Web-based decision support system to predict risk level of long term rice production

Imam Mukhlash\textsuperscript{1}, Ratna Maulidiyah\textsuperscript{1}, Sutikno\textsuperscript{2} and Budi Setyono\textsuperscript{1}

\textsuperscript{1}Department of Mathematics, Institut Teknologi Sepuluh Nopember Surabaya, Indonesia
\textsuperscript{2}Department of Statistics, Institut Teknologi Sepuluh Nopember Surabaya, Indonesia

E-mail: imamm@matematika.its.ac.id

Abstract. Appropriate decision making in risk management of rice production is very important in agricultural planning, especially for Indonesia which is an agricultural country. Good decision would be obtained if the supporting data required are satisfied and using appropriate methods. This study aims to develop a Decision Support System that can be used to predict the risk level of rice production in some districts which are central of rice production in East Java. Web-based decision support system is constructed so that the information can be easily accessed and understood. Components of the system are data management, model management, and user interface. This research uses regression models of OLS and Copula. OLS model used to predict rainfall while Copula model used to predict harvested area. Experimental results show that the models used are successfully predict the harvested area of rice production in some districts which are central of rice production in East Java at any given time based on the conditions and climate of a region. Furthermore, it can predict the amount of rice production with the level of risk. System generates prediction of production risk level in the long term for some districts that can be used as a decision support for the authorities.

1. Introduction

Agriculture is a very important sector for most developing countries. Indonesia has many advantages in terms of natural resources and has long been known as an agrarian country. Success factors of agriculture is not only influenced by natural resources, but also depends on several other aspects such as government policy and weather conditions. Appropriate decision-making is needed in agricultural planning, especially in risk management of rice production. The reason is to prevent from failure of rice production.

The decision making in agricultural planning can be done manually, but the results obtained will be much more optimal when using computer-based technology. Utilization of technology that can be used for this is to create a system that will be able to combine tools (tools) information systems and models to evaluate the various options. This system is known as Decision Support System (DSS) \cite{1}.

GIS-based DSS for agriculture has been developed in many countries, such as India. In 2004, Rao conducted a study on GIS-based DSS for estimating water demand in irrigation systems. As the result, DSS generates rapid estimates for various divisions of the irrigation in accordance to variation of needs \cite{2}. Research on GIS-based DSS has also conducted in Tanzania for identification of potential points in Rainwater Harvesting (RWH) by Mbilinyi in 2007. The result shows that applications developed from Spatial DSS are effective in identifying potential points for RWH technology \cite{3}.
In Indonesia, Sutikno has conducted research in 2014 about the relationship of climate change to agricultural production. From his research, OLS Regression Model to predict rainfall of a region is obtained. Furthermore, he also obtained Copula model to predict the harvested area of rainfall data [4]. Such models would be more useful if integrated into a web-based spatial decision support software, so the results can be represented in the form of easily accessible and understandable information.

Problems that will be discussed in this paper are about how to predict the risk level of rice production and how to develop a Spatial DSS to predict the risk level of rice production. From those issues, this research aims to know risk level of rice production in some districts in East Java, Indonesia at a certain time and produce a web-based software that will be able to do the processes for prediction. The data used in this research is secondary data which is obtained from the results of previous research by Sutikno. The parameters used in the Copula model are limited to the conditions and climate change of the region, and the system created is limited to providing only one decision support by showing districts with high risk in rice production at any given time.

2. Preliminaries
This part will discuss about some basic concepts about Decision Support Systems (DSS) and prediction models that are used in the research.

2.1. Decision Support Systems (DSS)
Turban, 1998 defines DSS as an interactive, flexible and adaptive system developed to support solutions of semi-structured or unstructured problem, in order to improve the quality of decision making. As shown in figure 1, there are four components of the subsystem on the DSS. They are data management, model management, knowledge management and user interface. The main characteristic of DSS is it uses at least one model. The model is a simplified representation or abstraction of simplified reality.

Sugumaran, 2011 defines Spatial Decision Support System (SDSS) as a system that combines between analytical tools with functions that are available in Geographic Information Systems (GIS) and models for evaluating options. SDSS has characteristics such as the ability to manage and analyze data including spatial data, support the resolution of the problem repeatedly (by iteration), the ability to combine models and data, the ability to simplify reporting, the ability to evaluate scenarios and options in decision making, data visualization capabilities as well as results such as drawings, diagrams, charts, maps, etc., provide support mainly on semi-structured or unstructured data, and easy to use and interactive user interface.

![Figure 1. The components of DSS [5].](image)
2.2. OLS and Copula Models
Climate change projection models often use general circulation models (GCM) output data that have produced by many institutions in developed countries, such as CSIRO, HadCM, ECHAM, GFDL, and UKMO. GCM is used in studies on the impact of diversity and climate change, as well as other climate studies. However, for areas with complex topography, along the coastline, and areas with heterogeneous land cover as in Indonesia, GCM model outputs are less sensitive (Wilby, 2004 referred in [4]). Therefore, it is necessary to set coordinate points (nine grids) and cropping GCM data on each grid to obtain the appropriate GCM output data.

Furthermore, data were tested using Kolmogorov-Smirnov. Kolmogorov-Smirnov test is used to test the normality of a data. The analysis result shows that the pattern of relationship between variables in the determination of rice production risk does not satisfy the normal distribution. This indicates that the shape of histogram tends to go down and has a slope (skewness) to the right, so the appropriate method for this case is using Copula [4].

Sutikno, 2014 already produced the constant of OLS regression model to predict rainfall from Principal Components (PC) data, which is shown in table 1. He also obtained Copula model constants to predict the harvested area of rainfall data, that some of the results are shown in table 2.

**Table 1. OLS Regression Model Constants to predict rainfall from Principal Components (PC) Data [4].**

| District    | C   | PC1 | PC2 | PC3 |
|-------------|-----|-----|-----|-----|
| Banyuwangi  | 150 | -0.981 | -5.56 | 0   |
| Bojonegoro  | 192 | -2.78  | -13.5 | 0   |
| Jember      | 220 | -1.41  | -14.6 | -12.3 |
| Lamongan    | -229 | 84.2   | 247  | 0   |
| Ngawi       | 278 | -4.75  | 10.2  | 0   |

**Table 2. Copula Model Constants to predict harvested area of rainfall data [4].**

| District    | Periods | Copula Model Constants |  |
|-------------|---------|------------------------|--|
|             | 1       | (Intercept) CH\text{Jan} CH\text{Feb} CH\text{Mar} CH\text{Apr} |  |
| Banyuwangi  | 10.8112 | 6e-04 -1e-04 -3e-04 2e-04 |  |
|             | 10.2994 | -0.0001 0.0013 -0.0013 -0.0002 |  |
|             | 10.2298 | 0.0000 6e-04 4e-04 3e-04 |  |
|             | 11.0727 | 3e-04 1e-04 -1e-04 -1e-04 |  |
| Bojonegoro  | 2       | (Intercept) CH\text{May} CH\text{Jun} CH\text{Jul} CH\text{Aug} |  |
|             | 9.7273  | 0.0044 0.0040 -0.0011 -0.0013 |  |
|             | 8.6694  | 0.0042 -0.0002 1e-03 1e-04 |  |
|             | 10.5826 | 0e+00 1e-04 -1e-04 2e-04 |  |
| Ngawi       | 2       | (Intercept) CH\text{May} CH\text{Jun} CH\text{Jul} CH\text{Aug} |  |
|             | 10.4984 | 4e-04 -5e-04 0.0008 -0.0006 |  |
|             | 9.3203  | 0.0016 0.0015 -3e-04 0.0003 |  |
3. Result and Discussion

3.1. Design of DSS Component Model

The DSS model consists of some subsystems illustrated in figure 2. These subsystems are:

a. Data management. It consists of internal data and external data. Internal data is data stored in the database, in the form of initial data to be processed. While external data is data outside the database, in the form of model constants obtained with the help of statistical software and spatial data.

b. Model management. The models used are regression models of OLS and Copula. Both are statistical model, so they belong to quantitative model. Each district has different constants in both model.

c. Model analysis and selection. It roles are to connect the data stored in database with the models used, then choose the appropriate model for each district.

d. User interface. It provides a view that enables users to communicate with the system.

3.2. Model process diagram

Process model diagram contains of program architecture as well as the appropriate model selection to be implemented in the data of each district. Analysis of program architecture is done to explain the structure of the program from the data that is ready to be processed until get the result of the risk level of rice production. The design will be presented in the form of the process model diagram in figure 3. This figure shows that the first step is to calculate the PC value by using PCA regression eigen value, both for historical data and scenario. The value of the PC is calculated by using equations (1), (2), and (3).

\[ PC1 = \sum_{i=1}^{9} grid_i \alpha_i \]  
\[ PC2 = \sum_{i=1}^{9} grid_i \beta_i \]  
\[ PC3 = \sum_{i=1}^{9} grid_i \gamma_i \]
To predict rainfall, we use OLS regression model coefficients. Rainfall is a function of PC, expressed in equation (4), with coefficient values using the data in table 1.

\[ CH = f(PC) \]  \hspace{1cm} (4)

After that, we can predict the harvested area by using Copula Model. Harvested area is a function of rainfall, expressed in equation (5), with coefficient values using the data in table 2.

\[ LP = f(CH) \]  \hspace{1cm} (5)

Production per period is calculated by multiplying harvested area with productivity. The formula is shown in equation (6).

\[ production_i = \sum_{i=1}^{3} LP_i \times productivity_i \]  \hspace{1cm} (6)

Then we calculate annual production for each district by summing production per period from period 1 (January-April) to period 3 (September-December), using equation (7).

\[ Total\ Production = \sum_{i=1}^{3} Production_i \]  \hspace{1cm} (7)

From the beginning process (calculate PC) until we obtain total production, the data consists of two types of scenario data, namely SRESA1B and SRESA2. Now we need to calculate the average production of both types of data for each year. Then we calculate the confidence interval. Confidence interval is an interval estimate accompanied by a level of confidence. How to calculate a confidence interval with a 90% chance of confidence is shown in equation (8), where \( s \) is the standard deviation.

\[ Confidence\ Interval = \bar{X} \pm 2s \]  \hspace{1cm} (8)

The standard deviation formula is expressed in equation (9).
Two values of confidence interval (called upper limit and lower limit of the confidence interval) already defined. After that, risk level of rice production is determined by the following rules:
- If the production value is within the confidence interval, then the district has medium risk level.
- If the production value is more than the upper limit of the confidence interval, then the district has low risk level.
- If production is less than the lower limit of the confidence interval, then the district has high risk level.

The final step is to represent the results in the form of a thematic map showing the predicted level of risk of rice production for each case study district.

3.3. Simulation Results
Some of the main results of the simulation process include land area prediction, productivity of rice, average production, and rice production risk level for each case study district. The results of production risk analysis show risk level of one of the districts shown as chart in figure 4. This figure shows prediction of rice production for Ngawi district from 2011 to 2050.

![Figure 4. Graph of rice production risk of one case study district.](image)

Furthermore, risk level also shown in the form of thematic map that depicted in figure 5. Each district area has color depends on the risk level: green for low risk, yellow for medium risk, and red for high risk. We can see form figure 5 that in 2019 Ngawi district has medium risk while other districts have low risk.
4. Conclusion and Future Work
In this paper, we develop a system which has successfully predicted rice production risk from 2011 to 2050 and then grouped it into three risk levels: low risk (safe), medium risk, and high risk (need to be aware). Rainfall prediction data was obtained by using OLS Regression model. Then from the predicted rainfall data, we obtain predicted harvest area by using Copula model. From the predicted data of harvested area, prediction of rice production is obtained. Furthermore, we can predict the amount of rice production. System generates risk level prediction for long term rice production. It can be used as one of the decision supports for the authorities.

In the next research, we will develop this system so that it will be able to deal with more complex issues, which the system will contain several models that can predict long-term and short-term rice production risk.

Acknowledgement
We would like to thank the Ministry of Research, Technology and Higher Education of the Republic of Indonesia for supporting this research.

References
[1] Sugumaran R and Degroote J 2011 Spatial Decision Support Systems Principles and Practices (Boca Raton: CRC Press Taylor and Francis Group, LLC)
[2] Rao N H, Brownee S M and Sarma P B S 2004 GIS-Based Decision Support System For Real Time Water Demand Estimation in Canal Irrigation Systems Current Science Vol. 87 No. 5
[3] Mbilinyi B P, Tumbo S D, Mahoo H F and Mikiramwinyi F O 2007 GIS-Based Decision Support System For Identifying Potential Sites For Rainwater Harvesting Physics and Chemistry of the Earth 32, issues 15-18, pp 1074–1081
[4] Sutikno, Kuswanto H and Ratih I D 2014 Journal of Mathematics and Statistics 10 (2), pp. 192-200
[5] Turban E and Aronson J E 1998 Decisions Support Systems and Intelligent Systems (New Jersey: Prentice-Hall International, Inc)