
PENERAPAN METODE ALGORITMA ID3 UNTUK PREDIKSI DIAGNOSA GAGAL GINJAL KRONIS (Studi Kasus: RSUD Sekarwangi Sukabumi)

¹Rusda Wajhillah, ²Nurul Diah Komala, ³Jamal Maulana

^{1,2} Universitas Bina Sarana Informatika

³ STMIK Nusa Mandiri

^{1,2} Jl. Cemerlang No.8 Sukakarya Warudoyong, Indonesia, Telp 0266621992

³ Jl. Damai No. 8, Warungjati Barat, Pasar Minggu, Jakarta Selatan

rusda.rwh@bsi.ac.id, dyahnurul23@gmail.com, jamal.jml@nusamandiri.ac.id

Abstract

The Iterative Dichotomiser 3 (ID3) algorithm is part of the classification method and is a type of method that can map or separate two or more different classes. One of the problems that can be solved using algorithm ID3 is the prediction of the diagnosis of chronic kidney disease. Chronic kidney disease is a failure of kidney function to maintain metabolism and fluid and electrolyte balance. Based on the classification performance measurement from 230 data training shows that the accuracy value reaches 96,08%. It can be concluded that the method of ID3 Algorithm is feasible to be used in predictive data on chronic kidney disease (CKD).

Keywords: ID3 Algoritim, Chronic Kidney Disease (CKD), Accuracy

Abstrak

Algoritma Iterative Dichotomiser 3 (ID3) adalah bagian dari metode klasifikasi dan merupakan jenis metode yang dapat memetakan atau memisahkan dua atau lebih class yang berbeda. Salah satu masalah yang dapat dipecahkan dengan menggunakan algoritma ID3 adalah prediksi diagnosa gagal ginjal kronis. Gagal ginjal kronis adalah kegagalan fungsi ginjal untuk mempertahankan metabolisme serta keseimbangan cairan dan elektrolit. Berdasarkan pengukuran kinerja klasifikasi dari data training 230 data menunjukkan bahwa nilai akurasi mencapai 96,08%. Dapat disimpulkan bahwa metode Algoritma ID3 layak untuk digunakan pada penelitian prediksi data gagal ginjal kronis (GGK)

Kata Kunci: Algoritma ID3, Gagal Ginjal Kronis (GGK), Akurasi

1. PENDAHULUAN

Ginjal merupakan organ penting yang berfungsi menjaga komposisi darah dengan mencegah menumpuknya limbah dan mengendalikan keseimbangan cairan dalam tubuh, menjaga level elektrolit seperti sodium, potasium dan fosfat tetap stabil [2]. Gagal ginjal kronis (GGK) adalah kegagalan fungsi ginjal untuk mempertahankan metabolisme serta keseimbangan cairan dan elektrolit akibat destruksi struktur ginjal yang progresif dengan manifestasi penumpukkan sisa metabolik (toksik uremik) di dalam darah [8].

Klasifikasi gagal ginjal kronis dibagi tiga stadium yaitu, stadium satu penurunan cadangan ginjal, pada stadium kadar kreatinin serum normal dan penderita asimtomatik, stadium dua insufisiensi ginjal, dimana lebih dari 75% jaringan telah rusak, dan kreatinin serum meningkat, stadium tiga gagal ginjal stadium akhir atau uremia (Firman, Mugianti, Sunarno, & Winarni, 2016:119). Tiga tindakan medis yang dilakukan dalam menangani penyakit gagal ginjal yaitu hemodialisis (cuci darah/HD), Peritoneal dialis dan Transplantasi ginjal (cangkok ginjal) [8].

Penyakit Ginjal Kronis di dunia saat ini mengalami peningkatan dan menjadi masalah kesehatan serius, hasil penelitian Global Burden of Disease tahun 2010 Penyakit Ginjal Kronis merupakan penyebab kematian peringkat ke 27 di dunia tahun 1990 dan meningkat menjadi urutan ke 18 pada tahun 2010 [6].

Hasil Riskesdas 2013 menunjukkan prevalensi meningkat seiring dengan bertambahnya umur. Dengan peningkatan tajam pada kelompok umur 34-44 tahun dibandingkan kelompok umur 25-34 tahun. Provinsi dengan prevalensi tertinggi adalah Sulawesi Tengah 0.5% diikuti Aceh, Gorontalo dan Sulawesi Utara masing-masing 0.4% [2].

Induksi Decision Tree (ID3) adalah bagian dari metode klasifikasi dan merupakan jenis metode yang dapat memetakan atau memisahkan dua atau lebih class yang berbeda (Kacung & Santoso, 2017:212). Metode Algoritma ID3 digunakan pada penelitian Diagnosa Penyakit Demam Berdarah dengan nilai akurasi 75,253% dan pada penelitian lain tahun 2017 tentang Minat Konsumen Dalam Membeli Mobil dihasilkan nilai akurasi 90% [1].

Pada penelitian ini akan dilakukan analisa tentang prediksi diagnosa gagal ginjal kronis pasien pada RSUD Sekarwangi dengan menggunakan metode klasifikasi data mining algoritma ID3.

2. METODOLOGI PENELITIAN

2.1. Teknik Pengumpulan Data

Metode pengumpulan data dengan melakukan pengamatan dan penelitian secara langsung kepada perawat RSUD Sekarwangi yang beralamat di Jl.Raya Siliwangi No.49 Cibadak Sukabumi. Wawancara dilakukan tanya jawab kepada narasumber, dalam hal ini yaitu perawat dan dokter penyakit dalam RSUD Sekarwangi. Studi pustaka dilakukan dengan mengumpulkan data dan teori melalui buku-buku yang ada di perpustakaan dan media cetak lainnya yang relevan dan berkaitan dengan penelitian Penerapan Metode Algoritma ID3 untuk Prediksi Diagnosa Gagal Ginjal Kronis Studi Kasus RSUD Sekarwangi Sukabumi.

2.2. Tahap Penelitian

Analisa data dilakukan melalui data rekap RSUD Sekarwangi dengan nilai rata-rata pasien positif GGK dan negatif GGK, data diolah dan diuji dalam pengujian pada Algoritma ID3. Selanjutnya dibuat aplikasi guna untuk evaluasi yang menghasilkan pengetahuan (*knowledge*) baru.

Pengujian data penelitian ini menggunakan model eksperimen *Cross Industry Standard Process for Data Mining* (CRISP-DM) yang terdiri dari enam tahapan, sebagai berikut [9]):

Tahap Pemahaman Bisnis (*Business Understanding*); pada tahap ini difokuskan dalam tujuan penelitian, menganalisa pasien penderita GJK diketahui secara dini agar dapat mengurangi jumlah resiko penderita GJK, maka hal ini dapat dibantu dengan menggunakan teknik klasifikasi *data mining* algoritma ID3.

Tahap Pemahaman Data (*Data Understanding*); data yang digunakan dalam penelitian ini merupakan data sekunder yang didapat dari RSUD Sekarwangi di Kabupaten Sukabumi didalam data tersebut dapat diketahui status pasien yang positif GJK dan negatif GJK yang terdiri dari 9 atribut *predictor* dan 1 atribut hasil.

Tahap Data Persiapan (*Preparation*); selanjutnya adalah penyeleksian atribut, melakukan pembersihan, transformasi, serta normalisasi data. Jumlah data yang diperoleh pada penelitian ini sebanyak 400 *record*, terdiri dari pasien positif GJK dan pasien negatif GJK, akan tetapi data tersebut masih mengandung duplikasi dan anomali atau inkonsisten data. Setelah dilakukan *preprocessing* data yang didapat dari data sekunder RSUD Sekarwangi sebanyak 400 *record* direduksi dengan menghilangkan duplikasi menjadi 328 *record*, 199 *record* positif GJK dan 129 negatif GJK, setelah itu data akan dibagi menjadi dua bagian yaitu 70% data untuk proses *training* dan 30% data untuk proses *testing*. Hasil dari perhitungan diatas didapat 230 untuk data *training* dan 98 untuk data *testing*.

Tahap Pemodelan (*Modeling*); tahap ini dilakukan setelah melakukan *data preparation*, dalam penelitian ini data dianalisa dengan menggunakan teknik data mining metode Algoritma ID3.

Tahap Evaluasi (*Evaluation*); setelah proses pencarian sistem rekomendasi dilakukan, hasil tersebut akan dievaluasi kembali dan digunakan sebagai bahan penunjang dalam pengambilan keputusan.

Tahap *Deployment*; pada tahap ini dilakukan proses pembuatan aplikasi berbasis desktop dari hasil penelitian dan penerapan metode Algoritma ID3 pada rumah sakit yang relevan untuk mendiagnosa atau memprediksi penyakit GJK dengan menggunakan data baru.

2.3 Instrument Penelitian

Penelitian ini menggunakan data sekunder sebanyak 400 *record*, berupa data pasien positif GJK sebanyak 253 dan pasien negatif GJK 147 diambil dari tahun 2015-2018 yang akan digunakan sebagai instrumen guna memperoleh data dalam proses mendiagnosa penyakit GJK. Perangkat lunak yang digunakan untuk menganalisis adalah *Rapid miner* dan untuk menguji *rule* algoritma menggunakan *Microsoft Visual Basic 2010*.

Terdapat 10 (sepuluh) atribut awal yang terdapat pada data pasien RSUD Sekarwangi di Kabupaten Sukabumi sebagai berikut; usia, tekanan darah, urine dipstick, serum creatinine, hemoglobin, hipertensi, diabetes mellitus, nafsu makan, anemia dan kelas. Atribut-atribut tersebut umumnya bertipe data *text* dan *binominal*.

2.4 Populasi dan Sample Penelitian

Metode pengumpulan data menggunakan data sekunder, data pasien GGK yang sudah direkap sebelumnya oleh perawat RSUD Sekarwangi di Kabupaten Sukabumi. Populasi dalam penelitian ini merupakan pasien GGK positif dan pasien GGK negatif tahun 2015-2018 yang berasal dari data internal RSUD Sekarwangi di Kabupaten Sukabumi.

Populasi data awal yang digunakan pada penelitian ini berjumlah 400 *instance* atau *record*. Tidak seluruh populasi dalam data dapat digunakan pengujian metode. Maka masih diperlukan tahap analisis data agar dapat diperoleh populasi data yang sesuai dengan kebutuhan perhitungan metode Algoritma ID3.

3. HASIL DAN PEMBAHASAN

3.1. Algoritma Iterative Dichotomiser 3 (ID3)

Langkah pertama sebelum melakukan pengolahan data adalah hitung jumlah *class* yang positif GGK, negatif GGK dan nilai *entropy* dari masing-masing *class* berdasarkan atribut yang ditentukan, serta bagi data menjadi data *training* dan data *testing*. Dalam penelitian ini, data dipartisi sebesar 70% untuk sampel *training* (pelatihan) atau sebanyak 230 data dan 30% untuk sampel *testing* (pengujian) atau sebanyak 98 data.

Dari data *training* diketahui jumlah kasusnya ada 230 *record*, pasien yang menderita positif GGK berjumlah 146 *record* dan negatif GGK sebanyak 84 *record*, sehingga nilai entropinya yaitu:

- a. Menghitung *entropy* S merupakan himpunan dari kelas positif dan negatif dengan rumus persamaan 1.

$$Entropy (S) = -\sum_i^n p_i \log_2 p_i$$

$$Entropy S_{(positif, negatif)} = -\left(\frac{146}{230}\right) \cdot \log_2\left(\frac{146}{230}\right) - \left(\frac{84}{230}\right) \cdot \log_2\left(\frac{84}{230}\right) = 0,946928968 \text{ bits}$$

- b. Menghitung frekuensi masing-masing kategori dengan rumus yang sama, sebagai contoh perhitungan atribut usia berdasarkan kelasnya.

Tabel 1 Frekuensi atribut usia sesuai kelasnya

Kategori Usia	Frekuensi		Total
	Positif	Negatif	
0	7	13	20
1	47	36	83
2	76	30	106
3	16	5	21
Total	146	84	230

- c. Menghitung nilai *entropy* pada atribut usia.

$$1) Entropy S_{(0)} = -\left(\frac{7}{20}\right) \cdot \log_2\left(\frac{7}{20}\right) - \left(\frac{13}{20}\right) \cdot \log_2\left(\frac{13}{20}\right) = 0,934068055$$

$$2) Entropy S_{(1)} = -\left(\frac{47}{83}\right) \cdot \log_2\left(\frac{47}{83}\right) - \left(\frac{36}{83}\right) \cdot \log_2\left(\frac{36}{83}\right) = 0,987292732$$

$$3) Entropy S_{(2)} = -\left(\frac{76}{106}\right) \cdot \log_2\left(\frac{76}{106}\right) - \left(\frac{30}{106}\right) \cdot \log_2\left(\frac{30}{106}\right) = 0,859531691$$

$$4) Entropy S_{(3)} = -\left(\frac{16}{21}\right) \cdot \log_2\left(\frac{16}{21}\right) - \left(\frac{5}{21}\right) \cdot \log_2\left(\frac{5}{21}\right) = 0,791858353$$

d. Menghitung nilai *information gain* pada atribut usia dengan rumus persamaan 2.

$$\begin{aligned} Gain_{(usia)} &= Entropy(S) - \sum_{v \in \{0,1,2,3\}} \frac{S_v}{S} Entropy(S_v) \\ &= Entropy(S) - \frac{20}{230} Entropy(S_0) - \frac{83}{230} Entropy(S_1) - \frac{106}{230} Entropy(S_2) - \frac{21}{230} Entropy(S_3) \\ &= 0,946928968 - \left(\frac{20}{230} \cdot 0,934068055\right) - \left(\frac{83}{230} \cdot 0,987292732\right) - \left(\frac{106}{230} \cdot 0,859531691\right) - \left(\frac{21}{230} \cdot 0,791858353\right) = \\ &0,040989653 \end{aligned}$$

Berikut ini adalah hasil perhitungan mencari nilai *entropy* dan *information gain* dari semua atribut untuk menentukan pemilah terbaik.

Tabel 2 Perhitungan Gain Keseluruhan Atribut

Node	Jumlah Kasus (J)	Negatif (S1)	Positif (S2)	Entropy (S)	Gain
Total	230	84	146	0,946928968	
Usia					0,040989653
0	20	13	7	0,934068055	
1	83	36	47	0,987292732	
2	106	30	76	0,859531691	
3	21	5	16	0,791858353	
Tekanan Darah					0,206786005
0	0	0	0	0	
1	49	0	49	0	
2	2	0	2	0	
3	141	57	84	0,973385435	
4	38	27	11	0,868040399	
Urine Dipstick					0,18285244
0	176	84	92	0,998509099	
1	47	0	47	0	
2	5	0	5	0	
3	2	0	2	0	
Serum Creatinine					0,002858594
0	229	84	145	0,948192952	
1	1	0	1	0	
Hemoglobin					0,360892176
0	52	47	5	0,456683632	
1	128	37	91	0,867507134	
2	50	0	50	0	
Hipertensi					0,423623476
0	101	0	101	0	
1	129	84	45	0,933025295	
Diabetes Mellitus					0,261958769
0	72	0	72	0	
1	158	84	74	0,997108517	
Nafsu Makan					0,143705096
0	186	84	102	0,99323382	

R6: If hipertensi = 0 and and hemoglobin = 1 and diabetes mellitus = 0 and tekanan darah = 3 and urine dipstic = 1 then positif.

R7: If hipertensi = 0 and and hemoglobin = 1 and diabetes mellitus = 0 and tekanan darah = 3 and urine dipstic = 0 and usia = 3 then negatif.

R8: If hipertensi = 0 and and hemoglobin = 1 and diabetes mellitus = 0 and tekanan darah = 3 and urine dipstic = 0 and usia = 2 and serum creatinine = 0 and nafsu makan = 0 and anemia = 0 then negatif.

R9: If hipertensi = 0 and and hemoglobin = 1 and diabetes mellitus = 0 and tekanan darah = 3 and urine dipstic = 0 and usia = 1 and nafsu makan = 1 then positif.

R10: If hipertensi = 0 and and hemoglobin = 1 and diabetes mellitus = 0 and tekanan darah = 3 and urine dipstic = 0 and usia = 1 and nafsu makan = 0 and anemia = 1 then positif.

R11: If hipertensi = 0 and and hemoglobin = 1 and diabetes mellitus = 0 and tekanan darah = 3 and urine dipstic = 0 and usia = 1 and nafsu makan = 0 and anemia = 0 and serum creatinine = 0 then negatif.

R12: If hipertensi = 0 and and hemoglobin = 1 and diabetes mellitus = 0 and tekanan darah = 3 and urine dipstic = 0 and usia = 0 and serum creatinine = 0 and nafsu makan = 0 and anemia = 0 then negatif.

R13: If hipertensi = 0 and and hemoglobin = 1 and diabetes mellitus = 0 and tekanan darah = 1 then positif.

R14: If hipertensi = 0 and and hemoglobin = 0 and tekanan darah = 4 then negatif.

R15: If hipertensi = 0 and and hemoglobin = 0 and tekanan darah = 3 and diabetes mellitus = 1 then positif.

R16: If hipertensi = 0 and and hemoglobin = 0 and tekanan darah = 3 and diabetes mellitus = 0 and usia = 3 then negatif.

R17: If hipertensi = 0 and and hemoglobin = 0 and tekanan darah = 3 and diabetes mellitus = 0 and usia = 2 and urine dipstic = 0 and serum creatinine = 0 and nafsu makan = 0 and anemia = 0 then negatif.

R18: If hipertensi = 0 and and hemoglobin = 0 and tekanan darah = 3 and diabetes mellitus = 0 and usia = 1 then negatif.

R19: If hipertensi = 0 and hemoglobin = 0 and tekanan darah = 3 and diabetes mellitus = 0 and usia = 0 then negatif.

R20: If hipertensi = 0 and and hemoglobin = 0 and tekanan darah = 1 then positif.

3.4. Desain User Interface

Dalam penerapan pada GUI yang dibuat, data yang digunakan sebanyak 32 *record* yang diambil pada 98 data *testing* secara *random*. Hasil yang diperoleh yaitu, 17 data positif GGK dan 15 data negatif GGK. Gambar 4 menggambarkan bentuk GUI untuk memprediksi GGK sebagai berikut:



Gambar 4. User Interface Sistem Prediksi GGK

3.5. Testing

Setelah didapatkan secara utuh hasil klasifikasi Algoritma ID3 berupa pohon keputusan, langkah selanjutnya adalah mengukur ketepatan hasil klasifikasi yang terbentuk. Ketepatan klasifikasi maupun kesalahan klasifikasi dirangkum dalam tabel matriks konfusi sebagai berikut:

Tabel 3 Hasil Matriks Konfusi Algoritma ID3 Menggunakan Data *Training*

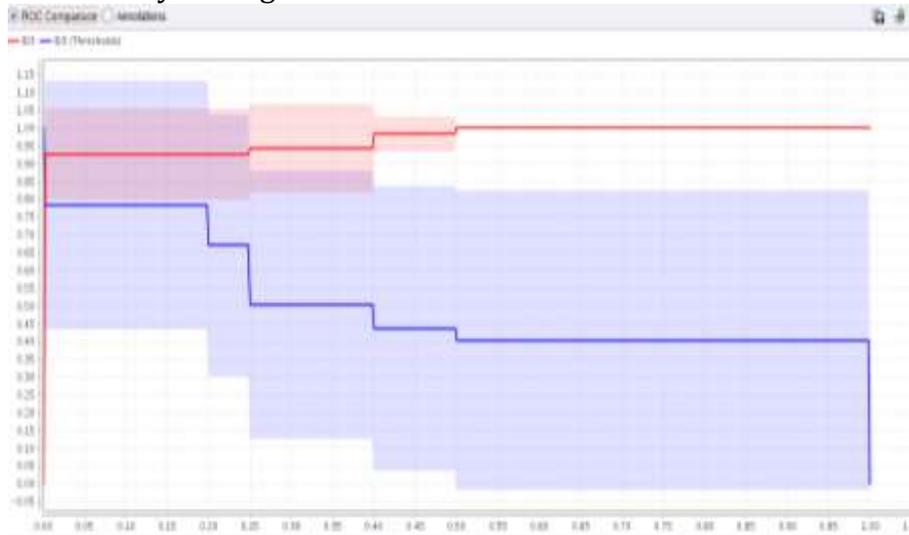
	Prediksi (Positif)	Prediksi (Negatif)	Total
(Positif)	137	9	146
(Negatif)	0	84	84
Total	137	93	230

Dapat dilihat pada tabel 3 bahwa sebanyak 137 kasus dengan status positif GGK dan sebanyak 84 kasus dengan status negatif GGK diklasifikasikan secara tepat. Kemudian sebanyak 9 kasus dengan status positif GGK diklasifikasikan kedalam status negatif GGK, sehingga hal ini disebut kesalahan klasifikasi dan 0 kasus dengan status negatif GGK diklasifikasikan kedalam status positif GGK maka hal ini juga disebut kesalahan klasifikasi.

Akurasi atau persentasi dari keseluruhan kasus yang diklasifikasikan secara tepat pada konstruksi pohon ini dihitung dengan persamaan 3 adalah:

$$\begin{aligned}
 \text{Akurasi} &= \frac{137+84}{230} \times 100\% \\
 &= \frac{221}{230} \times 100\% \\
 &= 96,08695652\% \\
 &= 96,08\%
 \end{aligned}$$

Model kurva ROC untuk Algoritma ID3 dengan menggunakan data *testing* bentuk dari kurvanya sebagai berikut:



Gambar 3 Kurva ROC Diagnosa GGK

4. SIMPULAN

Dari hasil penelitian dapat diambil kesimpulan bahwa model prediksi yang diperoleh dari perhitungan *information gain* dengan metode Algoritma ID3 ke prediksi data GGK berhasil, atribut-atribut yang dibentuk adalah usia, tekanan darah, urine dipstick, serum creatinine, hemoglobin, hipertensi, diabetes mellitus, nafsu makan dan anemia berpengaruh terhadap pohon keputusan dan tidak ada atribut yang tidak terpakai.

Berdasarkan pengukuran kinerja klasifikasi dari data *training* menunjukkan bahwa nilai akurasi klasifikasi mencapai 96,08%, sehingga metode Algoritma ID3 layak untuk digunakan pada penelitian prediksi data GGK.

Agar penelitian ini bisa ditingkatkan, ada beberapa saran yang diusulkan:

- Menambahkan jumlah data yang lebih besar dan atribut yang lebih banyak, sehingga hasil pengukuran yang akan didapatkan lebih baik lagi.
- Penelitian ini dapat dikembangkan dengan membandingkan algoritma *data mining* lainnya misalkan saja dengan metode *Support Vector Machine*, *K-Nearest Neighbor*, *CART* dan lainnya atau dapat mengoptimalkan parameter dengan *Particle Swarm Optimization*, *Genetic Algorithm* dan lainnya.

DAFTAR PUSTAKA

- [1] Arietya, W. W., Supriyatin, W., & Astuti, I, "Penentuan Minat Konsumen Dalam Membeli Mobil Menggunakan Algoritma Id3 Studi Kasus Daihatsu Wilayah Jakarta", Jurnal KNTIA, 4, E-25, 2017.
- [2] Emojion, "Situasi Penyakit Ginjal Kronis", *Infodatin* (pp. 1-10), Pusat Data dan Informasi, Jakarta, 2017.

- [3] Firman, R., Mugiarti, S., Sunarno, I., & Winarni, S, “**Kualitas Hidup Pasien Gagal Ginjal Yang Menjalani Hemodialisis**”, *Jurnal Ners dan Kebidanan*, 3(2355-052X), 118-122, Agustus 2016.
- [4] Ibrahim, I., Suryani, I., & Ismail, E, “**Hubungan Asupan Protein dengan Kadar Ureum dan Kreatinin pada Pasien Gagal Ginjal Kronik yang Sedang Menjalani Hemodialisa di Unit Hemodialisa RS PKU Muhammadiyah Yogyakarta**”, *Jurnal Nutrisia*, 19(1693-945X), 1-6, Maret 2017.
- [5] Kacung, S., & Santoso, B., “**Sistem Deteksi Dini Untuk Meningkatkan Performance Kelulusan Mahasiswa Dengan Id3 (Studi Kasus : Teknik Informatika Unitomo)**”, *Prosiding Penelitian DIPA*, XXXVI(978-602-61886-0-1), 212-216, 2017.
- [6] Moeloek, N. F., “**Upaya Peningkatan Promotif Preventif Bagi Kesehatan Ginjal di Indonesia.**” In M. R. Kesehatan, *Air Bagi Kesehatan* (pp. 1-34). Jakarta: Germas, 2018.
- [7] Nafi'iyah, N., & Faticah, C. (2018, Mei). METODE FUZZY ID3 UNTUK KLASIFIKASI BENTUK WAJAH MANUSIA MENGGUNAKAN DENTAL PANORAMIC. *Jurnal SPIRIT*, 10(2085 - 3092), 13-20.
- [8] Rosmalia, L., & Kusumadewi, S, “**Sistem Pendukung Keputusan Klinis Untuk Menentukan Jenis Gangguan Psikologi Pada Pasien Gagal Ginjal Kronis (GGK) yang Menjalani Terapi Hemodialisa**”, *informatika upgris*, 4(2460-4801), 11-21, 2018.
- [9] Shafique, U., & Qaiser, “**A Comparative Study of Data Mining Process Models (KDD , CRISP-DM and SEMMA)**”, *International Journal of Innovation and Scientific Research*, 12, 217–222, 2014.