

Perancangan Aplikasi Data Mining Untuk Menentukan Pasien Menderita Tifoid Dengan Metode Algoritma C4.5

Yusra Fadhilah¹, Muhammad Noor Hasan Siregar², Oberlin Siagian^{3*}

^{1,2,3} Universitas Graha Nusantara, Padangsidimpuan, Indonesia

¹yusra.fadilah18@gmail.com, ²noor.siregar@gmail.com, ^{3*}berlinsiagian99@gmail.com

^{*)} Email : yusra.fadilah18@gmail.com

Abstrak—Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Proses data mining ini akan diimplementasikan dengan menggunakan algoritma C.45. Algoritma C.45 dapat digunakan dalam pembentukan pohon keputusan tetapi lebih mengarah pada perhitungan probabilitas dari tiap-tiap record terhadap kategori-kategori tersebut atau untuk mengklarifikasi record dengan mengelompokkan ke dalam satu kelas. Setelah sebuah pohon keputusan dibangun maka dapat digunakan untuk mengklarifikasi record yang belum ada kelasnya. Dimulai dari node root menggunakan tes terhadap atribut dari record yang belum ada kelasnya, kemudian mengikuti cabang sesuai dengan proses pohon keputusan yaitu mengubah bentuk data (table) menjadi pohon (tree) kemudian merubah model pohon tersebut menjadi aturan. Berdasarkan penelitian yang telah dilakukan yang telah dibangun dapat menentukan pasien yang menderita penyakit tifoid.

Kata Kunci: Data Mining, Algoritma C.45, Decision Tree, PHP & MySql.

Abstract—Data mining is a term used to describe the discovery of knowledge in databases. This data mining process will be implemented using the C.45 algorithm. The C.45 algorithm can be used in the formation of decision trees but is more directed at calculating the probability of each record against these categories or to clarifying records by grouping them into one class. After a decision tree is built, it can be used to clarify records that do not yet have a class. Starting from the root node using tests on the attributes of records that do not have a class, then following the branches according to the decision tree process, which is to change the shape of the data (table) into a tree (tree) and then change the tree model into rules. Based on the research that has been done that has been built, it can determine which patients suffer from typhoid

Keywords: Data Mining, Algoritma C.45, Decision Tree, PHP & MySql.

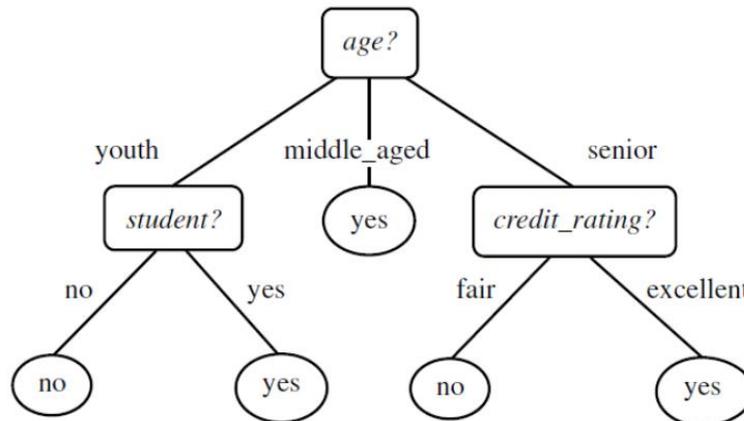
1. PENDAHULUAN

Penyakit demam Tifoid atau sering disebut dengan tifus abdominalis merupakan penyakit infeksi akut pada saluran pencernaan yang berpotensi menjadi penyakit sistemik yang disebabkan oleh salmonella [2]. Demam typhoid sendiri akan sangat berbahaya jika tidak segera ditangani secara baik dan benar, bahkan dapat menyebabkan kematian. Menurut data WHO (*World Health Organization*) memperkirakan angka insidensi di seluruh dunia sekitar 17 juta jiwa per tahun, angka kematian akibat demam typhoid mencapai 600.000 dan 70 % nya terjadi di Asia. Di Indonesia sendiri, penyakit typhoid bersifat *endemic*, menurut WHO angka penderita demam typhoid di Indonesia mencapai 81 % per 100.000 (Depkes RI, 2013). Berdasarkan data yang diperoleh dinas kesehatan sistem surveilans terpadu beberapa penyakit terpilih pada tahun 2010 penderita demam typhoid ada 44.422 penderita, termasuk urutan ketiga dibawah diare dan TBC selaput otak sedangkan pada tahun 2011 jumlah penderita tifoid di Sumatera Barat meningkat menjadi 46.142 penderita. Hal ini menunjukkan bahwa kejadian demam *typhoid* di Sumatera Barat termasuk tinggi. Masalah utama yang sering terjadi pada pasien penderita demam *typhoid* antara lain adalah demam, demam sering dijumpai, biasanya demam lebih dari seminggu, pada penderita demam *typhoid* juga ditemui masalah mual, muntah, nyeri abdomen atau perasaan tidak enak di perut, diare [4].

Secara umum data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining adalah proses yang menggunakan Teknik statistic, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [3].

Algoritma C4.5 merupakan struktur pohon dimana terdapat simpul yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal sampai tidak bias dibagi lagi. Algoritma C4.5 menggunakan konsep information gain atau entropy reduction untuk memilih pembagian yang optimal [5]. Pohon keputusan mirip sebuah struktur dimana terdapat *node* internal yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Pohon keputusan bekerja mulai dari akar paling atas, jika diberikan sejumlah data uji, misalnya X dimana kelas dari data X belum diketahui, maka pohon keputusan akan menelusuri mulai dari akar sampai node dan setiap

nilai dari atribut sesuai data X diuji apakah sesuai dengan aturan pohon keputusan, kemudian pohon keputusan akan memprediksi kelas dari tupel X [1]



Gambar 1. Konsep Pohon Keputusan

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 yaitu, [1]:

1. Menyiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama sebelum menghitung nilai gain dan atribut, hitung dahulu nilai entropy. Untuk menghitung nilai entropy digunakan rumus sebagai berikut :

$$\text{Entropi (S)} = \sum_i^k = 1 - p_i \log_2 p_i$$

Keterangan :

S = Himpunan Kasus

K = Jumlah Partisi S

Pi = Probabilitas yang didapat dari jumlah (ya/tidak) dibagi total kasus

3. Menghitung nilai gain menggunakan rumus :

$$\text{Gain (S,A)} = \text{Entropi (S)} - \sum_1^n = 1 - \frac{S_1}{S} * \text{Entropi (S}_1)$$

Keterangan :

S = Himpunan Kasus

A = Fitur

n = Jumlah partisi atribut A

|Si| = Proporsi Si terhadap S

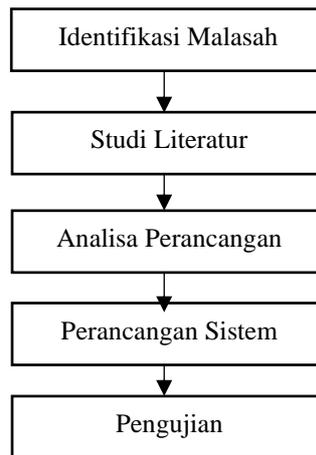
|S| = Jumlah kasus dalam S'

4. Ulangi langkah ke-2 hingga semua record terpartisi
5. Proses partisi pohon keputusan akan berhenti saat :
 - a. Semua record dalam simpul N mendapat kelas yang sama
 - b. Tidak ada atribut di dalam record yang dipartisi lagi
 - c. Tidak ada record di dalam cabang yang kosong

2. METODE PENELITIAN

2.1. Kerangka Penelitian

Dalam melakukan penelitian, salah satu langkah yang penting yaitu membuat kerangka pemikiran, kerangka pemikiran adalah merupakan suatu strategi untuk mencapai penelitian yang telah ditetapkan dan berperan sebagai pedoman atau penuntun peneliti seluruh proses penelitian.



Gambar 2. Kerangka Kerja Penelitian

1. Identifikasi masalah menentukan Batasan masalah yang diteliti dan memiliki konsistensi yang terarah, serta tercapainya tujuan dari penelitian ini
2. Studi literatur dilakukan agar peneliti mencari dan mempelajari sumber-sumber pengetahuan berupa jurnal-jurnal penelitian dan sumber pustaka otentik lainnya yang berkaitan dengan penelitian yaitu Algoritma C4.5, Data Mining, Typhoid dan Perancangan sistem.
3. Analisa Perancangan berupa kumpulan rangkaian data yang akan diolah dan di implementasikan ke dalam metode Algoritma C4.5
4. Perancangan sistem dilakukan untuk mendesain rancangan bentuk tampilan antarmuka, perancangan ini akan dilakukan perancangan program, perancangan sistem dilakukan setelah mendapat gambaran dengan jelas apa yang harus dikerjakan

3. HASIL DAN PEMBAHASAN

Khusus untuk penyakit Typhoid atribut yang digunakan adalah nama pasien, lingkungan, kondisi fisik dan makanan, dalam pengolahan data untuk memutuskan apakah seorang pasien menderita penyakit typhoid atau tidak.

Tabel 1. Data Pasien I

No	Nama Pasien	Lingkungan	Kondisi Fisik	Makanan
1	M. Ade Putra	Tidak Bersih	Kuat	Bergizi
2	Afdal Tanjung	Sangat Tidak Bersih	Sedang	Tidak Bergizi
3	Almaidi Saputra	Tidak Bersih	Kuat	Bergizi
4	Aldi Septian Alfino	Tidak Bersih	Sedang	Bergizi
5	Amin	Sangat Tidak Bersih	Sedang	Tidak Bergizi
6	Amrizal	Sangat Tidak Bersih	Lemah	Bergizi
7	Riski Fernandes	Tidak Bersih	Kuat	Bergizi
8	Disrizal	Bersih	Sedang	Tidak Bergizi
9	Doni Satria	Tidak Bersih	Lemah	Tidak Bergizi
10	Edo Rahmadani	Bersih	Kuat	Tidak Bergizi
11	Ema Fatmawati	Tidak Bersih	Kuat	Bergizi
12	Fadmah	Sangat Tidak Bersih	Sedang	Tidak Bergizi
13	Firmansyah	Tidak Bersih	Kuat	Bergizi
14	Hamzifah Rahmi	Tidak Bersih	Sedang	Bergizi
15	Jufriadi	Sangat Tidak Bersih	Sedang	Tidak Bergizi
16	Kahirudin	Sangat Tidak Bersih	Lemah	Tidak Bergizi
17	Nanda Saputra	Tidak Bersih	Kuat	Bergizi
18	Nawir	Bersih	Sedang	Tidak Bergizi
19	Rahman Syahputra	Tidak Bersih	Lemah	Tidak Bergizi
20	Rangga Ramadhan	Bersih	Kuat	Tidak Bergizi
21	Sukardi	Bersih	Sedang	Bergizi
22	Sesria Maharani	Sangat Tidak Bersih	Lemah	Tidak Bergizi

23	Suci Seviola	Tidak Bersih	Sedang	Bergizi
24	Putri Awalina	Tidak Bersih	Lemah	Tidak Bergizi
25	Pandi Agung	Bersih	Sedang	Bergizi
26	Pandra Pratama	Bersih	Sedang	Bergizi
27	Usnatun	Tidak Bersih	Lemah	Tidak Bergizi
28	Yuni Rahma	Sangat Tidak Bersih	Lemah	Tidak Bergizi
29	Yuliarti	Tidak Bersih	Kuat	Bergizi
30	Ziko Setiawan	Bersih	Kuat	Bergizi

Langkah – langkah algoritma C4.5 :

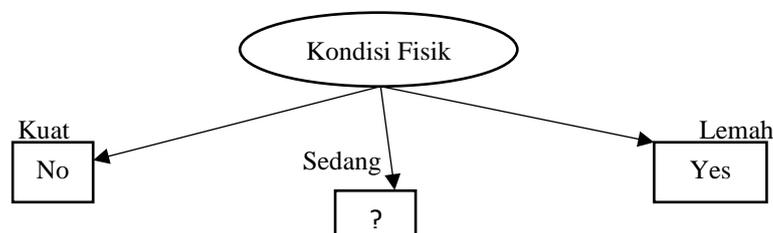
1. Pilih atribut sebagai akar (*root*).
 Dalam menentukan akar (*root*) perlu dilakukan pencarian nilai entropy dan nilai gain. Nilai *gain* yang tertinggi akan dijadikan sebagai akar (*root*)
2. Buat cabang untuk tiap-tiap nilai berdasarkan atribut yang ada, untuk membuat cabang dilakukan dengan pencarian nilai entrophynya
3. Bagi kasus kedalam cabang pada *decision tree*
4. Ulangi proses untuk semua cabang memiliki kelas yang sama.

Hal pertama yang dilakukan adalah melakukan pencarian *entropy*, kemudian dilanjutkan dengan pencarian gain dengan rumus yang telah disediakan. *Gain* yang tertinggi akan menjadi akar, dari hasil pencarian diatas diimplementasikan kedalam tabel seperti berikut :

Tabel 2. Pencarian Entropy Iterasi I

No	Node	Jumlah	Ya	Tidak	Entropy	Gain
1	Total	30	12	18	0.9709	
2	Makanan					0.2716
	Bergizi	16	2	14	0.5897	
	Tidak bergizi	14	10	4	0.8247	
3	Lingkungan					0.5681
	Bersih	8	0	8	0	
	Tidak BERSih	14	4	10	0.8246	
	Sangat Tidak Bersih	8	8	0	0	
4	Kondisi Fisik					0.5856
	Kuat	10	0	10	0	
	Sedang	12	4	8	0.9633	
	Lemah	8	8	0	0	

Dari tabel diatas atribut yang mempunyai nilai gain tertinggi dijadikan sebagai *node* akar, dimana kondisi fisik yang memiliki nilai *gain* tertinggi yang akan menjadi akar dari pohon keputusan sebagai berikut:



Gambar 3. Pohon Keputusan 1

Berdasarkan gambar diatas pohon keputusan dapat disimpulkan bahwa kondisi fisik mempengaruhi pasien yang menderita penyakit *typhoid*. Pasien yang memiliki kondisi fisik lemah maka pasien tersebut dikatakan dapat menderita penyakit *typhoid*, jika kondisi fisik pasien kuat maka pasien tersebut tidak menderita penyakit *typhoid*

Tabel 3. Data Pasien II

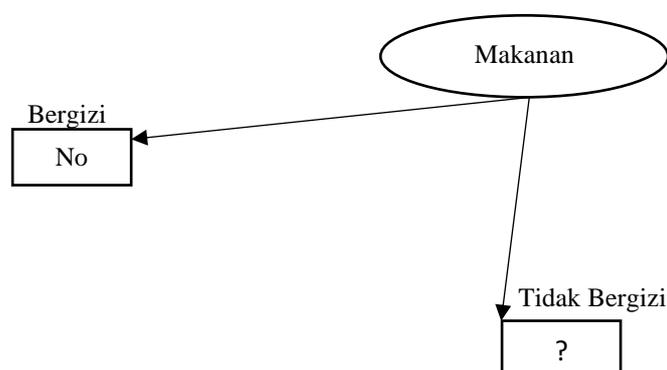
No	Nama Pasien	Kondisi Fisik	Lingkungan	Makanan	Keputusan
1	Afdal Tanjung	Sedang	Sangat Tidak Bersih	Tidak Bergizi	Ya
2	Almaidi Saputra	Sedang	Tidak Bersih	Bergizi	Tidak
3	Amin	Sedang	Sangat Tidak Bersih	Tidak Bergizi	Ya
4	Disrizal	Sedang	Bersih	Tidak Bergizi	Tidak
5	Fadmah	Sedang	Sangat Tidak Bersih	Tidak Bergizi	Ya
6	Hana Zifah	Sedang	Tidak Bersih	Bergizi	Tidak
7	Jufriadi	Sedang	Sangat Tidak Bersih	Tidak Bergizi	Ya
8	Nawir	Sedang	Bersih	Tidak Bergizi	Tidak
9	Sukardi	Sedang	Bersih	Bergizi	Tidak
10	Suci Seviola	Sedang	Tidak Bersih	Bergizi	Tidak
11	Pandra Pratama	Sedang	Bersih	Bergizi	Tidak

Kemudian dilanjutkan dengan pencarian *entropy* dan *gain* untuk iterasi II

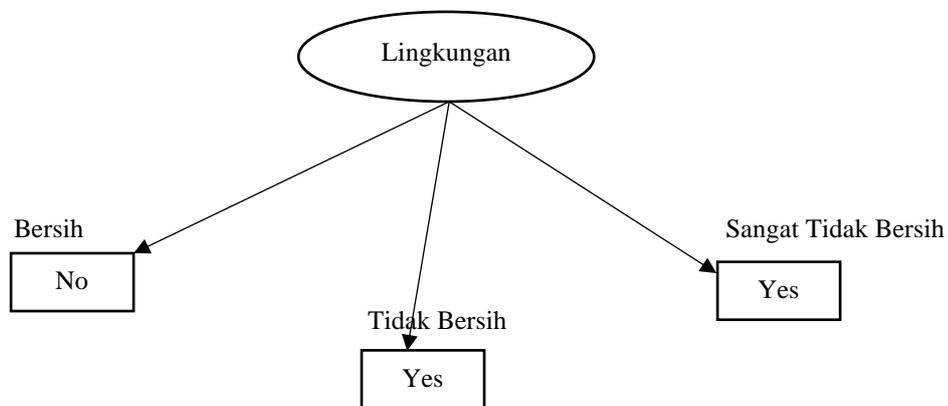
Tabel 4. Pencarian Entropy Iterasi II

No	Node	Jumlah	Ya	Tidak	Entropy	Gain
1	Total	12	4	8	0.9633	0
2	Lingkungan					0
	Bersih	5	0	5	0	
	Tidak Bersih	3	0	3	0	
	Sangat Tidak Bersih	4	4	0	0	
3	Makanan					0.4817
	Bergizi	6	0	6	0	
	Tidak Bergizi	6	4	2	0	

Dari tabel diatas atribut yang memiliki nilai *gain* tertinggi dijadikan sebagai *node* akar, dimana makanan yang memiliki nilai *gain* tertinggi yang akan menjadi akar dari pohon keputusan sebagai berikut :

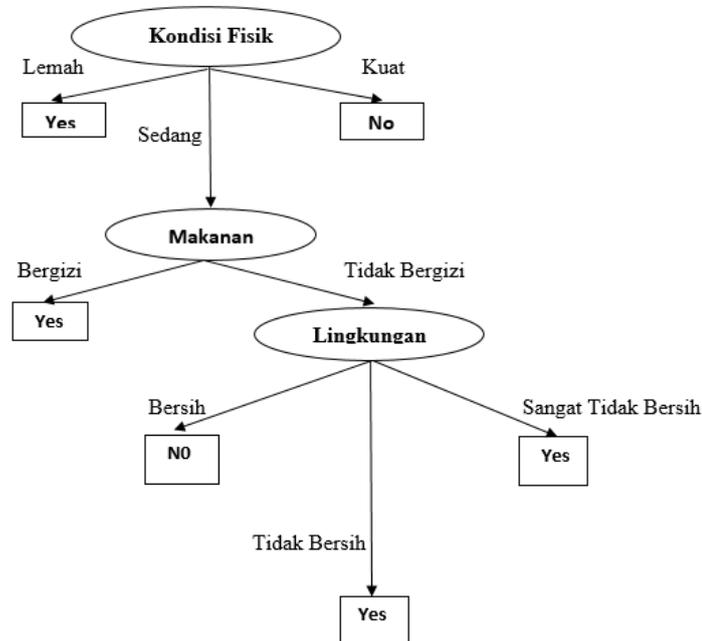

Gambar 4. Pohon Keputusan 2

Berdasarkan gambar diatas makanan juga mempengaruhi seorang pasien menderita penyakit *Typhoid*, jika pasien yang memiliki tanda- tanda *Typhoid* pada makanan tertentu dengan kategori makanan bergizi maka pasien tersebut tidak menderita penyakit *Typhoid*.



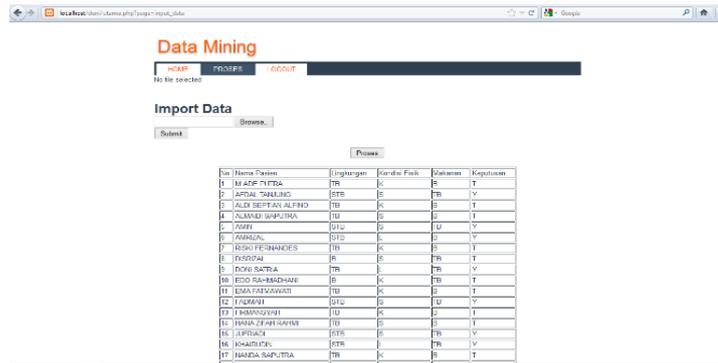
Gambar 5. Pohon Keputusan 3

Berdasarkan gambar diatas lingkungan juga mempengaruhi seorang pasien menderita penyakit *Typhoid*, jika lingkungan pasien tidak bersih maka pasien tersebut dapat dikatakan menderita penyakit *Typhoid*, jika lingkungan pasien sangat tidak bersih maka pasien tersebut menderita penyakit *Typhoid*, dan jika jika lingkungan pasien bersih maka pasien tersebut tidak menderita *Typhoid*. Dari kedua itersi tersebut dapat dihasilkan pohon keputusan sebagai berikut:

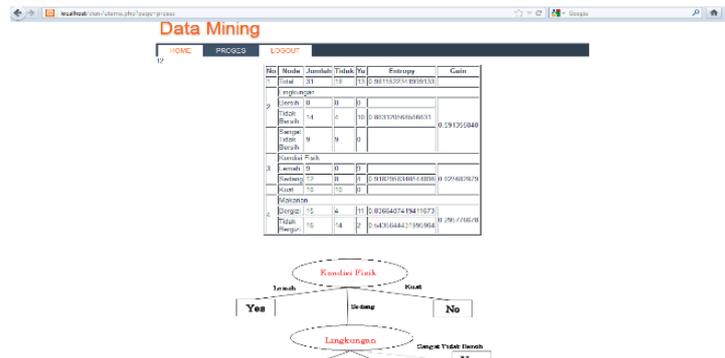


Gambar 6. Hasil Pohon Keputusan

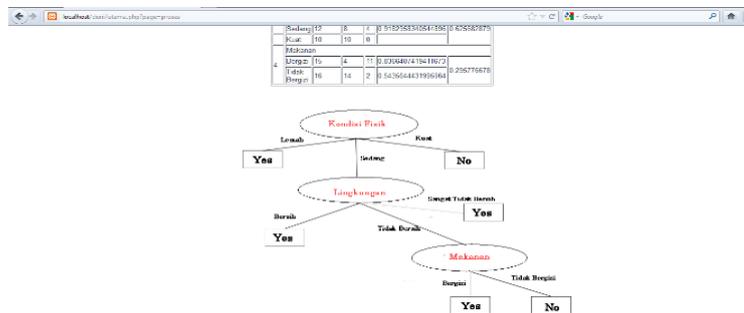
Pengujian dan implementasi bertujuan untuk melihat apakah aplikasi yang dirancang sudah sesuai dengan apa yang diinginkan atau belum, setelah dilakukannya pengujian dan implementasi maka kualitas sebuah aplikasi akan terlihat, tampilan aplikasi dapat dilihat step-by step yang menjelaskan tentang proses dimulainya sampai aplikasi ini selesai digunakan



Gambar 7. Halaman Import Data



Gambar 8. Halaman Proses dan Hasil



Gambar 9. Halaman Tampilan Pohon Keputusan Akhir

4. KESIMPULAN

Ada beberapa kesimpulan yang bias diambil dari penelitian ini, antara lain :

1. Dengan menerapkan aplikasi data mining , dokter dapat menentukan seorang pasien menderita penyakit typhoid atau tidak dengan cara melakukan panggilan data pasien yang ada pada database, kemudian dilakukan Analisa oleh aplikasi data mining
2. Dengan menerapkan aplikasi data mining dapat memprediksi pasien yang menderita penyakit Typhoid atau tidak dengan lebih cepat dan efisien. Penggunaan aplikasi data mining ini lebih cepat dan efisien dibandingkan dengan pendektiksaan secara manual yang memakan waktu dan kurangnya keefisien dalam mengolah data pasien. Dengan aplikasi data mining ini Dokter cukup memanggil data pasien yang telah ada dan dilakukan analisa oleh aplikasi data mining ini.
3. Dengan menerapkan aplikasi data mining dapat membantu Dokter dalam mengambil keputusan seorang pasien menderita penyakit Typhoid atau tidak menderita penyakit Typhoid. Dokter cukup memanggil data pasien didalam data base kemudian dilakukan analisa oleh aplikasi ini, maka akan mengeluarkan input berupa pohon keputusan dimana hasil akan terlihat apakah seorang pasien tersebut memiliki prediktor ataupun atribut- atribut yang menjadi penyebab penyakit Typhoid atau tidak dan aplikasi ini akan mengeksekusi data dengan pilihan ya atau tidak.
4. Dengan menerapkan aplikasi data mining dapat menghasilkan informasi yang bermanfaat bagi pihak rumah sakit dan pasien. Informasi yang dihasilkan bisa berupa faktor- faktor yang menjadi penyebab penyakit Typhoid dan cara pendeteksian penyakit Typhoid menggunakan aplikasi data mining.

REFERENCES

- [1] Kusriani dan Emha Taufiq Lutfi (2009). "Algoritma Data Mining" Yogyakarta : Andi
- [2] Cahyono, S. B (2010). Vaksinasi Cara Ampuh Cegah Penyakit Infeksi. Yogyakarta : Kanisisus
- [3] Turban (2009). "Sistem Informasi Manajemen". Pustaka Binaman Pressindo, Jakarta.
- [4] Nani dan Muzakir (2014). Kebiasaan Makan dengan Kejadian Demam typhoid pada anak, Journal of Pediatric Nursing
- [5] Han, J.,Kamber M., dan Pei, J. (2011) Data Mining Concepts and Techniques Third Edition. Waltham: Elsevier Inc.
- [6] Rosa A., S dan M., Shalahuddin (2013). Rekayasa Perangkat Lunak, Bandung : INFORMATIKA.
- [7] Mulyanto, Anur R., (2008). Rekayasa Perangkat Lunak 1, Jakarta : Direktorat Pembinaan Sekolah Menengah Kejuruan.
- [8] Aunur Rofiq Mulyanto (2008). Rekayasa Perangkat Lunak 2, Jakarta : Direktorat Pembinaan Sekolah Menengah Kejuruan.