

## IMPLEMENTASI DATA MINING PREDIKSI SISWA DALAM MEMPEROLEH BANTUAN DANA PENDIDIKAN MENGGUNAKAN ALGORITMA KLASIFIKASI DANA DAN NAÏVE BAYES

Dani

<sup>1</sup>Program Studi Teknik Informatika, Universitas Pamulang  
e-mail: <sup>1</sup>dosen02510@unpam.ac.id

### Abstract

*In educational institutions, students' grades can help the institution as a consideration in providing scholarships. The importance of data mining in the world of education can facilitate the decision-making process. Retrieving very large amounts of data will certainly pose a very large risk, so data mining that combines several of these methods can provide maximum results. Education is one component of life that can support a person's success towards a much better life. Providing this scholarship is very important, in order to give them peace of mind in participating in teaching and learning. However, sometimes scholarships are not given on target. Based on this problem, research was carried out regarding the selection of scholarship recipients to help increase the speed and accuracy in determining scholarships using the C4.5 and Naïve Bayes algorithm methods with the aim of measuring the level of accuracy and length of time for the process of determining scholarship recipients. In predicting After testing, the results obtained from the C4.5 algorithm have an accuracy of 54%. Better than the accuracy level of the Naïve Bayes algorithm of 93.33%, and calculations were carried out using Naïve Bayes. It is hoped that this research can help PERMATASARI 1 Islamic Vocational School in providing scholarships in determining the preferred criteria for providing scholarships, so that scholarship awards can be right on target.*

### Abstrak

Dalam institusi Pendidikan nilai siswa dapat membantu institusi sebagai bahan pertimbangan dalam memberikan beasiswa. Pentingnya data Mining dalam dunia pendidikan dapat memudahkan dalam proses pengambilan keputusan. Pengambilan data dalam jumlah yang sangat besar pastinya akan memberikan risiko yang sangat besar pula, sehingga data Mining yang menggabungkan beberapa metode ini dapat memberikan hasil secara maksimal. Pendidikan merupakan salah satu komponen kehidupan yang dapat menunjang keberhasilan seseorang menuju kehidupan yang jauh lebih baik lagi. Pemberian beasiswa ini sangat penting, guna memberikan ketenangan mereka dalam mengikuti belajar-mengajar. Namun terkadang pemberian beasiswa atau tidak tepat pada sasarannya. Dari permasalahan tersebut dilakukan penelitian mengenai seleksi penerimaan beasiswa agar membantu meningkatkan kecepatan dan ketepatan dalam penentuan beasiswa dengan menggunakan metode algoritma C4.5 dan naïve bayes dengan bertujuan untuk mengukur tingkat akurasi dan lama waktu proses penentuan penerima beasiswa. dalam memprediksi Setelah dilakukan pengujian, maka hasil yang didapat dari algoritma C4.5 memiliki akurasi sebesar 54%. Lebih baik dari tingkat akurasi dari algoritma naïve bayes sebesar 93.33%, dan dilakukan perbandingan menggunakan naïve bayes. Dengan adanya penelitian diharapkan dapat membantu SMK ISLAM PERMATASARI 1 dalam memberi beasiswa dalam menentukan kriteria yang lebih diutamakan untuk pemberian beasiswa, sehingga pemberian beasiswa dapat tepat sasaran.

Keywords: Data Mining; Algoritma Klasifikasi Dana; Naïve Bayes

### 1. PENDAHULUAN

Dalam institusi Pendidikan nilai siswa dapat membantu institusi sebagai bahan pertimbangan

dalam memberikan beasiswa. Kualitas dari siswa akan berpengaruh terhadap kualitas institusi pendidikan. Analisis dan pengolahan data nilai jika

dilakukan dengan cara manual akan memakan waktu lama. Selain itu, pengolahan data secara manual masih memungkinkan adanya kesalahan. Berdasarkan permasalahan tersebut, akan dilakukan prediksi bantuan dengan menggunakan algoritma klasifikasi.

Data Mining terdiri dari dua kata yaitu Data dan Mining. Data yaitu sekumpulan informasi berupa fakta, sedangkan Mining yang berarti menambang. Secara umum, data Mining adalah sebuah prosesi pengumpulan informasi atau data berupa fakta dalam jumlah yang besar. Prosesi data Mining seiring kali menggunakan metode statistika, matematika, hingga memanfaatkan teknologi artifisial intelligence. Dengan penggabungan beberapa metode ini dapat memudahkan serta menyempurnakan prosesi pengumpulan informasi. Selain daripada itu, data Mining berkaitan erat dengan database. Evolusi dari database itu sendiri berkembang mulai dari tahun 1960, 1970, 1980, 1990, hingga tahun 2000. Dari evolusi ini kita belajar, di tahun 2000 ke atas hal apa yang harus kita persiapkan untuk menunjang metode menjadi lebih baik.

Pentingnya data Mining dalam dunia pendidikan dapat memudahkan dalam prosesi pengambilan keputusan. Pengambilan data dalam jumlah yang sangat besar pastinya akan memberikan risiko yang sangat besar pula, sehingga data Mining yang menggabungkan beberapa metode ini dapat memberikan hasil secara maksimal. Maka dari itu jika dikaitkan dengan fungsi dari data Mining itu sendiri yaitu dapat memudahkan dalam prosesi pengambilan keputusan di ruang lingkup kecil ataupun besar.

Pendidikan merupakan salah satu komponen kehidupan yang dapat menunjang keberhasilan seseorang menuju kehidupan yang jauh lebih baik lagi. Terutama bagi anak yang ada dalam lingkup usia wajib belajar. akan tetapi tidak semua anak wajib belajar dapat mengikuti pendidikan, karena beberapa faktor penyebab. Oleh karena itu begitu penting nya mengatasi masalah biaya atau ekonomi bagi mereka yang kurang mampu. Beasiswa adalah pemberian berupa bantuan keuangan yang diberikan kepada perorangan yang bertujuan untuk digunakan demi keberlangsungan pendidikan yang ditempuh. Pemberi beasiswa, yaitu lembaga pemerintahan, perusahaan, atau yayasan. Pemberian beasiswa ini sangat penting, guna memberikan ketenangan mereka dalam mengikuti belajar- mengajar. Namun terkadang pemberian beasiswa tidak pada orang yang seharusnya menerima atau tidak tepat pada sasarannya. Dari permasalahan tersebut dilakukan penelitian mengenai seleksi penerimaan beasiswa agar membantu meningkatkan kecepatan dan

ketepatan dalam penentuan beasiswa dengan menggunakan metode algoritma C4.5 dan naïve bayes dengan bertujuan untuk mengukur tingkat akurasi dan lama waktu prosesi penentuan penerima beasiswa. Dalam memprediksi Setelah dilakukan pengujian, maka hasil yang didapat dari algoritma C4.5 memiliki akurasi sebesar 96.40%. lebih baik dari tingkat akurasi dari algoritma naïve bayes sebesar 95.11%, dan di lakukan perbandingan menggunakan naïve bayes.

Dengan adanya penelitian diharapkan dapat membantu SMK ISLAM PEIRMATASARI 1 dalam memberi beasiswa dalam menentukan kriteria yang lebih diutamakan untuk pemberian beasiswa, sehingga pemberian beasiswa dapat tepat sasaran. Seperti yang dituangkan dalam Undang-Undang Dasar 1945 pasal 31(1) bahwa tiap warga negara berhak mendapatkan pengajaran. Akan tetapi untuk mendapatkan pendidikan itu semakin tidak terjangkau atau sangatlah mahal, terutama bagi yang bertingkat ekonomi menengah kebaan, baik itu tingkat SD, SMP, SMA, bahkan perguruan tinggi. Beasiswa merupakan salah satu solusi untuk mengatasi masalah biaya bagi mereka yang kurang mampu. Menurut Kurniasih dalam Rainier (2017) beasiswa diartikan “sebagai bentuk penghargaan yang diberikan kepada individu agar dapat melanjutkan pendidikan ke jenjang yang lebih tinggi”. Sehingga dengan adanya beasiswa dapat menciptakan pengaruh yang positif bagi Mereka, sehingga tidak perlu lagi khawatir dengan tagihan biaya pendidikan dan Anak tersebut berusaha untuk mendapatkan nilai dan prestasi yang tinggi.

## **2. PENELITIAN YANG TERKAIT**

Pada penelitian sebelumnya dengan judul “ Analis kinerja algoritma C4.5 Dan Naïve Bayes untuk memprediksi prestasi siswa sekolah menengah kejuruan ”pada tahun 2019 yang dapat mewujudkan suasana belajar dan proses pembelajaran peserta didik mengembangkan potensi dirinya. Hasil dari penelitian ini dilakukan analisis dan komparasi dua metode klasifikasi data Mining yang memiliki karakteristik yang berbeda, maka hasil prediksi belajar menunjukkan bahwa algoritma Naïve Bayes memiliki nilai akurasi sebesar 95,67% dan nilai AUC sebesar 0,999 termasuk dalam excellent classification termasuk dalam excellent classification, untuk algoritma C4.5 memiliki nilai akurasi 90,91% dan nilai AUC sebesar

0,639 termasuk dalam keadaan poor classification.

Pada penelitian yang dilakukan oleh penulis dengan judul “Implementasi Data Mining dalam memperoleh bantuan dana menggunakan algoritma C4.5 dan Naïve Bayes di SMK ISLAM PERMATASARI 1” pada tahun 2023 dengan tujuan agar bisa mempermudah dalam mengolah data yang banyak, tepat sasaran saat memberikan bantuan dana terhadap siswa, dan efisien. Hasil dari penelitian ini berdasarkan data set yang di implementasikan dengan algoritma C4.5 diperoleh dengan nilai akurasi 54,00%. prediksi menggunakan metode Naïve Bayes memiliki nilai akurasi 93.33% yang lebih besar di bandingkan dengan nilai akurasi menggunakan algoritma C4.5.

Perbedaan antara penelitian yang sebelumnya dengan penelitian yang dilakukan penulis jelas berbeda, studi kasus yang berbeda dan hasil presentase yang dihasilkan berbeda karena rumusan masalah dan pemecahan masalah yang berbeda.

### 3. METODE PENELITIAN

#### Metode pengumpulan Data

##### Wawancara

Metode wawancara merupakan salah satu teknik pengumpulan data yang dilakukan dengan cara mengadakan Tanya jawab, baik secara terstruktur merupakan teknik wawancara yang dilakukan apabila peneliti telah memiliki informasi dari apa yang akan ditelitinya. Pewawancara juga telah menyiapkan pertanyaan tertulis serta alternatif jawabannya. Alat penunjang adalah tape reporter, gambar, dan brosur. maupun secara bebas (tidak terstruktur) peneliti tidak menggunakan pedoman wawancara yang sistematis dan lengkap untuk pengumpulan datanya. Biasanya di lakukan secara mendalam untuk mendapatkan jawaban responden atau narasumbernya, isi pembicaraan bergantung pada suasana wawancara.

##### Kuesioner

Kuesioner disebut juga angket adalah metode pengumpulan data, untuk memahami individu. Kuesioner berisi daftar pertanyaan tentang berbagai aspek. Peneliti akan mendapatkan berbagai macam data sesuai pertanyaan.

Ada 4 jenis kuesioner :

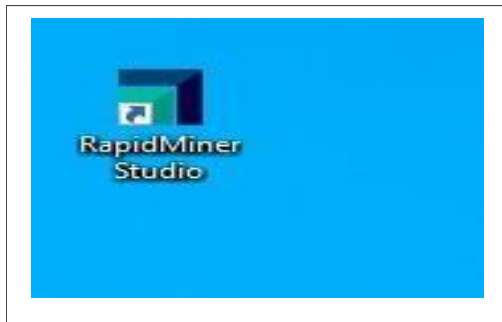
- Kuesioner Tertutup Sebuah kuesioner menggunakan pertanyaan yang tertutup, sehingga responden dapat memilih jawaban yang tersedia. Jawaban ini telah terikat, sehingga responden tidak dapat memberikan jawaban sebebasnya.
- Kuesioner Terbuka Pertanyaan terbuka di kuesioner memberi responden kebebasan jawaban. Responden dapat memberi tanggapan dari jawaban kuesioner secara tertulis. Kelebihan kuesioner terbuka yaitu peneliti mendapatkan data lebih bervariasi.
- Kuesioner Terbuka dan Tertutup Jenis kuesioner yang menggabungkan pertanyaan terbuka dan tertutup. Umumnya jenis kuesioner ini dipakai kepentingan bimbingan konseling. Data yang didapatkan memakai pertanyaan tertutup relatif pasti, sesuai fakta dan mudah diukur.
- Kuesioner Semi Terbuka Jenis kuesioner yang memberikan responden kesempatan memakai jawaban lain. Penggunaan jawaban alternatif jika jawaban tersedia tidak cocok. Kuesioner yang digunakan adalah kuesioner terbuka di kuesioner memberi responden kebebasan jawaban. Responden dapat memberi tanggapan melalui G-Form yang sudah disediakan oleh peneliti. Kelebihan kuesioner terbuka yaitu peneliti mendapatkan data lebih bervariasi.

#### Metode Pengembangan Sistem

Dalam metode pengembangan sistem penulisan skripsi ini, penulis menggunakan metode algoritma klasifikasi. Algoritma yang di gunakan C4.5 dan Naïve Bayes yang dimana kedua metode tersebut jika di gabungkan menghasilkan nilai yang maksimal. Dengan cara memanfaatkan tools Rapid Miner untuk membandingkan metode tersebut. Rapid miner merupakan salah satu tools yang dipakai dalam data Mining. RapidMiner memiliki kurang lebih 500 operator data Mining, termasuk input, output, data processing dan visualisasi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data Mining yang dapat diintegrasikan pada produknya sendiri.

### 4. HASIL DAN PEMBAHASAN

#### Aplikasi RapidMiner

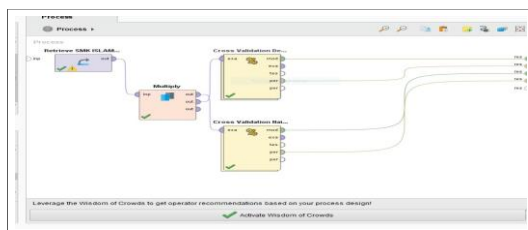


Gambar 1. Tampilan RapidMiner

Aplikasi ini berfungsi untuk penelitian digunakan untuk penelitian, pendidikan, pembuatan prototipe dengan cepat, dan pengembangan aplikasi serta mendukung semua Langkah prosesi pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi dan pengoptimalan.

Proses data Mining melibatkan matematika, statistika, data science hingga pemanfaatan teknologi Artificial Intelligence. Ada banyak tools yang bisa diaplikasikan untuk menunjang proses data Mining. Salah satu tools yaitu RapidMiner adalah tools data science open-source untuk melakukan analisis data. Mining, text Mining dan prediction analysis. RapidMiner memiliki kurang lebih 500 operator data Mining seperti input, output, data processing dan visualisasi. Di kalangan akademik RapidMiner digunakan oleh mahasiswa, dosen atau peneliti dengan latar belakang Sistem Informasi untuk pemodelan System berbasis kecerdasan buatan (optimasi, pengenalan pola gambar, teks, grafik, peramalan). Tools yang sejenis dibangun menggunakan Bahasa Java sehingga dapat dijalankan berbagai macam sistem operasi seperti Windows, Linux, Unix serta MacOS.

Desain Model Perbandingan Algoritma Klasifikasi

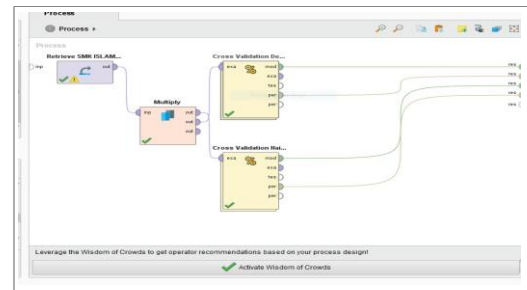


Gambar. 2 Desain Model Perbandingan Algoritma Klasifikasi

Desain model proses utama ini langsung menggunakan satu data set yang sama untuk dua algoritma yaitu algoritma C4.5 dan Naive

Bayes untuk memastikan konsistensi data set yang sama yang di proses oleh setiap algoritma.

Desain model Training Dan Testing Algoritma C4.5  
 Gambar .3 Design Training Dan Testing Algoritma



C4.5

Merupakan rincian disain model proses training dan testing untuk klasifikasi algoritma Decision Tree C4.5 yang di buat sesuai kebutuhan setiap peneliti.

Setelah kita melakukan View proses dan membuat proses training dan testing algoritma Decision Tree menghasilkan pohon keputusan yang dimana nilai raport mempengaruhi kelayakan siswa mendapatkan bantuan dana pendidikan.

Hasil Perbandingan Algoritma C4.5 Dan Naive Bayes

Tabel I Tabel Perbandingan

No	Algoritma	Tingkat Akurasi
1.	C4.5	54,00%
2.	Naive Bayes	93,33%

Hasil dari perbandingan tingkat akurasi tersebut menunjukkan bahwa dalam melakukan prediksi bantuan dana pendidikan algoritma C4.5 Dan Naive Bayes. Nilai tingkat akurasi Naive Bayes lebih tinggi di banding nilai tingkat akurasi Algoritma C4.5. Hasil perbandingan ini sesuai dengan penelitian oleh Fatmawati dan Listiana et al, dimana kinerja algoritma Naive Bayes memiliki tingkat akurasi yang lebih tinggi di banding algoritma Decision Tree.



## Kesimpulan Pengujian

Kesimpulan pengujian adalah suatu akhir dalam suatu pengujian, disini saya menggunakan pengujian black box dan pengujian user response (Kuesioner). Berikut kesimpulan dari pengujian saya :

### Kesimpulan pengujian black box

Berdasarkan hasil pengujian yang mengacu pada pengujian black box menunjukkan bahwa sistem RapidMiner dengan menggunakan beberapa pola tidak ada kesalahan. Dengan demikian, pengujian ini telah berhasil 100%.

### Kesimpulan pengujian white box

Berdasarkan hasil pengujian white box menghasilkan algoritma C4.5 54,00 % dan untuk naïve bayes menghasilkan 93,33 %. Dengan demikian, pengujian ini telah berhasil 100 %.

### Kesimpulan user response (kuesioner)

Kesimpulan hasil pengujian yang mengacu pada pengujian user response yang dilakukan pada google form yang di isi oleh 26 orang menunjukkan bahwa 88,5% menyatakan bahwa tidak ada kesulitan saat menginstal.

aplikasi RapidMiner, dan 80,8% menyatakan bahwa RapidMiner membuat efisien waktu saat memprediksi bantuan dana, dan 76,9% menyatakan bahwa RapidMiner mudah dipahami, dan 80,8% menyatakan bahwa RapidMiner sesuai kebutuhan SMK ISLAM PERMATASARI, dan 76,9% menyatakan bahwa RapidMiner membantu memprediksi bantuan dana pendidikan di SMK ISLAM PERMATASARI 1, dan 80% menyatakan bahwa Sistem Data Mining menggunakan RapidMiner sudah lebih canggih dengan adanya AI, dan 80,8% aktivitas pengambilan keputusan menjadi lebih simple dalam mengolah data yang banyak, dan 76,9 menyatakan bahwa RapidMiner memberikan tingkat akurasi yang tepat, dan 76,9 menyatakan bahwa tools RapidMiner salah satu yang diharapkan SMK ISLAM PERMATASARI 1, dan 80,8% RapidMiner akan berguna dan terpakai untuk kemajuan perkembangan SMK ISLAM PERMATASARI .

## 5. KESIMPULAN

Kesimpulan dari penelitian ini bahwa hasil prediksi bantuan dana pendidikan di SMK ISLAM PERMATASARI 1 berdasarkan data set yang di implementasikan dengan metode Algoritma C4.5 diperoleh nilai akurasi 54,00%, Prediksi menggunakan metode Naïve Bayes memiliki nilai akurasi 93,33% yang lebih besar di bandingkan dengan nilai akurasi menggunakan algoritma C4.5 Hasil kuesioner yang telah di uji cebakan oleh staff SMK ISLAM PERMATASARI 1 dan beberapa mahasiswa dan Masyarakat umum, menjadi sebuah jawaban dan pemecahan masalah yang ada di SMK ISLAM PERMATASARI 1, yaitu dapat berguna bagi pihak Yayasan dan memudahkan bagian staff kesiswaan untuk membuat prediksi bantuan dana pendidikan bagi siswa siswi dan memberikan pengaruh di era perkembangan teknologi yang sudah canggih sehingga kita bisa membuat sebuah keputusan dalam memprediksi bantuan dana siswa dengan waktu yang lebih efisien dan menjadi sebuah aktivitas yang sederhana dalam mengolah sebuah data yang banyak.

Dari pohon keputusan algoritma C4.5 dapat di simpulkan bahwa variabel atau kriteria yang berpengaruh dalam prediksi bantuan dana pendidikan di SMK ISLAM Permatasari 1 adalah pendapatan orang tua, dan alat transportasi dan Nilai Rapot.

## DAFTAR PUSTAKA

- [1] Afrianto, Jatmiko Eko, Endro Suseno, and Budi Warsito. 2020. "Sistem INFORMASIKLASIFIKASI SISWA Penerima KARTU Indonesia PINTAR (KIP) Menggunakan Metode Decision Tree Dengan ALGORITMA C4."
- [2] Ardiansyah, M., Sunyoto, A., & Luthfi, E. T. (2021). Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4. 5 untuk Klasifikasi Diabetes. *Edumatic J. Pendidik. Inform.*
- [3] Ariadi, F. (2020). Analisa Perbandingan Algoritma DT C. 45 dan Naïve Bayes Dalam Prediksi Penerimaan Kredit Motor. *KERNEL: Jurnal Riset Inovasi Bidang Informatika dan Pendidikan Informatika.*

- [4] Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4. 5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*.
- [5] Badriah, Siti, and Et Al. 2021. “Klasifikasi Algoritma C4. 5 Dalam Menentukan Penerima Bantuan Covid-19.” *Jurnal Informatika Polinema*.
- [6] Baskoro, Baskoro, Sriyanto Sriyanto, and Lilis Setya Rini. 2021. “Prediksi Penerima Beasiswa Dengan Menggunakan Teknik Data Mining Di Universitas Muhammadiyah Pringsewu.” In *Prosiding Seminar Nasional Darmajaya*.
- [7] Fatmawati, F., & Narti, N. (2022). Perbandingan Algoritma C4. 5 dan Naive Bayes Dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring. *JTIM: Jurnal Teknologi Informasi dan Multimedia*.
- [8] Fitriani, E. (2020). Perbandingan Algoritma C4. 5 Dan Naive Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. *Sistemasi*.
- [9] Handayani, K., Lisnawanty, L., Latif, A., Firdaus, M. R., & Hasan, F. N. (2021). Komparasi Algoritma C4. 5 dan Naive Bayes dalam Penentuan Status Kelayakan Donor Darah. *Sistemasi*.
- [10] Hendrian, Senna. 2018. “Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan.” Faktor Exacta.
- [11] Kadafi, Abdul Rahman. Perbandingan Algoritma Klasifikasi Untuk Penjurusan Siswa SMA. *Jurnal ELTIKOM: Jurnal Teknik Elektro, Teknologi Informasi Dan Komputer*, 2018.
- [12] Kamil, M., & Cholil, W. (2020). Analisis Perbandingan Algoritma C4. 5 dan Naive Bayes pada Lulusan Tepat Waktu Mahasiswa di Universitas Islam Negeri Raden Fatah Palembang. *Jurnal Informatika*.
- [13] Kohsasih, K. L., & Situmorang, Z. (2022). Analisis Perbandingan Algoritma C4. 5 dan Naive Bayes Dalam Memprediksi Penyakit Cerebrovascular. *Jurnal Informatika*.
- [14] Ladania, H. (2023). *ANALISIS PERBANDINGAN ALGORITMA C4. 5 DAN NAIVE BAYES UNTUK KLASIFIKASI DAERAH RAWAN PANGAN* (Doctoral dissertation, Universitas Tadulako).
- [15] Lita, Debby, and Et Al. 2021. “Penerapan Naive Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa.” *Jurnal Data Mining Dan Sistem Informasi*.
- [16] Marbun, S., Nugroho, N. B., & Ginting, R. I. (2022). Implementasi Data Mining Menggunakan Algoritma Naive Bayes Dalam Memprediksi Angka Kelahiran. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*.