Classification of the Mainstay Economic Region Using Decision Tree Method

Heru Ismanto[1], Azhari Azhari[2], Suharto Suharto[3], Lincolin Arsyad[4]
[1,2,3]Department of Computer Science and Electronics, Faculty of Mathematics and Natural Science,
Universitas Gadjah Mada, Yogyakarta, Indonesia
[1]Department of Informatics Engineering, Faculty of Engineering, Musamus University, Merauke, Indonesia
[4]Department of Economics, Faculty of Economics and Business, Universitas Gadjah Mada, Yogyakarta, Indonesia

Article Info

ABSTRACT

The development of the region cannot be separated from the concept of economic growth and the determination of the mainstay region as a regional center that is expected to have a positive impact on economic growth to the surrounding regions. In fact, the determination of the mainstay region is a difficult thing to do. Some cases of the determination of the mainstay region are mostly on the basis of the prerogative rights of the policy makers without carefully seeing the achievements of the development of a region. The objective of this study is to develop a classification model of the mainstay economic region using computational techniques. The decision tree methods of NBTree and J48 are used in this study and combined with Klassen typology. The results of this study show that J48 algorithm has better accuracy than NBTree in the formation process of decision tree. The accuracy of J48 is higher than NBTree i.e. 68.96%. The comparative result of the classification of the mainstay economic region between Klassen and J48 shows that there is a shift in the class position of the development quadrant. In Klassen classification, there are three regions that are categorized into the mainstay regions with advanced development and rapid growth (K1). Meanwhile, J48 results show that there is no region categorized into K1. However, the mainstay economic region on J48 is based on the level of development with the level below K1, i.e. K2. J48 classification results show that there are ten regencies that are categorized into the mainstay economic regions, namely Biak, Regency of Jayapura, Jayawijaya, Kerom, Merauke, Mimika, Nabire, Ndunga, Yapen, and the Municipality of Jayapura.

Corresponding Author:

Heru Ismanto,
Department of Computer Science and Electronics,
Faculty of Mathematics and Natural Science,
Universitas Gadjah Mada, Yogyakarta, Indonesia.
Email: heru@unmus.ac.id

1.   INTRODUCTION

The mainstay economic region is an region used as a barometer of the economic growth of a region so that it becomes the economic support for other regions. The determination of the mainstay economic region is usually conducted by looking at the achievements of the relevant regional development based on the data of gross regional domestic sector (GRDP). There are some approaches used to determine the mainstay economic region; one of them is Klassen typology. Klassen typology classifies the regions into four development quadrants. Quadrant I is a developed and rapid growth regions; Quadrant II is an advanced but depressed region, Quadrant III is a potential or developing region; and Quadrant IV is a relatively lagging region [1]. By seeing an region categorized into a particular development quadrant, the region which is the

mainstay economic region can be identified. The regions categorized into Quadrant I are usually used as the main mainstay economic region by the local government, the quadrant-II region becomes the second-level mainstay economic region, whereas the regions categorized into Quadrant III and IV are not categorized as a mainstay economic region. It means that those regions should be prioritized in further development activities.

Theoretically, Klassen is able to identify the mainstay economic region based on the results of the Gross Regional Domestic Product (GRDP) sector data clustering by looking at the development quadrant formed. However, the stages of clustering are very rigid and do not pay attention to the characteristics of the data and the distance between the data of its GRDP [2]. In addition, the clustering of the mainstay region with Klassen always selects the overall attributes of GRDP sector data owned by a region as a whole. Therefore, it takes much time to classify the mainstay economic region. This study was conducted in an attempt to provide an alternative approach to classify the mainstay economic region using decision tree computation technique. Decision Tree is formed from a set of data that form a smaller subset and interconnect between one to another attribute which form decision tree structure. In the process of forming a decision tree, it need the calculation of gain value to divide data with the same or similar instances into smaller subsets. Afterwards, the gain value calculation result is used to calculate the entropy value. This entropy value is used to determine which primary attributes are selected as the determinant of data classification, followed by other attributes that are arranged according to their entropy values. In clustering the mainstay economic region using Klassen, all attributes are seen as the same. Meanwhile, when it uses the decision tree, there is a selection of classified determinant attributes that are sorted by entropy value calculations repeatedly, so it can not be done using Klassen.

2. DECISION TREE

Decision tree is one of the data classification techniques that makes decision tree structure more easily understood [3]. Each internal node represents testing of an attribute, each branch represents output of the testing, and the leaf node represents classes or class distributions [4]-[5]. The topmost node is called root node. The root node will have some exiting edges, but it does not have an incoming edge. The internal node will have one incoming edge and some exiting edges. Meanwhile, the leaf node will only have one incoming edge and no exiting edge.

The decision tree is used to classify an unknown class sample into existing classes. The data test path will firstly go through the root node and finally go through the leaf node that will infer the class predictions of the data. The data attribute must be a categorical data; if it is continuous, the attribute must be discretized first [4].

This technique is widely used for classification of student exam passing grade [6, 7] identification of the risk of trauma in childbirth through patient data classification [8] as well as the classification of regional development level [9]. The followings are the explanation of NBTree and J48 techniques used in this study.

2.1. NBTree Algorithm

NBTree uses the frequency of a class appearing in the formation of a decision tree from a set of data. A study [10] states that the NBTree algorithm uses the Naive Bayes method to determine a leave tree while generating a decision tree. Below are the NBTree algorithms:
   a) Determine the initial conditions.
   b) Classify the data and calculate the value of spited node.
   c) Trim the tree that has been formed to evaluate the optimal tree and cross-validation error.
   d) Try it out using the test data of the tree and identify the terminal node based on the test data.
   e) Predict one step ahead using Naïve Bayes at the generated terminal node.

By assigning a set of instances to a node, the NBTree algorithm will evaluate the utility of split for each attribute. If the largest utility of all attributes is higher than the utility of the current node, the division of existing instances will be based on those attributes [11].

The utility of node is calculated by discretizing the existing data and calculating the estimation of 5-fold cross validation accuracy of the naïve-bayes usage at the node. Meanwhile, the utility of split is the weighted amount of the utility of node, where the weights assigned to a node are proportional to the number of instances derived by that node. The division is set up significantly if the relative reduction to error is better than 5% and there are at least 30 instances in the node. This is to avoid any division by small values [11].

An NBTree classifier specifies the class label of an instance by sorting it into a leaf and applying Naïve-Bayes in the leaf. The NBTree often achieves a higher degree of accuracy when compared to Naïve Bayesian classifier [12].

2.2. J48 Algorithm

The J48 algorithm is the result of the development of ID3 technique [13] and the determination of the decision tree root is conducted by looking at the gain and the ratio of the gain of an attribute. Below are the J48 algorithms:

a) Select an attribute as a root
b) Create a branch for each value
c) Divide the cases to the branches
d) Repeat the process for each branch so that all the cases on the branch have the same class

The J48 algorithm ignores the missing value, i.e. a value for a predictable item based on what is known about the attribute values in the other row. The basic idea of this algorithm is to divide data into range based on the attribute values for items found in training data sets. The J48 algorithm allows classification either through decision trees or rules generated from the formation of classifier [14].

3.   MAINSTAY REGION

Mainstay region is an region with greater economic growth potential compared to other regions [15]. This economic growth is usually determined by three important factors, namely: capital accumulation, population growth, and technological advancement owned by a region [16]. The existence of mainstay region is expected to have a positive impact on the economic growth for other regions surrounding. So far, the determination of mainstay region is usually conducted by the government through the decisions set forth in the National Spatial Planning Law [15]. However, it can also be determined based on the classification of development regions using Klassen approach [1].

4.   KLASSEN TYPOLOGY

Klassen typology is an approach used to look at the pattern of the economic development growth of a region [17]. Klassen divides the regions into four development quadrants as shown in Table 1.

Table 1. Classification of Economic Growth by Klassen Typology

Quadrant I (K1) developed and fast-growing regions	Quadrant II (K2) developed but depressed regions
Quadrant III (K3) Potential or developing regions	Quadrant IV (K4) relatively lagging regions

Advanced and rapidly growing sector (developed sector) is in Quadrant I. This quadrant is a quadrant of a specific sector growth rate in GRDP (si) which is greater than the sector growth rate in the regional GRDP as the reference (s) and has a sector contribution value to GRDP (ski) which is greater than the sector contribution to regional GRDP as the reference (sk). This classification is denoted with si> s and skis> sk.

Advanced but stagnant sector is in Quadrant II. This quadrant is a quadrant of a specific sector growth rate in GRDP that is smaller than the sector growth rate in the regional GRDP as the reference (s) but has greater sector contribution value to GRDP (ski) than the sector contribution to regional PDRB as the reference (sk). This classification is denoted with si <s and skis> sk.

Potential and developing sector is in (Quadrant III. This quadrant is a quadrant of a specific sector growth rate in GRDP (si) which is greater than the sector growth rate in the regional GRDP as the reference (s) but has a smaller sector contribution value to GRDP than the sector contribution to regional GRDP as the reference (sk). This classification is denoted with si> s and skis <sk.

Underdeveloped sector is in Quadrant IV. This quadrant is a quadrant of a specific sector growth rate in GRDP (si) which is smaller than the sector growth rate in the regional PDRB as the reference (s) and also has smaller sector contribution value to GRDP (ski) than the sector contribution to regional GRDP as the reference (sk). This classification is denoted with si <s and ski <sk.

5.   RESEARCH METHODOLOGY

The sample of this study is Papua Province in the eastern part of Indonesia. The study begins with the data collection of the provincial GRDP sector data. Sector data used are 2014 and 2015 data for 29 regencies in Papua Province. The next step is to divide the data into data training and testing. The data of

2014 are used as data training while data of 2015 are used as data testing. Furthermore, both types of data are classified based on Klassen typology to obtain the initial classification of the mainstay economic region. The next step is to establish the basic rules using decision tree techniques to obtain decision tree as a classification tool for the next mainstay economic region. Two decision tree techniques used in this study are NBTree and J48. Decision tree formed is tested to the data testing as well as to see the accuracy of the classification of the mainstay regions using the decision tree model. The decision tree technique with the highest level of accuracy is used as the foundation of the main rule in this study.

6. PROPOSED MODEL OF THE MAINSTAY ECONOMIC REGION CLASSIFICATION

This study develops a model of the mainstay economic region classification based on GRDP sector data owned by a region. Figure 1 shows the developed model. The developed model is a combination of decision tree and Klassen typology techniques as the basis for determining the classification of the mainstay economic region. GRDP sector data of a region in the period of previous *n* years are used to form decision tree using decision tree. GRDP sector data are then classified using Klassen typology to obtain the initial classification results. The result of this initial classification is used as data training for decision tree makers using decision tree. The next stage is to test the data testing to test the decision tree already formed. The main output of the developed model is the classification i.e of the mainstay economic region based on the value of the GRDP sector data owned by a region.
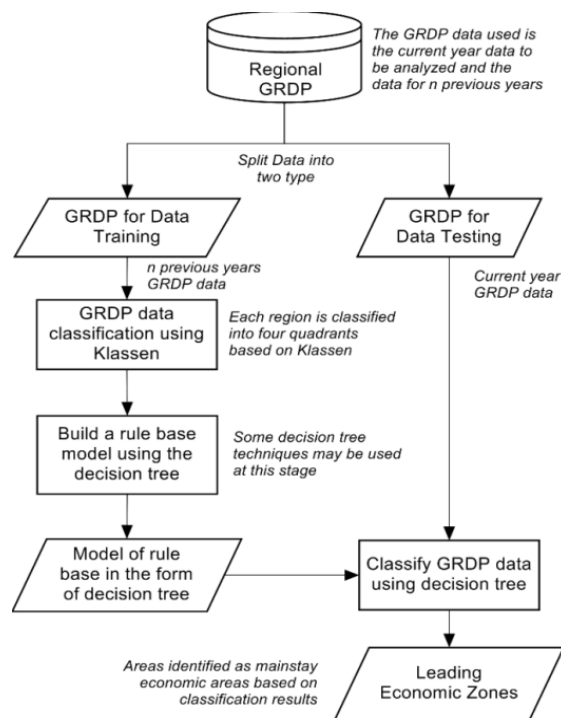
Figure 1. Proposed Model of Mainstay Economic Region Classification

7. RESULTS AND DISCUSSION

The initial phase of this study was conducted by classifying 29 regencies in Papua Province using Klassen. The main objective is to make it as data training on the formation of classification rules using decision tree techniques of NBTree and J48. Table 2 shows the results of the classification of 29 regencies in Papua Province based on Klassen.

The result of classification at early stage of this study shows that there are three (10.34%) regencies are categorized into the first quadrant (K1) i.e. the region with advanced development level and rapid growth. The regions categorized into K1 likely serve as the mainstay economic region are Jayapura and Paniai Regencies and Jayapura Municipality. It can be seen from the classification results showing that three regencies are categorized into regions with advanced development and rapid growth (K1). 24.13% of regencies in Papua Province are classified into advanced but depressed regions, 31.03% of regencies are

categorized into potential and developing regions, while the remaining 34.48% are regions with relatively underdeveloped development status. In the next stage, the results of this classification are then used as training data for the formation of decision tree using decision tree. As aforementioned before, two decision tree techniques used in this study are the NBTree and J48 algorithms. In this study, Weka tool is used for decision tree formation process.

Table 2. Classification of Development Quadrant in 29 Regencies in Papua Province

No	District	GRDP 2014	GRDP 2015	Quadrant
1	Asmat	788.328,61	831.082,49	K4
2	Biak	2.158.964,49	2.254.816,92	K2
3	Boven Digul	2.346.150,96	2.468.482,74	K2
4	Deiyai	425.336,88	471.671,60	K3
5	Dogiyai	366.619,98	392.533,37	K3
6	Intan Jaya	412.149,98	452.116,77	K3
7	Kab Jayapura	5.038.190,97	5.557.746,95	K1
8	Jayawijaya	2.416.172,11	2.578.258,76	K2
9	Kerom	1.224.239,70	1.308.614,70	K4
10	Lanyjaya	547.523,90	580.163,36	K4
11	Memberamo Raya	440.824,53	476.822,52	K3
12	Memberamo Tengah	344.236,30	366.598,58	K4
13	Mappi	953.121,31	1.018.560,21	K4
14	Merauke	5.252.312,30	5.586.617,68	K2
15	Mimika	51.013.497,45	54.326.848,32	K2
16	Nabire	4.143.384,63	4.421.359,00	K2
17	Ndunga	372.137,89	407.087,35	K3
18	Paniai	1.852.212,27	2.033.474,78	K1
19	Pegunungan Bintang	700.783,09	723.898,81	K4
20	Puncak Jaya	554.683,92	595.277,12	K3
21	Puncak	381.722,86	412.594,93	K3
22	Sarmi	991.923,83	1.057.063,76	K4
23	Supriori	404.556,82	417.100,97	K4
24	Tolikara	504.607,85	529.156,59	K4
25	Waropen	244,60	328,30	K3
26	Yahokimo	650.159,22	690.497,43	K4
27	Yalimo	347.173,15	378.228,06	K3
28	Yapen	1.615.976,20	1.708.539,10	K2
29	Kota Jayapura	9.434.791,40	10.251.863,96	K1

In algorithm testing, GRDP sector data both in 2014 and 2015 are used as class determinant of classification result. There are 18 attributes used, namely: agriculture, livestock, forestry, fishery (2014_S1 and 2015_S1); mining and extraction sectors (2014_S2 and 2015_S2); manufacturing industry sectors (2014_S3 and 2015_S3); electricity, gas and water sectors (2014_S4 and 2015_S4); construction sector (2014_S5 and 2015_S5) ; trade, hotels and restaurants sectors (2014_S6 and 2015_S6); transportation and communication sectors (2014_S7 and 2015_S7); finance, real estate and corporate services (2014_S8 and 2015_S8); and service sectors (2014_S9 and 2015_S9).

Test results of both algorithms show that J48 algorithm has a better accuracy than NBTree, shown in Table 3. From 29 data instances tested, 19 data are categorized into incorrectly classified instance, thus the inaccurate decision tree formed uses NBTRee of 65.57%. Meanwhile, for J48 algorithm, the inaccurate decision tree formed is smaller i.e. 31.03%. Table 3 shows the comparison of decision tree formation using NBTree and J48 algorithm seen from the value of classification accuracy, Kappa value, mean absolute error, and root mean square error.

Table 3. Comparison of NBTree and J48 Testing Results

Algorithms	J48	NB-Tree
Classification accuracy (%)	68.96	34.48
Kappa	0.555	0.155
Mean absolute Error	0.229	0.313
Root mean squared error	0.372	0.421

The formation of decision tree using J48 algorithm shows that the classification of mainstay economic region is more influenced by attribute of electricity, gas and water sectors for data of 2014 (2014_S4).

Figures 2a and 2b respectively show the decision tree formed and the rules generated from the decision tree formation process using J48.

<table>
<tr><td>2014_S4 <= 486
| 2014_S4 <= 67.049972: K3 (10.0/2.0)
| 2014_S4 > 67.049972: K4 (11.0/3.0)
2014_S4 > 486: K2 (8.0/2.0)</td></tr>
</table>

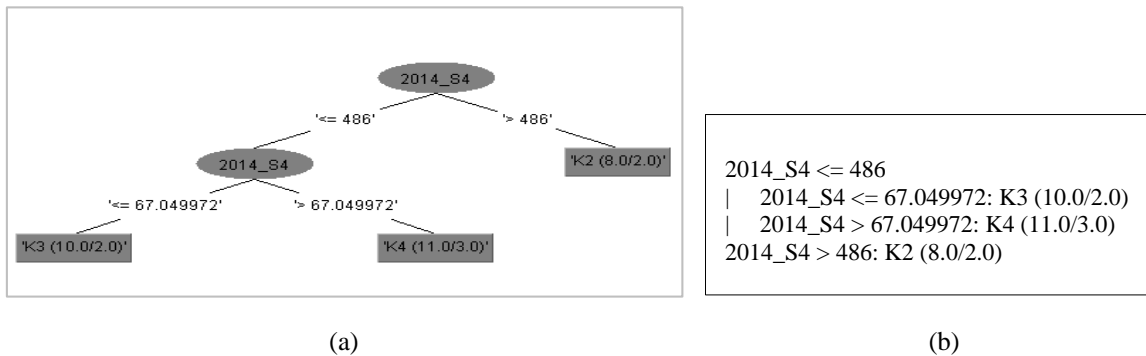(a)   (b)

Figure 2. (a) Decision Tree of J48 Results (b) Decision tree formation process using J48

The next stage is testing the data testing (2015 GRDP sector data) into the decision tree that is formed. Afterwards, the result of 2015 GRDP data classification using decision tree in Figure 2a is compared with the result of classification using Klassen which has been conducted earlier. Among the 29 regencies in Papua Province based on classification using decision tree to the data of 2015, there are three regions which are indicated as mainstay regions. In this case, the result of J48 decision tree has differences especially from the region identified as a mainstay economic region. The result of regional classification using J48 shows that there is no region which is categorized into K1 or region with advanced development level and rapid growth. Most regencies in Papua Province, based on the J48 classification, are categorized fall K2, K3, and K4. The measurement of accuracy using means square error to the Klassen result classification and J48 Decision Tree show that the accuracy level is 65.51%. Table 4 shows the comparison of classification results using Klassen and J48 Decision Tree.

Table 4. Comparison of Klassen and J48 Classifications

No	District	GRDP 2015	Klassen	J48
1	Asmat	831.082,49	K4	K4
2	Biak	2.254.816,92	K2	K2
3	Boven Digul	2.468.482,74	K2	K4
4	Deiyai	471.671,60	K3	K3
5	Dogiyai	392.533,37	K3	K3
6	Intan Jaya	452.116,77	K3	K3
7	Kab Jayapura	5.557.746,95	K1	K2
8	Jayawijaya	2.578.258,76	K2	K2
9	Kerom	1.308.614,70	K4	K2
10	Lanyjaya	580.163,36	K4	K3
11	Memberamo Raya	476.822,52	K3	K3
12	Memberamo Tengah	366.598,58	K4	K3
13	Mappi	1.018.560,21	K4	K4
14	Merauke	5.586.617,68	K2	K2
15	Mimika	54.326.848,32	K2	K2
16	Nabire	4.421.359,00	K2	K2
17	Ndunga	407.087,35	K3	K2
18	Paniai	2.033.474,78	K1	K4
19	Pegunungan Bintang	723.898,81	K4	K4
20	Puncak Jaya	595.277,12	K3	K4
21	Puncak	412.594,93	K3	K3
22	Sarmi	1.057.063,76	K4	K4
23	Supriori	417.100,97	K4	K4
24	Tolikara	529.156,59	K4	K4
25	Waropen	328,30	K3	K4
26	Yahokimo	690.497,43	K4	K4
27	Yalimo	378.228,06	K3	K3
28	Yapen	1.708.539,10	K2	K2
29	Kota Jayapura	10.251.863,96	K1	K2

In Table 4, there are 10 regencies with different classification results between Klassen and J48. The changes of Klassen classification position occurred in 10 regencies, namely: Boven Digul Regency is categorized into K2 in Klassen but K4 in J48; Jayapura Regency from K1 to K2; Kerom Regency from K4 to K2; Lanyjaya Regency from K4 to K3; Memberamo Tengah Regency from K4 to K3; Ndunga Regency from K3 to K2; Paniai Regency from K1 to K4; Puncak Jaya Regency from K3 to K4; Waropen Regency from K3 to K4; and Kota Jayapura from K1 to K2. Therefore, based on the results of J48 classification, there are 10 regencies which are categorized into the mainstay economic region with the level of advanced but depressed development (K2) as seen in Table 5.

Table 5. Regencies with Mainstay Economic Region

No	District	GRDP 2015	Klassen	J48
1	Biak	2.254.816,92	K2	K2
2	Kab Jayapura	5.557.746,95	K1	K2
3	Jayawijaya	2.578.258,76	K2	K2
4	Kerom	1.308.614,70	K4	K2
5	Merauke	5.586.617,68	K2	K2
6	Mimika	54.326.848,32	K2	K2
7	Nabire	4.421.359,00	K2	K2
8	Ndunga	407.087,35	K3	K2
9	Yapen	1.708.539,10	K2	K2
10	Kota Jayapura	10.251.863,96	K1	K2

Table 5 shows ten regencies, namely Biak, Jayapura, Jayawijaya, Kerom, Merauke, Mimika, Nabire, Ndunga, Yapen and Jayapura Municipality.

8. CONCLUSION

Based on the results of the study, decision tree techniques can be used as an alternative approach to determine the mainstay economic region. The results show that both Klassen and J48 decision trees indicated that Jayapura Municipality and Jayapura Regency are still the mainstay economic regions, although based on regional classification results, both are categorized into different class when they are classified with Klassen and J48. In addition, the accuracy level of 2015 GRDP sector data testing to the decision tree J48 shows that the accuracy is 65.51%. The results of Klassen show that there are three regencies that are categorized into the mainstay economic region. Meanwhile, the results of decision tree J48 show that there are 10 regencies that are categorized into the mainstay economic region. Therefore, decision tree technique, especially J48 algorithm, can be used as an alternative in classifying regions into certain mainstay regions. As a result, it can be used as policy making materials for local governments to determine the mainstay economic region.

REFERENCES

[1] Karsinah, Putri P.I., Rahayu K.N, Panjiputri A.F. The Profile of Pekalongan as a Center of Economic Growth at Tangkallangka Strategic Areas. *International Journal of Economics and Financial Issues*, 2016; Vol 6, Special Issue (S6).

[2] Munandar T.A., Azhari, Musdholifah A., Arsyad L. *Hierarchical Regional Disparities and Potential Sector Identification Using Modified Agglomerative Clustering*. Proceeding of The 1st Annual Applied Science and Engineering Conference (AASEC 2016).

[3] Ramani R. G, Shanthi S. *Classifier Prediction Evaluation in Modeling Road Traffic Accident Data,* Conference Proceeding of Computational Intelligence & Computing Research (ICCIC). 2012; Page(s): 1 – 4.

[4] Kamber, Jiawei H, Micheline. Data Mining: Concept and Techniques. CA USA. Academic Press. 2001.

[5] Niswatin R. K, Sanjaya A. Classification Of Category Selection Title Undergraduate Thesis Using K-Nearest Neighbor Method. 2017; *Indonesian Journal of Electrical Engineering and Computer Science.* Vol. 7. No. 3, pp. 846 ~ 854.

[6] Bresfelean V. P., Bresfelean M., Ghisoiu N. *Determining Students' Academic Failure Profile Founded on Data Mining Methods*. Conference Proceeding of 30th International Conference on Information Technology Interfaces. 2008; Page(s): 317 – 322.

[7] Hussain S, Dahan N.A, Ba-Alwi F.M, Ribata N. Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science.* 2018; Vol. 9, No. 2. February. pp. 447~459

[8] Slaughter G, Kurtz Z, desJardins M, Hu, P. F, Mackenzie C, Stansbury FRCA L, Stein D M, *Rediction of Mortality*. Proceeding of IEEE Biomedical Circuits and Systems Conference (BioCAS). 2012; Page(s) 1 – 4.

[9] Munandar T. A, Winarko E. Regional Development Classification Model using Decision Tree Approach. *International Journal of Computer Applications*, 2015; Volume 114 – No. 8, pp. 28 – 33.

[10] Mori, Hiroyuki. *Application of NBTree to Selection of Meteorological Variables in Wind Speed Prediction*. Conference Proceeding of Transmission & Distribution Conference & Exposition: Asia and Pacific. 2009; Page(s): 1 – 4.

[11] Kohavi R, *Scaling Up the Accuracy of Naïve-Bayes Classifiers : a Decision-Tree Hybrid*. AAAI. 1996.

[12] Zhao Y, Zhang Y. Comparison of decision tree methods for finding active objects. *Advances in Space Research*. 2008; 41(12), 1955–1959.

[13] Gayatri N, Nickolas S, Reddy A.V, Chitra R. *Performance Analysis Of Data Mining Algorithms For Software Quality Prediction*. Conference Proceeding of International Conference on Advances in Recent Technologies in Communication and Computing. 2009; Page(s) : 393 – 395.

[14] Goyal A, Mehta R. Performance comparison of Naïve Bayes and J48 classification algorithms. *International Journal of Applied Engineering Research*, 2012; 7(11 SUPPL.), 1389–1393.

[15] Aswandi H, Kuncoro M, Evaluasi Penetapan Kawasan Andalan : Studi Empiris di Kalimantan Selatan 1993-1998. *Jurnal Ekonomi dan Bisnis Indonesia*. 2002; Vol. 17, No. 1, pp. 27 – 45. in Bahasa.

[16] Todaro M. P. Economic Development, Seventh Edition. New York. Addition Wesley Longman. Inc. 2000.

[17] Sutarno, Kuncoro M. Ketimpangan Antar Kecamatan Di Kabupaten Banyumas 1993-2000. *Jurnal Ekonomi Pembangunan*, 2000; 8(2). 97–110. in Bahasa.