

Implementasi Metode CART untuk Klasifikasi Diagnosis Penyakit Hepatitis Pada Anak

Anna Hendri Soleliza Jones^{#1}, Muhchromin Sucron Makmun^{*2}

^{#12} *Program Studi Teknik Informatika, Fakultas Teknik Industri, Universitas Ahmad Dahlan
Jl. Ringroad Selatan, Kragilan, Tamanan, Kec. Banguntapan, Bantul, Daerah Istimewa Yogyakarta, 55191*

^{3.1} annahendri@tif.uad.ac.id

^{3.2} muhchromin1500018153@webmail.uad.ac.id

accepted on 29-06-2021

Abstract

Penyakit hepatitis adalah salah satu ancaman kesehatan utama di dunia. Hepatitis merupakan peradangan pada hati yang biasanya disebabkan oleh virus hepatitis. Berdasarkan hasil riset kesehatan dasar kementerian RI tahun 2014, diperkirakan 10 dari 100 orang Indonesia terinfeksi hepatitis. Menurut Direktur Jenderal Badan Organisasi Kesehatan Dunia (WHO), Tedros Adhanom Ghebreyesus hanya ada 1 dari 10 orang yang pernah melakukan tes hepatitis dan hanya 1 dari 5 orang yang mendapatkan pengobatan hepatitis yang tepat dimana hepatitis A justru lebih sering menyerang anak-anak, terutama yang tinggal di area dengan sanitasi rendah. Penelitian bertujuan untuk membangun sistem aplikasi berbasis komputer dalam menentukan klasifikasi diagnosis penyakit hepatitis dengan metode CART. Data yang digunakan merupakan data dua tahun terakhir dari RSUD Sei Bahar yaitu sebanyak 240 data. Prinsip dari metode CART adalah memilah seluruh amatan menjadi dua gugus amatan dan memilah kembali gugus amatan tersebut menjadi dua gugus amatan berikutnya. Hasil klasifikasi menggunakan metode CART sebagai pengetahuan menentukan penyakit hepatitis. Dengan menggunakan 35 data uji, dan analisis rekomendasi dari pakar, didapatkan bahwa metode CART dapat digunakan sebagai metode pengklasifikasian pada penyakit hepatitis dengan tingkat akurasi sebesar 94%.

Keywords: : CART, Data Mining, Diagnosis Penyakit, Penyakit Hepatitis, Sistem Pakar

I. PENDAHULUAN

Penyakit hepatitis adalah satu dari sekian banyak penyakit yang menjadi ancaman kesehatan utama di dunia. Gaya hidup ternyata dapat memicu virus hepatitis, demikian yang dikemukakan oleh dr. Irsan Hasan SpPD-KGEH, Ketua Perhimpunan Peneliti Hati Indonesia. Menurut hasil Riskesdas tahun 2013 sebagaimana dalam data *prevalensi* hepatitis provinsi, Bahwa jumlah orang yang di diagnosis menderita penyakit hepatitis oleh Tenaga Kesehatan pada tahun 2013 berdasarkan gejala-gejala yang ada, mengalami peningkatan 2 kali lipat dibandingkan dengan tahun 2007. Hal ini dapat memberikan petunjuk awal tentang upaya pengendalian dimasa lalu dan akan menjadi masalah dimasa yang akan datang apabila tidak segera dilakukan upaya-upaya yang serius. Risiko kerusakan hati bahkan bisa berkembang menjadi kanker ketika dewasa.

Hepatitis merupakan peradangan pada hati yang biasanya disebabkan oleh virus. Hepatitis A, B dan C sama-sama disebabkan oleh virus, yaitu Hepatitis Virus tipe A (HVA), Hepatitis Virus tipe B (HVB) dan Hepatitis Virus tipe C (HVC)[4][10]. Menurut Direktur Jenderal Badan Organisasi Kesehatan Dunia (WHO), Tedros Adhanom Ghebreyesus hanya ada 1 dari 10 orang yang pernah melakukan tes hepatitis dan hanya 1 dari 5

orang yang mendapatkan pengobatan hepatitis yang tepat. Secara global didapatkan sekitar 1,4 juta kasus baru infeksi hepatitis pertahunnya. Menurut CDC (*Centers for Disease Control and Prevention*), Hepatitis justru lebih sering menyerang anak-anak, terutama yang tinggal di area dengan sanitasi masi rendah. Hasil observasi yang dilakukan di RSUD Sei Bahar didapat 240 data penderita penyakit hepatitis dari tahun 2017 sampai tahun 2018, sedangkan hasil observasi bahwa hampir 70% setiap pasien yang datang tidak mengetahui mengenai gejala dan penyakit hepatitis, sehingga dapat meningkatkan prosentase penderita penyakit hepatitis.

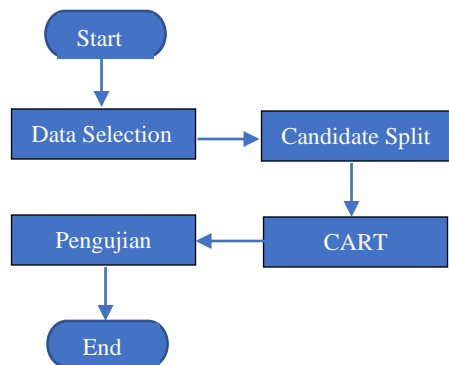
Penelitian ini menggunakan metode CART untuk melakukan klasifikasi menentukan penyakit hepatitis. Metode CART (*classification and Regression Tree*) merupakan metode untuk melakukan analisis klasifikasi. Prinsip dari klasifikasi adalah memilah seluruh amatan menjadi dua gugus amatan dan memilah kembali gugus amatan tersebut menjadi dua gugus amatan berikutnya, sampai diperoleh jumlah amatan minimum pada tiap-tiap gugus amatan berikutnya[14]. Metode CART belajar mempelajari data yang berasal dari masa lampau atau disebut juga data latihan, sehingga sistem akan berlatih dengan data latihan untuk mendapatkan pengetahuan [11].

Dalam membangun sebuah sistem pakar yang mampu mensimulasikan pengetahuan dan ketrampilan seorang pakar dipilihlah metode CART (*Classification and Regression Tree*) dalam penentuan sistemnya karena mampu untuk melakukan klasifikasi data kategorik[1][2]. Penelitian ini bertujuan membangun sistem aplikasi berbasis komputer untuk menentukan klasifikasi diagnosis penyakit hepatitis dengan metode CART yang memberikan rekomendasi cara penanganan penyakit hepatitis kepada pasien.[16].

II. METODOLOGI

2.1. Alur Penelitian

Alur penelitian yang dilakukan dimulai dengan pemilihan data yang digunakan dalam penelitian kemudian dilakukan *candidate split* yaitu membagi data menjadi data training dan data uji. Tahapan berikutnya mengimplementasikan metode CART kedalam data penelitian untuk menentukan pohon keputusan dan tahapan terkahir adalah pengujian untuk mengetahui akurasi hasil implementasi metode CART terhadap data dan kasus. Alur penelitian dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

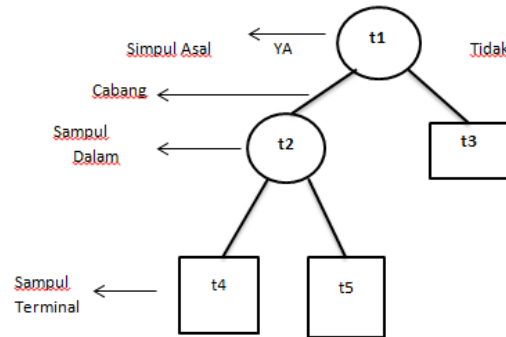
Basis pengetahuan yang digunakan dalam penelitian ini berupa data gejala hepatitis dan data penyakit hepatitis. Sedangkan mesin inferensi yang digunakan berupa penalaran maju atau *forward chaining*.

2.2. CART (*Classification and Regression Tree*)

Metode CART yaitu noktah keputusan yang selalu bercabang dua atau bercabang biner. CART dikembangkan untuk melakukan analisis klasifikasi. Prinsip dari klasifikasi adalah memilah seluruh amatan menjadi dua gugus amatan dan memilah kembali gugus amatan tersebut menjadi dua gugus amatan berikutnya, sampai diperoleh jumlah amatan minimum pada tiap-tiap gugus amatan berikutnya[5]. Langkah-langkah penentuan cabang.

1. susunlah calon cabang (*candidate split*). Penyusunan dilakukan terhadap seluruh variabel prediktor secara lengkap (*exhaustive*). Daftar yang berisi calon cabang disebut daftar calon cabang mutakhir.

2. Menilai kinerja keseluruhan calon cabang yang terdapat pada daftar calon cabang mutakhir dengan jalan menghitung nilai besaran kesesuaian $\Phi(s|t)$
3. Menentukan calon cabang manakah yang akan benar-benar dijadikan cabang dengan memilih calon cabang yang memiliki nilai kesesuaian $\Phi(s|t)$ terbesar. Jika tidak ada lagi noktah keputusan, pelaksanaan CART akan dihentikan. Namun, jika ada terdapat noktah keputusan, dilanjutkan dengan kembali ke langkah kedua, dengan terlebih dahulu membuang calon cabang yang sudah berhasil menjadi cabang sehingga mendapatkan daftar calon cabang mutakhir yang baru [15].



Gambar 2 Skema Cabang Pohon CART[19]

Kinerja dari setiap calon cabang akan diukur melalui ukuran kesesuaian *goodness* dengan rumus 1.

$$\Phi(s|t) = 2P_L P_R Q(s|t) \quad (1)$$

Untuk mendapatkan nilai $\Phi(s|t)$ didapat dari persamaan 2 dan nilai percabangan kiri dan kanan didapat dari persamaan 3 dan 4.

$$Q(s|t) = |P(j|tL) - P(j|tR)| \quad (2)$$

$$P_L = \frac{\text{calon cabang kiri } tL}{\text{data latihan}} \quad (3)$$

$$P_R = \frac{\text{calon cabang kanan } tR}{\text{data latihan}} \quad (4)$$

Keterangan:

S= calon cabang

t = noktah keputusan

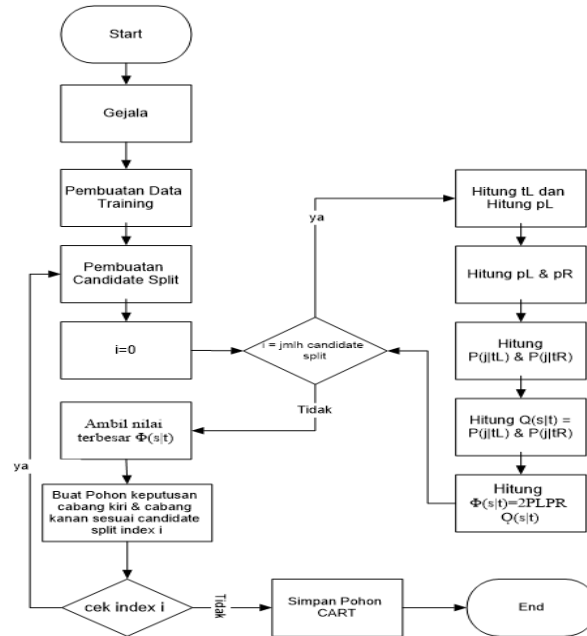
tL = cabang kiri dari noktah keputusan t

tR = cabang kanan dari noktah keputusan t

j = jumlah catatan

III. HASIL DAN DISKUSI

Metode CART akan menghasilkan nilai kesesuaian *goodness* $\Phi(s|t)$ disetiap node [15]. Nilai node digunakan untuk membuat pohon keputusan dan menentukan nilai-nilai disetiap gejala. Gejala yang digunakan untuk pembuatan data *training* [3][6]. Dari data *training* menentukan calon cabang (*candidate split*). Menentukan cabang mutakhir didapat dari nilai kesesuaian $\Phi(s|t)$ terbesar. Jika tidak ada lagi noktah keputusan, pelaksanaan metode CART dihentikan. jika masih ada terdapat noktah keputusan, kembali menghitung nilai kesesuaian $\Phi(s|t)$ terbesar, dengan terlebih dahulu membuang calon cabang yang sudah berhasil menjadi cabang sehingga mendapatkan daftar calon cabang mutakhir yang baru. Berikut *flowchart* perhitungan metode CART pada gambar 3.



Gambar 3 Flowchart metode CART

3.1. Data Training

Data yang didapat 240 data dari RSUD Sei Bahar. Penelitian menggunakan 35 data latih dengan lima variable, yaitu Kode Pasien, kode Gejala1, Kode Gejala2, Kode Gejala3, Penyakit.

Tabel 1. Data Training

kode pasien	Kode Gejala1	Kode Gejala2	Kode Gejala3	Hepatitis
P01	G01	G07	G06	A
P02	G02	G08	G13	B
P03	G03	G09	G14	A
...
P33	G04	G18	G08	B
P34	G04	G11	G13	B
P35	G04	G08	G17	A

3.2. Candidate Split

Didapat 23 calon cabang dari 35 data *training*. Ada dua kolom *candidate split* yaitu calon cabang kiri dan calon cabang kanan.

Tabel 2. Candidate Split

No	Calon Kiri	Calon kanan
1	G01	G02, G03, G04, G05, G06
2	G02	G01, G03, G04, G05, G06

...
22	G08	G06, G10, G12, G13, G14, G15, G16, G17
23	G10	G06, G08, G12, G13, G14, G15, G16, G17

Nilai P_L merupakan *candidate split* kiri terhadap data *training*. Jumlah G01, yaitu sebanyak 11 data, Kemudian dibagi 35 (data *training*).

$$P_L = 11/35 = 0,314$$

Nilai P_R merupakan *candidate split* kanan terhadap data *training*. Jumlah selain G01, yaitu sebanyak 24 data, Kemudian dibagi 35 (data *training*).

$$P_R = 24/35 = 0,685$$

$P(j|t_L)$ merupakan nilai probabilitas setiap penyakit pada data *training* yang bernilai calon cabang kiri. Jumlah Hepatitis G01 pada data P_L . Kemudian dibagi dengan jumlah data P_L .

$$\text{Hepatitis A} \quad 11/11 = 1$$

$$\text{Hepatitis B} \quad 0/11 = 0$$

$P(j|t_R)$ merupakan nilai probabilitas setiap penyakit pada data *training* yang bernilai calon cabang kanan. Jumlah Hepatitis selain G01 pada data P_R . Kemudian dibagi dengan jumlah data P_R .

$$\text{Hepatitis A} \quad 6/24 = 0,25$$

$$\text{Hepatitis B} \quad 18/24 = 0,75$$

Menentukan nilai $2P_L P_R$ yaitu dua dikali nilai P_L dikali nilai P_R .

$$2P_L P_R = 2 \times 0,314 \times 0,685 = 0,43102$$

$Q(s|t)$ merupakan nilai *consider* jenis penyakit hepatitis.

$$Q(s|t) = |1-0,25| + |0-0,75| = 1,5$$

$\Phi(s|t)$ merupakan nilai *goodness*. Proses perhitungan dilakukan disemua *candidate split*.

$$\Phi(s|t) = 0,43102 \times 1,5 = 0,64662$$

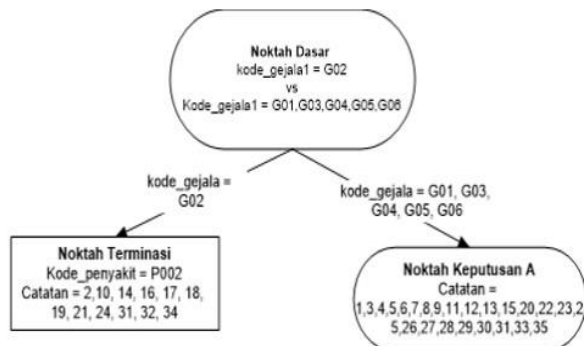
Berikut merupakan nilai $\Phi(s|t)$ disemua *candidate split* iterasi pertama disajikan pada tabel 3.

Tabel 3 Iterasi 1 CART

Split	P_L	P_R	$2P_L P_R$	$Q(s t)$	$\Phi(s t)$
1	0,287	0,714	0,40816	1,44	0,587755
2	0,342	0,657	0,45061	1,478260	0,666122
3	0,028	0,971	0,05552	1,058823	0,646530
...

21	0,114	0,885	0,202449	1,16129032	0,23510204
22	0,085	0,911	0,146939	1,0625	0,15812244
23	0,2	0,8	0,32	0,57142857	0,60735714

Pada iterasi 1 didapatkan nilai $\Phi(s|t)$ terbesar 0,666122 yaitu calon cabang urutan dua. Maka diperoleh calon cabang kiri G02 dan calon cabang kanan G01, G03, G04, G05, G06. Berikut pohon keputusan iterasi 1 pada gambar 4.



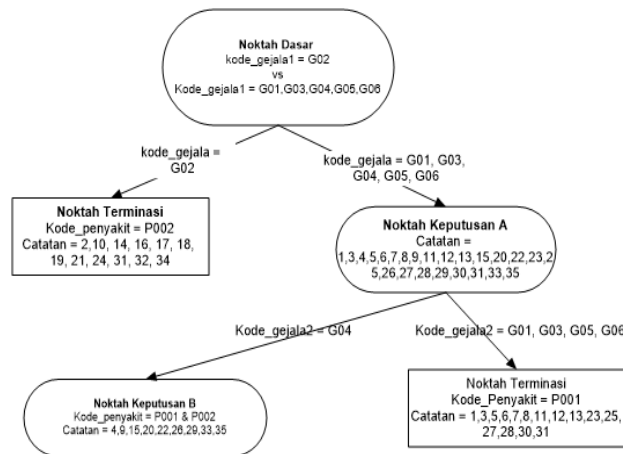
Gambar 4. Pohon Keputusan Iterasi 1

Karena nilai variabel prediktor bagi catatan-catatan cabang kanan masih ada yang memiliki penyakit hepatitis A dan B, maka dihasilkan noktah keputusan A.

Tabel 4 Iterasi 2

No	PI	Pr	2PLPR	Q (s t)	Φ (s t)
1	0,4782	0,52173	0,4990	1	0,49905
3	0,0434	0,95652	0,0831	0,545454	0,04536
4	0,3913	0,60869	0,4763	1,33333	0,635160
...
22	0,1304	0,8695	0,2268	1,7	0,38563
23	0,2608	0,7391	0,3856	0,2549	0,09829

Iterasi 2 didapat nilai terbesar yaitu 0,6351606 pada *split* 4.. Maka diperoleh calon cabang kiri G04 dan calon cabang kanan G01, G03, G05, G06. Berikut pohon keputusan iterasi 2 pada gambar 5.



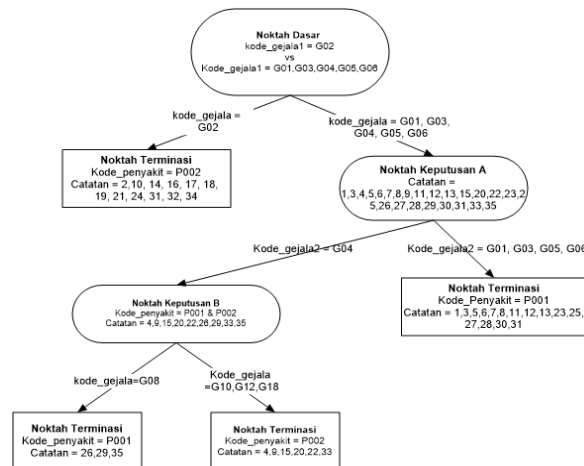
Gambar 5 Pohon Keputusan Iterasi 2

Karena nilai variabel prediktor masih terdapat penyakit hepatitis A dan hepatitis B. maka dihasilkan noktah keputusan B.

Tabel 5 Iterasi 3

No	PI	Pr	2PLPR	Q (s/t)	Φ (s/t)
8	0,3333	0,6666	0,444444	2	0,88888
10	0,1111	0,8888	0,197531	0,75	0,148148
11	0,3333	0,6666	0,444444	1	0,444444
...
16	0,1111	0,8888	0,197531	0,75	0,148148
18	0,1111	0,8888	0,197531	0,75	0,148148

Iterasi 3 didapat nilai terbesar yaitu 0,88888 pada *split* 8. Maka diperoleh calon cabang kiri G08 dan calon cabang kanan G10, G12, G18. Berikut pohon keputusan iterasi 3 pada gambar 6.



Gambar 6 Pohon Keputusan Iterasi 3

Karena nilai variabel prediktor seluruhnya penyakit hepatitis B, maka dihasilkan noktah terminasi. Sesuai namanya noktah terminasi tidak bisa bercabang lagi. Karena pada iterasi 3 tidak ada lagi noktah keputusan, iterasi dihentikan.[8]

3.3. Uji Vallidasi Confusiom Matrix

Pengujian menggunakan 35 data *training* hasil analisis pakar penyakit hepatitis. Data *traininig* tersebut dibandingkan dengan hasil klasifikasi yang dilakukan oleh sistem[12]. Hasil pengujian akurasi klasifikasi dapat dilihat pada tabel 6.

Tabel 6. Pengujian Akurasi Klasifikasi

No	Gejala	Hasil Analisis Pakar	Klasifikasi Oleh Sistem
1	Mudah lelah, Diare, Pusing	A	A
2	Mual, Lemah, sakit otot	B	B
3	Gatal Badan, Menggigil, Sakit kepala	A	A
4	Nafsu makan, nyeri sendi, kembung	B	B
5	Mudah lelah, Diare, Flu	A	A
6	Nyeri perut, lesu, demam	A	A
7	Mudah lelah, Mual ,Demam	A	A
8	Mudah Lelah, nyeri sendi, nyeri lambung	A	A
9	Nafsu makan, Lesu, Lemah	B	B
10	Mual, pegal linu , Sakit otot	B	B
11	Mudah lelah, Diare, Sakit otot	A	A
12	Pusing, Diare, Flu	A	A
13	mudah lelah, Pegal linu, Nyeri sendi	A	A
14	Mual ,lemah, nyeri sendi	B	B
15	Nafsu makan, Kembung, sakit otot	B	B
16	Mual, pegal linu, sakit otot	B	B
17	Nafsu makan, lemah, nyeri lambung	B	B
18	mual, lemah, sakit otot	B	B
19	Mual, Pegal linu, Sakit Otot	B	B
20	Nafsu makan, Lesu, lemah	B	B
21	Mual, pegal linu, sakit otot	B	B
22	Nafsu makan, kembung, nyeri sendi	B	B
23	Nafsu makan, pegal linu, nyeri sendi	A	B
24	Mual, pegal linu, sakit otot	B	B
25	Mudah lelah, diare, nyeri sendi	A	A
26	Nafsu makan, lemah, nyeri lambung	A	A

27	Mudah lelah, diare, flu	A	A
28	Mudah lelah, diare, nyeri sendi	A	A
29	Nafsu makan,lemah, nyeri lambung	A	A
30	Mudah lelah, diare, nyeri sendi	A	A
31	Mual, lemah , sakit otot	B	B
32	Mual, pegal linu ,sakit otot	B	B
33	Nafsu makan, lesu, lemah	B	B
34	Mual, pegal linu, sakit otot	B	B
35	Nafsu makan, Lesu, lemah	A	B

Hasil klasifikasi yang dilakukan oleh sistem dihitung akurasi dengan pengujian *confusion matrix*. Dapat dilihat pada tabel 7.

Tabel 7. Confusion Matrix

		True Values	
		A	B
Preidiction	A	15	1
	B	1	18

$$\text{Akurasi} = \frac{15+18}{15+1+1+18} = 0,94$$

$$\text{precision A} = \frac{15}{15+1} = 0,93$$

$$\text{precision B} = \frac{18}{1+18} = 0,94$$

Berdasarkan pengujian akurasi, didapat hasil akurasi data analisis pakar dengan sistem menggunakan metode CART sebesar 94% dengan *precision* hepatitis A sebesar 93%, dan *precision* hepatitis B sebesar 94%. Kesimpulan yang diperoleh dari pengujian akurasi adalah bahwa metode CART dapat digunakan sebagai metode pengklasifikasian pada analisis penyakit hepatitis karena tingkat akurasi yang tinggi.

V. KESIMPULAN

Dari penelitian yang telah dilakukan, dihasilkan sebuah *software* sistem pakar berbasis web untuk klasifikasi diagnosis penyakit hepatitis pada anak. Dengan menggunakan 240 data dan 35 data uji hasil klasifikasi menggunakan metode CART dan analisis rekomendasi dari pakar, didapatkan bahwa metode CART dapat digunakan sebagai metode pengklasifikasian untuk diagnosis penyakit hepatitis dengan tingkat akurasi hasil diagnosis sistem sebesar 94%.

REFERENSI

- [1] Abdul Aziz and Aziz Ur Rehman, "Detection of Cardiac Disease using Data Mining Classification Techniques" International Journal of Advanced Computer Science and Applications(IJACSA), 8(7), 2017.
- [2] B. Perez, C. Castellanos, and D. Correal, "Applying Data Mining Techniques to Predict Student Dropout: A Case Study," in *2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence, ColCACI 2018 - Proceedings*, 2018.
- [3] D. V. Paul, C. Nayagam, and J. D. Pawar, "Modeling Academic Performance using Subspace Clustering Algorithm," in *Proceedings - IEEE 8th International Conference on Technology for Education, T4E 2016*, 2017.
- [4] Fahad Ahmad, Kashaf Junaid and Ata ul Mustafa, "Relationship of Liver Enzymes with Viral Load of Hepatitis C in HCV Infected Patients by Data Analytics" International Journal of Advanced Computer Science and Applications(IJACSA), 9(11), 2018.
- [5] K. Kohli and S. Birla, "Data Mining on Student Database to Improve Future Performance," *Int. J. Comput. Appl.*, 2016.
- [6] M. Anoopkumar and A. M. J. Md Zubair Rahman, "A Review on Data Mining techniques and factors used in Educational Data Mining to predict student amelioration," in *Proceedings of 2016 International Conference on Data Mining and Advanced Computing, SAPIENCE 2016*, 2016.
- [7] Manikandan .R, Ramesh R."A Schematic Data Mining Approach for Web Page Recommendation Systems", Volume. 3 Issue. 10, October- 2018, International Journal of Innovative Science and Research Technology (IJISRT) ,www.ijisrt.com , ISSN - 2456-2165, PP:-489-492.
- [8] Mukhtar, Ismail."The use of Certainty Factor (CF) in Technostress Diagnatation Expert System", Volume. 4 Issue. 5, - 2019, International Journal of Innovative Science and Research Technology (IJISRT), www.ijisrt.com. ISSN - 2456-2165, PP:-727-732.
- [9] Novarina, A. T., & Santoso, E. (2018). Sistem Pakar Diagnosis Penyakit Hepatitis Menggunakan Metode Dempster Shafer. *Jurnal Pengermbananagan Teknologi Informasi Dan Ilmu Komputer*, 2(6), 2252–2258.
- [10] Ramdhani, A., Isnanto, R. R., & Windasari, I. P. (2015). Pengembangan Sistem Pakar Untuk Diagnosis Penyakit Hepatitis Berbasis Web Menggunakan Metode Certainty Factor. *Jurnal Teknologi Dan Sistem Komputer*, 3(1), 58–64. <https://doi.org/10.14710/JTSISKOM.3.1.2015.58-64>.
- [11] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, 2017.
- [12] Sri Winiarti, Herman Yuliansyah and Aprial Andi Purnama, "Identification of Toddlers' Nutritional Status using Data Mining Approach" International Journal of Advanced Computer Science and Applications(IJACSA), 9(1), 2018. <http://dx.doi.org/10.14569/IJACSA.2018.090122>.
- [13] Sahar F. Sabbeh, "Machine-Learning Techniques for Customer Retention: A Comparative Study" International Journal of Advanced Computer Science and Applications(IJACSA), 9(2), 2018.
- [14] Saranya.E, Maheswaran.T."IOT Based Disease Prediction and Diagnosis System for Healthcare Using Data Mining Techniques", Volume. 3 Issue. 10, October- 2018, International Journal of Innovative Science and Research Technology (IJISRT) ,www.ijisrt.com , ISSN - 2456-2165, PP:-718-721.
- [15] Susanto, S., & Suryani, D. (2010). *Pengantar Data Mining*. (Nikodemus WK, Ed.) (1st ed.). YOGYAKARTA: CV ANDI OFFSET.
- [16] Suwardika, G. (2017). Pengelompokan Dan Klasifikasi Pada Data Hepatitis Dengan Menggunakan Support Vector Machine (SVM), Classification And Regression Tree (Cart) Dan Regresi Logistik Biner, *J*, 183–191.