

IMPLEMENTASI DATA MINING PENYESUAIAN JENIS LENSA TERHADAP KEBUTUHAN PASIEN DENGAN MENGGUNAKAN ALGORITMA C4.5

Hosianna Saragih¹, Efori Buulolo², Fince Tinus Waruwu²

¹ Program Studi Teknik Informatika STMIK Budi Darma, Medan, Indonesia

² Dosen Tetap STMIK Budi Darma, Medan, Indonesia

^{1,2} Jln. Sisingamangaraja No. 338 Medan, Indonesia

Abstrak

Sulitnya memprediksi kebutuhan jenis lensa pada pasien, merupakan kendala yang dihadapi para pegawai di Optik Melawai dan membuat para pasien bingung dalam memilih jenis lensa yang sesuai dengannya. Sehingga para pegawai harus dapat mengambil keputusan yang tepat dan cepat guna memberikan pelayanan yang baik serta kepuasan kepada pasien dan membantu para pasien dalam memilih jenis lensa yang sesuai dengan mata pasien. Keputusan yang diambil harus mempertimbangkan dengan baik berdasarkan data - data yang dimiliki, terutama yang berkaitan erat dengan data pemeriksaan pasien. Penerapan algoritma C4.5 merupakan salah satu solusi yang digunakan dalam pemecahan masalah pada teknik klasifikasi yang memiliki karakteristik yaitu dengan proses penentuan nilai entropy dan nilai gain dari kemungkinan setiap kriteria yang menjadi acuan keputusan yang dilanjutkan dengan proses perangkingan dari hasil keputusan. Dalam memecahkan masalah yang ada, akan lebih simpel dan mudah dalam pengerjaannya, karena pengambilan keputusan yang sebelumnya kompleks dan sangat global, dapat diubah menjadi lebih simpel dan spesifik. Dari hasil pengujian, aplikasi yang dibangun maupun melakukan penerapan algoritma c4.5 dengan terbentuknya pohon keputusan sebagai bahan analisis keputusan memprediksi jenis lensa.

Kata Kunci: Jenis Lensa, Algoritma C4.5

Abstract

The difficulty of predicting the need for lens types in patients, is a constraint faced by employees in Opelics Melawai and make the patients confused in choosing the appropriate lens type. So that employees should be able to take the right decision and quickly in order to provide good service and satisfaction to patients and assist patients in choosing the type of lens that suits the patient's eyes. Decisions taken should be well considered based on the data - the data held, especially those that are closely related to patient examination data. Application of C4.5 algorithm is one of the solutions used in problem solving on classification techniques that have characteristics that is with the process of determining the value of entropy and the value of gain from the possibility of each criterion which is the reference of the decision followed by the ranking process of the decision. In solving the existing problem, it will be simpler and easier to work on, because the previous decision is complex and very global, can be changed to be more simple and specific. From the test results, the application is built and do the application of algorithm c4.5 with the formation of decision tree as a decision analysis materials predict the type of lens.

Keywords: Lens Type, Algorithm C4.5

1. PENDAHULUAN

Dengan kemajuan teknologi informasi dewasa ini, kebutuhan akan informasi yang akurat sangat dibutuhkan dalam kehidupan sehari-hari, sehingga informasi akan menjadi suatu elemen penting dalam perkembangan masyarakat saat ini dan waktu mendatang. Namun kebutuhan informasi yang tinggi kadang tidak diimbangi dengan penyajian informasi yang memadai, sering sekali informasi tersebut masih harus digali ulang dari data yang jumlahnya sangat besar. Kemampuan teknologi informasi untuk mengumpulkan dan menyimpan berbagai tipe data jauh meninggalkan kemampuan untuk menganalisis, meringkas dan mengekstrak pengetahuan dari data.

Sulitnya memprediksi kebutuhan jenis lensa pada pasien, merupakan kendala yang dihadapi para pegawai di Optik Melawai dan membuat para pasien bingung dalam memilih jenis lensa yang sesuai dengannya. Sehingga para pegawai harus dapat mengambil keputusan yang tepat dan cepat guna memberikan pelayanan yang baik serta kepuasan kepada pasien dan membantu para pasien dalam

memilih jenis lensa yang sesuai dengan mata pasien. Keputusan yang diambil harus mempertimbangkan dengan baik berdasarkan data - data yang dimiliki, terutama yang berkaitan erat dengan data pemeriksaan pasien. Memprediksi jenis lensa dengan menggunakan data - data yang telah ada di masa lalu untuk merekomendasikan kebutuhan lensa yang sesuai dengan pasien dimasa yang akan datang sangat diperlukan. Agar menjadi acuan untuk meningkatkan penjualan jasa dan produk melalui dengan memberikan hal - hal yang baru pada produk dan layanan terhadap pasien.

Penerapan algoritma C4.5 merupakan salah satu solusi yang sering digunakan dalam pemecahan masalah pada teknik klasifikasi yang memiliki karakteristik yaitu dengan proses penentuan nilai *entropy* dan nilai *gain* dari kemungkinan setiap kriteria yang menjadi acuan keputusan yang dilanjutkan dengan proses perangkingan dari hasil keputusan. Keluaran dari algoritma C4.5 yaitu berupa sebuah pohon keputusan (*decision tree*). Pohon keputusan adalah sebuah struktur yang dapat

digunakan untuk mengubah data menjadi pohon keputusan yang akan menghasilkan aturan-aturan keputusan. Dalam memecahkan masalah yang ada, akan lebih simpel dan mudah dalam pengerjaannya, karena pengambilan keputusan yang sebelumnya kompleks dan sangat global, dapat diubah menjadi lebih simpel dan spesifik. Dengan menerapkan algoritma C4.5 pada perhitungan manual, perhitungan yang dilakukan akan lebih mudah, karena akan mengeliminasi perhitungan-perhitungan yang tidak diperlukan dan *sample* diuji hanya berdasarkan kriteria atau kelas tertentu.

2. TEORITIS

2.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Sukma Putri Utari, "Implementasi Metode C4.5 untuk Menentukan Guru terbaik pada SMK 1 Percut Sei Tuan Medan", 2015).

2.2 Algoritma C4.5

The c4.5 algorithm is one of the data mining algorithms that included in the classification groups. c4.5 algorithms are used to form a decision tree.

Algoritma c4.5 adalah salah satu algoritma data mining yang termasuk dalam kelompok klasifikasi. Algoritma c4.5 digunakan untuk membentuk pohon keputusan. (Efori Buulolo, dkk, "C4.5 Algorithm To Predict the Impact of the Earthquake", 2017)

Sedang pohon keputusan dapat diartikan suatu cara untuk memprediksi atau mengklarifikasi yang sangat kuat. Pohon keputusan dapat membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan.

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut.

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera dalam persamaan 1 berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Keterangan:

S : himpunan kasus

A: atribut

n : jumlah partisi atribut A

|Si| : jumlah kasus pada partisi ke -i

|S| : jumlah kasus dalam S

Sementara itu perhitungan entropi dapat dilihat pada persamaan 2 berikut:

$$Entropy(S) = \sum_{i=1}^n - pi * \log_2 pi \quad (2)$$

Keterangan:

S : himpunan kasus

A: fitur

n : jumlah partisi S

pi: proporsi dari Si terhadap S

Untuk memudahkan penjelasan mengenai algoritma C4.5 berikut ini disertakan contoh kasus yang dituangkan dalam tabel 1.

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut.

1. Pilih Atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Tabel 1. Keputusan Bermain Tennis

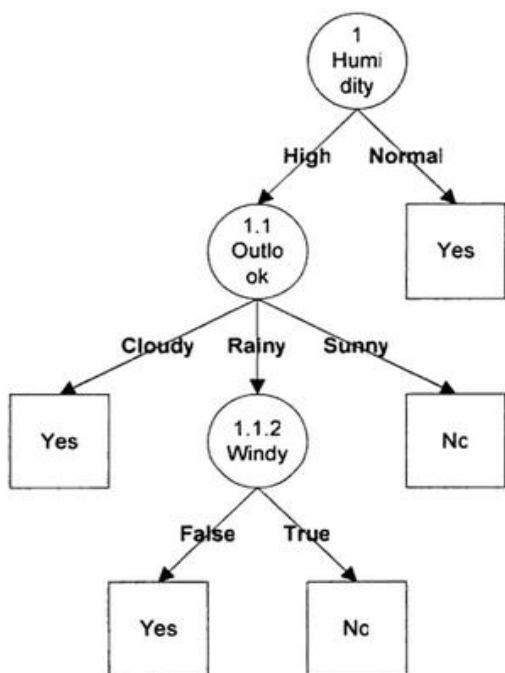
NO	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
1	Sunny	Hot	High	FALSE	No
2	Sunny	Hot	High	TRUE	No
3	Cloudy	Hot	High	FALSE	Yes
4	Rainy	Mild	High	FALSE	Yes
5	Rainy	Cool	Normal	FALSE	Yes
6	Rainy	Cool	Normal	TRUE	Yes
7	Cloudy	Cool	Normal	TRUE	Yes
8	Sunny	Mild	High	FALSE	No
9	Sunny	Cool	Normal	FALSE	Yes
10	Rainy	Mild	Normal	FALSE	Yes
11	Sunny	Mild	Normal	TRUE	Yes
12	Cloudy	Mild	High	TRUE	Yes
13	Cloudy	Hot	Normal	FALSE	Yes
14	Rainy	Mild	High	TRUE	No

Berikut hasil perhitungan node 1 yang dihasilkan melalui perhitungan Algoritma C4.5 pada tabel 2:

Tabel 2. Perhitungan Node 1

Node		Jml Kasus (S)	Tidak (S _i)	Ya (S _j)	Entropy	Gain
1	TOTAL	14	4	10	0.863120569	
	OUTLOOK					0.258521037
	CLOUDY	4	0	4		
	RAINY	5	1	4	0.721928095	
	SUNNY	5	3	2	0.970950594	
	TEMPERATURE					0.183850925
	COOL	4	0	4	0	
	HOT	4	2	2	1	
	MILD	6	2	4	0.918295834	
	HUMIDITY					0.370506501
	HIGH	7	4	3	0.985228136	
	NORMAL	7	0	7	0	
	WINDY					0.005977711
	FALSE	8	2	6	0.811278124	
	TRUE	6	4	2	0.918295834	

Berikut Pohon Keputusan yang diperoleh melalui perhitungan Algoritma C4.5 pada Gambar 2.1. (Kusrini, Emha Taufiq Luthfi, Algoritma Data Mining, 2009)



Gambar 1. Pohon Keputusan Hasil Perhitungan Node 1.1.2

3. ANALISA DAN PEMBAHASAN

Data yang digunakan dalam membentuk pohon keputusan adalah data pasien dengan kriteria-kriteria yang telah ditetapkan petugas laboratorium dalam pemeriksaan mata pasien pada Optik Malawai Medan.

Tabel 3. Hasil Test Laboratorium Pasien

No	Date	Name	Age	Spectacle Prescription	Astigmatism	Tear Production rate	Rekomended Lenses
1	16/01/2014	Ramadkan Sjeh	Young	Myope	No	reduced	none
2	08/03/2014	Yantisana Laowo	Young	Myope	No	normal	soft
3	24/05/2014	Sri Kuncoro	Young	Myope	Yes	reduced	none
4	15/07/2014	Hj. Zulaila	Young	Myope	Yes	normal	hard
5	06/08/2014	Margaretha Yado	Young	Hypermetrope	No	reduced	none
6	27/09/2014	Rosiana Tarigan	Young	Hypermetrope	No	normal	soft
7	17/10/2014	Lindobora Tinambunan	Young	Hypermetrope	Yes	reduced	none
8	19/12/2014	Desriany	Young	Hypermetrope	Yes	normal	hard
9	20/02/2015	Manto Efendi Sianturi	Pre-Presbyopic	Myope	No	reduced	none
10	06/04/2015	Iryelly	Pre-Presbyopic	Myope	No	normal	soft
11	04/05/2015	Sunardi Halomoan Hasibuan	Pre-Presbyopic	Myope	Yes	reduced	none
12	11/06/2015	Anna	Pre-Presbyopic	Myope	Yes	normal	hard
13	06/07/2015	Robia Flora	Pre-Presbyopic	Hypermetrope	No	reduced	none
14	17/08/2015	Susilawati	Pre-Presbyopic	Hypermetrope	No	normal	soft
15	02/11/2015	Dewi Syahputri	Pre-Presbyopic	Hypermetrope	Yes	reduced	none
16	10/12/2015	Siti Fatimah	Pre-Presbyopic	Hypermetrope	Yes	normal	none
17	10/01/2016	Rita Effida Lumbangaol	Presbyopic	Myope	No	reduced	none
18	03/03/2016	Riristiana Palpahan	Presbyopic	Myope	No	normal	none
19	21/05/2016	Marsita Parhusip	Presbyopic	Myope	Yes	reduced	none
20	12/07/2016	Anna Simbolon	Presbyopic	Myope	Yes	normal	hard
21	02/08/2016	Coky Mantap Siahaan	Presbyopic	Hypermetrope	No	reduced	none
22	24/09/2016	Asta Irdawati	Presbyopic	Hypermetrope	No	normal	soft
23	13/10/2016	Farida Hamzah	Presbyopic	Hypermetrope	Yes	reduced	none
24	10/12/2016	Dini H.Suci	Presbyopic	Hypermetrope	Yes	normal	none

Selanjutnya akan dilakukan praproses untuk menghasilkan data kasus yang siap untuk dibentuk menjadi pohon keputusan.

Praproses data meliputi :

1. Seleksi data
Untuk memilih data yang akan digunakan dalam penelitian ini, yaitu data pasien pada Optik Melawai Medan. Dari data yang ada, kolom yang diambil sebagai atribut/variabel dalam pembentukan pohon keputusan adalah :
 - a. Age
 - b. Spectacle Prescription
 - c. Astigmatism
 - d. Tear Production rate
 - e. Rekomended Lenses
2. Pembersihan data
Untuk membersihkan data, yaitu melengkapi data, menghapus data duplikat, menghilangkan noise.
3. Transformasi data

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk mendapatkan nilai *gain*, terlebih dahulu harus mencari nilai *entropy*. Dengan menggunakan dua persamaan berikut, seperti yang telah dijelaskan pada bab sebelumnya maka akan didapatkan nilai *entropy* dan *gain* yang digunakan sebagai akar dalam membuat pohon keputusan.

Tabel 4. Hasil Test Laboratorium Pasien

No	Age	Spectacle Prescription	Astigmatism	Tear Production rate	Rekomended Lenses
1	Young	Myope	No	reduced	none
2	Young	Myope	No	normal	soft
3	Young	Myope	Yes	reduced	none
4	Young	Myope	Yes	normal	hard
5	Young	Hypermetrope	No	reduced	none
6	Young	Hypermetrope	No	normal	soft
7	Young	Hypermetrope	Yes	reduced	none
8	Young	Hypermetrope	Yes	normal	hard
9	Pre-Presbyopic	Myope	No	reduced	none
10	Pre-Presbyopic	Myope	No	normal	soft
11	Pre-Presbyopic	Myope	Yes	reduced	none
12	Pre-Presbyopic	Myope	Yes	normal	hard
13	Pre-Presbyopic	Hypermetrope	No	reduced	none
14	Pre-Presbyopic	Hypermetrope	No	normal	soft
15	Pre-Presbyopic	Hypermetrope	Yes	reduced	none
16	Pre-Presbyopic	Hypermetrope	Yes	normal	none
17	Presbyopic	Myope	No	reduced	none
18	Presbyopic	Myope	No	normal	none
19	Presbyopic	Myope	Yes	reduced	none
20	Presbyopic	Myope	Yes	normal	hard
21	Presbyopic	Hypermetrope	No	reduced	none
22	Presbyopic	Hypermetrope	No	normal	soft
23	Presbyopic	Hypermetrope	Yes	reduced	none
24	Presbyopic	Hypermetrope	Yes	normal	none

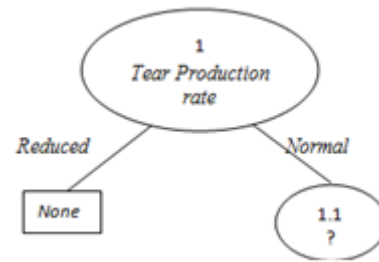
Setelah seluruh nilai *Entropy* dan *Gain* diperoleh, selanjutnya hasil dari perhitungan tersebut dimasukkan ke dalam Tabel 3.3, sebagai berikut :

Tabel 5. Perhitungan Node 1

Node		s	s1	s2	s3	Entropy	Gain
1	Total	24	15	5	4	1,326087	
	Age						0,039396
	Young	8	4	2	2	1,5	
	Pre-Presbyopic	8	5	2	1	1,298794	
	Presbyopic	8	6	1	1	1,061278	
	S. Prescription						0,039510
	Myope	12	7	2	3	1,384431	
	hypermetrope	12	8	3	1	1,188722	
	Astigmatism						0,377004
	No	12	7	5	0	0,979869	
	Yes	12	8	0	4	0,918296	
	T.Production						0,548794
	Reduced	12	12	0	0	0	
	Normal	12	3	5	4	1,554585	

Dari hasil Tabel 5 dapat diketahui bahwa atribut dengan gain tertinggi adalah *Tear Production rate. normal*, yaitu sebesar 0,548794. Dengan demikian, *Tear Production rate normal* menjadi *Node* akar.

Dari hasil tersebut dapat digambarkan pohon keputusan sementara tampak seperti Gambar 2.



Gambar 2. Pohon Node 1

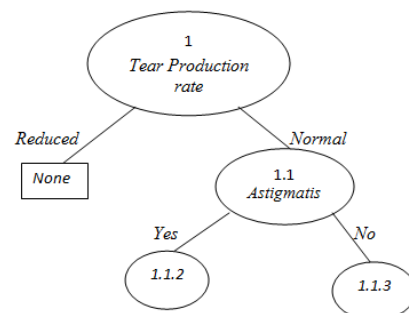
Untuk mencari *node* cabang berikutnya, maka harus dilakukan perhitungan kembali seperti yang tertera diatas, lakukan penghitungan nilai *gain* untuk tiap atribut. Hasil perhitungan ditunjukkan oleh Tabel 3.4 sebagai berikut:

Tabel 6. Perhitungan Node 1.1

Node		s	s1	s2	s3	Entropy	Gain
1.1	T.Production-normal	12	3	5	4	1,554585	
	Age						0,221252
	Young	4	0	2	2	1	
	Pre-Presbyopic	4	1	2	1	1,5	
	Presbyopic	4	2	1	1	1,5	
	S. Prescription						0,095437
	Myope	6	1	2	3	1,459148	
	hypermetrope	6	2	3	1	1,459148	
	Astigmatism						0,770426
	No	6	1	5	0	0,650022	
	Yes	6	2	0	4	0,918296	

Dari hasil Tabel 6. dapat diketahui bahwa atribut dengan gain tertinggi adalah *Astigmatism*, yaitu sebesar 0,770426. Dengan demikian, *Astigmatism* dapat menjadi *Node* 1.1.

Dari hasil tersebut dapat digambarkan pohon keputusan sementara tampak seperti Gambar 3.



Gambar 3. Pohon Node 1.1

Untuk mencari *node* cabang berikutnya, maka harus dilakukan perhitungan kembali seperti yang tertera diatas, lakukan penghitungan nilai *gain* untuk

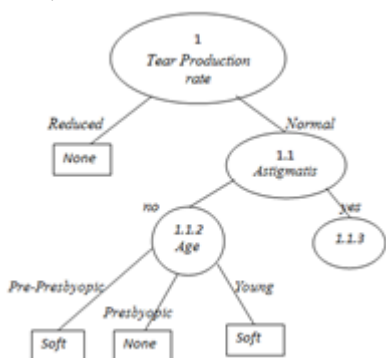
tiap-tiap atribut. Hasil perhitungan ditunjukkan oleh Tabel 7 sebagai berikut:

Tabel 7. Perhitungan Node 1.1.2

Node		s	s1	s2	s3	Entropy	Gain
1.1.2	T.Production-normal dan Astigmatism-no	6	1	5	0	0,650022	
	Age						0,316689
	Young	2	0	2	0	0	
	Pre-Presbyopic	2	0	2	0	0	
	Presbyopic	2	1	1	0	1	
	S. Presaption						0,190874
	Myope	3	1	2	0	0,918296	
	hypermetrope	3	0	3	0	0	

Dari hasil Tabel 7 dapat diketahui bahwa atribut dengan gain tertinggi adalah Age, yaitu sebesar 0,316689. Dengan demikian, Age dapat menjadi Node 1.1.2

Dari hasil tersebut dapat digambarkan pohon keputusan sementara tampak seperti gambar 4 dibawah ini:



Gambar 4. Pohon Node 1.1.2

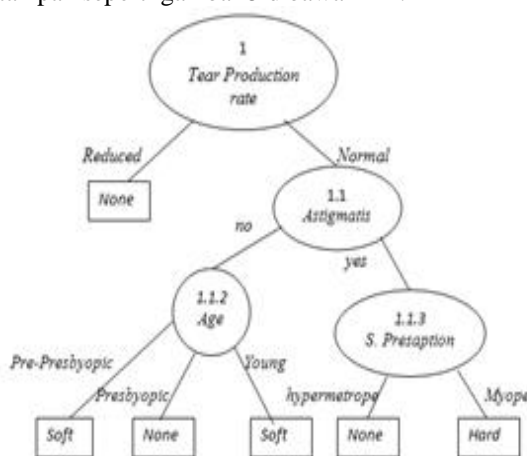
Untuk mencari node cabang berikutnya, maka harus dilakukan perhitungan kembali seperti yang tertera diatas. Hasil perhitungan ditunjukkan oleh Tabel 3.6 sebagai berikut :

Tabel 8. Perhitungan Node 1.1.3

Node		s	s1	s2	s3	Entropy	Gain
1.1.3	T.Production-normal dan Astigmatism=yes	6	2	0	4	0,918296	
	S. Presaption						0,459148
	Myope	3	0	0	3	0	
	hypermetrope	3	2	0	1	0,918296	

Dari hasil Tabel 3.6 dapat diketahui bahwa atribut dengan gainAge, yaitu sebesar 0,316689. Dengan demikian, Age menjadi Node 1.1.3 Dari hasil

tersebut dapat digambarkan pohon keputusan akhir tampak seperti gambar 5 dibawah ini :



Gambar 5. Pohon Node 1.1.3

Berdasarkan dari pohon keputusan yang terbentuk seperti pada gambar 5 maka dihasilkan rules sebagai berikut:

1. If Tear Production rate = reduced then Rekomended Lenses = none
2. If Tear Production rate = normal and Astigmatism=no and Age = Pre-Presbyopic then Rekomended Lenses = soft
3. If Tear Production rate = normal and Astigmatism=no and Age = Presbyopic then Rekomended Lenses = none or soft
4. If Tear Production rate = normal and Astigmatism=no and Age = Young then Rekomended Lenses = soft
5. If Tear Production rate = normal and Astigmatism=yes and Spectacle Presaption = hypermetrope then Rekomended Lenses = none
6. If Tear Production rate = normal and Astigmatism=yes and Spectacle Presaption = myope then Rekomended Lenses = hard

4. KESIMPULAN

Setelah dilakukan penelitian ini dan berdasarkan rumusan masalah yang ada, terdapat beberapa kesimpulan yaitu sebagai berikut :

1. Dengan menggunakan algoritma C4.5 untuk melakukan perhitungan, ternyata mampu menyelesaikan permasalahan para pegawai di Optik Malawai Medan dalam menyesuaikan jenis lensa pada pasien.
2. Dengan memanfaatkan data mining menggunakan algoritma C4.5 dalam menyesuaikan jenis lensa terhadap kebutuhan pasien, informasi yang dihasilkan bersifat klasifikasi yaitu mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan.

REFERENSI

- KUSRINI AND EMHA TAUFIQ LUTHFI, 2009.
Algoritma Data Mining, Theresia Ari Prabawati,
Ed. Yogyakarta, Yogyakarta: ANDI.
- SUKMA PUTRI UTARI, "**Implementasi Metode C4.5
untuk Menentukan Guru Terbaik pada SMK
1 Percut Sei Tuan Medan,**" *Pelita Informatika
Budi Darma*, vol. IX, no. 2301-9425, p. 3, April
2015.
- BUULOLO, E., SILALAH, N., FADLINA, & RAHIM, R.
(2017). **C4.5 Algorithm to Predict the Impact
of the Earthquake.** *International Journal of
Engineering Research & Technology (IJERT)* , 6
(02), 2.
- Buulolo, E. (2015). ALGORITMA APRIORI PADA
DATA PENJUALAN DI SUPERMARKET. In
*Seminar Nasional Inovasi dan Teknologi
Informasi 2015 (SNITI)* (pp. 4–7).
- Widayu, H., Nasution, S. D., Silalahi, N., & Mesran, M.
(2017). DATA MINING UNTUK
MEMPREDIKSI JENIS TRANSAKSI
NASABAH PADA KOPERASI SIMPAN
PINJAM DENGAN ALGORITMA C4.5.
MEDIA INFORMATIKA BUDIDARMA, 1(2)