VARIABLE CORRELATION ANALYSIS OF ENVIRONMENTAL CONDITIONS ON QUALITY OF BROILER PRODUCTION AT PT.MUSTIKA SUKSES NIAGATAMA

Adam Mukharil Bachtiar, S.Kom., M.T. [1], Bayu Firmawan Paoh[2]

[1,2] Teknik Informatika – Universitas Komputer Indonesia
Jl. Dipati Ukur No. 112-116, Bandung, 40132
E-mail : adammb@email.unikom.ac.id[1], bayupaoh@email.unikom.ac.id[2]

ABSTRACT

PT. Mustika Sukses Niagatama is a company engaged in the field of broiler farms. Increasing profits and customer satisfaction the company made efforts to improve the quality of broiler production. Efforts made were not optimal because the results obtained in the field is often different from the expected results. This happens because the company does not know the information linkage of environmental conditions of the cage to the quality of broilers.

A way to get information on the results of such analysis, is data mining because one of the functions of data mining is to perform correlation analysis. There are two types of correlation that can be analyzed in this research, namely one to one relationship (bivariate) and multiple to one relationship (multivariate). Based on the quantitative data, the used algorithm are Pearson (bivariate) moment, multiple linear regression (multivariate), canonical correlation (multivariate).

Correlation analysis using Pearson moment product algorithm, multiple linear regression and canonical correlation provides information in the form of correlation of each independent variable to dependent. Thus, the results of this study has helped the owner of the company get information linkage of environmental conditions of the cage to the quality of broilers at PT. Mustika Sukses Niagatama.

Keywords: Data mining, correlation analysis, Pearson moment product, multiple linear regression, canonical correlation.

1. INTRODUCTION

1.1. Background

PT. Mustika Sukses Niagatama is a company engaged in the field of broiler farms. To maintain customer loyalty and to increase profits, the company always maintains the quality of broiler. The company knows the quality of broiler by taking data from IoT device and also input data from prepared form then the system will calculate the quality of broiler using predetermined formula and visualized it to user. To improve the quality of broiler production, the company made several efforts such as increasing chicken weight, providing enough feed and keeping the chicken coop temperature. The efforts made were not optimal because the results obtained in the field are often different even though using the same method.

Based on interviews conducted with the owner of the cage is known that the company knows how to improve the quality of chicken production only from the articles on the internet and directly implement without knowing how much influence the environmental factors on the quality of chicken obtained. In addition, the running system only has a function to monitor the environmental conditions and quality of broiler, but the system does not provide information on how much influence the environmental factors on the quality of chicken. This is what causes when there is a problem with the environmental conditions of the cage such as high temperatures or other problems, can cause the quality of broiler decreased sharply.

In the science field of IT there is method to dig important information from a data, it is called data mining. Data mining is a process to extract important information from a data in the form of patterns and predict future trends based on the pattern [1]. In its science, data mining has many methods that are divided by function and purpose. In this study, researchers used correlation analysis method. Correlation analysis (correlation analysis) measures the proximity of the relationship between two or more variables. The technique considers the combined variation of the two measurements, which are not limited by the experimenter [2].

Based on the problem, data mining will be used by researchers in conducting the research in order to conduct correlation analysis of the environmental condition of the cage on the quality of broiler production. Therefore, based on the explanation of the facts and problems that have been described, the solution is obtained by conducting a correlation analysis of the environmental conditions of the cage

against the quality of broiler production at PT. Mustika Sukses Niagatama.

1.2. Research Purposes

The purpose of this study is to help the owner of the cage by providing information on the relevance of the environmental conditions of the cage to the quality of broiler production at PT. Mustika Sukses Niagatama.

2. MATERIALS AND METHODS

2.1. Research Method

The step of solving this problem in this study takes the reference of "Cross-Industry Standard for Data Mining (CRISP-DM)" [3]. The flow in this study can be seen on Image 1.

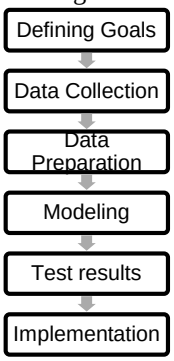


Image 1. Research step

2.2. Discussion

1. Defining Goals

Business goal to be achieved is to improve the quality of broiler so, the benefits can be gained by larger companies. The purpose of data mining is to obtain the following information.

a. Homeowner needs information on the effects of moisture on ammonia.
b. Homeowner needs information on the effects of temperature, humidity and ammonia according to the death of the broilers.
c. Homeowner needs information on the effects of temperature, humidity and ammonia on IP and the sale price of chickens.

2. Data Collection

The used data is from the previous harvest data on December 2016. Data sourced from Firebase Database. Then it is stored in the relational database. This is needed because Firebase Database has a weakness when the system is built requires complex queries. The explanation of each stored data can be seen on Table 1.

Table 1. Attribute explanation

Attribute	Explanation
Date	Date of data retrieval.
Temperature	Average cage temperature in each cage grid on that date.
Humidness	Average humidity enclosure in each cage grid on that date.
Ammonia	Rata-rata amonia kandang di setiap *grid* kandang di tanggal tersebut.
Death	Average chicken on that date.
IP	The value of the index performance earned on that date.
Sale Price	Predicted selling price of chicken obtained on that date.

3. Data Preparation

1. Data Selection

Based on the purpose of data mining in the process of determining the objectives of analyzing the relationship between environmental conditions variable to broiler quality variables, the attributes used are Temperature, Humidity, Ammonia, Death, IP, and Sale Price. Table 2 is the result of the process of the selection of attributes. The data is a sample of data to be used in this study.

Table 2. Data Selection Results

Temperature	Humidness	Ammonia	Death	IP	Sale Price
20	68	11	16	270	11626
24	64	9	15	273	11824
0	0	0	9	274	11870
21	65	10	15	275	11982
26	66	8	7	277	12099
25	62	14	7	279	12128
23	68	14	13	281	12244
22	63	12	13	283	12423
25	63	14	18	285	12657
21	71	13	11	287	13065
21	70	13	9	290	13182
23	71	13	15	292	14998
30	70	9	15	292	13928
20	61	8	14	294	13220
22	67	12	15	299	13298
26	68	9	8	301	13495
21	60	15	7	303	13529
24	68	12	15	310	13752
24	65	15	15	314	13823
23	72	13	16	319	14123
26	69	14	7	320	14526

2. Data Cleanup

To eliminate rows that have invalid values there are two processes namely cleaning and outliers analysis. For the process of cleaning the deleted data is invalid data. The minimum value and maximum value of a data is said to be valid biased views on Table 3.

Table 3. Minimum and maximum data limits are valid

No.	Attribute	Minimum Value	Maximum Value
1	Temperature	18 C	28 C
2	Humidness	60%	80%
3	Ammonia	0	20
4	Death	0	-
5	IP	0	-
6	Sale Price	0	-

To find outliers value using Equation (1).

$$t_{1,2} = \bar{x} \pm 2s \quad (1)$$

Where $\bar{x}$ is *mean* and *S* is the standard deviation. The results obtained can be seen in Table 4.

Table 4 *outliers* analysis Results

	Temperature	Humidness	Ammonia	Death	IP	Sale Price
$\bar{x}$	23,00	66,37	12,05	12,42	292,21	12860,21
S	2,03	3,55	2,27	3,73	15,60	904,04
t_1	27,06	73,46	16,60	19,88	323,42	14668,29
t_2	18,94	59,28	7,51	4,96	261,00	11052,13

The results of the cleanup data can be seen on Table 5.

Table 5. Data Cleanup Results

Temperature	Humidness	Ammonia	Death	IP	Sale Price
20	68	11	16	270	11626
24	64	9	15	273	11824
21	65	10	15	275	11982
26	66	8	7	277	12099
25	62	14	7	279	12128
23	68	14	13	281	12244
22	63	12	13	283	12423
25	63	14	18	285	12657
21	71	13	11	287	13065
21	70	13	9	290	13182
20	61	8	14	294	13220
22	67	12	15	299	13298
26	68	9	8	301	13495
21	60	15	7	303	13529
24	68	12	15	310	13752
24	65	15	15	314	13823
23	72	13	16	319	14123
26	69	14	7	320	14526

4. Modeling

a. Implementation of Pearson Moment Product Correlation Analysis

In this process, the correlation sought is bivariate correlation (one independent variable with one dependent variable). This method is used because the scale of data measurement used is ratio [4]. The method used to find the relationship between moisture to ammonia is the Person Moment Product Correlation. Steps from Pearson's Product Moment Correlation method, i.e. :

1. Determine the dependent variable and independent variable.

The purpose of data mining to be achieved is the owner of the cage needs information on the influence of humidity against ammonia, then got the independent variable is the variable moisture and the dependent variable is the ammonia variable.

2. Form a statistics table.

The establishment of statistical tables is done to facilitate the calculation of the correlation coefficient on the technique to be used. Where $\sum x_i$ is the sum of the temperature values and $\sum y_i$ is amount of moisture value. The results of statistical table calculations can be seen on Table 6.

Table 6. The calculation of statistics table

$\sum i$	$\sum x_i$	$\sum y_i$	$\sum x_i y_i$	$\sum x^2$	$\sum y^2$
19	1190	216	14290	78876	2684

3. Calculate the correlation coefficient calculation.

4. To calculate the correlation coefficient using Equation (2).

$$r = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{\{n(\sum x_i^2) - (\sum x_i)^2\}.\{n(\sum y_i^2) - (\sum y_i)^2\}}} \quad (2)$$

The correlation obtained coefficient is 0.0730.

5. Define the relationship criteria

Value $r_{x,y}$ which is valued at 0.0730 it can be concluded that humidity has a very weak effect on ammonia and is positive. When there is an increase in humidity then the ammonia will increase.

6. Visualize correlation results

After knowing the Criteria, the next step is to convey the information universally to be easy to understand. In a previous study, illustrating the relationship between two or more variables using a scatter plot [5]. Visualization using simple text, because the information conveyed in the form of meaning of correlation coefficient. To further facilitate the delivery of information obtained results then the user will get the information is whether the variables are calculated correlation has a correlation or not [6]. If it has a correlation it will convey the relationship between the variable whether it has a positive or negative relationship. A variable is said to have a correlation if the value of correlation coefficient is greater 0.4. So the result of the visualization of the correlation result is that moisture has no correlation to ammonia.

b. Implementation of Multiple Linear Regression Correlation Analysis

In this process, the correlations sought included multivariate correlations. This method is used when the number of independent variables is more than one with one dependent variable and ratio measurement scale [4]. The method used to find the relationship between temperature, humidity and ammonia to death is the correlation of multiple linear regression. The steps of the Linear Multiple Regression Correlation method, i.e. :

1. Determine the dependent variable and independent variable.

The purpose of data mining to be achieved is the owner of the cage needs information on the influence of temperature, humidity and ammonia to death, then get the independent variable that is variable temperature, humidity and ammonia. As for the dependent variable is the death variable.

2. Form a statistics table.

The establishment of statistical tables is done to facilitate the calculation of the correlation coefficient on the technique to be used. The results of statistical table calculations can be seen on Table 7.

Table 7 Multiple linear regression statistics table

$\sum_i$	19
Temperature ($\sum x_1$)	414
Humidness ($\sum x_2$)	1190
Ammonia ($\sum x_3$)	216
Death ($\sum Y$)	221
$\sum x_1^2$	9596
$\sum x_2^2$	78876
$\sum x_3^2$	2684
$\sum x_1.y$	5042
$\sum x_2.y$	14620
$\sum x_3.y$	2645

3. Search for b_i vector values.
   To get the value b_i obtained from Equation (3).

$$\begin{bmatrix} n & \sum X_1 & \sum X_2 & \sum X_3 \\ \sum X_1 & \sum X_1^2 & \sum X_2 X_1 & \sum X_2 X_1 \\ \sum X_2 & \sum X_2 X_1 & \sum X_2^2 & \sum X_2 X_1 \\ \sum X_3 & \sum X_2 X_1 & \sum X_2 X_1 & \sum X_3^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \\ \sum X_3 Y \end{bmatrix} \quad (3)$$

According from Table 7 and Equation(3), Then get the following matrix :

$$\begin{bmatrix} 18 & 414 & 1190 & 216 \\ 414 & 9596 & 27377 & 4970 \\ 1190 & 27377 & 78876 & 14290 \\ 216 & 4970 & 14290 & 2684 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 221 \\ 5042 \\ 14620 \\ 2645 \end{bmatrix}$$

Through the equation, then vector b can be calculated using the Gaussian elimination method. From the calculation, can be produced vector b, i.e.:

$$b = \begin{bmatrix} 21{,}420 \\ 0{,}4293 \\ 0.0690 \\ 0.0712 \end{bmatrix}$$

So that can be written Equation, i.e. :

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k \quad (4)$$

The result of the implementation of Equation (4).

$$Y = 21{,}420 + 0.4293 X_1 + 0.0690 X_2 + 0.0712 X_3$$

4. Define the relationship criteria.
   Based on these equations it can be determined the criteria of the relationship of variable X to Y, where :
   a. Temperature has a correlation has a moderate correlation rate with death (0.4293) and is positive, so if there is a rise in temperature then the chicken mortality rate will increase.
   b. Humidity has a very weak correlation rate with death (0.0690) and positive value, so that when the humidity rises then the death rate will rise as well.
   c. Ammonia has a very weak correlation rate with mortality (0.0714) and is positive, so if the ammonia rate drops then the chicken mortality rate will decrease.

5. Visualize correlation results.

The results of the visualization of the correlation to the user, are :
a. Temperature has a correlation to death. If there is an increase in temperature then the chicken mortality rate will increase.
b. Humidity has no correlation to death.
c. Ammonia has no correlation to death.

3. Implementation of Canonical Correlation Analysis
   In this process, the correlations sought included multivariate correlations. This method is used when the number of independent variables is at least one with the number of dependent variables more than one and the ratio measurement scale [4]. The method used to find the relationship between temperature, humidity and ammonia to IP and Selling Price is a canonical correlation. The steps of the Canonical Correlation method, i.e. :
   a. Determine the set of dependent and independent variables.
      The owner of the cage needs information on the effect of temperature, humidity and ammonia on IP and the selling price, then the result of independent variables here is temperature, humidity, ammonia and death. For the dependent variable is IP and the selling price.
   b. Calculate the R correlation matrix.
      The structure of the correlation matrix R can be seen on Equation (5).

$$R = \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix} \quad (5)$$

Where R_{xx}, R_{xy}, R_{yx}, and R_{yy} are :
R_{xx} : The correlation matrix for the group of variables x.
R_{yy} : The correlation matrix for the group of variables y.
R_{xy} : The correlation matrix between the group of variables x and y variables.
R_{yx} : The correlation matrix between the group of variables y and x variables.

Each of the correlation matrix R_{xx}, R_{xy}, R_{yy} and R_{yx} contains correlation coefficients $r_{x,y}$. $r_{x,y}$ that obtained from Equation (2).

As the results R_{xx} can be seen on Tabel 8.

Tabel 8. R_{xx} Calculation Results

R_{xx}	Temperature	Humidness	Ammonia
Suhu	1,000	0,057	0.024
Kelembapan	0,057	1,000	0,7303
Amonia	0,024	0,7303	1,000

R_{yy} calculation results can be seen on Table 9.

Table 9. R_{yy} Calculation Results

R_{yy}	IP	Sale Price
Ip	1,000	0,977
hargajual	0,977	1,000

R_{xy} calculation results can be seen on Table 10.

Table 10. R_{xy} Calculation Results

R_{xy}	IP	Sale Price
Suhu	0.2370	0.1835
kelembapan	0.2799	0.3092
Amonia	0.3953	0.3757

R_{yx} calculation results can be seen on Table 11.

Table 11. R_{yx} Calculation Results

R_{yx}	Temperature	Humidness	Ammonia
IP	0.2370	0.2799	0.3053
Sale Price	0.1835	0.3093	0.3767

c. Determine the value of λ, value a_i, and value b_i.

Value λ obtained from the eigen value of $R_{yy}^{-1}R_{yx}R_{xx}^{-1}R_{xy}$ and value b_i obtained from the eigen vector of $R_{yy}^{-1}R_{yx}R_{xx}^{-1}R_{xy}$. The matrix of matrix multiplication $R_{yy}^{-1}R_{yx}R_{xx}^{-1}R_{xy}$ is $\begin{bmatrix} 0.3819 & 0.2819 \\ -0.1192 & -0.0285 \end{bmatrix}$. The formula of obtaining eigenvalue can be seen on Equation (6).

$$\det(R_{yy}^{-1}R_{yx}R_{xx}^{-1}R_{xy} - \lambda \ \boldsymbol{I}) \qquad (6)$$

The obtained Eigen value is $\boldsymbol{\lambda_1} = 0.2689$ and $\boldsymbol{\lambda_2} = 0.0845$. Eigen value and eigen vector from matrix $R_{yy}^{-1}R_{yx}R_{xx}^{-1}R_{xy}$ can be seen in Table 12.

Table 12. Value b_i

i	λ^2	Λ	Eigen Vector *(b_i)*	
1	0.0723	0.2689	0.9282	-0.6880
2	0,0071	0.0845	-0.3721	0.7257

Value a_i obtained from $R_{xx}^{-1}R_{xy}bi$. The results can be seen on Table 13.

Table 13. Value a_i

i	a_i		
1	0.1396	0.1211	0.2144
2	-0.0319	0.0337	-0.0002

d. Interpret the results.

Interpretation is done to measure the relation of independent variable to dependent variable. For this step use canonical loading method. This method is chosen because it has good accuracy. To know the independent variable which has the biggest influence to the dependent variable, can use Equation (7).

$$Corr(X, Ui) = Rxx\ a_i \qquad (7)$$

The correlation relationship and the magnitude of the correlation of each dependent variable can be obtained using Equation (8).

$$Corr(Y, Vi) = Ryy\ b_i \qquad (8)$$

Value *i* which is used in a_i and b_i is *i* that has λ_i^2 with the greatest value. Value *i* that used is 1 then a_i that used is a_1 and b_i is b_1. Implementation of Equations (7) to obtain $Corr(X, Ui)$ are :

$$\begin{bmatrix} 1 & 0.057 & 0.024 \\ 0.057 & 1 & 0.073 \\ 0.024 & 0.073 & 1 \end{bmatrix} \begin{bmatrix} 0.1396 \\ 0.121 \\ 0.2144 \end{bmatrix} = \begin{bmatrix} 0.1517 \\ 0.1447 \\ 0.2266 \end{bmatrix}$$

Implementation of Equations (8) to obtain $Corr(Y, Vi)$ are :

$$\begin{bmatrix} 1 & 0.977 \\ 0.977 & 1 \end{bmatrix} \begin{bmatrix} 0.9282 \\ -0.688 \end{bmatrix} = \begin{bmatrix} 0.256 \\ 0.2188 \end{bmatrix}$$

The results of data recap correlation coefficient can be seen on Tabel 14.

Tabel 14. Recap the correlation coefficient data

Independent Variables	Independent Variable Correlation Coefficient on Dependent Variables	Correlation Coefficient of Dependent Variable to Independent Variables	
		IP	Sale Price
Temperature	0.1517	0.256	0.2188
Humidness	0.1447	0.256	0.2188
Ammonia	0.2266	0.256	0.2188

e. Finding relationship criteria.

Based on these equations it can be determined the criteria of the relationship set of variable X to the set of variables Y, where :

1. Temperature has a very weak effect (0.1517) on IP and Selling Price. If the temperature increases then the IP will increase and the Selling Price will increase.
2. Humidity has a very weak effect (0.1447) on IP and Selling Price. If the humidity increases then the IP will increase and the Selling Price will increase.

3. Ammonia has a weak influence (0.2266) on IP and Selling Price. If the ammonia happens, the IP will increase and the Selling Price will increase.

f. Visualize the final result

The results of the visualization of the correlation to the user, are :

1. Temperature has no correlation to IP and Selling Price.
2. Humidity has no correlation to IP and Selling Price.
3. Ammonia has no correlation to IP and Selling Price.

2.3. Accuracy Testing

Testing the measurement accuracy of the Pearson product moment algorithm algorithm, multiple linear regression and canonical correlation performed by the system. Testing accuracy measurements using sample testing testing types. The system will try to analyze the data that has been prepared and will be compared with the results of manual calculations. For data source using Table 2. then obtained the test results can be seen on Tabel 15.

Tabel 15. Tested Results

Statement	Manual	System
Effect of Humidity Against Ammonia	0.0730	0.0730344352
Effect of Temperature on Death	0.4293	0.5586541784
Effect of Humidity on Death	0.0690	0.0690433173
Effect of Ammonia on Death	0.0714	0.0714470088
Effect of Temperature on IP and Selling Price	0.1517	0.1517182435
Effect of Moisture Against IP and Selling Price	0.1447	0.1447126762
Effect of Ammonia on IP and Selling Price	0.2260	0.2266662426

3. RESULTS

3.1. Conclusion

The application of some processes in the preprocessing data phase has a major influence on the results to be gained. Then the system built has helped the owner of the cage to get information on the influence of environmental variables on the quality of broiler. Therefore, the results of this study has helped the owner of the company get information linkage of environmental conditions of the cage to the quality of broiler production at PT. Mustika Sukses Niagatama.

3.2. Suggestion

Here are some suggestions that can be used as consideration in the next research that is required special research on visualization of correlation analysis of the layman.

BIBLIOGRAPHY

[1] F. Lee and J. Santana, *Data Mining: Meramalkan Bisnis Perusahaan*. Jakarta: Elex Media Komputindo, 2010.

[2] G. A. Churchill and T. J. Brown, *Basic Marketing Research*, Fourth Edi. San Diego, California: Harcout, 2001.

[3] D. T. Larose, *Discovering Knowledge in Data*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2004.

[4] R. Kurniawan and B. Yuniarto, *Analisis Regresi – Dasar dan Penerapannya dengan R*. Jakarta: KENCANA PRENADA MEDIA, 2016.

[5] N. A. Syaripul and A. M. Bachtiar, "VISUALISASI DATA INTERAKTIF DATA TERBUKA PEMERINTAH PROVINSI DKI JAKARTA: TOPIK EKONOMI DAN KEUANGAN DAERAH," *J. Sist. Inf.*, vol. 12, no. 2, pp. 82–89, Nov. 2016.

[6] C. N. Knaflic, *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley, 2002.