PREDIKSI KELULUSAN NILAI KALKULUS MENGGUNAKAN NAÏVE BAYES

Annas Fajriansyah, Dadang Yusup, Kamal Prihandini
Program Studi Informatika, Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang
1910631170162@Student.unsika.ac.id

ABSTRAK

Kalkulus adalah salah satu mata pelajaran dasar yang perlu dipelajari di program studi teknik informatika fakultas ilmu komputer. Bagi Sebagian mahasiswa yang khususnya di fakultas Teknik informatika, Kalkulus merupakan mata pelajaran yang dianggap cukup sulit yang mengakibatkan mereka harus mengulang mata pelajaran kalkulus. Padahal mata pelajaran ini penting bagi mereka untuk itu, pada penelitian ini memprediksi pembelajaran kalkulus dengan menerapkan proses data mining menggunakan metode Naïve Bayes (kernel) untuk memprediksinya. Proses penelitian ini menerapkan Cross Industry Metodologi Starndard Process for Data Mining (CRIPS-DM) dengan menggunakan algoritma Naïve Bayes(kernel). Penelitian ini yang menggunakan data dari mahasiswa Angkatan 2016, 2017, 2018, dan 2019. Penelitian ini mencari akurasi yang lebih baik dan melihat kelas mana saja yang sangat berpengaruh kepada lulus tidak nya mahasiswa pada matakuliah kalkulus pada pengujian akan membandingkan antara rasio 60:40, 70:30, dan 80:20. Hasil dari penelitian ini memprediksi kelulusan pada mata pelajaran kalkulus dengan menggunakan algoritma naïve bayes dengan nilai akurasi tertinggi yaitu 93.83% pada rasio 60:40.

Kata kunci : *Naïve bayes, Kalkulus, Prediksi*

1. PENDAHULUAN

Pendidikan tinggi memiliki peranan yang krusial dalam menentukan saling sumber daya manusia di suatu negara. Salah satu tantangan utama dalam pendidikan tinggi adalah menjamin bahwa setiap siswa dapat menyelesaikan studi mereka dengan hasil yang memuaskan. Mata kuliah Kalkulus, yang merupakan bagian dari program studi Informatika dan beberapa jurusan lainnya, sering kali dianggap sebagai salah satu mata kuliah yang sulit oleh banyak mahasiswa. Tingkat kesulitan yang tinggi ini sering kali berdampak pada tingkat kelulusan siswa dalam mata kuliah tersebut.

Keberhasilan siswa dalam mata kuliah Kalkulus dipengaruhi oleh berbagai faktor, tidak hanya kemampuan akademik. Faktor-faktor lain seperti kehadiran, partisipasi dalam tugas, dan kesiapan mental juga berperan penting. Oleh karena itu, sangat penting untuk memprediksi kelulusan mahasiswa dalam mata kuliah ini dengan menggunakan data historis yang relevan, agar institusi pendidikan dapat memberikan intervensi atau dukungan yang sesuai untuk meningkatkan hasil akademik mereka.

Penelitian ini dilakukan untuk memprediksi kelulusan nilai kalkulus dengan menggunakan algoritma naïve bayes, karena menggunakan algoritma naïve bayes bisa melihat kemungkinan pada setiap kelasnya. Penelitian ini dilakukan karena masih banyak yang mengagap matakuliah kalkulus ini sulit dan mendapatkan nilai yang kurang memuaskan sehingga mereka mengulang kembali. Maka pada penelitian ini akan dilakukan prediksi untuk mengetahui apa saja yang berpengaruh agar mendapat nilai yang baik untuk lulus dalam mata kuliah kalkulus.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data Mining merupakan suatu istilah yang digunakan untuk penguraian data yang di ubah menjadi kan sebuah informasi pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terikat dari berbagai database besar [10].

2.2. Algoritma Naïve Bayes (NB)

Naïve Bayes merupakan metode prediksi yang berbasis probabilitas dan sederhana, yang didasarkan pada standar teorema Bayes dengan asumsi independensi yang kuat [2].

Algoritma *Naïve Bayes* termasuk salah satu teknik pembelajaran induktif yang paling efisien dan efektif dalam konteks *machine learning* dan data mining. Meskipun mengandalkan asumsi atribut independensi, kinerja *Naïve Bayes* dalam proses klasifikasi tetap kompetitif. Atribut independensi asumsi ini jarang terjadi dalam data nyata, namun meskipun asumsi tersebut tidak terpenuhi, kinerja pengklasifikasian *Naïve Bayes* tetap tinggi, seperti yang dibuktikan oleh berbagai penelitian empiris.

2.3. *Confusion Matrix*

Confusion matrix adalah alat untuk mengukur bentuk matriks yang digunakan untuk memperoleh tingkat akurasi klasifikasi untuk kelas algoritma yang digunakan. Pada confusion matrix ini didalamnya berisi suatu informasi tentang kelas yang aktual dan kelas yang diprediksi dari proses klasifikasi.

Menurut [6] Validasi digunakan untuk menentukan jenis model terbaik melalui confusion

matriks sebagai informasi tentang hasil klasifikasi aktual yang dapat diprediksi oleh suatu sistem, yang diukur melalui nilai akurasi, presisi, dan recall dengan menggunakan persamaan.

Menurut [3] Akurasi merupakan rasio prediksi benar positif dan negatif, Untuk menghitung rumus dari tingkat akurasi pada matriks dapat dilakukan melalui 2.3 berikut (Lamasigi, 2021).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Dimana :

TP = Data yang diprediksi benar dan bersifat positif
TN = Data yang diprediksi benar dan bersifat negative
FP = Data yang positif yang diprediksi sebagai data positif
FN = Data positif yang diprediksi sebagai data negatif

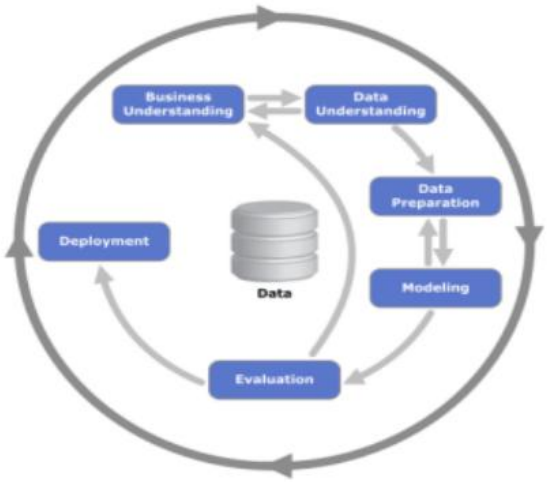
2.4. Rapidminer

Rapidminer adalah salah satu dari perangkat lunak yang digunakan untuk proses pengolahan pada penambangan data. Dalam penggunaan rapidminer text mining dilakukan berkisar untuk analisis teks, mengekstrasi suatu pola dari *big data* yang diperoleh, serta mengkombinasikan dengan berbagai metode seperti statistika, kecerdasan buatan, dan basis data [2].

Dengan menggunakan RapidMiner, para peneliti dan analis dapat memanfaatkan algoritma-algoritma machine learning yang canggih dan fitur - fitur analisis data untuk klasifikasikan dan mempredikasi sebuah data untuk mencari informasi. Penggunaan *software* RapidMiner untuk membangun model tanpa perlu membuat program karena semua alat pengolahan data yang dibutuhkan untuk mengolah data telah tersedia berbentuk operator. Data mining dapat digunakan dalam membantu proses pengambilan kebijakan. Salah satu model yang dapat digunakan adalah dengan melakukan prediksi dengan *regresi linear* [9].

3. METODE PENELITIAN

Tahapan dalam proses CRISP-DM menurut [5] meliputi tahapan berikut :



Gambar 1. CRISP-DM (*Cross Industry Standard Proses for Data Mining)*

CRISP-DM (*Cross Industry Standard Proses for Data Mining*) adalah metode yang di gunakan penulis dalam penelian disusun oleh konsorsium perusahaan yang didirikan oleh Komisi Eropa pada tahun 1996 dan telah ditetapkan sebagai proses standar dalam data mining[7].

Dalam tahapan business understanding merupakan tahap pemahaman substantif dari kegiatan penambangan data yang akan dilakukan termasuk kebutuhan akan penelitian. Untuk kegiatannya mencakup dari menentuan tujuan atau sasaran dari penelitian, pemahaman akan situasi, dan menerjemahkan tujuan dari penelitian ke dalam data mining

a. *Business Understanding*
   pada tahap ini dilakukan pemahaman substansi dari kegiatan data mining yang akan dilakukan serta kebutuhan dari perspektif bisnis.
b. *Data Understanding*
   Dalam tahapan pengumpulan data ini dilakukan kajian mempelajari terhadap data tersebut agar dapat dipahami dan dengan mudah dapat diproses ke tahap berikutnya. Kemudian data digunakan untuk mengindentifikasi dan mencari pengetahuan awal lebih lanjut.
c. *Data Preparation*
   Dalam tahapan data preparation dilakukan untuk menyiapkan struktur basis data yang dapat digunakan untuk proses data mining yaitu tahap pemodelan. Kegiatan yang dilakukan pada tahap ini mencakup proses data selection, proses data preprocessing, dan proses transformasi data apabila dibutuhkan.
d. *Modeling*
   Dalam tahapan ini melakukan penentuan teknik data mining yang akan digunakan, alat bantu untuk proses data mining, algoritma yang akan diterapkan, dan penentuan parameter dengan nilai optimal. Jika penyesuaian pada data diperlukan untuk proses data mining, maka yang harus dilakukan yaitu kembali ke tahap data *preparation* untuk memperbaikinya.
e. Evaluation
   Dalam tahapan *evaluation* ini melakukan suatu interpretasi hasil yang diperoleh dari proses data mining. Tahapan ini bertujuan agar dapat menyesuaikan model berdasarkan dengan tujuan yang diharapkan. Sehinnga dari hasil model tersebuat dapat diambil unuk melanjutkan ke tahap berikutnya yaitu *deployment*.
f. *Deployment*
   Tahap *deployment* merupakan tahapan dimana melakukan pembuatan laporan yang telah didapatkan berdasarkan dari hasil proses pengolahan data mining dengan mengenali pola pada data dalam proses data mining. Sehingga diperoleh pengetahuan dan informasi yang mudah untuk dimengerti.

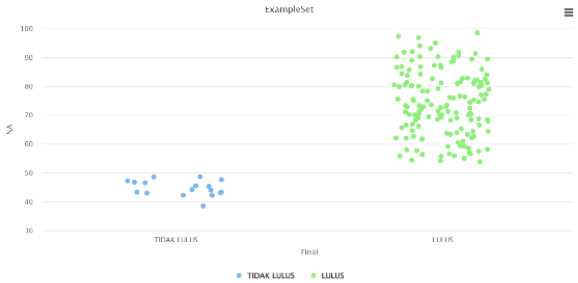
4. HASIL DAN PEMBAHASAN

Hasil Penelitian yang telas dilakukan yaitu Prediksi tinggkat kelulusan pada pembelajaran kalkulus pada mahasiswa Fakultas Ilmu Komputer pada Angkatan 2016-2019 yang memprediksi menjadi lulus dan tidak lulus yang menggunakan algoritma *Naïve Bayes* (NB). Tools yang digunakan untuk membantu proses pengolahan data yaitu rapidminer. *Naive Bayes* yang digunakan adalah Naïve Bayes yang digunakan untuk membuah sebuah diagram untuk menjadi diagram yang bisa menjadi gambaran untuk melihat seberapa besar tingkat kelulusan pada Laki – Laki dan Perempuan.

4.1. *Business Understanding*

Pada tahapan ini berfokus pada seperti yang dijelaskan pada pendahuluan penelitian ini, mahasiswa masih ada yang mengulang mata kuliah kalkulus setiap tahun. sehingga dapat mengetahui apa saya yang berpengaruh menjadi faktor yang mempengaruhi hasil belajar matakuliah kalkulus lulus dan tidak lulus nya mahasiswa. Penelitian ini memprediksi mengguanakan algoritma Naïve Bayes untuk mengetahui akurasi pada prediksi kelulusan matakuliah kalkulus.

Tahap awal yaitu dengan melihat mahasiswa yang mengulang dan tidak mengulang berdasarkan nilai akhirnya. plot dapat dilihat pada gambar 2.



Gambar 2. Plot

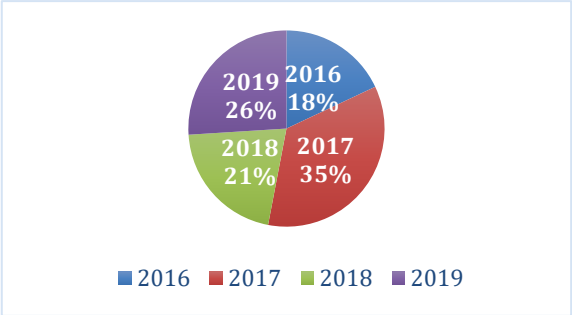
Dapat dilihat pada gambar 4.1 bahwa nilai akhir yang lebih dari 50 dikategorikan dengan pass sedangkan nilai akhir mahasiswa yang kurang dari 50 dikategorikan dengan fail.

4.2. *Data Understanding*

Data yang digunakan yaitu data yang diambil dari penelitian sebelumnya, penelitian sebelumnya mengambil dari nilai kalkulus yang dimiliki oleh dosen pengajar dan data penunjang yang diambil dari kuisioner yang telah disebarkan. Data kuisioner kemudian dikumpulkan dan diolah yang selanjutnya dataset dipakai untuk penelitian.

4.3. *Data Preparation*

Sebanyak 252 mahasiswa Program studi Teknik Informatika mengisi kuisioner yang telah disebarkan. Selanjutnya data dibersihkan dan menghasilkan data bersih sebanyak 204 mahasiswa, dari angkatan 2016 – 2019. Selanjutnya data yang sudah bersih diintregasi dengan data nilai menggunakan bantuan excel yang nantinya dataset dipakai untuk proses modeling. Berikut merupakan diagram dari jumlah mahasiswa setiap angkatan terhadap keseluruhan data kuisioner yang di lakukan pada penelitian sebelumnya yang ditampilkan pada gambar 3.



Gambar 3. Diagram data mahasiswa

Setelah melakukan pengecekan data diklasifikasikan menjadi 2 yaitu, Lulus dan Tidak Lulus yang ditampilkan pada table 2.

Tabel 2. Jumlah lulus dan tidak lulus

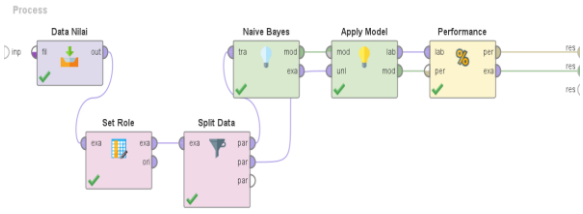
LULUS	TIDAK LULUS
183	21

Seperti yang bisa dilihat pada table 4. Sebanyak 204 mahasiswa yang mengikuti pembelajaran mata kuliah kalkulus dan terdapat 183 mahasiswa dinyatakan lulus dan 21 dinyatakan Tidak Lulus.

4.4. Modelling

Pada tahapan ini menerapkan Teknik pemodelan data mining yang telah ditentukan sebelumnya dan berdasarkan tujuan yang ingin dicapai. Pada tahap modelling yang harus dilakukan yaitu menentukan dan memilih Teknik pemodelan yang akan digunakan. Teknik pemodelan yang dipilih untuk digunakan yaitu prediksi dengan menggunakan Naïve bayes.

Setiap tahap menentukan Teknik pemodelan, tahp berikutnya adalah menerapkan Teknik pemodelan berdasarkan yang telah ditentukan sebelumya. Dalam membangun sebuah model pada rapidminer menggunakan algoritma Naïve bayes menggunakan tools yang ada di rapidminer yaitu Naïve bayes. Yang dimana data pembelajaran nilai kalkulus akan di uji menggunakan algoritman naïve bayes. Berikut adalah gambar model pada aplikasi rapidminer.



Gambar 4. Modelling

Pada Gambar 4 di atas adalah model pengujian yang akan diterapkan, yang mana nanti akan digunakan dalam melalukan 3 pengujian dengan data yang dibagi menjadi 60:40, 70:30, 80:20 masing masing akan dicoba menggunakan algoritman naïve bayes. Pada pebagian testing terdapat dua fitur apply model dan Performance. Fitur apply model digunkan untuk menerapkan model yang telah dilatih sebelumnya dan di uji coba dengan data test. Sedangkan pada fitur performance digunakan untuk mengevaluasi hasil kerja dari algoritma naïve bayes yang menghasilkan berupa parameter pengukur hasil dengan hasil akurasi.

Pada penelitian ini model bisa diliat pada gambar 4 terdapat set role yang dilakukan membuat set role pada atribut Final yang terbagi menjadi lulus dan tidak untuk menjadi acuan. Kemudian pembagian data menjadi data training dan data uji, setelah data sudah di bagi maka untuk data traning masuk kepada naïve bayes untuk pelajari setelah dipelajrai akan masuk ke tools apply model sedangkan untuk data uji akan langsung masuk ke tools apply model untuk pengujian dan yang terakhir tools performance itu digunakan itu mengukur sebuah akurasi atau performa algoritma naïve bayes pada data ini. Hasilnya seperti berikut :

a. Skenario 1

accuracy: 93.83%

	true TIDAK LULUS	true LULUS	class precision
pred. TIDAK LULUS	5	2	71.43%
pred. LULUS	3	71	95.95%
class recall	62.50%	97.26%	

Gambar 5. Hasil pengujian 60:40

Pada Skenario 1 itu dengan menggunakan rasio perbandingan *data traning* 60% dan *data testing* 40%. Hasil dari pengujian data tersebut menggunakan algoritma naïve bayes menghasilkan conffution matrix dengan record 76 dari 81 maka dihasilakan akurasi sebesar 93.83%.

b. Skenario 2

accuracy: 93.44%

	true TIDAK LULUS	true LULUS	class precision
pred. TIDAK LULUS	4	2	66.67%
pred. LULUS	2	53	96.36%
class recall	66.67%	96.36%	

Gambar 6. Hasil pengujian 70:30

Pada Skenario 2 itu dengan menggunakan rasio perbandingan *data traning* 70% dan *data testing* 30%. Hasil dari pengujian data tersebut menggunakan algoritma naïve bayes menghasilkan conffution matrix dengan record 57 dari 61 maka dihasilakan akurasi sebesar 93.44%.

c. Skenario 3

accuracy: 92.68%

	true TIDAK LULUS	true LULUS	class precision
pred. TIDAK LULUS	2	1	66.67%
pred. LULUS	2	36	94.74%
class recall	50.00%	97.30%	

Gambar 5. Hasil pengujian 80:20

Pada Skenario 1 itu dengan menggunakan rasio perbandingan *data traning* 80% dan *data testing* 20%. Hasil dari pengujian data tersebut menggunakan algoritma naïve bayes menghasilkan conffution matrix dengan record 38 dari 41 maka dihasilakan akurasi sebesar 92.68%.

4.5. *Evaluation*

Hasil yang diperoleh dari semua pengujian yang telah dilakukan menunjukan bahwa pembagian data traning dan data testing pada algoritma naïve bayes memiliki pengaruh terhadap hasil yang di inginkan. Pengujian dengan menerapkan perubahan pada partisi data traning dan data testing yang berbeda untuk menguji seberapa efektif performa yang dihasilkan Naïve bayes. Berikut adalah hasil prediksi dari algoritma naïve bayes mengenai nilai akurasi pada table 4.

Tabel 4. Hasil akuasi

Rasio	Akurasi
60:40	93.83%
70:30	93.44%
80:20	92.68%

Bisa dilihat pada gambar penelitian ini menunjukan bawa dari 3 rasio yang berbeda yang paling bagus untuk sebuah nilai akurasi nya yaitu rasio 80 : 20 yang menghasilkan 95.12%. Dengan demikian, maka dapat dilanjutkan ke tahap deployment.

4.6. *Deployment*

Penelitian ini melakukan prediksi Nilai pembelajaran kalkulus pada mahasiswa yang dikelompokan lulus dan tidak lulus dengan menerapkan metode algoritma naïve bayes. Adapun metodologi yang digunakan yaitu Cross-Industry Standart Process for data mining (CRISP-DM) yang tahapannya mencakup business understanding, data understanding, data preparation, modelling, evaluation, dan deployment. Serta dalam proses pengolahan data ang dilakukan menggunakan rapidminer.

Pada penelitian ini juga data dipecah menjadi 3 dipartisi, yaitu Partisi 60% *data traning* : 40% *data testing*, 70% *data traning* : 30% *data testing*, 80% *data traning* : 20% *data testing*. Setiap partisi semua di uji coba dalam mengevaluasi model performa dari penerapan algoritma *naïve bayes* dengan nilai akurasi. Hasil dari 3 pengujian terdapat nilai akurasi teringgi yaitu 93.83%.

5. KESIMPULAN DAN SARAN

Algoritma Naïve bayes mampu melakukan prediksi dari hasil pembelajaran kalkulus. Proses prediksi berdasarkan dengan menggunakan algoritma Naïve Bayes. Dengan menggunakan 13 atribut. Confusion Matrix adalah salah satu metode yang digunakan untuk mengukur performa hasil belajar kalkulus. Dengan menggunakan 3 kali percobaan, diperoleh hasil dari 60 : 40 mendapatkan hasil akurasi

93.83%, 70 : 30 mendapatkan hasil akurasi sebesar 93.44%, dan 80 : 20 mendapatkan hasil akurasi sebesar 92.68%. Maka hasil akurasi yang didapat tertinggi dimiliki oleh percobaan ke tiga dengan rasio 60:40 yang menghasilkan 93.83%. Diharapkan penelitian selanjutnya perlu diperhatikan mengenai kuatitas data yang dijadikan sebagai data latih dan penelitian selanjutnya menggunakan algoritma selanjutnya dan menggunakan data lain.

DAFTAR PUSTAKA

[1] Cahya, D., & Buani, P. (2021). Penerapan algoritma naïve bayes dengan seleksi fitur algoritma genetika untuk prediksi gagal jantung. *Jurnal Sains Dan Manajemen*, *9*(2). https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/11141.

[2] Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi. *Teknika*, *8*(1), 11–16. https://doi.org/10.34148/teknika.v8i1.95.

[3] Indahyanti, U., Azizah, N. L., & Setiawan, H. (2022). Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa. *Jurnal Sains Dan Informatika*, *8*(2). https://doi.org/10.34128/jsi.v8i2.459

[4] Lamasigi, Z. Y. (2021). DCT Untuk Ekstraksi Fitur Berbasis GLCM Pada Identifikasi Batik Menggunakan K-NN. *Jambura Journal of Electrical and Electronics Engineering*, *3*. https://ejurnal.ung.ac.id/index.php/jjeee/article/view/7113

[5] Nafisah Giustin, F., Nurina Sari, B., & Nur Padilah, T. (n.d.). APPLICATION OF C5.0 ALGORITHM IN PREDICTION OF LEARNING OUTCOMES IN CALCULUS SUBJECT. In *Journal of Applied Engineering and Technological Science* (Vol. 3, Issue 2).

[6] Nur, A., Rohim, A., Purnamasari, A. I., & Ali, I. (2024). KOMPARASI EFEKTIFITAS ALGORITMA C4.5 DAN NAÏVE BAYES UNTUK MENENTUKAN KELAYAKAN PENERIMA MANFAAT PROGRAM KELUARGA HARAPAN (STUDI KASUS : KECAMATAN CICALENGKA

[7] Ordila, R., Wahyuni, R., Irawan, Y., & Yulia Sari, M. (2020). Penerapan data mining untuk pengelompokan data rekam medis pasien berdasarkan jenis penyakit dengan algoritma clustering (Studi Kasus : Poli Klinik PT.Inecda). *Jurnal Ilmu Komputer*, *9*(2), 148–153. https://doi.org/10.33060/jik/2020/vol9.iss2.181

[8] Saragih, H., Buulolo, E., & Tinus Waruwu, F. (2017). Implementasi data mining penyesuaian jenis lensa terhadap kebutuhan pasien dengan menggunakan algoritma C4.5. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, *1*. http://ejurnal.stmik-budidarma.ac.id/index.php/komik

[9] Sholeh, M., Kumalasari Nurnawati, E., & Lestari, U. (2023). Penerapan data mining dengan metode regresi linear untuk memprediksi data nilai hasil ujian menggunakan rapidminer. *Jurnal Informatika Sunan Kalijaga)*, *8*(1), 10–21. https://archive.ics.uci.edu/ml/datasheets.php.

[10] Utomo, D. P., & Mesran, M. (2020). Analisis komparasi metode klasifikasi data mining dan reduksi atribut pada data set penyakit jantung. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *4*(2), 437. https://doi.org/10.30865/mib.v4i2.2080