

Penerapan Data Mining Klasifikasi Pada Calon Pelanggan Baru Indihome dengan C.45

Arfandi*, Agus Perdana Windarto, Ilham Syahputra Saragih

Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: arfandhi05@gmail.com

Abstrak—Pelanggan Indihome adalah orang yang menjadi pembeli produk yang telah dibuat dan dipasarkan oleh sebuah perusahaan, dimana orang ini bukan hanya sekali membeli produk tersebut tetapi berulang-ulang. Sedangkan Calon pelanggan Indihome adalah mereka yang belum menjadi pelanggan Anda (pembeli/pengguna jasa) namun dinilai berpotensi untuk segera jadi pelanggan untuk dalam waktu dekat atau waktu kedepannya. Dibutuhkan algoritma untuk mengklasifikasikan calon pelanggan baru indihome agar tidak ada kerugian antara kedua pihak. Algoritma C4.5 dipilih karena mampu mengklasifikasikan calon pelanggan baru indihome dengan menggunakan aplikasi *rapid miner* dan perhitungannya menggunakan *microsoft excel*. Dari perhitungan menggunakan kedua aplikasi tersebut diperoleh hasil antara lain Jika Penghasilan <2000000 dan Jenis Rumah = Kontrak, maka hasilnya Tidak Layak {Layak = 0 dan Tidak Layak = 7} dan hasil layak yaitu aturan Jika Pekerjaan = Wiraswasta dan Penghasilan > 2000000 maka hasilnya Layak (Layak = 54 dan Tidak Layak = 0).

Kata Kunci: Klasifikasi; Data Mining; Algoritma C4.5; Indihome

Abstract—Indihome customers are people who are buyers of products that have been made and marketed by a company, where this person not only buys the product once but repeatedly. Meanwhile, Prospective Indihome customers are those who have not become your customers (buyers/service users) but are considered to have the opportunity to become customers in the near future or in the future. An algorithm is needed to classify indihome's new prospective customers so that there is no loss between the two parties. The C4.5 algorithm was chosen because it is able to classify new prospective customers of Indihome by using the rapid miner application and the calculation using Microsoft Excel. From the calculations using the two applications, the results obtained include: If Income <2000000 and Type of House = Contract, then the result is Not Eligible {Eligible = 0 and Not Eligible = 7} and the result is feasible, namely the rule If Employment = Self Employed and Income > 2000000 then the result is Eligible (Eligible = 54 and Not Eligible = 0).

Keywords: Classification ;Data Mining; C4.5 Algorithm; Indihome

1. PENDAHULUAN

Algoritma C4.5 merupakan struktur pohon dimana terdapat simpul yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Algoritma C4.5 menggunakan konsep *information gain* atau entropy reduction untuk memilih pembagian yang optimal [1]. Konsep dasar *data mining* adalah menentukan informasi tersembunyi dalam sebuah basis data dan merupakan bagian dari *Knowledge Discovery in Databases (KDD)* untuk menemukan informasi dan pola yang berguna dalam data. *Data mining* mencari informasi baru, berharga dan berguna dalam sekumpulan data dengan melibatkan komputer dan manusia serta bersifat iteratif baik melalui proses yang otomatis ataupun manual [2]. Bagi pihak indihome tidaklah mudah dalam mengklasifikasi pelanggan baru maupun pelanggan lama agar tidak ada kerugian antara kedua pihak. Untuk mempermudah dalam proses pengklasifikasian penulis menggunakan algoritma C4.5 dalam mengklasifikasikan calon pelanggan baru indihome. Kelebihan utama Algoritma C4.5 dapat membuat pohon keputusan (decision tree) yang efisien menangani atribut tipe diskrit dan tipe diskrit- numerik, mudah untuk diinterpretasikan dan memiliki tingkat akurasi yang dapat diterima. Diharapkan penelitian ini dapat menganalisa penggunaan Algoritma C4.5 ini dapat mengukur dan mengklasifikasikan calon pelanggan baru Indihome sehingga tidak adanya kerugian pada kedua belah pihak, dan mengurangi kesalahan pemberian kelas jenis Indihome yang akan di pasang sehingga memaksimalkan data pelanggan yang belum optimal dalam memberi izin pemasangan Indihome dengan membagi data pelanggan ke dalam 3 kelompok yaitu kurang layak, layak, dan tidak layak

2. METODOLOGI PENELITIAN

Metode merupakan keseluruhan langkah ilmiah yang digunakan untuk mencari dan menemukan solusi dari suatu masalah. Metode Penelitian berkenaan dengan cara pengumpulan data yang selanjutnya data tersebut dapat menghantarkan pada pencapaian suatu tujuan, manfaat atau kegunaan tertentu [3].

2.1 Data Mining

Didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar [4].

2.2 Konsep Decision Tree

Decision tree merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode decision tree mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan. Aturan dapat dengan mudah

dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk basis data seperti Structure Query Language (SQL) untuk mencari record pada data tertentu [5].

2.3 Metode Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data yang dibutuhkan seperti studi literatur dan mencari berbagai sumber mulai dari perpustakaan, buku, prosiding atau jurnal sebagai media untuk bahan referensi dalam menentukan faktor, parameter dan label yang digunakan untuk penelitian.

2.4 Analisis Data

Dalam penelitian, data yang digunakan akan diolah dari hasil wawancara yang dilakukan selama 2 bulan, mulai bulan September sampai dengan Oktober 2019 sebanyak 600. Penulis mengambil sampel data dengan jumlah 100 sampel data. Berikut tabel pengambilan sampel data yang digunakan :

Tabel 1. Pengambilan Sampel Data Pelanggan

Semester	Jumlah Pelanggan September	Jumlah Pelanggan Oktober
September (2019)	296	45
Oktober (2019)	304	55
Jumlah	395	100

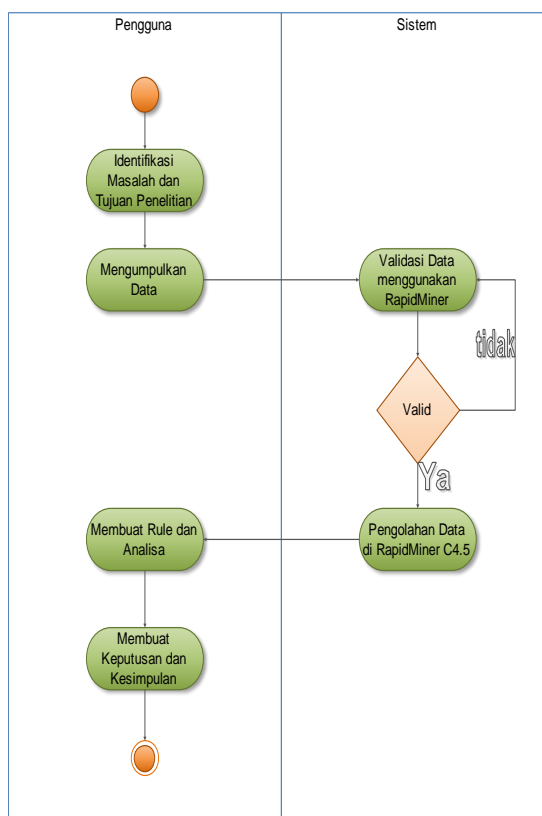
Data yang digunakan terdiri dari :

- Faktor Penghasilan,
- Faktor Pekerjaan,
- Faktor Status Rumah,
- Faktor Pelanggan (Pelanggan Baru/Sudah Pernah Masang Sebelumnya).

Setiap faktor terdiri dari pertanyaan yang diajukan ke pelanggan yang diberikan. Data yang digunakan menggunakan jenis statistik *deskriptif* dengan kepada Pelanggan Baru. Data diolah menggunakan *RapidMiner* dengan operator *Performance* yang berfungsi sebagai validasi dan reabilitas data untuk mencari keakuratan data.

2.5 Permodelan Activity Diagram

Alur kerja yang dilakukan penulis pada penelitian ini disajikan pada Gambar 1 dalam diagram aktivitas sebagai berikut:



Gambar 1. Activity Diagram Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Algoritma C4.5

Permasalahan yang sudah dianalisa, selanjutnya penulis menentukan faktor Klasifikasi Pada Calon Pelanggan Baru Indihome Dengan C.45 dengan menggunakan *Data Mining* algoritma C4.5. Algoritma C4.5 dilakukan dengan merancang atau membentuk pohon keputusan (*decision tree*) dengan memilih atribut pada variabel data sebagai akar (*node*), membuat cabang untuk setiap nilai, membagi kasus dalam cabang, dan mengulangi proses sampai semua cabang memiliki kelas yang sama. Perhitungan *Entropy* dan *Gain* dilakukan untuk membentuk suatu keputusan. Membuat pohon keputusan terlebih dahulu menghitung jumlah kasus untuk keputusan Layak dan jumlah kasus keputusan Tidak Layak. Berikut ini langkah-langkah dalam membentuk pohon keputusan dengan menggunakan Algoritma C4.5 :

- 1) Menghitung *Entropy* dari total data. Dari data penelitian diketahui jumlah kasus ada 80 *record*, diketahui status layak ada 70 *record* dan status tidak layak ada 10 *record* sehingga *Entropy* nilai akar total yang didapat:

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-70/80 . \log_2 (70/80)) + (-10/80 . \log_2 (10/80)) \\ &= 0,5436 \end{aligned}$$

- 2) Menghitung nilai *Entropy* pada atribut Pekerjaan dengan kategori ASN, jumlah kasus terdapat 20 *record*, terdapat status Layak sebanyak 2 *record* dan status Tidak layak sebanyak 0 *record*.

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-20/20 . \log_2 (20/20)) + (0) \\ &= 0 \end{aligned}$$

- 3) Seterusnya perhitungan *Entropy* pada Faktor Pekerjaan. Selanjutnya menghitung nilai *Gain* pada Atribut Faktor pekerjaan dengan kategori ASN terdapat 20 *record*, kategori Wiraswasta terdapat 54 *record*, dan kategori Wirausaha terdapat 6 *record* :

$$\begin{aligned} Gain(S,A) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\ &= (0,5436 - ((20/80)*0) + ((54/80)* 0,5564) + ((6/80) * 1)) \\ &= 0,0930 \end{aligned}$$

- 4) Selanjutnya Menghitung *SplitInfo* pada atribut Faktor Pekerjaan berdasarkan jumlah kasus pada Faktor Pekerjaan.

$$\begin{aligned} SplitInfo &= \sum_{i=1}^k p(v_i | s) \log_2 p(v_i | s) \\ &= -((20/80.\log_2 (20/80)) + (54/80.\log_2 (54/80) + (6/80.\log_2 \\ &\quad (6/80))) \\ &= 1,1630 \end{aligned}$$

- 5) Kemudian menghitung Nilai *GainRatio* Faktor Pekerjaan sebagai nilai yang digunakan sebagai pemecahan *Node* pada pohon keputusan.

$$\begin{aligned} RasioGain(s,j) &= \frac{Gain(s,j)}{SplitInfo(s,j)} \\ &= 0,0930 / 1,1630 \\ &= 0,0799 \end{aligned}$$

- 6) Menghitung nilai *Entropy* pada atribut Faktor Penghasilan dengan jumlah kasus terdapat 54 *record*, terdapat status Layak sebanyak 54 *record* dan status Tidak Layak sebanyak 0 *record*.

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-54/54 . \log_2 (54/54)) + (0)) = 0 \end{aligned}$$

- 7) Menghitung nilai *Entropy* pada atribut Penghasilan dengan kategori ASN, jumlah kasus terdapat 26 *record*, terdapat status Layak sebanyak 16 *record* dan status Tidak layak sebanyak 10 *record*.

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-16/26 . \log_2 (16/26)) + (-10/26 . \log_2 (10/26))) \\ &= 0,9612 \end{aligned}$$

- 8) Selanjutnya menghitung nilai *Gain* pada Atribut Faktor Penghasilan dengan kategori >2000000 terdapat 54 *record*, kategori <2000000 terdapat 26 *record*.

$$\begin{aligned} Gain(S,A) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\ &= 0,5436 - (((54/80)*0) + ((26/80)* 0,9612)) \\ &= 0,2312 \end{aligned}$$

- 9) Selanjutnya Menghitung *SplitInfo* pada atribut Faktor Penghasilan berdasarkan jumlah kasus pada Faktor Penghasilan.

$$\begin{aligned} SplitInfo &= \sum_{i=1}^k p(v_i | s) \log_2 p(v_i | s) \\ &= -((54/80.\log_2 (54/80)) + (26/80.\log_2 (26/80)) \\ &= 0,9097 \end{aligned}$$

- 10) Kemudian menghitung Nilai *Gain Ratio* Faktor Penghasilan sebagai nilai yang digunakan sebagai pemecahan *Node* pada pohon keputusan.

$$\begin{aligned} RasioGain(s,j) &= \frac{Gain(s,j)}{SplitInfo(s,j)} \\ &= 0,2312 / 0,9097 = 0,2541 \end{aligned}$$

Hasil dari perhitungan nilai *Entropy*, nilai *Gain*, *SplitInfo* dan *GainRatio* untuk tiap atribut dapat dilihat pada Tabel 2 sebagai berikut :

Tabel 2. Perhitungan *Gain Node 1*

No	Atribut	Jumlah Kasus	Layak	Tidak Layak	Entropy	Gain	Split Info	Gain Ratio
	Total	80	70	10	0,5436			
1	Pekerjaan					0,0930	1,1630	0,0799
	ASN	20	20	0				
	Wiraswasta	54	47	7	0,5564			
	Wirausaha	6	3	3	1,0000			
2	Penghasilan					0,2312	0,9097	0,2541
	>2000000	54	54	0				
	<2000000	26	16	10	0,9612			
3	Jenis Rumah					0,0825	0,8485	0,0972
	Pribadi	58	55	3	0,2937			
	Kontrak	22	15	7	0,9024			
4	Kelas Indihome					0,0387	1,2650	0,0306
	Ekonomi	42	34	8	0,7025			
	Premium	33	31	2	0,3298			
	Bisnis	5	5	0				

Dari Tabel 2 diketahui nilai *GainRatio* terbesar adalah atribut Penghasilan dengan nilai 0,2542, kategori Penghasilan <2000000 dengan nilai *Entropy* 0,9612, dan dijadikan simpul akar untuk mengulangi perhitungan nilai *Entropy* dan nilai *Gain* selanjutnya. Hasil perhitungan nilai *Entropy* dan nilai *Gain* sebagai Node 1.1 dapat dilihat pada Tabel 2 berikut :

Tabel 3. Perhitungan *Gain Node 1.1*

No	Atribut	Jumlah Kasus	Layak	Tidak Layak	Entropy	Gain	Split Info	Gain Ratio
1	Penghasilan <2000000	26	16	10	0,9612			
	Pekerjaan					0,1770	0,5159	0,3430
	ASN	0	0	0				
	Wiraswasta	23	16	7	0,8865			
2	Jenis Rumah					0,5014	0,8404	0,5967
	Pribadi	19	16	3	0,6292			
3	Kelas Indihome					0,1136	0,3912	0,2903
	Ekonomi	24	16	8	0,9183			
	Premium	2	0	2				
	Bisnis	0	0	0				

Dari Tabel 3 diketahui nilai *GainRatio* terbesar adalah atribut Jenis Rumah dengan nilai 0,5967 kategori Pribadi dengan nilai *Entropy* 0,6292, dan dijadikan simpul akar untuk mengulangi perhitungan nilai *Entropy* dan nilai *Gain* selanjutnya. Hasil perhitungan nilai *Entropy* dan nilai *Gain* sebagai Node 1.1.1. dapat dilihat pada Tabel 3 berikut :

Tabel 4. Perhitungan *Gain Node 1.1.1*

No	Atribut	Jumlah Kasus	Layak	Tidak Layak	Entropy	Gain	Split Info	Gain Ratio
1	Penghasilan <2000000 & Jenis Rumah Pribadi	19	16	3	0,6292			
	Pekerjaan					0,1525	0,2975	0,5126
	ASN	0	0	0				

No	Atribut	Jumlah Kasus	Layak	Tidak Layak	Entropy	Gain	Split Info	Gain Ratio	
2	Kelas Indihome	Wiraswasta	18	16	2	0,5033	0,1525	0,2975	0,5126
		wirusaha	1	0	1				
		Ekonomi	18	16	2	0,5033			
		Premium	1	0	1				
		Bisnis	0	0	0				

Dari tabel 4 diketahui nilai *GainRatio* terbesar adalah atribut Pekerjaan dengan nilai 0,5126 kategori Wiraswasta dengan nilai *Entropy* 0,5033, dan dijadikan simpul akar untuk mengulangi perhitungan nilai *Entropy* dan nilai *Gain* selanjutnya. Hasil perhitungan nilai *Entropy* dan nilai *Gain* sebagai Node 1.1.1.1 dapat dilihat pada Tabel 4 berikut :

Tabel 5. Perhitungan *Gain Node* 1.1.1.1

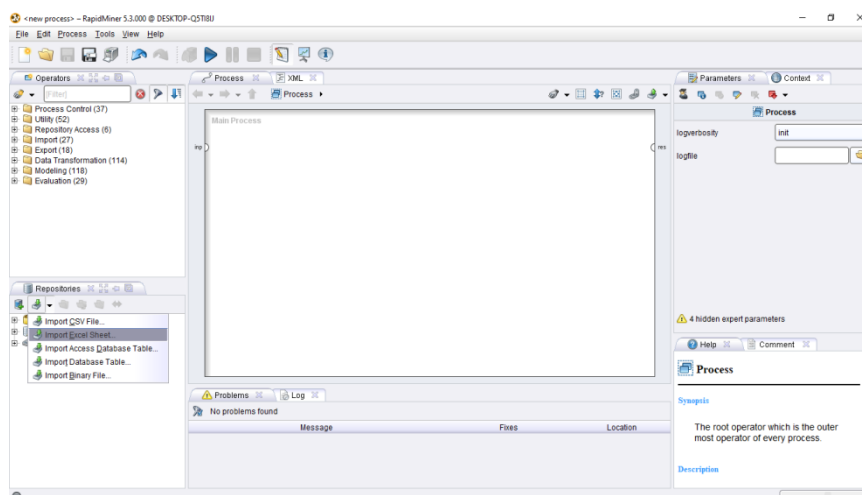
No	Atribut	Jumlah Kasus	Layak	Tidak Layak	Entropy	Gain	Split Info	Gain Ratio		
2	Kelas Indihome	Penghasilan <2000000 & Jenis Rumah Pribadi & Pekerjaan Wiraswasta		18	16	2	0,5033	0,1984	0,3095	0,6410
		Ekonomi	17	16	1	0,3228				
		Premium	1	0	1					
		Bisnis	0	0	0					

Dari tabel 5 diketahui nilai *GainRatio* terbesar adalah atribut Kelas Indihome dengan nilai 0,6410 dengan kategori Ekonomi dengan nilai *Entropy* 0,3228. Artinya bahwa *Node* 1.2.1 berhenti, sehingga didapat bentuk pohon keputusan yang dapat dilihat sebagai Aturan atau *Rule* model yang didapatkan adalah sebagai berikut :

- 1) Jika Penghasilan <2000000 dan Jenis Rumah = Kontrak, maka hasilnya Tidak Layak {Layak = 0 dan Tidak Layak = 7}
- 2) Jika Pekerjaan= Wiraswasta dan Kelas Indihome = Ekonomi maka hasilnya Layak (Layak = 16 dan Tidak Layak = 1)
- 3) Jika Pekerjaan= Wiraswasta dan Kelas Indihome = Premium maka hasilnya Tidak Layak (Layak = 0 dan Tidak Layak = 1)
- 4) Jika Pekerjaan= Wiraswasta dan Pekerjaan = Wirausaha maka hasilnya Tidak Layak (Layak = 0 dan Tidak Layak = 1)
- 5) Jika Pekerjaan= Wiraswasta dan Penghasilan > 2000000 maka hasilnya Layak (Layak = 54 dan Tidak Layak = 0)

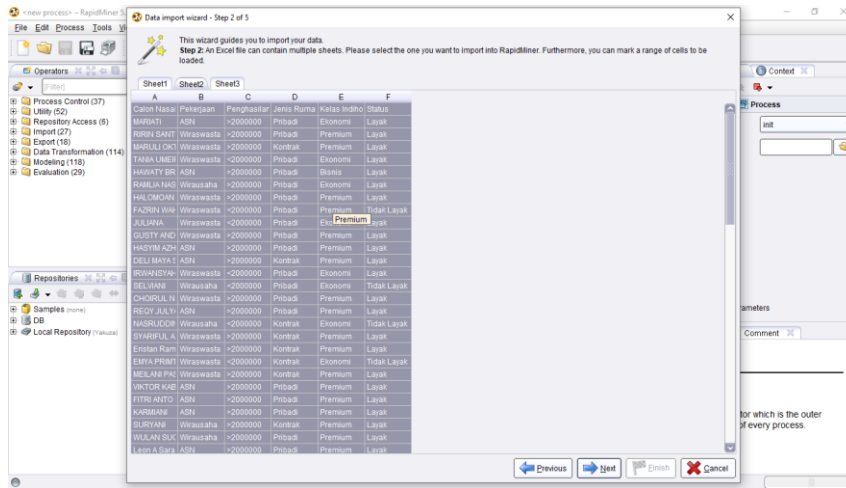
3.2 Implementasi Algoritma C4.5

Untuk membentuk model aturan penelitian, terlebih dahulu *importing* data dari data yang sudah diolah di *Microsoft Excel 2010* menggunakan perintah *Import Excel Sheet* pada *RapidMiner* agar bisa diakses menggunakan *software RapidMiner*. Dalam melakukan *import* data, dibutuhkan perintah seperti Gambar 2 berikut :



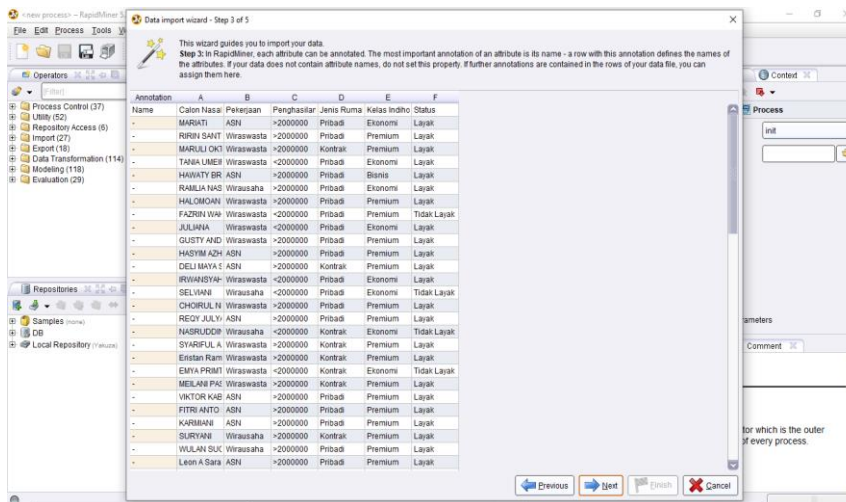
Gambar 2. Alur Proses *Import Data Step* 1

Data yang akan di *import* dalam bentuk *format Microsoft Excel*, pilih tool *Import Excel Sheet*. Fungsi *tool* ini untuk mengimpor data *format Microsoft Excel* dalam sebuah lembar (*sheet*) yang digunakan untuk penyeleksian dan pengolahan data. Pilih data yang akan digunakan untuk di *import* kemudian klik *Next* untuk mengelola data yang akan di *import*, seperti pada gambar 5 berikut :



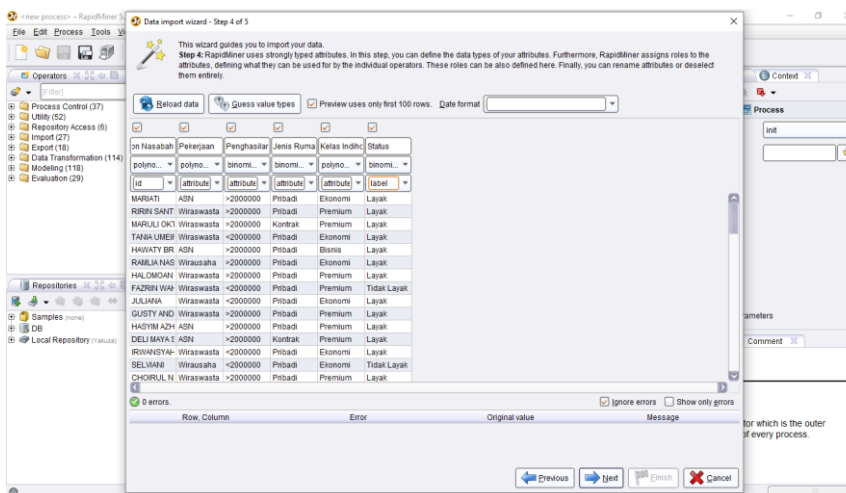
Gambar 3. Alur Proses *Import Data Step 2*

Tahapan untuk menentukan atribut dapat dilihat pada Gambar 4 berikut :

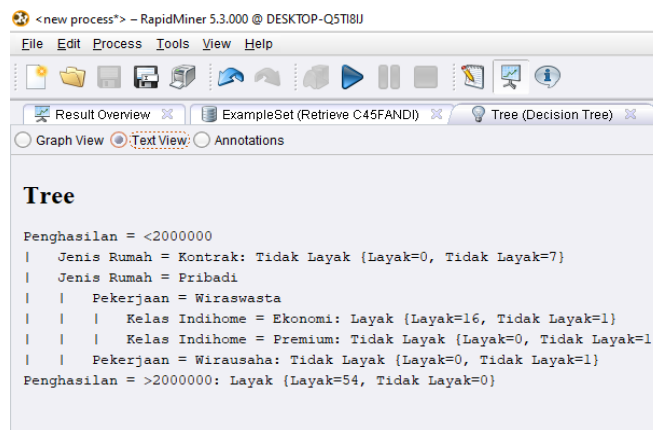


Gambar 4. Alur Proses *Import Data Step 3*

Tahapan selanjutnya dapat dilihat pada Gambar 5 berikut :



Gambar 5. Alur Proses *Import Data Step 4*

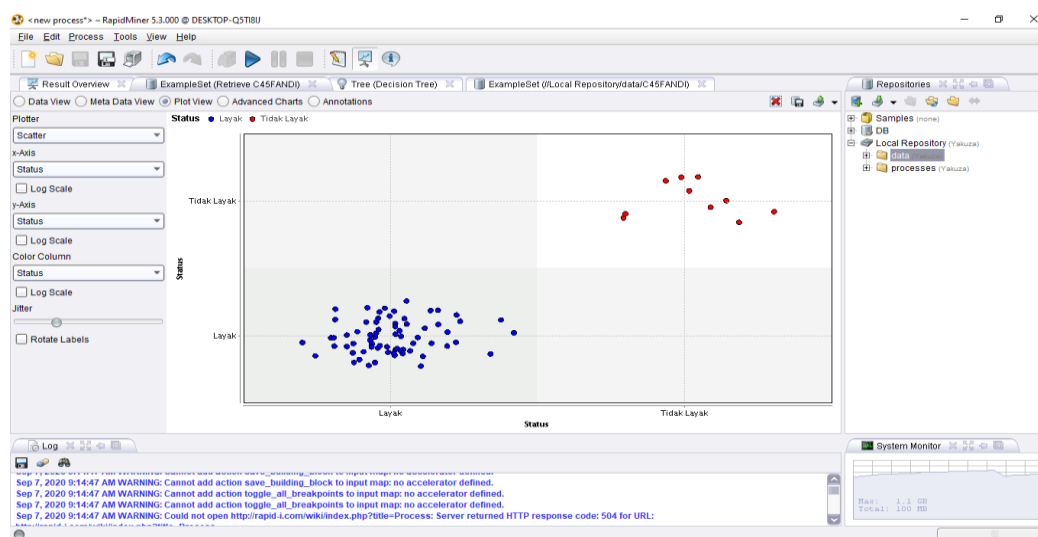


Gambar 6. Aturan atau *Rule* pada *RapidMiner*

Jika dilihat berdasarkan *text view* hasil pohon keputusan pada Gambar 8 peneliti membuat model aturan berupa teks berdasarkan Gambar 4.9 dengan penjelasan seperti berikut :

- Jika Penghasilan <2000000 dan Jenis Rumah = Kontrak, maka hasilnya Tidak Layak {Layak = 0 dan Tidak Layak = 7}
- Jika Pekerjaan= Wiraswasta dan Kelas Indihome = Ekonomi maka hasilnya Layak (Layak = 16 dan Tidak Layak = 1)
- Jika Pekerjaan= Wiraswasta dan Kelas Indihome = Premium maka hasilnya Tidak Layak (Layak = 0 dan Tidak Layak = 1)
- Jika Pekerjaan= Wiraswasta dan Pekerjaan = Wirausaha maka hasilnya Tidak Layak (Layak = 0 dan Tidak Layak = 1)
- Jika Pekerjaan= Wiraswasta dan Penghasilan > 2000000 maka hasilnya Layak (Layak = 54 dan Tidak Layak = 0)

Berdasarkan aturan atau model dari *text view* Gambar 6 dapat dijelaskan bahwa faktor yang mempengaruhi pada *node* pertama adalah Faktor Penghasilan, *node* kedua Faktor Jenis Rumah, *node* ketiga Faktor Pekerjaan, *node* keempat Faktor Kelas Indihome. Sehingga dari ketiga faktor harus ada.



Gambar 7. Aturan atau *Rule* pada *RapidMiner*

Gambar 7 dapat menjelaskan bahwa titik berwarna Biru (Layak) sebanyak 70 item dan titik berwarna Merah (Tidak Layak) sebanyak 10 item, artinya Pelanggan cenderung Layak mendapatkan pelayanan pemasangan Indihome.

4. KESIMPULAN

Diperoleh suatu model aturan yang dapat memperlihatkan persentase tertinggi dengan hasil tidak layak yaitu aturan Jika Penghasilan <2000000 dan Jenis Rumah = Kontrak, maka hasilnya Tidak Layak {Layak = 0 dan Tidak Layak = 7} dan hasil layak yaitu aturan. Jika Pekerjaan= Wiraswasta dan Penghasilan > 2000000 maka hasilnya Layak (Layak = 54 dan Tidak Layak = 0). Dengan diperolehnya hasil yang sama antara perhitungan manual dengan pengujian menggunakan *software RapidMiner*, maka dapat disimpulkan bahwa penelitian ini telah berhasil.

REFERENCES

- [1] A. H. Nasrullah, "Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out," *ILKOM Jurnal Ilmiah*, vol. 10, no. 2, p. 244, 2018, doi: 10.33096/ilkom.v10i2.300.244-250.
- [2] Y. S. Luvia, A. P. Windarto, S. Solikhun, and D. Hartama, "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa," *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, vol. 1, no. 1, p. 75, 2017, doi: 10.30645/jurasik.v1i1.12.
- [3] Muharto and A. Ambarita, "Metode Penelitian sistem informasi." deepublish, p. 151, 2016.
- [4] Y. iqbal maulana and Hendri, "Memprediksi Minat Pelanggan Membeli Produk Sepatu Menggunakan Metode Decision Tree Pada PT. Zoelindo Elite Anugerah," *Sentra Penelitian Engineering dan Edukasi*, vol. 10, no. 4, p. 70, 2018.
- [5] Asmira, "Penerapan Data Mining untuk Mengklasifikasi Pola Nasabah Menggunakan Algoritma C4,5 pada Bank BRI Unit Andounohu Kendari," *Jurnal Sistem Komputer dan Sistem Informasi*, vol. 1, no. 1, pp. 22–28, 2019.