensing the variables affecting price fluctuations into fewer factors might help improve the accuracy of the predictions related to increases or decreases in price [6][7]. Consequently, in experiment 2, the variables collected to predict the fluctuations in price were condensed into several factors. Then, the results were used to execute the logistic regression analysis for the purpose of assessing whether utilizing the results from factor analysis helps improving the accuracy of the predictions.

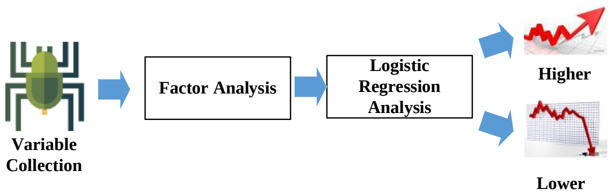

Fig. 3. Procedure of the experiment 2

III. Methods and Technologies

Based on the related works presented in section 2, this study designed a procedure to compare the analytical methods, as seen in Fig. 4, which includes (i) data collection, (ii) data processing, (iii) factor analysis, (iv) logistic regression analysis and (v) comparative results. These are discussed in more detail in the present section.

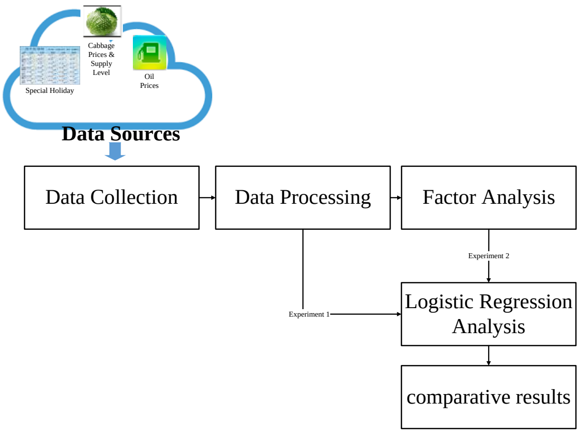

Fig. 4. Procedure of comparison of analytical methods

A. Data Collection

Based on the variables influencing the price fluctuations for cabbages presented in section 2-A, this study used web crawling as a technique for data gathering. The sources of the historical transaction data for cabbage, the weather conditions' data and the oil prices data are respectively from The Transaction Information Center of Wholesale Markets for Crops (directly translated), Central Weather Bureau Observation Data Inquire System and the official website of CPC Corporation, Taiwan. This study utilized some Python packages, such as requests, BeautifulSoup and pandas to carry out data collection. The function of the requests package is to get the data from the websites whereas the BeautifulSoup package is used to parse the HTML and retrieve the required information from the labels within HTML. Then the data from the three sources are combined into a spreadsheet by means of the pandas package. As for the data of special holidays including national holidays and religious holidays, they were labeled manually in the form of binary variables.

B. Data Processing

This study mainly used web crawling techniques to collect the data related to the variables affecting the fluctuations of cabbage prices. However, during the data collection, this study might encounter the problem of missing data. For example, there are a lot of missing data in several variable columns in the data for weather conditions due to reasons such as the malfunction in temperature sensors in some areas. The missing data in these columns are imputed with the mean of the values from the days before and after. For instance, the temperature data on January 20th were missing; thus, this study used the average temperature of Jan. 19th and 21st to impute the missing value. On top of that, this study deleted two-stroke oil, low sulfur boiler fuel oil 1% and special low sulfur fuel oil from the data of prices of oil because these three variables are almost constants, which make the correlation coefficients between them and other variables imponderable and thus interrupt the subsequent analysis.

C. Factor Analysis

Based on the influencing variables presented in section 2-A, this study collected many variables that affect the price for cabbages to rise or to fall. According to the literature review, there are common factors behind the variables which cause the price to fluctuate [8]. The reason is that the value of these variables is condensed into few factors. As a result, this study carried out factor analysis to identify the common factors behind the variables that influence the cabbage prices to either increase or decrease, designing a procedure of factor analysis as shown in Fig. 5. The description of the procedure is as follows.

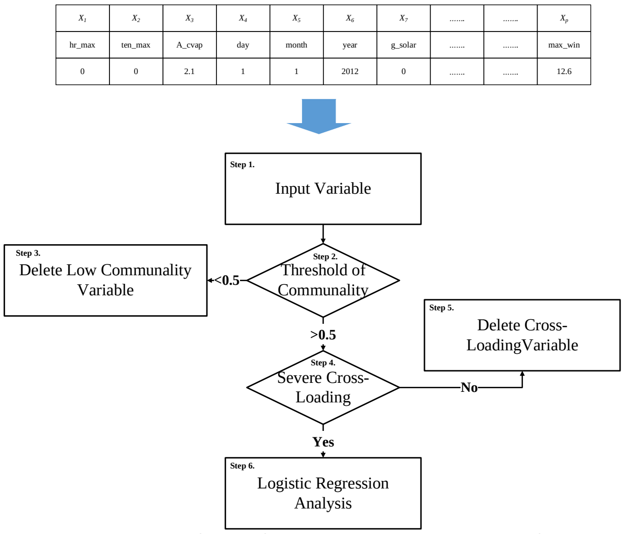

Fig. 5.Procedure to identify the factors behind the variables influencing cabbage prices fluctuations.

- Step 1: Input the influencing variables and form a data vector **x**, which contains elements x_1 to x_p corresponding to the values of all p variables affecting the increase or decrease of cabbage prices in this observation (i.e., $\mathbf{x} = (x_1, \ldots, x_p)^T$), as seen in Table 1. This study collected 45 variables affecting the fluctuations, so p equals 45 here.

TABLE I. VARIABLES THAT AFFECT THE PRICE FOR CABBAGES TO RISE OR TO FALL

X_1	X_2	X_3	X_4	X_5	X_6	X_7	....	....	X_p
one-hour maximum rainfall	ten-minute maximum rainfall	evapotranspiration	day	month	year	global solar radiation	....	....	Peak gust wind speed
0	0	2.1	1	1	2012	0	....	....	12.6

- Step 2: This study conducted factor analysis on the 45 variables that affect the increase or decrease of cabbage prices for the purpose of identifying the crucial factors influencing the price of cabbage. Among all 45 influencing variables, the ith variable's variance, say X_i, can be expressed by the following equation:

$$\mathrm{Var}(X_i) = \sigma_{ii} = \sum_{j=1}^{k} f_{ij}^2 + \psi_i \qquad (1)$$

for i = 1, … ,45 and k is the number of factors. f_{ij} is the (i,j) element of the factor loading matrix **F**, representing the weighting of the jth factor among all k factors with respect to the ith variable that affects the fluctuation of cabbage prices, whereas ψ_i represents the variance coming from the noise of the ith influencing variable itself. The $\sum_{j=1}^{k} f_{ij}^2$ part in (1) is called communality. Variables with high communality value share high common variance with other variables, while variables with low communality value share low common variance with other variables. As a result, the communality can be used as a basis for the deletion of variables that affect the price changes for cabbage. According to the rule of thumb, variables with communality less than 0.5 show lack of influence in the analysis. Hence, if any of the variables affecting the fluctuations has a communality value lower than 0.5, then execute step 3. In contrast, if all the variables that affect the price of cabbage have a communality value greater than 0.5, then move on to step 4.

- Step 3: This study deleted the variables influencing the price fluctuations with communality less than 0.5. After that, rerun step 1.

- Step 4: After executing the above steps, it remains 41 variables affecting the fluctuations of cabbage prices, which can be grouped into 8 factors. In order to identify the correlations between these variables and the factors, new factor loadings were obtained by applying rotations. The total sum of variance for columns $\sum_{j=1}^{k} v_j$ of the rotated factor loadings matrix $\boldsymbol{A} = [a_{ij}]$ can be derived from (2):

$$\sum_{j=1}^{8} v_j = \sum_{j=1}^{8}\left(\frac{1}{41}\sum_{i=1}^{41}\left(a_{ij}^2\right)^2 - \left(\frac{1}{41}\sum_{i=1}^{41} a_{ij}^2\right)^2\right) \quad (2)$$

The cross-loading problem occurs when there exists at least one of the variables highly correlated with two or more factors at the same time. If the cross-loading problem is observed, proceed to step 5. Otherwise, move on to step 6 if every variable that affects the fluctuations of cabbage prices is highly correlated with exactly one factor.

- Step 5: Variables with cross-loading problems were deleted in the study. After that, rerun step 1.

- Step 6: After executing the above steps, for every single observation, the values of all 41 variables can be condensed into factor scores of the 8 factors. This study then conducted logistic regression analysis with the factor scores derived from the result of factor analysis. Please refer to section 3-E for more details about building a logistic regression model to predict the fluctuations of cabbage prices.

D. *Logistic Regression Analysis*

Apart from the original dataset, based on the results of factor analysis, this study now have a new dataset composed of the factor scores. This study then constructed the predictive models for the price fluctuations using the two datasets. Because the price either increases or decreases, the response variable is considered binary therefore this study adopted logistic regression analysis. The data analyzed contained 45 variables which cause the price of cabbage to either increase or decrease, as well as a constant term, forming a 46-dimensional data vector $\boldsymbol{x} = (1, x_1, \dots, x_{45})^T$. The regression parameters vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{45})^T$ consists of the parameters for logistic regression curve. The logistic regression model applies the sigmoid function to limit the range of the output values in (0, 1) interval, as shown in Fig. 6 and (3):

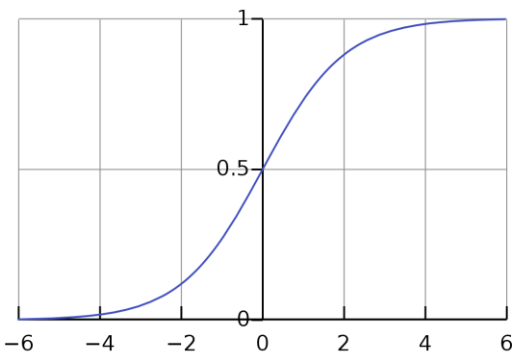

Fig. 5. Sigmoid function

$$y = \sigma(f(x)) = \frac{1}{1+e^{-f(x)}} = \frac{1}{1+e^{-\beta^T x}} \quad (3)$$

In this study, two logistic regression models were constructed using the statistical software SPSS in order to predict the fluctuations in cabbage prices. One of the models uses the original data which include all 45 variables affecting the fluctuations of cabbage prices as input, whereas the other model uses the factor scores derived from the results of the factor analysis, which is the information extracted from the variables influencing the price fluctuations for cabbages.

E. *Comparative Results*

In this study, two different logistic regression models were built to predict the price fluctuations in experiment 1 (original dataset: complete variables set) and experiment 2 (dataset modified by factor analysis: some variables deleted and condensed) respectively. Moreover, this study would like to investigate the different performances of the two models in predicting the fluctuations of cabbage prices. In order to realize the difference between experiment 1 and 2, this study compared the results of the two experiments by means of confusion matrix as shown in Table 2.

TABLE II. FORM OF CONFUSION MATRIX

Confusion Matrix of Price Rising or Falling		Predicted Value	
		0(fall)	1(rise)
Observed Value	0(fall)	TN	FP
	1(rise)	FN	TP

TP, TN, FP and FN stand for true positive, true negative, false positive and false negative respectively. Three ratios: accuracy, specificity and precision were taken as indicators of performance on predicting the fluctuations of cabbage prices. The three indicators were computed as presented in (4) to (6).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Specificity = \frac{TN}{TN+FP} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

IV. RESULTS AND DISCUSSIONS

In section 3-D, this study generated two different datasets from experiment 1 (original dataset: complete variables set) and experiment 2 (dataset modified by factor analysis: some variables deleted and condensed), then took them as inputs to train two logistic regression models to predict the fluctuations of cabbage prices. Detailed descriptions are as follow.

A. Building a Logistic Regression Model Using the Original Dataset to Predict the Fluctuations of Cabbage Prices

In the experiment 1, this study built a logistic regression model using the original dataset which includes all 45 variables affecting the increase or decrease of cabbage prices. The results of the predictions regarding the price fluctuations are shown in Table 3, with 68.0% accuracy, 72.1% Specificity and 70.1% precision. Then this study applied this model on the testing data, the results of the prediction are shown in Table 4, with 65.1% accuracy, 73.3% specificity and 67.5% precision.

TABLE III. CONFUSION MATRIX: ANALYZING THE RESULTS OF EXPERIMENT 1 IN TRAINING DATA

Classification Table[a]

	Observed		Predicted		
			train		Percentage Correct
			0(fall)	1(rise)	
Step 1	Train data	0(fall)	536	207	72.1
		1(rise)	275	486	63.9
	Percentage Correct		66.1	70.1	68.0

[a] The cut value is 0.5

TABLE IV. CONFUSION MATRIX: ANALYZING THE RESULTS OF EXPERIMENT 1 IN TESTING DATA

Predicted group * test Crosstabulation

	Observed		Predicted		
			test		Percentage Correct
			0(fall)	1(rise)	
Step 1	Test data	0(fall)	337	123	73.3
		1(rise)	195	256	56.8
	Percentage Correct		63.3	67.5	65.1

B. The Results of Factor Analysis

This study aimed at identifying the factors influencing the fluctuations of cabbage prices through factor analysis. As the rule of thumb, this study extracted 8 factors from the variables with eigenvalue greater than 1 and communality greater than 0.5. According to the rotated factor loading matrix, there is no cross-loading problem. The results of the factor analysis are shown in Table 5.

TABLE V. RESULTS OF FACTOR ANALYSIS

Number	Factors	Price Rising or Falling Influencing Variables for Cabbages
1	Prices of Oil	92-octane unleaded gasoline, 95-octane unleaded gasoline, 98-octane unleaded gasoline, diesel, kerosene, fishing boat fuel A, fishing boat fuel B, low sulfur fuel oil and low sulfur boiler fuel oil 0.5%
2	Temperature and Atmospheric Pressure	maximum temperature, minimum temperature, average temperature, dew point temperature, maximum pressure, minimum pressure, average pressure and sea-level pressure.
3	Solar Insolation and Humidity	solar insolation duration, solar insolation rate, global solar radiation, minimum relative humidity, average relative humidity, rainfall duration, total cloud cover and visibility.
4	Historical Prices of Cabbages	high price, low price, medium price and average price.
5	Rainfall	total rainfall, one-hour maximum rainfall and ten-minute maximum rainfall.
6	Wind Field	average wind speed, average wind direction, peak gust wind speed and peak gust wind direction.
7	Trade-related Variables	volume, open or not and days of week.
8	Special Holidays	national holidays or not and religious holidays or not.

This study replaced the values of the remaining 41 variables affecting the fluctuations of cabbage prices with the factor scores derived from the factor analysis to build a logistic regression model to predict the fluctuations of cabbage prices in the second experiment. The results of the fluctuations' prediction are shown in Table 6, with 55.9% accuracy, 58.3% specificity and 56.8% precision. Then this study applied this model on the testing data. The results of the predictions are shown in Table 7 with 67.2% accuracy, 78.9% specificity and 72.0% precision.

TABLE VI. CONFUSION MATRIX: ANALYZING THE RESULTS OF EXPERIMENT 2 IN TRAINING DATA

Classification Table[b]

	Observed		Predicted		
			train		Percentage Correct
			0(fall)	1(rise)	
Step 1	Train data	0(fall)	433	310	58.3
		1(rise)	354	407	53.5
	Percentage Correct		55.0	56.8	55.9

[b] The cut value is 0.5

TABLE VII. CONFUSION MATRIX: ANALYZING THE RESULTS OF EXPERIMENT 2 IN TESTING DATA

Predicted group * test Crosstabulation

	Observed		Predicted		
			test		Percentage Correct
			0(fall)	1(rise)	
Step 1	Test data	0(fall)	363	97	78.9
		1(rise)	201	250	55.4
	Percentage Correct		64.4	72.0	67.2

C. *Comparative Results*

Based on the results presented in section 4-A and 4-B, this study performed the comparative results of the predictive model for the fluctuation of cabbage prices using the original dataset as well as the factor scores derived from the factor analysis.

The three comparative indicators of the predictive models, namely specificity, precision and accuracy are presented in Table 8 and 9, separated by training data and testing data. The yellow-shaded part indicates that the model has better overall predictive performance.

TABLE VIII. COMPARATIVE RESULTS OF EXPERIMENT 1 AND 2 IN TERMS OF THE TRAINING DATA

	Train Data	Specificity	Precision	Accuracy
Experiment 1	Original Data	72.1%	70.1%	68.0%
Experiment 2	Data Modified by The Results of Factor Analysis	58.3%	56.8%	55.9%

TABLE IX. COMPARATIVE RESULTS OF EXPERIMENT 1 AND 2 IN TERMS OF THE TESTING DATA

	Test Data	Specificity	Precision	Accuracy
Experiment 1	Original Data	73.3%	67.5%	65.1%
Experiment 2	Data Modified by The Results of Factor Analysis	78.9%	72.0%	67.2%

V. CONCLUSIONS

This study collected data related to the variables affecting the fluctuation of cabbage prices including weather conditions, prices of oil and the historical transaction data of cabbages. By means of factor analysis, this study found that there are 8 factors behind the variables affecting the fluctuation of cabbage prices, which are: prices of oil, temperature and atmospheric pressure, solar insolation and humidity, historical prices for cabbages, rainfall, wind field, trade-related variables and special holidays.

This study also replaced the 45 original variables affecting the fluctuation of cabbage prices with the factor scores of the 8 factors mentioned above, and used them to construct a logistic regression model. Although being unable to fit the training data like the model based on the original dataset, the model that used the modified data referred to the results of factor analysis had better generalization ability and thus performed relatively well and stable in the testing data compared to the model built on the original dataset. Consequently, for various data in the real world, this study suggested that the farmers use the factor scores derived from the 8 factors found in the results of the factor analysis to predict the fluctuations of cabbages' price, instead of using all the 45 variables. After all, gathering information about the 8 factors is easier than collecting dozens of related variables.

ACKNOWLEDGMENT

This work was supported by Ministry of Science and Technology of Taiwan (R.O.C) under Grants MOST 108-2218-E-020-003.

REFERENCES

[1] H. Lee and A. Moon, "Development of yield prediction system based on real-time agricultural meteorological information," 16th International Conference on Advanced Communication Technology, Pyeongchang, 2014, pp. 1292-1295.

[2] S. F. Asnhari, P. H. Gunawan and Y. Rusmawati, "Predicting Staple Food Materials Price Using Multivariables Factors (Regression and Fourier Models with ARIMA)," 2019 7th International Conference on Information and Communication Technology (ICoICT), Kuala Lumpur, Malaysia, 2019, pp. 1-5.

[3] S. Mahida and B. Patel, "A Review of the Application of Data Mining Techniques for Vegetable Price Prediction", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, vol. 3, pp. 35-39, March-April-2018.

[4] E. Teimoury, H. Nedaei, S. Ansari and M. Sabbaghi, "A multi-objective analysis for import quota policy making in a perishable fruit and vegetable supply chain: A system dynamics approach", Computers and Electronics in Agriculture, ISSN : 0168-1699, vol. 93, pp. 37-45, 2013.

[5] Prashantha, S., Shravan, C. Y., Bharath, B., Bharghavachar, B. N., & Shilpa, B. L., "A SURVEY ON CROP ANALYSIS & AGRICULTURE COMMODITIES PRICE PREDICTION USING MACHINE LEARNING TECHNIQUES", International Journal For Technological Research In Engineering (IJTRE), ISSN : 2347-4718, vol. 7, pp. 6621-6624, March-2020.

[6] H. Ince and T. B. Trafalis, "Kernel principal component analysis and support vector machines for stock price prediction", IIE Transactions, vol. 39, pp. 629-637, 2007.

[7] Anish, C.M., Majhi, B., "Hybrid nonlinear adaptive scheme for stock market prediction using feedback FLANN and factor analysis", *J. Korean Stat. Soc.* **45**, pp. 64-76, 2016.

[8] "Factor Analysis on China's Real Estate Price: An Empirical Research of Panel Data", Journal of China University of Geosciences(Social Sciences Edition), ISSN : 1671-0169, vol. 10, pp. 98-103, 2010.

[9] Xu Yiting, Mu Yueying and Zhao Yousen, "Analysis of vegetable price changes and influencing factors in Beijing", Chinese Vegetables, ISSN : 1000-6346, vol. 1, no. 11, pp. 62-69, 2017.