EDUCATIONAL DATA MINING FOR MAPPING STUDENT ABILITY BASED ON SCHOOL LOCATION USING APRIORI METHOD CASE STUDY: SMK YPM SIDOARJO

[1*]M. MAHAPUTRA HIDAYAT, [2]RIFKI FAHRIAL ZAINAL, [3]ANDI ALFIAN EFENDI

Department of Informatics Engineering, Faculty of Engineering, Universitas Bhayangkara Surabaya

Jl. A Yani 114, Surabaya. Tel, 031-8285602

e-mail: [1]mahaputra@ubhara.ac.id, [2]rifki@ubhara.ac.id, [3]andialvian@gmail.com

[*]Corresponding author

ABSTRACT

Basically, Senior High School is programmed for those who continue to a higher level, while the provision of skills can be said to be non-existent. Vocational High Schools can produce quality graduates in terms of work skills, therefore currently many companies require graduates from Vocational Schools. A system is needed to map student abilities based on school location so that each school has an average picture of the ability of its students to further optimize performance in each school and find out the factors that influence the high and low grades of subjects in each school. The purpose of this study is to create an application to obtain useful information about mapping the value of subjects, especially English at YPM Vocational School in Sidoarjo with data mining techniques and Apriori method. From the results of system testing, it shows that there is associative rules that state that when the YPM 2 Taman Sidoarjo Vocational High School students in 2019 have an Indonesian National Examination score that exceed their KKM, then their English score also tends to be above their KKM score with a confidence of 97%. The oither result is that when YPM 2 Taman Sidoarjo Vocational High School students in 2019 have an Indonesian National Examination score that exceeds their KKM, then their Mathematics score also tends to be above the KKM score with a confidence of 94%..

Keywords: *Data Mining, Apriori, National Exam Score, School, SMK YPM Sidoarjo.*

1. INTRODUCTION

One of the educational establishments in charge of producing human resources with aptitudes, know-how, and competence is Vocational High School (SMK). This is done so that graduates can enhance their performance when entering the workforce. Government Regulation Number 29 of 1990 concerning Secondary Education provides a more detailed explanation of this type of education, stating that vocational secondary education is secondary education that places an emphasis on the development of students' competencies [1]. It is clear from the foregoing justification that vocational education is instruction meant to get pupils ready for the workforce. The uniqueness of vocational school education lies not only in the competitive learning experiences that prepare students for the business and industry (DU/DI) world, but also in the way that vocational schools are relevant to DU/DI in order to meet the school's objective of producing high-caliber graduates who can meet DU's demands. /IN. This is based on government policy that links and matches vocational school education with DU/DI, ensuring that both parties are aware of what DU/DI needs [2].

Industrial practice aims to provide the following benefits: (1) hands-on experience operating a production line; (2) comprehension of work attitudes and discipline through industrial work practices in production lines; (3) vocational competencies in line with the competency standards required by the industrial world; and (4) social competence, which includes collaborating with others to complete tasks and resolving conflicts at work.

Basically, senior high school is programmed for those who continue to a higher level, while the provision of skills (for senior high school) can be said to be non-existent. In contrast to the world of SMK, they are required to master skills and are expected to create their own jobs. Vocational schools can produce qualified graduates in terms of work skills; therefore, many companies currently need graduates from vocational schools. The Education Office

has suggested preferring SMK because it is more promising in the world of work [3], [4]. Based on the background that has been explained, a system is needed to map student abilities based on school location so that each school has an average picture of the ability of its students to further optimize performance in each school and find out the factors that influence the high and low grades of subjects in each school.

Numerous research has been conducted to map the student using different factors and techniques, such as the decision tree algorithm C.45 and the Naive Bayes classification algorithm. The findings demonstrate that data mining techniques are valuable not only for mapping out students according to demographic variables, level of activity, comprehension of the learning process, and so on, but also for predicting the performance level of student achievement, including GPA, test scores, length of study, and so forth [5].

The Apriori algorithm is well-known for its capacity to extract association rules from big datasets quickly and effectively. Through the examination of transactional data from SMK YPM Sidoarjo, the algorithm will detect commonly used itemsets and produce significant insights into well-liked student abilities. The school can use this information to inform its decisions about insight for student's skill control [6].

2. RESEARCH METHODS

The method used in designing this system is the Waterfall model. It is called Waterfall because the stages that are passed must wait for the completion of the previous stage and run systematically or sequentially. This model approach starts sequentially from the level of system requirements analysis, then goes to the design, coding, testing/verification and maintenance stages.

2.1 Data Mining

Data mining is mining or discovering new information by looking for certain patterns or rules from a very large amount of data [7]. Data mining is also referred to as a series of processes to explore added value in the form of knowledge that has so far been unknown manually from a data set. Data mining, often also referred to as knowledge discovery in database (KDD). KDD [8]is an activity that includes collecting, using historical data to find regularities, patterns or relationships in large data sets.

Data mining is the activity of finding interesting patterns from large amounts of data, data can be stored in databases, data warehouses, or other information stores. Data mining is related to other disciplines, such as database systems, data warehousing, statistics, machine learning, information retrieval, and high-level computing. In addition, data mining is supported by other sciences such as neural networks, pattern recognition, spatial data analysis, image databases, signal processing[9], [10]. Data mining is defined as the process of finding patterns in data. This process is automated or often semi-automatic. The pattern that is found must be meaningful and the pattern provides benefits, usually economic benefits. Large amounts of data are needed.

2.2 APRIORI

The apriori algorithm is one of the algorithms proposed by Agrawal and Srikant in 1994 which functions to determine frequent itemsets in boolean associations. This algorithm is tasked with monitoring the development of candidates from the itemset. Pseudo-Code Apriori Algorithm [11], [12]:

» Ck: Candidate itemset of size k;
» Lk : Frequent itemset of size k.
» L1 = {frequent items};
» for (k = 1; Lk !=0; k++) do begin
» Ck+1 = {candidates are built from
» for each transaction t loaded in the database do increase count of all candidates in Ck+1 contained in t
» Lk+1 = {candidates in Ck+1 with min_support}
» end
» return .k Lk;

2.3 System Flowchart

A flowchart is a part with certain symbols that describe the sequence of processes in detail with other processes in a system. The system flowchart is a chart that shows the workflow or what is being done in the system as a whole and explains the sequence of procedures in the system. In other words, this flowchart is a graphical description of the sequence of procedures that make up a system. The following is a flowchart with a system:
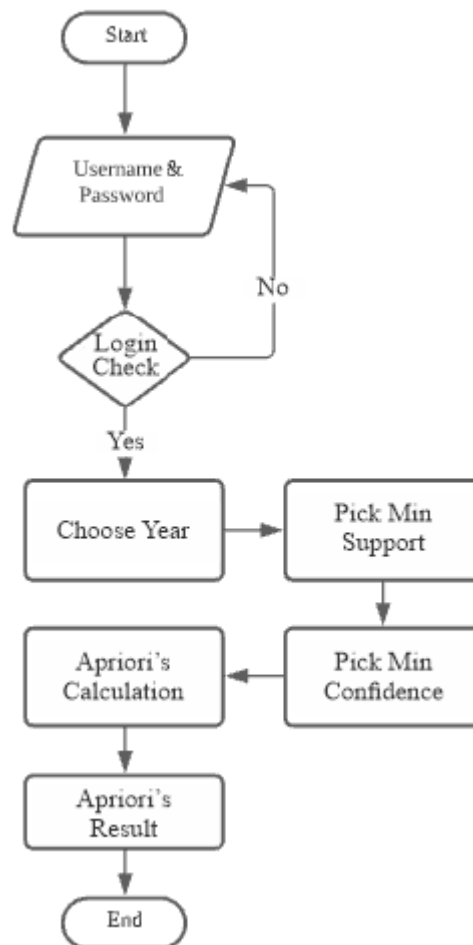
Figure 1. Flowchart of the Apriori System

From the flowchart above, it can be explained how to start the system by logging in by the user/admin, then selecting the year, determining the Min Support and Min Confidence values, resulting in an Apriori calculation.

2.4 Data Flow Diagram

A data flow diagram illustrates the movement of information within a system or process. It consists of data repositories, data inputs and outputs, and the several subprocesses the data goes through. Standardized nomenclature and symbols are used in the construction of DFDs to represent different entities and their relationships.

Data flow diagrams provide a visual representation of processes and systems that are difficult to explain in words alone. These diagrams can be used to plan out a new system's implementation or to map out an existing system and improve it. Seeing every component in detail makes it simple to spot inefficiencies and create the optimal system.
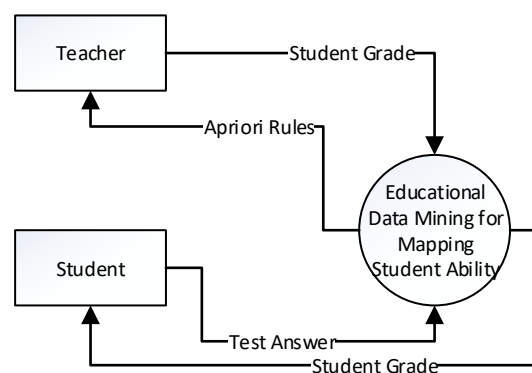

Figure 2. Data Flow Diagram on the System

Table 1. Student Conversion Table

No	Subject	KKM Value
1	Indonesian	≥ 60
2	English	≥ 55
3	Mathematic	≥ 50

3. RESULT AND DISCUSSION

Analysis is the stage of understanding an issue before taking an action or decision. Building a system needs to go through the analysis and design stages so that the system built can run as expected. At this stage an analysis of the current system (old system) and the system to be developed (new system) will be analyzed, analyzing system requirements and user needs.

3.1 Calculation of Apriori

Educational Data Mining System for Mapping Students' Ability Based on School Location (Case Study of YPM Sidoarjo Vocational High School) [12] This Apriori method is an application that will assist in viewing graphical data on the growth of National Exam scores in Indonesian, English and Mathematics in each YPM school in Sidoarjo according to the criteria existing by using the Apriori method applied in this application [13], [14]. In building this system required data as follows:

1. DKHUN data, namely:
    a) UN Pass Year
    b) Name of School.
    c) KKM standards for each subject.
    d) UN scores in Indonesian, English and Mathematics.

2. Student value conversion data

The student conversion value table is as follows:

Application of the Apriori method

* The first step is to determine the minimum support.
* Iteration 1: count items from Support (transactions that contain all items) by scanning the database for 1-itemset, after 1-itemset is obtained, from 1-itemset is it above the minimum Support, if it meets the minimum Support, 1-itemset it will be a high frequent pattern.
* Iteration 2: to get a 2-itemset, you have to do a combination of the previous k-itemset, then scan the database again to count the items that contain Support. itemset that meets the minimum Support will be selected as a high frequent pattern from the candidate.
* Set the k-itemset value of the Support that has met the minimum Support from the k-itemset.
* Do the process for the next iteration until there are no more k-itemsets that meet the minimum Support [15].

Minimum Support = 10%
Min Confidence = 60%

Table 2. Result data UN SMK YPM 2 Taman Sidoarjo

SMK YPM 2 TAMAN SIDOARJO				
No.	Name	BIN	BING	MAT
1	ACHMAD ZAKARIYA	72	44	37,5
2	AINUN NIKMATURROHMAH	68	46	27,5
3	AKHLAQUL ZAKARIA	76	38	37,5
4	AKHMAD NURIL DAFA	66	38	22,5
5	AKHSAN MAULANA	64	36	20
6	ALFITA FRY ANDINY	74	40	30
7	ALIF AKBAR ADHYATMA	78	70	30
8	ALIM FAKTA	50	30	22,5
9	AMIR SYAH	58	32	30
10	AZMIL ASIYAH NADHIFA	78	54	42,5
11	AZZAHRA NURIL FIRDAUS	78	68	32,5
12	BEBY GOLAM SYAH	78	36	42,5
13	CHELVIN WAHYU LUZYANDA	60	42	37,5

SMK YPM 2 TAMAN SIDOARJO				
No.	Name	BIN	BING	MAT
14	DAVID FEBRIADE	54	38	25
15	DEMAS OKTA HERDIAN S.	70	36	40
…				
…				
181	TRI INDAHSARI	78	72	30
182	WAHYU CATUR ANDINI	68	32	27,5
183	WIWIN YULIANTI	82	50	27,5
184	YASMIN BAYA'SUD	60	48	22,5

Table 3. Itemset 1 of 3 subjects

Itemset 1				
No	Item	Students	Support	Results
1	BIN	147	79%	Passed
2	BING	34	18%	Passed
3	MAT	18	9%	Failed

For the calculation of itemset 1

Number of students = 184 students
The number of students who scored Indonesian above the KKM = 147
Indonesian Language Value Support $= \frac{147}{184} = 0{,}798913043$
Indonesian Language Value Support = 79%
The number of students who scored English above the KKM = 34
English Value Support $= \frac{34}{184} = 0{,}184782609$
English Value Support = 18%
The number of students who get Mathematics scores above KKM = 18
Mathematical Value Support $= \frac{18}{184} = 0{,}097826087$
Support Math Value = 9%

For the calculation of itemset 2

Now count the number of students who have a combination of 2 Indonesian and English scores, which have scores above the specified KKM.

The number of students at SMK YPM 2 Taman Sidoarjo who have a combination score of Indonesian and English above the KKM = 33 students
Supports $= \frac{33}{184} = 0{,}179347826$
Support = 17%
And because the support value exceeds the minimum support value, the status is declared PASSED.

The number of students at SMK YPM 2 Taman Sidoarjo who have a combination score of Indonesian and Mathematics above the KKM = 17 students
Supports $= \frac{17}{184} = 0{,}092391304$
Support = 9%
And because the support value does not exceed the minimum support value, the status is declared FAILED.

The number of students at SMK YPM 2 Taman Sidoarjo who have a combination score of Indonesian and Mathematics above KKM = 7 students
Supports $= \frac{7}{184} = 0{,}038043478$
Support = 3%
And because the support value does not exceed the minimum support value, the status is declared FAILED.

Table 4. Itemset 2 from 2 subjects combination

No	Combination		Students	Support	Results
1	BIN	BING	33	17%	Passed
2	BIN	MAT	17	9%	Failed
3	BING	MAT	7	3%	Failed

Table 5. Itemset 3 from 3 subjects combination

No	Combination			Students	Support	Status
1	BIN	BING	MAT	7	4%	Failed

Table 6. Associative Rules

	Associative Rules			Support Itemset 2	Support Itemset 1	Confidence	Status
1	BIN	with	BING	33	34	97%	Passed
2	BIN	with	MAT	17	18	94%	Passed

For the calculation of itemset 3.
Now count the number of students who have a combination of the three grades for Indonesian, English and Mathematics subjects, which have scores above the specified KKM.

The number of students at SMK YPM 2 Taman Sidoarjo who have a combination score of Indonesian, English and Mathematics above the KKM = 7 students

$$\text{Supports} = \frac{7}{184} = 0{,}038043478$$

Support = 4%
And because the support value does not exceed the minimum support value, the status is declared FAILED.

3.2 Testing Result and Analysis

Because Itemset 3 does not meet the minimum support, it can be concluded that the association rules are as follows.

Associative Rules between Indonesian and English
Itemset Support 1 = 34
Support Itemset 2 = 33

$$\text{Confidence} = \frac{\text{Support Itemset 2}}{\text{Support Itemset 1}}$$

$$\text{Confidence} = \frac{33}{34} = 0{,}97$$

Confidence = 97%

Associative Rules between Indonesian and Mathematics
Itemset Support 1 = 18
Itemset Support 2 = 17

$$\text{Confidence} = \frac{\text{Support Itemset 2}}{\text{Support Itemset 1}}$$

$$\text{Confidence} = \frac{17}{18} = 0{,}94$$

Confidence = 94%

Both associative rules show that when a student in YPM 2 Taman Sidoarjo Vocational High School have an Indonesian National examination score that exceeds their KKM, then their English and Mathematics score also tends to be above their KKM score. These results will guide the YPM 2 Taman Sidoarjo Vocational High School to focus their teaching more in Indonesian field to gain increase score in the English and Mathematics score.
This research open an abundant of possibilities on another level of research. One of the research is to test several new test score to found another apriori rules that give another story. Another research can be executed in another school to make

sure if the existing apriori rules are applied to another school.

4. CONCLUSIONS

The conclusion is: (1) If YPM 2 Taman Sidoarjo Vocational High School students in 2019 have an Indonesian National Examination score that exceeds their KKM, then their English scores also tends to be above their KKM score with a confidence of 97%. (2) If YPM 2 Taman Sidoarjo Vocational High School students in 2019 have an Indonesian National Examination score that exceeds their KKM, then their Mathematics score also tends to be above the KKM score with a confidence of 94%.

REFERENCES

[1]   A. Susanto, D. Jatmoko, Y. Widiyono, J. Purwanto, R. Milyuantara, and U. M. Purworejo, (2023). "*Analisis Minat dan Hasil Belajar Siswa SMK dengan Pemanfaatan Media Pembelajaran Berbasis Aplikasi Lectora Inspire*," Jurnal Pendidikan, vol. 11, no. 02, pp.338-344.

[2]   P. H. Hasyim, and A.P. Abadi, (2023). "*Pengaruh Motivasi Belajar Terhadap Kemampuan Pemahaman Konsep Matematis Siswa SMK*," Journal on Education, vol. 06, no. 01, pp. 4877–4883.

[3]   Y. Zalukhu, H.S. Harefa, E. P.P. Zebua, and T. Harefa, (2023). "*An Analysis of Factors Influence Students' Speaking Performance at the Eleventh Grade of SMK Negeri 1 Gunungsitoli in 2022/2023*," Jurnal Edumaspul, vol. 7, no. 2, pp. 3757-3765.

[4]   F. Hansun, A. Mewengkang, O.E.S. Liando, (2023). "*Hubungan Motivasi Dengan Prestasi Belajar Siswa Kelas X SMK Negeri 1 Ratahan*,", Edutik, Vol. 3, No. 3, pp.300-307.

[5]   G. Simarmata and C. V. R. Sinaga, (2023). "*Motivasi dalam Pembelajaran Matematika Siswa di SMK Teladan Tanah Jawa*," PENDIPA Journal of Science Education, vol. 7, no. 2, pp. 299–303, doi: 10.33369/pendipa.7.2.299-303.

[6]   M.A. Ridla, F. Fajriyanto, and M. Marzuqi, (2023). "*Implementasi Algoritma Apriori untuk Menentukan Pola Transaksi Penjualan Berbasis Web*," JTIM : Jurnal Teknologi Informasi dan Multimedia, vol. 5, no. 3, pp. 196–207, doi: 10.35746/jtim.v5i3.399.

[7]   E.T. Naldy, and A. Andri, (2021). "*Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN*," Jurnal Nasional Ilmu Komputer, Vol. 2, No. 2, pp.89-101.

[8]   F. Alghifari and D. Juardi, (2021). "*Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes*," Jurnal Ilmiah Informatika, Vol. 9, No. 2, pp.75-81.

[9]   B. S. Purnomo and P. T. Prasetyaningrum, (2021). "*Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan di Kota Yogyakarta Menggunakan Metode K-Means*," Journal of Computer Science and Technology, Vol. 1, No. 1, pp.27-32.

[10]   S. Rohman, Z. Al Azis, and M. Hidayat, (2023). "*Analisis Pola Pembelian Produk Pada Toko Madu Mukti Dengan Algoritma Apriori*," Journal of Informatiom System and Computer, Vol. 3, No. 1, pp.39-60.

[11]   R. Kurniawan and R. Yusuf, (2023). "*Penerapan Metode Asosiasi Dengan Algoritma Apriori Untuk Mendukung Strategi Promosi*," *Edusaintek: Jurnal Pendidikan, Sains dan Teknologi*, vol. 10, no. 2, pp. 514–528, Feb. 2023, doi: 10.47668/edusaintek.v10i2.786.

[12]   A. Ardianti, N. Novriyenni, L.A.N. Kadim, (2023). "*Penerapan Algoritma Apriori Menentukan Korelasi Data Penjualan Pupuk (Studi Kasus : PT. Karunia Rotorindo Tani)*," *Jurnal Manajemen Informatika Jayakarta*, vol. 3, no.3, pp. 292–301, doi: 10.52362/jmijayakarta.v3i3.1196.

[13]   E. Prayitno and D. F. Sari, (2022). "*Implementasi Algoritma Apriori Untuk Pola Kombinasi Pembelian Barang*," Jurnal Cakrawala Ilmiah, vol. 2, no. 2, pp.691-696.

[14]   I. A. Ashari, A. Wirasto, D.N. Triwibowo, and P. Purwono, (2022). "*Implementasi Market Basket Analysis dengan Algoritma Apriori untuk Analisis Pendapatan Usaha Retail*," MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 21, no. 3, pp. 701–709, doi: 10.30812/matrik.v21i3.1439.

[15]   W. Dari, (2022). "*Analisis Metode Apriori Untuk Memprediksi Persediaan Barang Pada Warung*," INSOLOGI: Jurnal Sains dan Teknologi, vol. 1, no. 4, pp. 438–447, doi: 10.55123/insologi.v1i4.807.