

Pemilihan Program Studi Menggunakan Algoritma Fp-Growth dan J48

Norhikmah^{1*}, Rumini²

¹Program Studi Sistem Informasi, Universitas AMIKOM Yogyakarta, Indonesia

²Program Studi Teknik Informatika, Universitas AMIKOM Yogyakarta, Indonesia

Email: ¹hikmah@amikom.ac.id, ²rumini@amikoml.ac.id,

INFORMASI ARTIKEL

Histori artikel:

Naskah masuk, 25 Nopember 2021

Direvisi, 19 Pebruari 2022

Diiterima, 19 Pebruari 2022

Kata Kunci:

J48,

Fp-Growth,

Program Studi,

Assosiation rule

ABSTRAK

Abstract- The study program is an absolute thing and must be chosen by prospective new students when registering to become a student, the lack of knowledge of the courses offered by the campus becomes a difficulty in determining study programs that match their interests. The first research method of analyzing data uses the association rule to get the relationship of study programs with other study programs, and subsequently classifies the study program using decision tree j48, and is followed by examiners using confusion matrix. The results of this study are to use the association rule Fp-Growth method obtained 10 best rules, namely for the first rule D3 Informatics with International Relations. Of the 10 rules obtained as a reference or basis for determining classes on the j48 algorithm, from the results of the 10 rules, analyzed according to the rules applied in Amikom, then obtained into 11 classes where the rules are based on the origin of natural science or social studies schools. By using the j48 algorithm, 99.8% accuracy is obtained with the highest hierarchy in the decision tree, namely the D3 Informatics study program and for the origin school, namely IPS high school

Abstrak- Program studi merupakan suatu hal yang mutlak dan harus dipilih oleh calon mahasiswa baru ketika akan mendaftar menjadi seorang mahasiswa, kurangnya pengetahuan terhadap program studi yang ditawarkan oleh kampus menjadi suatu kesulitan dalam menentukan prodi yang sesuai dengan minat. Metode penelitian yang pertama menganalisis data menggunakan *assosiation rule* untuk mendapatkan keterkaitan program studi dengan program studi yang lain, dan selanjutnya mengklasifikasi program studi menggunakan decision tree j48, dan dilanjutkan dengan pengujian menggunakan *confusion matrix*. Hasil dari penelitian ini adalah menggunakan association rule metode Fp-Growth didapatkan 10 rule terbaik yaitu untuk rule pertama D3 Informatika dengan Hubungan Internasional. Dari 10 rule yang didapatkan sebagai acuan atau dasar dalam menentukan kelas pada algoritma j48, dari hasil 10 rule tersebut, dianalisis sesuai aturan yang diterapkan dipmb Amikom maka diperoleh menjadi 11 kelas yang dimana rule tersebut berdasarkan asal sekolah IPA atau IPS. Dengan menggunakan algoritma j48 didapatkan akurasi 99,8% dengan hirarki tertinggi dalam pohon keputusan yaitu program studi D3 Informatika dan untuk sekolah asal yaitu SMA IPS

Copyright © 2022 LPPM - STMIK IKMI Cirebon

This is an open access article under the CC-BY license

Penulis Korespondensi:

Norhikmah

Program Studi Sistem Informasi,

Universitas AMIKOM Yogyakarta

Jln.Ring Road Utara, Kel.Condong Catur,Kec.Depok,Kab Sleman,Prop .Yogyakarta

Email: hikmah@amikom.ac.id

1. Pendahuluan

Program studi merupakan suatu hal yang mutlak yang harus dipilih oleh calon mahasiswa baru ketika akan mendaftar menjadi seorang mahasiswa, ketika pilihan prodi sudah dipilih minimal 4 pilihan, maka dilanjutkan dengan ujian masuk, hasil ujian masuk tersebut menjadi penentu apakah diterima keprodi yang menjadi pilihan utama atau masuk ke pilihan prodi alternatif. Kurangnya pemahaman pengetahuan calon mahasiswa baru terhadap informasi program studi yang ditawarkan, sehingga mahasiswa mengalami kesulitan dan cenderung bingung dalam memilih program studi apa saja yang menjadi prioritas untuk dipilih menjadi pilihan terbaik, maka dari itu dibutuhkan peran kampus dalam hal ini karyawan pmb yang dibantu oleh data yang sudah diolah menjadi informasi dalam memberikan rekomendasi pilihan prodi yang tepat. Pemilihan konsentrasi sangat berpengaruh terhadap akademik mahasiswa, kebanyakan mahasiswa memilih konsentrasi bukan karena pertimbangan yang matang akan tetapi mengikuti tren pilihan teman-temannya, sehingga tidak sesuai dengan kemampuannya[1]. Sekumpulan data item set yang sering muncul secara bersamaan, data dicari menggunakan *association rule*. Jika mahasiswa memilih prodi z maka kemungkinan memilih prodi x data tersebut bisa dijadikan pihak kampus untuk menjadi sebuah informasi[2]. salah satu fungsi *association analyst* yaitu algoritma *FP-Growth*. Adapun *FPGrowth* sendiri adalah alternatif algoritma yang dapat digunakan dalam hal menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sekumpulan data[3]. Algoritma untuk klasifikasi salah satunya menggunakan metode c45 yang terdiri dari kumpulan data yang banyak, dan membentuk pohon keputusan[4]. menggunakan metode data mining dan machine learning yakni algoritma J48 Decision Tree dapat digunakan untuk membantu para pembuat kebijakan maupun para pemangku kepentingan untuk mengambil keputusan[5].

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar[6]. Terdapat beberapa istilah lain yang memiliki makna sama dengan data mining, yaitu *Knowledge discovery in databases (KDD)*, ekstraksi pengetahuan (*knowledge extraction*), Analisa data/pola (*data/pattern analysis*), kecerdasan bisnis (*business intelligence*) dan data *archaeology* dan data *dredging*[7].Masalah-masalah yang sesuai untuk diselesaikan dengan teknik *data mining* dapat dirincikan dengan Memerlukan keputusan yang bersifat

knowledgebased. Mempunyai lingkungan yang berubah, Metode yang ada sekarang bersifat sub-optimal, Tersedia data yang bisa diakses, cukup dan relevan dan Memberikan keuntungan yang tinggi jika keputusan yang diambil tepat[8].

Berdasarkan fungsionalitasnya, tugas-tugas data mining dikelompokkan kedalam enam kelompok adalah a). Klasifikasi (*classification*), mengeneralisasi struktur yang diketahui untuk diaplikasikan pada data-data baru. Misalkan, klasifikasi penyakit kedalam sejumlah jenis. b). Klusterisasi (*clustering*), mengelompokkan data yang tidak diketahui label kelasnya, kedalam sejumlah kelompok tertentu sesuai ukuran kemiripannya. c). Regresi, menemukan suatu fungsi memodelkan data dengan galat (kesalahan prediksi) seminimal mungkin. d). Deteksi anomali (*anomaly detection*), mengidentifikasi data yang tidak umum, dapat berupa *outlier* (pencilan), perubahan atau deviasi yang mungkin sangat penting dan perlu investigasi lebih lanjut. e). Pembelajaran aturan asosiasi (*association rule learning*) atau pemodelan kebergantungan (*dependency modeling*), mencari relasi/hubungan antar variable. f). Perangkuman (*summarization*), menyediakan representasi data yang lebih sederhana, meliputi visualisasi dan pembuatan laporan[9].

Klasifikasi data terdiri dari 2 langkah proses, yang Pertama adalah *learning (fase training)*, dimana algoritma klasifikasi dibuat untuk menganalisa data *training*, lalu direprestasikan dalam bentuk *rule* klasifikasi. Proses yang Kedua adalah klasifikasi, dimana data tes digunakan untuk memprediksi atau memperkirakan akurasi dari *rule* klasifikasi [10] *Association Rule* adalah *Association rule* adalah teknik untuk menemukan aturan asosiatif antara suatu kombinasi *item*. Contoh aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu[11]. *Fp Growth* adalah algoritma pencarian *frequent itemsets* yang didapat dari dengan menjelajahi *tree* dari bawah menuju ke atas, algoritma *Fp-Growth* merupakan pengembangan dari algoritma *Apriori*. Sehingga kekurangan dari algoritma *apriori* diperbaiki di algoritma *Fp-Growth*. Algoritma ini menentukan *frequent itemset* yang berakhiran *suffix* tertentu dengan menggunakan metode *divide and conquer* untuk memecah *problem* menjadi *subproblem* yang lebih kecil[11].

Tree atau pohon keputusan banyak dikenal sebagai bagian dari *Graph*, yang termasuk dalam irisan bidang ilmu otomata dan teori bahasa serta

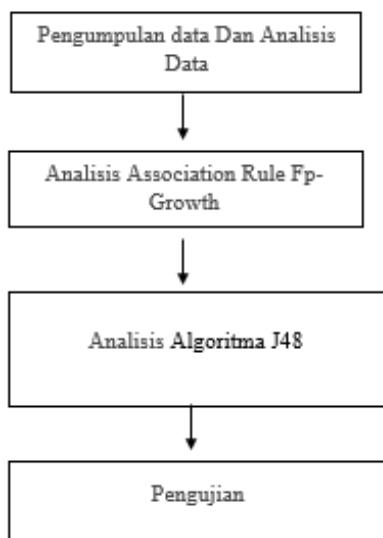
matematika diskrit. *Tree* sendiri merupakan graf tak-berarah yang terhubung, serta tidak mengandung sirkuit[12].

J48 adalah salah satu jenis *classifier* pada metode klasifikasi dalam data mining dan bagian dari C4.5 decision tree yang sederhana. C4.5 membangun sebuah pohon keputusan berdasarkan pada seperangkat *input* data yang berlabel. Pohon keputusan adalah model prediksi menggunakan pohon keputusan atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan[13].

Confusion matrix berisi informasi tentang actual dan prediksi klasifikasi dilakukan oleh sistem klasifikasi. Kinerja sistem seperti itu biasanya dievaluasi menggunakan data dalam matrix[14]. Tujuan dalam penelitian ini adalah untuk mengetahui keterkaitan antar program studi dengan menggunakan *association rule*, dan memberikan rekomendasi pilihan program studi yang sesuai dengan minat menggunakan algoritma decision tree J48

2. Metode

Menggunakan teknik *exprimnt* dalam menganalisis permasalahan dalam memilih program studi, dengan tahapan berikut:



Gambar 1. Alir Diagram Penelitian

Berikut Penjelasan dari gambar 1:

- Analisis aata dari data internal AMIKOM yaitu Data PMB dari tahun 2017 -2019
- Analisis Association Rule (Fp-Growth)

Analisis data menggunakan *Association rule* untuk mengetahui pola atau rule pemilihan dan *support* dan *confidence* pilihan program studi menggunakan algoritma *growth*. Untuk menentukan nilai *support* dapat diperoleh dengan persamaan sebagai berikut [11]:

$$\text{Support A} = \frac{\text{Jumlah Transaksi A}}{\text{Total Transaksi}} \quad (1)$$

$$\text{Support } (A \rightarrow B) = P(A \cap B) \quad (2)$$

$$\text{Support } (A \cap B) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}} \quad (3)$$

Sedangkan persamaan untuk mencari *Confidence*:

$$\text{Confidence } (A \rightarrow B) = P(A | B) \quad (4)$$

$$\text{Support } (A | B) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Jumlah Transaksi A}} \quad (5)$$

2.3 .Analisis Algoritma J48

Penelitian dengan menerapkan metode j48 untuk mengklasifikasi pemilihan program studi. Berikut ini tahapan algoritme J48 :

- Menyiapkan data training
- Menentukan akar dari pohon.
- Menghitung nilai Gain melalui rumus $\text{Entropy}(S) = \sum_{i=1}^n -p_i * \log_2 p_i$ (6)
- Ulangi langkah ke-2 hingga semua tupel terpartisi dengan menggunakan rumus
- Gain $(S,A) = S - \sum_{i=1}^n \frac{|S_i|}{|S|} * S_i$ (7)
- Proses partisi pohon keputusan akan berhenti saat semua tupel dalam node N mendapat kelas yang sama dan atau tidak ada atribut di dalam tupel yang dipartisi lagi dan atau tidak ada tupel didalam cabang yang kosong.

2.4 Pengujian menggunakan *Confusion matrix*. Beberapa istilah[14]:

1. True positive (TP)

Jika hasil dari prediksi adalah p dan aktualnilainya juga p, maka itu disebut true positive (TP).

2. False positive (FP)

Jika nilai aktualnya adalah n maka dikatakan salah positif

3. Precision and Recall

Precision adalah sebagian kecil dari instance yang diambilrelevan, sedangkan *recall* adalah sebagian kecil dari contoh yang relevan yang diambil. Oleh

karena itu baik *Precision and recall* erdasarkan pada pemahaman dan ukuran relevansi.

4. *Precision* dapat dilihat sebagai ukuran ketepatan atau kualitas, sedangkan *Recall* adalah ukuran kelengkapan atau kuantitas.

High recall berarti suatu algoritma mengembalikan sebagian besar hasil yang relevan. Presisi tinggi berarti suatu algoritma mengembalikan hasil yang lebih relevan daripada tidak relevan.

$$Precision = \frac{tp}{tp + fP} \quad (8)$$

$$Recall = \frac{tp}{tp + fp} \quad (9)$$

Weka adalah perangkat lunak yang memuat algoritma machine learning yang digunakan dalam menyelesaikan tugas-tugas atau analisa pada data mining[15][16].

3. Hasil Pembahasan

3.1 Analisis Data

Data yang digunakan berasal dari data pilihan dan kelulusan PMB tahun ajaran 2017-2020 dengan jumlah keseluruhan 13448 data. Tahapan dalam penyeleksian data:

- 1). Pengelompokkan data berdasarkan Program studi.
- 2). Menjumlahkan pilihan program studi berdasarkan tahun, hasil penyeleksian terlihat pada tabel 1.

Tabel 1. Data Total Pilihan Program Studi

No	Program Studi	Total
1	D3-Teknik Informatika (D3TI)	2124
2	D3-Manajemen Informatika (MI)	2530
3	S1-Informatika (IF)	6021
4	S1-Sistem Informasi (SI)	8024
5	S1-Bachelor Information Technology (BCIT)	796
6	S1-Bachelor Informatics (BI)	327
7	S1-Bachelor Information System (BCIS)	636
8	S1-Bachelor Communication Science	123
9	S1-Teknologi Informasi (SITI)	6804
10	S1-Teknik Komputer (Tekkom)	4570
11	S1-Arsitektur	1886
12	S1-Geografi	1263
13	S1-Perencanaan Wilayah dan kota	1828
14	S1-Ekonomi	1935
15	S1-Kewirausahaan	2477
16	S1-Akuntansi	1700

No	Program Studi	Total
17	S1-Ilmu Pemerintahan (IP)	2034
18	S1-Hubungan Internasional (HI)	2492
19	S1-Ilmu Komunikasi (Ilkom)	5898

3.2 Analisis Association Rule Fp-Growth

Analisis Fp-Growth dilakukan untuk mengetahui program studi yang paling banyak dipilih, dan menentukan *Frequent Itemset* yang mana selanjutnya akan digunakan sebagai penentuan *rule* dalam algoritma J48.

Tabel 2. Data Support Program Studi

No	Program Studi	Support
1	D3-Teknik Informatika	0,157942
2	D3-Manajemen Informatika	0,188132
3	S1-Informatika	0,447725
4	S1-Sistem Informasi	0,596669
5	S1-Bachelor Information Technology	0,059191
6	S1-Bachelor Informatics	0,024316
7	S1-Bachelor Information System	0,047293
8	S1-Bachelor Communication Science	0,009146
9	S1-Teknologi Informasi	0,505949
10	S1-Teknik Komputer	0,339827
11	S1-Arsitektur	0,140244
12	S1-Geografi	0,093917
13	S1-Perencanaan Wilayah dan kota	0,135931
14	S1-Ekonomi	0,143888
15	S1-Kewirausahaan	0,184191
16	S1-Akuntansi	0,126413
17	S1-Ilmu Pemerintahan	0,151249
18	S1-Hubungan Internasional	0,185306
19	S1-Ilmu Komunikasi	0,438578

Tabel 2 merupakan hasil dari perhitungan persamaan 1-3, yang dimana hasil support tertinggi diperoleh program studi sistem informasi yaitu sebesar 59,67%.

Tabel 3. Hasil Confidence Prodi

No	Nama Prodi			C
1	D3-TI	S1-HI		0,96
2	S1-IF	S1-IP	S1-Tekkom	0,96
3	S1-IF	S1-Ekonomi	S1-Tekkom	0,95
4	S1-IP	S1-Tekkom		0,93
5	S1-	S1-Tekkom		0,92

No	Nama Prodi			C
	Ekonomi			
6	S1-Tekkom	S1-Ekonomi	S1-IF	0,92
7	D3-MI	Hubungan Internasional		0,92
8	S1-IF	S1-Kewirausahaan	S1-Tekkom	0,91
9	S1- S1-Tekkom	S1-IP		0,91
10	SI-HI	S1-Ekonomi	S1-Tekkom	0,91

Tabel 3 memperlihatkan nilai *confidence* 2 program studi, dengan menggunakan persamaan (4), dilanjutkan perhitungan *Fp Growth* menggunakan *tools* Weka 3.8

3.3 Analisis Algoritma J48

Berdasarkan hasil rule yang diperoleh dengan menggunakan perhitungan *Fp-growth* yang menjadi dasar dalam menentukan kelas pada penentuan data training dalam perhitungan algoritma j48, akan tetapi karena Universitas Amikom memiliki aturan harus memilih 4 program studi dan ada syarat pilihan yang berlaku maka dianalisis ulang seperti yang terlihat hasil akhir rulenya pada tabel 4 sebagai berikut:

Tabel 4. Nama Kelas Program Studi

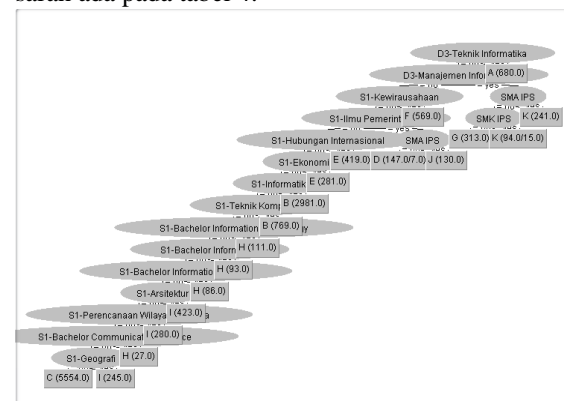
Nama Kelas	Keterangan Program Studi
A	D3 Informatika, Hubungan Internasional, Informatika, Ilmu Pemerintahan
B	Informatika, Teknik Komputer, Ekonomi, Ilmu Pemerintahan
C	Sistem Informasi, Akuntansi, Ilmu Komunikasi, Teknologi Informasi
D	Ilmu Pemerintahan, Teknik Komputer, Informatika, Hubungan Internasional
E	Ekonomi, Teknik Komputer, Informatika, Hubungan Internasional
F	Kewirausahaan, Informatika, D3 Teknik Informatika, Teknik Komputer
G	D3 Manajemen, Hubungan Internasional, D3 Teknik Informatika, Teknik Informatika
H	BCIT, BCIS, BCS, BI
I	Teknologi Informasi, Arsitektur, Geografi, Perencanaan
J	Ilmu Pemerintahan, hubungan Internasional, ilmu komunikasi, Ekonomi
K	D3 Manajemen, Sistem Informasi, Teknologi Informasi, Ilmu Komunikasi

Tabel 5. Contoh Dataset

No	Pendidikan	Minat	Saran
1	SMA IPA	S1-Sistem Informasi	C
2	SMA IPA	S1-Informatika	B

3	SMA IPA	S1-Geografi	I
4	SMA IPA	S1-Arsitektur	I
5	SMA IPA	S1-Sistem Informasi	C
6	SMA IPA	S1-Informatika	B
7	SMA IPA	S1-Teknik Komputer	B
8	SMA IPA	S1-Informatika	B
9	SMA IPA	S1-Informatika	B

Tabel 5 adalah contoh dataset yang digunakan sebagai data pada metode j48, dengan keterangan saran ada pada tabel 4.



Gambar 1. Pohon Keputusan

Pada Gambar 1 diatas adalah gambaran pohon keputusan yang dimana yang menjadi akarnya adalah prodi Teknik informatika dan SMA IPS.

```

S1-Teknik Komputer = no
S1-Bachelor Information Technology = no
S1-Bachelor Informatics = no
S1-Bachelor Information System = no
S1-Arsitektur = no
S1-Perencanaan Wilayah dan kota = no
S1-Bachelor Communication Science = no
S1-Geografi = no: C (5554.0)
S1-Geografi = yes: I (245.0)
S1-Bachelor Communication Science = yes: H
S1-Perencanaan Wilayah dan kota = yes: I (280.0)
S1-Arsitektur = yes: I (423.0)
S1-Bachelor Information System = yes: H (86.0)
S1-Bachelor Informatics = yes: H (93.0)
S1-Bachelor Information Technology = yes: H (111.0)
S1-Teknik Komputer = yes: B (769.0)
S1-Informatika = yes: B (2981.0)
S1-Ekonomi = yes: E (281.0)
S1-Hubungan Internasional = yes: E (419.0)
S1-Ilmu Pemerintahan = yes
SMA IPS = no: D (147.0/7.0)
SMA IPS = yes: J (130.0)
S1-Kewirausahaan = yes: F (569.0)
D3-Manajemen Informatika = yes
SMA IPS = no
SMK IPS = no: G (313.0)
SMK IPS = yes: K (94.0/15.0)
SMA IPS = yes: K (241.0)
D3-Teknik Informatika = yes: A (680.0)
    
```

Gambar 2. Hasil J48

Gambar 2 diatas menceritakan contoh prodi ilmu pemerintahan ada 2 pilihan jika SMA IPA no maka pilihannya kelas D jika Iya pilihannya kelas J, sedangkan kelas E ada 2 pilihan prodi yang memilih yaitu prodi ekonomi dan hubungan internasional.

4 Kesimpulan

Kesimpulan dari penelitian ini adalah menggunakan association rule metode *Fp-Growth* didapatkan 10 rule terbaik yaitu untuk rule pertama D3 Informatika dengan Hubungan Internasional. Dari 10 rule yang didapatkan sebagai acuan/dasar dalam

menentukan kelas pada algoritma j48, dari hasil 10 rule tersebut, dianalisis sesuai aturan yang diterapkan dipmb Amikom maka diperoleh menjadi 11 kelas atau rule berdasarkan asal sekolah IPA atau IPS. Dengan menggunakan algoritma j48 didapatkan akurasi 99,8% dengan hirarki tertinggi dalam pohon keputusan yaitu program studi D3 Informatika dan untuk sekolah asal adalah SMA IPS.saran untuk penelitian selanjutnya mengukur tingkat keberhasilan dari program studi yang dipilih.

UCAPAN TERIMA KASIH

Ucapan Terimakasih Kepada Direktorat LPPM Universitas AMIKOM Yogyakarta yang sudah memberikan dana hibah kepada penelitian ini.

Daftar Pustaka

- [1] J. Aranda and W. A. G. Natasya, "Penerapan Metode K-Means Cluster Analysis Pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk Mahasiswa International Class Stmik Amikom Yogyakarta," *Semasteknomedia Online*, vol. 4, no. 1, pp. 4-2-1, 2016.
- [2] D. Arisandi and Nofriandi, "Analisa Pola Pemilihan Program Studi Bagi Calon Mahasiswa DI Universitas Abdurrah Menggunakan Association Rule," *J. Teknol. dan Inf. UNIVRAB*, vol. 1, no. Januari, 2016.
- [3] S. Kurniawan, W. Gata, and H. Wiyana, "Analisis Algoritma FP-Growth Untuk Rekomendasi Produk Pada Data Retail Penjualan Produk Kosmetik (Studi Kasus: MT Shop Kelapa Gading)," *Semin. Nas. Teknol. Inf. dan Komun. 2018 (SENTIKA 2018)*, vol. 2018, no. 8, pp. 61-69, 2018.
- [4] B. Novianti, T. Rismawan, and S. Bahri, "Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak)," *J. Coding, Sist. Komput. Untan*, vol. 04, no. 3, pp. 75-84, 2016.
- [5] F. J. Kaunang, "Penerapan Algoritma J48 Decision Tree Untuk Analisis Tingkat Kemiskinan di Indonesia," *CogITo Smart J.*, vol. 4, no. 2, p. 348, 2019.
- [6] Kusrini and Luthfi, *Algoritma Data Mining*. Penerbit Andi Yogyakarta, 2009.
- [7] D. T. Larose, *Discovering Knowledge in Data : An Introduction to Data Mining*. John Willey & Sons, Inc, 2005.
- [8] Herera and Francisco, *Data Mining and Soft Computing, Dept. of Computer Science and A.I. University of Granada, Spain.*, 2010.
- [9] U. Fayyad, *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- [10] S. Tan, P.N. and V. M., Kumar, *Introduction to Data Mining, 1st Ed, Pearson Education*. Boston San Fransisco New York., 2006.
- [11] Jiawei Han, Micheline Kembar, and J. Pie, *Data mining: Data mining concepts and techniques (third Edition)*. MK Morgan Kaufmann, 2014.
- [12] M. R., *Matematika Diskrit*. Bandung: Informatika Bandung, 2010.
- [13] S. Diwandari and N. A. Setiawan, "Perbandingan Algoritme J48 Dan Nbtreet Untuk Klasifikasi Diagnosa Penyakit Pada Soybean," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2015, no. Sentika, pp. 205-212, 2015.
- [14] A. Goyal and R. Mehta, "Performance comparison of Naïve Bayes and J48 classification algorithms," *Int. J. Appl. Eng. Res.*, vol. 7, no. 11 SUPPL., pp. 1389-1393, 2012.
- [15] E. Frank, M. Hall, G. Holmes, R. Kirkby, B. Pfahringer, and I. H. Witten, *Weka-a machine learning workbench for data mining n Data mining" and knowledge discovery handbook*. ed: Springer, 2009.
- [16] I. H. Witten, E. Frank, M. A. Hall, and and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.