

Price Prediction of Chili Commodities in Bandung Regency Using Bayesian Network

Putri Nuvaisiyah ^{#1}, Fhira Nhita ^{*2}, Deni Saepudin ^{#3}

#School Of Computing, Telkom University

** School Of Computing, Telkom University*

#School Of Computing, Telkom University

Telekomunikasi Street, Terusan Buah Batu Bandung, 40257, Indonesia

¹ putrinuvaisiyah@student.telkomuniversity.ac.id

² fhiranhita@telkomuniversity.ac.id

³ denisaepudin@telkomuniversity.ac.id

Abstract

Chili is one of the agricultural commodities consumed by Indonesian people. Market data in recent years show that chili prices tend to fluctuate as supply and demand changes. One of the impacts of chili price changes for farmers is the production cost is higher than the selling price. In addition to supply and demand changes, the weather is also indicated as a factor of price changes due to the weather being considered by farmers to grow chili. Price prediction is needed to determine the condition of chili prices in the future to help farmers in making decisions to plant at the right time. One method that can be used to make prediction is Data Mining classification method. In this paper, Bayesian network algorithm was used as Data Mining classification method to predict the price of chili commodity in Bandung Regency based on weather information and classified the price into economic class and not economic class. The result shows that the prediction model obtained by the Bayesian Network gives a system's performance for precision and recall that is 0.92 and 1 respectively with average accuracy of 83.5% in classifying the price.

Keywords: Bayesian Network, Data Mining, classification, price prediction.

I. INTRODUCTION

INDONESIA is a country whose population is still dependent on agriculture. One of the most consumed agricultural commodities by Indonesian people is chilli. Chili is consumed as food in Indonesia and it also widely consumed in Bangladesh (Hossen, 2016). Chili contains many vitamins that are good for the body such as vitamin A, vitamin B, vitamin C and iron (Parle & Kaura, 2013). The high demand for chili makes chili prices fluctuate (Firdaus, & Gunawan, 2013) and becomes one of the reasons for the high inflation rate in Indonesia (Mariyono, Joko, & Sumarno, 2016). Price fluctuations certainly have a negative impact for both farmers and the community. With the farmers selling price lower than the production cost makes farmers loss (Ditakristy, Saepudin, & Nhita 2016), which causes the price of chilli is not economic for farmers. Economic price for farmers is if the production cost incurred is less than the selling price. Otherwise, if the production cost is higher than the selling price then the price is not economic. In addition to changes in supply and demand, weather is also indicated as the cause of the chili's price changes. Weather plays an important role in the process of chili planting. Chili plants require enough water content to have good quality results. With the weather changing conditions of course it is very influential on chili plants. Therefore, price prediction becomes very important and necessary. With the prediction of the price, farmers could know the condition of chili prices in the future so it can helps farmers in determining the right planting time.

Data mining classification methods can be used to make predictions. There have been many studies that used data mining methods to do predictions as in the study Ramesh and Vardhan (2013), which predicts the number of crops production using SVM, KNN and ANN algorithms. The algorithms were also used in making price predictions in research (Kaur, Gulati, & Kundra, 2014). In this paper the algorithm used is Bayesian Network where it has been widely applied to make predictions. Bayesian network is a probabilistic graphical model that shows the relationship between variables that affect each other of an event in the form of Directed Acyclic Graph. It can also be used to handle incomplete data. In the previous study Eisuke, Harada, and Mizuno (2012), Bayesian network was used to predict NIKKEI and Toyota stock prices with an average maximum error of 30% for NIKKEI and 20% for Toyota. In addition, the Bayesian Network is also used to predict the weather where the accuracy obtained from the overall test scenario gives a good accuracy that is more than 80% (Sharma & Goyal, 2015). In the field of agriculture, Bayesian Network has been widely used as in research Chawla et al., (2016), Gandhi, Armstrong and Petkar (2016), Rasmussen, Madsen and Lund (2014), Newlands and Tomley-Smith (2011), Pérez-Ariza, Nicholson, and Flores (2014), who conducted research on the prediction of diseases in plants, crops prediction and also used as a risk management tool.

In this paper, Bayesian Network is used to predict the price of chilli commodity in Bandung Regency based on weather information. The data used in this research are the historical data of chili prices and weather data of Bandung Regency. This study provides output in the form of prices that have been classified into economic and not economic classes.

II. BAYESIAN NETWORK

Bayesian Network is a probabilistic graphical model (PGM) based on the probability calculation of each random variable (Heckerman, 1997). Random variables in the Bayesian Network are represented in the form of Directed Acyclic Graph (DAG) that consists of nodes and edges. Each random variable is a node in DAG while the relationship between variables is denoted by an edge.

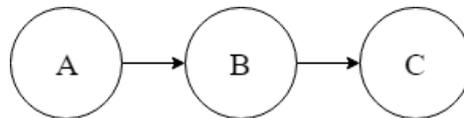


Fig. 1. Simple DAG (Sharma & Goyal, 2015)

In Figure 1 node A, node B and node C show the random variables and the edges show the relation between them. In Figure 1, there is an edge from node A to node B. This means that node A affects node B or in other words node A is the parent of node B. As for node C is the child of node B. Each node has a conditional probability table (CPT) containing the probability of a node given its parent. Suppose in the DAG containing of n nodes and its parent denoted by X . To calculate the probability of all nodes and its parents or called joint probability can be calculated using the equation below (Sharma & Goyal, 2015).

$$P(n_1, n_2, \dots, n_m) = \prod_{i=1}^m P(n_i | X_i) \quad (1)$$

III. RESEARCH METHOD

The aim of this paper is to predict the price of chilli commodity in Bandung Regency with the output of chilli prices that have been classified into economic class and not economic class. The design system that used in this paper is shown in Table I.

TABLE I
DESIGN SYSTEM

Design System
Input : Historical monthly data of chili prices and monthly weather data of Bandung Regency from year 2014-2016
Output : Price classification and performance evaluation
Preprocessing Data : Data discretization (detailed in section IIIB.1) Data partition (detailed in section IIIB.3)
Bayesian Network Algorithm : Structure learning to build the structure of Bayesian Network (detailed in section III.C) Parameter learning to calculate the conditional probability table (detailed in section III.D) Inference (detailed in section III.E)
Performance Evaluation Confusion Matrix

A. Dataset

The data used in this paper are historical monthly data of chilli commodity prices year 2014-2016 obtained from Department of Trade Bandung Regency as shown in Figure 2 and weather data of Bandung Regency obtained from Meteorology Climatology and Geophysics Agency (BMKG).

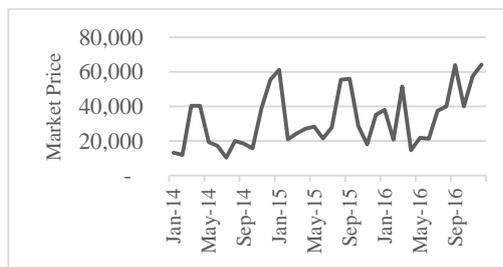


Fig. 2. Market price chart

There are 7 attributes that used for input data, the weather attributes consists of Solar Radiation (S), Wind Speed (W), Temperature (T), Relative Humidity (H), Evaporation (E), Rainfall (R) and Price (P) attribute. The sample data is shown in this following Table II.

TABLE II
SAMPLE DATA

No	S	W	T	H	E	R	P
1	42	4	22.5	82	3.3	309	13,161
2	47	4	22.9	80	3.6	89	12,000
3	52	3	23.3	82	3.4	421.7	40,323
4	58	3	23.7	81	3.7	217.6	40,323

B. Preprocessing

At the preprocessing stage, the data will be discretized and partitioned into training and testing data. Discretization is performed on all attributes because the type of the attributes is continuous as shown in Table II while Bayesian Network can work well on discrete attribute. The discretization done in this paper consists of three ways that will be described below.

1) Discretization

Equal width is one of the discretization techniques that can be used to discretized continuous data. Equal width discretization divides the data into n categories with the same range denoted by w (Liu, Hussain, Tan, & Dash, 2002). The maximum and minimum value of the attributes are needed. The equation (2) below used for equal width discretization.

$$w = \frac{(Maximum - Minimum)}{n} \quad (2)$$

Solar Radiation, Temperature, Relative Humidity and Evaporation were discretized into 3 categories (Nhita, 2013). As for Wind Speed attribute was discretized into 2 categories due to uneven distribution of data if it was discretized into 3 categories. As for Rainfall attribute was discretized by grouping the data into 4 categories based on BMKG rainfall category. The category is shown in Table III below (Nhita, 2013).

TABLE III
RAINFALL CATEGORY

Rainfall	Category
$\leq 20\text{mm}$	1
$\leq 50\text{mm}$	2
$\leq 100 \text{ mm}$	3
$>100 \text{ mm}$	4

2) Price Classification

In this paper chili prices is classified into economic class and not economic class. The class is determined by comparing the farmers price with the future value of production cost denoted by V . The production cost is the cost incurred during the chili planting process such as costs for seed, fertilizer, pesticide, land and machine. The farmers price is 50% of the market. This information obtained from direct interview with Association of Vegetable Farmers in Bandung Regency. As for determining the economic price, the monthly inflation rate (i) and the lowest price of chili commodity (Lp) are required. In this paper the lowest price used is 10,000 according to the interview with the farmers. The future value is calculated using this following equation (Ditakristy, Saepudin, & Nhita, 2016; Capiński & Zastawniak, 2003).

$$V = Lp \times e^{-i} \quad (3)$$

The inflation rate from year 2014-2016 reached 8.36%, 3.35% and 3.02% respectively based on the information obtained from Indonesian Central Bureau of Statistics with the average of the inflation rate is 4.91%. It uses a continuous interest rate so that the average inflation is 4% in a year which is 0.0033 for monthly inflation rate. After calculate the future value as in (3), chili prices then will be classified by comparing the farmers price and the future value to obtained economic price. The classification is shown in the Table IV below.

TABLE IV
PRICE CLASSIFICATION

No	Price	Class
1	Farmers Price \leq Future Value Price	Not Economic
2	Farmers Price $>$ Future Value Price	Economic

Figure 3 below shows the comparison between farmers price and economic price that used in this paper which shows that all of the data are mostly classified into economic class.

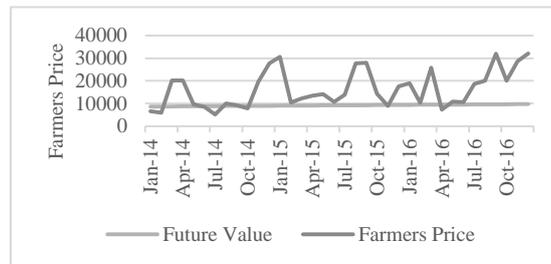


Fig. 3. The comparison chart between farmers price and economic price

From Figure 3 above it can be seen that the future value is almost linear, but in fact this value may increase or decrease depending on the cost of production incurred. For example if there is an increase in land prices, then the production costs incurred will be higher.

3) Data Partition

After the data has been discretized then partition is done to divide the data for training and testing. The Table V below is the partition of training and testing data and its class distribution.

TABLE V
DATA PARTITION

Data	Data Partition	Class distribution	
		Not Economic	Economic
January 2014 – November 2015	Training Data	22%	78%
December 2015 – November 2016	Testing Data	8%	92%

C. Structure Learning Bayesian Network

The aim of learning structure is to construct the structure of Bayesian Network. The Bayesian Network structure can be build manually based on domain knowledge or from data. To build the Bayesian Network structure from data, a special algorithm is required. In this paper the Bayesian Network structure was built using

the K2 algorithm. The K2 algorithm is a search and scoring algorithm for finding the best Bayesian Network structure with the initial assumption that all nodes do not have a parent. The algorithm was introduced by Cooper and Herskovits (1992). In the K2 algorithm all nodes were ordered and the maximum number of parent (k) for each node must be initialized. In this paper we selected $k = 1,2,3$ to avoid complex structures because the good structure of Bayesian Network are less complex but can clearly represent relationships between attributes. Suppose that in a database consisting of n cases, node x has a number of values denoted by r_x and the number of instantiation of parent denoted by v_x , the number of cases in attribute x with z^{th} value and the parent with y^{th} value defined by N_{xyz} . The K2 algorithm will search for the parent π_x that maximizes the value of the node with the following equation (Ruiz, 2009).

$$f(x, \pi_x) = \prod_{y=1}^{v_x} \frac{(r_x - 1)!}{(N_{xy} + r_x - 1)!} \prod_{z=1}^{r_x} N_{xyz!} \quad (4)$$

The K2 algorithm is described in this following Table VI (Ruiz, 2009).

TABLE VI
K2 ALGORITHM

K2 Algorithm
Input :
Number of node (x)
Dataset
Ordering node
Maximum number of parent (k)
Output :
Set of parent for each node (π)
Algorithm :
For i = 1 to x do
$\pi(i) = \{ \}$
Score = $f(i, \pi_i)$ as in (4)
If pred(i) = { } then (pred is predecessor node i)
Stop
else
While the number of $\pi(i) < k$ do
ScoreNew = $f(i, \pi_i \cup \{pred\})$
If ScoreNew > Score then
Score = ScoreNew
$\pi(i) = \pi(i) \cup \{pred\}$
End if
End while
End if
Output ($\pi(i)$)
End for

D. Parameter Learning Bayesian Network

Learning parameters Bayesian Network is the process of calculating the conditional probability table based on the structure of the graph that has been built. In this paper the calculation of conditional probability table is calculated using Maximum A Posterior (MAP) calculation to avoid zero probability. Suppose the probability of x node given its parent denoted by θ_{xyz} , N_{xyz} is the number of cases with the value of x node is z and the value of its parent is y, with $\alpha_{xy} = \sum \alpha_{xyz}$ and $N_{xy} = \sum N_{xyz}$ where α_{xyz} is a bias probability as in (6) where r_x is the number of value in node x and q_x is the number of instantiation of parent of node x, the MAP equation is as follow (Heckerman, 1998).

$$\theta_{xyz} = \frac{\alpha_{xyz} + N_{xyz}}{\alpha_{xy} + N_{xy}} \tag{5}$$

$$\alpha_{xyz} = \frac{0.1}{q_x \times r_x} \tag{6}$$

E. Inference Bayesian Network

Inference on the Bayesian Network is a calculation to determine the class probability of an event based on existing structure and CPT. Inference on the Bayesian Network uses the basis of Bayes's theorem describing the probability of two events such as in (7) which describes the probability of event X with the condition that event Y occurs first (Sarshar, Granmo, Radianti, & Gonzalez, 2013).

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \tag{7}$$

In this case, the probability of economic class and not economic class is calculated based on existing graph structure and CPT. The class with a greater probability value become final decisions.

F. Performance Evaluation

Performance evaluation is needed to analyze and evaluate the performance of the system in classifying prices. The confusion matrix consists of rows and columns that contain the amount of data that has been classified. With confusion matrix, the amount of data predicted correctly or wrongly is known. Here is an example of confusion matrix (Han, Pei, & Kamber, 2011).

TABLE VII
CONFUSION MATRIX

Predicted	Actual	
	Positive	Negative
Positive	PP	NP
Negative	PN	NN

In Table VII, PP is the number of positive class that were correctly predicted as a positive class by system, NP is the number of negative class that were predicted as a positive class by system, PN is the number of Positive class that were predicted as a negative class by system and NN is the number of negative class that were correctly predicted as a negative class by system. From the confusion matrix above, we can calculate pecision, recall, F1 score and accuracy to know the system's performance.

1) *Precision*

Precision is a percentage that shows the exactness of system in predicting the data into positive class. Precision value can be calculated using this equation (7) where PP and NP obtained from the confusion matrix as shown in Table VII (Han, Pei, & Kamber, 2011).

$$Precision = \frac{PP}{PP + NP} \quad (8)$$

2) *Recall*

Recall shows the sensitivity of the system in predicting the positive class. Recall can be calculated using this equation (8) where PP and PN obtained from the confusion matrix as shown in Table VII (Han, Pei, & Kamber, 2011).

$$Recall = \frac{PP}{PP + PN} \quad (9)$$

3) *F1-Score*

F1-Score is a value that shows the system's performance in doing prediction. The greater value of F1-Score generated, the better the system's performance. F1-Score can be calculated using this following equation (10) (Han, Pei, & Kamber, 2011), where precision is obtained from equation (5) and recall is obtained from equation (6).

$$F1\ Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (10)$$

4) *Accuracy*

Accuracy is a percentage of the system that was predict the data correctly. The equation below (10) is used to calculate accuracy (Han, Pei, & Kamber, 2011).

$$Accuracy = \frac{(PP + NN)}{(PP + NP + PN + NN)} \quad (11)$$

IV. RESULTS AND DISCUSSION

The research was done by using 2 scenarios that is dataset scenario and learning scenario of Bayesian Network's structure. Dataset scenarios uses data with all attributes and uses rainfall and price attributes. For learning scenario of Bayesian Network's structure we selected k = 1, 2, 3 as a maximum number of parent for Bayesian Network's structure.

A. *Dataset Scenario*

This study used 2 scenarios of dataset as detailed in Table VIII. The aim of this scenario is to analyze the effect of using certain attributes to predict chili prices. The dataset scenario is as follow.

TABLE VIII
DATASET SCENARIO

Scenario	Attributes
1	All attributes
2	Rainfall and price

B. Results of Performance Evaluation

The resulting structure and dataset scenario were evaluated to know the system performance. The Table IX below shows the performance evaluation of the scenarios.

TABLE IX
PERFORMANCE EVALUATION

Scenario	k	Testing Data				Training Data				Average Accuracy
		Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy	
1	1	0.9	0.82	0.86	75%	0.83	0.83	0.83	74%	74.5%
2		0.92	1	0.96	92%	0.78	1	0.88	78%	85%
1	2	0.92	1	0.96	92%	0.85	0.94	0.89	82%	87%
1	3	0.9	0.82	0.86	75%	1	0.94	0.97	96%	85.5

Scenario 2 can not be done by using k = 2, 3 because the number of attribute used are 2 attributes that is Rainfall and Price. So that the maximum number of parent must be 1. For the average accuracy of training and testing data are 82.5% and 83.5% respectively.

1) The influence of k analysis

The dataset scenario were used to create a Bayesian Network structure. From the results obtained by choosing k = 1,2 and 3 produce different structures. Here is the explanation of each structure generated.

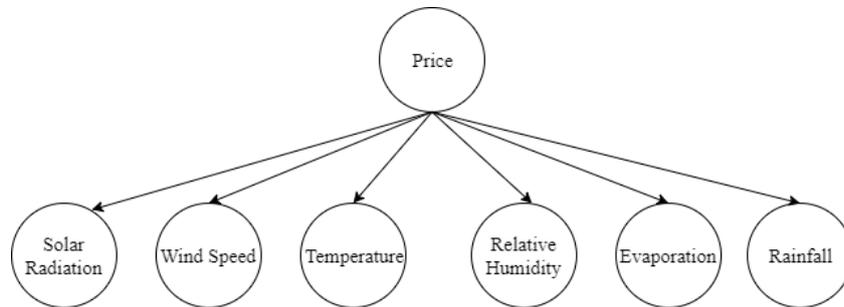


Fig. 4. Structure of Bayesian Network with k = 1

The inference using the structure in Figure 4 is as follow (Witten, Frank, Hall, & Pal, 2016).

$$Pr(P|S, W, T, H, E, R) = \frac{Pr(S|P) \times Pr(W|P) \times Pr(T|P) \times Pr(H|P) \times Pr(E|P) \times Pr(R|P) \times Pr(P)}{Pr(S, W, T, H, E, R|P) + Pr(S, W, T, H, E, R|\sim P)} \tag{12}$$

From Figure 4 it can be seen that the price node is directly related to other nodes which in Figure 4 shows that the price node becomes the parent of all nodes. This shows that there is a relationship between price variable and weather variables. Where price variables can affect other variables. As for the other node which is weather variables there is no direct relationship. As for the structure k = 2 produce the structure as follows.

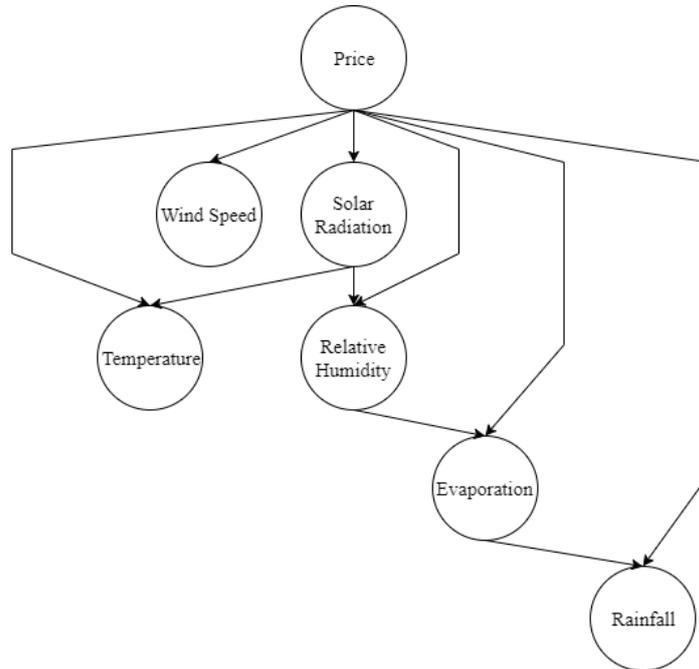


Fig. 5. Structure of Bayesian Network with k = 2

The inference using the structure in Figure 5 is as follow (Witten, Frank, Hall, & Pal, 2016).

$$Pr(P|S, W, T, H, E, R) = \frac{Pr(S|P) \times Pr(W|P) \times Pr(T|S, P) \times Pr(H|S, P) \times Pr(E|H, P) \times Pr(R|E, P) \times Pr(P)}{Pr(S, W, T, H, E, R|P) + Pr(S, W, T, H, E, R|\sim P)} \quad (13)$$

Figure 5 gives the same result as Figure 4 which price node becomes the parent of all nodes which implies a direct edge between the price node and other nodes. The difference is in the weather variables. In Figure 5 the weather variables are connected. This means that there is a relation between them. For example, evaporation variable has a direct relation to rainfall variable where there is an edge from evaporation node to rainfall node. So the value of evaporation variable will affect the value of rainfall variable. As for the structure with k = 3 produces the structure as follows.

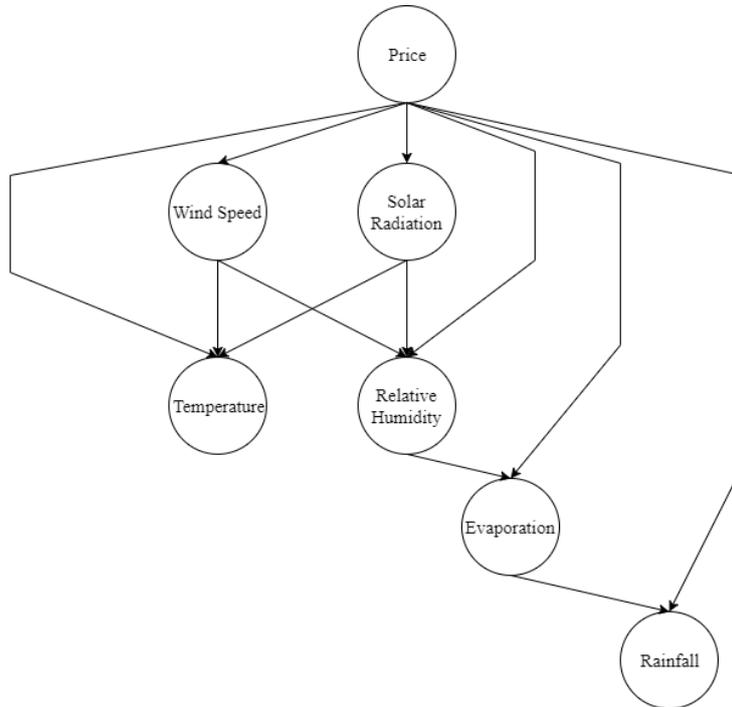


Fig. 6. Structure of Bayesian Network with k = 3

The inference using the structure in Figure 6 is as follow (Witten, Frank, Hall, & Pal, 2016).

$$Pr(P|S, W, T, H, E, R) = \frac{Pr(S|P) \times Pr(W|P) \times Pr(T|S, W, P) \times Pr(H|S, W, P) \times Pr(E|H, P) \times Pr(R|E, P) \times Pr(P)}{Pr(S, W, T, H, E, R|P) + Pr(S, W, T, H, E, R|\sim P)} \quad (14)$$

The resulting structure with k = 3 is not much different from the resulting structure k = 2. The difference is in the temperature and humidity nodes which the two variables are both influenced by the wind speed and solar radiation variables. As for the structure of dataset scenario 2 is as follow.

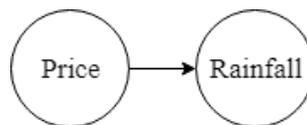


Fig. 7. Structure of Bayesian Network for dataset scenario 2

The inference using the structure in Figure 7 is as follow (Witten, Frank, Hall, & Pal, 2016).

$$Pr(P|R) = \frac{Pr(R|P) \times Pr(P)}{Pr(R|P) + Pr(R|\sim P)} \quad (15)$$

The result of the structure shows that the price node affects the rainfall node similar to the structure in the dataset scenario 1.

From the results of the structure obtained using K2 Algorithm, the overall structure produced is good. The relationship of each attribute is well illustrated. From the classification performance using the structure obtained, from Table IX it can be seen that in training data, the more the number of k then the performance

increases. However, if it seen from the testing data, for $k = 3$ has lower performance than $k = 2$. The performance value of data testing for $k = 2$ for precision is 0.92, recall is 1 F1-score is 0.96 with 92% accuracy. From the results of training and testing data, the best graph structure is $k = 2$ with a dataset scenario 1 that shows greater accuracy than $k = 1$ in training data and greater accuracy than $k = 1$ and $k = 3$. This difference caused by the uneven distribution of data on each attribute. Therefore the number k has an influence on the classification result. With the increasing number of k , the more parent a node can have. The more parent of a node affects the more combination of conditional probability. Overall the best structure for Bayesian Network is $k = 2$ which described the relation of all variables.

2) *Dataset Analysis*

Based on the results in Table VIII, each dataset scenario gives a good performance. Scenario 1 is the best dataset scenario for training which gives 1 and 0.94 for precision and recall respectively using $k = 3$. While the scenario 1 with $k = 2$ is the best for testing with the precision and recall values are 0.92 and 1 respectively. This means that each scenario can be used to predict the chili prices. Here is the confusion matrix of scenario 1 with $k = 2$.

TABLE X
 CONFUSION MATRIX OF SCENARIO 1 WITH $k=2$

Predicted	Actual			
	Training		Testing	
	Economic	Not Economic	Economic	Not Economic
Economic	17	3	11	1
Not Economic	1	2	0	0

From Table X it can be seen that the system mostly predict the price into economic class. This means that the price of chili in the next month is still classified into an economic class for farmers. So farmers can plant chili next month. The prediction class where the economic class is dominant can be caused by due to the uneven distribution of data used in training data and testing data. Where in the training data or testing the majority of the class is an economic class. In addition, the amount of data used was not much. With uneven data distribution and lack of amount of data used make the system difficult to find prediction patterns.

V. Conclusion

Weather variables can be used to predicting chili prices because there is a relation between weather variables and chili prices. As shown in the learned structure of Bayesian Network there is an edge from node price to other nodes which is weather variables. The best structure of Bayesian Network has 2 maximum number of parent which described the relation between every node. Price predictions using the Bayesian Network algorithm provides a good performance with precision, recall and F1 score for testing data are 0.92, 1, 0.96 respectively. As for the training data it gives 0.85, 0.94, 0.89 respectively with the average accuracy of testing is 83.5%. The number of parent of node affects the performance of the built system. The more number of parent on a node the more probability combinations will be calculated and this could happen for prediction errors. The uneven distribution of data also affects the performance of the system. In this paper, the data used is not much and the economic class is dominant. This distribution makes the system less understand the prediction pattern and predicts most of the data into economic class.

ACKNOWLEDGMENT

The authors would like to thank the School of Computing Telkom University for supporting this study and also to the government of Bandung Regency who has helped this study by providing data.

REFERENCES

- Capiński, M., & Zastawniak, T. (2003). *Mathematics for Finance: An Introduction to Financial Engineering*. Springer.
- Chawla, V., Naik, H. S., Akintayo, A., Hayes, D., Schnable, P., Ganapathysubramanian, B., & Sarkar, S. (2016). A Bayesian Network approach to County-Level Corn Yield Prediction using historical data and expert knowledge. arXiv preprint arXiv:1608.05137.
- Cooper, G. F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine learning*, 9(4), 309-347.
- Ditakristy, M. L., Saepudin, D., Nhita, F. (2016). Analysis And Implementation Of Radial Basis Function Neural Network In Forecasting Of Agricultural Commodity Prices. e-Proceeding of Engineering Vol.3, No.1 April 2016 (pp.1130)
- Eisuke, K. I. T. A., Harada, M., & Mizuno, T. (2012). Application of Bayesian Network to stock price prediction. *Artificial Intelligence Research*, 1(2), 171.
- Firdaus, M., & Gunawan, I. (2013). Integration among regional vegetable markets in Indonesia. *Journal of ISAAS*, 8(2), 106-116
- Gandhi, N., Armstrong, L. J., & Petkar, O. (2016, September). Predicting Rice crop yield using Bayesian networks. In *Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on* (pp. 7105-71010). IEEE.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier
- Heckerman, D. (1997). Bayesian networks for data mining. *Data mining and knowledge discovery*, 1(1), 79-119
- Heckerman, D. (1998). A tutorial on learning with Bayesian networks. *Nato Asi Series D Behavioural And Social Sciences*, 89, 301-354.
- Hossen, S. M. (2016). A Statistical Analysis on Production of Chili and Its' Prospect in Bangladesh
- Kaur, M., Gulati, H., & Kundra, H. (2014). Data mining in Agriculture on crop price prediction: Techniques and Applications. *International Journal of Computer Applications*, 99(12), 1-2
- Liu, H., Hussain, F., Tan, C. L., & Dash, M. (2002). Discretization: An enabling technique. *Data mining and knowledge discovery*, 6(4), 393-423.
- Mariyono, Joko, and Sumarno Sumarno. "Chilli production and adoption of chilli-based agribusiness in Indonesia." *Journal of Agribusiness in Developing and Emerging Economies* 5.1 (2016): 57-75.
- Newlands, N. K., & Townley-Smith, L. (2011). Predicting energy crop yield using bayesian networks. In *Proceedings of the Fifth IASTED International Conference* (Vol. 712, pp. 015-116).
- Nhita, F. (2013, March). A rainfall forecasting using fuzzy system based on genetic algorithm. In *Information and Communication Technology (ICoICT), 2013 International Conference of* (pp. 111-115). IEEE.
- Parle, M., & Kaura, S. (2013). A hot way leading to healthy stay. *Int Res J Pharm*, 3(6), 21-25.
- Pérez-Ariza, C. B., Nicholson, A. E., & Flores, M. J. (2012). Prediction of coffee rust disease using bayesian networks. In *Proceedings of the Sixth European Workshop on Probabilistic Graphical Models* (pp. 259-266).
- Ramesh, D., & Vardhan, B. V. (2013). Data mining techniques and applications to agricultural yield data. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(9), 3477-80.
- Rasmussen, S., Madsen, A. L., & Lund, M. (2014). Bayesian network as a modelling tool for risk management in agriculture (No. 2014/13). University of Copenhagen, Department of Food and Resource Economics.
- Ruiz, C. (2009). Illustration of the K2 algorithm for learning Bayes net structures. Department of Computer Science WPI, Bayesian Network Power Constructor, Worcester, MA.
- Sarshar, P., Granmo, O. C., Radianti, J., & Gonzalez, J. J. (2013, April). A Bayesian network model for evacuation time analysis during a ship fire. In *Computational Intelligence in Dynamic and Uncertain Environments (CIDUE), 2013 IEEE Symposium on* (pp. 100-107). IEEE.
- Sharma, A., & Goyal, M. K. (2015, November). Bayesian network model for monthly rainfall forecast. In *Research in Computational Intelligence and Communication Networks (ICRCICN), 2015 IEEE International Conference on* (pp. 241-246). IEEE.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann

