

PAPER • OPEN ACCESS

Data Mining Technique as Majors Support System Management with Classification Approach

To cite this article: Basri *et al* 2019 *J. Phys.: Conf. Ser.* **1244** 012004

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Data Mining Technique as Majors Support System Management with Classification Approach

Basri*

Informatics Study Program, Universitas Al Asyariah Mandar, West Sulawesi, Indonesia

Muhammad Siddiq, Rosmawati Tamin

Information System Study Program, Universitas Al Asyariah Mandar, West Sulawesi, Indonesia

Sulihin Azis

Education Faculty, Universitas Al Asyariah Mandar, West Sulawesi, Indonesia

*Corresponding author: basri@unasman.ac.id

Abstract. Majors systems in an educational institution both at the primary level and higher education are the main thing. The academic potential of students becomes one of the main factors in supporting the success of learning. Even in an admission system, entrance test cannot be separated from the potential academic test. The academic potential test is only used as a standard of acceptance, and very rarely even never used to be information that can classify students into a specialization. Analysis of potential academic data with various variables will be very much so that required Data Mining techniques to mining of the larger data. The purpose of this research is to implement of Data Mining technique with classification approach to analyze academic potency in the major supports system management. The classification technique is applied as the object of research and measured its performance to obtain information according to the needs and objectives of the study. The results showed that the implementation of classification techniques could provide information that indicates the academic potential in major support system management, so this research can be a reference to create a system of potential academic search in general and can be implemented as an applicable tool in major support system management.

Keywords: data mining, the classification approach, majors support system, management

1. Introduction

Big Data in an educational institution is inevitable and has become a trend at this time. The development of information technology becomes the trigger to collect data and accommodate the big data into storage database. One of the challenges that exist today is how to utilize the Big Data in an educational institution as one of the tools in policy making (Abdul Fattah Mashat et al., 2013). Utilization of the data would require techniques to process databases with the concept of mining information or in other terms Knowledge Discovery in Databases [1].



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

In the academic world of higher education, for example, a particular set of academic policy information was abstracted from various data derived from integrated academic applications, thus demonstrating the much-needed pattern for the Executive Information System in adopting management policies [2]. One of the policies that can utilize data mining techniques is System Majors in the process of admission of new students with data classification approach. As conducted by Kaur, P. et al., that data classification technique was a promising thing in the future, especially for education [3]. Data classification is intended to classify students into several categories so that policies issued by the Executive Information System can be more targeted.

Problems that arise after the observation, the admission system collaboration in Indonesia, generally it was placing the graduate student on specific majors consisting of several options [4] so that by the system, prospective students are forced to create a sequence of choices based on alternative possibilities there and pushed in one of them. Whereas the academic potential of students can be a decision-making material and it has become one of the success factors if compared to the many admission test results [5]. In the outline of this research utilizes Decision Support System technique and data mining classification to process data on an admission system of students in a sample of the educational institution. Analysis of academic potential was done through the history of learning achievement in the previous level so can be reached standardized management system model. Also, this study made a recommendation of design management admission system based on system requirement analysis so can be adjusting to the major's system.

2. Literature Review

2.1 Educational Data Mining (EDM)

Many applications of Data Mining (DM) are utilized in various fields such as marketing, industry, accounting and finance, telecommunication and much more [6]. Likewise in the education institution, DM was used to explore the added value of knowledge that has not been known manually from a collection of education data, where activities include collection, use of data, historical to find regularities, patterns or relationships in large datasets [7]. In educational management, DM was widely used in analyzing data to generate useful policy recommendations for the development of educational institutions, including to analyze the subject's needs with the curriculum [8]. Also, in predicting the performance of students in the form of cluster mapping [9][3], and making it useful for institutions to issue policies that support the improvement of student achievement [10]. In the admission system, regarding majors were not separated from utilizing DM technology, so the decision of educational institutions in classifying the results of graduation would adjust the potential of students [11].

2.2 Majors Support System Management

The Admission system is essential for an educational institution to evaluate prospective students and recommend majors according to their potential and interests. Analysis of the academic history of students in the previous level was essential data compared with the result of the entrance test [5]. Research that evaluates the form of admission tests as by Anderson, K., who found that the results of tests by students cannot accurately measure the quality of prospective participants from where they came from [12]. The use of intelligent systems that can analyze data and look for academic history correlation was a workable solution. The pattern of data relationships based on academic experience data could be an interesting reference to provide the best recommendations in the admission system [13]. The use of Data Mining as an admission system was the best alternative and recommended to be applied [14].

Research for the recommendation of majors by implementing DM classification technique had been done by Swastina, which recommends Decision Tree C4.5 Algorithm with 93.31% accuracy as majors system management model that can adjust the history of academic ability, personality and talent prospective students [15]. Meanwhile, previous research using DM technique in the major's system obtained accuracy in the training process of 93.75% and on the testing process of 80.95% by

using the Naïve Bayes Classification technique [16]. Although it has not reached the perfect result, the system could be done well enough to analyzed the academic history of the students, so it could be a model in applying scheme systems.

2.3 *K-Nearest Neighbors Classification (KNN)*

KNN is an algorithm that classifies objects based on attributes and training data [17]. This algorithm uses the object adjacency system as a predictive value. Classification modeling with KNN algorithm will find some k objects or training points that are closest to the query point that is the target of classification. This method uses the concept of supervised learning so that it will find new patterns in the data set that already exists in the database [18]. Classification is done by measuring the proximity distance of each object so that in this model the Euclidean Distance formula is used [19].

2.4 *Classification Tree (CT)*

Classification Tree is a classification method that is widely implemented in data mining techniques. This method classified objects or records in a data set to produce a decision tree and converted it into a classification rule [20]. This technique consists of a collection of decision nodes, which branch down from the root node to the leaf node. To be able to run this classification model requires target variables, variable training data and target attribute classes must be diacritical. One of the advantages obtained from this method is a decision-making process that was previously very complex to be simpler. Classification trees with various variants in decision-making systems have been widely implemented, including the development of a prediction system model for dengue severity characterization based on early clinical and laboratory indicators [21]. The development of the Classification Tree model as a classification system with the right model combination will get the correct prediction accuracy [22].

3. Methodology

The method used in this research is an analytic experiment involving open interview questionnaire on respondents who come from psychology and workers in the field of counseling education and analyzing data using computation approach. The flow used in the study as shown in Figure 1.

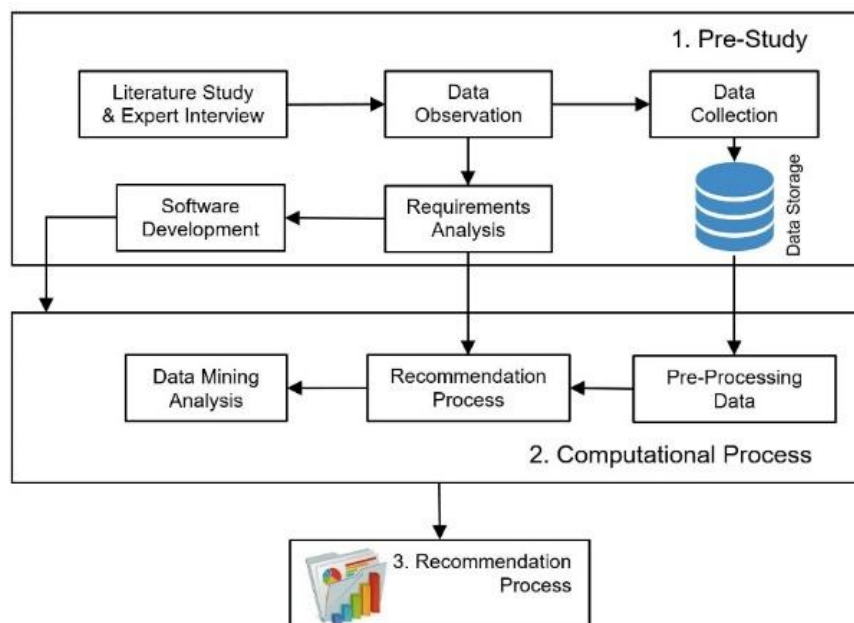


Figure 1. Research Flowchart.

3.1 Pre-Study

The first step taken in this research was studying literature to collect materials, references and state of the art from research conducted. Also, interviews with psychologists and counseling counselors to find out of the theory of specialization in the ideal scoring system.

Furthermore, the observation of data to be adjusted to the requirement analysis system that will be developed, so that the system can be a reference in the management of majors in an educational institution. The final stage of the pre-study was application development process where using the waterfall model [23]. The data to be processed in this research are 352 data derived from the data processing of majors with the conventional calculating system [4]. The data is selected into Science Study, Social Study and Language Study in High School Model.

3.2 Computational Process

This stage was done after pre-processing data to adjust the requirement of applications. After the data was by the application compatibility, the next stage was the recommendation process. The stage of the recommendation process was done by using the technique Decision Support System (DSS). It was common and easy to do and can adjust the management needs of majors regarding weighting. The method chosen in this step is the Simple Additive Weighting (SAW) method because it can select the decision according to the criterion data [24] and by the previous interview result. The weighting used as presented in Table 1.

Table 1. Criteria weighting

Alternatif	Value of Range Criteria		
	50 - 69	70 - 79	>= 80
Math			
Science Study	0,25	0,25	0,5
Social Study	0,25	0,5	1
Language Study	0,25	0,25	0,5
English Language			
Science Study	0,5	0,5	0,75
Social Study	0,25	0,25	0,5
Language Study	0,25	0,5	0,75
Indonesia Language			
Science Study	0,5	0,75	0,75
Social Study	0,25	0,75	0,75
Language Study	0,25	0,75	0,75
Natural Science			
Science Study	0,75	0,25	0,5
Social Study	0,75	0,25	0,75
Language Study	0,75	0,5	0,75

Recommendations output of DSS applications will be analyzed using data mining techniques as a technique that can measure and validate the recommendations. The method used to validate the SAW DSS results by using the Classification model such as K-Nearest Neighbors Classification (KNN) and Classification Tree (CT) with XLMiner Application [25]. The division of data processing by DM technique was divided into the Training process by 176 records (50%), Validation by 106 records (30%) and on Test Data by 70 records (20%) of 352 data overall data sampled in this study.

3.3 Report & Recommendation

At this stage, the analysis of DM processing results show in the graphical capture, so can be analyzed based on the difference of accuracy based on the Confusion Matrix with ROC technique [26]. Then

compared to the previous conventional system. At this stage will also be presented the recommendation process model of majors based on the analysis of system requirements.

4. Findings

4.1. Majors Decision System

As the results of observations and interviews obtained information that the development of the majors model of each educational institution was using a local model recommendation, depending on the needs of these institutions. In some of the developed models, in 2016 through collaboration admission website system (<https://siap-ppdb.com>) it was known that the selection process generally follows the following models:

4.1.1. *Selection based on the rating result of the total value of school examination,*

4.1.2. *The order of determination of selection results based on the number of values and order of choice of school,*

4.1.3. *If there is a similar value at the end of the quota of acceptance, then it was determined based on the academic value in sequence (Indonesia Language, English Language, Natural Science, and Mathematics)*

4.1.4. *If still the same, the system used the time comparison to start registering.*

In addition to the graduation mechanism, a selection rule of more than one choice is selected in priority order. So if students are deemed not to pass on the first option, it would be pass with the next primary option. Although, it is not a priority option. So this case will be very detrimental to a new student. In the model management system that follows the design of the result of system requirements analysis, it is created a DSS Application that will provide recommendation and classification of majors by academic potential based on the value of learning outcomes history. The application designing interfaces using VB.NET 2013 programming. Application Dashboard screenshot as Figure 2.



Figure 2. Dashboard majors decision support system

Application development designed consists of input data masters to enter attribute data into the database. Application with master data report used to exported data from the classification process and further analyzed. Applications that are designed certainly prioritize the principle of accessibility with the waterfall system development model with white box and black box testing techniques [27], so the system is free from programming logic errors and system functionality errors. DSS application designing was done by following the SAW technique as shown in the Context Diagram in Figure 3.

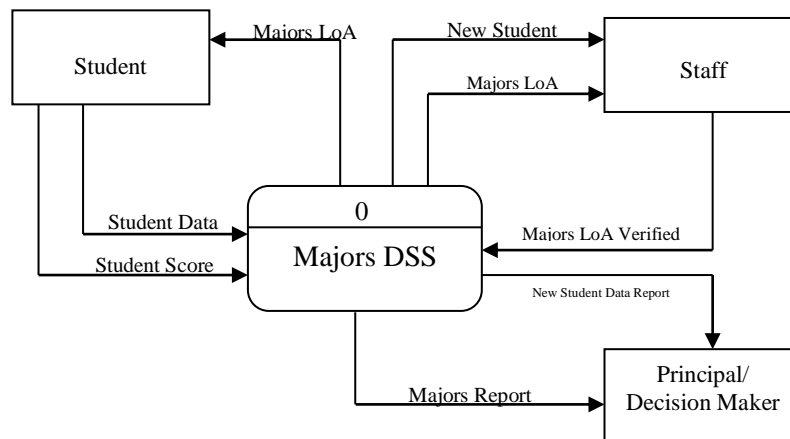


Figure 3. Context diagram majors decision support system

Figure 3 shows the process of data communication flow from the system to each system entity. In the system based on Figure 2, there are three types of entities, namely student as an object that has the main attribute. Also, the system entity also has staff as system users who will input and perform system analysis. Principal as the decision maker will monitor each classification result and be tasked with making decisions from the result of the system.

4.2. Data Mining Analysis

Data processing from Majors DSS is analyzed using the classification technique of KNN, and CT by comparing with the results of data recommendations on the conventional counting system on SIAP-PPDB system. The results of the classification as shown are shown in Table 2.

Table 2. Data mining analysis result

Metrics	SIAP-PPDB						Majors DSS					
	Training		Validation		Testing		Training		Validation		Testing	
	a	b	a	b	a	b	a	b	a	b	a	b
KNN	176	100	64	60,38	35	50	176	100	98	92,45	62	88,57
CT	169	96,02	54	50,94	36	51,42	176	100	104	98,11	68	97,14

a = correct number, b = percentage of correct number

The results of the analysis using the KNN and CT classification techniques shows that Majors DSS was better than the SIAP-PPDB framework mechanism. This can be seen from the accuracy that obtained in both classification techniques. The highest average value was at each stage of data processing both during Training, Validation, and Testing. Judging from the processing of data testing which consists of 70 records, the KNN method can provide accuracy of 88.57% in Majors DSS while in SIAP-PPDB only 50%, as well as on the CT method was able to provide 97.14% accuracy on Majors DSS while in SIAP-PPDB only 51.42%. The differences can be seen visually in the following graphic.

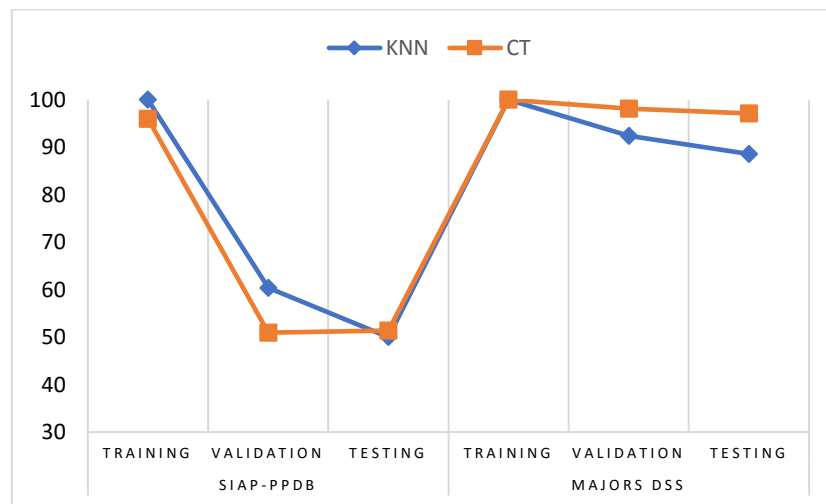


Figure 4. Graphics of majors decision accuracy

While the classification and recommendation each major in the data testing can be seen in table 3 below.

Table 3. Correct-error analysis per majors

Metrics	SIAP-PPDB		Majors DSS	
	Correct	Error	Correct	Error
Science Study				
KNN	21	10	27	4
CT	24	7	30	1
Social Study				
KNN	7	19	23	3
CT	6	20	26	0
Language Study				
KNN	7	6	12	1
CT	6	7	12	1

The analysis of the data each department shows that the conventional classification system shows a high degree of misguided error when viewed from the study history of the students. As an example of the KNN classification in Science Study, the departmental recommendation error was 10 out of 31 data records. Of the ten error data records 9 of which are recommended to the Social Study and 1 to the Language Study, although based on the method of data mining classification of students was more suitable to be in Science Study. That was quite significant when compared to Majors DSS that only make error 4 data record.

5. Conclusions

In general, based on observations and interviews, a good Majors System is a system that only accepts the primary interests of students without many alternative options, because otherwise, it will have an impact on the wrong recommendation. The analysis using Majors DSS developed suggests that analyze the potential outcomes of student' study at the previous level provide information on potential learning that will be appropriate in the selection of majors of interest, unlike with the admission Majors System who have been used. This is illustrated from data processing by Data Mining technique using K-Nearest Neighbors Classification (KNN) and Classification Tree (CT) method so that this research can be implemented as an applicable tool in majors support system management.

References

- [1] S. Charles, L. Arockiam, and V. A. Kumar, "Deriving association between learning behavior and programming skills," in *Communications in Computer and Information Science*, 2011, vol. 142 CCIS, pp. 96–103.
- [2] E. Prasetyo, L. Edi Nugroho, and M. Nurtiantara Aji, "Perancangan Data Warehouse Sistem Informasi Eksekutif untuk Data Akademik Program Studi," *J. Nas. Tek. Elektro dan Teknol. Inf. UGM*, vol. 1, no. 3, pp. 13–20, 2012.
- [3] P. Kaur, M. Singh, and G. S. Josan, "Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector," in *Procedia Computer Science*, 2015, vol. 57, pp. 500–508.
- [4] Siap-ppdb, "Aturan & Prosedur | PPDB SMA Jalur Reguler Kabupaten Polewali Mandar," 2016. .
- [5] R. Sawyer, "Usefulness of high school average and ACT scores in making college admission decisions.," *ACT Res. Rep. Ser.* 2010-2, 2010.
- [6] K. Bunkar, U. K. Singh, B. Pandya, and R. Bunkar, "Data mining: Prediction for performance improvement of graduate students using classification," *2012 Ninth Int. Conf. Wirel. Opt. Commun. Networks*, pp. 1–5, 2012.
- [7] H. Yulianton, "Data Mining untuk Dunia Bisnis Keputusan Informasi," *J. Teknol. Inf. Din.*, vol. XIII, no. 1, pp. 9–15, 2008.
- [8] V. Kumar and A. Chadha, "An empirical study of the applications of data mining techniques in higher education," ... *J. Adv. Comput. Sci.* ..., vol. 2, pp. 80–84, 2011.
- [9] A. P. Alfiani and F. A. Wulandari, "Mapping Student's Performance Based on Data Mining Approach (A Case Study)," *Agric. Agric. Sci. Procedia*, vol. 3, pp. 173–177, 2015.
- [10] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," in *Procedia Computer Science*, 2015, vol. 72, pp. 414–422.
- [11] N. Hermanto, "Sistem Pendukung Keputusan Menggunakan Metode Simple Additive Weighting (Saw) Untuk Menentukan Jurusan Pada Smk Bakti Purwokerto," *Semin. Nas. Teknol. Inf. Komun. Terap. 2012 (Semantik 2012)*, vol. 2012, no. Semantik, pp. 52–62, 2012.
- [12] K. Anderson, X. Gong, K. Hong, and X. Zhang, "Do selective high schools improve student achievement? Effects of exam schools in China," *China Econ. Rev.*, vol. 40, pp. 121–134, 2016.
- [13] R. C. Atkinson and S. Geiser, "Reflections on a Century of College Admissions Tests," *Educ. Res.*, vol. 38, no. 9, pp. 665–676, 2009.
- [14] M. I. Al-Twijri and A. Y. Noaman, "A New Data Mining Model Adopted for Higher Institutions," in *Procedia Computer Science*, 2015, vol. 65, pp. 836–844.
- [15] L. Swastina, "Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa," *Gema Aktual.*, vol. 2, no. 1, p. 6, 2013.
- [16] Basri, "ALGORITMA NAÏVE BAYES CLASSIFIER (NBC) SEBAGAI SISTEM PENUNJANG KEPUTUSAN PEMILIHAN KONSENTRASI," 2014, pp. 342–346.
- [17] X. Wu et al., "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.
- [18] T. Mitsa, *Temporal data mining*. Chapman and Hall/CRC, 2010.
- [19] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.
- [20] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- [21] K. Phakhounthong et al., "Predicting the severity of dengue fever in children on admission based on clinical features and laboratory indicators: application of classification tree analysis," *BMC Pediatr.*, vol. 18, no. 1, p. 109, 2018.
- [22] A. T. Kashani and A. S. Mohaymany, "Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models," *Saf. Sci.*, vol. 49, no. 10, pp. 1314–1320, 2011.

- [23] Y. Bassil, "A Simulation Model for the Waterfall Software Development Life Cycle," *Int. J. Eng. Technol.*, vol. 2, no. 5, pp. 2049–3444, 2012.
- [24] A. Reza Afshari and A. R. Afshari, "Selection of construction project manager by using Delphi and fuzzy linguistic decision making," *J. Intell. Fuzzy Syst.*, vol. 28, no. 6, pp. 2827–2838, 2015.
- [25] N. Kumar, "Data Mining for Business Intelligence--Concepts, Techniques, and Applications in Microsoft Office Excel® with XLMiner®," *J. Qual. Technol.*, vol. 44, no. 1, pp. 81–83, 2012.
- [26] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.
- [27] S. Madgunda, U. Suman, G. S. Praneeth, and R. Kasera, "Steps in requirement stage of waterfall model," *Int. J. Comput. Math. Sci.*, pp. 86–87, 2015.