

Penerapan Algoritma Naïve Bayes dan ID3 untuk Memprediksi Segmentasi Pelanggan pada Penjualan Mobil

Julianto Lemantara¹

¹ Program Studi/Jurusan S1 Sistem Informasi, Universitas Dinamika

Email: julianto@dinamika.ac.id

Abstrak: Memprediksi segmentasi pelanggan merupakan salah satu cara untuk memperluas peluang penjualan mobil dengan mengklasifikasikan calon pelanggan sehingga dapat meningkatkan pendapatan perusahaan. Sampai saat ini banyak pendekatan pemasaran hanya berorientasi pada produk. Pendekatan produk ini ternyata berdampak terhadap sulitnya mengidentifikasi pelanggan dengan tepat. Untuk itu diperlukan sebuah metode atau cara dalam memprediksi segmentasi pelanggan penjualan mobil. Perbandingan kinerja algoritma Naive Bayes dan ID3 bertujuan mengukur tingkat akurasi terbaik dari masing-masing algoritma dalam memprediksi segmentasi pelanggan penjualan mobil. Data yang diperoleh sebanyak 840 *dataset* yang terbagi atas data *training* sebanyak 700 *dataset* dan data *testing* sebanyak 140 *dataset*. Dengan menggunakan Weka 3.8.4, hasil perbandingan algoritma Naïve Bayes dan ID3 menunjukkan bahwa algoritma ID3 menghasilkan nilai akurasi 96,43% sedangkan Naïve Bayes menghasilkan nilai akurasi 56,43%. Dari data tersebut diketahui bahwa metode ID3 dinilai lebih baik daripada metode Naïve Bayes dalam memprediksi segmentasi pelanggan penjualan mobil.

Kata Kunci: Algoritma ID3, Naïve Bayes, Segmentasi Pelanggan.

Abstract: Predicting customer segmentation is one way to widen car sales opportunities by classifying potential customers so that it can increase company revenue. So far, the existing marketing approach is only product oriented. The product approach turned out to have an impact on the difficulty of correctly identifying customers. Therefore we need a method or way to predict customer segmentation of car sales. Comparison of the performance of the Naive Bayes and ID3 Algorithms aims to measure the best level of accuracy of each algorithm to be applied in the prediction of car sales customer segmentation. The data obtained were 840 datasets and divided into training data of 700 datasets and testing data of 140 datasets. By using Weka 3.8.4, the comparison results of the Naive Bayes and ID3 Algorithms showed that ID3 Algorithm gives an accuracy value of 96.43%, while Naïve Bayes Algorithm gives an accuracy value of 56.43%. From these data it is known that the ID3 method is considered better than the Naive Bayes method in predicting customer segmentation of car sales.

Keywords: Customer Segmentation, ID3 Algorithm, Naïve Bayes

PENDAHULUAN

Pelanggan merupakan suatu faktor yang sangat penting bagi kelangsungan hidup perusahaan, baik perusahaan dagang, perusahaan jasa, maupun perusahaan manufaktur. Perusahaan dapat mengalami kerugian yang sangat besar jika ditinggalkan para pelanggan [1]. Pelanggan bagi perusahaan otomotif adalah salah satu aset yang utama. Oleh karena itu, berbagai cara harus ditempuh perusahaan agar pelanggan tetap loyal dan melanjutkan pembelian mobil secara berkelanjutan. Perusahaan harus mampu mengenali pelanggan yang potensial sehingga perusahaan dapat mempertahankan pelanggan dengan lebih mudah. Tujuan dari mempertahankan pelanggan (*customer retention*) ini adalah agar pelanggan tidak berhenti membeli produk, apalagi berpindah ke perusahaan kompetitor (*churn*) [2].

Selama ini banyak perusahaan masih menggunakan pendekatan pemasaran berorientasi kepada produk sehingga banyak perusahaan hanya menaruh perhatian utama pada produk yang dibuat. Pendekatan produk ini ternyata dapat berdampak

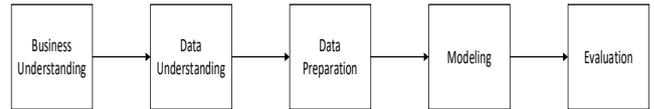
terhadap sulitnya melakukan identifikasi pelanggan dengan tepat. Hal tersebut mengubah sudut pandang perusahaan mengenai siklus perusahaan. Apabila sebelumnya perusahaan mempunyai sudut pandang berorientasi produk, sekarang berganti dengan sudut pandang berorientasi pelanggan [3]. Bagian pemasaran dari suatu perusahaan mempunyai peranan penting dalam menghadapi kompetisi yang semakin hari semakin ketat. Pada umumnya, perusahaan yang memiliki orientasi penjualan produk ke pelanggan menghadapi masalah dalam bidang pemasaran. Survei pasar diperlukan dalam rangka memperoleh informasi mengenai kebutuhan dan permintaan spesifik dari pelanggan. Oleh karena itu, perusahaan wajib mempertimbangkan karakteristik dari masing-masing pelanggan. Segmentasi pelanggan adalah salah satu strategi pemasaran yang bisa ditempuh perusahaan guna mempertahankan pelanggan yang potensial serta memperoleh keuntungan yang maksimal [4]. Hal ini penting dilakukan untuk pengembangan perusahaan dengan menjalin hubungan baik antara perusahaan dan pelanggan.

Prediksi segmentasi pelanggan penjualan mobil sangat diperlukan oleh perusahaan untuk mengetahui segmentasi pelanggan yang potensial. Berbagai pendekatan prediksi atau klasifikasi telah dikembangkan beberapa tahun terakhir ini, salah satunya yaitu pendekatan data mining [5][6]. Klasifikasi termasuk *supervised learning* yang pada prosesnya memerlukan label guna mengekstraksi model yang dipakai dalam melakukan prediksi [7]. Proses segmentasi pelanggan pernah dilakukan dengan metode Naïve Bayes pada penelitian [8]. Hasil penelitian menunjukkan Naïve Bayes mampu melakukan prediksi pelanggan loyal dengan tingkat akurasi sebesar 97,27%, *recall* sebesar 96,98%, dan *precision* sebesar 100%. Penelitian terkait prediksi pelanggan juga pernah dilakukan, terutama dalam hal memprediksi *churn* berdasarkan kelas pelanggan retail [9]. Hasil penelitian tersebut menunjukkan Naïve Bayes lebih baik dalam hal prediksi *churn* daripada *decision tree* C4.5. Tingkat akurasi Naïve Bayes untuk semua kelas pelanggan yaitu 83.49%, sedangkan C4.5 memiliki akurasi 80.6%. Penelitian sejenis terkait segmentasi pelanggan pernah dilakukan pula dengan metode *decision tree* J-48, Zero-R dan Naïve Bayes. Hasil uji coba menunjukkan aplikasi segmentasi pelanggan menggunakan metode *decision tree* J-48 memiliki tingkat *error rate* sebesar 5,8769%, metode Zero-R memiliki tingkat *error rate* sebesar 11,6985%, dan metode Naïve Bayes memiliki tingkat *error rate* sebesar 11,944% [10]. Hasil penelitian ini berbeda dengan penelitian-penelitian yang disebutkan sebelumnya karena Naïve Bayes justru mempunyai akurasi lebih buruk dan *error rate* yang lebih tinggi.

Penelitian terkait prediksi dengan teknik klasifikasi ini sudah banyak dilakukan. Ada pula penelitian yang menerapkan algoritma Naïve Bayes dan ID, meskipun penelitian ini tidak terkait dengan prediksi segmentasi pelanggan. Penelitian yang dimaksud adalah penelitian tentang klasifikasi penyakit diabetes Mellitus [11]. Penelitian tersebut menghasilkan temuan bahwa Naïve Bayes memiliki akurasi lebih baik dari ID3. Naïve Bayes memiliki akurasi 76%, sedangkan ID3 memiliki akurasi 74%. Pada penelitian ini, algoritma atau metode klasifikasi yang digunakan sama seperti penelitian Nurdiana dan Algifari, yaitu Naïve Bayes dan ID3, tetapi diterapkan untuk memprediksi segmentasi pelanggan penjualan mobil berdasarkan data yang ada.

METODE

Dalam pengolahan data, penelitian ini menggunakan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Adapun tahapan yang harus dilakukan pada metode CRISP-DM adalah *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, dan *Evaluation* [12]. Untuk lebih jelasnya, tahapan penelitian dengan metode CRISP-DM dapat dilihat di gambar 1.



Gambar 1. Tahapan Penelitian

Business Understanding

Business Understanding merupakan suatu proses memahami tujuan bisnis dan melakukan penerjemahan tujuan bisnis ke dalam tujuan penggalian data. Penelitian memerlukan pengetahuan guna mengetahui segmentasi pelanggan penjualan mobil yang bisa digunakan dalam membantu manajemen perusahaan untuk melakukan prediksi segmentasi pelanggan. Dengan begitu, pihak manajemen perusahaan dapat mengambil kebijakan promosi penjualan yang sesuai dan tepat sasaran.

Data Understanding

Pada tahap *Data Understanding*, proses pengumpulan data dilakukan dengan sebaik mungkin. Setelah itu, dilakukan proses analisis data dan evaluasi kualitas data. Penelitian mengambil sumber data sekunder yang berasal dari *Kaggle Machine Learning and Data Science Community*. Data ini merupakan data kuantitatif yaitu data pelanggan mobil, dengan 9 atribut dan 1 label. Atribut yang digunakan adalah *ever married*, *graduated*, *profession*, *spending score*, *work experience*, *age*, *gender*, *family size*, *var1*, sedangkan label yang digunakan adalah *segmentation*.

Data Preparation

Hal yang dilakukan pada tahap ini meliputi pemilihan data dan pengelompokan atribut yang dibutuhkan dalam melakukan prediksi segmentasi pelanggan penjualan mobil. Adapun atribut yang digunakan pada penelitian ini dapat dilihat di tabel 1.

Tabel 1. Atribut Segmentasi Pelanggan

Atribut	Keterangan	Nilai
Gender	Jenis kelamin	1. Male
		2. Female
Ever Married	Status pernikahan	1. Yes
		2. No
Age	Usia (satuan: tahun)	1. 10 – 19
		2. 20 – 29
		3. 30 – 39
		4. 40 – 49
		5. 50 – 59
		6. 60 – 69
Graduated	Status kelulusan dari pendidikan	7. 70 – 79
		8. 80 – 89
		9. 90 – 99
Profession	Bidang profesi	1. Healthcare

Atribut	Keterangan	Nilai
		2. <i>Engineer</i>
		3. <i>Lawyer</i>
		4. <i>Entertainment</i>
		5. <i>Artist</i>
		6. <i>Doctor</i>
		7. <i>Homemaker</i>
		8. <i>Marketing</i>
		9. <i>Executive</i>
<i>Work Experience</i>	Pengalaman kerja (satuan: tahun)	1. 0 – 4 2. 5 – 9 3. 10 – 14
<i>Spending Score</i>	Nilai tingkat pengeluaran pelanggan	1. <i>High</i> 2. <i>Average</i> 3. <i>Low</i>
<i>Family Size</i>	Jumlah anggota keluarga, termasuk pelanggan	1. 1 – 3 2. 4 – 6 3. 7 – 9
Var 1	Kategori anonim untuk pelanggan	1. Var 1 2. Var 2 3. Var 3 4. Var 4 5. Var 5 6. Var 6 7. Var 7
<i>Segmentation</i>	Target: segmentasi pelanggan	1. A 2. B 3. C 4. D

Modeling

Setelah melakukan tahap pengumpulan data dan peninjauan penelitian-penelitian terdahulu, maka penelitian ini memutuskan untuk menerapkan algoritma *Naïve Bayes* dan ID3 dalam mprediksi segmentasi pelanggan penjualan mobil. Perangkat lunak yang digunakan untuk pengolahan data pada penelitian ini yaitu Weka 3.8.4. Untuk model validasi klasifikasi digunakan *Supplied Test Set* yang berguna untuk menguji seberapa baik prediksi yang dihasilkan dari *data training* [13]. Data yang diolah dalam penelitian ini sejumlah 840 data, kemudian dibagi ke dalam data *training* sejumlah 700 dan data *testing* sejumlah 140. Untuk mendapatkan sampel data *training* atau data latih terbaik, penelitian ini melakukan beberapa kali penarikan sampel hingga didapatkan tingkat akurasi model yang tinggi.

Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah teknik pengklasifikasian yang menggunakan metode statistik dan probabilitas [14]. Algoritma ini memprediksi peluang di masa mendatang berdasarkan pada pengalaman di masa lampau [15]. Salah satu keunggulan dari algoritma ini yaitu tidak memerlukan jumlah data

training yang terlalu banyak untuk dapat melakukan proses pengklasifikasian atau prediksi [15]–[17]. Algoritma ini juga mudah digunakan karena alur perhitungan yang relatif pendek dan sederhana [16]. Algoritma ini sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks daripada yang diekspektasikan [17]. Secara umum, metode *Naïve Bayes* menggunakan persamaan 1 [18].

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \tag{1}$$

Keterangan:

X : data *sample* dengan kelas (label) yang tidak diketahui

H : hipotesis bahwa data X merupakan suatu kelas spesifik

P(H/X) : peluang hipotesis H berdasarkan kondisi X (*posteriori probability*)

P(H) : peluang hipotesis H (*prior probability*)

P(X) : peluang data *sample* yang diamati

P(X/H) : peluang X berdasarkan kondisi hipotesis H

Untuk menghindari nilai probabilitas nol pada proses prediksi yang menyebabkan *Naïve Bayes* tidak bisa melakukan klasifikasi sebuah data inputan secara baik, maka terdapat teknik yang bisa diterapkan yaitu *Laplace Correction*. *Laplace* ini merupakan suatu teknik yang menambah nilai 1 pada setiap kombinasi atribut [19]. Pada penerapan dengan volume data yang besar, teknik *laplace* ini memiliki akurasi yang tinggi dikarenakan tidak membuat selisih (perbedaan) yang signifikan pada estimasi probabilitas. *Laplace correction* dapat dihitung dengan persamaan 2 [19][20].

$$P(H|X) = \frac{P(H|X)P(H)+1}{P(X)+|V|} \tag{2}$$

Keterangan:

P(X) : peluang data *sample* yang diamati

|V| : banyaknya *term* pada data uji yang digunakan

Algoritma Decision Tree ID3

Decision tree merupakan algoritma yang sering digunakan dalam banyak penelitian dikarenakan algoritma ini mudah diinterpretasikan berkat strukturnya yang sederhana [21]. Salah satu algoritma untuk membuat pohon keputusan (*decision tree*) ialah algoritma *Iterative Dichotomiser 3* (ID3). Masukan dari algoritma ID3 ini adalah sebuah basis data dengan beberapa variabel atau atribut [22].

Proses klasifikasi dilakukan dari akar (*root*), lalu bercabang-cabang hingga diperoleh *node* daun (*leaves*). *Leaves* ini menunjukkan hasil akhir dari klasifikasi. Setiap objek yang diklasifikasi pada pohon keputusan harus diuji nilai entropinya. Entropi ini menunjukkan ukuran yang digunakan untuk mengetahui homogenitas dari kumpulan data [23]. *Entropy* digunakan untuk menilai seberapa penting sebuah *node* [24]. Perhitungan nilai *entropy* dapat menggunakan persamaan 3. Sesudah memperoleh nilai entropi,

pemilihan atribut dilakukan berdasarkan nilai *information gain* yang terbesar. *Information gain* merupakan tolok ukur pemilihan atribut untuk dijadikan sebagai akar maupun *node* pada pohon keputusan [23]. Nilai *Information gain* dihitung dengan persamaan 4. Entropi (S) = 0, jika semua contoh dalam S berada di kelas yang sama
 Entropi (S) = 1, jika jumlah contoh positif sama dengan jumlah contoh negatif dalam S
 $0 < \text{Entropi (S)} < 1$, jika jumlah contoh positif dan negatif dalam S berbeda

$$\text{Entropi (S)} = \sum_{j=1}^k - P_j \log_2 P_j \quad (3)$$

Keterangan:

- S : himpunan atau dataset kasus
- k : jumlah partisi S
- P_j : peluang suatu kelas dibagi total kasus

$$\text{Gain (A)} = \text{Entropi (S)} - \sum_{i=1}^k \frac{|S_i|}{|S|} \times \text{Entropi}(S_i) \quad (4)$$

Keterangan:

- S : ruang (data) sampel yang digunakan *training*
- A : atribut
- [S_i] : jumlah sampel untuk nilai V
- [S] : jumlah seluruh sampel data
- Entropi (S_i) : entropi untuk sampel-sampel yang memiliki nilai *i*

Evaluation

Pada tahap *evaluation*, hal yang dilakukan yaitu menguji model yang sudah dibuat dan mengevaluasi keakuratan dan generalisasinya. Tahap ini mengukur sejauh mana model dapat memenuhi tujuan bisnis.

HASIL DAN PEMBAHASAN

Pada bagian ini memuat hasil implementasi algoritma Naïve Bayes dan ID3 untuk memprediksi segmentasi pelanggan penjualan mobil. Pada penelitian ini digunakan *Supplied Test Set* untuk mencari kesalahan klasifikasi model pada data. Pada tabel 2 dapat dilihat hasil *Supplied Test Set*, bahwa tingkat kebenaran dari hasil klasifikasi menggunakan algoritma ID3 sebanyak 135 dari 140 data sehingga diperoleh tingkat akurasi model sebesar 96,4286%. Hasil klasifikasi menggunakan algoritma ID3 ini memiliki tingkat kesalahan sejumlah 5 dari 140 data, yaitu sebesar 3,5714%.

Tingkat akurasi ini juga didukung oleh hasil *Precision* dan *Recall* yang disajikan pada tabel 2. *Precision* adalah *metric* untuk mengukur kinerja sistem dalam mendapatkan data yang relevan. Dalam penelitian ini, nilai *Precision* yang diperoleh untuk label A yaitu 0,971, untuk label B yaitu 0,889, untuk label C yaitu 0,976, dan untuk label D yaitu 0,979. Sementara itu, *Recall* adalah *metric* untuk mengukur proporsi positif asli yang diprediksi secara benar sebagai positif. Nilai *Recall* yang diperoleh sebesar 0,917 untuk label A, 1,000 untuk label B, 1,000 untuk label C, dan 0,958 untuk label D.

Tabel 2. *Confusion Matrix* Algoritma ID3

	A	B	C	D	<i>Precision</i>
A	33	1	1	1	0,971
B	0	16	0	0	0,889
C	0	0	40	0	0,976
D	1	1	0	46	0,979
<i>Recall</i>	0,917	1,000	1,000	0,958	

Pada tabel 3 dapat dilihat hasil *Supplied Test Set*, bahwa tingkat kebenaran dari hasil klasifikasi dengan algoritma Naïve Bayes berjumlah 79 dari 140 data sehingga diperoleh tingkat akurasi model sebesar 56,4286%. Hasil klasifikasi menggunakan algoritma Naïve Bayes memiliki tingkat kesalahan sejumlah 61 dari 140 data, yaitu sebesar 43,5714%.

Tabel 1. *Confusion Matrix* Algoritma Naïve Bayes

	A	B	C	D	<i>Precision</i>
A	13	3	12	8	0,394
B	4	6	6	0	0,500
C	3	3	30	4	0,566
D	13	0	5	30	0,714
<i>Recall</i>	0,361	0,375	0,750	0,625	

Tingkat akurasi ini didukung pula oleh hasil *Precision* dan *Recall* yang diperlihatkan pada tabel 3. Pada penelitian ini diperoleh nilai *Precision* pada label A yaitu 0,394 pada label B yaitu 0,500 pada label C yaitu 0,566 pada label D yaitu 0,714. Pada penelitian ini diperoleh nilai *Recall* 0,361 pada label A, 0,375 pada label B, 0,750 pada label C, dan 0,625 pada label D.

Jika dibandingkan antara hasil klasifikasi ID3 dengan hasil klasifikasi Naïve Bayes untuk prediksi segmentasi pelanggan pada penjualan mobil ini, maka ID3 memiliki akurasi, *precision*, dan *recall* yang jauh lebih baik. Dengan demikian, ID3 lebih cocok digunakan pada studi kasus segmentasi pelanggan penjualan mobil ini. Namun, pada kasus penelitian lain seperti penelitian [11], Naïve Bayes justru memiliki hasil klasifikasi atau hasil prediksi yang lebih baik daripada ID3 dalam hal klasifikasi diabetes. Pada penelitian [9], Naïve Bayes juga mempunyai tingkat akurasi yang lebih tinggi dari *decision tree* C4.5 (bukan ID3) dalam hal segmentasi pelanggan. Hal tersebut dikarenakan pada *decision tree* C4.5 terdapat *pruning* atau pemangkasan cabang pohon sehingga *decision tree* menjadi tidak utuh, tetapi proses komputasi lebih cepat daripada ID3 yang utuh. Secara umum, tidak dapat disimpulkan bahwa *decision tree* lebih baik daripada Naïve Bayes, demikian pula sebaliknya. Hal tersebut sangat bergantung pada karakteristik dan kualitas data, baik data training maupun data testing yang digunakan pada setiap penelitian.

Dengan penerapan algoritma *Naïve Bayes* dapat diperoleh aturan (*rule*) yang bisa digunakan untuk memprediksi segmentasi pelanggan penjualan mobil. Hasil aturan (*rule*) sebagai berikut:

$P(C_i)$
 $P(\text{Segmentation} = \text{"A"}) = 185/700 = 0,26420455$
 $P(X|C_i)$
 $P(\text{Gender} = \text{"Male"} | \text{Segmentation} = \text{"A"}) = 95/185 = 0,51336898$
 $P(\text{Gender} = \text{"Female"} | \text{Segmentation} = \text{"A"}) = 90/185 = 0,48663102$

$P(\text{Ever Married} = \text{"No"} | \text{Segmentation} = \text{"A"}) = 73/185 = 0,39037433$
 $P(\text{Ever Married} = \text{"Yes"} | \text{Segmentation} = \text{"A"}) = 112/185 = 0,60962567$

Hasil dari perhitungan lebih lanjut terkait metode *Naïve Bayes* dengan *software Weka* dapat dilihat pada gambar 2.

Class A: P(C) = 0.26420455										Class B: P(C) = 0.20880682																																																																					
Attribute Gender																																																																															
Male										Female																																																																					
0.51336898					0.48663102					0.49324324					0.50675676																																																																
Attribute Ever_Married																																																																															
No										Yes																																																																					
0.39037433					0.60962567					0.29054054					0.70945946																																																																
Attribute Age																																																																															
20 - 29										30 - 39										60 - 69										40 - 49										50 - 59										10 - 19										70 - 79										80 - 89									
0.20725389		0.31088083		0.06735751		0.19689119		0.11917098		0.02072539		0.03108808		0.04663212		0.0974026		0.24675325		0.11688312		0.27922078		0.14935065		0.00649351		0.05844156																																																			
Attribute Graduated																																																																															
No										Yes																																																																					
0.35828877					0.64171123					0.35135135					0.64864865																																																																
Attribute Profession																																																																															
Healthcare Engineer Lawyer Entertainment Artist Doctor Homemaker Marketing Executive																																																																															
0.10824742		0.1443299		0.07731959		0.1443299		0.27835052		0.08247423		0.02061856		0.05670103		0.08762887		0.05806452		0.1483871		0.10322581		0.10322581		0.38709677		0.07741935		0.038709680																																																	
Attribute Work_Experience																																																																															
0 - 4										5 - 9										10 - 14										0 - 4																																																	
0.71957672		0.21164021		0.06349206		0.00529101		0.82		0.14666667		0.02666667		0.00666667		0.44966443		0.32214765		0.22818792																																																											
Attribute Spending_Score																																																																															
Low										Average										High																																																											
0.71276596		0.15957447		0.12765957		0.44966443		0.32214765		0.22818792																																																																					
Attribute Family_Size																																																																															
4 - 6										1 - 3										7 - 9																																																											
0.18617021		0.78723404		0.02659574		0.31543624		0.66442953		0.02013423																																																																					
Attribute Var_1																																																																															
Cat_4										Cat_6										Cat_7										Cat_3										Cat_1										Cat_2										Cat_5																			
0.1875		0.58333333		0.046875		0.09895833		0.03125		0.046875		0.00520833		0.1503268		0.62745098		0.02614379		0.07189542		0.02614379		0.07189542		0.02614379																																																					

Gambar 2. Hasil WEKA Menggunakan Metode Naïve Bayes

Dengan penerapan algoritma ID3 dapat diperoleh aturan (*rule*) dari pohon keputusan yang bisa digunakan untuk memprediksi segmentasi pelanggan penjualan mobil. Hasil aturan (*rule*) dari struktur pohon keputusan ID3 dapat dilihat pada tabel 4 dan tabel 5. Hasil dari perhitungan metode ID3 lebih lanjut dengan

software Weka dapat dilihat pada gambar 3, gambar 4, gambar 5, gambar 6, dan gambar 7. Kelima gambar tersebut menunjukkan model pohon keputusan dalam menentukan segmentasi pelanggan, apakah pelanggan termasuk segmen A, B, C, atau D.

Tabel 4. Nilai Entropi ID3

Total Kasus	Sum (A)	Sum (B)	Sum (C)	Sum (D)	Entropi Total
700	185	146	182	187	1,993057049

Tabel 5. Contoh Perhitungan Manual ID3

Node	Atribut	Nilai	Sum (Nilai)	Sum (A)	Sum (B)	Sum (C)	Sum (D)	Entropi	Gain
1	Gender	Male	361	95	72	96	98	1,989570130	
		Female	339	90	74	86	89	1,995804083	
									0,000467904
	Ever Married	Yes	412	113	104	141	54	1,926872811	
		No	288	72	42	41	133	1,820188446	
									0,110077234
	Age	10 - 19	25	3	0	1	21	0,764114055	
		20 - 29	147	39	14	15	79	1,648409124	
		30 - 39	170	59	37	31	43	1,958007498	
		40 - 49	151	37	42	51	21	1,935361927	
		50 - 59	93	22	22	46	3	1,646107965	
		60 - 69	52	12	17	19	4	1,830881236	
		70 - 79	38	5	8	16	9	1,875849474	
		80 - 89	24	8	6	3	7	1,921789711	
		90 - 99	0	0	0	0	0	0	
									0,20417274
	Graduated	Yes	423	119	95	149	60	1,928558435	
		No	277	66	51	33	127	1,824024962	
									0,105864002
	Profession	Healthcare	128	20	8	16	84	1,442240426	
		Engineer	70	27	22	7	14	1,851512852	
		Lawyer	52	14	15	9	14	1,974696992	
		Entertainment	69	27	15	14	13	1,928901171	
		Artist	228	53	59	104	12	1,734113678	
		Doctor	60	15	11	16	18	1,978294650	
		Homemaker	17	3	5	1	8	1,713078843	
		Marketing	30	10	3	4	13	1,770894436	
		Executive	46	16	8	11	11	1,956007704	
								0,226925300	
Work Experience	0 - 4	558	135	122	157	144	1,993933644		
	5 - 9	123	39	21	22	41	1,933274829		
	10 - 14	19	11	3	3	2	1,639321050		
								0,019407224	
Spending Score	High	96	23	33	26	14	1,938906954		
	Average	185	29	47	96	13	1,681601778		
	Low	419	133	66	60	160	1,877382790		
								0,158978784	
Family Size	1 - 3	479	147	98	120	114	1,984529586		
	4 - 6	210	34	46	60	70	1,949860596		
	7 - 9	11	4	2	2	3	1,936260028		
								0,019686681	
Var 1	Cat_1	16	5	3	4	4	1,977217001		
	Cat_2	36	8	10	5	13	1,921740619		
	Cat_3	56	18	10	12	16	1,962757047		
	Cat_4	96	35	22	9	30	1,862374542		
	Cat_5	8	0	3	1	4	1,405639062		
	Cat_6	463	111	95	143	114	1,984182624		
	Cat_7	25	8	3	8	6	1,91326969		
								0,039808628	

Profession = Healthcare	Graduated = No	Family_Size = 7 - 9:	Graduated = Yes:	Profession = Engineer
Age = 20 - 29	Work_Experience =	null	A	Age = 20 - 29
Var_1 = Cat_4	0 - 4	Graduated = Yes	Family_Size = 7 - 9: null	Family_Size = 4 - 6
Graduated = No	Family_Size = 4 -	Work_Experience = 0 -	Work_Experience = 5 - 9	Var_1 = Cat_4: D
Family_Size = 4 - 6: D	6: D	4: A	Gender = Male: D	Var_1 = Cat_6: D
Family_Size = 1 - 3	Family_Size = 1 -	Work_Experience = 5 -	Gender = Female	Var_1 = Cat_7: null
Gender = Male: D	3: D	9: C	Graduated = No	Var_1 = Cat_3: null
Gender = Female: D	Family_Size = 7 -	Work_Experience = 10	Family_Size = 4 - 6:	Var_1 = Cat_1: null
Family_Size = 7 - 9: null	9: null	- 14: null	Family_Size = 1 - 3:	Var_1 = Cat_2: D
Graduated = Yes	Work_Experience =	Work_Experience = 0 -	Family_Size = 7 - 9:	Var_1 = Cat_5: null
Gender = Male: B	5 - 9: null	4 : null	A	Family_Size = 1 - 3
Gender = Female	Work_Experience =	Gender = Female: D	Family_Size = 7 - 9:	Work_Experience = 0 - 4
Ever_Married = No	10 - 14: D	Var_1 = Cat_1: D	null	Var_1 = Cat_4
Work_Experience =	Work_Experience =	Var_1 = Cat_2	Graduated = Yes: D	Ever_Married = No: D
0 - 4: B	0 - 4 : null	Gender = Male: C	Work_Experience = 10 - 14	Ever_Married = Yes: A
Work_Experience =	Graduated = Yes	Gender = Female: D	Ever_Married = No: A	Var_1 = Cat_5: A
5 - 9: A	Ever_Married = No	Var_1 = Cat_5: D	Ever_Married = Yes: B	Var_1 = Cat_7: null
Work_Experience =	Family_Size = 4 - 6	Age = 30 - 39	Work_Experience = 0 - 4 :	Var_1 = Cat_3: null
10 - 14: null		Var_1 = Cat_4	null	Var_1 = Cat_2: null
Work_Experience =	Work_Experience = 0 - 4: D	Family_Size = 4 - 6	Var_1 = Cat_7: B	Var_1 = Cat_2: A
0 - 4 : null		Gender = Male: D	Var_1 = Cat_3	Var_1 = Cat_5: null
Ever_Married = Yes: A	Work_Experience = 5 - 9: D	Gender = Female	Gender = Male: A	Work_Experience = 5 - 9
Var_1 = Cat_5		Graduated = No: C	Gender = Female: D	Graduated = No: A
Spending_Score = Low	Work_Experience = 10 - 14: null	Graduated = Yes: D	Var_1 = Cat_1: C	Graduated = Yes: B
Gender = Male		Family_Size = 1 - 3: A	Var_1 = Cat_2	Work_Experience = 10 - 14:
Family_Size = 4 - 6	Work_Experience = 0 - 4 : null	Family_Size = 7 - 9: A	Gender = Male: A	D
Work_Experience =	Family_Size = 1 -	Var_1 = Cat_6	Gender = Female: C	Work_Experience = 0 - 4 :
0 - 4: D	3: D	Work_Experience = 0 - 4	Var_1 = Cat_5: D	null
Work_Experience =	Family_Size = 7 -	Family_Size = 4 - 6	Age = 60 - 69: null	Family_Size = 7 - 9: D
5 - 9: D	9: null	Ever_Married = No	Age = 40 - 49	Age = 30 - 39
Work_Experience =	Ever_Married = Yes:	Graduated = No: D	Var_1 = Cat_4: null	Var_1 = Cat_4
10 - 14: null	D	Graduated = Yes: D	Var_1 = Cat_6	Work_Experience = 0 - 4
Work_Experience =	Spending_Score = Average:	Ever_Married = Yes: B	Gender = Male	Gender = Male: D
0 - 4 : null	A	Family_Size = 1 - 3	Ever_Married = No: D	Gender = Female: A
Family_Size = 1 - 3	Spending_Score = High:	Ever_Married = No	Ever_Married = Yes: A	Work_Experience = 5 - 9: D
Work_Experience =	null	Graduated = No: D	Gender = Female: A	Work_Experience = 10 - 14:
0 - 4: D	Var_1 = Cat_7	Graduated = Yes	Var_1 = Cat_7: D	null
Work_Experience =	Family_Size = 4 - 6: D	Graduated = No: D	Var_1 = Cat_3: null	Work_Experience = 0 - 4 :
5 - 9: D	Family_Size = 1 - 3: A	Gender = Male: D	Var_1 = Cat_1: C	null
Work_Experience =	Family_Size = 7 - 9: C	Gender = Female:	Var_1 = Cat_2: D	Var_1 = Cat_6
10 - 14: null	Var_1 = Cat_3	D	Var_1 = Cat_5: null	Graduated = No
Work_Experience =	Gender = Male	Ever_Married = Yes	Age = 50 - 59: D	Ever_Married = No: C
0 - 4 : null	Graduated = No	Gender = Male: D	Age = 10 - 19: D	Ever_Married = Yes: B
Family_Size = 7 - 9: D	Family_Size = 4 - 6: C	Gender = Female	Age = 70 - 79: null	Graduated = Yes: A
Gender = Female	Family_Size = 1 - 3: D	Graduated = No: A	Age = 80 - 89: null	Var_1 = Cat_7: null

Gambar 1. Hasil Pohon Keputusan Menggunakan Metode ID3 (1)

Var_1 = Cat_3: null	Var_1 = Cat_1: null	Age = 80 - 89	Spending_Score =	Age = 80 - 89: null
Var_1 = Cat_1: A	Var_1 = Cat_2: B	Graduated = No	High: null	Family_Size = 7 - 9: null
Var_1 = Cat_2: B	Var_1 = Cat_5: null	Spending_Score = Low	Graduated = Yes: B	Spending_Score = Average
Var_1 = Cat_5: null	Spending_Score = High: B	Gender = Male	Gender = Female: D	Age = 20 - 29: A
Age = 60 - 69	Age = 10 - 19: null	Var_1 = Cat_4: B	Age = 50 - 59	Age = 30 - 39: null
Var_1 = Cat_4: B	Age = 70 - 79	Var_1 = Cat_6: D	Graduated = No: B	Age = 60 - 69: C
Var_1 = Cat_6	Gender = Male	Var_1 = Cat_7: null	Graduated = Yes: A	Age = 40 - 49
Graduated = No: A	Graduated = No: A	Var_1 = Cat_3: null	Age = 10 - 19: null	Family_Size = 4 - 6
Graduated = Yes: B	Graduated = Yes: B	Var_1 = Cat_1: null	Age = 70 - 79: null	Graduated = No: C
Var_1 = Cat_7: D	Gender = Female: C	Var_1 = Cat_2: null	Age = 80 - 89: null	Graduated = Yes: B
Var_1 = Cat_3: A	Age = 80 - 89: null	Var_1 = Cat_5: null	Var_1 = Cat_6	Family_Size = 1 - 3: A
Profession = Lawyer	Profession = Lawyer	Gender = Female: A	Spending_Score = Low	Family_Size = 7 - 9: null
Var_1 = Cat_1: null	Var_1 = Cat_2: B	Spending_Score = Average:	Family_Size = 4 - 6	Age = 50 - 59
Var_1 = Cat_2: null	Age = 20 - 29: A	A	Gender = Male: C	Family_Size = 4 - 6: A
Var_1 = Cat_5: null	Age = 30 - 39: null	Spending_Score = High: D	Gender = Female	Family_Size = 1 - 3
Age = 40 - 49	Age = 60 - 69	Graduated = Yes	Age = 20 - 29: C	Gender = Male: C
Var_1 = Cat_4	Spending_Score = Low	Graduated = No: D	Age = 30 - 39: B	Gender = Female
Ever_Married = No: A	Gender = Male: D	Work_Experience = 0 - 4	Age = 60 - 69: null	Work_Experience =
Ever_Married = Yes: B	Gender = Female: A	Gender = Male: B	Age = 40 - 49: null	0 - 4: C
Var_1 = Cat_6	Spending_Score = Average: C	Gender = Female	Age = 50 - 59: null	Work_Experience =
Spending_Score = Low	Spending_Score = High	Spending_Score = Low:	Age = 10 - 19: null	5 - 9: A
Gender = Male: A	Graduated = No: A	A	Age = 70 - 79: null	Work_Experience =
Gender = Female	Graduated = Yes	Spending_Score =	Age = 80 - 89: null	10 - 14: null
Ever_Married = No: A	Gender = Male: B	Average: null	Family_Size = 1 - 3	Work_Experience =
Ever_Married = Yes: D	Gender = Female: A	Spending_Score =	Age = 20 - 29	0 - 4 : null
Spending_Score = Average	Age = 40 - 49: null	High: A	Gender = Male: A	Family_Size = 7 - 9: null
Graduated = No: D	Age = 50 - 59	Work_Experience = 5 - 9:	Gender = Female: D	Age = 10 - 19: null
Graduated = Yes: C	Gender = Male: A	null	Age = 30 - 39	Age = 70 - 79: C
Spending_Score = High:	Gender = Female: B	Work_Experience = 10 - 14:	Work_Experience = 0 -	Age = 80 - 89: null
null	Age = 10 - 19: null	A	4: D	Spending_Score = High
Var_1 = Cat_7: A	Age = 70 - 79	Work_Experience = 0 - 4 :	Work_Experience = 5 -	Gender = Male
Var_1 = Cat_3: B	Graduated = No	null	9: A	Work_Experience = 0 - 4:
Var_1 = Cat_1: null	Spending_Score = Low: A	Profession = Entertainment	Work_Experience = 10	C
Var_1 = Cat_2: B	Spending_Score = Average:	Var_1 = Cat_4	- 14: A	Work_Experience = 5 - 9:
Var_1 = Cat_5: null	null	Age = 20 - 29: D	Work_Experience = 0 -	B
Age = 50 - 59	Spending_Score = High	Age = 30 - 39: A	4 : null	Work_Experience = 10 -
Spending_Score = Low	Gender = Male: B	Age = 60 - 69: null	Age = 60 - 69	14: null
Gender = Male: A	Gender = Female: D	Age = 40 - 49	Gender = Male: A	Work_Experience = 0 - 4
Gender = Female	Graduated = Yes	Gender = Male	Gender = Female	: null
Graduated = No: A	Gender = Male: D	Ever_Married = No: A	Graduated = No: A	Gender = Female: A
Graduated = Yes: B	Gender = Female	Ever_Married = Yes	Graduated = Yes: D	Var_1 = Cat_7
Spending_Score = Average	Spending_Score = Low: B	Graduated = No	Age = 40 - 49: D	Ever_Married = No: A
Var_1 = Cat_4: B	Spending_Score =	Spending_Score =	Age = 50 - 59: A	Ever_Married = Yes
Var_1 = Cat_6: C	Average: null	Low: A	Age = 10 - 19: null	Age = 20 - 29: D
Var_1 = Cat_7: null	Spending_Score = High:	Spending_Score =	Age = 70 - 79: null	Age = 30 - 39: null
Var_1 = Cat_3: null	B	Average: A		

Gambar 2. Hasil Pohon Keputusan Menggunakan Metode ID3 (2)

Age = 60 - 69: null	Age = 20 - 29: B	Work_Experience = 0 - 4: C	Work_Experience = 0 - 4: A	Work_Experience = 0 - 4: null
Age = 40 - 49: null	Age = 30 - 39: A	Work_Experience = 5 - 9: A	Work_Experience = 5 - 9: A	Work_Experience = 5 - 9: null
Age = 50 - 59: B	Age = 60 - 69: null	Work_Experience = 10 - 14: null	Work_Experience = 10 - 14: null	Work_Experience = 10 - 14: null
Age = 10 - 19: null	Age = 40 - 49	Work_Experience = 0 - 4: null	Work_Experience = 0 - 4: null	Work_Experience = 0 - 4: null
Age = 70 - 79: null	Gender = Male	Family_Size = 1 - 3	Age = 40 - 49	Family_Size = 1 - 3: C
Age = 80 - 89: null	Graduated = No: A	Ever_Married = No	Graduated = No: B	Family_Size = 7 - 9: null
Var_1 = Cat_3	Graduated = Yes: A	Work_Experience = 0 - 4	Graduated = Yes	Work_Experience = 10 - 14: null
Age = 20 - 29: A	Gender = Female: A	Graduated = No: B	Ever_Married = No	Work_Experience = 0 - 4: null
Age = 30 - 39	Age = 50 - 59: A	Graduated = Yes	Gender = Male: C	Gender = Female: C
Spending_Score = Low: D	Age = 10 - 19: null	Gender = Male: B	Gender = Female	Age = 10 - 19: null
Spending_Score = Average	Age = 70 - 79: C	Family_Size = 4 - 6: A	Family_Size = 4	Age = 70 - 79: B
Graduated = No: B	Age = 80 - 89: null	Gender = Male: B	Family_Size = 1	Age = 80 - 89: null
Graduated = Yes: C	Var_1 = Cat_6	Gender = Female: D	Family_Size = 7	Var_1 = Cat_3
Spending_Score = High: null	Age = 20 - 29	Family_Size = 7 - 9:	Family_Size = 1	Age = 20 - 29: null
Age = 60 - 69: B	Work_Experience = 0 - 4	Family_Size = 4 - 6: A	Family_Size = 4	Age = 30 - 39: A
Age = 40 - 49	Family_Size = 1 - 3	Ever_Married = No	Family_Size = 1	Age = 60 - 69: null
Gender = Male: C	Gender = Male: B	Work_Experience = 5 - 9	Family_Size = 7	Age = 40 - 49
Gender = Female: B	Gender = Female: D	Graduated = No: A	Family_Size = 1	Work_Experience = 0 - 4
Age = 50 - 59: null	Family_Size = 7 - 9:	Graduated = Yes	Family_Size = 4	Gender = Male: A
Age = 10 - 19: null	Work_Experience = 5 - 9	Gender = Male: B	Family_Size = 7	Gender = Female
Age = 70 - 79: null	Gender = Male: A	Gender = Male: B	Family_Size = 1	Ever_Married = No: A
Age = 80 - 89: null	Gender = Female	Gender = Female: B	Family_Size = 4	Ever_Married = Yes: B
Var_1 = Cat_1	Ever_Married = No:	Work_Experience = 10 - 14: A	Family_Size = 1	Work_Experience = 5 - 9: B
Age = 20 - 29: null	A	Work_Experience = 0 - 4: null	Family_Size = 7	Work_Experience = 10 - 14: null
Age = 30 - 39: B	Ever_Married = Yes:	Ever_Married = Yes	Family_Size = 1	Work_Experience = 0 - 4: null
Age = 60 - 69: null	A	Gender = Male	Family_Size = 4	Age = 50 - 59: A
Age = 40 - 49: A	Work_Experience = 10 - 14: A	Work_Experience = 0 - 4: C	Family_Size = 1	Age = 10 - 19: null
Age = 50 - 59: null	Work_Experience = 10 - 14: A	Work_Experience = 5 - 9: A	Family_Size = 7	Age = 70 - 79: null
Age = 10 - 19: null	Work_Experience = 0 - 4: D	Work_Experience = 10 - 14: null	Family_Size = 1	Age = 80 - 89: null
Age = 70 - 79: null	Work_Experience = 0 - 4: null	Work_Experience = 0 - 4: null	Family_Size = 7	Var_1 = Cat_1: A
Age = 80 - 89: null	Age = 30 - 39	Work_Experience = 0 - 4: null	Family_Size = 4	Var_1 = Cat_2
Var_1 = Cat_2	Family_Size = 4 - 6	Work_Experience = 5 - 9: C	Family_Size = 1	Age = 60 - 69
Age = 20 - 29: null	Ever_Married = No	Work_Experience = 10 - 14: null	Family_Size = 7	Gender = Male
Age = 30 - 39: null	Work_Experience = 0 - 4: D	Work_Experience = 0 - 4: null	Family_Size = 1	Graduated = No: B
Age = 60 - 69: null	Work_Experience = 5 - 9: C	Work_Experience = 0 - 4: null	Family_Size = 7	Graduated = Yes: B
Age = 40 - 49: B	Work_Experience = 10 - 14: A	Gender = Female: C	Family_Size = 1	Gender = Female: C
Age = 50 - 59: A	Work_Experience = 0 - 4: null	Family_Size = 7 - 9: C	Family_Size = 4	Age = 40 - 49
Age = 10 - 19: null	Ever_Married = Yes	Age = 60 - 69	Family_Size = 1	Family_Size = 4 - 6: B
Age = 70 - 79: D	Work_Experience = 10 - 14: null	Ever_Married = No: B	Family_Size = 7	Family_Size = 1 - 3: A
Age = 80 - 89: null	Work_Experience = 0 - 4: null	Ever_Married = Yes	Family_Size = 1	Family_Size = 7 - 9: null
Var_1 = Cat_5: C	Profession = Artist	Gender = Male	Family_Size = 4	Age = 50 - 59
Spending_Score = Low	Spending_Score = Low	Gender = Male	Family_Size = 7	Gender = Male: D
Var_1 = Cat_4	Var_1 = Cat_4	Gender = Male	Family_Size = 1	Gender = Female: C

Gambar 3. Hasil Pohon Keputusan Menggunakan Metode ID3 (3)

Work_Experience = 0 - 4: C	Family_Size = 4 - 6: C	Var_1 = Cat_7	Age = 60 - 69	Family_Size = 4 - 6: D
Work_Experience = 5 - 9: A	Family_Size = 1 - 3: A	Gender = Male: B	Gender = Male	Family_Size = 1 - 3: A
Work_Experience = 10 - 14: null	Family_Size = 7 - 9: null	Gender = Female: C	Graduated = No: B	Family_Size = 7 - 9: null
Work_Experience = 0 - 4: null	Gender = Female: C	Var_1 = Cat_3: C	Graduated = Yes: B	Graduated = Yes
Gender = Female: A	Var_1 = Cat_7: null	Var_1 = Cat_1: null	Gender = Female: C	Family_Size = 4 - 6
Var_1 = Cat_5	Var_1 = Cat_3: A	Var_1 = Cat_2: B	Age = 40 - 49	Gender = Male: A
Age = 20 - 29: D	Var_1 = Cat_1: null	Age = 50 - 59	Family_Size = 4 - 6: B	Gender = Female: B
Age = 30 - 39: B	Var_1 = Cat_2: B	Graduated = No: C	Family_Size = 1 - 3: A	Family_Size = 1 - 3
Age = 60 - 69: null	Var_1 = Cat_5: null	Graduated = Yes: A	Family_Size = 7 - 9: null	Work_Experience = 0 - 4
Age = 40 - 49: B	Age = 40 - 49	Var_1 = Cat_6	Age = 50 - 59	Spending_Score = Low
Age = 50 - 59: null	Var_1 = Cat_4	Gender = Male	Gender = Male: D	Ever_Married = No
Age = 10 - 19: null	Gender = Male: B	Graduated = No: C	Gender = Female: C	Gender = Male: B
Age = 70 - 79: null	Gender = Female: A	Graduated = Yes	Age = 10 - 19: null	Gender = Male: C
Age = 80 - 89: null	Var_1 = Cat_6	Graduated = No: C	Age = 70 - 79: C	Gender = Male: D
Spending_Score = Average	Graduated = No	Family_Size = 4 - 6: C	Age = 80 - 89: A	Profession = Doctor
Age = 20 - 29	Gender = Male: A	Family_Size = 1 - 3	Age = 20 - 29	Gender =
Var_1 = Cat_4: B	Gender = Female: B	Work_Experience = 0 - 4: C	Var_1 = Cat_4: B	Var_1 = Cat_4: B
Var_1 = Cat_6	Graduated = Yes	Work_Experience = 5 - 9: C	Var_1 = Cat_6	Var_1 = Cat_6
Graduated = No: D	Gender = Male	Work_Experience = 10 - 14: null	Family_Size = 4 - 6	Ever_Married = Yes: D
Graduated = Yes: B	Gender = Female	Work_Experience = 0 - 4: null	Gender = Male: C	Spending_Score = Average: A
Var_1 = Cat_7: null	Family_Size = 4 - 6	Family_Size = 7 - 9: null	Gender = Female: D	Spending_Score = High: null
Var_1 = Cat_3: C	Work_Experience = 5 - 9: null	Work_Experience = 10 - 14: C	Work_Experience = 0 - 4	Work_Experience = 5 - 9: null
Var_1 = Cat_1: null	Work_Experience = 10 - 14: C	Gender = Female: C	Graduated = No: D	Work_Experience = 10 - 14: A
Var_1 = Cat_2: null	Work_Experience = 0 - 4: null	Var_1 = Cat_7: C	Graduated = Yes: D	Work_Experience = 0 - 4: null
Var_1 = Cat_5: null	Family_Size = 1 - 3: C	Var_1 = Cat_3: null	Work_Experience = 5 - 9: C	Work_Experience = 10 - 14: A
Age = 30 - 39	Family_Size = 7 - 9:	Var_1 = Cat_1: C	Work_Experience = 0 - 4	Family_Size = 7 - 9: null
Graduated = No: B	Gender = Female	Var_1 = Cat_2: C	Family_Size = 4 - 6	Var_1 = Cat_7: null
Graduated = Yes	Family_Size = 4 - 6: C	Var_1 = Cat_5: null	Gender = Male: C	Var_1 = Cat_3
Work_Experience = 0 - 4	Family_Size = 1 - 3	Age = 10 - 19: null	Gender = Female: D	Graduated = No: C
Gender = Male: C	Work_Experience = 0 - 4: C	Age = 70 - 79: C	Work_Experience = 0 - 4	Graduated = Yes: B
Gender = Female: B	Work_Experience = 5 - 9: B	Age = 80 - 89: null	Work_Experience = 5 - 9: C	Var_1 = Cat_1: null
Work_Experience = 5 - 9: C	Work_Experience = 10 - 14: B	Spending_Score = High	Work_Experience = 10 - 14: A	Var_1 = Cat_2: B
Work_Experience = 10 - 14: B	Work_Experience = 0 - 4: null	Age = 20 - 29: null	Work_Experience = 0 - 4	Var_1 = Cat_5: null
Work_Experience = 0 - 4: null	Work_Experience = 10 - 14: null	Age = 30 - 39	Work_Experience = 5 - 9: C	Age = 60 - 69
Age = 60 - 69	Work_Experience = 0 - 4: null	Work_Experience = 0 - 4: A	Family_Size = 4 - 6: C	Gender = Male
Var_1 = Cat_4: null	Family_Size = 7 - 9:	Work_Experience = 5 - 9: B	Family_Size = 1 - 3: D	Family_Size = 4 - 6
Var_1 = Cat_6	Work_Experience = 0 - 4: null	Work_Experience = 10 - 14: null	Family_Size = 7 - 9: null	Var_1 = Cat_4: null
Gender = Male	Family_Size = 4 - 6: A	Work_Experience = 0 - 4: null	Var_1 = Cat_6	Var_1 = Cat_6: B
	Family_Size = 7 - 9:	Work_Experience = 0 - 4: null	Graduated = No	Var_1 = Cat_7: null
				Var_1 = Cat_3: C

Gambar 4. Hasil Pohon Keputusan Menggunakan Metode ID3 (4)

Var_1 = Cat_1: null	Age = 60 - 69: null	Graduated = Yes: A	Var_1 = Cat_7: null	Var_1 = Cat_7: null
Var_1 = Cat_2: null	Age = 40 - 49: null	Var_1 = Cat_1: null	Var_1 = Cat_3: B	Var_1 = Cat_3: B
Var_1 = Cat_5: null	Age = 50 - 59: null	Var_1 = Cat_2: null	Var_1 = Cat_1: B	Var_1 = Cat_2: A
Family_Size = 1 - 3: C	Age = 10 - 19: null	Var_1 = Cat_5	Var_1 = Cat_2: A	Var_1 = Cat_2: A
Family_Size = 7 - 9: null	Age = 70 - 79: null	Gender = Male: D	Var_1 = Cat_5: null	Var_1 = Cat_5: null
Gender = Female: B	Age = 80 - 89: null	Gender = Female: B	Age = 60 - 69	Age = 10 - 19: A
Age = 40 - 49	Spending_Score = Average: D	Age = 30 - 39	Graduated = No	Age = 70 - 79
Spending_Score = Low	Spending_Score = High: B	Graduated = No: D	Family_Size = 4 - 6: null	Family_Size = 4 - 6: null
Graduated = No: A	Var_1 = Cat_7: null	Graduated = Yes: A	Family_Size = 1 - 3: C	Family_Size = 1 - 3: C
Graduated = Yes	Var_1 = Cat_3: D	Age = 60 - 69: D	Family_Size = 7 - 9: B	Family_Size = 7 - 9: D
Gender = Male: C	Var_1 = Cat_1: null	Age = 40 - 49	Graduated = Yes: A	Age = 80 - 89
Gender = Female	Var_1 = Cat_2: B	Spending_Score = Low: D	Age = 40 - 49	Spending_Score = Low: A
Ever_Married = No: A	Var_1 = Cat_5: null	Spending_Score = Average: null	Family_Size = 4 - 6	Spending_Score = Average: null
Ever_Married = Yes: C	Graduated = Yes	Spending_Score = High	Spending_Score = Low: null	Spending_Score = Average: null
Spending_Score = Average	Work_Experience = 0 - 4	Graduated = No: A	Spending_Score = Average: C	Spending_Score = High
Family_Size = 4 - 6	Gender = Male: D	Graduated = Yes: B	Spending_Score = High	Graduated = No: B
Gender = Male: B	Gender = Female: A	Age = 50 - 59	Graduated = No: A	Graduated = Yes: C
Gender = Female: C	Work_Experience = 5 - 9: C	Gender = Male: A	Graduated = Yes: B	
Family_Size = 1 - 3	Work_Experience = 10 - 14: A	Gender = Female: B	Family_Size = 1 - 3	
Graduated = No: D	Work_Experience = 0 - 4 : null	Age = 10 - 19	Spending_Score = Low: D	
Graduated = Yes: A	Profession = Marketing	Ever_Married = No: A	Spending_Score = Average: A	
Family_Size = 7 - 9: null	Age = 20 - 29	Ever_Married = Yes: C		
Spending_Score = High: B	Var_1 = Cat_4: D	Age = 70 - 79	Spending_Score = High	
Age = 50 - 59	Var_1 = Cat_6	Ever_Married = No: D	Var_1 = Cat_4: null	
Var_1 = Cat_4: null	Ever_Married = No	Ever_Married = Yes: A	Var_1 = Cat_6	
Var_1 = Cat_6: C	Family_Size = 4 - 6	Age = 80 - 89: null	Graduated = No: C	
Var_1 = Cat_7: A	Gender = Male	Profession = Executive	Graduated = Yes: C	
Var_1 = Cat_3: B	Work_Experience = 0 - 4: D	Age = 20 - 29	Var_1 = Cat_7: null	
Var_1 = Cat_1: C	Work_Experience = 5 - 9: C	Work_Experience = 0 - 4	Var_1 = Cat_3: null	
Var_1 = Cat_2: null	Work_Experience = 10 - 14: null	Graduated = No: D	Var_1 = Cat_1: null	
Var_1 = Cat_5: null	Work_Experience = 0 - 4: null	Graduated = Yes: A	Var_1 = Cat_2: D	
Age = 10 - 19: D	Gender = Female: D	Work_Experience = 5 - 9: D	Var_1 = Cat_5: null	
Age = 70 - 79: B	Family_Size = 1 - 3: C	Work_Experience = 10 - 14: A	Family_Size = 7 - 9: null	
Age = 80 - 89: null	Family_Size = 7 - 9: null	Work_Experience = 0 - 4 : null	Age = 50 - 59	
Profession = Homemaker	Ever_Married = Yes: A	Age = 30 - 39	Var_1 = Cat_4: C	
Graduated = No	Var_1 = Cat_7: C	Var_1 = Cat_4: D	Var_1 = Cat_6	
Var_1 = Cat_4: B	Var_1 = Cat_3	Var_1 = Cat_6	Gender = Male	
Var_1 = Cat_6	Graduated = No	Graduated = No: A	Graduated = No	
Spending_Score = Low	Ever_Married = Yes: A	Graduated = Yes	Family_Size = 4 - 6: C	
Age = 20 - 29: D	Var_1 = Cat_7: C	Spending_Score = Low: D	Family_Size = 1 - 3: A	
Age = 30 - 39	Var_1 = Cat_3	Spending_Score = Average: C	Family_Size = 7 - 9:	
Family_Size = 4 - 6: D	Graduated = No	null	null	
Family_Size = 1 - 3: B	Gender = Male: D	Spending_Score = High: B	Graduated = Yes: C	
Family_Size = 7 - 9:	Gender = Female: A	B	Gender = Female: A	
null				

Gambar 5. Hasil Pohon Keputusan Menggunakan Metode ID3 (5)

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian mengenai prediksi segmentasi pelanggan dapat disimpulkan bahwa Algoritma ID3 dinilai lebih cocok digunakan untuk memprediksi segmentasi pelanggan penjualan mobil. Hal ini dikarenakan ID3 mempunyai tingkat akurasi yang tinggi, yaitu 96,4286%. Sementara itu, algoritma Naïve Bayes dinilai masih kurang cocok digunakan dalam memprediksi segmentasi pelanggan penjualan mobil dikarenakan memiliki tingkat akurasi yang kurang baik, yaitu sebesar 56,4286%. Hasil penelitian ini dapat membantu manajemen perusahaan untuk memprediksi segmentasi pelanggan penjualan mobil, sehingga dapat mengambil kebijakan promosi penjualan yang sesuai dan tepat sasaran.

DAFTAR PUSTAKA

[1] M. Arifin, "Business Intelligence untuk Customer Churn Telekomunikasi," in *Prosiding SNATIF ke-1*, 2014, no. 1, pp. 279–286.
 [2] N. W. Wardani, G. R. Dantes, and G. Indrawan, "Prediksi Customer Churn dengan Algoritma Decision Tree C4.5," *J. Resist.*, vol. 1, no. 1, pp. 16–24, 2018.
 [3] S. Sudriyanto, "Clustering Loyalitas Pelanggan dengan Metode RFM (Recency, Frequency, Monetary) dan Fuzzy C-Means," in *Prosiding SNATIF ke-4*, 2017, pp. 815–822.
 [4] H. A. Indriyani, "Algoritma Decision Tree (ID3)

dan K-Nearest Neighbour (K-NN) dalam Prediksi Customer Churn Berdasarkan Segmentasi Pelanggan (Studi Kasus Radity Craft Tasikmalaya)," Universitas Siliwangi, 2019.

[5] I. Oktania and A. A. Supianto, "Perbandingan Teknik Klasifikasi dalam Data Mining untuk Bank Direct Marketing," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 567, 2018, doi: 10.25126/jtiik.201855958.
 [6] J. R. Otukei and T. Blaschke, "Land Cover Change Assessment Using Decision Trees, Support Vector Machines and Maximum Likelihood Classification Algorithms," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 12, no. SUPPL. 1, pp. 27–31, 2010, doi: 10.1016/j.jag.2009.11.002.
 [7] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques Third Edition*, 3rd ed. Waltham.; Elsevier, 2012.
 [8] N. W. Wardani et al., "Prediksi Pelanggan Loyal Menggunakan Metode Naïve Bayes Berdasarkan Segmentasi Pelanggan dengan Pemodelan RFM," *J. Manaj. dan Teknol. Inf.*, vol. 12, no. 2, pp. 113–124, 2022.
 [9] N. W. Wardani and N. K. Ariasih, "Analisa Komparasi Algoritma Decision Tree C4.5 dan Naïve Bayes untuk Prediksi Churn Berdasarkan Kelas Pelanggan Retail," *Int. J. Nat. Sci. Eng.*, vol. 3, no. 3, p. 103, 2019, doi: 10.23887/ijnse.v3i3.23113.

- [10] Maghfirah, T. B. Adji, and N. A. Setiawan, "Menggunakan Data Mining untuk Segmentasi Customer pada Bank untuk Meningkatkan Customer Relationship Management (CRM) dengan Metode Klasifikasi (Algoritma J-48, Zero-R dan Naive Bayes)," in *Prosiding Seminar Nasional Sains dan Teknologi*, 2015, vol. 1, no. 1, pp. 65–70.
- [11] N. Nurdiana and A. Algifari, "Studi Komparasi Algoritma ID3 dan Algoritma Naive Bayes untuk Klasifikasi Penyakit Diabetes Mellitus," *Infotech J.*, vol. 6, no. 2, pp. 18–23, 2020.
- [12] D. A. Munawwaroh and A. H. Primandari, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Lila Ibu Hamil Berpotensi Gizi Kurang," *J. Ilm. Pendidik. Mat.*, vol. 10, no. 2, pp. 367–380, 2022.
- [13] A. Muzakir and R. A. Wulandari, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016, doi: 10.15294/sji.v3i1.4610.
- [14] T. R. Patil and S. . Sherekar, "Performance Analysis of Naive Bayes dan J48 Classification Algorithm for Data Classification," *Int. J. Comput. Sci. Appl.*, vol. 6, no. 2, pp. 256–261, 2013.
- [15] M. F. Rifai, H. Jatnika, and B. Valentino, "Penerapan Algoritma Naive Bayes pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)," *Petir J. Pengkaj. dan Penerapan Tek. Inform.*, vol. 12, no. 2, pp. 131–144, 2019, doi: 10.33322/petir.v12i2.471.
- [16] H. F. Putro, R. T. Vulandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes untuk Klasifikasi Pelanggan," *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500.
- [17] A. Ghassa, A. Wahyudi, S. O. Tampubolon, N. A. Putri, E. Rasywir, and D. Kisbianty, "Penerapan Data Mining Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap INDIHOME," *J. Inform. dan Rekayasa Komput.*, vol. 2, no. 2, pp. 240–247, 2022.
- [18] N. Ramadhani and N. Fajarianto, "Sistem Informasi Evaluasi Perkuliahan dengan Sentimen Analisis Menggunakan Naive Bayes dan Smoothing Laplace," *JSINBIS (Jurnal Sist. Inf. Bisnis)*, vol. 10, no. 2, pp. 228–234, 2020.
- [19] D. Indrajaya, "Sistem Pendukung Keputusan Perizinan Santri Menggunakan Metode Naive Bayes Classifier dengan Laplace Correction," *J. Simki-Techsain*, vol. 2, no. 12, pp. 1–6, 2018.
- [20] I. Listiowarni, "Implementasi Naive Bayesian dengan Laplacian Smoothing untuk Peminatan dan Lintas Minat Siswa SMAN 5 Pamekasan," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 8, no. 2, pp. 124–129, 2019, doi: 10.32736/sisfokom.v8i2.652.
- [21] C. J. Mantas and J. Abellán, "Credal-C4.5: Decision Tree Based on Imprecise Probabilities to Classify Noisy Data," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4625–4637, 2014, doi: 10.1016/j.eswa.2014.01.017.
- [22] N. Nafi'iyah and C. Fatichah, "Metode Fuzzy ID3 untuk Klasifikasi Bentuk Wajah Manusia Menggunakan Dental Panoramic," *J. SPIRIT*, vol. 10, no. 1, pp. 13–20, 2018.
- [23] A. Sifaunajah and R. D. Wahyuningtyas, "Penggunaan Algoritma ID3 untuk Klasifikasi Data Calon Peserta Didik," *CSRID J.*, vol. 14, no. 2, pp. 103–112, 2022.
- [24] H. Hikmatulloh, S. Susilawati, and S. Nurajizah, "Peningkatan Akurasi pada Algoritma ID3 Menggunakan Operator Bagging dalam Mendiagnosa Kesehatan Kehamilan," *Indones. J. Comput. Inf. Technol.*, vol. 6, no. 2, pp. 92–96, 2021.