

Pemanfaatan Algoritma Decision Tree ID3 Bagi Manajemen Bimbel Untuk Menentukan Faktor Kelulusan Pada Sekolah Kedinasan

Juwita Tetra Marani Aliyah Nazanah^{1*}, Muhammad Ihsan Jambak²

¹Fakultas Ilmu Komputer, Sistem Informasi, Universitas Sriwijaya, Palembang, Indonesia

²Fakultas Ilmu Komputer, Manajemen Informatika, Universitas Sriwijaya, Palembang, Indonesia

Email: ¹*09031382025152@student.unsri.ac.id, ²jambak@unsri.ac.id

Email Penulis Korespondensi: 09031382025152@student.unsri.ac.id

Abstrak—Saat ini bimbingan belajar kedinasan menjadi pilihan bagi para siswa untuk menunjang persiapan mereka ke jenjang pendidikan selanjutnya, tidak hanya sebatas membimbing belajar saja, biasanya manajemen bimbingan belajar menyiapkan strategi tertentu agar siswanya bisa lolos seleksi di kedinasan yang dituju. Sekolah kedinasan di Indonesia saat ini cukup populer dan diminati banyak siswa, hal ini disebabkan oleh keunggulan yang dimiliki sekolah kedinasan, yaitu biaya pendidikan yang relatif lebih murah bahkan gratis, dinaungi lembaga negara dan mempunyai kesempatan lebih besar untuk langsung bekerja setelah lulus. Minat yang tinggi ini menyebabkan tingginya persaingan masuk ke sekolah kedinasan. Salah satu kedinasan yang paling diminati saat ini adalah Politeknik Statistika STIS. Seleksi Pemilihan Mahasiswa Baru (SPMB) pada Politeknik Statistika STIS melewati banyak tahapan. Sehingga hal ini menjadi kepentingan bagi para manajemen bimbingan belajar untuk mengetahui faktor apa saja yang menentukan kelulusan pada seleksi tersebut. Ketidaktahuan manajemen bimbingan belajar dalam mengetahui faktor kelulusan dapat menyebabkan kurangnya efektif strategi dan proses belajar para siswa bimbingan belajar. Oleh karena itu, pada penelitian ini dilakukan penerapan ilmu data mining yaitu mengklasifikasikan data hasil SPMB Politeknik Statistika STIS tahun 2022 dengan algoritma decision tree ID3 yang bertujuan untuk mengetahui faktor utama yang menentukan mahasiswa yang lulus. Kemudian, hasil penelitian bisa menjadi pendukung keputusan manajemen bimbingan belajar dalam membuat strategi maupun evaluasi kedepannya. Sehingga siswa bimbingan belajar mendapatkan strategi pembinaan yang paling tepat dan sesuai berdasarkan hasil dari penelitian ini. Dataset dianalisis dengan Data Mining menggunakan Algoritma Decision Tree ID3. Berdasarkan penelitian yang dilakukan pada data adalah nilai Kappa sebesar 1.000, Accuracy sebesar 100% , Recall sebesar 100.000%, Classification Error sebesar 0.00 dan Precision sebesar 100% .

Kata Kunci: Data Mining; Decision Tree; Klasifikasi; Seleksi Kompetensi Dasar; ID3

Abstract—At present official tutoring is an option for students to support their preparation for the next level of education, not only limited to guiding study, usually tutoring management prepares certain strategies so that students can pass the selection at the intended official service. Official schools in Indonesia are currently quite popular and in demand by many students, this is due to the advantages of official schools, namely the cost of education is relatively cheaper and even free, under the auspices of state institutions and have a greater opportunity to work immediately after graduation. This high interest causes high competition to enter service schools. One of the most popular services today is the STIS Statistics Polytechnic. The New Student Selection (SPMB) at the STIS Statistics Polytechnic went through many stages. So that this is of interest to the tutoring management to find out what factors determine the passing of the selection. Ignorance of tutoring management in knowing the passing factor can lead to a lack of effective strategies and learning processes for tutoring students. Therefore, in this study the application of data mining science was carried out, namely classifying data from the 2022 STIS Statistics Polytechnic SPMB results using the ID3 decision tree algorithm which aims to find out the main factors that determine which students graduate. Then, the results of the research can be a support for tutoring management decisions in making strategies and future evaluations. So that tutoring students get the most appropriate and appropriate coaching strategy based on the results of this study. The dataset was analyzed by Data Mining using the ID3 Decision Tree Algorithm. Based on the research conducted on the data, the Kappa value is 1.000, the Accuracy is 100%, the Recall is 100.000%, the Classification Error is 0.00 and the Precision is 100% .

Keywords: Data Mining; Decision Tree; Classification; Basic Competency Selection; ID3

1. PENDAHULUAN

Saat ini sudah banyak bimbingan belajar yang dipercaya dapat menunjang persiapan siswanya agar bisa lulus ke kedinasan yang ingin dituju, bahkan sering ditemui bimbingan belajar yang berani menjamin kelulusan bagi siswa bimbingannya. Bimbingan belajar tentunya berani memberikan jaminan kelulusan tersebut karena sudah mempersiapkan strategi tertentu dengan riset menggunakan data yang aktual. Hal ini dilatarbelakangi karena sekolah kedinasan memiliki banyak keunggulan dibandingkan perguruan tinggi lainnya, sekolah kedinasan dinilai lebih unggul karena biaya pendidikan yang lebih terjangkau, prospek kerja terjamin, dinaungi lembaga negara dan dipersiapkan lebih mendalam dalam mengasah softskill masing-masing sehingga setelah lulus para lulusan sudah terampil dan siap bekerja. Karena keunggulan tersebut menyebabkan minat pada sekolah kedinasan menjadi meningkat dan tentunya memperketat persaingan dalam seleksi. Dengan demikian karena persaingan yang ketat tadi menimbulkan persaingan bagi manajemen pusat bimbingan untuk dapat mengetahui faktor-faktor apa saja yang menentukan kelulusan peserta sehingga bisa mengefektifkan cara belajar calon peserta seleksi.

Salah satu kedinasan dengan peminat terbanyak adalah Politeknik Statistika STIS. Proses seleksi dari Politeknik Statistika STIS memiliki banyak tahapan, antara lain Seleksi Kompetensi Dasar (SKD) yang merupakan tahapan seleksi berdasarkan parameter hasil penilaian dari Tes Karakter Kepribadian (TKP), Tes Wawasan Kewarganegaraan (TWK), dan Tes Intelegensi Umum (TIU). Dilanjutkan dengan seleksi matematika dan seleksi kesehatan. Permasalahannya, dengan banyaknya tahapan seleksi tersebut, menyebabkan proses seleksi tidak dapat ditebak secara langsung faktor utama penilaian didalamnya. Tentunya hal ini menjadi kepentingan bimbingan belajar untuk mengetahui hal apa yang menjadi

penentu utama kelulusan siswa bimbingannya agar dapat mempersiapkan strategi yang tepat dan sesuai untuk siswa bimbingannya. Terlebih lagi ada banyak tahapan pada seleksi ini dan belum diketahui faktor utama apa yang menentukan kelulusan siswa. Jika manajemen bimbingan belajar menganalisis data seleksi tersebut secara manual, tentunya akan sulit sekali dan memerlukan waktu yang lama karena banyaknya data yang ada. Sehingga diperlukan suatu pengetahuan sebagai pendukung keputusan strategi dalam pembinaan siswa bimbingan belajar untuk mengefektikan strategi belajar dan mengurangi potensi kegagalan siswa bimbingan belajar dengan memanfaatkan data yang sudah ada. Sebab itu, pada penelitian ini digunakan data hasil Seleksi Pemilihan Siswa Baru (SPMB) pada Politeknik Statistika STIS tahun 2022 untuk mendapatkan gambaran pengetahuan bagi manajemen bimbingan belajar agar dapat mempersiapkan strategi terbaik dari hasil pengetahuan yang ada dengan mengetahui faktor kelulusan pada seleksi kedinasan yang pada kasus ini adalah Politeknik Statistika STIS. Pada penelitian ini sendiri memanfaatkan ilmu data mining untuk mengambil pengetahuan yang ada dari data seleksi tersebut.

Data mining adalah analisis data (biasanya *big data*) untuk menemukan koneksi dan kesimpulan yang jelas yang sebelumnya tidak diketahui secara manual, dari data ini biasanya diketahui sebuah pola, kebiasaan dan perilaku tertentu dengan cara yang saat ini dapat dipahami dan bermanfaat bagi pemilik data [1]–[4]. Dari pengertian para ahli tersebut, dapat disimpulkan bahwa untuk mengolah data seleksi dengan jumlah yang cukup banyak pada penelitian ini bisa menggunakan data mining untuk mendapatkan pengetahuan didalam data hasil SPMB Politeknik Statistika STIS ini. Klasifikasi sendiri adalah salah satu metode dari data mining yang dapat dimanfaatkan dalam penelitian ini. Klasifikasi merupakan teknik data mining yang paling sederhana dan banyak digunakan [5]. Teknik klasifikasi juga terbukti sebagai teknik data mining yang paling bermanfaat dibidang data untuk Pendidikan[6]. Klasifikasi sendiri adalah proses mencari himpunan model (fungsi) sehingga dapat mendeskripsikan, membedakan kelas-kelas data atau konsep yang bertujuan untuk memprediksi data yang belum memiliki kelasnya[7]–[9]. Pada penelitian ini menggunakan algoritma data mining, yaitu *decision tree*. Metode ini sering digunakan dan hasilnya bisa dianalisis dengan mudah dan cukup banyak diimplementasikan di berbagai bidang. Decision Tree sendiri menghasilkan akurasi yang baik untuk mengklasifikasikan data dengan jumlah yang sangat banyak. Algoritma Decision Tree telah digunakan di beberapa penelitian *data mining*. Salah satunya, dilakukan penelitian oleh Siregar dkk [10] untuk mengklasifikasi mahasiswa berdasarkan kemampuan ekonomi untuk biaya kuliah. Dihasilkan tujuh kategori dari klasifikasi tersebut. Algoritma Decision Tree ini mendapatkan nilai akurasi sebesar 80.52%. Penelitian serupa sebelumnya [11] menggunakan Algoritma ID3 untuk melakukan Klasifikasi Data Calon Peserta Didik dan menghasilkan dua klasifikasi yaitu sebanyak 50,72% peserta didik diklasifikasikan tidak mampu dan sebanyak 49.27% diklasifikasikan mampu dari total 345 peserta didik. Penelitian serupa lainnya [12] menggunakan klasifikasi dengan Algoritma ID3 untuk mengetahui tingkat kepuasan pada sarana laboratorium komputer dan menghasilkan klasifikasi dengan nilai akurasi sebesar 0,95-1,00 yang sangat baik. Algoritma ID3 melakukan prosedur pencarian secara menyeluruh pada semua kemungkinan yang ada pada pohon keputusan [13]. Dengan algoritma ID3 kita dapat melihat faktor-faktor kemungkinan yang dapat mempengaruhi alternatif-alternatif keputusan tersebut ,disertai dengan mendapatkan estimasi akhir jika kita mengambil alternatif keputusan tersebut[14]. Berdasarkan beberapa hasil penelitian tersebut, maka penelitian ini dapat juga menggunakan algoritma Decision Tree ID3 untuk mendapatkan pengetahuan dari data hasil SPMB Politeknik Statistika STIS 2022

Sebagai langkah dan solusi yang diajukan untuk menyelesaikan uraian di atas, penelitian ini bertujuan melakukan penerapan ilmu data mining untuk mengklasifikasikan data hasil SPMB Politeknik Statistika STIS pada tahun 2022 agar dapat mengetahui faktor utama penentu karakteristik peserta yang lulus. Kemudian, hasil penelitian bisa menjadi pendukung keputusan manajemen bimbingan belajar dalam menentukan strategi pembinaan yang paling tepat dan sesuai kepada siswa bimbingan belajarnya. Sehingga siswa yang mendapatkan pembinaan sesuai dari bimbingan belajar diharapkan dapat meningkatkan kemampuan dan menerapkan strategi terbaik dari bimbingan belajar untuk dapat lulus ke kedinasan yang dituju. Pada akhirnya, hal ini dapat membangun tingkat kepercayaan para siswa dan orang tua siswa.kepada manajemen bimbingan belajar.

2. METODOLOGI PENELITIAN

2.1 Algoritma Decision Tree ID3

Decision Tree adalah pohon keputusan yang digunakan dalam analisis pemecahan masalah untuk memetakan berbagai alternatif pemecahan masalah. Decision tree adalah salah satu metode dari klasifikasi dengan menggunakan struktur pohon dimana setiap node mewakili atribut, cabangnya mewakili nilai dari atribut, dan daun mewakili kelas[11] . Decision tree dapat juga dikatakan salah satu metode klasifikasi yang paling sering digunakan karena mudah untuk dipahami oleh manusia [11], [15]. Konsep dasar decision tree adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan/rule . Pada decision tree terdapat 3 jenis node, yaitu:

1. Root Node, adalah node paling atas, tidak ada masukan pada node ini dan bisa tidak mempunyai keluaran atau memiliki keluaran lebih dari satu.
2. Internal Node, adalah node percabangan, hanya terdapat satu input dan mempunyai output minimal dua pada node ini.
3. Leaf Node atau Terminal Node, adalah node akhir, pada node ini hanya terdapat satu input dan tidak mempunyai output.

Algoritma ID3 atau Iterative Dichotomiser 3 merupakan salah satu algoritma yang dapat digunakan dalam membuat sebuah decision tree dengan cara mencari semua kemungkinan yang ada pada pohon keputusan dengan menggunakan struktur hierarki untuk pembelajaran supervised [11], [15], [16]. J. Ross Quinlan menemukan metode ini pada tahun 1979 dengan memanfaatkan Teori Informasi milik Shanon. Alur kerja dari decision tree dimulai dari root node hingga leaf node dan dilakukan secara rekursif, dimana pada setiap percabangannya menyatakan suatu kondisi yang harus dipenuhi dan disetiap ujung pohon menyatakan kelas dari suatu data sehingga bentuk data yang berupa tabel dirubah menjadi model pohon (tree), lalu model pohon yang diubah akan menjadi rule atau aturan yang dihasilkan [11].

2.2.1 Entropy dan Gain

Entropy adalah rumus yang digunakan sebagai alat menghitung homogenitas dari sebuah sample atau contoh [5]. Entropy merupakan suatu ukuran ketidakpastian yang diasosiasikan dengan variabel acak yang memiliki nilai antara 0-1 sebagai representasi dari informasi yang diketahui. Setelah mengetahui nilai entropy, kita akan mengetahui seberapa banyak informasi yang ada dari node tersebut.

Persamaan yang digunakan untuk mengetahui nilai entropy dari suatu data :

$$Entropy(S) = \sum_{i=1}^J - p_i \log_2 p_i \quad (1)$$

Keterangan :

S = Himpunan kasus

J = Banyaknya partisi S

Pi = Proporsi Si terhadap S

Setelah didapatkan nilai entropy pada kumpulan data, atribut akan dipilih dengan memperhatikan nilai information gain yang paling tinggi. Information Gain adalah penguatan informasi yang digunakan sebagai ukuran pemilihan atribut untuk dijadikan akar dalam pohon keputusan. Untuk mencari nilai dari information gain, menggunakan persamaan :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^J \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2)$$

Keterangan

S = Himpunan kasus

J = Banyaknya partisi S

A = Atribut

|Si| = Keseluruhan kasus untuk pada nilai i

|S| = Keseluruhan kasus pada S

Entropy |Si| = Entropi untuk sample – sample yang memiliki nilai i

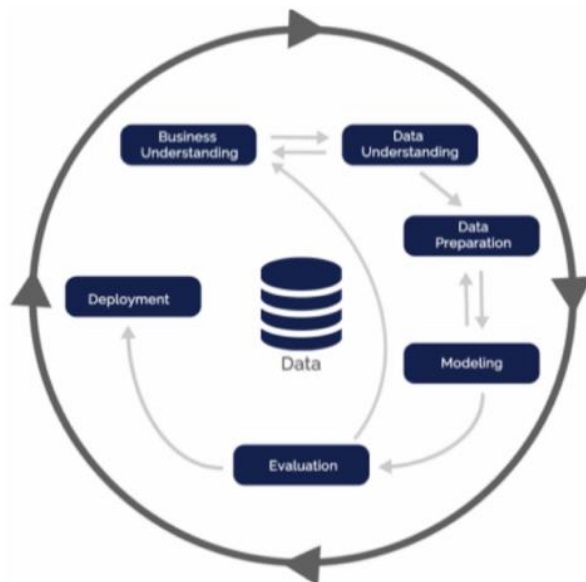
2.2 Alur Kerja Decision Tree ID3

Algoritma ID3 melakukan langkah pencarian pada semua kemungkinan pohon keputusan secara menyeluruh. Secara ringkas, langkah kerja algoritma ID3 adalah sebagai berikut [15]:

- Masukkan data training dan variabel
- Menghitung Entropy dan Information Gain dari setiap variabel tujuan dari pengukuran nilai information gain adalah untuk memilih variabel yang akan menjadi cabang saat pembentukan pohon keputusan.
- Gunakan variabel dengan nilai information gain yang paling tinggi.
- Bentuk simpul yang berisi variabel tersebut.
- Information gain akan terus diulangi proses perhitungannya sampai semua data telah termasuk dalam kelas yang sama. Variabel yang telah dipilih tidak diikuti lagi dalam perhitungan nilai Information gain.

2.3 CRISP DM

Pada penelitian ini menggunakan logika kerja data mining CRISP DM. CRISP DM dapat digunakan untuk proses standar yang cocok pada data mining yang prinsip utamanya adalah problem solving strategi. Oleh karena itu, penelitian ini menggunakan prosedur atau proses standar dengan CRISP DM sebagai logika logika berpikir data mining. CRISP DM sendiri memiliki enam alur yaitu, Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation dan Deployment [1].



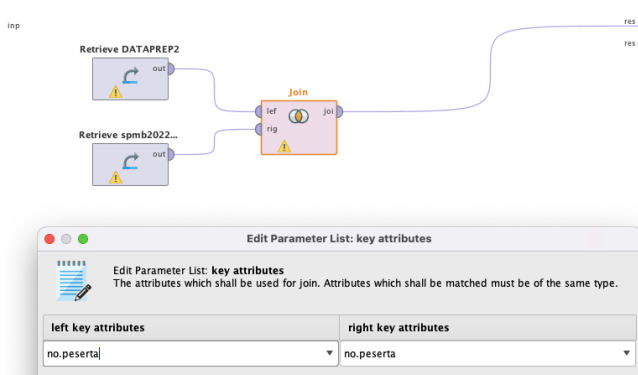
Gambar 1. Alur dari CRISP DM [17]

2.3.1 Business Understanding

Penelitian ini bertujuan untuk mengetahui karakteristik atau factor kelulusan dari siswa yang diterima pada proses seleksi, dalam penelitian ini ingin mencari tau karakteristik siswa yang lulus untuk menjadi evaluasi manajemen bimbingan belajar dalam mempersiapkan siswa bimbingan belajar agar bisa mengatur strategi kedepannya dari data SPMB Politeknik Statistika STIS.

2.3.2 Data Understanding

Pada penelitian ini data yang digunakan adalah data peserta SPMB Politeknik Statistika STIS 2022 dan merupakan data publik. Dataset terdiri dari lima file dataset yaitu data formasi peserta, data berisi skor dan hasil SKD, lalu data nilai matematika dan psikologi data, hasil akhir berdasarkan test kesehatan dan terakhir peserta yang lolos cadangan. Dengan data ini diharapkan dapat menentukan factor karakteristik penentu kelulusan. Parameter pemilihan kelulusan adalah SKD (SKD total, SKD TIU, SKD TWK, SKD TKP.), Akademik (Matematika), lokasi formasi. Dari lima file data set tersebut penulis menggabungkan 5 data tersebut menjadi satu kesatuan data yang sama, menggunakan fitur *join* pada rapid miner. Yang dimana data yang terpisah tersebut digabungkan menjadi satu berdasarkan no.peserta yang ada dalam setiap file. Lalu hasil dari join tersebut diberi nama DATAPREP2 dan dilakukan join ulang untuk data selanjutnya.



Gambar 2. Proses Operator Join pada Rapid Miner

Process dilakukan berulang sampai akhirnya semua file SKD, Nilai Matematika dan Hasil akhir digabungkan menjadi satu kesatuan untuk masuk dalam process cleansing dan agar lebih mudah diolah.

Tabel 1. Tabel Atribut Data SPMB Politeknik Statistika STIS

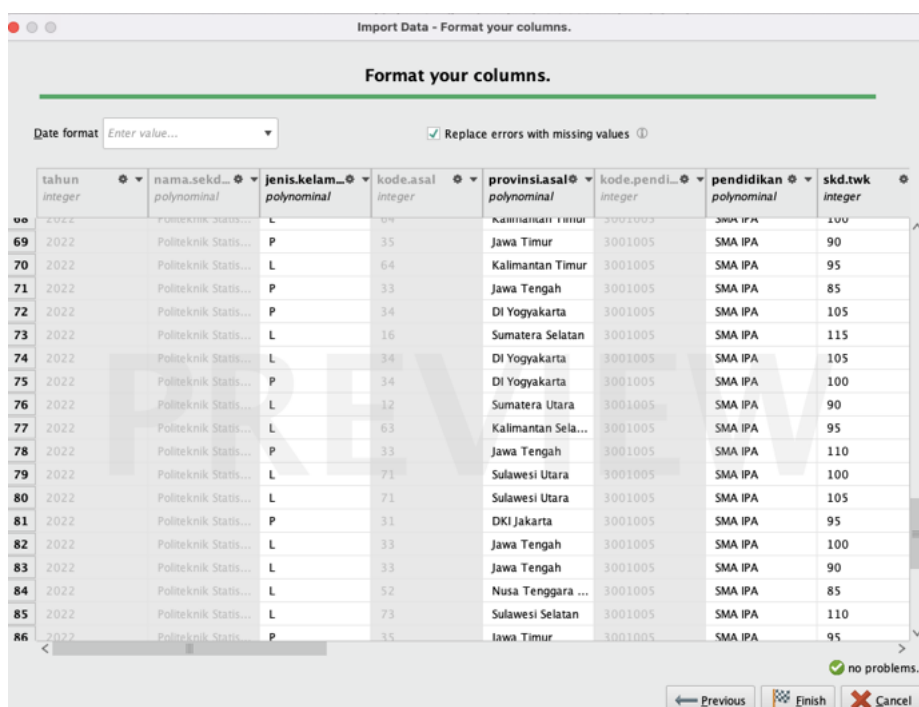
No	Atribut	Tipe Data
1	No.ujian	Integer (ID)
2	hasil.akhir	Polynomial (label)
3	Jenis.kelamin	Binominal
4	Skd.twk	Integer
5	Skd.tiu	Integer

6	Skd.tkp	Integer
7	Nama.prodi	Polynomial
8	Lokasi.formasi	Polynomial
9	Mtk.nilai	Integer
10	Tahun	Integer
11	Nama.sekdin	Polynomial
12	Kode.asal	Integer
13	Provinsi.asal	Polynomial
14	Kode.pendidikan	Integer
15	Pendidikan	Polynomial
16	skd.hasil	Integer
17	kode.prodi	Integer
18	psi.keterangan	Polynomial
19	mtk.nilai	Integer
20	mtk.keterangan	Polynomial
21	t2.keterangan	Polynomial
23	peringkat.cadangan	Integer
24	kode.formasi	Integer
25	no.peserta	Polynomial
26	skd.total	Integer

Pada penelitian ini agar dapat fokus maka data yang digunakan kami batasi sehingga data yang digunakan untuk pengklasifikasian hasil SPMB Politeknik Statistika STIS 2022 terdiri atas 735 data dengan 26 atribut. Untuk mengolah data tersebut proses klasifikasi dilakukan dengan menggunakan RapidMiner versi 10.

2.3.3 Data Preparation

Persiapan Data adalah memanipulasi dataset sebelum digunakan pada model. Salah satu persiapan data adalah pembersihan data (*Data Cleaning*) yaitu memperbaiki data yang tidak konsisten dan menghapus redundansi data. Disini penulis akan menghilangkan data dari file gabungan *data set* sebelumnya dengan cara meng-*exclude* kan data yang tidak mendukung untuk klasifikasi, seperti tahun, kode asal, kode Pendidikan, dll dan data yang bersifat duplikat pada RapidMiner seperti gambar dibawah ini. Proses ini dilakukan otomatis di *RapidMiner* untuk memudahkan dan mempercepat proses.



Gambar 3. Proses pembersihan data dengan cara meng-*exclude* kan atribut yang tidak perlu

Untuk table tahun di *exclude* dikarenakan tahun semua sama yaitu tahun 2022. Tabel nama sekdin juga diexclude dikarenakan semua sama yaitu Politeknik Statistika STIS. Kode Asal tidak diperlukan karena sudah ada provinsi asal. Kode Pendidikan tidak diperlukan dikarenakan sudah menggunakan table Pendidikan Kode.prodi tidak diperlukan karena sudah menggunakan kolom nama.prodi. Skd.hasil tidak diperlukan karena kita hanya membutuhkan hasil akhir. Psi.keterangan, mtk.keterangan, t2 keterangan tidak diperlukan karena akan terlihat di hasil akhir. Kolom peringkat

cadangan tidak diperlukan sehingga di exclude. Kolom kode formasi juga tidak diperlukan dikarenakan sudah menggunakan kolom lokasi formasi. Lalu, no.peserta tidak diperlukan karena sama dengan no ujian yang menjadi *id*. Berikut hasil data yang terpilih dari banyak-nya jenis formasi yang ada pada data peserta SPMB Politeknik Statistika STIS 2022 dan akan diujikan menggunakan *Data Mining*.

Tabel 2. Atribut data yang terpilih setelah pembersihan data

No	Atribut	Tipe Data
1	No.ujian	Integer (ID)
2	hasil.akhir	Polynomial (Label)
3	Skd.twk	Integer
4	Skd.tiu	Integer
5	Skd.tkp	Integer
6	Nama.prodi	Polynomial
7	Lokasi.formasi	Polynomial
8	Mtk.nilai	Integer

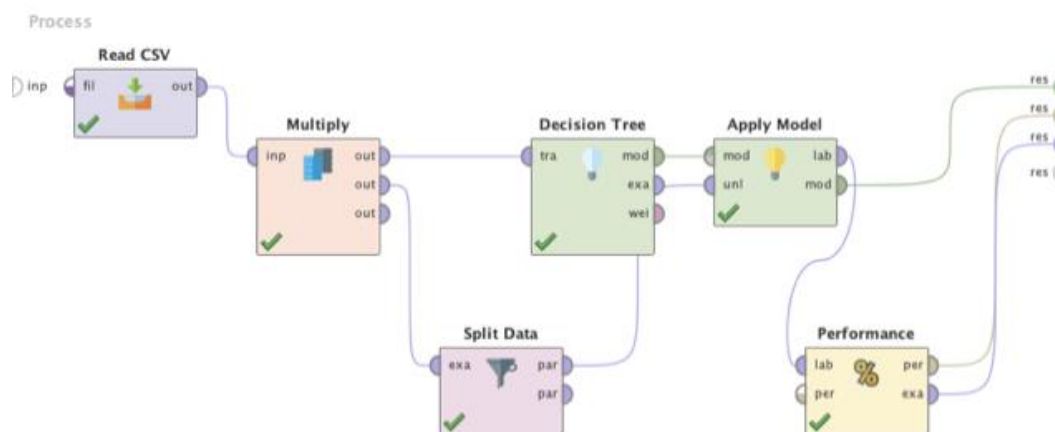
Dengan persiapan dan seleksi data ini memudahkan dan meringankan beban memori pada *Rapid Miner* sehingga dengan memilih kolom penting saja akan lebih mempercepat proses klasifikasi pada penelitian ini dan memudahkan kita untuk mendapatkan pengetahuan dari *dataset* tersebut.

3. HASIL DAN PEMBAHASAN

Fase ke-empat dalam CRISP-DM yaitu Modeling yaitu memasukkan data kepada algoritma *Decision Tree* mengingat metode klasifikasi adalah metode yang bersifat *supervised* maka data terlebih dahulu dilakukan split data untuk memisahkan data training dan data testing.

3.1 Modeling Algoritma Decision Tree ID3 pada Rapid Miner

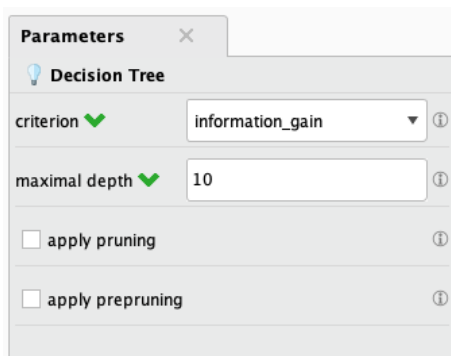
Pada Rapid Miner diterapkan model dengan menggunakan beberapa operator yaitu *Split Data*, *Decision Tree*, *Set Role*, *Split Data*, *Performance*, dan *Apply Model*. Model yang dibuat dapat dilihat pada gambar dibawah ini.



Gambar 4. Rancangan Model Algoritma Decision Tree untuk Klasifikasi SPMB dengan RapidMiner

Pada Gambar 4 terdapat beberapa operator yang diterapkan, seperti operator *Read CSV*, untuk membaca file CSV pada data penelitian ini. Operator *Split Data* digunakan untuk membagi data menjadi partisi data antara data *training* dan *testing* dengan rasio 0.9 dan 0.1. Operator *Decision Tree* adalah model yang diterapkan pada penelitian ini. Dengan *criterion* yang digunakan pada *Decision Tree* adalah *information gain* (ID3) dengan me-nonaktifkan *pruning* dan *prepruning*. Operator *Performance* diterapkan untuk mengevaluasi kinerja dari model yang digunakan penulis. Evaluasi kinerja yang digunakan yaitu, *Kappa*, *Accuracy*, *Recall* dan *Classification Error*. Terakhir operator *Apply Model* digunakan untuk menerapkan model *Decision Tree* pada dataset penelitian.

Algoritma ID3 pada Rapid Miner sendiri disebut Information Gain. Information Gain terletak pada parameter pada kriteria pada operator decision tree. Maka dari itu, kita harus memilih kriteria Information Gain untuk menggunakan Algoritma ID3 pada Rapid Miner. Kegunaan kriteria sendiri adalah memilih atribut untuk dilakukan pemisahan. Nilai yang dihasilkan pemisahan dioptimalkan sehubungan dengan masing-masing kriteria yang dipilih. Information Gain sendiri merupakan Entropi dari semua Atribut dihitung dan atribut dengan entropi terkecil dipilih untuk dipisahkan. Metode ini memiliki bias terhadap pemilihan Atribut dengan jumlah nilai yang banyak.

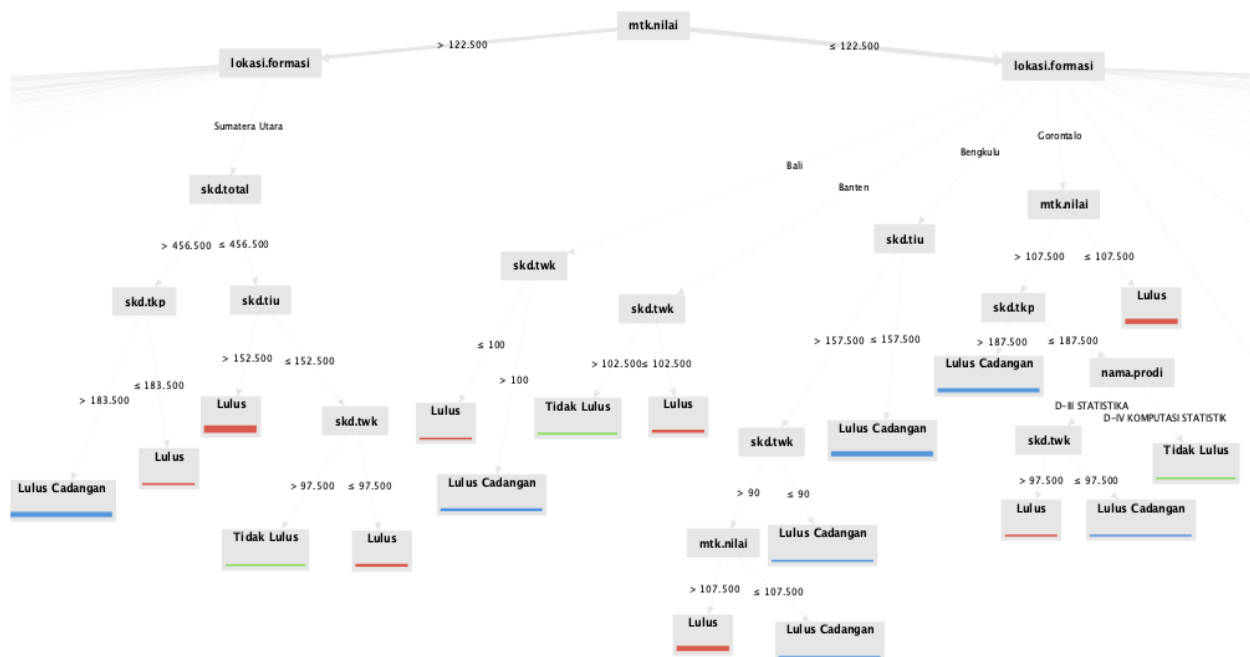


Gambar 5. Pemilihan Kriteria Information Gain pada RapidMiner

Pada operator algoritma decision tree ini, kita tidak perlu menggunakan apply pruning dan apply prepruning karena hasilnya pada data ini akan lebih maksimal.

Row No.	no.ujian	hasil.akhir	prediction(...)	confidence(...)	confidence(...)	confidence(...)	jenis.kelam...	provinsi.asal	pendidikan	skd.twk	skd.tiu
1	1102897	Lulus Cadan...	Lulus Cadan...	1	0	0	L	Aceh	MA IPA	105	155
2	1800331	Tidak Lulus	Tidak Lulus	0	1	0	P	Lampung	SMA IPA	110	140
3	3305159	Lulus	Lulus	0	0	1	P	Jawa Tengah	SMA IPA	95	150
4	1205569	Lulus Cadan...	Lulus Cadan...	1	0	0	L	Sumatera Ut...	MA IPA	110	140
5	1210637	Lulus	Lulus	0	0	1	P	Sumatera Ut...	SMA IPA	80	170
6	1201444	Lulus	Lulus	0	0	1	P	Sumatera Ut...	SMA IPA	95	140
7	1104727	Lulus Cadan...	Lulus Cadan...	1	0	0	L	Aceh	SMA IPA	95	120
8	1201020	Lulus Cadan...	Lulus Cadan...	1	0	0	P	Sumatera Ut...	SMA IPA	100	165
9	1200797	Lulus	Lulus	0	0	1	L	Sumatera Ut...	SMA IPA	95	145
10	1202676	Lulus	Lulus	0	0	1	P	Sumatera Ut...	SMA IPA	100	140
11	1301399	Lulus	Lulus	0	0	1	P	Sumatera Ba...	SMA IPA	110	145
12	1308138	Lulus	Lulus	0	0	1	P	Sumatera Ba...	SMA IPA	115	125
13	1302457	Tidak Lulus	Tidak Lulus	0	1	0	L	Sumatera Ba...	SMA IPA	115	145
14	1300671	Lulus	Lulus	0	0	1	L	Sumatera Ba...	SMA IPA	95	135
15	1308819	Lulus Cadan...	Lulus Cadan...	1	0	0	P	Sumatera Ba...	SMA IPA	95	145
16	3208524	Lulus	Lulus	0	0	1	P	Jawa Barat	SMA IPA	100	150
17	1403581	Lulus Cadan...	Lulus Cadan...	1	0	0	P	Riau	MA IPA	110	150
18	1201255	Lulus	Lulus	0	0	1	P	Sumatera Ut...	SMA IPA	95	150

Gambar 6. Tabel Hasil Prediksi pada RapidMiner



Gambar 7. Hasil dari Klasifikasi menggunakan Algoritma Decision Tree pada SPMB Politeknik Statistika STIS

Pada struktur pohon pembagian kelas menjadi Lulus, Lulus Cadangan dan Tidak Lulus berdasarkan nilai MTK, formasi dan TIU, TWK, TKP, dan Total.

Pada penelitian ini, diketahui faktor-faktor utama penentu karakteristik dari mahasiswa yang lulus adalah nilai matematika, lokasi formasi dan diikuti dengan SKD tertentu sesuai dengan lokasi formasi. Hal ini dapat dilihat dari akar pertama dan kedua dari pohon keputusan atas nilai matematika dan lokasi formasi.

Seperti yang diketahui dari hasil klasifikasi dapat dilihat bahwa faktor utama sekali dalam kelulusan adalah nilai matematika, hal ini sejalan dengan maksud dari Politeknik Statistika STIS itu sendiri. Karena ilmu yang didalami di Politeknik Statistika STIS ini sendiri adalah ilmu statistika yang mana berkecimpungan dalam dunia matematika. Sehingga bisa dijadikan evaluasi untuk para manajemen bimbingan belajar untuk lebih mengarahkan siswanya untuk fokus kepada pembelajaran dan pemberian materi terhadap konsep-konsep dari matematika itu sendiri agar nilai matematika para siswa bisa lebih meningkat dan memperbesar peluang kelulusan.

Lalu, factor kedua adalah lokasi formasi, hal ini disebabkan oleh jumlah dan kuota formasi yang ada sehingga bisa menjadi acuan strategi kedua bagi para manajemen bimbingan belajar untuk mengevaluasi apakah standarisasi penilaian sudah sesuai dengan lokasi formasi dan menjadi pendukung keputusan untuk penilaian dan penyusunan strategi yang tepat kedepannya. Sehingga ditemukan jawaban, mengapa ada skor tertentu yang lebih rendah bisa lolos, tetapi skor yang lebih besar tidak lolos, hal ini disebabkan oleh lokasi formasi karena dipengaruhi oleh banyaknya jumlah formasi. Diketahui juga dari klasifikasi diatas disetiap lokasi formasi ternyata memiliki faktor penilaian yang berbeda lagi sesuai dengan lokasi yang dipilih.

3.2 Rule Base

Rule base adalah hasil deskripsi dari sebuah *decision tree* yang berhasil dibuat. Rule base ini membantu para pengguna dan memudahkan pengguna dalam membaca hasil dari data yang di klasifikasi [11].

Setelah data diolah dan dijalankan di Rapid Miner dengan menggunakan Algoritma Decision Tree ID3 maka hasil rule base yang didapatkan adalah sebagai berikut :

```
mtk.nilai > 122.500
| lokasi.formasi = Aceh
| | nama.prodi = D-III STATISTIKA
| | | skd.twk > 102.500: Tidak Lulus {Lulus Cadangan=0, Tidak Lulus=1, Lulus=0}
| | | skd.twk ≤ 102.500: Lulus {Lulus Cadangan=0, Tidak Lulus=0, Lulus=3}
| | nama.prodi = D-IV KOMPUTASI STATISTIK
| | | skd.tkp > 183
| | | | skd.tkp > 190: Lulus {Lulus Cadangan=0, Tidak Lulus=0, Lulus=4}
| | | | skd.tkp ≤ 190
| | | | | skd.twk > 105: Lulus {Lulus Cadangan=0, Tidak Lulus=0, Lulus=2}
| | | | | skd.twk ≤ 105: Tidak Lulus {Lulus Cadangan=0, Tidak Lulus=2, Lulus=0}
| | | | skd.tkp ≤ 183
| | | | | skd.twk > 97.500: Lulus {Lulus Cadangan=0, Tidak Lulus=0, Lulus=1}
| | | | | skd.twk ≤ 97.500: Lulus Cadangan {Lulus Cadangan=2, Tidak Lulus=0, Lulus=0}
| | | nama.prodi = D-IV STATISTIKA: Lulus {Lulus Cadangan=0, Tidak Lulus=0, Lulus=11}
| | lokasi.formasi = Bali: Lulus {Lulus Cadangan=0, Tidak Lulus=0, Lulus=2}
| | lokasi.formasi = Banten
| | | skd.tkp > 193
| | | | skd.twk > 120: Lulus {Lulus Cadangan=0, Tidak Lulus=0, Lulus=1}
| | | | skd.twk ≤ 120: Lulus Cadangan {Lulus Cadangan=2, Tidak Lulus=0, Lulus=0}
| | | | skd.tkp ≤ 193
| | | | | skd.twk > 117.500
| | | | | | skd.tkp > 183.500: Tidak Lulus {Lulus Cadangan=0, Tidak Lulus=2, Lulus=0}
| | | | | | skd.tkp ≤ 183.500: Lulus {Lulus Cadangan=0, Tidak Lulus=0, Lulus=1}
| | | | | skd.twk ≤ 117.500: Lulus {Lulus Cadangan=0, Tidak Lulus=0, Lulus=2}
| | lokasi.formasi = Bengkulu
| | | mtk.nilai > 132.500
| | | | skd.total > 446
| | | | | skd.total > 452: Lulus {Lulus Cadangan=0, Tidak Lulus=0, Lulus=4}
| | | | | skd.total ≤ 452: Tidak Lulus {Lulus Cadangan=0, Tidak Lulus=2, Lulus=0}
... dst
```

3.3 Evaluasi Performa Model Decision Tree ID3

Nilai performa akan didapat dari model *Decision Tree* yang dirancang untuk mengklasifikasikan hasil SPMB. Performa dilihat menurut *Confusion Matrix*. Diterapkan beberapa performa antara lain: *recall*, *classification error*, *kappa*, *accuracy* dan *error*. Berikut *Confusion Matrix* yang dihasilkan.

Tabel 3. Tabel *Confusion Matrix* pada penelitian

	true Lulus	true Tidak Lulus	true Lulus Cadangan	class precision
gred. Lulus	438	0	0	100%
gred. Tidak Lulus	0	91	0	100%
gred. Lulus Cadangan	0	0	206	100%
Class recall	100%	100%	100%	

Berdasarkan [18], [19] untuk menghitung accuracy berdasarkan *Confussion Matrix* adalah sebagai berikut :

a. Accuracy

$$\frac{(TP+TN)}{TP+TN+FP+FN} \times 100\% = \frac{438+91+206}{438+0+0+0+91+0+0+0+206} \times 100\% = \frac{735}{735} \times 100\% = 1 \times 100\% = 100\%$$

b. Classification Error

$$1 - \max \frac{735}{735} = 1 - 1 = 0$$

c. Precision

Precision Lulus

$$\frac{TP}{TP+FP} \times 100\% = \frac{438}{438+0+0} \times 100\% = 1 \times 100\% = 100\%$$

Precision Tidak Lulus

$$\frac{TP}{TP+FP} \times 100\% = \frac{91}{91+0+0} \times 100\% = 1 \times 100\% = 100\%$$

Precision Lulus Cadangan

$$\frac{TP}{TP+FP} \times 100\% = \frac{206}{206+0+0} \times 100\% = 1 \times 100\% = 100\%$$

d. Recall

$$\frac{TP}{TP+FN} \times 100\% = \frac{438}{438+0+0} \times 100\% = \frac{438}{438} \times 100\% = 1 \times 100\% = 100\%$$

e. Kappa Accuracy

Perkalian silang data

$$= (438 \times 438) + (91 \times 91) + (206 \times 206) = 191844 + 8281 + 42436 = 242561$$

Kappa Accuracy

$$\frac{735 \times 735 - 242561}{|(735^2) - 242561|} \times 100\% = 1 \times 100\% = 100\%$$

Nilai performa *accuracy* adalah nilai keakuratan pada model yang diterapkan. Didapatkan nilai akurasi 100% yang mana artinya model sangat baik untuk hasil prediksi. Nilai performa *Classification error* adalah kesalahan pada nilai yang didapatkan dari model yang diterapkan. Performa *kappa* adalah normalisasi dari nilai *accuracy*. Didapatkan Nilai *Kappa* yaitu 1 yang artinya model ini bisa disebut sempurna. Selain itu, pada model penelitian ini didapatkan nilai kesalahan 0% dimana artinya kesalahan tidak ada pada model yang digunakan. Nilai *Precision* dan *Recall* yang didapatkan adalah 100%.

4. KESIMPULAN

Berdasarkan hasil penelitian dengan menggunakan model *Decision Tree* pada klasifikasi hasil SPMB Politeknik Statistika STIS untuk mengetahui faktor utama yang menyebabkan mahasiswa yang lulus. Pada penelitian ini, berkaitan dengan rumusan masalah maka kesimpulan yang dapat diambil adalah sebagai berikut, diketahui bahwa faktor utama dari kelulusan mahasiswa di STIS terdapat pada nilai matematika, lalu disusul dengan lokasi formasi. Hal ini dikarenakan nilai matematika dan lokasi formasi menjadi akar pertama dan akar kedua pada akar *decision tree* tersebut. Sehingga ditemukan jawaban, mengapa ada skor tertentu yang lebih rendah bisa lolos, tetapi skor yang lebih besar tidak lolos, hal ini disebabkan oleh lokasi formasi karena dipengaruhi oleh banyaknya jumlah formasi. Diketahui juga dari klasifikasi diatas disetiap lokasi formasi ternyata memiliki faktor penilaian yang berbeda lagi sesuai dengan lokasi yang dipilih. Dimana hasil penelitian dengan model *Decision Tree ID3* ini terbukti menghasilkan performa akurasi dengan nilai *Classification Error* = 0%, *Kappa* = 1.000, *Accuracy* = 100%, dan *Recall* = 100% yang artinya akurat berdasarkan nilai evaluasi pada model penelitian ini. Diharapkan manajemen bimbingan belajar bisa mendapatkan gambaran lebih untuk

mengarahkan dan mengevaluasi pembelajaran siswa bimbingan belajar agar sesuai dengan standar penilaian lokasi formasi yang dipilih siswa.

Selain itu, penelitian ini dapat dikembangkan melalui implementasi algoritma *data mining* lainnya dan nilai performanya dapat dibandingkan satu sama lain, sehingga dapat diketahui performa dari algoritma lainnya dan bisa dijadikan rujukan untuk penelitian berikutnya.

REFERENCES

- [1] M. North, *Data mining for the masses*, Second., vol. second edition. 2016.
- [2] C. C. Aggarwal, *Data Mining*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8.
- [3] J. Duque, F. Silva, and A. Godinho, "Data Mining applied to Knowledge Management," *Procedia Comput Sci*, vol. 219, pp. 455–461, 2023, doi: 10.1016/j.procs.2023.01.312.
- [4] P. Bidang Komputer Sains dan Pendidikan Informatika, D. Akademi Perekam dan Informasi Kesehatan Iris Padang Jl Gajah Mada No, and S. Barat, "Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Yuli Mardi".
- [5] R. D. Syah, "METODE DECISION TREE UNTUK KLASIFIKASI HASIL SELEKSI KOMPETENSI DASAR PADA CPNS 2019 DI ARSIP NASIONAL REPUBLIK INDONESIA," *Jurnal Ilmiah Informatika Komputer*, vol. 25, no. 2, pp. 107–114, 2020, doi: 10.35760/ik.2020.v25i2.2750.
- [6] A. B. E. D. Ahmed and I. S. Elaraby, "Data Mining: A prediction for Student's Performance Using Classification Method," *World Journal of Computer Application and Technology*, vol. 2, no. 2, pp. 43–47, Feb. 2014, doi: 10.13189/wjcat.2014.020203.
- [7] A. H. Nasrullah, "Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out," *ILKOM Jurnal Ilmiah*, vol. 10, 2018.
- [8] A. H. Nasrullah, "IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS," vol. 7, no. 2, 2021, [Online]. Available: <http://ejournal.fikom-unasman.ac.id>
- [9] S. Bahri and A. Lubis, "METODE KLASIFIKASI DECISION TREE UNTUK MEMPREDIKSI JUARA ENGLISH PREMIER LEAGUE," vol. 2, no. 1, 2020.
- [10] B. Siregar, E. Budhiarti Nababan, N. Sagala, U. Andayani, and Fahmi, "Tuition Single Classification using Decision Tree Method and C4.5," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2019. doi: 10.1088/1742-6596/1175/1/012105.
- [11] A. Sifaunajah and R. D. Wahyuningtyas, "Penggunaan Algoritma ID3 Untuk Klasifikasi Data Calon Peserta Didik," *CSRID (Computer Science Research and Its Development Journal)*, vol. 14, no. 2, p. 103, Sep. 2022, doi: 10.22303/csrid.14.2.2022.103-112.
- [12] I. R. Munthe and V. Sihombing, "Klasifikasi Algoritma Iterative Dichotomizer (ID3) untuk Tingkat kepuasan pada Sarana Laboratorium Komputer."
- [13] D. Kinasih Widiyati, M. Wati, and H. Santoso Pakpahan, "Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara," 2018.
- [14] I. Srimenganti et al., *Implementasi Algoritma Decision Tree (ID3) Untuk Penyakit Campak*. SENTER 2018: Seminar Nasional Teknik Elektro 2018, 2018.
- [15] D. Kinasih Widiyati, M. Wati, and H. Santoso Pakpahan, "Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara," 2018.
- [16] I. R. Munthe and V. Sihombing, "Klasifikasi Algoritma Iterative Dichotomizer (ID3) untuk Tingkat kepuasan pada Sarana Laboratorium Komputer," *Jurnal Teknologi dan Ilmu Komputer Prima (JUTIKOMP)*, vol. 1, no. 2, pp. 27–34, Oct. 2018, doi: 10.34012/jutikomp.v1i2.237.
- [17] Y. Suhandi, I. Kurniati, and S. Norma, "Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik," *Jurnal Teknologi Informatika dan Komputer*, vol. 6, no. 2, pp. 12–20, Sep. 2020, doi: 10.37012/jtik.v6i2.299.
- [18] P. Kasih, "Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara," *Innovation in Research of Informatics (INNOVATICS)*, vol. 1, no. 2, Oct. 2019, doi: 10.37058/innovatics.v1i2.918.
- [19] R. Iriane, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Pangan Hewan Menggunakan Metode K-Nearest Neighbor," *KLIK: KAJIAN ILMIAH INFORMATIKA DAN KOMPUTER*, vol. 3, no. 5, pp. 509–515, 2023, [Online]. Available: <https://djournals.com/klik>