International Journal of Scientific Research and Engineering Development-– Volume 2 Issue 6, Nov- Dec 2019

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Predicting Student Ability in Competition in Work World Using Naïve Bayes Method in SMKN 4 Pandeglang

MuhamadJuwayni*, BambangKrismalela**

*(SMKN 4 Pandeglang, andPandeglang

Email: m.juwayni@gmail.com)

** (SMKN 4 Pendeglang, andPandeglang

Email: bambang.krismalela83@gmail.com)

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

Abstract:

ThehighunemploymentrateinBantenProvinceisbasedondatafromBPSwhich reached9.28%nationally.ThispercentagebecomesBanten Provinceasthehighest ranked province in Indonesia. Thehigh rate ofunemployment thatcauses vocationalschoolgraduatestomaketheroleofvocationalschoolsasaschoolthat aimsto producegraduatesreadyforworkis in the spotlight.Veryfastlessonsfrom vocationalschoolswitheducationalfactors.Schoolsneedtodoevaluations that canbeusedintheworkplace. Evaluationcan bedonefromtheresultsofthe measurementofstudents'abilitiesbasedonacademicgradescurrentlytheschool doesnotuse academicvaluedatato predictstudents'abilities.Onthatbasis,the researchwas conducted to predictstudents'abilities byusing datamining techniques,statistical methodsthatcanbeusedbyNaïveBayes.Dataminingtoextractinformationfromavailabledata.Fromthe results ofthestudyproducedadataminingmodelwiththeNaïveBayesalgorithm,amodel thatwaslater madeinthemakingofprototypestopredict students'abilityto competeinthe worldofwork.Theresults ofdataminingmodeltestingwith confusionmatrixresultedin96.0725%accuracy.

Keywords ***—information system, code igniter, assessment, processing data.***

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

I. INTRODUCTION

The development of information technology is currently happening so fast. Every agency or organization can take advantage of developments in technology and information in supporting business progress and the goals to be achieved. The development of technology and information and supported by good quality human resources is a key factor in the success of an organization in the era of free trade.

Banten Province was established in 2000 as a regional division of West Java. Banten Province is a province that borders directly with the capital city of Jakarta. At present the population of Banten province is based on data from the national statistical agency in 2016 of 12,203,148 inhabitants. According to data from the national statistics agency in 2017, Banten province currently occupies the second position in Indonesia as a province with an open unemployment rate with a percentage of unemployment as much as 9.28%.

Table I. Provincial Open Unemployment Rate in Indonesia

Provinsi	2014		2015		2016		2017	
	Feb	Agu	Feb	Agu	Feb	Agu	Feb	Agu
Maluku	6.59	10.51	6.72	9.93	6.98	7.05	7.77	9.29
Banten	9.87	9.07	8.58	9.55	7.95	8.92	7.75	9.28
JawaBarat	8.66	8.45	8.40	8.72	8.57	8.89	8.49	8.22
SulawesiUtara	7.27	7.54	8.69	9.03	7.82	6.18	6.12	7.18
KepulauanRiau	5.26	6.69	9.05	6.20	9.03	7.69	6.44	7.16

The high unemployment rate in Banten province is inversely proportional to the large number of industries in Banten province. According to data from the Central Statistics Agency of Banten Province, industry figures in 2014 increased from the previous period in 2013, where the number of industries increased from a total of 1,570 to 1,682. Increasing industry does not reduce unemployment, this is certainly a job for the Banten provincial government in reducing unemployment. One of the factors that increase the unemployment rate is the lack of absorption of local labor by industries in Banten, where many graduates in Banten Province are not in accordance with industry needs. One of the graduates who accounted for the largest unemployment rate in Banten was a graduate of vocational high school or vocational school.

Indeed vocational high schools are places to create reliable human resources, who are ready to compete in the world of work. However, in reality vocational high school graduates in Banten Province contributed the highest unemployment rate after junior high school graduates. Comparison of the highest number of unemployment contributors can be seen from the following chart:

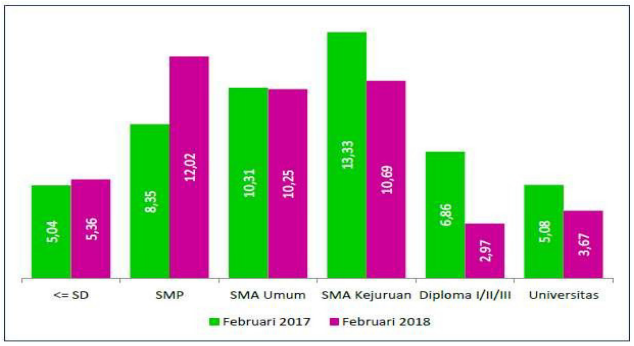

Figure 1.Banten Province open unemployment rate

Vocational high schools are chosen as a means of students and students in demanding knowledge and as provisions in finding work after graduation. Vocational High Schools provide several competency skills that can be selected by prospective students according to their interests. Some of the competency expertise offered in the digital era is now becoming a much sought after competency by companies in increasing productivity and achieving corporate goals, for example majoring in computer network engineering and software engineering.

In addition to debriefing in teaching and learning activities that take place at school, other activities in vocational high schools in improving student competency and quality in accordance with the curriculum are Internship. Internship as an activity held by vocational high schools and in collaboration with the industrial world. Internship is an education and training activity carried out in the business or industrial world in an effort to improve the quality of vocational high school students as a provision for the future to enter the workforce.

Internship is carried out for approximately 3 months by students and students, internship must be planned carefully by the school so that it can be applied correctly by the supervisor in the field in giving grades to participants internship. The value obtained will be used as the value stated on the Job Training Certificate (PKL) that will be obtained by the participating students. Based on the vocational high school curriculum, in the apprenticeship assessment process is divided into three things, namely the work ethic, the domain of attitude and social relations.

II. Research Method

A. Data Mining

From a business perspective, data mining is to achieve deeper analysis on a large number of enterprise data according to the established enterprise business goals, the aim is to discover the unknown, hidden potential rules and convert them into a corresponding model, thus supporting business decision support activities[1].

According to (Turban, et al, 2005) in Kusrini and EmhaTautiq[2] data mining is a term used to describe the discovery of knowledge in a database. Data mining is a process that uses statistical techniques, mathematics, artificial intelligence and machine learning to extract and identify information that is useful for related knowledge from various databases.

Then according to FajarAstutiHermawan [3] Data mining is a process that employs one or more computer learning techniques (machine learning) to automatically analyze and extract knowledge. Other definitions include induction-based learning (induction-based learning) is the process of forming general concept definitions which are carried out by observing specific examples of the concepts to be learned. Knowledge Discovery in Database (KDD) is the application of scientific methods to data mining. In this context data mining is one step in the process of KDD.

According to Sigit A. and Yuita A.S [4] states that Data mining discusses the extraction or collection of useful information from a data set. Information usually collected is hidden patterns in the data, relationships between data elements, or modeling for the purpose of data deepening.

B. Knowledge Discovery in Database

Knowledge Discovery in Database (KDD) is a computer-aided process to explore and analyze large amounts of data sets and extract useful information and knowledge. Data mining tools predict future behavior and trends, enabling businesses to make proactive and knowledge-based decisions. Data mining tools are able to answer business problems that have traditionally been too long to solve. Data mining tools roam the database in search of hidden patterns, finding predictive information that experts may miss because it is beyond their expectations[5].

The stages of the Process of Knowledge Discovery in Database can be broadly explained as follows[6]:

1. Selection
   The selection or selection of data from a set of operational data needs to be done before the information gathering stage in KDD begins. The selected data will be used for data mining process, stored in file form, separate from the operational database.
2. Preprocessing
   Before the data mining process can be carried out, it is necessary to clean up the data that is the focus of KDD. The cleaning process includes removing duplicate data, checking inconsistent data, and correcting errors in the data, such as typographical errors. The enrichment process is also carried out, which is the process of enriching existing data with data or other information that is relevant and needed for KDD, such as external data or information.
3. Transformation
   Converting data to an extension format that is suitable for processing in data mining. Some data mining methods require special data formats before they can be processed in data mining techniques. For example some standard methods such as association analysis and clustering can only accept categorical data input. Therefore data in the form of numerical numbers that continue to be divided into several intervals.
4. Data mining
   Data mining process of looking for interesting patterns or information in selected data using certain techniques or methods. Techniques, methods in data mining vary greatly, the selection of the right method is very dependent on the overall goals and processes of KDD.
5. 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 involves checking whether the pattern or information found is contrary to the facts or hypotheses that existed before. The stages - stages in KDD can be seen in Figure II - 1 below:

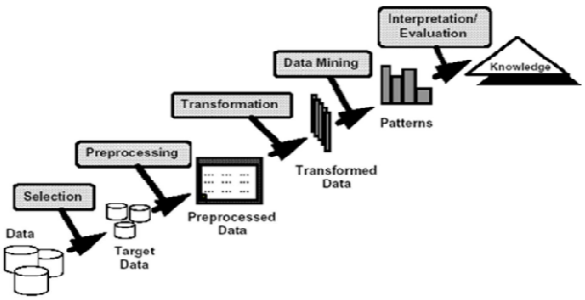

Figure 2. Knowledge Discovery Process in Database.

C. Naive Bayes Method

Bayesian Classification is a statistical classification that can be used to predict the probability of membership of a class. Bayesian Classification is based on the Bayes theorem which has the same classification capabilities as decision trees and neural networks. Bayesian Classification is proven to have high accuracy and speed when applied to large databases with data [2].

Naive Bayes is one of the applications of Bayes theorem in classification, Naive Bayes is based on the assumption of simplification that the attribute values are conditionally independent of each other if given an output value[7]. Bayes' theorem has the following general form:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

From the formula above as a basis for Bayesian theory as problem solving, we must know in advance several things including:

X: Sample data that has an unknown class (label).

H: The hypothesis that x is class specific data (label).

P (H | X): H hypothesis probability based on condition X (Posterior Probability).

P (H): Hypothesis probability H (Prior probability).

P (X | H): Probability of X based on conditions on hypothesis H.

P (X): Probability of X.

D. K-Nearest Neighbors Method

According to Kusrini and EmhaTaufq[2] Nearest Neighbors is an approach to find cases by calculating the closeness between new cases and old cases, which is based on matching the weights of a number of existing features.

Algortima K-Nearest Neighbors or commonly abbreviated as KNN has different ways of working compared to other classification methods. Classification methods in general will form a model, which is a function of mapping input output from training data, and using a model that has been formed to estimate the output of a new input.

The K-Nearest Neigbors method works based on the assumption that a data will have the same class or category as the data around it. This concept is called the concept of neighborhood. Figures II - 2 show the distribution of data that has two types of categories, namely positive and negative categories.

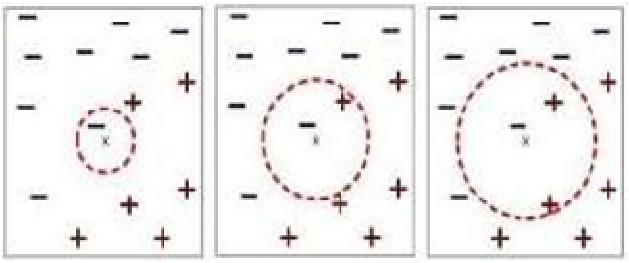

Figure 3. Illustration of Nearest Neighbors

III. RESULT AND DISCUSSION

The purpose of this study is to apply the Naive Bayes Method to create models in predicting the ability of graduate students to compete in the world of work. Based on these research objectives, the research method that will be used in this study is applied research methods. The results of the research conducted can be directly implemented to solve the problems faced[8].

This study uses the literature study method as an initial step by studying the theoretical basis of classification data mining using the Naive Bayes Method in several other literature and references. References include data from journals, the internet, e-books, and other documents related to this research.

In this study the method used is a quantitative method where the calculation process is carried out in accordance with the existing formula based on the method used to obtain a decision [8]. Calculations are performed on student data to obtain predictions of students' ability to compete in the world of work.

A. Research variable

The process of classifying student data in this study will be carried out by the Naive Bayes Method, for the attributes used in this study include:

1. Report Card Value

An attribute that contains the student report card grades. Report card grades are sorted into 4 categories, based on the 2013 curriculum. New

grades are to make it easier to do data transformation.

2. National Examination Score (UN / UNBK)

An attribute that contains a student's National Examination score. The UN score data is sorted into 4 categories, based on the 2013 curriculum. New grades are to make it easier to do data transformation.

3. Vocational Competency Examination (UKK)

It is an attribute of the vocational competency examination of students. The UKK value data is sorted into 4 categories based on the 2013 curriculum. New values are to make it easier when doing data transformation.

4. Industry

Internship attribute is the average value obtained by students during the internship process. Internship values are categorized into 4 categories.

5. Skills

The value of skills is an attribute of the value of skills possessed by students. Skill score data is sorted based on 2013 Curriculum. New grades are made to make it easier to do data transformation.

To make the classification used 4 attributes, each attribute has a value as a reference in the classification process, the following value scale for the attribute used:

Table II. Value Scale

RentangAngka	Huruf
86-100	A
62-85	B
38-61	C
0-37	D

B. Evaluation and Validation of Results

The process of experimentation and model testing uses parts of existing datasets. All datasets are then tested by the method proposed in the Weka 3.8.2 application. Evaluation of the formed model will be done by measuring accuracy. Accuracy is measured using the Confusion Matrix table.

Confusion matrix will describe the results of accuracy ranging from positive positive predictions, positive positive predictions, negative negative predictions that are true, and negative negative predictions. So that the model formed can be directly tested with randomly separated data with four variables.

C. Application of the Selected Method

1. Prototype Testing with User Acceptance Test

User Acceptance Test is a method of testing applications from the user side. User Acceptance Tests conducted on prototypes of data mining applications to determine the level of user acceptance of applications made. In user acceptance test testing, researchers conduct testing by adapting the characteristics of ISO 9126. The stages of testing are as follows:Determination of Testing Points

There are 4 (four) test points used in prototype data mining in predicting students' abilities, namely:

Table III. System Testing Points

No	Variabel	Sub Variabel	Indikator Pengukuran	Butir Uji
1.	Functionality	Suitability	Kesesuaian sistem dengan kebutuhan	1
		Accuracy	Keakuratan informasi yang dihasilkan oleh sistem	2
		Security	Keamanan data dan pengguna	3
		Interoperability	Integritas & akses sistem dengan perbedaan teknologi yang digunakan	4
		Compliance	Kesesuaian sistem dengan peraturan yang berlaku	5
2.	Reliability	Maturity	Rendahnya tingkat kesalahan dalam sistem	6
		Fault Tolerance	Kemampuan untuk berfungsi seperti biasa	7
		Recoverability	Kemampuan sistem untuk mengatasi kesalahan yang terjadi	8
3.	Usability	Understandibility	Kemudahan sistem untuk dipahami	9
		Learnability	Kemudahan sistem untuk dipelajari	10
		Operability	Kemudahan sistem untuk dioperasikan	11
		Attractiveness	Kamampuan sistem untuk menarik user	12
4.	Efficiency	Time Behavior	Kecepatan respon dan waktu pengolahan	13
		Resource Behavior	Kesesuaian penggunaan sumber daya	14

2. Characteristics and Rating Scale

There are 5 (five) criteria in the final results of testing prototype data mining to predict students' abilities, the scale of scores is the weight of the percentage value of the actual

score of each assessment characteristic. The following rating score scale:

Table IV. Score Scale

Number of Scores	Criteria
20% - 36%	Not Good
36.01% - 52%	Poor
52.01% - 68%	Sufficient
68.01% - 84%	Good
84.01% - 100%	Very Good

The score criterion is the weighting of the percentage value of the actual score of each characteristic.

$$\%\ skor\ aktual = \frac{Skor\ Aktual\ Karakteristik}{Skor\ Ideal}\ x\ 100\ \%$$

D. Test result

Based on the results of the calculation of the results of the questionnaire given by the user, the conclusion is that the prototype to predict students' ability to compete in the work world, as follows:

Table V. System test results by user

No	Variabel	Value	Average
1	Functionality	76%	78,1%
2	Reability	78,3%	
3	Usability	83,1%	
4	Efficiency	75%	

From the table of system test results by the user above shows that the results of the system test by the user for 4 (four) characteristics, namely Functionality, Reability, Usability, and Efficiency. The average score is 78.1%. Based on table V shows that the prototype is good value.

E. Research Implications

In research using data mining, not all cases can be solved by one method. Therefore the method chosen for one case may not solve the problem in another case. Therefore for each different case it is necessary to do research again on other data mining methods.

From the research results obtained using the Naive Bayes Method, each has a different level of accuracy in each dataset. But on average the accuracy and AUC levels in the Naive Bayes Method are quite high because they are above 90%. Implications for these findings cover the aspects of the system, managerial aspects and further research aspects.

1. System Aspects

   To implement the results of this study requires a good system support, so interested parties can use the results of this study to predict students' ability to compete in the world of work. Therefore we need adequate facilities and infrastructure consisting of hardware, software and other infrastructure in order to provide the best results, hardware and software specifications that can be used in this study have the minimum specifications as shown in the following table:

   Table VI. Components of system requirements

Component requirement	Spesification
Processor	Corei5
Memory	4GB
Operating System	Windows7/8/10
Hardisk	500GB
Monitor	14Inchi

2. Managerial Aspects

   From the results of research evaluations, the Naive Bayes Method shows good accuracy in classifying data so that this method can be used as a solution to predict students in competing in the world of work. That way, the school is able to know the pattern of students who are able to compete and are unable to compete in the world of work.

   By predicting students who are able and unable to compete, schools can evaluate learning and teaching, and improve student competencies according to their respective vocational. So that in the future every graduate can compete in the

world of work and the high unemployment rate can be overcome.

IV. CONCLUSIONS

Based on the discussion of the results of the research discussed in the previous chapter, this research concludes as follows:

1. Data mining models to form a classification of students' abilities using the Naive Bayes Method and evaluated with AUC (Area Under Curve) with ROC produces 96.07% accuracy with 0.9621 precision and 0.8805 recall.
2. Based on the classification model that has been designed, the processing of data mining uses prototype as a reference in the process of calculating the prediction of students' abilities based on the attributes that have been determined.
3. Testing prototype data mining using a user acceptance test can be well received by users with an average percentage value of 78.1%.

ACKNOWLEDGMENT

Thank you to Allah SWT who always marries mercy and love for his servants, and thanks to all colleagues who provide support and enthusiasm to finish writing this journal article, and for us MuhamadJuwayni andBambangKrismalela as authors of this journal can be useful according to our individual needs.

REFERENCES

[1] F. Alfiah, B. W. Pandhito, A. T. Sunarni, D. Muharam, and R. Matusin, “Data Mining Systems to Determine Sales Trends and Quantity Forecast Using Association Rule and CRISP-DM Method Abstract :,” vol. 4, no. 1, pp. 186–192, 2018.

[2] Amir Hamzah (2014), "Sentiment Analysis untukMemanfaatkanSaran KuesionerdalamEvaluasiPembelajarandenganMenggunakan Naive Bayes Classifier (NBC)", Prosiding Seminar NasionalAplikasi Sains &Teknologi (SNAST).

[3] ArdianaDkk (2010),"Kompetensi SDM UKM danPengaruhnyaTerhadapKinerja UKM Surabaya",JurnalManajemendanKewirausahaan. Vol.12 No.1. Maret 2010.

[4] Aria MustofaHidayatdanModammadSyafrullah (2017),"Metode Naive Bayes dalamAnalisisSentimenuntukKlasifikasiPadaLayanan Internet PT. XYZ", Jurnal TELEMATIKA MKOM Vol.9 No.2 Juli 2017.

[5] AriefJananto (2013) “Metode Naive Bayes untukMencariPerkiraanWaktuStduiMahasiswa”, JurnalTeknologiInformasi DINAMIKA Volume 18, No.1 hal 9-16.

[6] David HartantoKamagidanSengHansun (2014),"Implementasi Data Mining denganMetode C4.5 untukMemprediksi Tingkat KelulusanMahasiswa", ULTIMATICS, Vol.VI No.1.

[7] EgiBadarSambanidanFitriNuraeni (2017), "PenerapanMetode C4.5 UntukKlasifikasiPolaPenjurusan di SekolahMenengahKejuruan (SMK) Kota Tasikmalaya", CSRID Journal, Vol.9 No.3.Doi: 10.22303/csrid.9.3.2017.143-151.

[8] EkaSabnadanMuhardi (2016). "Penerapan Data Mining UntukMemprediksiPrestasiAkademikMahasiswaBerdasarkanDosen, Motivasi, Kedisiplinan, Ekonomi, danHasilBelajar", JurnalCoreIT, Vol.2,No.2.

[9] Hera WasiatidanDwiWijayanti (2014), "SistemPendukungKeputusanPenentuanKelayakanCalonTenagaKerja Indonesia MenggunakanMetode Naïve Bayes", Indonesian Journal on Networking and Security, Vol.3 No.2.

