Department Recommendations for Prospective Students Vocational High School of Information Technology with Naïve Bayes Method

Dyna Marisa Khairina[1)], Fajar Ramadhani[2)], Septya Maharani[3)], Heliza Rahmania Hatta[4)]
Computer Science, Faculty of Mathematics and Natural Science
Mulawarman University
Samarinda, Indonesia
dyna.ilkom@gmail.com[1)], fajar.nyappy@gmail.com[2)], septyamaharani@gmail.com[3)], heliza_rahmania@yahoo.com[4)]

***Abstract*—The selection of the appropriate department in Vocational High School gives the big difference of the ability in thinking for the students. Most students tend to follow their friends in choosing the departments so that the students are possible to feel incompatible with the departments followed and then fail. A student needs to find the department that is suitable to the interest, ability and talent of the student. Each student has the ability to think differently and different talents as well. Naive Bayes methods are used as decision support to provide recommendations for consideration in the selection of departments appropriately in accordance with the interests, abilities, and talents tendency of students by using reference data to make decisions. Naïve Bayes is a classification with a method of probability and statistics. Bayes' approach in classification is to find the highest probability with attributes input. All data are entered for calculating the percentage probability in accordance with the criteria in order to obtain the recommendation of appropriate department for prospective students. The criteria used is 4 (four) criteria. The result obtained from the research is the department election system to give the suitable recommendations for the prospective students in considering to decision making.**

Keywords—decision making; naïve bayes; vocational high school

I. INTRODUCTION

As students continue their formal education to vocational high schools there will be little difficulty in determining the choice. There are various options of majors when the students enter the vocational level. The selection of the appropriate departments in vocational provide a large difference in the way of thinking of the students. It is common that some vocational students do not get along with the majors chosen so that they fail and lose the time, energy and mind. The students need to find a department that matches the interests, abilities, and talents. Each student has the ability to think differently and different talent to do something.

This study makes a system that implements a method to provide the best recommendation for the in choosing the department for the candidate of the students majoring in IT based on specified criteria. Naïve Bayes methods is the classification with the methods of probability and statistics. Bayes' approach in classification is to find the highest probability with attributes input. The calculation of the percentage of probability to choose majors in accordance with the interests, abilities and talents tendency of students based on standard criteria can produce more accurate recommendation majors for prospective students.

II. DECISION SUPPORT SYSTEMS

Decision Support System (DSS) is a computer-based information system that performs the approaches to produce variety of alternative decision to assist particular party in addressing the problem by using data and model. A DSS only provides an alternative decision and next is handed over to the user to make a decision [1].

The concept of Decision Support System (DSS) is a computer-based interactive system that helps making decision and utilizes data and models to solve problems that are unstructured and semi-structured. Decision Support Systems (DSS) is designed to support all stages of decision making, which starts from the stage of identifying problems, selecting the relevant data, determine the approach used in the process of decision making and the activities of evaluation of alternative selection [2].

The decision support system characteristics [3]:

1. The decision support system is designed to help decision makers in solving problems that are semi-structured or unstructured.

2. The decision support system is designed in such a way so that it can be used easily by people who do not have the basic capability of operating high level computer.

3. Decision support system is designed by emphasizing on the flexibility and high adaptability. So it is easily adapted to a variety of environmental changes that occur in the user's needs.

III. NAÏVE BAYES

Naïve Bayes is classification with a probability method and statistics expressed by the British scientist Thomas Bayes. Naïve Bayes, for each class of decisions, calculates the

probability on condition that the class of decisions is correct, where vector is object information [4].

Steps to the calculation of the method of Naïve Bayes:

1. Determining criteria.
2. Creating a table of rules.
3. Calculating the probability of occurrence of each value to the attributes of each criterion.
4. Calculating the value of each criterion.
5. Calculating the value of probability.

Probability involved in producing a final estimate is calculated as the sum of the frequencies of "master" decision table. Bayes approach when classifying is to search the highest probability (V_{MAP}) with attributes input ($a_1, a_2, a_3, \ldots, a_n$).

$$vMAP = \arg\max P\ (\ v_j\ |a_1, a_2, a_3, \ldots, a_n\ )\ v_j \in V \quad (1)$$

Bayes theorem itself is originated from the formula:

$$P\ (A \mid B\ ) = \frac{P\ (B \cap A\ )}{P(B)} \quad (2)$$

Where P (A | B) means the chances of A if the condition of B is known.

Then from the equation (2) is found that:

$$P(B \cap A) = P(B \mid A)P(A) \quad (3)$$

So we get *bayes* theorem:

$$P(A \mid B) = \frac{P(B|A)P(A)}{P(B)} \quad (4)$$

From equation (4) can be written as:

$$V_{MAP} = \arg max \frac{P(a_1,a_2,a_3,\ldots,a_n|v_j)P(v_j)}{P(a_1,a_2,a_3,\ldots,a_n)} \quad (5)$$

Because the value of $P(a_1, a_2, a_3, \ldots, a_n)$ is constant for all v_j, then this equation can be written as:

$$V_{MAP} = \arg\max P(a_1, a_2, a_3, \ldots, a_n|v_j)P(v_j)\ \ v_j \in V \quad (6)$$

To calculate $P(a_1, a_2, a_3, \ldots, a_n |v_j)$ can be more difficult because of the number of terms $P(a_1, a_2, a_3, \ldots, a_n |v_j)$ can be very large. This is due to the number of terms is equal to the sum of all combinations of word position multiplied by the number of existing categories.

Naïve Bayesian Classifier simplifies this by assuming that the features contained inside are not mutually dependent or independent, each word is independent of each other. In other words:

$$P(a_1, a_2, a_3, \ldots, a_n|v_j) = \prod_i P(a_i|v_j) \quad (7)$$

By substituting equation (6) with equation (7) will result in:

$$v_{MAP} = arg\ \max_{v_j \in V} P(v_j) \prod_i P(a_i|v_j) \quad (8)$$

$P(v_j)$ can be interpreted as opportunities of j category. Where $|kriteria_j|$ is the number of criteria on the category j and |jum_kejadian| is the number of events that have been in the rules. While $P(a_i|v_j)$ is an opportunity a_i if it is known that v_j circumstances, in a sense, is an opportunity to criteria i, , in j category criteria. It can be written:

$$P(v_j) = \frac{|kriteria_j|}{|jum\_kejadian|} \quad (9)$$

Where n_i is the number of occurrences in the criteria a_i in v_j category. And $|kriteria\_ v_j|$ is the number of the existing criteria in the category v_j.

$$P(a_i|v_j) = \frac{n_i}{|\ kriteria\_v_j|} \quad (10)$$

[5].

IV. NAÏVE BAYES FOR PROBLEM SOLVING

The steps in the department selection process for prospective students majoring in Vocational High School using Naïve Bayes methods are:

1. Determining the criteria in the election department system.

Determining the type of criteria used in the calculation using the method of Naïve Bayes becomes the first stage in the election system of the majors because the criteria becomes essential requirements in determining the selection of majors. From these criteria, it will result competent department according to desired criteria. The criteria used in the major election system can be seen in Table 1.

2. Creating a table of rules

In this step, the criteria in Table 1 are organized in the form of a table of rules, where the Naïve Bayes method uses only 4 rules. Table of rules in the department election system can be seen in Table 2.

TABLE I. CRITERIA FOR ELECTORAL SYSTEMS DEPARTMENT

Criteria	Description
MTK	Value of Mathematics
ENG	Value of English
Minat	Interests of Prospective Students
Bakat	Talent of Prospective Students

TABLE II. TABLE RULES

No	MTK	ENG	MINAT	BAKAT	DEPARTMENT
1	>75	>75	RPL	Multimedia	RPL
2	>75	>75	RPL	Programing	RPL
3	>75	>75	RPL	Teknik Komputer	RPL
4	>75	>75	MM	Multimedia	MM
5	>75	>75	MM	Programing	MM
6	>75	>75	MM	Teknik Komputer	MM
7	>75	>75	TKJ	Multimedia	TKJ
8	>75	>75	TKJ	Programing	TKJ
9	>75	>75	TKJ	Teknik Komputer	TKJ
10	>75	>75	RPL	No	RPL
11	>75	>75	MM	No	MM
12	>75	>75	TKJ	No	TKJ
13	>75	70-75	RPL	Multimedia	RPL
14	>75	70-75	RPL	Programing	RPL
15	>75	70-75	RPL	Teknik Komputer	RPL
16	>75	70-75	MM	Multimedia	MM
17	>75	70-75	MM	Programing	RPL
18	>75	70-75	MM	Teknik Komputer	MM
19	>75	70-75	TKJ	Multimedia	TKJ

20	>75	70-75	TKJ	Programing	TKJ
21	>75	70-75	TKJ	Teknik Komputer	TKJ
22	>75	70-75	RPL	No	RPL
23	>75	70-75	MM	No	MM
24	>75	70-75	TKJ	No	TKJ
25	70-75	>75	RPL	Multimedia	RPL
26	70-75	>75	RPL	Programing	RPL
27	70-75	>75	RPL	Teknik Komputer	TKJ
28	70-75	>75	MM	Multimedia	MM
29	70-75	>75	MM	Programing	MM
30	70-75	>75	MM	Teknik Komputer	MM
31	70-75	>75	TKJ	Multimedia	TKJ
32	70-75	>75	TKJ	Programing	TKJ
33	70-75	>75	TKJ	Teknik Komputer	TKJ
34	70-75	>75	RPL	No	RPL
35	70-75	>75	MM	No	MM
36	70-75	>75	TKJ	No	TKJ
37	70-75	70-75	RPL	Multimedia	MM
38	70-75	70-75	RPL	Programing	RPL
39	70-75	70-75	RPL	Teknik Komputer	TKJ
40	70-75	70-75	MM	Multimedia	MM
41	70-75	70-75	MM	Programing	MM
42	70-75	70-75	MM	Teknik Komputer	MM
43	70-75	70-75	TKJ	Multimedia	TKJ
44	70-75	70-75	TKJ	Programing	TKJ
45	70-75	70-75	TKJ	Teknik Komputer	TKJ
46	70-75	70-75	RPL	No	RPL
47	70-75	70-75	MM	No	MM
48	70-75	70-75	TKJ	No	TKJ

3. Calculating the probability of occurrence of each value of each criterion.

At this stage, each of the values of the criteria will be made into a table on the probability of occurrence of each value. Table of probability of the occurrence of the value for the criterion of mathematical values can be seen in Table 3. Table probability of the occurrence of the value for the criterion of the value of the English can be seen in Table 4. Table probability of the occurrence of the value for the criterion of interest can be seen in Table 5. Table probability of the occurrence of the value for the criterion of talent can be seen in Table 6.

TABLE III. TABLE PROBABILITY MATH SCORES

Mathematical probability value						
Value	***The number of events in select***			***Probability***		
	RPL	MM	TKJ	RPL	MM	TKJ
>75	9	7	8	9/14	7/16	8/18
70-75	5	9	10	5/14	9/16	10/18
Amount	14	16	18	1	1	1

TABLE IV. TABLE PROBABILITY VALUE ENGLISH

English probability value						
Value	***The number of events in select***			***Probability***		
	RPL	MM	TKJ	RPL	MM	TKJ
>75	7	8	9	7/14	8/16	9/18
70-75	7	8	9	7/14	8/16	9/18
Amount	14	16	18	1	1	1

TABLE V. TABLE PROBABILITY VALUE INTERESTS

The probability of the value of Interests						
Value	***The number of events in select***			***Probability***		
	RPL	MM	TKJ	RPL	MM	TKJ
RPL	13	1	2	13/14	1/16	2/18
MM	1	15	0	1/14	15/16	0/18
TKJ	0	0	16	0/14	0/16	16/18
Amount	14	16	18	1	1	1

TABLE VI. TABLE PROBABILITY VALUE TALENT

The probability of the value of Talent						
Value	***The number of events in select***			***Probability***		
	RPL	MM	TKJ	RPL	MM	TKJ
Programing	5	3	4	5/14	3/16	4/18
Multimedia	3	5	4	3/14	5/16	4/18
Teknik Komputer	2	4	6	2/14	4/16	6/18
No	4	4	4	4/14	4/16	4/18
Amount	14	16	18	1	1	1

TABLE VII. DATA TEST SYSTEMS

Criteria	**Data Prospective Students**
Name	Fajar Ramadhani
MTK	72
ENG	92
Minat	TKJ
Bakat	Multimedia

4. Calculating the value of RPL likelihood, MM likelihood, and likelihood TKJ.

RPL likelihood values, MM likelihood, and TKJ likelihood are taken from the table of probability of occurrence of each attribute, where the final result of the value is used for the probability value.

From the data of the prospective student presented in Table 7, the likelihood value of RPL, MM likelihood, and likelihood TKJ can be calculated, to calculate this value, data are taken from the table of the probability of occurrence of every criterion from each criterion:

Calculation of Data of Prospective Students

Likelihood RPL = 5/14 x 7/14 x 0/14 x 3/14 x 14/48

$$= \frac{0}{1843968} = 0$$

Likelihood MM = 9/16 x 8/16 x 0/16 x 5/16 x 16/48

$$= \frac{0}{3145728} = 0$$

Likelihood TKJ = 10/18 x 9/18 x 16/18 x 4/18 x 18/48

$$= \frac{103680}{5038848} = 0.0206$$

5. Calculating the probability value

Calculating the probability value can be calculated by performing normalization from every likelihood RPL, MM likelihood, and TKJ likelihood so the number of earned value = 1, where the biggest value is considered suitable to be a suggestion of the majors that will be chosen.

Calculation of Probability Value of the Prospective Students Data

$$RPL = \frac{\text{Value likelihood RPL}}{\text{Value likelihood RPL} + \text{Value likelihood MM} + \text{Value likelihood TKJ}}$$

$$= \frac{0}{0+0+0{,}0206}$$

$$= \frac{0}{0{,}0206} = \mathbf{0}$$

$$MM = \frac{\text{Value likelihood MM}}{\text{Value likelihood RPL} + \text{Value likelihood MM} + \text{Value likelihood TKJ}}$$

$$= \frac{0}{0+0+0{,}0206}$$

$$= \frac{0}{0{,}0206} = \mathbf{0}$$

$$TKJ = \frac{\text{Value likelihood TKJ}}{\text{Value likelihood RPL} + \text{Value likelihood MM} + \text{Value likelihood TKJ}}$$

$$= \frac{0{,}0206}{0+0+0{,}0206}$$

$$= \frac{0{,}0206}{0{,}0206} = \mathbf{1}$$

From the data above, RPL probability value is 0, the value of probability MM is 0, and the value Probability Probability TKJ is 1. TKJ probability value is bigger than the value of Probability of RPL and the value of Probability of MM, so the result that will be obtained is the department suggested according to the criteria, which is TKJ.

User runs this system by entering the data of prospective students, and then user comes into the process of selecting majors in accordance with the criteria required or assigned. For prospective students form data input can be seen in Figure 1.

After the data of prospective students are successfully incorporated into the system in accordance with the existing criteria, the selection process is then performed by applying the Naïve Bayes method as the department recommendation to prospective students. Figure 2 displays the data that has been processed in the election based on the data of prospective students, which results the data of department election that is selected as the recommendation of the department suitable with the required criteria. The detail result of naïve Bayes calculation can be seen in Figure 3.

From the Naive Bayes calculation results in Figure 3, can be seen that the probability value from those three outputs, it is found that TKJ probability value is bigger than RPL and MM so that the recommendation department suggested from the data sample is TKJ department.

Fig. 1. Input Data Prospective Students

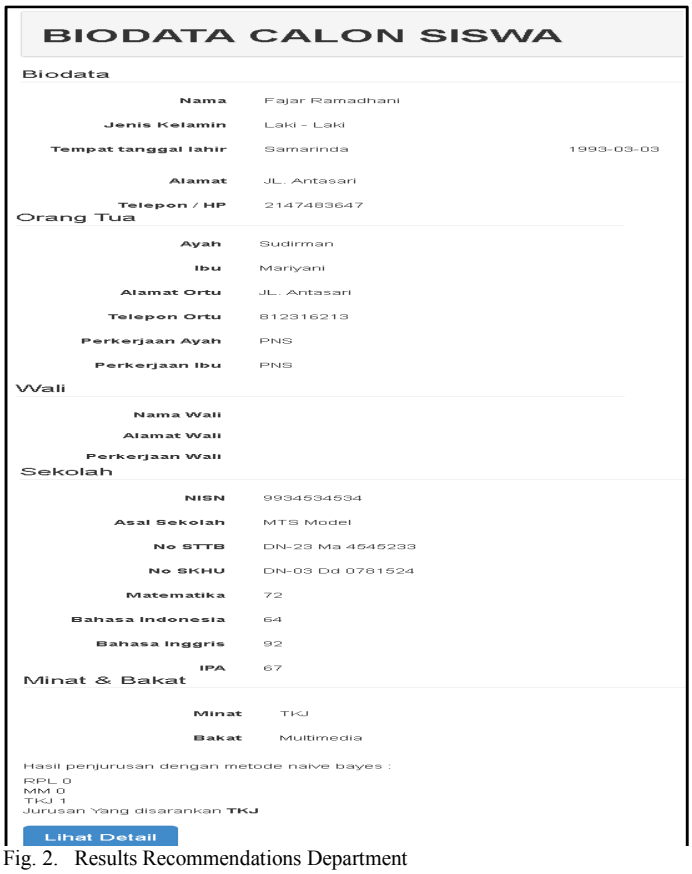

Fig. 2. Results Recommendations Department

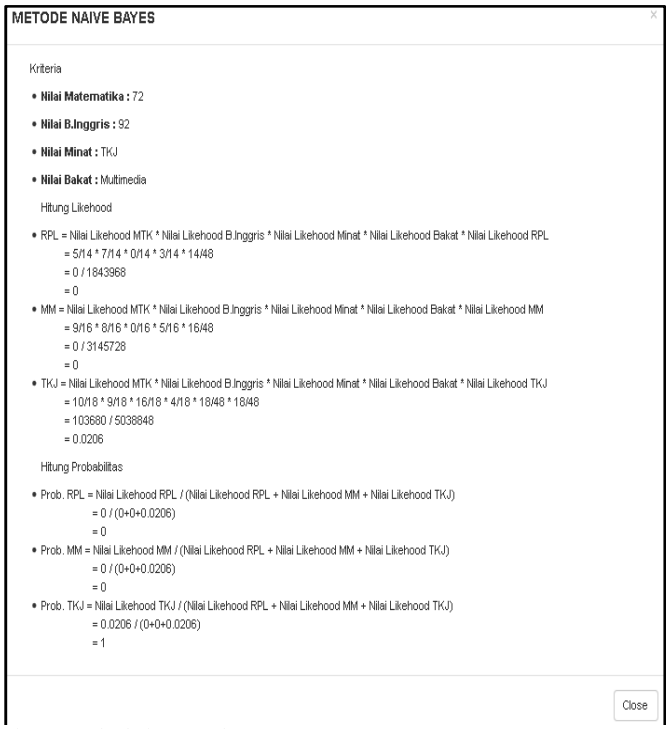

Fig. 3. Calculation Results Naïve Bayes

V. CONCLUSION

Based on the results of research about the election of the department for the prospective students of Vocational High School majoring in Information Technology with the Naïve Bayes method, it is resulted the system test data of probability value according to the sample given where it is resulted the calculation with the probability value, that the TKJ has bigger value compared to the two other majors. The highest value of the probability of the department is suggested as a recommended department for the prospective student. This department election system runs based on the criteria determined along with the table of rules that becomes the reference data in considering to choose the department.

REFERENCES

[1] Daihani, D, U, "Komputerisasi Pengambilan Keputusan," Ghalia Indonesia, Bogor, 2001.

[2] Kosasi, S, "Sistem Penunjang Keputusan (Decision Support System)," Pontianak, 2002.

[3] Turban, E, "Decision Support System and Inteligent Systems," Volume 1. Issue 7, Andi Offset, Yogyakarta, 2005.

[4] Olson, David L. and Dursun, Delen, "Advanced Data MiningTechniques," USA: Springer-Verlag Berlin Heidelberg, 2008.

[5] Anugroho, P, "Klasifikasi Spam Email dengan Metode Naïve Bayes Classifier" Thesis Teknologi Informasi, Institut Teknologi Sepuluh Nopember, 2010.