

Implementasi Data Mining Penerimaan Mahasiswa Baru Universitas Graha Karya Muara Bulian

Azwar Anas^{1}, Akhmadi², Ade Jermawinsyah Zebua³*

Program Studi Sistem Informasi, Universitas Graha Karya Muara Bulian¹

Program Studi Manajemen, Universitas Graha Karya Muara Bulian^{2,3}

Jl. Gajah Mada Muara Bulian-Indonesia

azwarzayn@gmail.com¹, h.akhmadijambi@gmail.com², adejermawinsyahzebua9@gmail.com³

Submitted : 28/02/2023; Reviewed : 08/04/2023; Accepted : 03/09/2023; Published : 31/10/2023

Abstract

The choice of prospective students on the registration form produces a unique. Piles of forms with combinations of program study options will have no meaning for UGK MB if detailed mining and analysis is not carried out. Difficulties in answering past data stacks, which are often also known as chunks of data of this kind, can be ported with data mining (Data Mining). The amount of data studied is 135 forms. The Apriori Algorithm analysis used is processing data. The results show that for 1-itemset, the highest frequency occurs in the Information Systems variable with a value of 58, 43% support and 100% confidence, while the lowest frequency occurs in Accounting with a value of 14, 10% support and 100% confidence. For the 2-item set, the association rules If the first choice is the PGSD study program then the second choice is the TIP study program with a frequency value of 16, 12% support and 53% confidence and this is the highest association rule. Meanwhile, the lowest association rule is the rule. If the first choice is the Agro Technology study program, then the second choice is the Information Systems study program with a frequency of 4, 3% support and 20% confidence.

Keywords: Mining, apriori, support, confidence, student.

Abstrak

Pilihan calon mahasiswa pada formulir pendaftaran menghasilkan pola yang unik. Tumpukan formulir dengan kombinasi pilihan program studi tersebut tidak akan memiliki makna bagi UGK MB apabila tidak dilakukan penambangan dan analisis secara rinci. Permasalahan untuk menjawab tumpukan data masa lalu yang sering juga dikenal dengan istilah bongkahan data semacam ini dapat dipecahkan dengan penambangan data (*Data Mining*) jumlah data yang diteliti adalah 135 formulir. Analisis Algoritma Apriori digunakan untuk mengolah data. Hasil penelitian memperlihatkan bahwa untuk 1-itemset, frekuensi tertinggi terjadi pada variabel Sistem Informasi dengan nilai 58, support 43% dan confidence 100%, sedangkan frekuensi terendah pada Akuntansi dengan nilai 14, support 10% dan confidence 100%. Untuk 2-itemset, aturan asosiasi Jika pilihan pertama program studi PGSD maka pilihan kedua adalah program studi TIP dengan nilai frekuensi 16, support 12% dan confidence 53% dan ini menjadi aturan asosiasi tertinggi. Sedangkan aturan asosiasi terendah adalah aturan Jika pilihan pertama program studi Agro Teknologi maka pilihan kedua program studi Sistem Informasi dengan frekuensi 4, support 3% dan confidence 20%.

Kata Kunci : Mining, Apriori, support, confidence, mahasiswa.

1. Pendahuluan

Universitas Graha Karya Muara Bulian (UGK MB) merupakan hasil penggabungan dua sekolah tinggi yaitu Sekolah Tinggi Ilmu Pertanian (STIP) Graha Karya dan Sekolah Tinggi Ilmu Ekonomi (STIE) Graha Karya Muara Bulian. Universitas ini berada di bawah naungan Yayasan Pendidikan Serentak Bak Regam Muara Bulian. Izin penggabungan UGK MB dikeluarkan berdasarkan Surat Keputusan Menteri Pendidikan Kebudayaan, Riset dan Teknologi Republik Indonesia Nomor: 562/E/O/2022 tanggal 27 Juli 2022 tentang Izin Penggabungan Sekolah Tinggi Ilmu Pertanian Graha Karya Muara Bulian Jambi di Kabupaten Batang Hari dan Sekolah Tinggi Ilmu Ekonomi Graha Karya di Kabupaten Batang Hari menjadi Universitas Graha Karya Muara Bulian di Kabupaten Batang Hari Provinsi Jambi. Setelah diserahterimakan Surat Keputusan tentang izin operasional UGK MB, hingga kini telah aktif melaksanakan Tri Darma Perguruan Tinggi sebagaimana perguruan tinggi lain di Indonesia.

Pelaksanaan Tri Darma Perguruan Tinggi tidak akan bisa dilakukan apabila unsur-unsur yang ada tidak lengkap. Salah satu unsur terpentingnya adalah keberadaan mahasiswa. Bagaimana mungkin pelaksanaan pendidikan dan pengajaran sebagai darma pertama bisa dilaksanakan jika tidak ada mahasiswa sebagai objek kegiatan. Oleh karena itu, UGK MB secara gencar melaksanakan promosi kepada masyarakat luas dan calon mahasiswa khususnya untuk bergabung menjadi mahasiswa.

Jumlah program studi yang ada di UGK MB adalah 9, yaitu S1 Manajemen, S1 Akuntansi, D3 Akuntansi, S1 Agribisnis, S1 Agro Teknologi, S1 Teknologi Industri Pertanian, S1 Sistem Informasi, S1 Ilmu Hukum dan S1 Pendidikan Guru Sekolah Dasar. Kesembilan program studi tersebut adalah gabungan 4 program studi lama dan 5 program studi baru dan semuanya telah terakreditasi BAN PT.

Sistem penerimaan mahasiswa baru di UGK MB yaitu mahasiswa memilih program studi sebanyak 3 pilihan, jika tidak lulus pada pilihan pertama, maka akan ditawarkan ke pilihan kedua, jika tidak lulus pada pilihan kedua, akan ditawarkan pilihan ketiga.

Pilihan calon mahasiswa pada formulir pendaftaran menghasilkan pola yang unik karena setiap calon mahasiswa memiliki pilihan yang berbeda sesuai latar belakang, cita-cita dan biaya perkuliahan dengan tetap mengacu pada 9 program studi yang ditawarkan.

Tumpukan formulir dengan kombinasi pilihan program studi tersebut tidak akan memiliki makna bagi UGK MB apabila tidak dilakukan penambangan dan analisis secara rinci. Padahal dengan melakukan penambangan data akan didapatkan pengetahuan dan informasi yang berguna bagi UGK MB dan pihak pengambil keputusan lainnya untuk menyikapi proses penerimaan mahasiswa baru di tahun berikutnya. Permasalahan untuk menjawab tumpukan data masa lalu yang sering juga dikenal dengan istilah bongkahan data semacam ini dapat dipecahkan dengan penambangan data (*Data Mining*) [1]. Ada banyak algoritma yang bisa digunakan dalam *data mining* sesuai jenis data dan pengetahuan yang ingin digali. Istilah lain dari *data mining* adalah *Pattern Recognition* [2]. Untuk menggali kaidah asosiasi (*Association Rule*) maka dikenal algoritma K-Means [3], algoritma *Market Basket Analysis* (MBA)[4], algoritma *Frequent Pattern Tree* [5], dan algoritma Apriori [6].

Penelitian kali ini penulis memilih algoritma apriori. Proses penggalian kaidah asosiasi algoritma ini menggunakan dua parameter yaitu *support* (daya dukung) dan *confidence* (tingkat kepercayaan)[7].

Rumus yang digunakan dalam algoritma apriori [4]:

$$\text{Support, } s(X \rightarrow Y) = \frac{\sigma(XUY)}{N} \quad (1)$$

$$\text{Confidence, } c(X \rightarrow Y) = \frac{\sigma(XUY)}{\sigma X} \quad (2)$$

Suatu *association rule* dianggap baik jika nilai dukungan lebih besar dari *minimum support* dan juga nilai kepercayaan lebih besar dari *minimum confidence* [8].

Beberapa penelitian terkait yang telah dihasilkan diantaranya:

1. Penelitian yang dilakukan pada Jumbo Travel Medan) [9] . Penelitian ini menggunakan algoritma apriori untuk melihat kecenderungan konsumen dalam membeli tiket pesawat. Hasil penelitian menunjukkan kombinasi pembelian tiket pesawat Air Asia dengan dukungan mencapai 50% dan kepercayaan mencapai 66,67%. Sedangkan kombinasi pembelian tiket pesawat Air Asia dan Garuda dengan dukungan mencapai 50% dan kepercayaan mencapai 85,71%.
2. Penelitian pada mata kuliah pilihan menggunakan algoritma apriori [10]. Penelitian ini bertujuan untuk melihat pola mata kuliah pilihan yang sering muncul secara bersamaan. Hasil penelitian menunjukkan kombinasi mata kuliah Pemasaran, SKB dan MSDM memiliki nilai tertinggi yaitu *support* 12% dan *confidence* 85%.
3. Penelitian pada Dealer Honda Rumbai) [11]. Penelitian ini bertujuan untuk melihat kesetiaan konsumen terhadap merek kendaraan yang dibeli. Hasil penelitian menunjukkan Jika Merek kendaraan adalah Tiger maka yang membeli adalah Karyawan Laki-laki dengan *support* 10% dan *confidence* 62,5%.

4. Penelitian pada Barbar Warehouse[12]. Algoritma yang digunakan adalah data mining dengan data penjualan produk di Barbar Warehouse. Data yang digunakan adalah data 3 bulan, data diolah menggunakan aplikasi Rapid Miner. Hasil penelitian menyimpulkan beding (seprei) dan watches (kacamata) adalah yang paling banyak terjual.

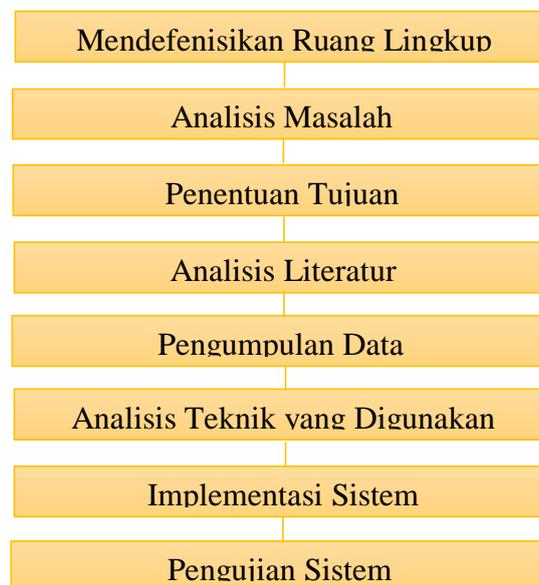
2. Metode Penelitian

Pada penelitian ini digunakan metode *library research* atau penelitian pustaka. Algoritma apriori penulis pilih karena merupakan bagian dari algoritma *data mining* yang sesuai dengan jenis data dan pengetahuan yang akan dihasilkan. Sedangkan *software data mining* yang digunakan adalah *Weka*.

Setiap melaksanakan penelitian, selayaknya menggunakan metode yang benar dan sesuai agar menghasilkan penelitian yang baik. Metode penelitian adalah langkah-langkah atau tahapan dalam menghasilkan suatu solusi terhadap masalah penelitian. Berdasarkan metode yang disusun, diharapkan penelitian berjalan dengan sistematis dan ilmiah.

2.1. Kerangka Penelitian

Berikut adalah tahapan kerangka penelitian yang dilakukan.



Gambar 1. Kerangka Penelitian

Berikut adalah uraian langkah kerja penelitian :

1. Mendefinisikan Ruang Lingkup Masalah
Langkah pertama adalah menentukan ruang lingkup dan batasan masalah, agar penelitian terfokus pada masalah yang akan dibahas.
2. Analisis Masalah
Analisis masalah dilakukan untuk menetapkan masalah penelitian yang telah ditentukan sebelumnya atau batasannya. Pada bagian ini digambarkan proses untuk menggali kaidah asosiasi penerimaan mahasiswa baru menggunakan algoritma apriori. Selain itu juga diuraikan berbagai kebutuhan data untuk menggali *rule asosiasi* yang diinginkan. Dari hasil analisis tersebut, diharapkan algoritma ini mampu menampilkan pola penerimaan mahasiswa baru dengan baik.
3. Penentuan Tujuan
Hasil pemahaman dari masalah, selanjutnya ditentukan tujuan penelitian ini. Tujuan yang adakan dicapai menjawab masalah-masalah sebelumnya.
4. Analisis Literatur
Agar tujuan tercapai, maka dianalisis berbagai referensi yang sesuai dengan masalah penelitian. Selanjutnya referensi-referensi diseleksi sesuai kebutuhan penelitian. Referensi bersumber dari jurnal ilmiah, buku dan internet yang relevan dengan penelitian.

5. Pengumpulan Data
 Tahapan pengumpulan data dimulai dengan melakukan observasi secara langsung pada objek penelitian agar masalah yang ada diketahui dengan jelas. Selanjutnya dilakukan wawancara dengan tujuan mendapatkan data-data pendukung dan informasi lainnya. Studi kepustakaan dilakukan dengan cara mempelajari berbagai buku yang terkait dengan masalah penelitian. Penelitian ini membutuhkan data-data penerimaan mahasiswa baru Universitas Graha Karya Muara Bulian yang penulis amati dengan jumlah sebanyak 135 data.
6. Analisis Teknik yang digunakan
 Langkah selanjutnya adalah analisis teknik dalam pengolahan data. Pada penelitian ini penulis menggunakan algoritma apriori.
7. Implementasi Sistem
 Selanjutnya adalah implementasi pada *software data mining Weka*. Pada *software Weka* dipilih algoritma apriori. Ketika implementasi pada sistem berjalan dengan baik, maka data dianggap akurat.
8. Pengujian Sistem
 Pengujian sistem dilakukan sebagai perbandingan antara perhitungan manual dan komputerisasi pada *software*, berikut tahapan pengujian yang akan dilakukan :
 - a. Pengujian data penerimaan mahasiswa baru dengan rumus algoritma apriori.
 - b. Berikutnya implementasi data pada *software Weka* dengan memilih algoritma apriori sebagai metode analisis.
 - c. Terakhir adalah membandingkan antara data yang diolah secara manual dengan komputerisasi, apabila hasil yang diperoleh tidak jauh berbeda, dapat disimpulkan hasil yang dicapai telah benar.

3. Hasil Penelitian

3.1. Data Penerimaan Mahasiswa Baru

Penulis menggunakan data dalam penelitian ini sebanyak 135 data penerimaan mahasiswa baru UGK MB. Adapun variabel yang digunakan adalah semua data mahasiswa baru program studi yaitu M (Manajemen), Akt (Akuntansi), Agb (Agribisnis), Agt (Agroteknologi), TIP (Teknologi Industri Pertanian), SI (Sistem Informasi), H (Ilmu Hukum) dan PGSD (Pendidikan Guru Sekolah Dasar). Variabel tersebut dipadukan untuk dilihat kaidah asosiasinya berdasarkan nilai *support* dan *confidence* yang memenuhi *threshold*.

Tabel 1. Data Penerimaan Mahasiswa Baru

No.	M	AKT	AGB	AGT	SI	PGSD	H	TIP
1	Y	Y	T	T	T	T	T	T
2	T	Y	T	T	T	Y	T	T
3	Y	T	T	T	Y	T	T	T
4	T	T	T	T	Y	Y	T	T
5	Y	T	T	T	T	T	Y	T
6	T	T	T	T	Y	T	Y	T
7	T	Y	Y	T	T	T	T	T
8	T	T	T	Y	Y	T	T	T
9	T	T	T	Y	T	T	T	Y
10	Y	T	T	T	Y	T	T	T
11	T	T	T	T	T	Y	T	Y
12	T	T	T	T	Y	T	T	Y
13	T	Y	Y	T	T	T	T	T
14	Y	T	T	T	Y	T	T	T
15	Y	T	T	T	T	T	Y	T
16	T	T	T	T	T	T	Y	Y
17	T	T	T	T	Y	T	Y	T
18	Y	T	T	T	T	Y	T	T
19	T	T	T	T	T	Y	T	Y

20	Y	T	T	Y	T	T	T	T
21	T	T	T	T	Y	T	T	Y
22	T	T	Y	T	Y	T	T	T
23	T	T	T	T	T	Y	T	Y
24	Y	T	T	Y	T	T	T	T
25	T	Y	Y	T	T	T	T	T
26	T	T	Y	T	T	Y	T	T
27	T	T	T	T	Y	T	Y	T
28	T	T	T	T	T	T	Y	Y
29	T	T	T	T	Y	T	Y	T
135	T	T	T	T	Y	T	T	Y
Jml	44	14	21	20	58	30	37	46

Berdasarkan tabel di atas, dapat dilihat jumlah mahasiswa yang memilih program studi Manajemen 44 orang, Akuntansi 14 orang, Agribisnis 21 orang, Agroteknologi 20 orang, Sistem Informasi 58 orang, PGSD 30 orang, Ilmu Hukum 37 orang dan TIP 46 orang.

3.2. Analisis Algoritma Apriori

Berdasarkan tabel 1 di atas, berikut disajikan nilai frekuensi data, *support* dan *confidence*.

Tabel 2. Nilai Frekuensi, Support dan Confidence

No.	Variabel	Frekuensi	Support	Confidence
1	Manajemen	44	33%	100%
2	Akuntansi	14	10%	100%
3	Agribisnis	21	16%	100%
4	Agro Teknologi	20	15%	100%
5	Sistem Informasi	58	43%	100%
6	PGSD	30	22%	100%
7	Hukum	37	27%	100%
8	Teknologi Industri Pertanian	46	34%	100%

Berdasarkan tabel 2 di atas, frekuensi tertinggi terjadi pada variabel Sistem Informasi dengan nilai 58, *support* 43% dan *confidence* 100%, sedangkan frekuensi terendah pada Akuntansi dengan nilai 14, *support* 10% dan *confidence* 100%. Selanjutnya adalah penghitungan nilai frekuensi, *support* dan *confidence* untuk 2-itemset sebagaimana tabel 3 berikut.

Tabel 3. Nilai Frekuensi, Support dan Confidence 2-itemset

No.	2-itemset	Frekuensi	Support	Confidence
1	Jika memilih Manajemen maka memilih Sistem Informasi	13	10%	30%
2	Jika memilih Manajemen maka memilih Hukum	12	9%	27%
3	Jika memilih Agribisnis maka memilih Sistem Informasi	7	5%	33%
4	Jika memilih Agro Teknologi maka memilih Sistem Informasi	4	3%	20%

5	Jika memilih Agro Teknologi maka memilih TIP	6	4%	30%
6	Jika memilih Sistem Informasi maka memilih Hukum	16	12%	28%
7	Jika memilih Sistem Informasi maka memilih TIP	15	11%	26%
8	Jika memilih PGSD maka memilih TIP	16	12%	53%
9	Jika memilih Hukum maka memilih TIP	9	7%	33%

Berdasarkan tabel 3 di atas, terdapat aturan Jika pilihan pertama program studi PGSD maka pilihan kedua adalah program studi TIP dengan nilai frekuensi 16, *support* 12% dan *confidence* 53% dan ini menjadi aturan asosiasi tertinggi untuk 2-itemset. Sedangkan aturan asosiasi terendah untuk 2-itemset adalah aturan Jika pilihan pertama program studi Agro Teknologi maka pilihan kedua program studi Sistem Informasi dengan frekuensi 4, *support* 3% dan *confidence* 20%.

3.3. Pengujian Pada Software Weka

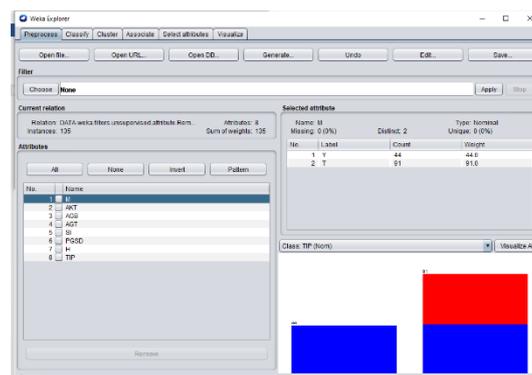
Tahapan dalam melakukan pengujian pada *software data mining Weka* adalah sebagai berikut:

1. Tampilan *user interface* Weka



Gambar 2. User Interface Weka

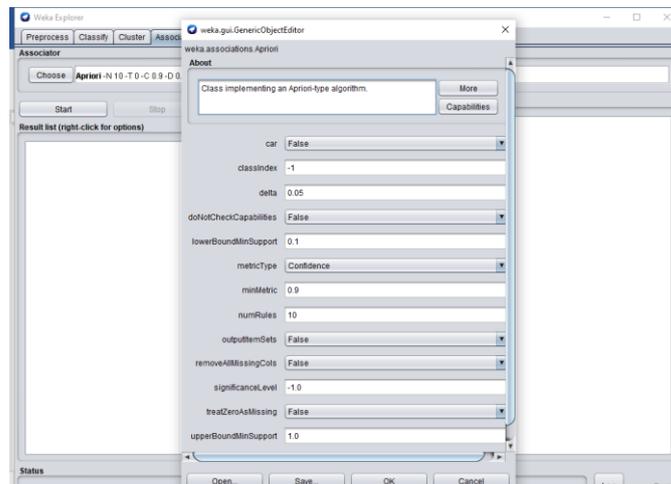
2. Proses pemanggilan data dengan format csv (*comma delimited*) melalui menu *Explorer*.



Gambar 3. Proses Pemanggilan Data

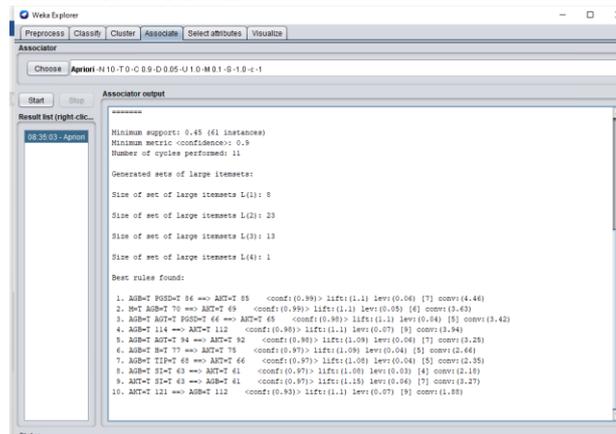
Berdasarkan gambar 3 di atas, terdapat 8 variabel data yaitu M (Manajemen), AKT (Akuntansi), AGB (Agribisnis), AGT (Agro Teknologi), SI (Sistem Informasi), PGSD (Pendidikan Guru Sekolah Dasar), H (Hukum) dan TIP (Teknologi Industri Pertanian). Kemudian terdapat 44 data dengan kategori Y (Ya) dan 91 data dengan kategori T (Tidak).

3. Pemilihan algoritma Apriori melalui menu *Associate* dan penetapan nilai *Minimum Support* dan *Minimum Confidence*



Gambar 4. Pemilihan Algoritma Apriori dan penetapan Nilai Minimum Support dan Minimum Confidence

4. Tampilan hasil *Association Rule* Terbaik



Gambar 5. Hasil Rule Association Terbaik

4. Penutup

Tingkat akurasi pengujian menggunakan *software Weka* dibandingkan dengan perhitungan manual mencapai 100%. Untuk 1-itemset, frekuensi tertinggi terjadi pada variabel Sistem Informasi dengan nilai 58, *support* 43% dan *confidence* 100%, sedangkan frekuensi terendah pada Akuntansi dengan nilai 14, *support* 10% dan *confidence* 100%. Untuk 2-itemset, aturan asosiasi Jika pilihan pertama program studi PGSD maka pilihan kedua adalah program studi TIP dengan nilai frekuensi 16, *support* 12% dan *confidence* 53% dan ini menjadi aturan asosiasi tertinggi. Sedangkan aturan asosiasi terendah adalah aturan Jika pilihan pertama program studi Agro Teknologi maka pilihan kedua program studi Sistem Informasi dengan frekuensi 4, *support* 3% dan *confidence* 20%.

5. Referensi

- [1] F. Wang *et al.*, "Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns," *Energy Convers. Manag.*, vol. 171, pp. 839–854, 2018, doi: 10.1016/j.enconman.2018.06.017.
- [2] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi:

- 10.25077/teknosi.v5i1.2019.17-24.
- [3] N. T. Suswanto, P. Chyan, and V. Putri, "Using K-Means Algorithm to Investigate Community Behavior in Treating Waste toward Smart City," vol. 11, no. 4, 2021.
- [4] A. Anas, Azwar; Jermawinsyah Zebua, "Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat," *J. Ilm. MEDIA SISFO*, vol. 16, no. 1, pp. 54–61, 2022, doi: 10.35957/jatasi.v7i2.195.
- [5] B. S. Hasugian, "Penerapan Metode Association Rule Untuk Menganalisa Pola Pemakaian Bahan Kimia Di Laboratorium Menggunakan Algoritma FP-Growth (Studi Kasus di Laboratorium Kimia PT . PLN (Persero) Sektor Pembangkitan Belawan Medan) Buyung Solihin Hasugian Universitas," *J. Ilmu Komput. dan Inform.*, vol. 6341, no. November, pp. 56–69, 2019.
- [6] D. Ai, H. Pan, X. Li, Y. Gao, and D. He, "Association rule mining algorithms on high-dimensional datasets," *Artif. Life Robot.*, vol. 23, no. 3, pp. 420–427, 2018, doi: 10.1007/s10015-018-0437-y.
- [7] M. Abdel-Basset, M. Mohamed, F. Smarandache, and V. Chang, "Neutrosophic association rule mining algorithm for big data analysis," *Symmetry (Basel)*, vol. 10, no. 4, pp. 1–19, 2018, doi: 10.3390/sym10040106.
- [8] J. M. Luna, M. Ondra, H. M. Fardoun, and S. Ventura, "Optimization of quality measures in association rule mining: An empirical study," *Int. J. Comput. Intell. Syst.*, vol. 12, no. 1, pp. 59–78, 2018, doi: 10.2991/ijcis.2018.25905182.
- [9] Z. Wardah, "Implementasi Data Mining Pada Penjualan Tiket Pesawat Menggunakan Algoritma Apriori (Studi Kasus : PT. Pesona Ceria Travel)," vol. 2, no. 2, pp. 31–39, 2017.
- [10] A. Anas and B. Darma, "Algoritma Asosiasi untuk Mendapatkan Pola Mata Kuliah Pilihan STIE-GK Muara Bulian," *J. Edik Inform.*, vol. 6, no. 1, pp. 1–12, 2019.
- [11] W. Choiriah, "Penggunaan Algoritma Apriori Data Mining Untuk Mengetahui Tingkatkesetiaan Konsumen (Brand Loyalty) Terhadap Merek Kendaraan Bermotor (Studi Kasus Dealer Honda Rumbai)," *Pengguna. Algoritm. Apriori Data Min. Untuk Mengetahui Tingkat Kesetiaan Konsum. (Brand Loyal. Terhadap Merek Kendaraan Bermotor (Studi Kasus Deal. Honda Rumbai)*, vol. 7, no. 1, pp. 44–52, 2016.
- [12] A. F. Lestari and M. Hafiz, "Penerapan Algoritma Apriori Pada Data Penjualan Barbar Warehouse," *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 1, p. 96, 2020, doi: 10.35314/isi.v5i1.1317.