UNIKOM

UNIVERSITAS KOMPUTER INDONESIA

Book of Abstract

INCITEST

International Conference on Informatics Engineering, Science & Technology

May 9th , 2018

Bandung - Indonesia

Published by :

IOP Publishing
Conference Series

Indexed by :

Scopus

IET Inspec

Compendex

INCITEST
International Conference on Informatics
Engineering, Science & Technology

BOOK OF ABSTRACTS

INCITEST

International Conference on Informatics,
Engineering, Science and Technology

Challenges of Science and Technology in the 21st Century

Universitas Komputer Indonesia

9 May 2018
BANDUNG - INDONESIA

Published by :

IOP Publishing
Conference Series
Materials Science and Engineering

Indexed by :

Scopus IET Inspec Ei Compendex

BOOK OF ABSTRACTS

INCITEST

International Conference on Informatics, Engineering, Science and Technology

Theme :
Challenges of Science and Technology in the 21st Century

Date : 9 May 2018

Venues : Grand Tjokro Hotel
Bandung - Indonesia

Organizer :
UNIVERSITAS KOMPUTER INDONESIA

Publication :
IOP Indexed by Scopus

INCITEST 2018

INTERNATIONAL CONFERENCE ON INFORMATICS, ENGINEERING, SCIENCES AND TECHNOLOGY

LIST OF COMMITEE

Advisory Board and Scientific Committee:

Prof. Dr. Ir. H. Denny Kurniadie, M.Sc – Universitas Komputer Indonesia
Prof. Rongtau Hou – Nanjing University Of Information Science and Technology, China
Assoc Prof M. Roil Bilad – Universiti Teknologi Petronas, Malaysia
Assoc Prof. Zulfan Adi – Universiti Teknologi Petronas, Malaysia
Dr. Ade Gafar Abdullah – Universitas Pendidikan Indonesia
Dr. Eng. Asep Bayu Dani Nandiyanto – Universitas Pendidikan Indonesia
Syeilendra Pramuditya, PhD – Institut Teknologi Bandung, Indonesia
Dr. Ing. Ana Hadiana M.Eng.Sc. – Lembaga Ilmu Pengetahuan Indonesia
Dr. Eng. Farid Triawan – Tokyo Institute of Technology, Japan
Dr. Supeno Mardi Susiki Nugroho, ST.,M.T – Institut Teknologi Surabaya
Dr. Yeffrie Handoko Putra – Universitas Komputer Indonesia
Dr. Yuzrila Y Kerloza – Universitas Komputer Indonesia
Dr. Andi Harapan, M.T – Universitas Komputer Indonesia
Dr. Dhini Dewiyanti Tantarto, M.T – Universitas Komputer Indonesia
Dr. Henny, ST.,M.T – Universitas Komputer Indonesia
Dr. Salmon Priadji Martana, ST.,M.T – Universitas Komputer Indonesia
Dr. Y. Djoko Setiarto, ST.,M.T – Universitas Komputer Indonesia
Irfan Dwiguna Sumitra, M.Kom., Ph.D – Universitas Komputer Indonesia
Dr. Yackob Astor, S.T., M.T – Politeknik Negeri Bandung
Dr. Rer. Nat. I Gusti Ngurah Agung Suryaputra, S.T., M.Sc. – Universitas Pendidikan Ganesha
Dr. Eng. Suranto – Universitas Pembangunan Nasional Veteran Yogyakarta
Prof. Dr. Muhammad Ali Ramdhani, M.T – UIN Sunan Gunung Djati Bandung
Sriadhi, M.Pd., M.Kom., Ph.D – Universitas Negeri Medan
Dr. Astri Rinanti, MT – Universitas Trisakti, Jakarta
Dr. Juniastel Rajagukguk, M.Si – Universitas Negeri Medan
Dr. Lilik Anifah, M.T – Universitas Negeri Surabaya
Dr. Rino A Nugroho – Universitas Sebelas Maret
Dr. Ir. Rudy Laksmono, M.T – Universitas Pertahanan Sentul Bogor

Organizing Commitee:

Dr. Lia Warlina.
Dr. Poni Sukaesih Kurniati, S.IP., M.Si.
Bobi Kurniawan, S.T., M.Kom
Senny Luckyardi, S.P

INCITEST 2018
INTERNATIONAL CONFERENCE ON INFORMATICS, ENGINEERING, SCIENCES AND TECHNOLOGY

[ABS-145] Prediction Student Eligibility In Vocation School With Naïve-Byes Decision Algorithm

L. Melian* and A.Nursikuwagus
Department of Information System, Faculty of Technic and Computer Science
Indonesia Computer University
*lusi.melian@email.unikom.ac.id

Abstract

Prediction is one area in data mining that discuss about classification. One method of algorithm in prediction is Naïve-Byes Decisions Algorithm. This research is proposed about decision in eligibility for accepting and rejecting student in vocational school. Variable will be counted by Naive-Byes Decisions algorithm that content of final exam, report, psychology, interview, and competency. Execution for the first step is transforming each variable into ranging values. Step forward is continued to calculate possibility value for every variable. In 270 student, the research has been resulted the validation test is 199 same as actual condition. Precision is 96,1%, recall is 99,3%.. The result is meaning that the algorithm has ability to get the decision in accepting manner with 74,87% accuracy. The value gives the reasoning that has not all submitting students are accepted by the school. Meanwhile, the algorithm has giving fair process for accepting the candidate. The impact of this research is the school can predict how many students is accepted and rejected. So, school will prepare for how many rooms should be available.

Topic: 5. Sciences

Author details

Print Email

Nursikuwagus, Agus

View potential author matches

http://orcid.org/0000-0001-8435-7522

Affiliation(s):

Universitas Komputer Indonesia, Bandung, Indonesia View more

Subject area: Engineering Materials Science Computer Science

Profile actions

Edit author profile

Connect to ORCID

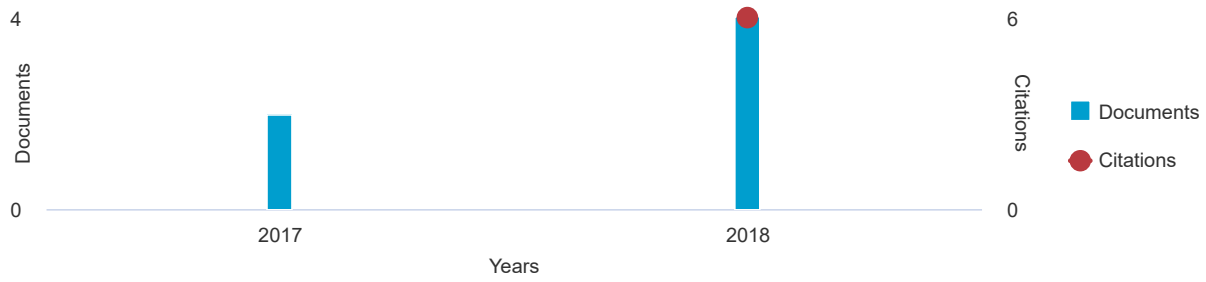
Alerts

Set citation alert

Set document alert

Documents by author	Total citations	*h*-index:
6	6 by 6 documents	1

Document and citation trends:



6 Documents | Cited by 6 documents | 10 co-authors | Topics

Preview users can view an author's latest 10 documents. View 73 references >

Set document alert

Document title	Authors	Year	Source	Cited by
Prediction Student Eligibility in Vocation School with Naïve-Byes Decision Algorithm Open Access View abstract ⌄ Related documents	MeLian, L., Nursikuwagus, A.	2018	IOP Conference Series: Materials Science and Engineering	0
Computational model of student competency analysis in fuzzy topsis method Open Access View abstract ⌄ Related documents	Nursikuwagus, A., MeLian, L., Permatasari, D.	2018	IOP Conference Series: Materials Science and Engineering	0
Rank computation model for distribution product in Fuzzy Multiple Attribute Decision Making View abstract ⌄ Related documents	Fenny, S.R., Nursikuwagus, A., Hartono, T.	2018	Telkomnika (Telecommunication Computing Electronics and Control)	1
A schedule optimization of ant colony optimization to arrange scheduling process at certainty variables Open Access View abstract ⌄ Related documents	Sidik, R., Fitriawati, M., Mauluddin, S., Nursikuwagus, A.	2018	International Journal of Advanced Computer Science and Applications	0
A mamdani fuzzy model to choose eligible student entry Open Access View abstract ⌄ Related documents	Nursikuwagus, A., Baswara, A.	2017	Telkomnika (Telecommunication Computing Electronics and Control)	5
Implementation ID3 algorithm to predict children achievement in response (case study children playgroup school) View abstract ⌄ Related documents	Agus, N., Lusi, M., Deasy, P.	2017	Journal of Engineering and Applied Sciences	0

Preview users can view an author's latest 10 documents.

⌃ Top of page

Prediction Student Eligibility in Vocation School with *Naïve-Byes* Decision Algorithm

IOP Conf. Series: Materials Science and Engineering **407** (2018) 012140 doi:10.1088/1757-899X/407/1/012140

Prediction Student Eligibility in Vocation School with *Naïve-Byes* Decision Algorithm

L Melian* and A Nursikuwagus

Department of Information System, Faculty of Technic and Computer Science, Universitas Komputer Indonesia

*lusi.melian@email.unikom.ac.id

Abstract. This research is proposed about decision in eligibility for accepting and rejecting student in vocational school. Prediction is one area in data mining that discuss about classification. One method of algorithm in prediction is Naïve-Byes Decisions Algorithm. Variable will be counted by Naive-Byes Decisions algorithm that content of final exam, report, psychology, interview, and competency. Execution for the first step is transforming each variable into ranging values. Step forward is continued to calculate possibility value for every variable. In 270 student, the research has been resulted the validation test is 199 same as actual condition. Precision is 96,1%, recall is 99,3%. The result is meaning that the algorithm has ability to get the decision in accepting manner with 74,87% accuracy. The value gives the reasoning that has not all submitting students are accepted by the school. Meanwhile, the algorithm has giving fair process for accepting the candidate. The impact of this research is the school can predict how many students is accepted and rejected. So, school will prepare for how many rooms should be available.

1. Introduction

Prediction is a model used for data sustainability using existing data. Prediction models are used for decision making on the information provided. Development of prediction model, in the realm of data mining, is very helpful to solve the problem of classification. Purpose of data mining is to find patterns against dataset presented. Performance student is a description of the ability of students' knowledge to be able to follow to advanced level. This information is often used to assist schools in accepting new students. *Humera Sahija* explains that use of data mining techniques helps to recognize student abilities in the subjects taken [1]. Semester assessment is used as a reference to provide treatment to students about the results of the value obtained each semester. *Mokhairi Makhtar* in his research answered the problem of applying *Naïve Bayes* to extract hidden information from students [2]. Result obtained that the Naïve Bayes algorithm is able to extract the hidden information with the accuracy of 73.4%. In the study of *Lalit Dole and Jayant Rajurkar*, explains that *Naïve Bayes* can predict students' success rates through a variety of ways [3].

Problem in this research is how to determine of new students by using the quantitative measure that students have. Prediction of graduation is very important for students as it will help the provision of facilities and infrastructure for new students. Determination of graduation that has been running still uses the calculation of the average value that is on the students. This calculation has not been able to extract hidden information on prospective students. This causes students who excel at certain objects

may not necessarily follow the learning in vocational schools. Recognition of information from students is very important, because it will review how far the students are successful in learning. Data mining is a widely used technique for extracting knowledge from existing dataset information. This approach is knows as machine learning that can be classified into supervise learning and unsupervised learning. Stages of data mining enrol consists of three parts of pre-processing data, Mining, and Post processing [1].

Purpose of this study is to obtain prediction of graduation from students who will follow vocational school by assessment on student performance such as final exam score, competency score, report score, physical test, interview test, and psychology test. Prediction will be done by applying a proven Naïve Bayes Algorithm in classifying the hidden information in a data set [1-8]. Impact of this research is candidate will predict about accepting and rejecting in vocational school.

2. Methodology

Pre-processing data is the beginning of the process of data mining. At this stage is preparing the dataset to be processed by using mining techniques. This stage also performs feature selection that is to reduce the dimensions (parameters) and remove the information that is irrelevant to improve the accuracy of the results [2].

Mining is the step of applying techniques or methods to get the pattern or knowledge of the dataset used. The application of mining techniques in this research is by using Naïve Bayes Algorithm [1-8]. Method approach in this research has used Naïve Bayes Algorithm [1-8]. We have three sequences process that is Pre-Processing, Mining, and Post Processing. Each step in mining process will be included Naïve Bayes model. Dataset is taken from survey in vocational school. Amount of dataset is 270 students. Pre-process that has occupied is cleaning dataset from violation and error value. In mining step, we have worked by applying Naïve Bayes model into dataset tend to get pattern of dataset. Final step in mining model is Post processing. Post processing is model to gain decision from evaluation dataset. Result from final step is model that can be used into next feature dataset. At Figure 1 is picturing steps of mining in Naïve Bayes model (Figure 1).

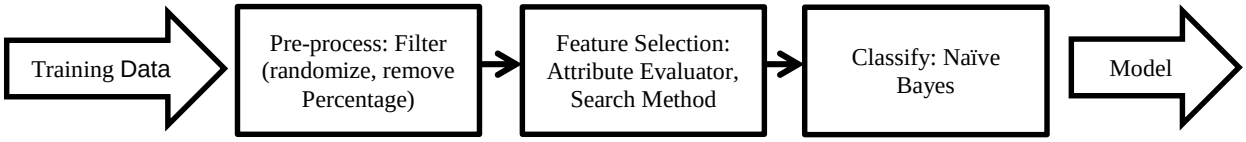


Figure 1. Naïve bayes model for student competence.

2.1. Naïve bayes algorithm

Naïve Byes was first introduced by Thomas Bayes. Bayes Classification provides practical learning, knowledge and data to be reviewed so that it can be combined. Data generation was done by using simulation test on WEKA 3.6.11.

The model used in Naïve Bayes consists of test data, model Naïve Bayes, and Results of prediction. The Bayes Theorem is as follows [7]:

$$P(h|D\ ) = P(D|h\ )\ P\ (h)[P(D)]^{-1} \quad (1)$$

Explain the formula is P (h) is a previous occurrence caused by h, P(D) training dataset in actual condition D, P(h|D) is event h caused by D, P(D|h) is D event caused by h. Calculation of events in the class is defined by dataset can be calculated by Naïve Bayes Classifier. The events in the Ci can be determined if and only if:

$$P(C_i \mid x) > P\left(C_j \mid x\right) \quad for\ \ 1 \le j \le m\ and\ j <> i \quad (2)$$

To determine the maximum value then used the rule:

$$P(C_i \mid x) = P(X \mid C_i)\, P(C_i)[P(X)]^{-1} \quad (3)$$

While determining the maximum likelihood then used the following rules:

$$P(X \mid C_i) = P(X \mid C_i)\, P(C_i) \quad (4)$$

2.2. Precision and recall

Precision is the closeness of the result to the actual event. This precision will give you the corresponding measure of the results given during the mining process. Precision is also a standardized measure in determining the document or dataset, whether it relates to the intended purpose or not. Precision can be calculated using the following formula [1-8]:

$$\boldsymbol{Precision} = \boldsymbol{TP}\,[\boldsymbol{TP} + \boldsymbol{FP}]^{-1} \quad (5)$$

Another calculation that gives good results is Recall. Recall is a measure of the significance of the existing document to the given result. Recall can be calculated using:

$$\boldsymbol{Recall} = \mathbf{TP}\,[\mathbf{TP} + \mathbf{FN}]^{-1} \quad (6)$$

Calculating *Precision* and *Recall* can be composed by confusion matrix. Content of confusion matrix is value from how many result is corrected with real condition. Matrix can be constructed and looked at (Table 1).

Table 1. Composition of confusion matrix.

	Condition : A	Not A
Test says accepted A	True positive (TP)	False positive (FP)
Test says accepted not A	False negative (FN)	True negative (TN)

3. Results and discussion

Objective in this research is discovered prediction model by Naïve Bayes model. Dataset that used in this research is almost 270. We have process not all data, only 199 dataset. The data is collected by survey. This survey is tailor from vocation school that can be accepted student.

In this subsection, we explain about results that have tailored. Refer to Figure 1, pre-processing step has done by evaluating dataset from error value and inconsistence. We have 270 dataset, and final dataset is no more error value. In pre-processing, we have normalize dataset that conversion from numeric value into perception in range like highest, high, moderate, and low. Conversion dataset can be seen at Table 2. (Table 2) views about rules to convert original dataset into normalize dataset. At Table 2, we are engineered transformation to convert original value into categorical data. We proposed some limitation classification of range. This activity is worked because algorithm just receives value in classification value.

Table 2. Rules of conversion in dataset.

No	Attribute	Description	Values
1	Final Exam	Final exam after student complete learning	highest >= 35 ; high >=30 and <35; moderate >=25 and < 30; low <= 25
2	Competency Test	Competency Assessment	highest >= 80 ; high >=70 and <80; moderate >=60 and < 70; low <= 60
3	Grade Report	Grade report	highest >= 80 ; high >=70 and <80; moderate >=60 and < 70; low <= 60
4	Body Test	Body endurance Assessment	highest >= 80 ; high >=70 and <80; moderate >=60 and < 70; low <= 60
5	Interview	Interview Assessment	highest >= 80 ; high >=70 and <80; moderate >=60 and < 70; low <= 60
6	Psychology Test	Psychology Assessment	highest >= 80 ; high >=70 and <80; moderate >=60 and < 70; low <= 60

At (Table 3) shows about original data and at (Table 4) is result table from conversion between original data into classification value.

Table 3. An example original dataset.

No	Name	Final Exam	Compe-tency Test	Report	Body test	Inter-view	Psy-chology Test	Ave.
1	DADANG WASISTO	30.74	83.33	79.84	80.00	85.00	50.00	68.15
2	TRI ANDI KUSUMAH	28.21	73.33	79.52	80.00	85.00	80.00	71.01
3	ANI SURYANI	31.66	83.33	79.16	70.00	65.00	80.00	68.19
4	WINNY FITRIANI	25.62	63.33	79.04	70.00	65.00	50.00	58.83
5	RIKA SILVIA	27.67	66.67	77.68	70.00	65.00	80.00	64.50

Table 4. An Example normalized dataset.

Name	X1	X2	X3	X4	X5	X6	Recommen-dation
Dadang Wasisto	High	High	High	Higher	Higher	Low	Accepted
Tri Andi Kusumah	High	High	High	Higher	Higher	Higher	Accepted
Ani Suryani	High	High	High	High	Higher	Higher	Accepted
Winny Fitriani	Moderate	High	High	High	Higher	Higher	Rejected
Rika Silvia	Moderate	Moderate	High	High	Higher	High	Accepted

Several data mining algorithms for classification problems have been made such algorithms as Decision Trees, K-Nearest Neighbour, and Naïve Bayes. To predict the Naïve Bayes Algorithm, mining process can be done if the data is no longer, no noise or null value. Steps taken on the naïve Bayes as wrote in subsection 2.1. Purpose of mining using Naïve Bayes is to predict students eligible or not to entry in vocation school. It has two classifications in target that is P(Accepted) and P(Rejected). P(accepted) is 0.7463 and P (rejected) is 0.2537. Through generating, the process is noticed such as True Positive (TF) and False Positive (FP). Result TF and FP can be seen at (Table 5-6).

Table 5. True positive and false positive classification in WEKA process.

	TP-Rate	FP-Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.993	0.12	0.961	0.993	0.977	0.999	Accepted
	0.88	0.007	0.978	0.880	0.926	0.999	Rejected
Weighted Avg.	0.965	0.092	0.965	0.965	0.964	0.999	

Table 6. Evaluation in data training set.

Evaluation	Calculation	Percentage
Correctly Classified Instances	192	96.4824 %
Incorrectly Classified Instances	7	3.5176 %
Kappa statistic	0.9033	
Mean absolute error	0.1098	
Root mean squared error	0.1826	
Relative absolute error	29.0772 %	
Root relative squared error	42.1073 %	
Total Number of Instances	199	

At Table 6, evaluation Naïve Bayes is worked by WEKA. WEKA is mining software that created by Waikato University and free license. We have 270 dataset and only 199 dataset that we used to generate the model. To validate the model, we should be separated dataset in two parts that is accepted and rejected manner. At Table 7, *Correctly Classified instance* is reach 192, and *incorrectly* only 7 classes. We are concluded result with confusion matrix at (Table 7) [1].

Table 7. Class results are presented in confusion matrix.

Classification	Accepted	Rejected
Accepted By Test	TF = 148	FP = 1
Rejected by test	FN = 6	TN = 44

At Table 5, precision dataset has been shown is 0.961 for Accepted and 0.978 for Rejected. This value can be stated that dataset have significant in model. In the other hand, calculated recall is got 0.993 for accepted and 0.880 for rejected. Calculating accuracy can be calculated in formula (5) as follow: $\boldsymbol{Accuracy} = \mathbf{TF} + \mathbf{FP}\ [Total\ Dataset]^{-\mathbf{1}} = (148 + 1)[199]^{-1} = 0.7487$. Accuracy can be written as closeness to actual data. Accuracy is 0.7487; it means we have closeness to actual data 74,87% .

Final result in process is depending on providing dataset. Naïve byes guarantee, classification that has gained, is have dependence among data. At Table 8 shows comparison among Naïve Byes, J48, and Decision (Table 8) [1-3, 7, 9].

Table 8. Comparison result among Naïve Byes, J48, and decision table.

Evaluation on Training Set	Naïve Bayes	J48	Decision Table
Correctly Classified Instances	96.48%	89.95%	87.94%
Incorrectly Classified Instances	3.52%	10.05%	12.06%
Kappa statistic	0.9033	0.7398	0.6246
Mean absolute error	0.1098	0.1402	0.2016
Root mean squared error	0.1826	0.2648	0.293
Relative absolute error	29.0772	37.14%	53.40%
Root relative squared error	42.11%	61.04%	67.55%
Accuracy	0.7487	0.7487	0.7487
Precision	0.961	0.945	0.865
Recall	0.993	0.919	0.993
Total Number of Instances	199	199	199

4. Conclusion

Prediction is one area of data mining to classify the desired target. In the dataset of 270, total of 199 datasets are expressed in accordance with the actual dataset. Calculation obtained using precision is 96.1%, recall is 99.3%, and accuracy is 74.87%. Results obtained can be said that dataset have an actual to dataset. Thus, the prediction with Naïve Bayes is 74.87% against the actual dataset. For comparison with other algorithms, it produces same accuracy like Naïve Bayes. Testing with the J48 and Decision Table has a result that approximates actual dataset with same accuracy is 74.87%. Impact from the research is prediction can be approximate how many student pass in exam.

References

[1] Amjad A S 2016 "Educational Data Mining & Student's Performance Prediction," *Int. J. Adv. Com. Sci. and App.* **7** (5) 212.
[2] Ekta J and Roopesh S 2017 *Data Mining: Document Classification using Naive Bayes Classifier Data Mining: Document Classification using Naive Bayes Classifier* **167** 97-115.
[3] Fadhilah A, Nur Hafieza I and Azwa A A 2015 "The Prediction of Students' Academic Performance Using Classification Data Mining Techniques," *J. App. Math. Sci.* **9** 6415.
[4] Humera S, Raniah Z, and Kavitha Z 2015 "Prediction of Students Performance in Semester Exams using a Naïve Bayes Classifier," *Int. J. of. Ino. Res. in. Sci. and. Eng. Tech*. **4** (1) 98-123.
[5] Lalit D and Jayant R 2014 "A Decision Support System for Predicting Student Performance," *Int. J. of. Ino. Res in Sci. and. Eng. Tech.* **2** (12) 7232.
[6] Masud K and Rashedur M R 2013 "Decision Tree and Naïve Bayes for classification and Generation of Actionable Knowledge for Direct Marketing," *J. Sof. Eng. and. App* **6** 196.
[7] Mokhairi M, Hasnah N and Syadiah N W S 2005 *"Analysis on Students Performance Using Naïve Bayes Classifier" J. of. The. And. App. Inf. Tec.* **95** 16-26.
[8] P V Praveen S 2013 "A Comparative Study for Predicting Student's Academic Performance Using Bayesian Network Classifiers," *Iosr J. of. Eng.* **3** 2-37.
[9] Rajeswari R P, Kavitha J and Aradhana 2017 *Text Classification for Student Data Set using Naive Bayes Classifier and KNN Classifier* **43** 1-8.