

Student Identification Based on Patterns of Association For Student Misbehaviour Using Frequent Pattern Growth Algorithms

Erlinda¹, Dwipa Junika Putra², Mourend Devegi³

¹Teknik Informatika, Fakultas Teknik, Universitas Islam Kuantan Singingi, Indonesia

²Bisnis Digital, Fakultas Ekonomi Bisnis Dan Ilmu Sosial, Universitas Perintis Indonesia, Indonesia

³Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas PGSI, Indonesia

Article Info

Article history:

Received 05 28, 2023

Revised 06 11, 2023

Accepted 06 22, 2023

Keywords:

Data Mining

Student Violations

Association Rule

Fp-Growth

Frequent Itemset

ABSTRACT

Student infractions are incidents often committed by students who break the rules at school. This naturally worries school authorities and overwhelms them with student misbehavior. Student rule-breaking is a common problem that can interfere with a safe and orderly learning environment. The more students break the rules, the greater the impact on several aspects, including student achievement, discipline, suboptimal teaching and learning activities, and students' social lives outside of school. Identifying students who are prone to rule violations can help school officials implement more effective prevention programs. Data mining is a process of extracting information from large data sets to discover patterns and relationships hidden within them. This study aims to identify frequent student infractions using the Frequent Pattern Growth algorithm. The Frequent Pattern Growth (FP - growth) algorithm is used to generate frequent itemsets that are then used in the association rules process. The association rules process aims to find rules or relationships between violations based on the discovered Frequent Itemsets. This process is influenced by predefined minimum support and minimum confidence values. A Minimum Support value of 30% and a Minimum Confidence value of 50% are used to obtain rules with a sufficiently high confidence level. It is expected that the identification results from this study will provide a better understanding of the types of violations commonly committed by students in school. This information can be used by school officials to develop more effective prevention strategies and focus on.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Erlinda

Teknik Informatika, Fakultas Teknik

Kuantan Singingi, Indonesia

Email: erlinda120015@gmail.com

© The Author(s) 2023

1. Introduction

Artificial intelligence (AI) is a field of computer technology that aims to create systems capable of mimicking and surpassing human cognitive abilities. Using advanced algorithms and computers, artificial intelligence can learn, analyse and make decisions based on the data provided. The purpose of developing artificial intelligence is to increase the efficiency, productivity and adaptability of systems in various fields that can assist humans in decision-making [1][2][3][4].

A process that uses statistical, mathematical, artificial intelligence (AI), and machine learning techniques to extract and identify useful information from large databases is called data mining [4][5][6][7].

Juvenile delinquency and crime cannot be separated from the context of socio-cultural conditions of the era of globalization, because each period is distinctive and presents a kind of challenge, especially for the younger generation, so young people respond with a distinctive reaction to the existing social and cultural stimuli [8][9].

Adolescence is a time of psychological shock, a time of transition or a wavering bridge that connects a dependent childhood with a mature, independent adulthood. When the teenager already feels responsible for himself. He is able to account for every action and accept the philosophy of life of the society in which he lives. According to the changes of feelings he has gone through. It is undeniable that these teenagers are potentially already religious [10].

Student delinquency is an act that often causes community, school, and family unrest. Thus, student delinquency actually refers to a behavior in the form of deviation or violation of established norms, and in terms of the law, delinquency is a violation of the law that cannot be subjected to criminal law in relation to its age [3][11].

Islamic counseling not only helps in overcoming students' problems related to learning, but also touches the religious aspects of students, as religion plays a very important role in life, from bad deeds to useful lives, both in worldly life and in the hereafter [11][12].

In order to address the privacy concerns in data mining, the subfield of data mining called privacy preserving data mining (PPDM) has developed greatly in recent years. The purpose of PPDM is to protect sensitive information from unwanted or unauthorized disclosure and to preserve the utility of the data. The considerations for PPDM are twofold. First, sensitive raw data, such as individual ID card numbers and mobile phone numbers, should not be used directly for mining. Second, sensitive mining products, the disclosure of which would lead to an invasion of privacy, should be excluded [13][14][5].

Kernel-based data mining (DKBDM), such as distributed support vector engines. Among the few known data destruction problems, those related to insider attacks have increased significantly, making them one of the fastest growing types of security breaches. Once considered a minor issue, insider attacks have risen to become one of the top three data breaches. Research related to insiders includes only limited kernel-based data mining distributions, leading to significant vulnerabilities in developing protections for cooperating organisations. Previous work often fails due to multifactor models that are rather limited in scope and implementation [15][16][17][18].

Motion capture is an important technique with a wide range of applications in fields such as computer vision, computer animation, film production and medical rehabilitation. Even with a professional motion capture system. The raw data obtained largely contains the inevitable noises and outliers. Many methods have been developed to differentiate the data, although this problem remains a challenge due to the high complexity of human movements and the diversity of real-world situations. Data driven Robust Human is an approach to motion denoising that identifies spatio-temporal patterns and structural sparsity in motion data [19].

Empirical data analysis methods based on random matrix theory (RMT) and time series analysis are proposed for energy systems. In the context of Big Data in energy systems, there is a great need for new mathematical tools to describe and analyse Big Data. The results showed that empirical data from power systems modelled by RMT in time series have high sensitivity to dynamically characterised system states as well as efficiency observations in system analysis compared to conventional equation-based methods [20][21][22].

The frequent extraction of route patterns from personal trajectory data is the basis for location awareness and location services. However, most approaches are only capable of finding short, incomplete route patterns. A new approach is proposed to find route patterns, often based on abstractions of trajectories. First, path partitioning, location extraction, data simplification and general segment detection are used for abstract path data, converting these trajectories into Common Temporal Segment (STS) sequences and generating a 1-frequency frequency item set. Then a pattern search algorithm was proposed based on the temporal-spatial proximity relationship of STAR [23][24][25][26].

In this study, it will discuss one of the algorithms in Data Mining, namely the Frequent Pattern-Growth (FP-Growth) algorithm. This algorithm is part of the association technique in data mining. Fp-Growth is used as an alternative to determine data that occur frequently in a data set (Frequent Itemset). The feature of Fp-Growth is a data structure used in a tree called fp-Tree. Using Fp-Tree can extract Frequent Itemset from Fp-Tree [24]. The purpose of using this Fp-Growth algorithm in this case is to determine the behavior patterns of SMA Negeri 1 Lembang Jaya students who are often conduct the violation.

2. Research Methods

The research methodology is a step that must be carried out in order to make the preparation of the research easier and also to serve as a guide for the researcher in carrying out the research. The methodology used in the preparation of this study is presented in a framework. The form of the framework description in this study starts with the identification of the problems, problem analysis, setting the objectives, studying the literature, data collection, analysis and processing of the data, implementation, testing the results, and decision making. The phase of the framework are designed to ensure that the research is purposeful and that the objectives set in the study are achieved.

The objective of this study is to identify students based on association patterns of student violations using frequent pattern growth algorithms that can serve as a reference for teachers when making decisions about student violations in school.

This framework presents the steps taken to solve the problems under discussion. The framework of this research can be illustrated in the picture below :

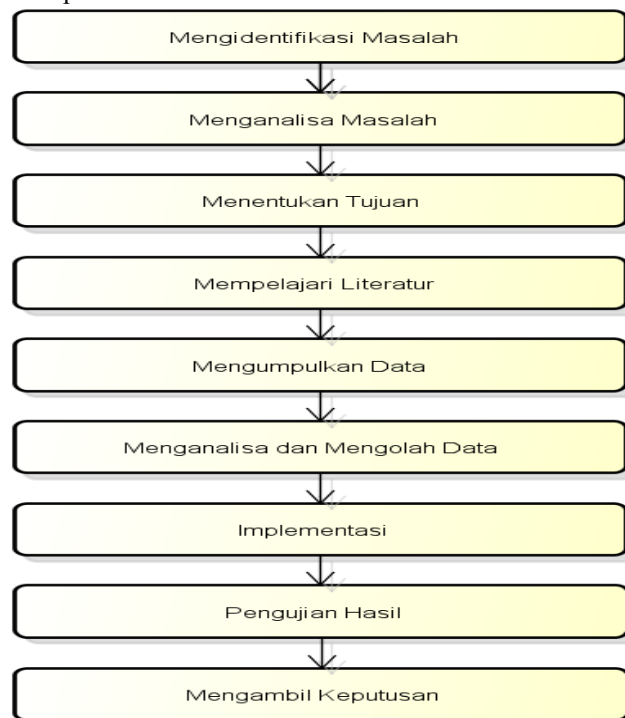


Figure 1. Framework

1 Result

3.1 Introduction

The system implementation phase is one of the phase in the system development life cycle. This phase is the phase of putting the information system so that it is ready for use. In order for a system to run properly, it must first be determined where it will be implemented.

3.2 System Implementation

In implementation, a computer is needed, to operate the computer itself also requires three supporting components as follows:

1. Hardware

Hardware used to design or run application programs that have been created. It is a computer unit that has memory as a storage medium, and a printer unit as a report printer.

2. Software

To run the designed application program must use some supporting software.

3.3 WEKA

WEKA is a package of practical Machine Learning Tools. "WEKA stands for "Waikato Environment For Knowledge Analysis", which was created at the University of Waikato, New Zealand for research, education and various applications.

- a. Create a File with Notepad and save the file in Arff format based on the prepared violation data as figure 2

```
@relation Pelanggaran@attribute Laki-Laki
{1}@attribute 'Pelanggaran Sedang' {1}
@attribute 'Koto Anau' {1}@attribute IPA2
{1}@attribute 'Pelanggaran Berat' {1}
@attribute 'Tanah Sirah' {1}@attribute IPS2
{1}@data1,1,1,1,1,?,?,?
1,1,1,1,?,?,?
1,1,1,1,?,?,?
1,1,?,?,?,1,1
1,?,?,?,1,1,?
1,?,?,?,1,1,?
1,1,1,?,?,?,1
1,1,1,?,?,?,1
?,1,?,1,?,?,?
?,1,?,1,?,?,?
1,?,1,?,1,?,?
1,?,1,?,1,?,?
1,1,?,?,?,?,?
```

Figure 2. File creation with Notepad

- b. After creating files with Notepad, then test the results with WEKA (WEKA initial display)

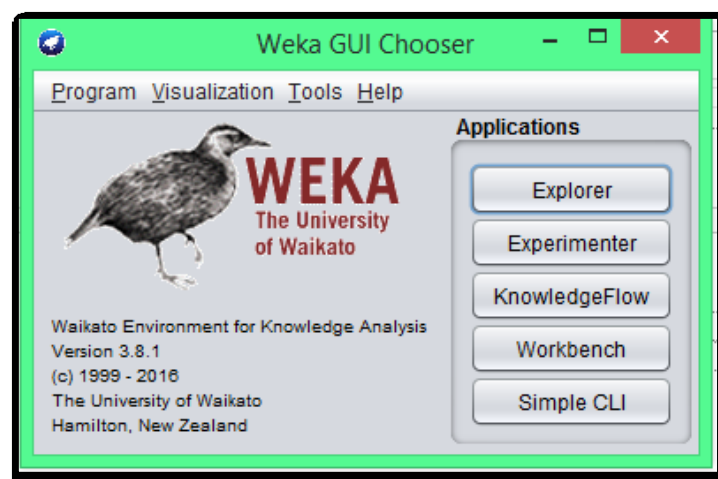


Figure 3. WEKA Initial Display

On the initial display, when the WEKA application is opened, it looks like figure 5.2. WEKA has four main menus and four buttons. The four main menus are programs, visualization, Tools, and Help.

- c. Click Tools, choose Arffviewer, then it looks like figure 4.

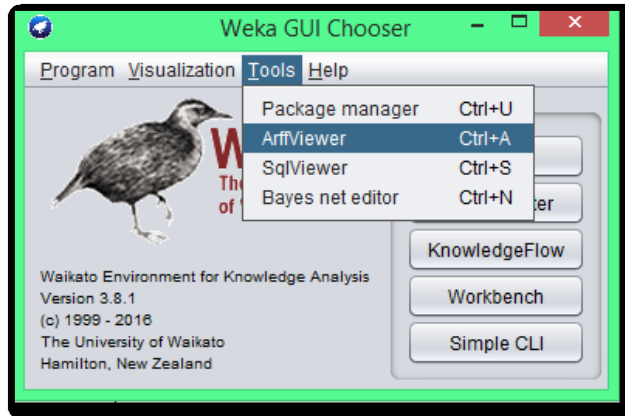


Figure 4. Arffviewer display

Arffviewer is used to display Arff Files in student violation data Spreadsheet format.

- d. Click File, choose Open then it looks like figure 5.

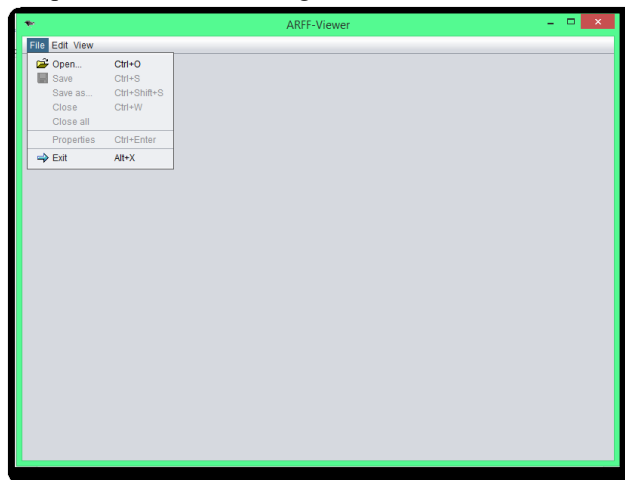


Figure 5. File Retrieval

In this phase is a way to retrieve violation data files that have been changed through Notepad with Arff format from folders stored on the saved partition.

- e. Inputted file display on WEKA as shown in figure 6.

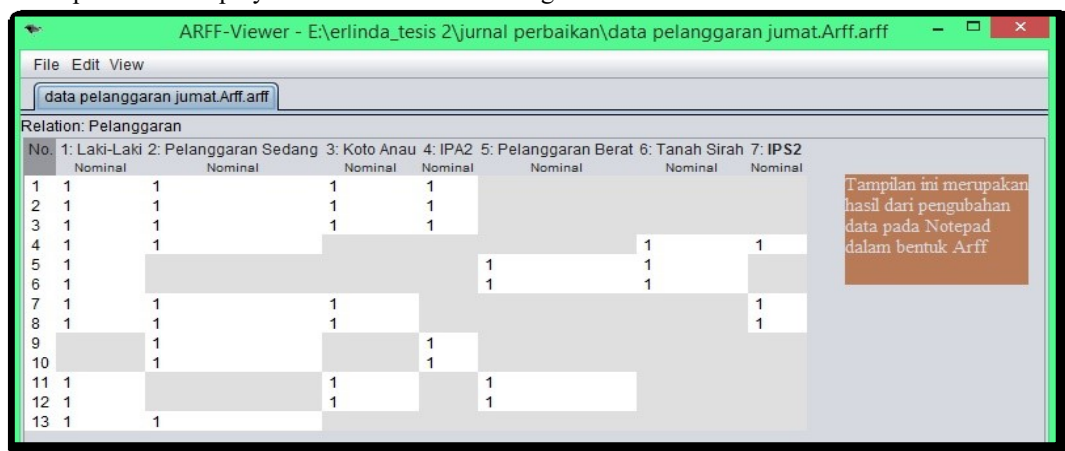


Figure 6. File Display on WEKA

- f. Then open Explorer, it looks like on figure 7.

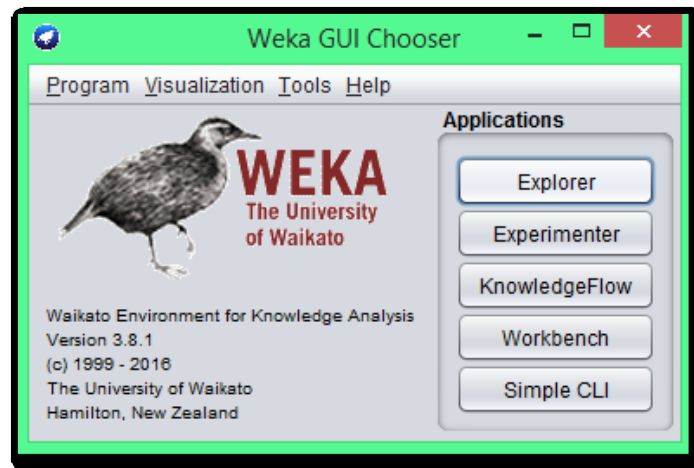


Figure 7. Display for opening Explorer

The Explorer menu is used to test student violation data to get the results of the Rule.

- g. Then click Open File and find the Arff format data savings, and click Open, as shown in figure 8.

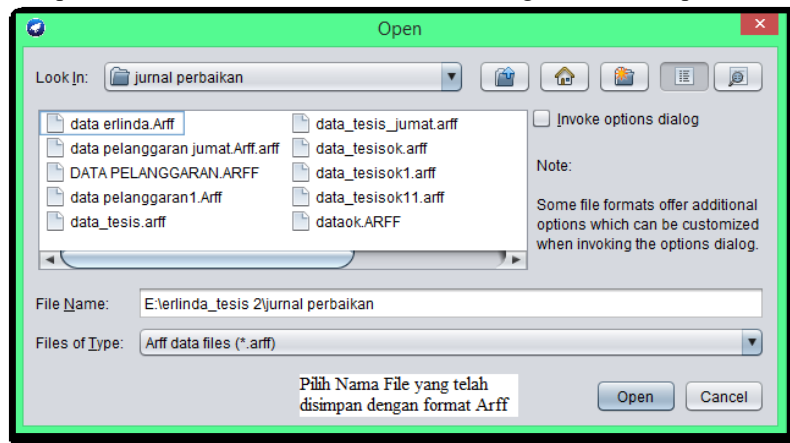


Figure 8. Display for opening files

- h. Then select Association. Click Associate, and Choose FP-Growth like figure 9.

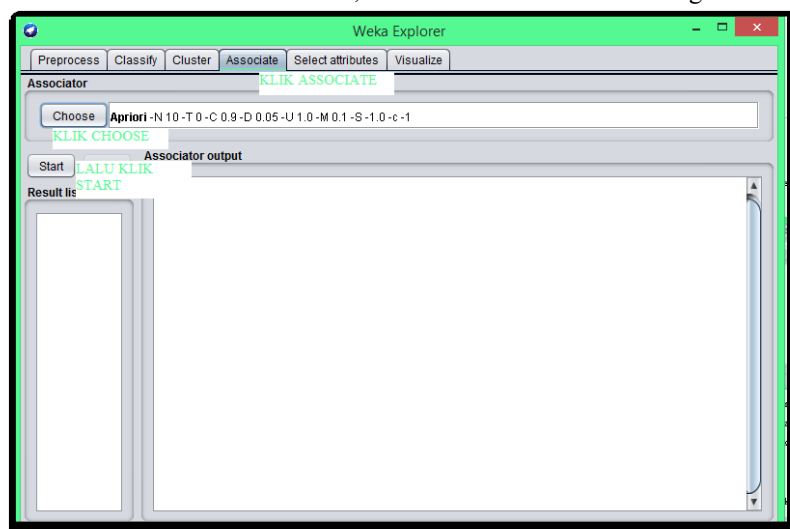


Figure 9. Display for association selection

On the figure above, it is used to select the Association and make settings that will be used in testing student violation data.

- i. Then select the Association and setting. Click FP-Growth to make settings, then click Ok as shown in figure 10.

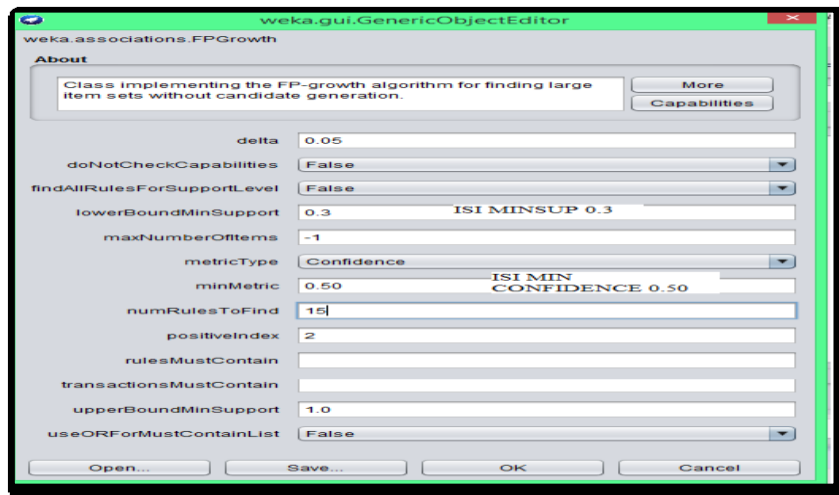


Figure 10. Display for Association Selection and Settings

- j. An then click Start to get the Output as figure 11.

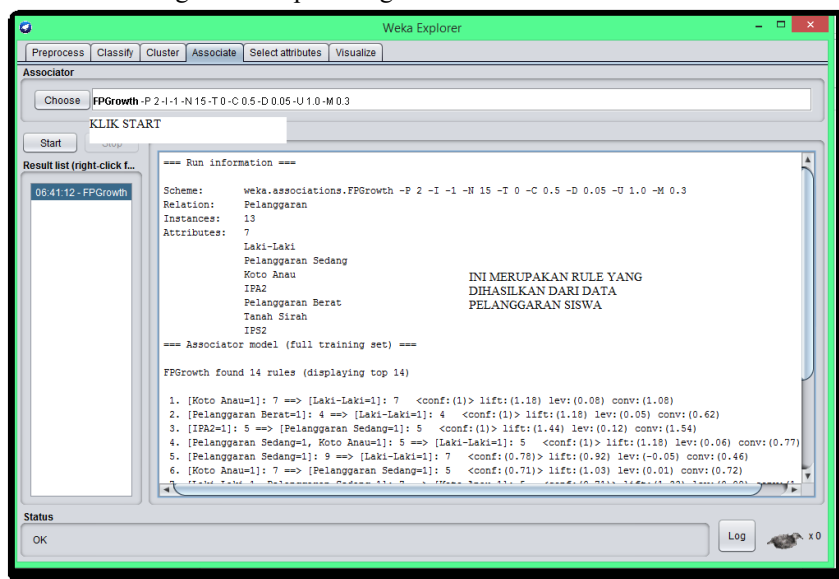


Figure 11. Display of the output result

As shown above is the result of student violation rules produced according to the Frequent Growth pattern algorithm used which is useful for making the right decision.

4. Conclusion

Based on the results of the analyzes and tests performed, the following can be concluded:

1. The greater the minimum support, the fewer rules are generated..
2. The Frequent Pattern Growth data mining method can be used to identify the violations frequently committed by students at SMA 1 Lembang Jaya, Solok..
3. The patterns obtained can be used to help guidance counselors make decisions about violations that occur.

References

1. Prasad GR. Intelligent (artificial) marketing: A data mining perspective. *Int J Manag IT* 2019;9(May).
2. Putra DJ, Nofriadi N, Erlinda E. Implementation of Fuzzy Logic Using Mamdani Method to Determine The Quantity of Bag Production (Case Study In Roman Indah Padang Bag Factory). *J Teknol DAN OPEN SOURCE*. 2022;5(1).
3. Iqbal M, Khoirunnisa N. Design And Construction Of An Expert System For Early Detection of Mental Illness in COVID-19 Patients And Handling Using Certainty Factor Method. *J Teknol Dan Open Source*. 2021;4(2):216–26.
4. Malik RA, Putra DJ, Firmansyah W. Implementation of K-Means Clustering Algorithm in Measuring Learning Media Evaluations Result by Student (Case Study Introductory Information Technology Course for Bachelor Degree of Digital Business Upertis). 2016. 1–23 p.
5. Anggun Pastika Sandi, Vina Widya Ningsih. Implementasi Data Mining Sebagai Penentu Persediaan Produk Dengan Algoritma Fp-Growth Pada Data Penjualan Sinarmart. *J Publ Ilmu Komput dan Multimed*. 2022;1(2):111–22.
6. Hasanah NN, Purnomo AS. Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus : Perpustakaan Politeknik LPP Yogyakarta). *J Teknol Dan Sist Inf Bisnis*. 2022;4(2):300–11.
7. Aranski AW, Handoko K. Data Mining Dalam Pengelompokan Nilai Iq Siswa. *J Teknol Dan Open Source*. 2019;2(2):13–22.
8. Qalbi N, Ibrahim. Senioritas dan Perilaku Kekerasan di Kalangan Mahasiswa (Studi Kasus pada Mahasiswa Fakultas Ilmu Sosial Universitas Negeri Makassar). *Soc J Soc Sci Humanit*. 2021;1(1).
9. Wahyuni T. SISTEM PAKAR UNTUK MENGIDENTIFIKASI MASALAH PSIKOLOGI REMAJA MENGGUNAKAN METODE INFERENSI FORWARD CHAINING BERBASIS ANDROID. *J-ENSITEC*. 2016;2(02).
10. Riadi D. Peran Lingkungan Pendidikan Islam. *Eliya, I A S Int Semin Islam Stud* , IAIN Bengkulu. 2019;
11. Triandini M, Defit S, Nurcahyo GW. Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia. *J Inf dan Teknol*. 2021;3:167–73.
12. Syafarudi, DKK. *Dasar-Dasar Bimbingan Dan Konseling Telaah Konsep , Teori Dan Praktik*. Perdana Publishing, 2019.
13. Et. al. PRP. Efficient Model for Privacy Preserving Classification Of Data Streams. *Turkish J Comput Math Educ*. 2021;12(2).
14. Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan, Yong Ren. Information Security in Big Data: Privacy and Data Mining. *IEEE Access*. 2014;2.
15. Sasongko TB, Arifin O. Implementasi Metode Forward Selection pada Algoritma Support Vector Machine (SVM) dan Naive Bayes Classifier Kernel Density (Studi Kasus Klasifikasi Jalur Minat SMA). *J Teknol Inf dan Ilmu Komput*. 2019;6(4).
16. Wang PS, Lai F, Hsiao HC, Wu JL. Insider Collusion Attack on Privacy-Preserving Kernel-Based Data Mining Systems. *IEEE Access*. 2016;4.
17. AHMAD ADRI. Implementasi Data Mining Menggunakan Algoritma Apriori. *Pap Knowl Towar a Media Hist Doc*. 2021;6(2):1–77.
18. Anwar MT, Permana DRA. Perbandingan Performa Model Data Mining untuk Prediksi Dropout Mahasiswa. *J Teknol dan Manaj*. 2021;19(2).
19. Feng Y, Ji M, Xiao J, Yang X, Zhang JJ, Zhuang Y, et al. Mining Spatialoral Patterns and Structural Sparsity for Human Motion Data Denoising. *IEEE Trans Cybern*. 2015;45(12).
20. Mai TT, Bezbradica M, Crane M. Learning behaviours data in programming education: Community analysis and outcome prediction with cleaned data. *Futur Gener Comput Syst*. 2022;127.
21. Han B, Luo L, Sheng G, Li G, Jiang X. Framework of Random Matrix Theory for Power System Data Mining in a Non-Gaussian Environment. *IEEE Access*. 2016;4.
22. Han F, Li Y. Moment Bounds for Large Autocovariance Matrices Under Dependence. *J Theor Probab*. 2020;33(3).
23. Tahir M, Sitompul N. Penerapan Algoritma Fp-Growth Dalam Menentukan Kecenderungan Mahasiswa Mengambil Mata Kuliah Pilihan. *Netw Eng Res Oper*. 2021;6(1):59.
24. Budianto H, Riana J. Penerapan Data Mining menggunakan algoritma Fp-Growth Untuk Penentuan Strategi Promosi Fakultas Ilmu Komputer Universitas Kuningan. *J Cloud Inf*. 2020;5(1):22–9.

25. Indahsari D. Penerapan Metode Asosiasi Dengan Algoritma FP-Growth Pada Data Transaksi PT John Tampi Group. *Comput Based Inf Syst J.* 2021;9(2):1–9.
26. Fu X, Huang J, Zeng D, Huang Y, Ding X, Paisley J. Removing rain from single images via a deep detail network. In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017.* 2017.