RESEARCH ARTICLE OPEN ACCESS

Determination of Student Department Using Method C4.5 Case Study of SMK Bina Mandiri Teluknaga Kabupaten Tangerang

Sahrul, Ade Octaviansyah, Ajie Maulana, Mohammad Syafrullah

Master Program in Computer Science, Information System Technology, Budi Luhur University
Jl. Raya Ciledug, North Petukangan, Kebayoran Lama, South Jakarta 12260.

----------**********----------

Abstract:

At this time parents and students are still confused with the majors they will take to complete education at the vocational school level. Many of them (students) choose a specialization field following the choices taken by the majority of their classmates who are familiar, regardless of the student's academic achievement factors. This has an impact on the incompatibility of areas of interest with the interests and skills of these students. As a result many students who have difficulty when completing education in vocational school. Many cases found that the selection of majors that are not in accordance with the ability, personality, interests and talents can influence students in following the learning in its class. The impact of student achievement decreased and could even occur drop out of school. Data Mining using C4.5 Algorithm method has greater accuracy value than Naïve Bayes method that is equal to 74,53% while Naïve Bayes accuracy value 67,89% with difference value equal to 6,64%, Rule yielded from method C4. 5 can be applied to the making of the student's determination system at SMK Bina Mandiri Teluk Naga. So it is expected to facilitate the management of SMK Bina Mandiri Teluk Naga in taking policy for students who are still confused in the determination of the majors .

Keywords — **Data Mining, Determination Department, Classification, Algorithm C4.5, Decision Tree, Weka.**

----------**********----------

1. INTRODUCTION

At SMK Bina Mandiri Teluk Naga which is located at Kali Jaya Street RT 02/08 Village Kampung Melayu Barat Teluk Naga District has 2 courses: Accounting, Multimedia. In the process of student learning in SMK Bina Mandiri Teluk Naga, a lot of students who do not know will choose what study program will be selected, the majority of them is to follow friends - friends without seeing talent and ability that should be a benchmark in choosing the course majors. The problems are: 1. The determination of the Department is not based on the value of the National Examination and Vocational Test only by choice. 2. Many students move majors in the next semester, due to wrong in choosing majors. To overcome the various things described in the background and problems of the above research, it can be formulated problem formulation that will be discussed in this research is how the implementation of data mining using C4.5 method to assist new students in the determination of the majors to be selected on SMK Bina Mandiri Teluk Naga Kabupaten Tangerang.

In this case the researcher tries to make a system/application that can direct their majors according to their talents and abilities by using Data Mining taken from the graduation score of SMP/MTs, for the last 3 years with attributes consist of: NISN (National Student Identification Number), student name, Gender, National Examination Score (Indonesia, Mathematics, English, Natural Sciences (IPA), Vocational test scores (Academic, Psychotest,

Physical) and Choice of Department (Accounting and Multimedia). The data is processed using C4.5 method in predicting what majors will become the student's choice.

This research was conducted by K. Sumathi, PhD, S.Kannan, PhD and K. Nagarajan (2016) conducted research on Data Mining: Analysis of the student database using Classification Techniques. In this paper, student databases were analyzed using J48 classification algorithms and predicted placement-related information based on academic results. Data mining techniques are also used in analyzing students' academic results in various aspects as well as predicting the reasons. Once the data is retrieved from the relevant source, the classification algorithm can be applied to categorize the data. The selection of attribute features from large data sets plays an important role in the efficiency of the algorithm Liliana Swastina, 2013 conducted research with the title: Application of Algorithm C4.5 for Determination of Student Department.

In this research suggests that Decision Tree Algorithm predicts more accurate than C4.5 in determining the suitability of department and recommendation of student department. And it is concluded that the Decision Tree C4.5 algorithm is accurately applied for the determination of suitability of student majors with the accuracy level of 93.31% and the recommendation accuracy of the department of 82.64%. The implementation of Decission Tree C4.5 is expected to provide solutions for students and can help STMIK Indonesia in determining the appropriate majors that will be pursued by students during the study so that opportunities for success in higher education studies. Fina Nasari (2014) also conducted a research with the title Application of Algorithm C4.5

In Selection of sibject areas of Program System Information Studies at STMIK Potensi Utama Medan and get the results of conclusions based on calculations using C4.5 algorithm obtained dominant factor someone choosing the specialization is based on the value of JK with data matching level up to 80,14% and this research variable still see the data of value and gender, for its development need to be seen also interest and talent from student who will choose specialization so that selection of specialization will be more appropriate.

2. THEORETICAL BASIS

2.1 Data mining

Data mining is the process of analyzing large amounts of data to find a pattern that will be used for decision making. Data mining is a semi-automatic process that uses mathematical statistical techniques, artificial intelligence, and machine learning to extract and identify useful useful and useful knowledge information stored in large databases (Turban et al, 2005). According to Gartner Group data mining is a process of finding meaningful relationships, patterns and intelligence by checking in a large set of data stored in storage using pattern recognition techniques such as statsitik and math techniques (Larose, 2006). Knowledge Discovery In Database (KDD) process can be explained as follows:

1. Data Selection. Selection of data from a set of operational data needs to be done before the stage of extracting information in KDD begins. Selection results that will be used for data mining processes are stored in a file, separate from the operational database. 2. Pre-processing/Cleaning. Before the process of data mining can be implemented, cleaning process should be done on the data that became the focus of KDD. The cleaning process includes removing data duplication, checking inconsistent data, and correcting data errors, such as typographical errors.
2. Transformation Coding is a transformation on the data that has been selected, so the data is appropriate for data mining process. The coding process in KDD is a creative process and depends on the type or pattern of information to be searched in the database.
3. Data mining is the process of finding patterns or interesting information in selected data by using a particular technique or method. Techniques, methods, or algorithms in data mining very widely. The choice of the appropriate method or algorithm depends heavily on the purpose and process of KDD as a whole.
4. Interpretation/Evaluation. The pattern of information generated from the data mining process needs to be displayed in a form that is easily understood by interested parties. This stage is part of the KDD process called interpretation. This stage includes examining whether the pattern or information found is against the previous fact or hypothesis.

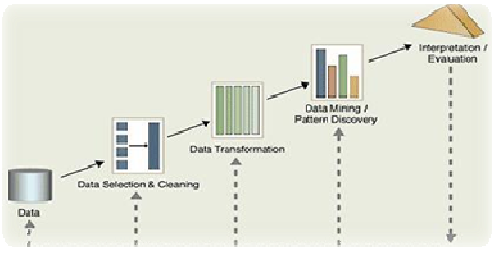

Figure 1. Flow of Information in data mining

2.2 Algorithm C4.5

The C4.5 algorithm is one of a series of algorithms for clarification problems in machine learning and data mining (Wu, 2009). The C4.5 algorithm was designed by J. Ross Quinlan, a research engineer, creating a decision tree algorithm known as ID3 (Iterative Dichotomiser) is called as such because the C4.5 algorithm is descended from the popular ID3 approach within the decision tree decision tree). Decision tree is a set of systematically arranged questions, where each question will continue on another question and so on until it stops at the leaf label which means the class of the variable. This set of questions is illustrated in the form of a very easy-to-understand tree diagram. In the tree diagram, the root of the tree is described as the first question and any branching that appears will be called a tree branch consisting of testing against the value in the attributes tested. The existing branches will branch out until the last branch is called a leaf. Leaf is a type of data label that is being tested, can be said as a result of classification or data prediction results (Wu, 2009). Several stages in making a decision tree with C4.5 algorithm (Kusrini and Lutfi, 2009) are:

1. Preparing training data. Data training is usually taken from historical data that have occurred before and have been grouped into certain classes.
2. Determining the roots of the tree. The root will be taken from the selected attribute, by calculating the gain value of each attribute, the highest gain value will be the first root. Before calculating the gain of the attribute, first calculate the entropy value:

$$\text{Entropy (S)} = \sum_{i=1}^{n} -pi * \log 2pi$$

Information :
S: The set of cases
A: Attribute
N: Number of partitions
S Pi: The proportion of stage Si to S

3. Then calculate the Gain value with the Gain information method:

$$\text{Gain (S, A)} = \text{Entropy(S)} = \sum_{i=1}^{n} * \frac{[Si]}{[S]} * \text{Entropy(S)}$$

4. Repeat step 2 until all tuples are partitioned.
5. The decision tree partition process will stop when:
   a. All tuples in node N get the same class.
   b. No attributes in the tuple are partitioned again
   c. There are no tuples in the empty branch

3. SYSTEM DESIGN AND APPLICATIONS

In this research data obtained from the management of SMK Bina MandiriTeluk Naga that is new student data for the year of registration 2012, 2013, and 2014 which has passed the amount of 271 students consisting of 192 programs majoring in Accounting and 79 programs Multimedia majors.

Then the attribute selection is done to select which attributes are needed from the dataset used in the process of analyzing the determination of the student majors so that less data is taken. The attributes consist of: Student Number (NIS), Name, Sex, National Examination (Indonesia Language,English Language, Mathematics, Science), Academic, Psychotesty, Physical Score and Choice of Department (Accounting and Multimedia).

Table 1: Dataset Samples

No	NISN	Nama	L/P	Nilai Ujian Nasional				Nilai Tes Jurusan			Pilihan Jurusan	
				B.Ind.	B.Ing.	Matematika	IPA	Akademik	Psikotes	Fisik	Akuntansi	Multimedia
1	9999614611	ADINI	L	76	38	58	60	65	75	64	1	
2	9994990063	ADI TIA	L	58	58	38	95	78	52	63	1	
3	9982685084	AGUNG SARDJU PAMUNGKAS	L	66	36	33	48	65	52	72	1	
4	0002160253	AYUNDA NURUL KHOLIFAH	P	76	38	58	60	45	42	54	1	
5	0002146286	DEA KHOLILAH	P	58	76	73	88	80	55	76	1	
6	9994198317	ERIKA	P	44	44	38	48	54	65	72	1	
7	0003792473	ERWIN MAHARAJA	L	64	68	33	85	78	53	48	1	
8	0002125639	FITRIATUL AMALIYATI	P	82	68	58	53	65	52	72	1	
9	0002202350	HOTIFAH LATYUBA	P	50	46	38	33	45	42	54	1	
10	0002147969	IKA NURLIANA	P	82	68	58	53	80	55	76	1	
11	9992587608	ISWANTO	L	84	50	65	73	54	65	72	1	
12	9996433334	LINDA THEODORA	P	50	42	38	38	78	53	48	1	
13	9981842824	MARIO	L	38	38	33	28	65	75	64	1	
14	9992600464	MUIN	L	78	40	30	48	78	52	63	1	
15	9992068872	NADHA NATALIA	P	76	66	62	72	65	52	72	1	
16	9992622539	NATALIA	P	76	38	58	60	45	42	54	1	
17	0002125634	NOVIANA	P	76	82	68	65	80	55	76	1	
18	0003985840	NUR RATNASARI	P	68	74	45	68	54	65	72	1	
19	9992567582	SERLI OKTAPIANI	P	40	40	30	33	78	53	48	1	
20	9992566451	SINTIA DEWI	P	76	38	58	60	65	75	64	1	

The data class used for data mining is prepared (preprocessing) so that it has a bionominal or polynominal class according to a rule that has been created based on its data value. But before the data in the form of value (number) is classified with the form of Alphabet.

Table 2: Value Classification

Score	Classification
90-100	A
80-90	B
70-80	C
60-70	D
<59	E

Table 3: Pieces of Preprocessing Results (20 records)

No	NISN	Nama	L/P	Nilai UN Bhs. Indonesia	Nilai UN Bhs. Inggris	Nilai UN Matematika	Nilai UN IPA	Nilai Tes Akademik	Nilai Tes Psikotes	Nilai Tes Fisik	Pilihan Jurusan
1	0005027541	SELVIANE	P	C	E	E	D	D	C	D	Akuntansi
2	9925837809	SHERLY	P	C	E	E	D	C	E	D	Akuntansi
3	0017153399	TANIA	P	D	E	E	E	D	E	C	Akuntansi
4	9962706154	TEDY SALIM	L	E	E	E	E	E	E	E	Akuntansi
5	0002202367	UMMAIROH	P	B	D	E	E	B	E	C	Akuntansi
6	0002146442	VICKY RAFIKA	P	D	E	E	D	E	D	C	Akuntansi
7	0002160168	ZEKI NURKOPIFAH	P	B	D	D	E	C	E	E	Akuntansi
8	9992622483	ALDI DARMAWAN	L	D	D	D	E	D	C	D	Multimedia
9	9992622534	ALDIANSYAH RAHMAH PUTRA	L	D	E	C	D	C	E	D	Multimedia
10	0002144183	ANDI HARTAWAN WIJAYA SALIM	L	E	E	E	D	D	E	C	Multimedia
11	0002147961	CAMELIA SAFITRI	P	C	D	E	E	E	E	E	Multimedia
12	0002140546	DIMAS FANI RAHMAN	L	C	E	E	C	D	C	D	Multimedia
13	0002141082	DONNY BAMBANG APRILIANTO	L	D	D	E	E	C	E	D	Multimedia
14	0018911387	ENDI SUHENDI	L	D	C	E	D	D	E	C	Multimedia
15	0002140583	ERWIN RAMADHAN	L	E	E	E	E	E	E	E	Multimedia
16	0002931080	HERYANTO	L	E	E	E	E	E	D	C	Multimedia
17	0003130618	IKBAL ADE SAPUTRA	L	D	E	E	E	C	E	E	Multimedia
18	9981846441	ILHAM DAHMARA	L	E	D	C	E	D	C	D	Multimedia
19	0002129305	INDRA AGUS NUGRAHADI	L	E	E	E	E	C	E	D	Multimedia
20	9992200895	MARTIN	L	D	D	C	B	D	E	C	Multimedia

3.1 Application of Algorithm C4.5

Manual calculation using algorithm C4.5 by using 271 sample training data that has been done variable selection. The following steps - steps data classification using C4.5 algorithm:

1. Prepare training data. The data used is 271 sample training data.
2. Calculate the number of students who choose to major in Accounting and Multimedia majors.
3. Calculate the total entropy value where it is known that the number of students who choose Accounting majors amounted to 192 and students who chose the Multimedia department amounted to 79.

   Entropy (S) = $\sum_{i=1}^{n} -pi.\log_2 pi$

   = (-192/271) * $\log_2$(192/271) + (-79/271) * $\log_2$(79/271)

   = **0,8707**

4. Calculate the gain value for each attribute. Then find the highest gain value. Attribute with the highest gain value then the attribute is used as root (root). For example, the value of gain for attribute value of UN Indonesian Language:

   $$\text{Gain}(S,A) = \text{Entrophy}(s) - \sum_{i=1}^{n} \frac{|si|}{|s|} * \text{Entrophy}(si)$$

   = 0,8707 - ((0/0*0) + (57/44*0,7746) + (38/28*0,8315) + (56/36*0,9403) + (120/84*0,8813))

   = **0,0412**

From the calculation of entropy and gain values for all attributes performed to obtain the highest gain value that will be used as the root. The calculation results are shown in the following table.

Table 4 :. Result of calculation of entropy and gain value with C4.5 algorithm.

No	Atribut	NilaiAtribut	JumlahKasus Total	JumlahKasus Multimedia	JumlahKasusAkutansi	Entropy	Information Gain	Split Info	Gain Ratio
1	Total		271	79	192	0.8707			0
2	nilai_un_bhs_indo	A	0	0	0	0	0.0066	1.861	0.0035
3	nilai_un_bhs_indo	B	57	13	44	0.7746	0.0066	1.861	0.0035
4	nilai_un_bhs_indo	C	38	10	28	0.8315	0.0066	1.861	0.0035
5	nilai_un_bhs_indo	D	56	20	36	0.9403	0.0066	1.861	0.0035
6	nilai_un_bhs_indo	E	120	36	84	0.8813	0.0066	1.861	0.0035
7	nilai_un_bhs_ing	A	0	0	0	0	0.003	1.0118	0.003
8	nilai_un_bhs_ing	B	1	0	1	0	0.003	1.0118	0.003
9	nilai_un_bhs_ing	C	6	2	4	0.9183	0.003	1.0118	0.003
10	nilai_un_bhs_ing	D	72	23	49	0.9038	0.003	1.0118	0.003
11	nilai_un_bhs_ing	E	192	54	138	0.8571	0.003	1.0118	0.003
12	nilai_un_mtk	A	0	0	0	0	0.0099	0.8442	0.0117
13	nilai_un_mtk	B	0	0	0	0	0.0099	0.8442	0.0117
14	nilai_un_mtk	C	6	4	2	0.9183	0.0099	0.8442	0.0117
15	nilai_un_mtk	D	51	14	37	0.8479	0.0099	0.8442	0.0117
16	nilai_un_mtk	E	214	61	153	0.8622	0.0099	0.8442	0.0117
17	nilai_un_ipa	A	1	0	1	0	0.0152	1.3026	0.0117
18	nilai_un_ipa	B	3	1	2	0.9183	0.0152	1.3026	0.0117
19	nilai_un_ipa	C	56	20	36	0.9403	0.0152	1.3026	0.0117
20	nilai_un_ipa	D	29	4	25	0.5788	0.0152	1.3026	0.0117
21	nilai_un_ipa	E	182	54	128	0.8772	0.0152	1.3026	0.0117
22	nilai_tes_akademik	A	0	0	0	0	0.0106	1.835	0.0058
23	nilai_tes_akademik	B	19	4	15	0.7425	0.0106	1.835	0.0058
24	nilai_tes_akademik	C	84	21	63	0.8113	0.0106	1.835	0.0058
25	nilai_tes_akademik	D	93	34	59	0.9472	0.0106	1.835	0.0058
26	nilai_tes_akademik	E	75	20	55	0.8366	0.0106	1.835	0.0058
27	nilai_tes_psikotes	A	0	0	0	0	0.022	1.6519	0.0133
28	nilai_tes_psikotes	B	3	0	3	0	0.022	1.6519	0.0133
29	nilai_tes_psikotes	C	82	21	61	0.8208	0.022	1.6519	0.0133
30	nilai_tes_psikotes	D	89	35	54	0.9669	0.022	1.6519	0.0133
31	nilai_tes_psikotes	E	97	23	74	0.7902	0.022	1.6519	0.0133
32	nilai_tes_fisik	A	0	0	0	0	0.0412	1.5893	0.0259
33	nilai_tes_fisik	B	2	0	2	0	0.0412	1.5893	0.0259
34	nilai_tes_fisik	C	59	26	33	0.9898	0.0412	1.5893	0.0259
35	nilai_tes_fisik	D	99	17	82	0.6616	0.0412	1.5893	0.0259
36	nilai_tes_fisik	E	111	36	75	0.909	0.0412	1.5893	0.0259

Based on the calculation of entropy and gain in Table-4, the attribute value of Physical Test has the highest gain value of 0.0412, so the attribute of Physical Test Value is the root node in the decision tree. Perform entropy and gain calculation until last root.

4. RESULT AND DISCUSSION

Test the system using the Weka 3.8 tool. The result of classification process with C4.5 algorithm yields decision of knowledge of rules or rule as follows:

1. If Physical Test Value = B then Accounting
2. If Value of Physical Tests = E then Accounting
3. If Value of Physical Tests = D then Accounting
4. If Value of Physical Test = C and Test Value psychotest = B then Accounting
5. If Physical Test Value = C and Test Value psychotest = E then Accounting
6. If Value of Physical Test = C and Test Value psychotest = C then Multimedia
7. If Value of Physical Test = C and Test Value psychotest = D and Value of UN English Language = B then Accounting
8. If Value of Physical Test = C and Test Value psychotest = D and Value of UN English Language = D then Multimedia
9. If Physical Test Value = C and Test Value psychotest = D and Value of UN English Language = C then Accounting
10. If Value of Physical Test = C and Test Value psychotest = D and Value of UN English Language = C then Accountin.

4.1 Comparison with Method C4.5 with Mete Naïve Bayes

The author also tries to compare the C4.5 method with the Naive Bayes Method using the same tools of Weka, the value of Correctly Classified Instances is 74.5387% with the C4.5 method. While with method of Naive Bayes Correctly Classified Instances equal to 67.1587%. For that the value of Accuracy is greater with C4.5 method. The comparison can be seen in the following figure:

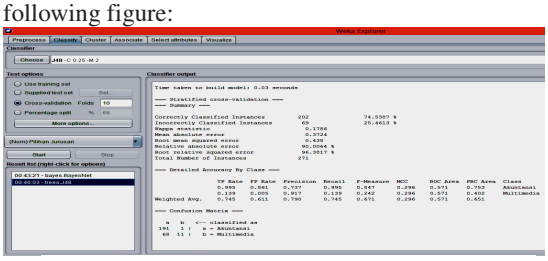

Figure 2 Results of Method C4.5 with Weka tools

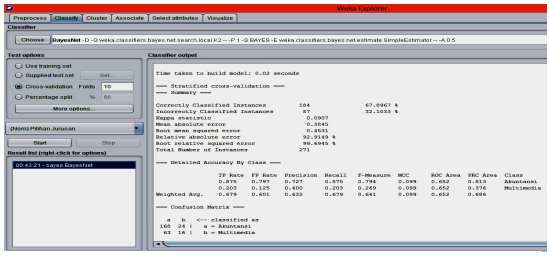

Figure 3 Results of Naïve Bayes Method with Weka tools

4.2 Evaluation Model with Confusion

Matrix Evaluation of the Confusion Matrix model will form a matrix consisting of accurancy, precision and recall, then testing is done on the training data already prepared into the Confusion Matrix. First, testing Confusion Matrix using C4.5 algorithm can be seen in the table below.

Table 5: Confusion Matrix with C4.5 Algorithm

	Identified Accounting	Identified Multimedia
Accounting	191	1
Multimedia	68	11

$$Precision = \left(\frac{d}{b+d}\right) x100\%$$

Precision = 11/(1+11) x 100%
Precision = 0,9166 = 91,66%

$$Recall = \left(\frac{d}{c+d}\right) x100\%$$

Recall = 11/(68+11) x 100%
Recall = 0,1392 = 13,92%

$$Accuracy = \left(\frac{a+d}{total\ sample}\right) x100\%$$

Accurancy = (191+11)/(191+1+68+11) x 100%
Accurancy = 0,7453 = **74,53**

Based on the calculation, the value of data accuracy on the algorithm C4.5 of 74.53% is a calculation of training data, consisting of 271 records data, 191 data classified Accounting and 1 data classified Multimedia, but apparently Accounting, 68 data correctly classified Accounting and 11 data predicted Accounting turns Multimedia.

The second test of Confusion Matrix using Naïve Bayes method can be seen in the table below:

Table 6: Confusion Matrix with Naïve Method

	Identified Accounting	Identified Multimedia
Accounting	168	24
Multimedia	63	16

$$Precision = \left(\frac{d}{b+d}\right) x100\%$$

Precision = 16/(24+16) x 100%
Precision = 0,4 = 40 %

$$Recall = \left(\frac{d}{c+d}\right) x100\%$$

Recall = 16/(63+16) x 100%
Recall = 0,2025 = 20,25 %

$$Accuracy = \left(\frac{a+d}{total\ sample}\right) x100\%$$

Accurancy = (168+16)/(168+24+63+16) x 100%
Accurancy = 0,6789 = **67,89%**

Based on the calculation, the value of data accuracy in Naïve Bayes Method of 67.89% is a calculation of training data, consisting of 271 records of data, 168 classified data Accounting and 24 data classified Multimedia, but apparently Accounting, 63 data are correctly classified Accounting and 16 data predicted Accounting turned out Multimedia. From result comparison of Method C4.5 with Naïve Bayes Method from 271 dataset by using WEKA tool obtained data as follows:

Table 7: Comparison of Method C4.5 with Naïve Bayes Method

	Method C4.5	Method Naïve Bayes
Precision	91,66%	40 %
Recall	13,92%	20,25 %
Accurancy	74,53%	67,89%

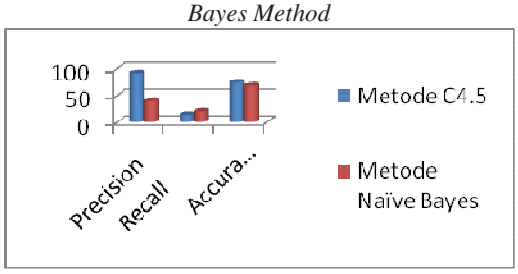

Figure 3: Comparison Chart Method C4.5

From the evaluation and validation results obtained that the C4.5 method has a value of accuracy and performance is higher than the Naïve Bayes method. Rule result from C4.5 method applied to the making of application for determination of student majors at SMK Bina Mandiri Teluk Naga by using programming language PHP and MySQL as database.

5. CONCLUSIONS AND SUGGESTIONS

5.1 CONCLUSION

From the research that has been done, then drawn the conclusion as follows:

1. Data mining by using the attribute of National Exam Score (Indonesia Language, Mathematics, English and Science) and Vocational Test Score (Psychotest Test Score, Academic Test Score, Physical Test Value) using Algorithm C4.5 method has greater accuracy value than with Naïve Bayes method that is equal to 74,53% while Naïve Bayes accuracy value 67,89% with difference value equal to 6,64%, and Rule generated from method C4.5 can be applied to making system of determination of new student department so that determination majors no longer by choice.
2. With the making of the system determining the majors with data mining using C4.5 method, can minimize the students moved majors, so the chances of success in the study at SMK Bina Mandiri Teluk Naga can be achieved well.

5.2 SUGGESTIONS

But there are some things that need to suggest authors for the development of this research, among others, such as testing data by adding some new data attributes. Can also be done by comparison with other algorithm methods that support testing existing data, so that can be obtained a better accuracy rate.

REFERENCES

[1] K. Sumathi, PhD 2016 K. Sumathi, PhD., S.Kana PhD., K Nagarajan, Data Mining: Analysis of student database using Classification Techniques, *International Journal of Computer Applications (0975 – 8887) Volume 141 – No.8, May 2016*

[2]Liliana Swastina2013 *Liliana Swastina, Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa Jurnal Gema Aktualita,Vol. 2 No. 1,Juni 2013*

[3] Heri Susanto, 2014 Heri Susanto, *Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan Dan Prestasi Masa Lalu Jurnal Pendidikan Vokasi, Vol 4, Nomor 2, Juni 2014*

[4] FinaNasar, 1993 Fina Nasar, *Penerapan Algoritma C4.5 Dalam Pemilihan Bidang Peminatan Program Studi Sistem Informasi Di Stmik Potensi Utama Medan. Seminar Nasional Informatika 2014*

[6] Yohana Nugraheni, 2011, Yohana Nugraheni Data Mining Dengan Metode Fuzzy Untuk

Customer Relationship Management (Crm) Pada Perusahaan Retail, mahasiswa Universitas Udayana Denpasar

[7] David Hartanto Kamagi 2014 David Hartanto Kamagi, Seng Hansun, *Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa,ULTIMATICS, Vol. VI, No. 1\Juni 2014ISSN 2085-4552*

[8] Kadek Juni Arta 2016 Kadek Juni Arta1, Gede Indrawan2, Gede Rasben Dantes 3, *Data Mining Rekomendasi Calon Mahasiswa Berprestasi Di Stmik Denpasar Menggunakan Metode Technique For Others Reference By Similarity To Ideal Solution, ISSN: 2303-3142 Vol. 5, No. 2, Oktober 2016*

[9] Mujib Ridwan 2016 Mujib Ridwan, Hadi Suyono, dan M. Sarosa.*Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier Jurnal EECCIS Vol.7, No. 1, Juni 2013*

[10] Eka Sabna 2016 Eka Sabna, Muhardi, *Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen, Motivasi, Kedisiplinan, Ekonomi, dan Hasil Belajar Jurnal CoreIT, Vol.2, No.2, Desember 2016 ISSN: 2460-738X*