

Analisis Kepuasan Pelanggan Menggunakan Algoritma C5.0

Dandi Junandi Purba^{1,*}, Muhammad Ridwan Lubis², Zulia Almaida Siregar²

¹ STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

² AMIK Tunas Bangsa, Pematangsiantar, Indonesia

Email: dandi.junandi3344@gmail.com,

Email Penulis Korespondensi: dandi.junandi3344@gmail.com

Abstrak—Air merupakan salah satu kebutuhan yang sangat diperlukan oleh makhluk hidup terutama manusia. Dalam kehidupan sehari-hari manusia tidak terlepas dari kebutuhan akan air. PDAM merupakan salah satu unit usaha milik daerah, yang bergerak dalam distribusi air bersih bagi masyarakat umum. Namun dalam perjalanannya seiring mendapat keluhan dari masyarakat atau pelanggan, keluhan masyarakat tentang semakin sulitnya untuk mendapatkan air bersih tampaknya masih menjadi kendala yang sepenuhnya belum dapat diatasi, misalnya saja daerah Sumatera Utara tepatnya di Kecamatan Gunung Malela. Di satu pihak permintaan masyarakat akan air bersih semakin meningkat, namun kualitas pelayanan yang diberikan belum sebanding dengan pemenuhan permintaan masyarakat tersebut. Maka dari itu penelitian ini dilakukan untuk mengetahui kepuasan pelanggan terhadap PDAM Tirta Lihou di Kecamatan Gunung Malela. Banyak cabang ilmu komputer yang mempelajari berbagai algoritma salah satunya metode data mining, SPK, JST, dan lain-lain. Penelitian ini menggunakan metode data mining dengan algoritma C5.0 untuk menyelesaikan masalah faktor yang mempengaruhi kepuasan pelanggan terhadap PDAM Tirta Lihou Kecamatan Gunung Malela.

Kata Kunci: Air; PDAM Tirta Lihou; Kepuasan Pelanggan; Data Mining; Algoritma C5.0

Abstract—Water is one of the necessities that are needed by living things, especially humans. In everyday life, humans cannot be separated from the need for air. PDAM is a regional-owned business unit, which is engaged in the distribution of clean water for the general public. However, along the way, along with receiving complaints from the public or customers, public complaints about the increasing difficulty of getting clean water are still an obstacle that can be completely overcome, for example in the North Sumatra area, precisely in Gunung Malela District. On the one hand, public demand will increase, but the quality of services provided is not comparable to meeting the demands of the community. Therefore, this study was conducted to determine customer satisfaction with PDAM Tirta Lihou in Gunung Malela District. Many branches of computer science study various algorithms, one of which is data mining methods, DSS, ANN, and others. This study uses a data mining method with the C5.0 algorithm to solve the problem of factors that affect customer satisfaction with PDAM Tirta Lihou, Gunung Malela District.

Keywords: Water; PDAM Tirta Lihou; Customer Satisfaction; Data Mining; Algorithm C5.0

1. PENDAHULUAN

Air bersih merupakan kebutuhan yang sangat mendasar bagi manusia yang berdampak langsung pada kesehatan dan kesejahteraan fisik. Dalam kehidupan sehari-hari manusia tidak terlepas dari kebutuhan akan air, oleh karena itu penyediaan air bersih sangat diperlukan bagi masyarakat. PDAM TirtaLihou merupakan salah satu unit usaha milik daerah, yang bergerak dalam distribusi air bersih bagi masyarakat umum, Namun dalam perjalanannya seiring mendapat keluhan dari masyarakat atau pelanggan, keluhan masyarakat tentang semakin sulitnya untuk mendapatkan air bersih tampaknya masih menjadi kendala yang sepenuhnya belum dapat diatasi. Misalnya saja daerah Sumatera Utara tepatnya di Kecamatan Gunung Malela. Di satu pihak permintaan masyarakat akan air bersih semakin meningkat, namun kualitas pelayanan yang diberikan belum sebanding dengan pemenuhan permintaan masyarakat tersebut. Maka dari itu penelitian ini dilakukan untuk mengetahui kepuasan pelanggan terhadap PDAM Tirta Lihou di Kecamatan Gunung Malela. Dengan menggunakan Algoritma C5.0 sumber data yang digunakan pada penelitian adalah dengan melakukan observasi dan wawancara yaitu dengan membagikan 200 kuisioner pada pelanggan PDAM TirtaLihou di kec. Gunung Malela. Hasil dari penelitian ini diharapkan mampu menentukan kepuasan pelanggan PDAM TirtaLihou agar pelayanan dapat ditingkatkan. Adapun penelitian terdahulu yang menjadi salah satu acuan dari penelitian yang dilakukan oleh Putu Eki Ruziqa Maris dengan judul “Analisis Kepuasan Pelanggan Menggunakan Algoritma C4.5” yang bertujuan untuk mengetahui dan menganalisa Kepuasan Pelanggan Telkomsel dikota Semarang. Dari penelitian tersebut, dihasilkan keputusan untuk menentukan kepuasan pelanggan yang terlihat dari faktor harga kekuatan sinyal dan pelayanan pegawai Telkomsel [1].

Penelitian yang dilakukan oleh Basrah Saidani, Lisa Monita Lusiana, Shandy Aditya dengan judul “Analisis Pengaruh Kualitas Website dan Kepercayaan Terhadap Kepuasan Pelanggan Dalam Membentuk Minat Pembelian Ulang Pada Pelanggan Shopee” yang bertujuan untuk menganalisis pengaruh kualitas website terhadap kepuasan pelanggan, kepercayaan terhadap kepuasan pelanggan, kualitas website terhadap minat pembelian ulang, kepercayaan terhadap minat pembelian ulang dan kepuasan pelanggan terhadap minat pembelian ulang. Dari penelitian tersebut, dihasilkan keputusan bahwa responden merasa puas dilihat dari segi penggunaan Aplikasi Shopee yang mudah baik dari segi penggunaan secara umum maupun dari segi transaksi jual beli [2]. Penelitian yang dilakukan oleh Ibnu Mubarak, Fitri Marisa, Dwi Purnomo, dengan judul “Implementasi Algoritma Apriori Untuk Menentukan Kepuasan Mahasiswa Terhadap Pelayanan SIM Online Universitas Widyagama Malang” yang bertujuan untuk menentukan kepuasan mahasiswa terhadap pelayanan sim online. Dari penelitian tersebut dihasilkan suatu aplikasi yang dibangun menggunakan algoritma apriori untuk mengoptimalkan waktu pemrosesan pada saat pelayanan sim online sedan berlangsung [3]. Dari penelitian ini dihasilkan sebuah pohon keputusan yang dapat dijadikan sebagai tolak ukur dalam menganalisis tingkat kepuasan terhadap pelayanan pelanggan PDAM TirtaLihou. Diharapkan hasil dari penelitian ini dapat memberikan masukan kepada PDAM

TirtaLihou agar dapat memperbaiki system pelayanan terhadap konsumen, dalam upaya meningkatkan mutu pelayanan yang luas dan berkelanjutan.

2. METODOLOGI PENELITIAN

2.1 Proses Pengumpulan data

Penelitian ini dilakukan di PDAM Tirta Lihou yang beralamat di Bangun KM 17 Kabupaten Simalungun, yang mana dalam penelitian ini penulis melakukan pengumpulan data dengan waktu 1 minggu dari tanggal 08 Maret s.d 15 Maret 2021.

2.2 Analisa Data

Analisis data diperlukan dalam mengolah data agar menjadi suatu informasi sehingga data tersebut dapat dipahami dan bermanfaat dalam mencari faktor yang dominan terhadap menurunnya kesadaran masyarakat pada penerapan protokol kesehatan. Pada penelitian ini terdapat 4 atribut yang digunakan yaitu : Kualitas Pelayanan (C1), Perbaikan (C2), dan Kelancaran (C3).

2.3 Algoritma C5.0

AlgoritmaC5.0 adalah salah satu algoritma klasifikasi data mining yang khususnya diterapkan pada teknik *decision tree*.C5.0 merupakan penyempurnaan algoritma sebelumnya yang dibentuk oleh *Ross Quinlan* padatahun 1987, yaitu ID3 dan C4.5. Dalam memilih atribut untuk pemecah objek dalam beberapa kelas harus dipilih atribut yang menghasilkan *information gain* paling besar [7].

Adapun strategi pengembangan *decision tree* dengan menggunakan AlgoritmaC5.0 adalah sebagai berikut:

- Pada tahap awal, *tree* digambarkan sebagai *node* tunggal yang merepresentasikan *training set*.
- Jika sampel selanjutnya berisi kelas yang sama, maka *node* tersebut menjadi *leaf* dan diberi kelas tersebut.
- Jika tidak, algoritma dengan menggunakan ukuran berbasis *entropy (information gain)* akan memilih variabel prediktor yang akan memisahkan *record* ke dalam kelas-kelas individual, variabel tersebut menjadi variabel tes atau keputusan pada *node* tersebut.
- Cabang dikembangkan untuk tiap nilai yang diketahui dari variabel tes, dan sampel dipartisi berdasarkan cabang tersebut.
- Algoritma menggunakan proses yang sama secara rekursif membentuk *decision tree*.
- Partisi rekursif berakhir hanya ketika satu dari kondisi-kondisi berikut terpenuhi:
 - Seluruh *record* pada *node* tertentu memiliki kelas yang sama.
 - Tidak ada atribut yang tersisa pada *record* yang dapat dipartisi lebih lanjut. Dalam kasus ini suara mayoritas digunakan *node* tersebut menjadi *leaf node* dan dilabeli dengan kelas yang menjadi mayoritas dalam *record* yang ada [6].
- Tidak ada *record* untuk cabang variabel variabel tes. Dalam kasus ini, *leaf* terbentuk dengan mayoritas kelas sebagai label *record* tersebut.

Untuk menentukan akar dari pohon keputusan ditentukan oleh *gain* yang tertinggi, sebelum menemukan *gain* terlebih dahulu menghitung *entropy* keseluruhan dan *entropy* dari setiap atribut seperti yang terlihat pada rumus dibawah :

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

Ket :

- S = Himpunan Kasus;
 S1 = jumlah sampel;
 n = jumlah partisi atribut;
 Pi = proporsi kelas.

Setelah *entropy* diperoleh selanjutnya menentukan *gain* dari setiap atribut, dan mencari *gain* yang tertinggi untuk dijadikan akar dari cabang pohon keputusan seperti rumus dibawah ini :

$$Gain(S,A) = Entropy(S) - \sum_{t=1}^n \frac{[S1]}{[S]} * Entropy(S1) \quad (2)$$

Ket :

- A = atribut;
 S = himpunan kasus;
 [S1] = jumlah kasus pada partisi ke-I;
 [S] = jumlah kasus dalam S.

3. HASIL DAN PEMBAHASAN

3.1 Pengolahan Data

Langkah-langkah dalam pembentukan pohon keputusan pada tingkat Kepuasan Pelanggan PDAM Tirta lihou dilakukan pada yang tertera pada tabel 1. Untuk perhitungan AlgoritmaC5.0 dapat diuraikan sebagai berikut :

a) Menghitung entropy total :

$$\text{Entropy [Total]} = (-116/200) * \text{LOG}_2 (116/200) + (-84/200) * \text{LOG}_2 (84/200) = 0.981453895$$

b) Menghitung entropy dan gain (C1) :

$$\text{Entropy [Tidak Jernih]} = (-16/75) * \text{LOG}_2 (16/75) + (-59/75) * \text{LOG}_2 (59/75) = 0.747806158$$

$$\text{Entropy [Jernih]} = (-64/78) * \text{LOG}_2 (64/78) + (-14/78) * \text{LOG}_2 (14/78) = 0.6789539$$

$$\text{Entropy [Sangat jernih]} = (-36/47) * \text{LOG}_2 (36/47) + (-11/47) * \text{LOG}_2 (11/47) = 0.784992089$$

c) Gain

$$\text{Gain} = (0.981453895) - (75/200) * 0.747806158 - (78/200) * 0.6789539 - (47/200) * 0.784992089 = 0.251761424$$

Hal yang sama dilakukan terhadap seluruh atribut sehingga diperoleh hasil seperti pada tabel 1 dibawah ini :

Tabel 1. Hasil Perhitungan *Node 1.*

Node 1.	Jumlah	Puas	Tidak Puas	Entropy	Gain
Total	200	116	84	0.981453895	
Kualitas Air (C1)					0.251761424
Keruh	75	16	59	0.747806158	
Jernih	78	64	14	0.6789539	
Sangat Jernih	47	36	11	0.784992089	
Kelancaran (C2)					0.039051177
Tidak Lancar	67	28	39	0.980467931	
Lancar	133	88	45	0.923227009	
Pelayanan (C3)					0.581944589
Tidak Sopan	76	5	71	0.350010593	
Sopan	102	90	12	0.522559375	
Sangat Sopan	22	22	0	0	
Perbaikan (C4)					0.285567229
Tidak Tanggap	71	13	58	0.686808893	
Tanggap	100	82	18	0.680077046	
Sangat Tanggap	29	22	6	0.772627491	

Pada hasil perhitungan diatas nilai atribut tertinggi adalah C3 yang memiliki *gain* tertinggi **0.581944589**. Dimana C3 terdiri dari 3 kelas atribut yaitu Tidak Sopan, Sopan, dan Sangat Sopan. Berdasarkan *entropy* dari ketiga kelas atribut diatas, kelas atribut Tidak Baik dan Sopan mengklasifikasikan kasus yaitu Tidak Puas dan Sangat Sopan mengklasifikasikan kasus menjadi satu keputusan yaitu Puas. node 1.1 belum mendapatkan hasilnya 0 sedangkan pelayanan sopan menghasilkan puas karena sudah mendapatkan hasil 0 dari hasil perhitungan, begitu juga dengan pelayanan sangat sopan menghasilkan puas karena hasilnya 0. Selanjutnya pada kelas atribut Tidak Baik dapat ditunjukkan pada tabel 2 berikut ini:

Tabel 2. Hasil perhitungan *Node 1.1*

Node1.1	Jumlah	Puas	Tidak Puas	Entropy	Gain
Pelayanan-Tidak Baik	76	5	71	0.350010593	
Kualitas Air					0.148125156
Keruh	48	0	48	0	
Jernih	18	5	13	0.852405179	
Sangat Jernih	10	0	10	0	
Kelancaran					0.000602389
Tidak Lancar	33	2	30	0.370117999	
Lancar	43	3	41	0.33351464	
Perbaikan					0.062475323
Tidak Tanggap	51	1	50	0.139232999	
Tanggap	21	4	17	0.702466551	
Sangat Tanggap	4	0	4	0	

Dari hasil perhitungan pada Tabel 2 atribut yang menjadi *node* cabang dari (C3) Tidak Baik adalah (C1) Kualitas Air dengan *gain* tertinggi sebesar **0,148125156**. Nilai kelas atribut Keruh, Jernih dan Sangat Jernih menghasilkan keputusan Puas. Akan tetapi nilai kelas atribut Jernih belum diperoleh hasilnya, maka akan dilakukan perhitungan lebih lanjut.

Tabel 3. Hasil Perhitung *Node* 1.1.1

Node 1.1.1	Jumlah	Puas	Tidak Puas	Entropy	Gain
Pelayanan-Tidak Sopan, Kualitas Air –Jernih	18	5	13	0.852405179	0.027141125
Kelancaran					
	Tidak Lancar	10	2	0.721928095	0.569840383
	Lancar	8	3	0.954434003	
Perbaikan					0.569840383
	Tidak Tanggap	13	1	0.391243564	
	Tanggap	4	4	0	
	Sangat Tanggap	1	0	1	

Dari hasil perhitungan tabel 3 yang menjadi *node* dari cabang Pelayanan – Tidak sopan dan Kualitas Air –Jernih adalah perbaikan dengan *gain* tertingi **0,569840383** dengan kelas atribut Tidak Tanggap, Tanggap dan Sangat Tanggap. Kelas atribut Tanggap menghasilkan keputusan puas sedangkan Sangat Tanggap dan Tidak Tanggap menghasilkan keputusan Tidak Puas. Kelas atribut Tidak tanggap belum diperoleh hasil keputusan. Oleh karena itu kelas atribut Tidak Tanggap dilakukan perhitungan lebih lanjut. Perhitungan terus berlanjut sampai setiap atribut tidak dapat dibagi lagi atau bernilai nol. Hasil perhitungan akhir dapat dilihat pada tabel 4 berikut :

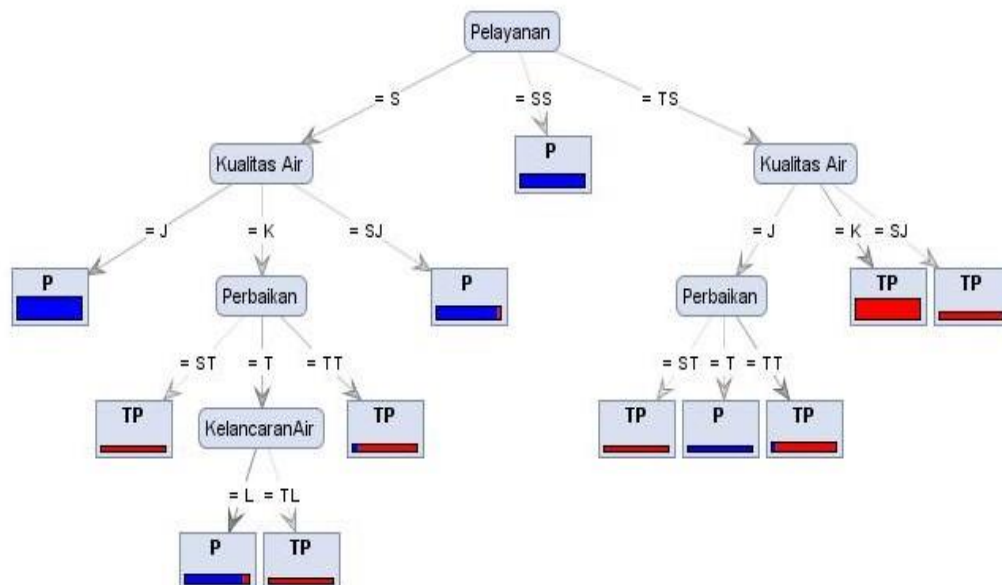
Tabel 4. Hasil perhitungan *Node* 1.1.1.2

Node 1.1.1.2	Jumlah	Puas	Tidak Puas	Entropy	Gain
Pelayanan-Tidak Baik, Kualitas Air-Jernih,Perbaikan-Tanggap	4	4	0	0	0
Kelancaran					
	Tidak Lancar	2	2	0	
	Lancar	2	2	0	0

Dari perhitungan tabel 4 di atas telah memperoleh keputusan Puas pada pada kelas atribut lancar. Dengan demikian node 1.1.1.2 adalah *node* cabang terakhir yang terbentuk. Maka proses perhitungan telah selesai dilakukan.

3.2 Hasil Percobaan

Hasil pengolahan data dengan model pohon keputusan sesuai dengan *software rapidminer*, dapat dilihat pada gambar 1 sebagai berikut :



Gambar 1. Hasil *decision tree*

Pada gambar 1 menghasilkan bentuk pohon keputusan yang berisikan variabel pelayanan puas. Variabel perbaikan menghasilkan – Tidak puas karena. Variabel Kualitas air menghasilkan – puas Sedangkan variabel Kualitas air menghasilkan – Tidak puas. Selanjutnya didapat pola pohon keputusan akhir setelah melakukan perhitungan dan pengujian data pada masing – masing atribut dengan AlgoritmaC5.0 sebagai berikut :

```

Tree
Pelayanan = S
|  Kualitas Air = J: P {P=54, TP=0}
|  Kualitas Air = K
|  |  Perbaikan = ST: TP {P=0, TP=1}
|  |  Perbaikan = T
|  |  |  KelancaranAir = L: P {P=11, TP=1}
|  |  |  KelancaranAir = TL: TP {P=0, TP=1}
|  |  |  Perbaikan = TT: TP {P=1, TP=8}
|  |  Kualitas Air = SJ: P {P=24, TP=1}
Pelayanan = SS: P {P=22, TP=0}
Pelayanan = TS
|  Kualitas Air = J
|  |  Perbaikan = ST: TP {P=0, TP=1}
|  |  Perbaikan = T: P {P=4, TP=0}
|  |  Perbaikan = TT: TP {P=1, TP=12}
|  Kualitas Air = K: TP {P=0, TP=48}
|  Kualitas Air = SJ: TP {P=0, TP=10}
    
```

Gambar 2. Deskripsi decision tree

Dari keputusan pada gambar 1 diperoleh rules sebagai berikut :

Tabel 5. Rules yang dihasilkan

No	Rules	Keputusan
1	Jika pelayanan sopan, Kualitas air jernih maka hasilnya	Puas
2	Jika pelayanan sopan, Kualitas air Keruh, Perbaikan Sangat Tanggap maka	Tidak Puas
3	Jika Pelayanan Sopan, Kualitas air Keruh, Perbaikan Tanggap, Kelancaran air Lancar maka	Puas
4	Jika pelayanan sopan, Kualitas air keruh, Perbaikan Tanggap, Kelancaran air Tidak Lancar maka	Tidak Puas
5	Jika pelayanan sopan, Kualitas air keruh, Perbaikan Tidak Tanggap maka	Tidak Puas
6	Jika Pelayanan Sopan, Kualitas air Sangat Jernih maka	Puas
7	Jika Pelayanan Sangat Sopan maka	Puas
8	Jika Pelayanan Tidak Sopan, Kualitas air Jernih, Perbaikan Sangat Tanggap maka	Tidak Puas
9	Jika Pelayanan Tidak Sopan, Kualitas air Jernih, Perbaikan Tanggap maka	Puas
10	Jika Pelayanan Tidak Sopan, kualitas air Jernih, Perbaikan Tidak Tanggap maka	Tidak Puas
11	Jika Pelayanan Tidak Sopan, Kualitas air Keruh maka	Tidak Puas
12	Jika Pelayanan Tidak Sopan, Kualitas air Sangat jernih maka	Tidak Puas

Berikut ini tingkat akurasi yang didapatkan :

accuracy: 98.33%			
	true P	true TP	class precision
pred. P	35	1	97.22%
pred. TP	0	24	100.00%
class recall	100.00%	96.00%	

Gambar 3. Nilai Accuracy

Berikut ini Performance yang di dapatkan :

```

PerformanceVector
PerformanceVector:
accuracy: 98.33%
ConfusionMatrix:
True:  P      TP
P:      35     1
TP:     0      24
    
```

Gambar 4. Nilai Performance Vektor

Berdasarkan pengolahan data menggunakan *software Rapidminer* didapat nilai akurasi 98.33%. Dimana model yang telah dibentuk diuji tingkat akurasinya dengan memasukkan data uji yang berasal dari data training dengan menggunakan *split validation* pada aplikasi *rapidminer 5.3*.

4. KESIMPULAN

Dari uraian pada bab sebelumnya, maka penulis mengambil beberapa kesimpulan yaitu, Dari hasil penelitian yang dilakukan dengan menggunakan datakepuasan pelanggan PDAM Tirtalihou, implementasi AlgoritmaC5.0 untuk menentukan kepuasan pelanggan PDAM Tirtalihou dapat diterapkan. Hasil uji coba yang dilakukan dengan *software RapidMiner 5.3* memperoleh nilai akurasi 98.33%

REFERENCES

- [1] Ruziqa, "Analisis Kepuasan Pelanggan Menggunakan Algoritma C4 . 5 Eki Ruziqa Maris," pp. 1-14, 2019.
- [2] B. Saidani *et al.*, "Analisis Pengaruh Kualitas Website," vol. 10, no. 2, pp. 425-444, 2019.
- [3] I. Mubarok *et al.*, "Kepuasan Mahasiswa Terhadap Pelayanan Sim Online," 2019.
- [4] S. Husen, K. Khotimah, and E. Sutisna, "Membangun Kepuasan Pelanggan Untuk Mengatasi Gap Antara Pengaruh Customer Experience Terhadap Loyalitas Pelanggan Pada Ud. Anwarudin Dan Apms ...," *J. Bus. ...*, 2021.
- [5] S. Wulandari, P. Poningsih, and W. Saputra, "Penerapan Metode Naive Bayes dalam Menentukan Pengaruh Penasihat Akademik pada Kelulusan Mahasiswa Tingkat Akhir," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 661, 2019.
- [6] F. Hadi, S. Kom, and M. Kom, "Penerapan Data Mining Dalam Menganalisa Pemberian Pinjamana Dengan Menggunakan Metode Algoritma C5 . 0 (Studi Kasus : Koperasi Jasa Keuangan Syariah Kelurahan Lambung Bukik)," vol. 4, no. 2, pp. 214-223, 2017.
- [7] M. Pardede, E. Bululolo, and E. Ndruru, "Implementasi Algoritma C5.0 Pada Kelulusan Peserta Ujian Kemahiran Berbahasa Indonesia (Ukbi) Pada Balai Bahasa Sumatera Utara," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 64-72, 2019.
- [8] M. Afdal and M. Rosadi, "Penerapan Association Rule Mining Untuk Analisis Penempatan Tata Letak Buku Di Perpustakaan Menggunakan Algoritma Apriori," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 99, 2019.