



Analisis Pola Minat Siswa Lulusan SMU/SMK Untuk Melanjutkan Kuliah dengan Menggunakan Algoritma C4.5

Yuli Septya Ayunda, Dedy Hartama, Muhammad Ridwan Lubis, Indra Gunawan, Muhammad Rafai*

Teknik Informatika, STIKOM Tunas Bangsa, Pematang Siantar, Indonesia

Email: ¹yuliseptyaayunda@gmail.com, ²dedyhartama@amiktunasbangsa.ac.id, ³muhammadirdwanlubis76@gmail.com,

⁴indra@amiktunasbangsa.ac.id, ⁵*rafaimhd123@gmail.com

Email Penulis Korespondensi: rafaimhd123@gmail.com

Abstrak—STIKOM Tunas Bangsa merupakan salah satu perguruan tinggi swasta terbaik yang ada di kota Pematangsiantar. Dengan banyaknya perguruan tinggi yang ada di kota Pematangsiantar tentunya menimbulkan persaingan dalam menarik minat calon mahasiswa baru. Agar dapat tetap mempertahankan jumlah mahasiswa tiap tahunnya pihak perguruan tinggi harus tahu faktor apa yang menjadi pendorong utama bagi calon mahasiswa/i untuk melanjutkan studinya di STIKOM Tunas Bangsa. Tujuan penelitian ini adalah untuk mengetahui faktor utama siswa lulusan SMU/SMK dalam melanjutkan studi di STIKOM Tunas Bangsa. Data diperoleh dari kuisioner yang kemudian diolah menggunakan data Mining dengan Algoritma C4.5 dan diuji dengan software Rapid Miner. Berdasarkan hasil penelitian dapat disimpulkan bahwa ada enam (6) pola yang dihasilkan pada kasus peminatan terhadap calon mahasiswa yang akan berkuliah di STIKOM Tunas Bangsa, dimana tiga (3) dari rules tersebut menghasilkan keputusan minat dan tiga (3) rules menghasilkan keputusan tidak minat. Salah satu rules yang dihasilkan ialah Jika Rekomendasi = Pernah, dan Kesesuaian Jurusan = Sesuai, maka hasilnya Minat. Rekomendasi dari orang sekitar menjadi faktor pendukung utama untuk siswa lulusan SMU/SMK untuk melanjutkan kuliah di STIKOM Tunas Bangsa Pematangsiantar.

Kata Kunci: Pola Minat; C4.5; Data Mining; Rapid Miner; Analisis

Abstract—STIKOM Tunas Bangsa is one of the best private universities in the city of Pematangsiantar. With so many universities in the city of Pematangsiantar, it certainly creates competition in attracting prospective new students. In order to maintain the number of students each year, universities must know what factors are the main drivers for prospective students to continue their studies at STIKOM Tunas Bangsa. The aim of this research is to determine the main factors for high school/vocational school graduate students in continuing their studies at STIKOM Tunas Bangsa. Data was obtained from questionnaires which were then processed using data mining with the C4.5 algorithm and tested with Rapid Miner software. Based on the research results, it can be concluded that there are six (6) patterns produced in cases of interest in prospective students who will study at STIKOM Tunas Bangsa, where three (3) of these rules result in interest decisions and three (3) rules result in no interest decisions. One of the resulting rules is if Recommendation = Never, and Suitability of Department = Appropriate, then the result is Interest. Recommendations from local people are the main supporting factor for high school/vocational school graduate students to continue studying at STIKOM Tunas Bangsa Pematangsiantar.

Keywords: Interest Patterns; C4.5; Data Mining; Rapid Miner; Analysis

1. PENDAHULUAN

Pendidikan memiliki peranan yang sangat penting bagi suatu bangsa, biasanya tingkat kemajuan suatu bangsa dipengaruhi oleh kualitas pendidikan bangsa tersebut. Pendidikan akan menghasilkan sumber daya manusia yang berkualitas baik dari segi spiritual, intelegensi, kemampuan dan pendidikan adalah sebuah proses untuk menghasilkan generasi penerus bangsa (Anggraeni et al., 2021; M.SYARIFUDDIN YUSUF, 2021; Nathan & Scobell, 2012; Yekti Pawestri, 2022). Pendidikan tinggi ialah jenjang pendidikan setelah sekolah menengah yang meliputi program diploma, program sarjana, program magister, program doktor, program profesi, serta program spesialis, yang diselenggarakan oleh perguruan tinggi. Perkembangan sebuah perguruan tinggi, salah satunya dilihat berdasarkan jumlah mahasiswa yang diperoleh setiap tahunnya. Perguruan tinggi terutama untuk swasta, perolehan mahasiswa setiap tahunnya menjadi faktor utama dalam berkembangnya perguruan tinggi tersebut (Adenuddin Alwy, 2022; Joni et al., 2021; Ninh et al., 2020; Ramdhan & Nufriana, 2019; Sholihin & Ariyani, 2023).

Penerimaan mahasiswa/i baru merupakan kegiatan tahunan yang dilaksanakan setiap perguruan tinggi dan perguruan tinggi swasta. Kegiatan penerimaan mahasiswa/i baru tentu saja merupakan bagian terpenting dalam kehidupan perguruan tinggi untuk mempertahankan jumlah mahasiswa/i serta meningkatkan calon mahasiswa/i. Kota Pematangsiantar memiliki banyak perguruan tinggi, dimana hal tersebut menjadi tantangan bagi tiap-tiap perguruan tinggi dalam menarik minat calon mahasiswa/i untuk mendaftar di perguruan tinggi tersebut karena dengan banyaknya perguruan tinggi yang ada tentu saja calon mahasiswa/i juga memiliki banyak pilihan untuk menentukan dimana mereka akan berkuliah. Sangat penting untuk pihak kampus untuk mengetahui pola minat siswa lulusan SMU/SMK untuk melanjutkan kuliah serta faktor utama yang mendorong minat calon mahasiswa/i untuk berkuliah sehingga mampu bersaing dengan perguruan tinggi lainnya. Data yang digunakan pada penelitian ini adalah data yang diperoleh dari kuisioner yang dibagikan penulis kepada siswa SMU/SMK dengan jumlah 80 responden. Penyelesaian masalah pada penelitian ini penulis menggunakan algoritma C4.5 (Abdelsamea et al., 2022; Alam Jusia et al., 2019; Aswathy & Jagannath, 2020; Dharma et al., 2019; Ouddai et al., 2023; Seyedan & Mafakheri, 2020; Tahirović & Krivić, 2023; Yuli Mardi, 2019). Algoritma C4.5 merupakan salah satu metode dalam membuat decision tree berdasarkan training sample yang telah tersedia. Algoritma C4.5 memiliki kelebihan pada klasifikasi pohon keputusan dari pada algoritma lainnya. Algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah

memberikan kesan, memiliki tingkat akurasi yang tepat dan dapat diterima, efisien dalam menangani atribut bertipe diskret serta numerik (Dr. V. Suma, 2020; Hakkoum et al., 2023; Srisook et al., 2022; Tolan et al., 2023; Zeng et al., 2019).

Beberapa penelitian sebelumnya yang menjadi rujukan penelitian ini diantaranya : Penelitian Prediksi Minat Calon Mahasiswa memilih perguruan Tinggi Menggunakan Algoritma C4.5 dari 25 data sampel sebagai training dataset dengan 5 data penilaian. Hasil dari prediksi pada penelitian ini menghasilkan akurasi 100% dengan 120 testing dataset. Penelitian kecenderungan calon mahasiswa memilih perguruan tinggi dengan menggunakan metode analisis faktor. Berdasarkan analisis dan pembahasannya menyimpulkan bahwa dari 25 variabel menghasilkan 5 faktor yang mewakili mereka , yakni faktor kepercayaan terhadap lembaga, faktor biaya pendidikan, faktor layanan dan fasilitas, faktor promosi internal, serta faktor aksesibilitas.

Dari penelitian ini nantinya akan memberikan model pola dalam bentuk decision tree agar dapat mengetahui model marketing terhadap calon mahasiswa/i yang akan mendaftar ke STIKOM Tunas Bangsa dengan melihat faktor yang paling dominan.

2. METODOLOGI PENELITIAN

2.1 Pendahuluan

Metodologi penelitian ini akan menggambarkan bagaimana pengumpulan data yang akan digunakan. Pengumpulan data yang akan dibahas diantaranya lokasi penelitian, waktu penelitian, data yang akan digunakan, dan proses pengolahan data. Proses penelitian akan digambarkan dalam bentuk flowchart dan di perkuat dengan penjelasan terkait dengan penelitian yang dilakukan.

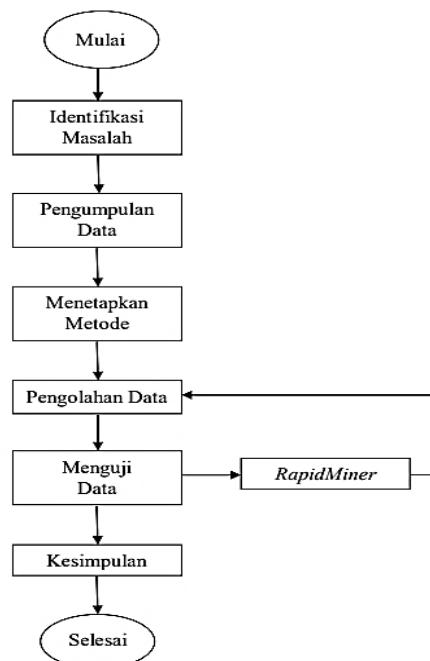
Tahap ini dimulai dengan data yang kita miliki. Data yang terkumpul dalam bentuk spreadsheet file excel 2007.xls dari kuisioner yang telah dibagikan penulis. Tujuan dari penelitian ini ialah untuk merancang sebuah data mentah menjadi kumpulan data olahan yang sesuai dengan keinginan dengan menggunakan algoritma C4.5 untuk melihat pola minat lulusan SMU/SMK dalam melanjutkan kuliah.

2.1.1 Lokasi dan Waktu Penelitian

Penelitian ini dilakukan pada STIKOM Tunas Bangsa yang beralamat di Jalan jenderal Sudirman No.1,2,&3, Kecamatan Siantar Barat, Kota Pematangsiantar, Provinsi Sumatera Utara. Penulis melakukan riset selama 10 hari mulai 09 Juni 2022 sampai dengan 18 Juni 2022.

2.1.2 Rancangan Penelitian

Rancangan penelitian digunakan untuk menjelaskan proses kerja dalam menyelesaikan masalah penelitian. Alur penelitian ini akan digambarkan dalam sebuah rancangan flowchart pada gambar 1 berikut:



Gambar 1. Rancangan Penelitian

Gambar 1 menjelaskan rancangan penelitian untuk menentukan pola minat siswa lulusan SMU/SMK untuk melanjutkan kuliah dengan menggunakan metode C4.5 yang terdiri dari :



1. Identifikasi Masalah
Mengidentifikasi masalah yang terkait dengan pola minat siswa lulusan SMU/SMK dalam melanjutkan kuliah.
2. Pengumpulan Data
Data pada penelitian ini diperoleh dengan cara membagikan kuisioner yang dibagikan kepada para siswa lulusan SMU/SMK.
3. Menetapkan Metode
Menetapkan metode untuk mencari solusi masalah. Pada penelitian ini penulis menggunakan teknik data mining dengan algoritma C4.5.
4. Pengolahan Data
Data yang sudah dikumpul akan diolah agar dapat digunakan
5. Menguji Data
Pengujian Data dilakukan dengan menggunakan software RapidMiner versi 5.3
6. Kesimpulan
Kesimpulan didapat dari penyelesaian masalah yang menjadi hasil dari penelitian ini

Berikut ini adalah data yang didapat dari kuisioner yang telah dibagikan oleh penulis kepada siswa lulusan SMU/SMK dalam bentuk .xls yang dapat dilihat pada tabel 1 dibawah ini:

Tabel 1. Data Awal Penelitian

CM	Kesesuaian Dengan Jurusan	Akreditasi	Biaya Kuliah	Prospek Bekerja	Fasilitas	Lokasi	Rekomen dasi	Minat
1	Tidak Sesuai	Unggul	Mahal	Ya	Mendukung	Dekat	Tidak	Tidak
2	Sesuai	Unggul	Mahal	Ya	Mendukung	Dekat	Tidak	Tidak
3	Sesuai	Unggul	Terjangkau	Ya	Mendukung	Dekat	Tidak	Tidak
4	Sesuai	Unggul	Terjangkau	Ya	Mendukung	Jauh	Pernah	Tidak
5	Sesuai	Unggul	Terjangkau	Ya	Mendukung	Dekat	Pernah	Tidak
....
80	Tidak Sesuai	Unggul	Terjangkau	Ya	Mendukung	Dekat	Pernah	Minat

Data penelitian yang akan digunakan menggunakan Algoritma C4.5 dari tabel 1 tersebut dapat dilihat pada tabel 2 dibawah ini:

Tabel 2. Dataset Penelitian

Field Data	Kategorikal Data	Type Data
Minat	Label	Nominal
Kesesuaian dengan jurusan	Atribut	Nominal
Akreditasi	Atribut	Nominal
Biaya Kuliah	Atribut	Nominal
Prospek Kerja	Atribut	Nominal
Fasilitas	Atribut	Nominal
Lokasi	Atribut	Nominal
Rekomendasi	Atribut	Nominal

2.1.3 Prosedur Pengumpulan Data

Dalam upaya untuk memperoleh data dan informasi yang diperlukan dalam penulisan skripsi ini penulis membagikan kuisioner kepada siswa lulusan SMU/SMK agar dijawab. Data yang diperoleh yaitu sejumlah 80 data calon mahasiswa/i. Adapun atribut yang digunakan ialah : Kesesuaian Dengan Jurusan, Akreditasi, Biaya kuliah, Prospek kerja, Fasilitas, Lokasi, Rekomendasi.

2.2 Analisis Data

Analisa data yang dilakukan pada penulisan ini terdiri dari :

2.2.1 Alat Analisis Data

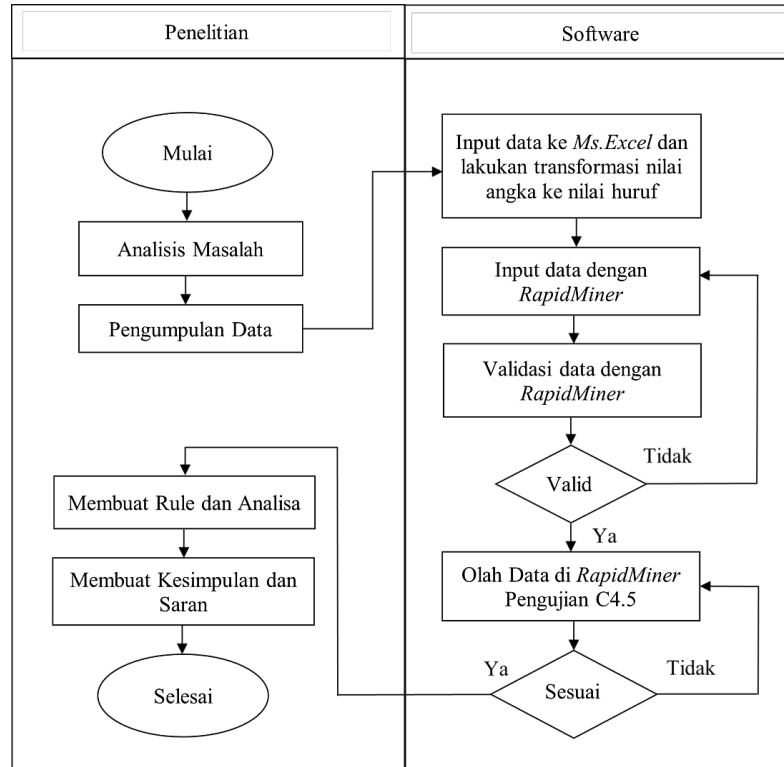
Dalam penelitian ini penulis menggunakan software RapidMiner 5.3 untuk menguji data yang sebelumnya terlebih dahulu diproses menggunakan microsoft excel. RapidMiner merupakan suatu perangkat lunak yang bersifat terbuka. RapidMiner dapat digunakan untuk analisis prediksi.

2.2.2 Instrumen Penelitian

Dalam penelitian ini penulis melakukan pengumpulan data dari kuisioner yang sebelumnya telah dibagikan kemudian memproses data dengan algoritma C4.5 sehingga ditemukan pola minat siswa lulusan SMU/SMK dalam melanjutkan kuliah.

2.2.3 Diagram Aktifitas Kerja Penelitian

Untuk rancangan alur ataupun menentukan pola penelitian yang dilakukan agar memiliki hasil yang sesuai dengan harapan penulis, dibutuhkan diagram aktifitas kerja penelitian sebagai alur atau jalan dalam penelitian ini. Berikut ini ialah diagram aktifitas kerja penelitian yang dapat dilihat pada gambar 2 di bawah ini:



Gambar 2. Diagram Aktifitas Kerja Penelitian

Gambar 2 memaparkan diagram aktifitas kerja penelitian ini dengan penjelasan yaitu peneliti menganalisis masalah yang ada pada penerimaan mahasiswa baru, mengumpulkan data dari koisioner yang dibagikan kepada siswa/i lulusan SMU/SMK, kemudian mengekstra data dalam bentuk Ms.excel dan memvalidasi data dengan menggunakan RapidMiner untuk mengukur target error, dari hasil informasi yang diberikan aplikasi Rapidminer maka penulis membuat pola decision tree.

3. HASIL DAN PEMBAHASAN

3.1 Pendahuluan

Dalam implementasi hasil akhir dari penerapan Algoritma C4.5 dibagi ke dalam dua tahap yaitu proses perhitungan manual menggunakan Algoritma C4.5 dan penyesuaian hasil perhitungan manual melalui pengujian menggunakan software RapidMiner 5.3. Data yang digunakan adalah hasil kuesioner yang tertera pada tabel 1 sebagai input untuk membuat model aturan menggunakan Algoritma C4.5.

3.2 Pengolahan Data

Perhitungan Algoritma C4.5 untuk memperoleh model aturan pohon keputusan data diuraikan sebagai berikut:

Langkah 1: Menghitung jumlah kasus, jumlah kasus untuk keputusan minat dan jumlah kasus untuk keputusan tidak minat.

Langkah 2: Menghitung Entropy dari semua kasus dan kasus yang dibagikan berdasarkan kelas atribut dengan persamaan (1). Selanjutnya dilakukan perhitungan Gain untuk masing masing atribut dengan persamaan (2). Berikut ini adalah perhitungan nilai Entropy dan Gain.

1. Menghitung entropy Total:

$$Entropy [Total] = \left(\left(-\frac{55}{80} \right) \times \log_2 \left(\frac{55}{80} \right) \right) + \left(\left(-\frac{25}{80} \right) \times \log_2 \left(\frac{25}{80} \right) \right) = 0,896038233$$

2. Menghitung entropy dan gain Kesesuaian dengan jurusan :

a. Entropy :

$$[Kesesuaian - Sesuai] = \left(\left(-\frac{44}{53} \right) \times \log_2 \left(\frac{44}{53} \right) \right) + \left(\left(-\frac{9}{53} \right) \times \log_2 \left(\frac{9}{53} \right) \right) = 0,657272979$$



$$[\text{Kesesuaian – Tidak Sesuai}] = \left(\left(-\frac{13}{27} \right) \times \log_2 \left(\frac{13}{27} \right) \right) + \left(\left(-\frac{14}{27} \right) \times \log_2 \left(\frac{14}{27} \right) \right) = 0,999010271$$

b. *Gain* :

$$[\text{Kesesuaian Jurusan}] = 0,896038233 - \left(\left(\frac{53}{80} \right) \times 0,657272979 \right) + \left(\left(\frac{27}{80} \right) \times 0,999010271 \right) = 0,123428918$$

3. Menghitung *entropy* dan *gain* Akreditasi :

a. *Entropy* :

$$[\text{Akreditasi – Unggul}] = \left(\left(-\frac{44}{67} \right) \times \log_2 \left(\frac{44}{67} \right) \right) + \left(\left(-\frac{23}{67} \right) \times \log_2 \left(\frac{23}{67} \right) \right) = 0,927926262$$

$$[\text{Akreditasi – Baik}] = \left(\left(-\frac{11}{13} \right) \times \log_2 \left(\frac{11}{13} \right) \right) + \left(\left(-\frac{2}{13} \right) \times \log_2 \left(\frac{2}{13} \right) \right) = 0,619382195$$

b. *Gain* :

$$[\text{Akreditasi}] = 0,896038233 - \left(\left(\frac{67}{80} \right) \times 0,927926262 \right) + \left(\left(\frac{13}{80} \right) \times 0,619382195 \right) = 0,018250382$$

4. Menghitung *entropy* dan *gain* Kesesuaian dengan jurusan :

a. *Entropy* :

$$[\text{Kesesuaian – Sesuai}] = \left(\left(-\frac{44}{53} \right) \times \log_2 \left(\frac{44}{53} \right) \right) + \left(\left(-\frac{9}{53} \right) \times \log_2 \left(\frac{9}{53} \right) \right) = 0,657272979$$

$$[\text{Kesesuaian – Tidak Sesuai}] = \left(\left(-\frac{13}{27} \right) \times \log_2 \left(\frac{13}{27} \right) \right) + \left(\left(-\frac{14}{27} \right) \times \log_2 \left(\frac{14}{27} \right) \right) = 0,999010271$$

b. *Gain* :

$$[\text{Kesesuaian Jurusan}] = 0,896038233 - \left(\left(\frac{53}{80} \right) \times 0,657272979 \right) + \left(\left(\frac{27}{80} \right) \times 0,999010271 \right) = 0,123428918$$

5. Menghitung *entropy* dan *gain* Akreditasi :

a. *Entropy* :

$$[\text{Akreditasi – Unggul}] = \left(\left(-\frac{44}{67} \right) \times \log_2 \left(\frac{44}{67} \right) \right) + \left(\left(-\frac{23}{67} \right) \times \log_2 \left(\frac{23}{67} \right) \right) = 0,927926262$$

$$[\text{Akreditasi – Baik}] = \left(\left(-\frac{11}{13} \right) \times \log_2 \left(\frac{11}{13} \right) \right) + \left(\left(-\frac{2}{13} \right) \times \log_2 \left(\frac{2}{13} \right) \right) = 0,619382195$$

b. *Gain* :

$$[\text{Akreditasi}] = 0,896038233 - \left(\left(\frac{67}{80} \right) \times 0,927926262 \right) + \left(\left(\frac{13}{80} \right) \times 0,619382195 \right) = 0,018250382$$

6. Menghitung *entropy* dan *gain* Biaya Kuliah :

a. *Entropy* :

$$[\text{Biaya Kuliah – Terjangkau}] = \left(\left(-\frac{42}{52} \right) \times \log_2 \left(\frac{42}{52} \right) \right) + \left(\left(-\frac{10}{52} \right) \times \log_2 \left(\frac{10}{52} \right) \right) = 0,706274089$$

$$[\text{Biaya Kuliah – Murah}] = \left(\left(-\frac{13}{28} \right) \times \log_2 \left(\frac{13}{28} \right) \right) + \left(\left(-\frac{15}{28} \right) \times \log_2 \left(\frac{15}{28} \right) \right) = 0,99631652$$

b. *Gain* :

$$[\text{Biaya Kuliah}] = 0,896038233 - \left(\left(\frac{52}{80} \right) \times 0,706274089 \right) + \left(\left(\frac{28}{80} \right) \times 0,99631652 \right) = 0,088249293$$

7. Menghitung *entropy* dan *gain* Prospek Bekerja :

a. *Entropy* :

$$[\text{Prospek Bekerja – Ya}] = \left(\left(-\frac{54}{78} \right) \times \log_2 \left(\frac{54}{78} \right) \right) + \left(\left(-\frac{24}{78} \right) \times \log_2 \left(\frac{24}{78} \right) \right) = 0,89049164$$

$$[\text{Prospek Bekerja – Tidak}] = \left(\left(-\frac{1}{2} \right) \times \log_2 \left(\frac{1}{2} \right) \right) + \left(\left(-\frac{1}{2} \right) \times \log_2 \left(\frac{1}{2} \right) \right) = 1$$

b. *Gain* :

$$[\text{Prospek Bekerja}] = 0,896038233 - \left(\left(\frac{78}{80} \right) \times 0,89049164 \right) + \left(\left(\frac{2}{80} \right) \times 1 \right) = 0,002808883$$

8. Menghitung *entropy* dan *gain* Fasilitas

a. *Entropy* :

$$[\text{Fasilitas – Mendukung}] = \left(\left(-\frac{4}{4} \right) \times \log_2 \left(\frac{4}{4} \right) \right) + \left(\left(-\frac{25}{76} \right) \times \log_2 \left(\frac{25}{76} \right) \right) = 0,913847287$$

$$[\text{Fasilitas – Tidak}] = \left(\left(-\frac{4}{4} \right) \times \log_2 \left(\frac{4}{4} \right) \right) + \left(\left(-\frac{0}{4} \right) \times \log_2 \left(\frac{0}{4} \right) \right) = 0$$

b. *Gain* :

$$[\text{Fasilitas}] = 0,896038233 - \left(\left(\frac{76}{80} \right) \times 0,913847287 \right) + \left(\left(\frac{4}{80} \right) \times 0 \right) = 0,02788331$$

9. Menghitung *entropy* dan *gain* Lokasi :

a. *Entropy* :

$$[\text{Lokasi – Jauh}] = \left(\left(-\frac{4}{4} \right) \times \log_2 \left(\frac{4}{4} \right) \right) + \left(\left(-\frac{25}{78} \right) \times \log_2 \left(\frac{25}{78} \right) \right) = 0,9090225156$$

$$[\text{Lokasi – Dekat}] = \left(\left(-\frac{30}{43} \right) \times \log_2 \left(\frac{30}{43} \right) \right) + \left(\left(-\frac{13}{43} \right) \times \log_2 \left(\frac{13}{43} \right) \right) = 0,884115122$$

b. *Gain* :

$$[\text{Lokasi}] = 0,896038233 - \left(\left(\frac{37}{80} \right) \times 0,9090225156 \right) + \left(\left(\frac{43}{80} \right) \times 0,884115112 \right) = 0,02788331$$

10. Menghitung *entropy* dan *gain* Rekomendasi :

a. *Entropy* :

$$[\text{Rekomendasi} - \text{Pernah}] = \left(\left(-\frac{44}{49} \right) \times \log_2 \left(\frac{44}{49} \right) \right) + \left(\left(-\frac{5}{49} \right) \times \log_2 \left(\frac{5}{49} \right) \right) = 0,598100125$$

$$[\text{Rekomendasi} - \text{Tidak}] = \left(\left(-\frac{11}{31} \right) \times \log_2 \left(\frac{11}{31} \right) \right) + \left(\left(-\frac{20}{31} \right) \times \log_2 \left(\frac{20}{31} \right) \right) = 0,938315352$$

b. *Gain* :

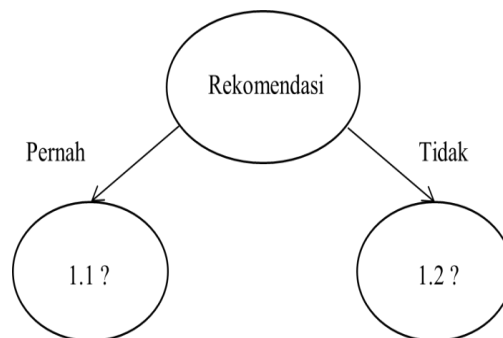
$$[\text{Rekomendasi}] = 0,896038233 - \left(\left(\frac{49}{80} \right) \times 0,598100125 \right) + \left(\left(\frac{31}{80} \right) \times 0,938315352 \right) = 0,24123915$$

Dibawah ini merupakan hasil perhitungan nilai *entropy* dan *gain* yang akan ditunjukkan pada tabel 3 berikut.

Tabel 3. Hasil Perhitungan nilai *entropy* dan *gain*

		Jml Kasus (S)	Ya (S1)	Tidak (S2)	Entropy	Gain
Node 1		80	55	25	0.896038233	
Kesesuaian Dengan Jurusan	Sesuai	53	44	9	0.657272979	
	Tidak Sesuai	27	13	14	0.999010271	
						0.123428918
Akreditasi	Unggul	67	44	23	0.927926262	
	Baik	13	11	2	0.619382195	
	Cukup	0	0	0	0	
						0.018250382
Biaya Kuliah	Terjangkau	52	42	10	0.706274089	
	Mahal	28	13	15	0.99631652	
						0.088249293
Prospek Bekerja	Ya	78	54	24	0.89049164	
	Tidak	2	1	1	1	
						0.002808883
Fasilitas	Mendukung	76	51	25	0.913847287	
	Tidak	4	4	0	0	
						0.02788331
Jarak	Jauh	37	25	12	0.909022156	
	Dekat	43	30	13	0.884115122	
						0.000403607
Rekomendasi	Pernah	49	44	5	0.475431646	
	Tidak	31	11	20	0.938315352	
						0.24123915

Berdasarkan tabel 3 proses pemodelan pohon keputusan berdasarkan hasil perhitungan algoritma C4.5 dari perhitungan node 1 diperoleh pohon keputusan yang dapat dilihat pada gambar 3 berikut ini:



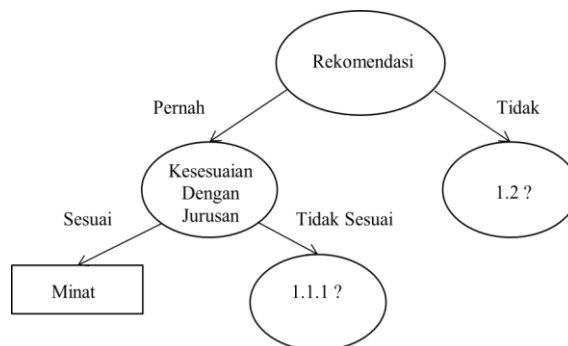
Gambar 3. Pohon keputusan *node 1*

Langkah 3: Dari hasil perhitungan pada Tabel 3 diperoleh nilai atribut tertinggi adalah Rekomendasi dengan gain sebesar 0.24123915. Maka atribut rekomendasi dipilih sebagai node akar. Untuk kelas atributnya pernah dan tidak pernah belum diperoleh hasilnya, maka perlu di lakukan perhitungan lebih lanjut. Hasil perhitungannya dapat dilihat pada tabel 4 berikut ini.

Tabel 4. Hasil Perhitungan *Node 1.1*

			Jml Kasus (S)	Ya (S1)	Tidak (S2)	Entropy	Gain
Rekomendasi	–		49	44	5	0.475431646	
Pernah		Sesuai	35	35	0	0	
Kesesuaian Jurusan	Dengan	Tidak Sesuai	14	9	5	0.940285959	0.206778515
		Unggul	40	35	5	0.543564443	
		Baik	9	9	0	0	
Akreditasi		Cukup	0	0	0	0	0.03170557
		Terjangkau	38	37	1	0.175565026	
Biaya Kuliah		Mahal	11	7	4	0.945660305	0.126988088
		Ya	48	43	5	0.482066148	
Prospek Bekerja		Tidak	1	1	0	0	0.003203583
		Mendukung	47	42	5	0.488908591	
Fasilitas		Tidak	2	2	0	0	0.006478508
		Jauh	19	17	2	0.485460761	
Lokasi		Dekat	30	27	3	0.468995594	0.184261368

Hasil perhitungan pada tabel 5, atribut yang menjadi node cabang dari Rekomendasi = pernah, kesesuaian dengan jurusan memiliki nilai gain tertinggi sebesar 0.206778515. Hasil perhitungan dari node 1.1 diperoleh pohon keputusan yang dapat dilihat pada gambar 4 berikut ini:



Gambar 4. Pohon keputusan *node 1.1*

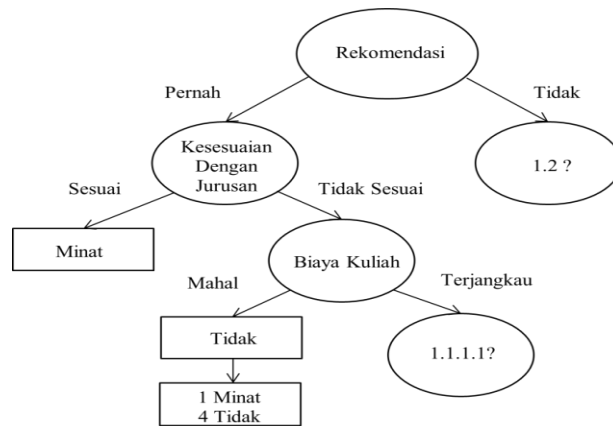
Untuk nilai kelas atribut sesuai sudah memperoleh satu keputusan yaitu minat, sehingga tidak perlu dilakukan perhitungan lagi dan kelas atribut tidak sesuai masih perlu dilakukan perhitungan. Berikut ini hasil perhitungan Rekomendasi = pernah dan kesesuaian dengan jurusan = tidak sesuai, ditunjukkan pada Tabel 5 berikut ini:

Tabel 5. Hasil Perhitungan *Node 1.1.1*

			Jml Kasus (S)	Ya (S1)	Tidak (S2)	Entropy	Gain
Kesesuaian - Tidak Sesuai			14	9	5	0.940285959	
		Unggul	11	6	5	0.994030212	
		Baik	3	3	0	0	
Akreditasi		Cukup	0	0	0	0	0.159262221
		Terjangkau	9	8	1	0.503258335	
Biaya Kuliah		Mahal	5	1	4	0.721928095	0.358931281
Kesesuaian - Tidak Sesuai			14	9	5	0.940285959	
Prospek Bekerja		Ya	14	9	5	0.940285959	
		Tidak	0	0	0	0	

		Jml Kasus (S)	Ya (S1)	Tidak (S2)	Entropy	Gain
						0
Fasilitas	Mendukung	13	8	5	0.961236605	0.047709111
	Tidak	1	1	0	0	
Lokasi	Jauh	3	1	2	0.918295834	0.184261368
	Dekat	11	8	3	0.845350937	

Hasil perhitungan pada tabel 5, atribut yang menjadi node cabang dari Rekomendasi = pernah dan kesesuaian dengan jurusan = tidak sesuai, biaya kuliah memiliki nilai gain tertinggi sebesar 0.358931281. Hasil perhitungan dari node 1.1.1 diperoleh pohon keputusan yang dapat dilihat pada gambar 5 berikut ini:



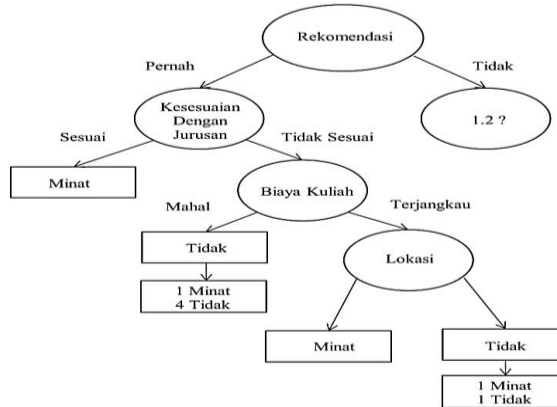
Gambar 5. Pohon keputusan *node* 1.1.1

Untuk nilai pada kelas atribut mahal, jumlah keputusan minat = 1 dan tidak = 4 oleh karena itu kelas atribut mahal ditetapkan memiliki nilai tidak. Nilai pada kelas terjangkau belum diperoleh hasilnya, maka perlu di lakukan perhitungan lebih lanjut. Berikut ini hasil perhitungan rekomendasi = pernah, kesesuaian dengan jurusan = tidak sesuai, dan biaya kuliah = terjangkau dapat dilihat pada tabel 6 berikut ini.

Tabel 6. Hasil Perhitungan *Node* 1.1.1.1

		Jml Kasus (S)	Ya (S1)	Tidak (S2)	Entropy	Gain
Biaya Kuliah-Terjangkau						
		9	8	1	0.503258335	
Akreditasi	Unggul	6	5	1	0.650022422	0.069910054
	Baik	3	3	0	0	
	Cukup	0	0	0	0	
ProspekBekerja						
	Ya	9	8	1	0.503258335	0
	Tidak	0	0	0	0	
Fasilitas						
	Mendukung	8	7	1	0.543564443	0.020089941
	Tidak	1	1	0	0	
Lokasi						
	Jauh	2	1	1	1	0.184261368
	Dekat	7	7	0	0	

Hasil perhitungan pada Tabel 6, atribut yang menjadi node cabang dari rekomendasi = pernah, kesesuaian dengan jurusan = tidak sesuai, dan biaya kuliah = terjangkau, lokasi memiliki nilai gain tertinggi sebesar 0.184261368. Nilai atribut dekat memperoleh satu keputusan yaitu minat sehingga tidak perlu dilakukan perhitungan lebih lanjut, nilai atribut jauh memiliki nilai minat = 1 dan tidak = 1. Hasil perhitungan dari node 1.1.1.1 diperoleh pohon keputusan yang dapat dilihat pada gambar 6 berikut ini:



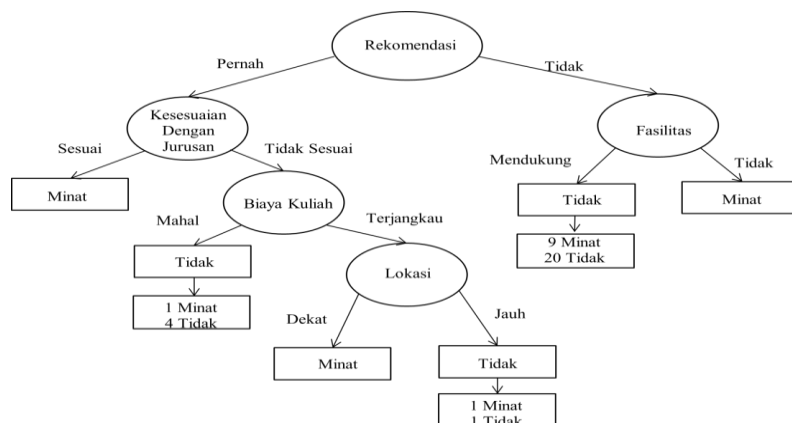
Gambar 6. Pohon keputusan *node* 1.1.1.1

Untuk node 1.2 dengan node cabang Rekomendasi = tidak pernah memiliki hasil perhitungan seperti tabel 7 berikut :

Tabel 7. Hasil Perhitungan *Node* 1.2

		Jml Kasus (S)	Minat (S1)	Tidak (S2)	Entropy	Gain
Rekomendasi-Tidak		31	11	20	0.938315352	
Kesesuaian Dengan Jurusan	Sesuai	17	7	11	0.933476405	
	Tidak	13	4	9	0.89049164	
	Sesuai					0.052976958
Akreditasi	Unggul	27	8	19	0.876716289	
	Baik	4	3	1	0.811278125	
	Cukup	0	0	0	0	
						0.070042697
Biaya Kuliah	Terjangkau	14	5	9	0.940285959	
	Mahal	17	6	11	0.936667382	
						-0.02994417
Prospek Bekerja	Ya	30	11	19	0.948078244	
	Tidak	1	0	1	0	
						0.020820278
Fasilitas	Mendukung	29	9	20	0.893571102	
	Tidak	2	2	0	0	
						0.102393999
Jarak	Jauh	18	8	10	0.99107606	
	Dekat	13	3	10	0.779349837	
						0.036027708

Hasil perhitungan pada Tabel 7, atribut yang menjadi node cabang dari rekomendasi = tidak pernah , fasilitas memiliki nilai gain tertinggi sebesar 0.102393999. Hasil perhitungan dari node 1.1.1.1 diperoleh pohon keputusan yang dapat dilihat pada gambar 7 berikut ini:



Gambar 7. Pohon keputusan *node* 1.2

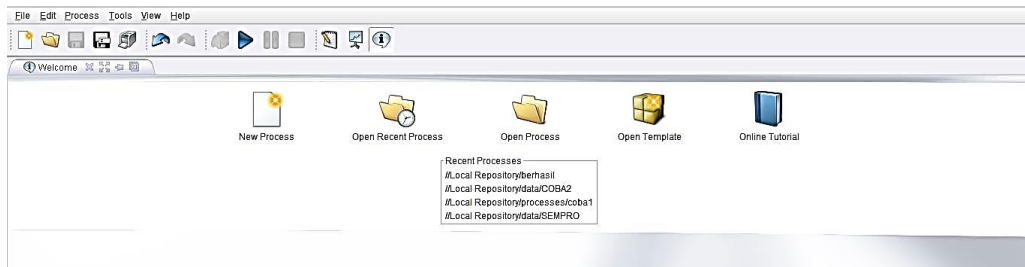
Nilai atribut tidak memiliki satu keputusan yaitu minat, untuk atribut mendukung memiliki nilai 9 minat dan 20 tidak sehingga diputuskan nilai atribut mendukung ialah tidak. Berdasarkan pohon keputusan pada gambar 4.5. terdapat 6 (enam) rules yang dapat di jadikan sebagai referensi dalam menganalisis pola minat siswa lulusan SMU/SMK untuk melanjutkan kuliah yaitu 3 (tiga) rules keputusan minat dan 3 (tiga) rules keputusan tidak,

1. Jika Rekomendasi = Pernah, dan Kesesuaian Dengan Jurusan = Sesuai maka hasilnya Minat.
2. Jika Rekomendasi = Pernah, Kesesuaian Dengan Jurusan = Tidak, dan Biaya Kuliah = Mahal maka hasilnya Tidak.
3. Jika Rekomendasi = Pernah, Kesesuaian Dengan Jurusan = Tidak, Biaya Kuliah = Terjangkau, dan Lokasi = Dekat maka hasilnya Minat.
4. Jika Rekomendasi = Pernah, Kesesuaian Dengan Jurusan = Tidak, Biaya Kuliah = Terjangkau, dan Lokasi = Jauh maka hasilnya Tidak.
5. Jika Rekomendasi = Tidak, dan Fasilitas = Mendukung maka hasilnya Tidak.
6. Jika Rekomendasi = Tidak, dan Fasilitas = Tidak maka hasilnya Minat.

3.3 Pengolahan Data Menggunakan Software Rapidminer 5.3.

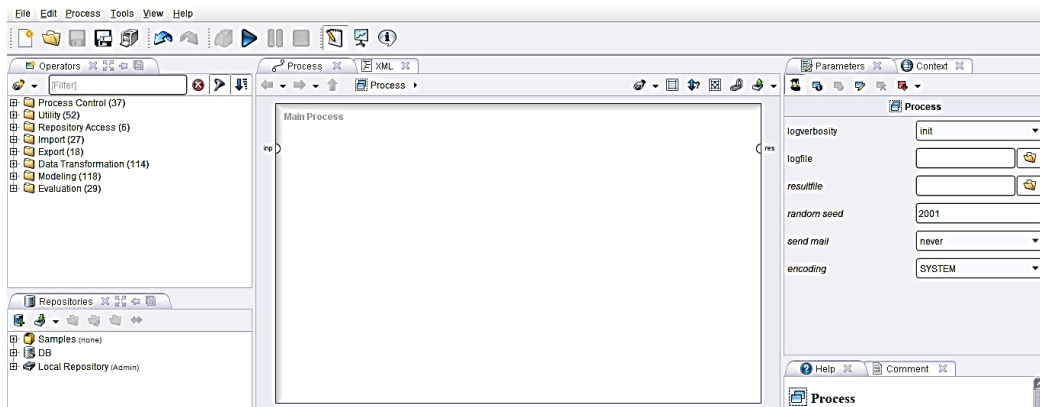
Hasil perhitungan manual telah di dapatkan sebelumnya, maka selanjutnya melakukan pengujian menggunakan Software Rapidminer 5.3 dengan urutan proses sebagai berikut :

- a. Langkah 1 : Membuka software RapidMiner, dimana tampilan utamanya dapat kita lihat pada Gambar 8.



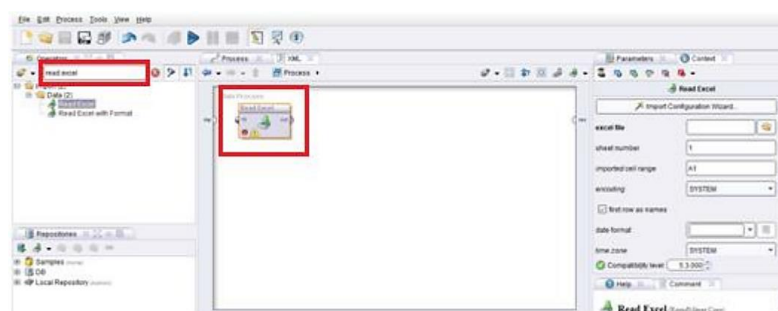
Gambar 8. Tampilan utama RapidMiner Studio 5.3.

- b. Langkah 2: Pilih New Process untuk membuat proses baru dan melakukan pengujian data. Berikut ini adalah tampilan New Process yang ditampilkan pada Gambar 9.



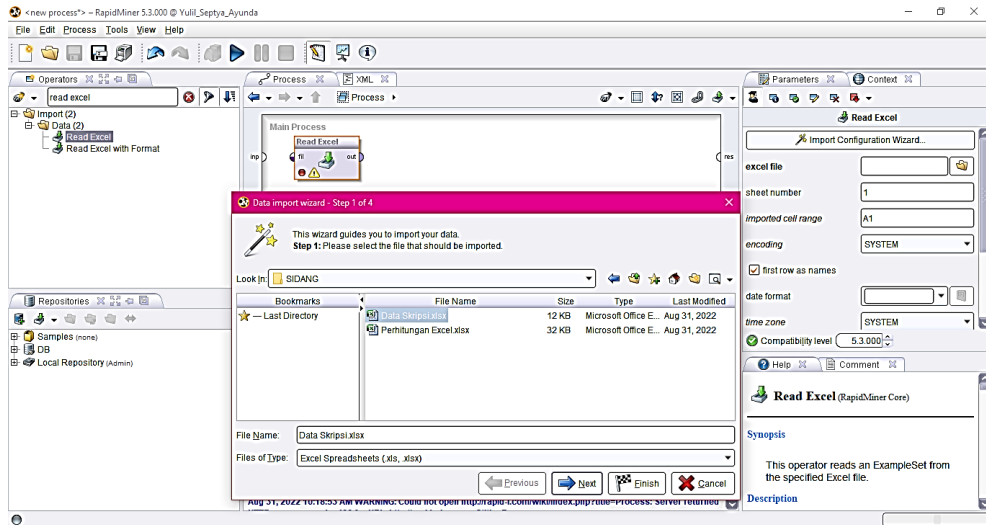
Gambar 9. Tampilan new process

- c. Langkah 3 : Gunakan operator Read Excel untuk menginput data yang akan diuji. Untuk menemukan operator kita dapat mencarinya dengan mengetikkan “nama_operator” yang dicari pada kolom [filter] di bagian Operators. Selanjutnya drag and drop operator Read Excel ke lembar kerja main process. Tahapan ini dapat dilihat pada Gambar 10 berikut ini :



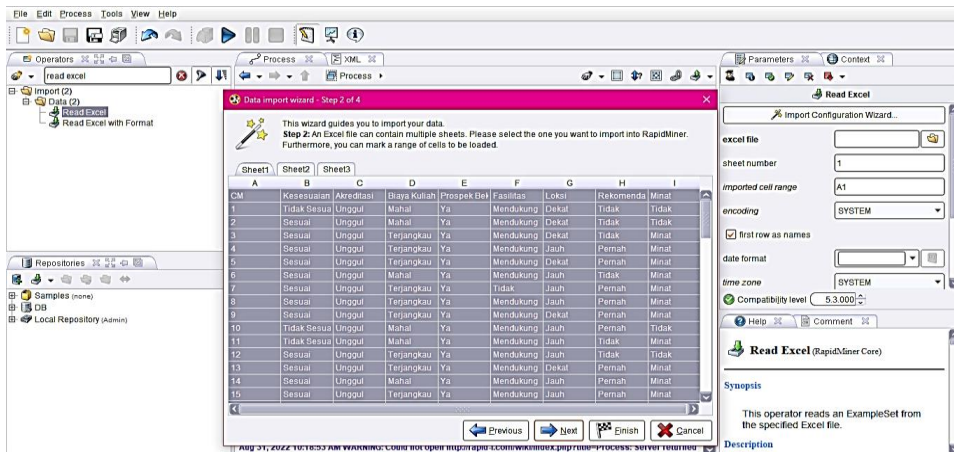
Gambar 10. Input data excel

d. Langkah 4 :Menginput data excel pada operator Read Excel. Berikut ini adalah tahapan input data yang ditampilkan pada Gambar 11 dibawah ini :



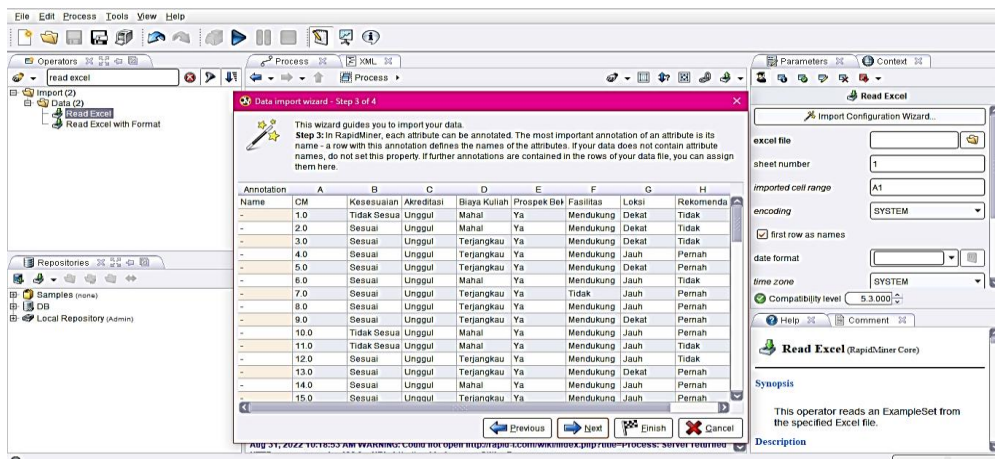
Gambar 11. Tampilan import data tahap 1

Gambar 11 diatas dapat dijelaskan sebagai berikut : pilih *Import Configuration Wizard* yang terdapat pada sub parameter *Read Excel*. Kemudian akan muncul tampilan jendela *Data Import Wizard*, lalu pada bagian look in pilih tempat kita menyimpan data. Selanjutnya pilih file name data yang akan digunakan lalu klik Next. Berikut ini adalah tampilan data yang diimport:



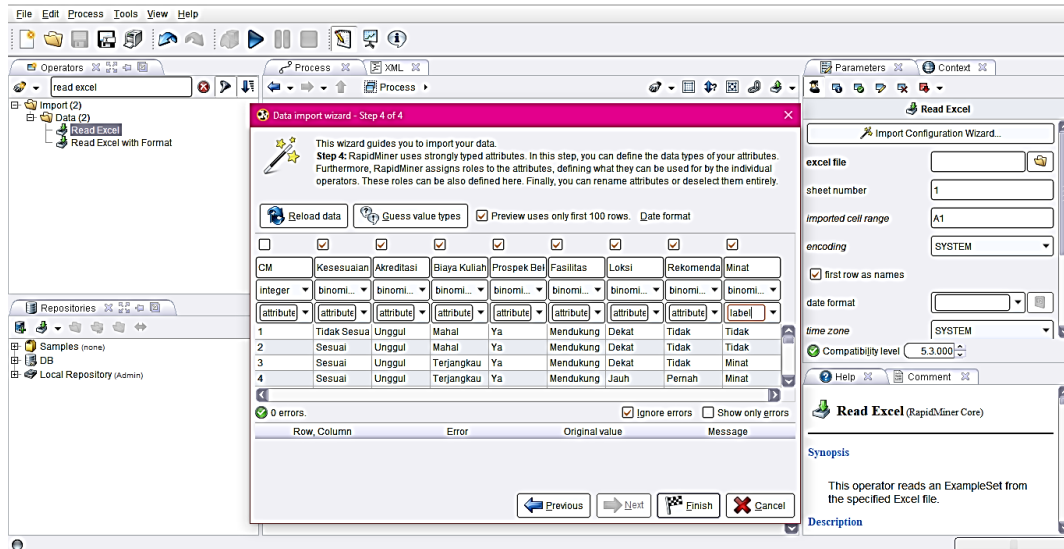
Gambar 12. Tampilan import data tahap 2

Gambar 12 diatas, menunjukkan isi data yang akan di proses. Pilih sheet data serta seleksi data yang akan di proses lalu klik next. Selanjutnya akan muncul tampilan hasil seleksi yang ditunjukkan pada Gambar 13 lalu klik Next.



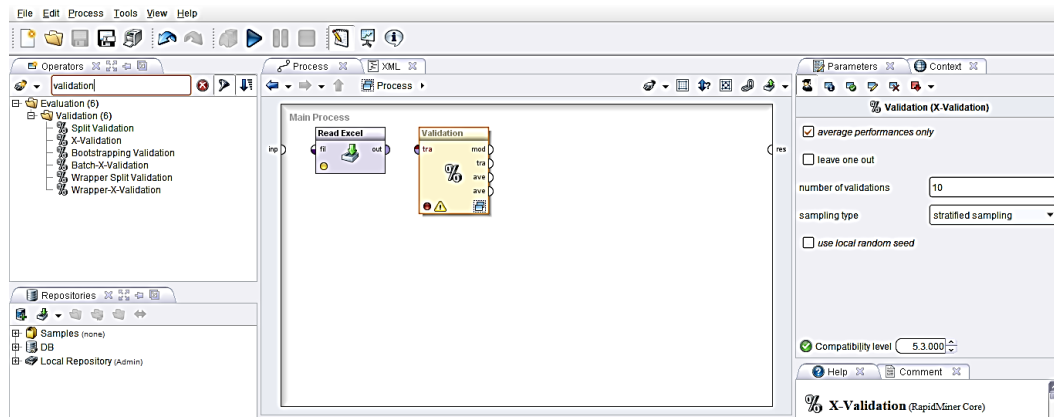
Gambar 13. Tampilan import data tahap 3

Pada import data tahap 4 dilakukan pemilihan tipe data dan menentukan tipe data “label” dapat dilihat pada Gambar 14, lalu pilih Finish.



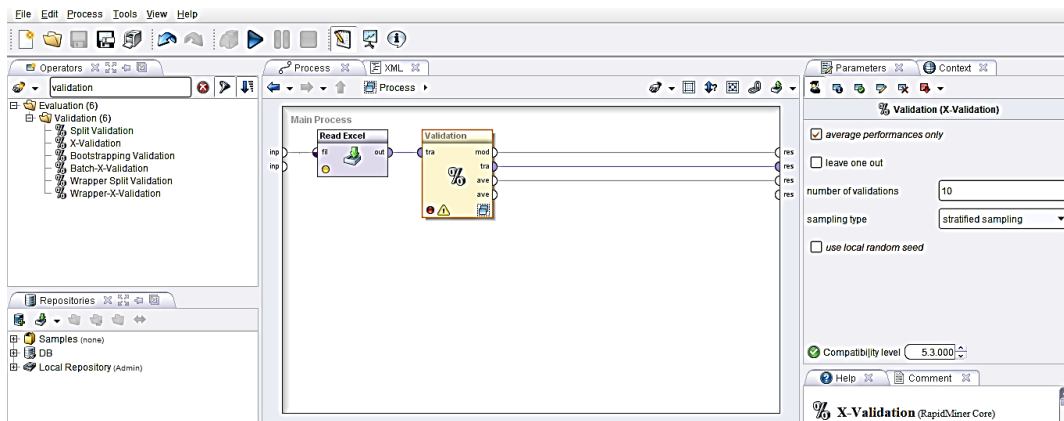
Gambar 14. Tampilan import data tahap 4

e. Langkah 5 : Data yang sudah berhasil diimport kemudian diolah ke dalam decision tree dengan model pengolahan data sebagai berikut : Proses tahap awal pemodelan pengolahan data yaitu dengan menambahkan operator X Validation seperti pada Gambar 15.



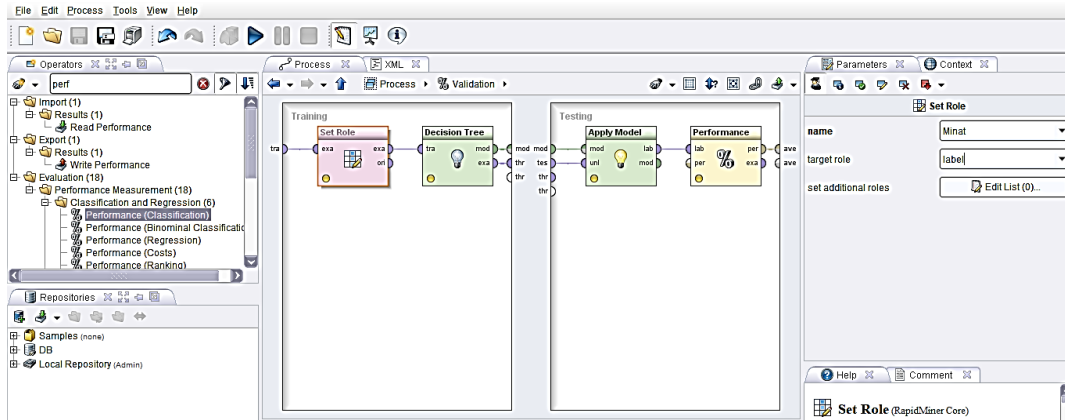
Gambar 15. Input Operator x Validation

Lalu menghubungkan operator read excel dengan operator X validation seperti pada Gambar 16.



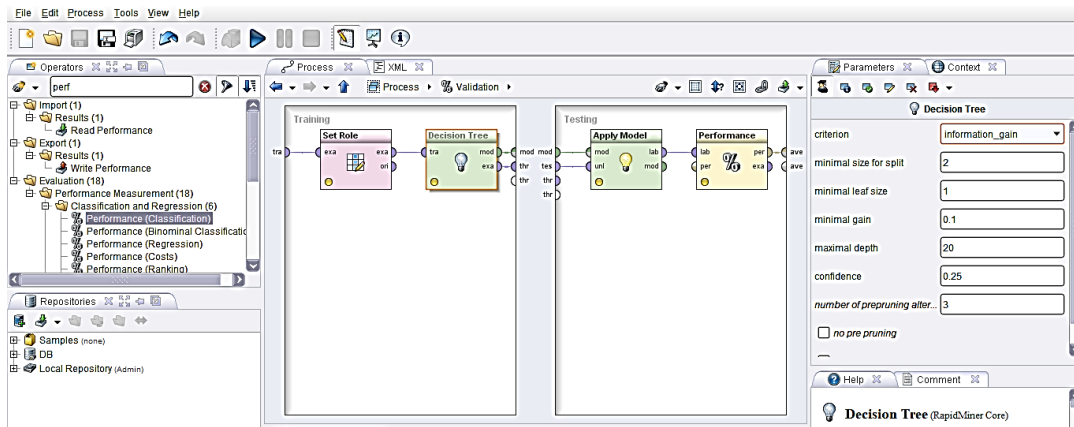
Gambar 16. Konektivitas antara data dan model pohon keputusan (1)

Pada proses tahap kedua yaitu membuat aturan model pohon keputusan didalam operator X validation dengan cara double klik pada operator X validation, kemudian akan tampil halaman operator validation yang terbagi atas area training dan testing yang ditunjukkan pada Gambar 17 berikut :



Gambar 17. Konektivitas antara data dan model pohon keputusan (2)

Selanjutnya *drag and drop* operator *Set Role* dan *decision tree* pada area *training*, operator *Apply Model* dan operator *performance (classification)* pada area *testing* kemudian hubungkan setiap operator seperti pada Gambar 17. Agar didapat hasil yang sesuai antara perhitungan manual dengan hasil pengujian menggunakan *software RapidMiner*, perlu dilakukan penyesuaian terhadap parameter pada *decision tree*. Untuk penyesuaian terhadap *parameters* dapat dilihat pada Gambar 18 berikut.

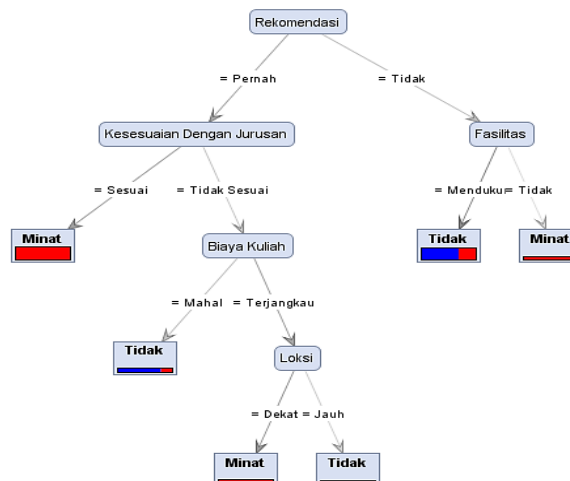


Gambar 18. Parameter Decision Tree

f. Langkah 6: Setelah semua data dan operator model pohon keputusan berhasil terkoneksi (tidak ada *error* pada koneksi operator di halaman *validation* maupun *process*) kemudian klik tombol *Run* pada toolbox untuk memperoleh hasil klasifikasi pohon keputusan Algoritma C4.5.

3.4 Hasil Pengolahan Data

Hasil pengolahan data dengan model pohon keputusan sesuai dengan software *RapidMiner*, dapat dilihat pada Gambar 19 sebagai berikut :



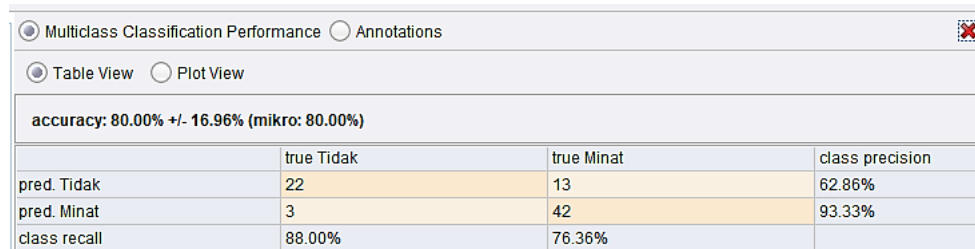
Gambar 19. Decision Tree pada *RapidMiner*

Gambar 19 merupakan pohon keputusan yang dihasilkan pada RapidMiner, penyederhanaan dari gambar diatas dapat dilihat pada Rule Decision Tree gambar 20 berikut:

```
Tree
Rekomendasi = Pernah
| Kesesuaian Dengan Jurusan = Sesuai: Minat {Tidak=0, Minat=35}
| Kesesuaian Dengan Jurusan = Tidak Sesuai
| | Biaya Kuliah = Mahal: Tidak {Tidak=4, Minat=1}
| | Biaya Kuliah = Terjangkau
| | | Lokasi = Dekat: Minat {Tidak=0, Minat=7}
| | | Lokasi = Jauh: Tidak {Tidak=1, Minat=1}
Rekomendasi = Tidak
| Fasilitas = Mendukung: Tidak {Tidak=20, Minat=9}
| Fasilitas = Tidak: Minat {Tidak=0, Minat=2}
```

Gambar 20. Rule Decision Tree pada RapidMiner

Hasil penerapan Algoritma C4.5 menggunakan software RapidMiner dengan operator X Validation diperoleh nilai accuracy yaitu sebesar 80% dan classification_error sebesar 16.96% artinya bahwa rule yang dihasilkan tingkat kebenarannya mendekati 100%, seperti yang terlihat pada Gambar 21.



	true Tidak	true Minat	class precision
pred. Tidak	22	13	62.86%
pred. Minat	3	42	93.33%
class recall	88.00%	76.36%	

Gambar 21. Nilai Akurasi Algoritma C4.5

Gambar 21 dapat dijelaskan bahwa label Tidak mempunyai nilai 13 dengan Class Precision sebesar 62.86% dan prediksi label Minat mempunyai nilai 42 dengan Class Precision sebesar 93.33%.

3.5 Pembahasan

Hasil yang dilakukan peneliti dalam perhitungan Algoritma C4.5 diperoleh 6 model aturan atau rule untuk siswa lulusan SMU/SMK untuk kuliah di STIKOM Tunas Bangsa Pematangsiantar. Model aturan dalam bentuk pohon keputusan yang diperoleh oleh peneliti dapat dilihat pada Gambar 3.17. Berdasarkan pengolahan data menggunakan software RapidMiner didapat nilai accuracy sebesar 80.00%, artinya bahwa rule yang dihasilkan tingkat kebenarannya mendekati 100%. Parameters yang digunakan pada decision tree telah dilakukan penyesuaian terhadap kriteria decision tree yang digunakan pada RapidMiner yaitu information gain, maksimal depth = 20, confidence = 0,25, minimal gain = 0,1, minimal leaf size = 1, minimal size for split = 2, number of pruning alt = 3. Artinya hasil proses yang dilakukan peneliti pada perhitungan Algoritma C4.5 dan Rapidminer diperoleh hasil yang sama dan sesuai. Sehingga pengujian dengan RapidMiner dapat dikatakan berhasil dan dapat menemukan pohon keputusan pada kasus analisis pola minat siswa lulusan SMU/SMK untuk melanjutkan kuliah di STIKOM Tunas Bangsa Pematangsiantar.

4. KESIMPULAN

Berdasarkan pembahasan pada bab sebelumnya dapat disimpulkan bahwa ada enam (6) pola yang dihasilkan pada kasus peminatan terhadap calon mahasiswa yang akan berkuliah di STIKOM Tunas Bangsa, dimana tiga (3) dari rules tersebut menghasilkan keputusan minat dan tiga (3) rules menghasilkan keputusan tidak minat. Salah satu rules yang dihasilkan ialah Jika Rekomendasi = Pernah, dan Kesesuaian Jurusan = Sesuai, maka hasilnya Minat. Rekomendasi dari orang sekitar menjadi faktor pendukung utama untuk siswa lulusan SMU/SMK untuk melanjutkan kuliah di STIKOM Tunas Bangsa Pematangsiantar.

REFERENCES

- Abdelsamea, M. M., Zidan, U., Senousy, Z., Gaber, M. M., Rakha, E., & Ilyas, M. (2022). A survey on artificial intelligence in histopathology image analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(6), 1–44. <https://doi.org/10.1002/widm.1474>
- Adenuddin Alwy, M. (2022). Manajemen Sumber Daya Manusia Di Era Digital Melalui Lensa Manajer Sumber Daya Manusia Generasi Berikutnya. *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi*,



- Dan Pendidikan*, 1(10), 2265–2276. <https://doi.org/10.54443/sibatik.v1i10.334>
- Alam Jusia, P., Muhammad Irfan, F., & Dinamika Bangsa Jambi Jl Jend Sudirman Thehok Jambi, S. (2019). Clustering Data Untuk Rekomendasi Penentuan Jurusan Perguruan Tinggi Menggunakan Metode K-Means. *Jurnal IKRA-ITH Informatika*, 3(3), 75.
- Anggraeni, L. D., Toby, Y. R., & Rasmada, S. (2021). Analisis Asupan Zat Gizi Terhadap Status Gizi Balita. *Faletehan Health Journal*, 8(02), 92–101. <https://doi.org/10.33746/fhj.v8i02.191>
- Aswathy, M. A., & Jagannath, M. (2020). Performance Analysis of Segmentation Algorithms for the Detection of Breast Cancer. *Procedia Computer Science*, 167, 666–676. <https://doi.org/10.1016/j.procs.2020.03.333>
- Dharma, R., Harahap, A. S. H., Dalimunthe, L., & Defit, S. (2019). Sistem Pembuatan Keputusan Penetapan Calon Sertifikasi Dosen Menggunakan Analitical Hierarchy Process (AHP). *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 130. <https://doi.org/10.30645/senaris.v1i0.16>
- Dr. V. Suma. (2020). Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics. *Journal of Soft Computing Paradigm*, 2(3), 153–159. <https://doi.org/10.36548/jscp.2020.3.002>
- Hakkoum, H., Idri, A., Abnane, I., & Fernades-Aleman, J. (2023). Does Categorical Encoding Affect the Interpretability of a Multilayer Perceptron for Breast Cancer Classification? *Proceedings Ofthe 12th International Conference on Data Science, Technology and Applications (DATA 2023)*, Data, 351–358. <https://doi.org/10.5220/0012084800003541>
- Joni, J., Rosario, M., & Pahlevi, M. R. (2021). Perancangan Sistem Informasi Peminjaman Barang Pada Chara Center Jambi. *Jurnal Ilmiah Mahasiswa Sistem Informasi*, 3(2), 168–181.
- M.SYARIFUDDIN YUSUF. (2021). IMPLEMENTASI METODE ANALYTIC HIERARCHY PROCESS (AHP) DALAM KELAYAKAN PENGAJUAN KREDIT PADA BMT UMMAT SEJAHTERA ABADI JEPARA. *Jurnal Al-Ilmi*, 13(1), 110–117.
- Nathan, A. J., & Scobell, A. (2012). Model Algoritma K-nearest Neighbor untuk memprediksi kelulusan mahasiswa. *Foreign Affairs*, 91(5), 1–9.
- Ninh, D. B., He, J., Trung, V. T., & Huy, D. P. (2020). An effective random statistical method for Indoor Positioning System using WiFi fingerprinting. *Future Generation Computer Systems*, 109, 238–248. <https://doi.org/10.1016/j.future.2020.03.043>
- Ouddai, G., Hamdi, I., & Ghezala, H. Ben. (2023). A Comparative Study of BRISK, ORB and DAISY Features for Breast Cancer Classification. *International Conference on Pattern Recognition Applications and Methods, 1(Icpram)*, 964–970. <https://doi.org/10.5220/0011902200003411>
- Ramadhan, N. A., & Nufriana, D. A. (2019). Rancang Bangun Dan Implementasi Sistem Informasi Skripsi Oline Berbasis WEB. *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, 1(02), 1–12. <https://doi.org/10.46772/intech.v1i02.75>
- Seyedan, M., & Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *Journal of Big Data*, 7(1). <https://doi.org/10.1186/s40537-020-00329-2>
- Sholihin, I., & Ariyani, F. (2023). Perancangan Sistem Informasi Pendaftaran Anggota Baru Berbasis Web Pada UKMI Ar-Rahman Universitas Teknokrat Indonesia. *Jurnal Data Mining Dan Sistem Informasi*, 4(1), 23. <https://doi.org/10.33365/jdmsi.v4i1.2618>
- Srisook, N., Tuntoolavest, O., Danphitsanuparn, P., Pattana-anake, V., & Joseph, F. J. J. (2022). Convolutional Neural Network Based Nutrient Deficiency Classification in Leaves of Elaeis guineensis Jacq. *International Journal of Computer Information Systems and Industrial Management Applications*, 14, 19–27.
- Tahirović, E., & Krivić, S. (2023). Interpretability and Explainability of Logistic Regression Model for Breast Cancer Detection. *Proceedings Ofthe 15th International Conference on Agents and Artificial Intelligence (ICAART2023)*, 3(Icaart), 161–168. <https://doi.org/10.5220/0011627600003393>
- Tolan, J., Yang, H.-I., Nosarzewski, B., Couairon, G., Vo, H., Brandt, J., Spore, J., Majumdar, S., Haziza, D., Vamaraju, J., Moutakanni, T., Bojanowski, P., Johns, T., White, B., Tieceke, T., & Couprie, C. (2023). *Sub-meter resolution canopy height maps using self-supervised learning and a vision transformer trained on Aerial and GEDI Lidar*.
- Yekti Pawestri, A. (2022). *Implementasi Text Mining Dalam Perbaikan Kualitas Aplikasi Google Classroom Berdasarkan Review Pengguna*. 1–62.
- Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika. Jurnal Edik Informatika*, 2(2), 213–219.
- Zeng, J., Ben, K., Li, X., & Zhang, X. (2019). Fast code clone detection based on weighted recursive autoencoders. *IEEE Access*, 7, 125062–125078. <https://doi.org/10.1109/ACCESS.2019.2938825>