International Journal of Engineering & Technology, 7 (4.15) (2018) 421-423

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Application of the Naïve Bayes Algorithm for Student Graduation Analysis

Erick Akhmad Fahmi Alfa'izy [1], Khairil Anam [1], Naidah Naing[2], Rosanita Tritias Utami[3], Nur Anim Jauhariyah[4], Ahmad Munib Syafa'at[4], Lely Ana Ferawati Ekaningsih4*, Mohammad Roesli[5], Yanna Ika Pratiwi[5], Yeni Ika Pratiwi[5]

[1] Department of Informatics Engineering, Universitas Maarif Hasyim Latif, Sidoarjo, Indonesia
[3]Universitas Muslim Indonesia
[4]STKIP PGRI Tulungagung Indonesia
[5]Institut Agama Islam Darussalam Blokagung
[6] Universitas Merdeka Surabaya
**Corresponding author E-mail: lafwens@gmail.com*

Abstract

Design an analysis system to find out graduation by comparing previous data and existing data to overcome errors in a college system. By taking data records that are already available to be processed using the naïve Bayes algorithm. This research was conducted at Universitas Maarif Hasyim Latif. In this case, the object of research is to analyze the data of students with naïve Bayes algorithms to find out their graduation. For sampling the data taken is the previous Faculty of Law Student data to be used as training data, to retrieve the entire data using data records that are already available in the Directorate of Information Systems. That the naïve Bayes algorithm can be used in the classification of data in the form of a string or textual. This is based on researchers' trials in taking examples of calculations that have been done before. To compare the results of the classification of graduation analysis using the naïve Bayes algorithm testing is done with a sample of data in the form of training data compared to data testing. From the calculations that have been made, the accuracy is 77.78%.

Keywords: *Naïve Bayes algorithm, Application for analysis, algorithm, graduation analysis.*

1. Introduction

Students are intellectuals who have high mobility and are a great asset that must be owned by a country. Young students are a very productive phase in moving and contributing to work. But the role of students at this time is inferior, because of the many lecture systems that are less supportive and not run optimally. The system at a college also sometimes experiences a lack of data which often affects student graduation. Because there are still system errors and still use conventional methods.

According to Alfa Saleh in his research, the naïve Bayes method was used to predict the electricity usage of each household, out of 60 data that were tested obtained 78.3333% accuracy for seven predictions, of which 60 data contained 47 household electricity usage data that were successfully classified with the right [1], [2]. Design an analysis system to find out a student's graduation with available data and apply a naïve bayes algorithm to compare the value of a student with a sample of data that has been prepared is one solution to overcome problems in a college system [3][4].

Based on the abundance of student data and data on the number of student graduations, hidden information can be known by processing student data so that it is useful for the university. Data processing students need to be done to find out relevant information in the form of new knowledge (knowledge discovery), for example, information about classifying student data based on profiles and academic data. The unique experience can help the university to carry out a classification of student graduation rates to determine a strategy to increase graduation in the following years [5]–[8].

2. Method

This research was conducted at Maarif Hasyim Latif University because there was no graduation analysis system in the information system on the campus; this research was conducted for four months. In this case, the object of research is to analyse the data of students with naïve bayes algorithms to find out their graduation. For sampling the data taken is the previous Faculty of Law Student data to be used as training data, to retrieve the entire data using data records that are already available in the Directorate of Information Systems [9]–[11].

3. Result and Discussion

The main problem raised from this research is how to build a system that can analyze web-based student data from inputting Student Main Number by the user to the system using the Naïve Bayes algorithm. The programming language of the system is built using the PHP language, the software for writing program code uses Sublime Text version 3.0 and for storing data using the XAMPP bundle as MySQL and as a local server.

As a first step of the system needs analysis process, the purpose of this research will be re-described as the previous chapter is to

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

make a web-based student graduation level analysis system using the Naïve Bayes algorithm. Then this system will be a software that can analyze data from user input (input in the form of Student Main Number) in the column that has been provided so that it is straightforward for users to carry out an analysis process.

In analyzing, this system compares the value of the data entered by the user with a date testing that has been set into the system using the Naïve Bayes algorithm. To find out how this algorithm works in the order, the author will explain in detail the process analysis. The stages that will be carried out to start the naïve Bayes algorithm in the graduation analysis system are as follows:

1. Input made by the user is in the form of Student Identification Number
2. Student ID numbers that have been inputted will process data mining on MySQL and will display related data
3. If the Student Identification Number inputted is declared correct, the results of the data mining will be processed using the naïve Bayes algorithm.

From the results of the performed tests using Rapid Miner training data that has been set into the system has an accuracy rate of 77.78%.

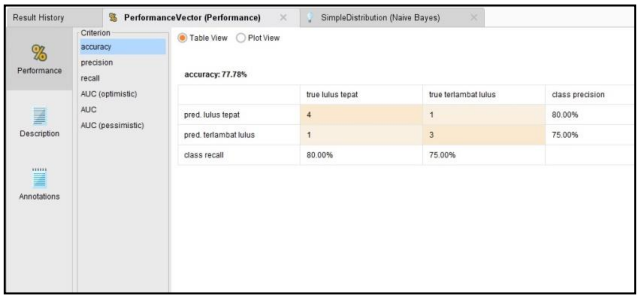


Fig. 1. Rapid Miner Accuracy Results

Comparison of naïve Bayes algorithm calculations with the system and using the same training data.

	status mahasiswa	jenis kelamin	status nilai	status tunggakan	analisa kelulusan
1	mahasiswa	perempuan	Tidak Ada Nilai D atau E	Ada tunggakan	lulus tepat
2	mahasiswa	perempuan	Tidak Ada Nilai D atau E	Ada tunggakan	lulus tepat
3	pekerja	laki-laki	Tidak Ada Nilai D atau E	Ada tunggakan	lulus tepat
4	pekerja	laki-laki	Ada Nilai D atau E	Tidak ada tunggakan	terlambat lulus
5	mahasiswa	laki-laki	Tidak Ada Nilai D atau E	Tidak ada tunggakan	lulus tepat
6	mahasiswa	perempuan	Tidak Ada Nilai D atau E	Ada tunggakan	lulus tepat
7	mahasiswa	perempuan	Tidak Ada Nilai D atau E	Tidak ada tunggakan	lulus tepat
8	pekerja	perempuan	Ada Nilai D atau E	Ada tunggakan	terlambat lulus
9	mahasiswa	laki-laki	Ada Nilai D atau E	Ada tunggakan	terlambat lulus
10	pekerja	perempuan	Ada Nilai D atau E	Ada tunggakan	terlambat lulus

Fig.2. Training Data

From the training data above, if you know the data that contains :

1	perempuan	mahasiswa	tidak ada nilai D atau E	Ada tunggakan	Status kelulusan??

Fig. 3. Data Unknown Graduation Status

From the data above the unknown status of the graduation attribute can be calculated as follows:

P (right pass) = 6/9 = 0.66
P (too late) = 5/9 = 0.55
P (students | pass right) = 4/6 = 0.66
P (student | graduated late) = 1/5 = 0.2
P (workers | pass right) = 1/6 = 0.166
P (worker | late in graduation) = 3/5 = 0.6
P (female | pass right) = 3/6 = 0.5
P (women | pass late) = 2/5 = 0.4
P (male | pass right) = 2/6 = 0.33
P (male | late in graduation) = 2/5 = 0.4
P (There is no D or E | pass right) = 5/6 = 0.833
P (There is no D or E value? Too late) = 1/5 = 0.2
P (there is a D or E | pass right) = 1/6 = 0.166
P (there is a value of D or E | late passing) = 4/5 = 0.8
P (no arrears | pass right) = 2/6 = 0.33
P (no arrears | late passing) = 1/5 = 0.2
P (there is an arrears | pass right) = 3/6 = 0.5
P (there is arrears | late in graduation) = 3/5 = 0.6

Posterior Passed right = P (passed right) x P (student | passed right) x P (worker | passed right) x P (female | passed right) x P (male | passed right) x P (There is no D value or E | pass right) x P (there is a D or E | pass right) x P (no arrears | pass right) x P (there are arrears | pass right)

= 0.66 x 0.66 x 0.166 x 0.5 x 0.33 x 0.833 x 0.166 x 0.33 x 0.5 = 0.000272218

Posterior Late graduation = P (late in graduation) x P (student | late in graduation) x P (worker | late in graduation) x P (female | late in graduation) x P (male | late in graduation) x P (No D value or E | late graduating) x P (there is a D or E value | late passing) x P (no arrears | late passing) x P (there are delinquency | late in graduation)

= 0.55 x 0.2 x 0.6 x 0.4 x 0.4 x 0.2 x 0.8 x 0.2 x 0, 6 = 0.000202752

Because P (passed right)> P (too late to graduate), the results of the calculation above indicate that the graduation status that is not yet known is the right pass. From the manual calculation above, let's compare it to the calculation by the system that has been made.

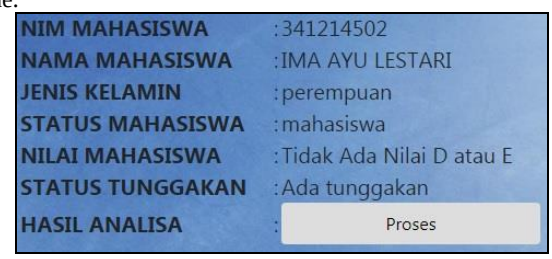


Fig. 4. Data that will be processed

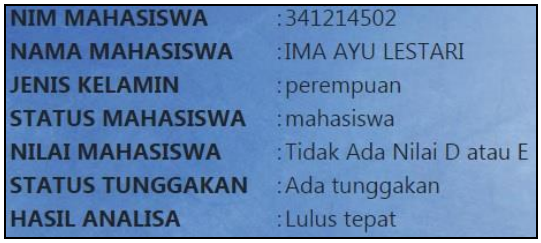


Fig. 5. System Process Results

Based on the results of the above trials it can be concluded that this system still has responses that are still not correct because training data is entered into the system. So the system keeps comparing the data inputted with training data that is prepared into the system.

4. Conclusion

In this study, it can be concluded that the naïve bayes algorithm can be used in the classification of data in the form of string or textual. This is based on researchers' trials in taking examples of calculations that have been done before. To compare the results of the classification of graduation analysis using the naïve Bayes algorithm testing is done with a sample of data in the form of training data compared to data testing. based on the results of the analysis in this study, obtained naïve bayes accuracy for 39 data of law students with the stage of comparing training data and testing data has an accuracy of 77.78%.

References

[1] A. Saleh, "Implementasi metode klasifikasi naive bayes dalam memprediksi besarnya penggunaan listrik rumah tangga," *Creat. Inf. Technol. J.*, vol. 2, no. 3, pp. 207–217, 2015.

[2] H. Saleh, D. Suryadi, and J. A. Dahlan, "Promoting students' mathematical problem-solving skills through 7e learning cycle and hypnoteaching model," in *Journal of Physics: Conference Series*, 2018, vol. 948, no. 1, p. 12037.

[3] M. B. N. Wajdi *et al.*, "Advancement of E-Book Through

Automation System in Department of Islamic Education (STAI) Miftahul Ula Nganjuk," *Int. J. Eng. Technol.*, vol. 7, no. 3.6, pp. 438–441, Jul. 2018.

[4] R. Rahim *et al.*, "Prototype File Transfer Protocol Application for LAN and Wi-Fi Communication," *Int. J. Eng. Technol.*, vol. 7, no. 2.13, pp. 345–347, 2018.

[5] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *J. EECCIS*, vol. 7, no. 1, pp. 59–64, 2013.

[6] Y. S. Nugroho, "Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro," *Dian Nuswantoro Fak. Ilmu Komput. Skripsi*, 2014.

[7] S. Andini, "Klasifikasi Dokument Teks Menggunakan Algoritma Naive Bayes dengan Bahasa Pemograman Java," *J. Teknol. Inf. Pendidik.*, vol. 6, no. 2, pp. 140–147, 2013.

[8] A. Jananto, "Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa," *Din. Teknol. Inf.*, vol. 18, no. 1, 2013.

[9] A. Y. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes," in *Advances in neural information processing systems*, 2002, pp. 841–848.

[10] D. D. Lewis, "Naive (Bayes) at forty: The independence assumption in information retrieval," in *European conference on machine learning*, 1998, pp. 4–15.

[11] A. McCallum and K. Nigam, "A comparison of event models for naive bayes text classification," in *AAAI-98 workshop on learning for text categorization*, 1998, vol. 752, no. 1, pp. 41–48.