

Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5

Sanni Ucha Putri¹, Eka Irawan², Fitri Rizky³

^{1,2,3}STIKOM Tunas Bangsa, Pematangsiantar – Indonesia

Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara

¹sanniuchaputri6@gmail.com, ²ekairawanstikomtunasbangsa.ac.id,

³fitri_rizki@sitkomb.ac.id

Abstract

Diabetes is a worldwide health problem with an estimated 120 million sufferers. This figure will increase if there is ignorance of the general public about the factors that can trigger diabetes. In this study, the aim of this research is to make a prediction model using Data Mining Algorithm C4.5 which produces a decision tree and tests carried out using Rapidminer so that diabetes prevention can be done as soon as possible. In this study, there are several classification attributes, namely body weight, age, blood pressure, pulse and blood sugar levels. The results of this study will be used as a reference to be able to see whether a person is at risk of diabetes or not based on predetermined attributes.

Keywords : Diabetes, Data Mining, C.45 Algorithm

Abstrak

Penyakit diabetes merupakan masalah kesehatan di seluruh dunia dengan perkiraan dari 120 juta penderita. Angka tersebut akan semakin meningkat jika adanya ketidaktahuan masyarakat umum tentang faktor-faktor yang dapat memicu terkena penyakit diabetes. Pada penelitian ini bertujuan untuk membuat model prediksi menggunakan Data Mining Algoritma C4.5 yang menghasilkan sebuah pohon keputusan serta pengujian yang dilakukan dengan menggunakan Rapidminer agar pencegahan terhadap penyakit diabetes dapat dilakukan segera mungkin. Dalam penelitian ini terdapat beberapa atribut klasifikasi yaitu berat badan, usia, tekanan darah, denyut nadi dan kadar gula darah. Hasil dari penelitian ini akan dijadikan sebagai acuan untuk dapat melihat apakah seseorang beresiko terkena penyakit diabetes atau tidak berdasarkan atribut yang telah ditetapkan.

Kata Kunci : Penyakit Diabetes, Data Mining, Algoritma C4.5

1. Pendahuluan

Data mining merupakan suatu rangkaian atau tindakan untuk menemukan hubungan yang memiliki arti melalui pola dan kecenderungan dalam himpunan besar data yang tersimpan dengan menggunakan metode. “Teknik, metode atau algoritma di dalam data mining sangat beragam. Data mining dikenal dengan nama Knowledge Discovery in Database (KDD) dalam jurnal ilmiah” [1] sehingga hasil dari data mining dapat digunakan untuk memperbaiki pengambilan keputusan dimasa yang akan datang. Adapun proses yang umumnya dilakukan oleh data mining meliputi deskripsi, prediksi, estimasi, klasifikasi, *clushtering* dan asosiasi. Dalam data mining prediksi dan klasifikasi banyak digunakan untuk menganalisis suatu data yang dapat menggambarkan kelas data atau untuk memprediksi data di masa depan.

Algoritma C4.5 merupakan suatu Algoritma yang dapat digunakan untuk melakukan klasifikasi atau pengelompokkan pada dataset. Algoritma C4.5 memiliki ide dasar yaitu pembentukan pohon keputusan (*Decision Tree*). Dimana Pohon

keputusan termasuk metode klasifikasi dan prediksi yang terkenal karena sangat mudah dipahami dengan bahasa alami sehingga dapat diinterpretasikan dengan cepat. "Pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki *gain* tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai sumbu atribut klasifikasi" [2]. Sehingga pohon keputusan dapat membantu memecahkan masalah dalam pengambilan keputusan mengenai alternatif-alternatif yang ada pada masalah tersebut.

Keterlambatan diagnosis penyakit diabetes menjadi salah satu penyebab meningkatnya jumlah penderita diabetes. Pasien yang sudah meninggal akibat komplikasi sebelum adanya penegakkan diagnosa. Salah satu penyebab keterlambatan penegakkan diagnosis adalah dengan banyaknya faktor-faktor yang berpengaruh terhadap pilihan yang ada atau beragamnya variabel-variabel. Diabetes menjadi penyakit kronis yang berbahaya dengan ditandai tingginya kadar gula (glukosa) darah atau diatas nilai normal (rata-rata). Jika penyakit ini tidak dapat dikontrol dan ditangani dengan baik, akan menimbulkan berbagai penyakit komplikasi yang membahayakan nyawa penderita. Peningkatan penyakit diabetes setiap tahun selalu bertambah. Menurut *International Diabetes Federation (IDF)* Jumlah penderita penyakit diabetes disetiap negara meningkat dan usia terbanyak orang dengan penyakit diabetes berada di usia antara 49-59 tahun. Pada umumnya penderita penyakit diabetes mengakibatkan penyakit *cardio vascular* dua sampai empat kali lebih banyak dibandingkan dengan yang bukan penderita penyakit diabetes. Dampak yang diakibatkan penyakit diabetes tersebut terjadi karena kerusakan *vacular radio* serta akibat dari penyakit *retinopati dan neuropati*, oleh karena itu maka dilakukan prediksi lebih awal dengan melihat atribut-atribut pendukung.

Dalam sebuah penelitian sudah banyak percobaan yang telah dilakukan dengan memanfaatkan teknik *data mining* dalam memprediksi penyakit menggunakan berbagai algoritma seperti *Naive Bayes*, *Decision Tree*, *SVM*, *J48* dan lain-lain. Diantaranya telah dapat hasil penelitian [3] dengan menggunakan penerapan algoritma *C4.5* dalam memprediksi gangguan autisme pada anak dapat disimpulkan bahwa metode klasifikasi *data mining* algoritma *C4.5* menghasilkan akurasi sebanyak 72% dengan melakukan pengujian metode *cross validation* pada aplikasi *rapidminer*. Adapun penulis [4] menerapkan *data mining* untuk klasifikasi prediksi penyakit ISPA (*Infeksi Saluran Pernapasan Akut*) dengan Algoritma *Decision Tree (ID3)* yang menghasilkan atribut penyakit ISPA yang paling berpengaruh pertama adalah gizi sebesar 240.9 kemudian atribut usia sebesar 10.7, maka atribut yang tidak berpengaruh seseorang terkena penyakit ISPA adalah jenis kelamin.

Penelitian ini melakukan prediksi secara diagnostik apakah pasien menderita penyakit diabetes dengan menerapkan algoritma *C4.5* dan melihat beberapa faktor-faktor pendukung. penulis menggunakan teknik *data mining* klasifikasi untuk memprediksi seseorang mengidap penyakit diabetes atau tidak. Hasil dari penelitian ini nantinya diharapkan dapat digunakan oleh pihak penyedia layanan kesehatan dalam menentukan status diabetes seseorang.

2. Metodologi Penelitian

2.1. Data Mining

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Data mining mulai ada sejak 1990-an sebagai cara yang benar dan tepat untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antara data untuk melakukan pengelompokkan ke dalam satu atau lebih cluster sehingga objek - objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu

dengan lainnya. Data mining merupakan bagian dari proses penemuan pengetahuan dari basis data Knowledge Discovery in [5].

2.2. Pohon Keputusan (*Decision Tree*)

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Dimana metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang dapat mempresentasikan aturan-aturan. Aturan tersebut dapat dengan mudah dipahami ataupun dijelaskan dengan ulang karena memiliki bahasa alami yang dapat dimengerti. Pohon keputusan juga berguna untuk menjelajahi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. “Pohon keputusan memiliki fleksibilitas (mudah digunakan) yang membuatnya sangat menarik, terutama karena menyajikan keuntungan dari proses pengubahan konsep menjadi gambar yang telah disajikan sangat mudah dimengerti dimana cabang-cabang dari pohon merangkum klasifikasi” [6].

2.3. Algoritma C4.5

Algoritma C4.5 merupakan metode yang menjadi pilihan pertama dan sering digunakan dalam pengembangan Data Mining karena kecepatannya dalam proses klasifikasi atau segmentasi dan bersifat prediktif. Algoritma C4.5 salah satu metode untuk membuat *Decision Tree* berdasarkan *training* data yang telah disediakan. “Kelebihan algoritma C4.5 adalah metode pohon keputusan yang dapat menghindari munculnya permasalahan dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi keputusan yang dihasilkan” [7].

3. Hasil dan Pembahasan

Dalam penelitian, data sampel yang digunakan adalah 49 data pasien penyakit diabetes dari RSUD. Dr. Djasamen Saragih Pematangsiantar. Setelah data yang diperlukan diperoleh, kemudian ditentukan atribut-atribut yang diperlukan pada penelitian ini adalah pemeriksaan awal pasien, antara lain : Usia Penyakit Diabetes, Tekanan Darah (Nm/Hg) Penyakit Diabetes, Denyut Nadi (xx/Menit) Penyakit Diabetes, Berat Badan (Kg) Penyakit Diabetes, dan Kadar Gula Darah (mg/dl) Penyakit Diabetes. Sedangkan kondisi pasien yang diprediksi yaitu terkena penyakit diabetes atau tidak merupakan variabel *output*. Berikut ini adalah data-data pasien yang sudah diklasifikasikan berdasarkan atribut diatas yang akan diolah dengan menggunakan perhitungan manual C4.5.

Tabel 1. Klasifikasi Penyakit Diabetes

No	Usia (tahun)	Tekanan Darah (NmHg)	Denyut Nadi (x/menit)	Berat Badan (kg)	kadar gula darah (mg/dl)	Variabel
1	Old	Rendah	Tinggi	Normal	Tinggi	Positif
2	Old	Tinggi	Tinggi	Normal	Tinggi	Positif
3	Old	Tinggi	Tinggi	Normal	Tinggi	Positif
4	Old	Tinggi	Tinggi	Tinggi	Tinggi	Positif
5	Old	Tinggi	Tinggi	Tinggi	Tinggi	Positif
6	Old	Tinggi	Tinggi	Tinggi	Tinggi	Positif
7	Old	Tinggi	Tinggi	Normal	Tinggi	Positif
8	Old	Tinggi	Normal	Normal	Tinggi	Positif
9	Old	Tinggi	Tinggi	Tinggi	Tinggi	Positif
10	Old	Tinggi	Tinggi	Tinggi	Tinggi	Positif
11	Old	Tinggi	Tinggi	Tinggi	Tinggi	Positif
12	Old	Rendah	Tinggi	Tinggi	Rendah	Negatif

No	Usia (tahun)	Tekanan Darah (NmHg)	Denyut Nadi (x/menit)	Berat Badan (kg)	kadar gula darah (mg/dl)	Variabel
13	Old	Tinggi	Tinggi	Normal	Rendah	Negatif
...
48	Medium	Tinggi	Tinggi	Normal	Rendah	Negatif
49	Old	Tinggi	Tinggi	Normal	Tinggi	Positif

Data yang sudah diklasifikasikan kemudian diolah menggunakan *RapidMiner* menggunakan *Perfomance* yang berfungsi sebagai validasi dan reabilitas data untuk mencari keakuratan data. Data yang akurat maka akan dilakukan pengolahan data untuk mencari hasil dari penyakit diabetes dengan menggunakan *RapidMiner* dan mengambil keputusan dari hasil pengolahan data menggunakan *RapidMiner* yang dilakukan.

Dalam teknik algoritma C4.5 memilih atribut sebagai *node* (akar) dengan mencari terlebih dahulu jumlah kasus keseluruhan, jumlah kasus untuk keputusan positif diabetes dan jumlah kasus untuk keputusan negatif diabetes serta *entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut usia, tekanan darah, denyut nadi, berat badan dan kadar gula darah. Setelah itu dilakukan perhitungan *Gain* untuk masing-masing atribut. Hasil perhitungan ditunjukkan oleh tabel berikut ini.

Tabel 2. Perhitungan *Node 1*

<i>Node</i>		(S)	(S1)	(S2)	<i>Entropy</i>	<i>Information Gain</i>
1	Total	49	33	16	0,911342375	
	Usia					0,022857303
	<i>Young</i>	0	0	0	0	
	<i>Medium</i>	3	1	2	0,918295834	
	<i>Old</i>	46	32	14	0,886540892	
	Tekanan Darah					0,063592748
	Rendah	6	2	4	0,918295834	
	Normal	4	2	2	1	
	Tinggi	39	29	10	0,821280941	
	Denyut Nadi					0,023881347
	Rendah	0	0	0	0	
	Normal	2	2	0	0	
	Tinggi	47	31	16	0,925225327	
	Berat Badan					0,136251472
	<i>Under Weight</i>	2	0	2	0	
	<i>Avarage</i>	26	15	11	0,982858689	
	<i>Over weight</i>	21	18	3	0,591672778	
	Kadar Gula Darah					0,406244936
	Rendah	7	0	7	0	
	Normal	5	1	4	0,721928095	
	Tinggi	37	32	5	0,571354974	

Kolom *entropy* pada baris total dihitung dengan persamaan (1) sebagai berikut :

$$Entropy (total) = ((-33/49)*LOG_2(33/49)+(-16/49)*LOG_2(16/49))$$

$Entropy (total) = 0,911342375$. Kemudian pada nilai *Gain* pada baris usia dihitung dengan persamaan (2) sebagai berikut:

$$Gain (total, usia) = (0,911342375)-((0/49)*0)-((3/49)*0,918295834)-((46/49)*0,886540892)$$

$$Gain (total, usia) = 0,022857303$$

Dari hasil perhitungan pada tabel 2 diperoleh atribut yang menjadi *node* (akar) adalah kadar gula darah dimana terdiri dari tiga sub atribut yaitu rendah, normal dan tinggi. Berdasarkan nilai *entropy* dari ketiga sub atribut diatas, hanya sub atribut rendah yang memperoleh satu keputusan. Maka perlu dilakukan perhitungan lebih lanjut. Untuk hasil perhitungan selanjutnya Algoritma C4.5 dapat ditunjukkan pada tabel berikut ini.

Tabel 3. Perhitungan *Node* 1.1

<i>Node</i>	Total	(S)	(S1)	(S2)	<i>Entropy</i>	<i>Information Gain</i>
1.1	Kadar Gula Darah – NORMAL	5	1	4	0,721928095	
	Usia					0
	<i>Old</i>	5	1	4	0,721928095	
	<i>Medium</i>	0	0	0	0	
	<i>Young</i>	0	0	0	0	
	Tekanan Darah					0,721928095
	Tinggi	4	0	4	0	
	Normal	0	0	0	0	
	Rendah	1	1	0	0	
	Denyut Nadi					0
	Tinggi	5	1	4	0,72192805	
	Normal	0	0	0	0	
	Rendah	0	0	0	0	
	Berat Badan					0,321928095
	<i>Over Weight</i>	2	0	2	0	
	<i>Avarage</i>	2	1	1	1	
	<i>Under weight</i>	1	0	1	0	

Dari hasil perhitungan diatas pada tabel 3, atribut yang menjadi *Node* cabang dari atribut kadar gula darah normal adalah tekanan darah. Dimana sub atribut dari tekanan darah yang terdiri dari tekanan darah tinggi, tekanan darah normal dan tekanan darah rendah sudah mengklasifikasikan kasusnya masing-masing. Tekanan darah tinggi menghasilkan keputusan negatif sedangkan tekanan darah rendah menghasilkan keputusan positif. Namun pada tekanan darah normal tidak memiliki jumlah kasus sehingga tidak perlu dihitung. Untuk hasil perhitungan selanjutnya pada *Node* cabang dari atribut kadar gula darah Tinggi dapat ditunjukkan pada tabel 4.

Tabel 4. Perhitungan *Node* 1.2

<i>Node</i>	Total	(S)	(S1)	(S2)	<i>Entropy</i>	<i>Information Gain</i>
1.2	Kadar Gula Darah- TINGGI	37	32	5	0,5713549	
	Usia					0,005744867
	<i>Old</i>	36	31	5	0,5813214	
	<i>Medium</i>	1	1	0	0	
	<i>Young</i>	0	0	0	0	
	Tekanan Darah					0,120040467
	Tinggi	31	29	2	0,3451173	
	Normal	4	2	2	1	
	Rendah	2	1	1	1	
	Denyut Nadi					0,011664508
	Tinggi	35	30	5	0,5916727	
	Normal	2	2	0	0	
	Rendah	0	0	0	0	

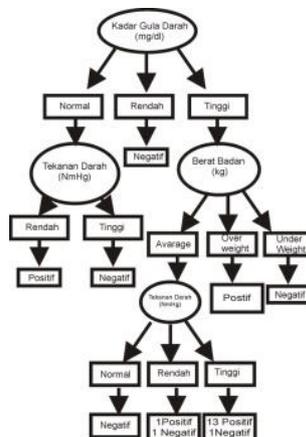
Node	Total	(S)	(S1)	(S2)	Entropy	Information Gain
1.2	Kadar Gula Darah - TINGGI	37	32	5	0,5713549	
	Berat Badan					0,199579809
	Over Weight	18	18	0	0	
	Avarage	18	14	4	0,7642045	
	under weight	1	0	1	0	

Dari hasil perhitungan diatas pada tabel 4, atribut yang menjadi *Node* cabang dari atribut kadar gula darah tinggi adalah berat badan. Dimana sub atribut dari berat badan yang terdiri dari *Over Weight*, *Avarage* dan *Under Weight*. *Avarage* belum menghasilkan keputusan klasifikasi sehingga harus dihitung kembali. Sedangkan *under weight* dan *Over Weight* sudah menghasilkan satu keputusan. Pada perhitungan selanjutnya, sisa dari *node* cabang berat badan adalah *Avarage*. Berikut ini adalah hasil perhitungan untuk menghasilkan klasifikasi berat badan avarage.

Tabel 5. Perhitungan Node 1.2.1

Node	Total	(S)	(S1)	(Si)	Entropy	Information Gain
1.2.2	Kadar Gula Darah - TINGGI-BB-Avarage	18	14	4	0,764204506	
	Usia					0,02080717
	Old	17	13	4	0,787126586	
	Medium	1	1	0	0	
	Young	0	0	0	0	
	Denyut Nadi					0,04306839
	Tinggi	16	12	4	0,811278124	
	Normal	2	2	0	0	
	Rendah	0	0	0	0	
	Tekanan Darah					0,36435714
	Tinggi	14	13	1	0,371232326	
	Normal	2	0	2	0	
	Rendah	2	1	1	1	

Dari hasil perhitungan diatas pada tabel 5, atribut yang menjadi *Node* cabang dari Berat Badan *Avarage* adalah Tekanan Darah. Dimana sub atribut dari Tekanan Darah adalah Tinggi, Normal dan Rendah. Masing-masing sub kriteria sudah menghasilkan klasifikasi. Dari hasil tersebut dapat digambarkan pohon keputusan akhir tampak seperti gambar 1. berikut ini.



Gambar 1. Pohon Keputusan hasil Node akhir

Dari perhitungan diatas terdapat delapan *rules* yang dapat dijadikan sebagai referensi dalam memprediksi penyakit diabetes. Adapun aturan *rules* diatas berupa lima *rules* keputusan positif dan tiga *rules* keputusan negatif dapat dijabarkan melalui narasi sebagai berikut :

- Jika Kadar Gula Darah (mg/dl) = Normal dan Tekanan Darah (NmHg) = Rendah, maka Hasil = Positif
- Jika Kadar Gula Darah (mg/dl) = Normal dan Tekanan Darah (NmHg) = Tinggi, maka Hasil = Negatif
- Jika Kadar Gula Darah (mg/dl) = Rendah, maka Hasil = Negatif
- Jika Kadar Gula Darah (mg/dl) = Tinggi dan Berat Badan (kg)= *Avarage* dan Tekanan Darah (NmHg) = Normal, maka Hasil = Negatif
- Jika Kadar Gula Darah (mg/dl) = Tinggi dan Berat Badan (kg)= *Avarage* dan Tekanan Darah (NmHg) = Rendah, maka Hasil = Positif (Positif=1 dan Negatif=1)
- Jika Kadar Gula Darah (mg/dl) = Tinggi dan Berat Badan (kg) = *Avarage* dan Tekanan Darah (NmHg) = Tinggi, maka Hasil = Positif (Positif=13 dan Negatif=1)
- Jika Kadar Gula Darah (mg/dl) = Tinggi dan Berat Badan (kg)= *Over Weight*, maka Hasil = Positif (Positif =18 dan Negatif=0)
- Jika Kadar Gula Darah (mg/dl) = Tinggi dan Berat Badan (kg)= *Under Weight*, maka Hasil = Negatif

Pengujian terhadap hasil perhitungan manual menggunakan *software RapidMiner* dilakukan melalui beberapa proses hingga diperoleh hasil pemrosesan data menggunakan *software rapidminer* didapat hasil akurasi penerapan algoritma C4.5 yaitu 90,00%. Untuk melihat nilai akurasi dapat dilihat pada gambar 2 berikut ini :

accuracy: 90.00% +/- 10.00% (mikro: 89.80%)			
	true Positif	true Negatif	class precision
pred. Positif	31	3	91.18%
pred. Negatif	2	13	86.67%
class recall	93.94%	81.25%	

Gambar 2. Nilai Akurasi Algoritma C4.5

PerformanceVector	
PerformanceVector:	
accuracy: 90.00% +/- 10.00% (mikro: 89.80%)	
ConfusionMatrix:	
True: Positif Negatif	
Positif: 31 3	
Negatif: 2 13	
classification_error: 10.00% +/- 10.00% (mikro: 10.20%)	
ConfusionMatrix:	
True: Positif Negatif	
Positif: 31 3	
Negatif: 2 13	

Gambar 3. Nilai Performance Vector Algoritma C4.5

Berdasarkan pengolahan data menggunakan *software RapidMiner* didapat nilai akurasi sistem sebesar 90,00%, artinya bahwa *rule* yang dihasilkan tingkat kebenaran mendekati 100%. Dimana Jumlah benar prediksi positif sebanyak 31 data, salah prediksi positif sebanyak 3 data, jumlah benar prediksi negatif sebanyak 2 data sedangkan salah prediksi negatif sebanyak 13 data. Prediksi positif dari seluruh data menghasilkan 90,00% seseorang terkena penyakit diabetes. Parameter yang digunakan pada *decission tree* yang digunakan pada rapidminer yaitu *information gain*, *minimal size for split=2*, *minimal leaf size=1*, *minimal gain=1*, *maximal depth=20*, *confidence=0.25* dan *number of pruning alt=3*. Sesuai dengan ketentuan tersebut maka hasil perhitungan manual dengan rapidminer dapat

dikatakan berhasil dan dapat menemukan pohon keputusan pada kasus penyakit diabetes.

4. Kesimpulan

Adapun kesimpulan yang dapat diambil pada Implementasi Data Mining untuk penyakit diabetes dengan menggunakan Algoritma C4.5 adalah sebagai berikut :

- a) Implementasi Data Mining dengan menggunakan Algoritma C4.5 dapat membantu peneliti dalam prediksi penyakit diabetes berdasarkan variabel-variabel yang diperlukan yaitu usia, tekanan darah, denyut nadi, berat badan dan kadar gula darah. Pada penelitian ini menghasilkan jumlah prediksi positif sebanyak 36 data dan jumlah prediksi negatif sebanyak 2. Prediksi positif dari seluruh data menghasilkan 90,00%.
- b) Dengan menerapkan Algoritma C4.5 pada software Rapidminer menghasilkan hasil yang sama pada perhitungan manual. Berdasarkan pengolahan data menggunakan *software RapidMiner* didapat nilai akurasi sistem sebesar 90,00%, artinya bahwa *rule* yang dihasilkan tingkat kebenaran mendekati 100%.

Daftar Pustaka

- [1] 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.
- [2] M. F. Arifin And D. Fitriana, "Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : Pt Atria Artha Persada," *J. Telekomun. Dan Komput.*, Vol. 8, No. 2, Pp. 87–101, 2018, Doi: 10.22441/Incomtech.V8i1.2198.
- [3] B. Sugara, D. Widyatmoko, B. S. Prakoso, And D. M. Saputro, "Penerapan Algoritma C4.5 Untuk Deteksi Dini Autisme Pada Anak," *Semin. Nas. Teknol. Inf. Dan Komun.*, Vol. 2018, No. Sentika, Pp. 87–96, 2018.
- [4] A. E. Pramadhani, "2613-7473-1-Pb," Vol. 2, 2014.
- [5] S. R. Ningsih, I. S. Damanik, A. P. Windarto, H. S. Tambunan, J. Jalaluddin, And A. Wanto, "Analisis K-Medoids Dalam Pengelompokan Penduduk Buta Huruf Menurut Provinsi," *Pros. Semin. Nas. Ris. Inf. Sci.*, Vol. 1, No. September, P. 721, 2019, Doi: 10.30645/Senaris.V1i0.78.
- [6] A. S. Sunge *Et Al.*, "Prediksi Kompetensi Karyawan Menggunakan Algoritma C4 . 5 (Studi Kasus : Pt Hankook Tire Indonesia)," Vol. 2018, No. Sentika, Pp. 23–24, 2018.
- [7] Ari Budiono, "Pohon Keputusan Dengan Algoritma C4.5 1.," Pp. 3–4, 2017.