A decision tree algorithm for predicting amount of batik tulis lasem production by decision support system to support financial feasibility

T Khotimah[1], R Nindyasari[1] and N Ermawati[2]

[1]Department of Informatics Engineering, University of Muria Kudus, Gondang Manis, Kudus, 59327, Indonesia
[2]Department of Accounting, University of Muria Kudus, Gondang Manis, Kudus, 59327, Indonesia

Abstract. Decision Support System for financial feasibility of the *Batik Tulis* Lasem Industry can process input data based-on Assumptions Value, Investment Costs, and Operational Costs into output in the form of recapitulation of Fund Sources, Projection of Profit and Loss and Projections of Cash Flows and value determinants of business feasibility such as: Net Present Value (NPV), Internal Rate of Return (IRR), and PayBack Period (PP). By this application, the researchers did a simulation asumtion value of *Batik Tulis* Lasem amount production. It was used 9 categories of *Batik Tulis* Lasem products, such as: *Batik* 1 Color Cheap, *Batik* 1 Color Medium, *Batik* 1 Color Expensive, *Batik* 2 Colors Cheap, *Batik* 2 Colors Medium, *Batik* 2 Colors Expensive, *Batik* 3 Colors Cheap, *Batik* 3 Colors Medium, and *Batik* 3 Colors Expensive. Simulation results are used as data in this study. The purpose of this study is to classify financial feasibility based-on the number of products produced. The algorithm used is Decision Tree. The data used is 100 records. Rule generated from the Decision Tree algorithm has an accuracy value of 95%. From the resulting the Decision Tree, there are 3 important rules can be taken in determining the amount of production.

1. Introduction

Batik Tulis Lasem Industry is located in Babagan Village, Lasem, regency of Rembang, Central Java. There are many *Batik Tulis* productive houses in Babagan. Productive houses in Babagan are owned by indigenous people and Chinese [1]. One of the well-known *Batik Tulis* industries is the productive house of Sumber Rejeki. It is owned by Sri Winarti, an indigene. She also become the chairman of the Sarwo Endah -*Kelompok Usaha Bersama* (KUB). At the Sarwo Endah gallery, each member can showcase and market their products[2].

To facilitate the *Batik Tulis* Lasem industry in analyzing financial feasibility in the context of submitting capital assistance, the researchers built a Decision Support System for the financial feasibility of the *Batik Tulis* Lasem Industry. This system can process input data based on Assumption Values, Investment Costs, and Operational Costs to be output in the form of recapitulation of Fund Sources, Projected Profit Losses and Projections of Cash Flow and the value of determinants of business feasibility such as: Net Present Value (NPV), Internal Rate of Return (IRR), and Pay Back Period (PP). From the three values produced, it can be determined the feasibility to add capital loan.

The next problem is how much production must be produced to meet the targets that are feasible or not feasible for the *Batik Tulis* production industry. In *Batik Tulis* industries managed by Sri Winarti, a production period is given a minimum of 150 products. *Batik Tulis* product itself is classified based on several things, such as materials, colors, motifs, prices, and etc. How many products must be produced by each product classification? Here, the researchers limit the classification of products into 9 categories: *Batik* 1 Color Cheap, *Batik* 1 Color Medium, *Batik* 1 Color Expensive, *Batik* 2 Colors Cheap, *Batik* 2 Colors Medium, *Batik* 2 Colors Expensive, *Batik* 3 Colors Cheap, *Batik* 3 Colors Medium, and *Batik* 3 Colors Expensive.

In this paper, the researchers present data simulation using 9 product categories, structure analysis decision tree that applied in Rapid Miner tools, and then the design of hierarchical trees and use them for predicting the amount of production. This prediction result will be used as data input on The Financial Feasibility by Decision Support System to determine financial feasibility status and real amount of production for each category in the *Batik Tulis* gallery Sarwo Endah as one of the feature for decision tree models. It increases the accuracy of amount prediction.

2. Literature Review

There are many algorithm for the classification technique had been used in the fields of data mining. These algorithm are decision tree, rule-based method, memory-based learning, Bayesian Networks, Neural Network and Support Vector Machines. A comparative study has been implemented using three classification algorithms are Support Vector Machine (SVM), Fuzzy-SVM, and decision tree [3]. In the other hand the researchers also shows a comparative study between decision tress and KNN for classification technique[4].

Based on the both of comparative study previously, decision tree algorithm has variation in produced results, because the results depend on method to choose best attribute and criteria used to grow the tree. Decision Tree used to predict a class for a given attribute and KNN algorithm only used to find similar value. Based on the research by A. Trabelsi et al., Decision tree algorithm can be used to solving classification problems with attribute values and class label that applied on data imperfection [5].

In other research, this algorithm can applied to multi criteria inventory classification (MCIC) based on supervised classifiers [6]. The training data using a sample items from simulating inventory control system. Many other problems using Decision tree to make a model classification to evaluate the productive potential of biometric during harvest the sugarcane crop [7]. The classification technique aims to a prediction that accurate for oil production in order to operations effective cost [8]. A. Tike and S. Tavarageri develop research to predict prices into a medical price prediction system [9]. N. Nwulu compares decision tree with machine learning algorithms such as Random Forest, linier regression, and Gradient-boosted trees to prediction the crude oil price and the result experiment achieves high accuracy[10]. In The agriculture field, M. Shakoor et al. using supervised machine learning to prediction agricultural production on farming in Bangladesh [11].

Next research, S. Rajeswari and K. Suthendran develop classification model for agricultural data of the soil fertility level. This model is activated on Design of Smart Information System (DSIS) using Global Positioning System(GPS) to identify the user location [12]. In the healthcare domain also using decision tree for prediction of heart disease and produce prediction with high accuracy and sensitivity[13]. In drilling industry decision tree algorithm used to predicting lost circulation that applied to develop smart system for estimating amount of lost circulation, and this algorithm most accurate method relative to other method are ANN, ANFIS and MLP-NN [14]. From the various research about implementation decision tree algorithm previously, so in this research will used Decision tree algorithm. If we look the data will be used is more appropriate with decision tree. Next to measure prediction accuracy in this research, now available open source software is rapid miner [15]. The software can show performance of the classification models with accuracy measure.

3. Methodology

The Method of the research can be seen in Figure 1.

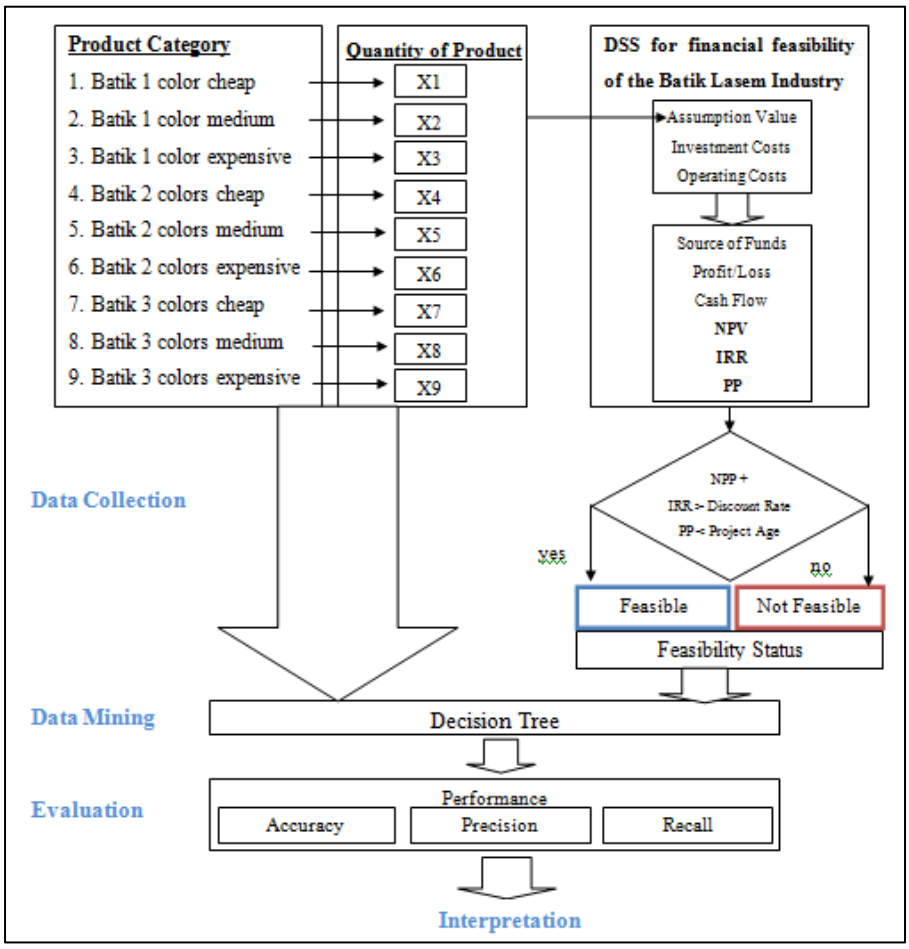

Figure1. The Research Method of Decision Tree

The research steps are as follow:

1. Data Collection

   The Data is collected by conducting simulations using the Decision Support System application for the financial feasibility of the *Batik Tulis* Lasem Industry. The variables used consist of 2 types of variables: class variable and independens variables. Feasibility Status is class variable and other variables are:

   a. quantity of product from *Batik* 1 color cheap
   b. quantity of product from *Batik* 1 color medium
   c. quantity of product from *Batik* 1 color expensive
   d. quantity of product from *Batik* 2 colors cheap
   e. quantity of product from *Batik* 2 colors medium
   f. quantity of product from *Batik* 2 colors expensive
   g. quantity of product from *Batik* 3 colors cheap
   h. quantity of product from *Batik* 3 colors medium
   i. quantity of product from *Batik* 3 colors expensive

2. Data Mining

   In processing data mining, the algorithm used is Decision Tree. Tree formation is done by using the index gini criteria.

3. Evaluation
   Evaluation is done to determine the value of accuracy, precision, and recall of the algorithm used.
4. Interpretation
   Interpretation is done by analyzing the tree formed from the data mining process by Decision Tree. The analysis is used to take the decision.

The determination of class variable of Feasibility Status was done by some tools [16]:

a. Net Present Value (NPV)
   If NPV positve, the Feasibility Status is Feasible. The other, the Feasibility Status is Not Feasible.
b. Internal Rate of Return (IRR)
   If IRR > Discount Rate, the Feasibility Status is Feasible. The other, the Feasibility Status is Not Feasible.
c. Payback Period (PP)
   If PP < Project Age, the Feasibility Status is Feasible. The other, the Feasibility Status is Not Feasible.

4. Results and discussion

The data used in this study were 100 records. Data obtained from simulations of the Assumption Value of the production of *Batik Tulis* Lasem. The simulation application uses the Decision Support System for the financial feasibility of the *Batik Tulis* Lasem Industry. Simulation is done by entering the number of *Batik Tulis* Lasem productions which consists of 9 categories, such as: *Batik* 1 Color Cheap, *Batik* 1 Color Medium, *Batik* 1 Color Expensive, *Batik* 2 Colors Cheap, *Batik* 2 Colors Medium, *Batik* 2 Colors Expensive, *Batik* 3 Colors Cheap, *Batik* 3 Colors Medium, and *Batik* 3 Colors Expensive. The results of value NPV, IRR, and PP can be seen in Simulation Data in **Table 1**

Table 1. Simulation Data

Batik 1 color			*Batik* 2 colors			*Batik* 3 colors			Financial Feasibility			Feasibility Status
Cheap	Medium	Expensive	Cheap	Medium	Expensive	Cheap	Medium	Expensive	NPV	IRR	PP	
20	20	10	20	20	10	20	20	10	1.110.840.137	150,48	0,61	Feasible
20	15	15	20	15	15	20	15	15	1.357.807.604	151,11	0,515	Feasible
.												
.												
.												
90	50	10	0	0	0	0	0	0	-70.389.345	135,56	3,936	NotFeasible

Data of 100 records consisted of 57 records with Feasibility Status are Feasible and 43 records with Feasibility Status are Not Feasible. From this, the Data Mining is processed using the Decision Tree algorithm. The criteria is index Gini. Data mining processing uses Rapid Miner tools. Figure 2 shows the structure of the analysis Decision Tree in Rapid Miner. Figure 3 shows the Tree of Decision Tree Result.

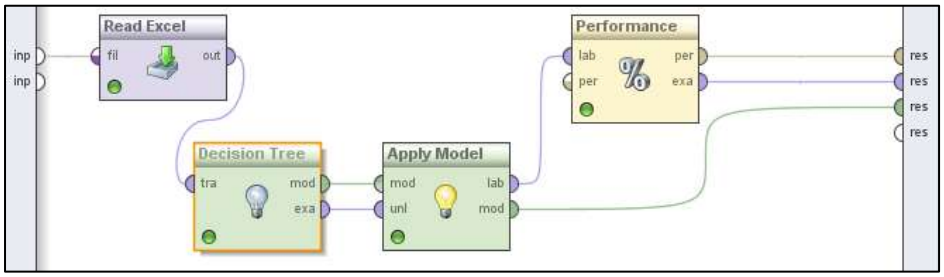

Figure 2. Analysis Structure of Decision Tree

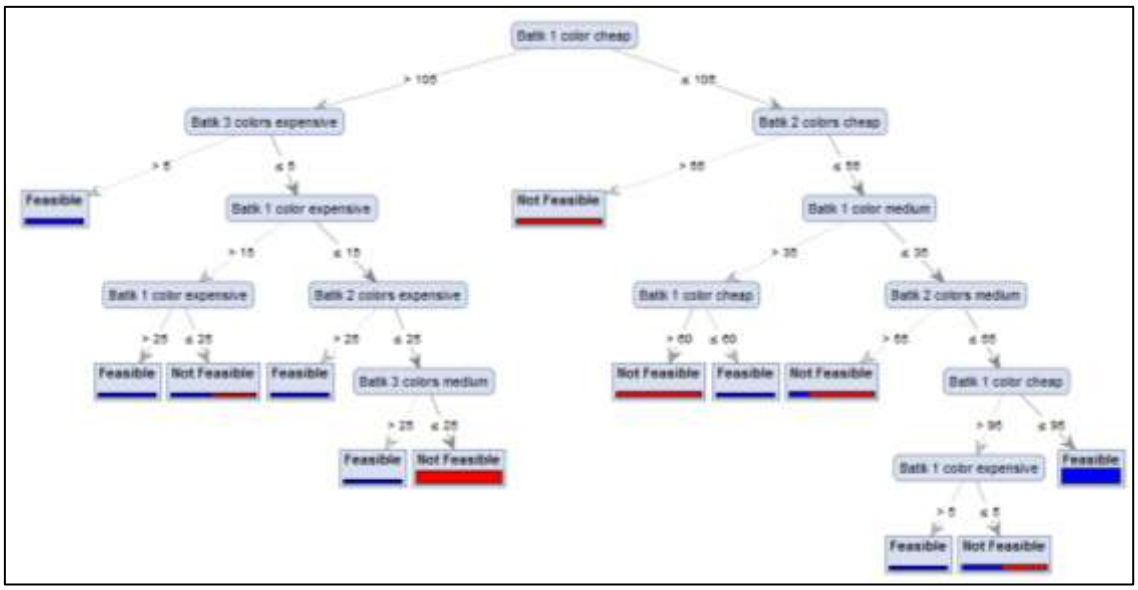

Figure 3. The Result of Decision Tree

accuracy: 95.00%

	true Feasible	true Not Feasible	class precision
pred. Feasible	52	0	100.00%
pred. Not Feasible	5	43	89.58%
class recall	91.23%	100.00%	

Figure 4. Evaluation Result of Performance Value

Figure 4 shows the results of the evaluation of the performance values of the Decision Tree algorithm. Accuracy value is 95%, class precision for Feasible prediction is 100% and class precision for Not Feasible prediction is 89.58%. The class recall value for true Feasible is 91.23% and the class recall value for true Not Feasible is 100%.

5. Conclusion

The research is using the Decision Tree algorithm. For the conclusion of the research, there are 3 rules produced to determine the number of products that must be produced by *Batik Tulis* Lasem industry players. To obtain Feasibility Status equal to Feasible, in a production period there are 150 products, at least must comply with the following rules:

1. If *Batik* 1 Color Cheap are produced with a minimum number of 106 products, it also must be produced *Batik* 1 Color Expensive and or *Batik* 2 Colors Expensive and or *Batik* 3 Colors Medium with a minimum number of 26 products.

2. If *Batik* 1 Color Cheap are produced with range 96-105 products, it also must be produced *Batik* 1 Color Expensive with a minimum number of 6 products.
3. If *Batik* 1 Color Cheap are produced with a maximum number of 60 products, it also must be produced *Batik* 1 Color Medium with a minimum number of 36 products.

Acknowledgments
This paper is part of research funded by Directorate General Higher Education, Ministry of Research, Technology and Higher Education of the Republic of Indonesia, year 2019.

References
[1] E R Kridarso 2018 *IOP Conf. Ser.:Earth Environ. Sci*. **106** 102020
[2] Khotimah T and Nindyasari R 2017 *J. Mantik Penusa* **1** 71–5
[3] P Pandey and A Jain 2016 *3rd Int. Conf. Comput. Sustainable Global Dev. (INDIACom)*(New Delhi, India: IEEE) pp 3620-24
[4] M Mohanapriya and J Lekha Mrs 2018 *J. Phys.: Conf. Ser*.**1142** 012011
[5] Trabelsi A, ElouediZ and Lefevre E 2019 *J. Fuzzy Sets Syst.* **366** 46-62
[6] Lolli F, Ishizaka A, Gamberini R, Balugani E and Rimini B 2017 *J. Procedia Manuf*.**11** 1871–81
[7] Rossi Neto J, de Souza ZM, de Medeiros Oliveira SR, Kölln OT, Ferreira DA, Carvalho JLN, et al 2017 Sugar Tech **19** 662-8
[8] Li X, Chan CW, Nguyen HH 2013 *J. Pet. Sci. Eng* **104** 11–6
[9] Tike A and Tavarageri S 2017 *2017IEEE Int. Conf. on Big Data (Big Data)* (Boston, MA, USA: IEEE)
[10] Nwulu N I 2017 *2017Int. Artif. Intell. Data Process. Symp.(IDAP)*(Malatya, Turkey: IEEE)
[11] Shakoor M T, Rahman K, Rayta S N and Chakrabarty A 2017 *1st Int. Conf. on Next Gener. Comput. Appl.* (Mauritius: IEEE)
[12] Rajeswari S and Suthendran K. 2019 *J. Comput. Electron. Agric.* **156** 530-9
[13] Mathan K, Kumar P M, Panchatcharam P, Manogaran G and Varadharajan R 2018 *J. Des. Autom. Embedded Syst*. **22** 225-42
[14] Sabah M, Talebkeikhah M, Agin F, Talebkeikhah F and Hasheminasab E 2019 *J. Pet. Sci. Eng*. **177** 236-49
[15] Naik A and Samant L 2016 *J. Procedia Comput. Sci.***85** 662-8
[16] Nur Agung Mulyana Halim *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.***508** 012021