

## IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK MENENTUKAN CALON PENERIMA BEASISWA DI SMK YPM 14 SUMOBITO JOMBANG

Wahyuningsih<sup>2)</sup>, Budiman<sup>2)</sup>, Izzatul Umami<sup>3)</sup>

<sup>123</sup>Fakultas Teknik, Universitas Darul ‘Ulum, Jl. Gus Dur No.29A Mojongapit Jombang  
email: [aningwahyuningsih0@gmail.com](mailto:aningwahyuningsih0@gmail.com), [budimanjombang1@gmail.com](mailto:budimanjombang1@gmail.com), [izzatulummuslimi@gmail.com](mailto:izzatulummuslimi@gmail.com)

### ABSTRACT

*Scholarships are grants in the form of financial assistance given to individuals that aim to be used for the continuation of the education achieved. With this scholarship program, it is hoped that it can help students who have problems in financing. The selection of scholarship recipients at SMK YPM 14 Sumobito still experiences problems in decision making, because the assessment process is not always decided based on definite considerations but the policy of the decision makers who ultimately determine the scholarship recipients. This is because there is no method that can predict prospective scholarship recipients. Eligibility of prospective scholarship recipients is determined by applying the Naïve Bayes method. This method was chosen because it was able to study the previous case data used as test data. After testing with the Naïve Bayes algorithm using the Rapidmaner tool, the results were 90.48% accuracy, 96.88% precision, 83.33% recall each. So that it can be said as a good key to apply for prospective scholarship recipients.*

**Keywords:** (scholarship assistance, Data Mining, Naïve Bayes).

### ABSTRAK

Beasiswa adalah pemberian berupa bantuan keuangan yang diberikan kepada perorangan yang bertujuan untuk digunakan demi keberlangsungan pendidikan yang ditempuh. Dengan adanya program Beasiswa ini diharapkan dapat membantu siswa yang memiliki masalah dalam hal pembiayaan. Seleksi penerimaan beasiswa di SMK YPM 14 Sumobito masih mengalami kendala pada pengambilan keputusan, karena proses penilaiannya tidak selalu diputuskan berdasarkan perhitungan yang pasti tetapi kebijakan dari pembuat keputusan yang akhirnya menentukan penerima beasiswa. Hal ini dikarenakan belum ada sebuah metode yang dapat memprediksi calon penerima beasiswa. Kelayakan calon penerima beasiswa ditentukan dengan menerapkan metode Naïve Bayes. Metode ini dipilih karena mampu mempelajari data kasus sebelumnya yang digunakan sebagai data uji. Setelah dilakukan pengujian dengan algoritma Naïve Bayes tersebut menggunakan tools Rapidmaner didapatkan hasil yaitu nilai accuracy 90,48%, precision 96,88%, recall 83,33% disetiap algoritmanya. Sehingga dapat disimpulkan algoritma Naïve Bayes merupakan algoritma yang baik untuk diterapkan untuk calon penerima beasiswa.

**(Kata Kunci :** Bantuan beasiswa, Data Mining, Naïve Bayes)

*This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license*



### PENDAHULUAN

Pendidikan merupakan kebutuhan manusia yang sangat penting karena pendidikan mempunyai tugas untuk

menyiapkan Sumber Daya Manusia (SDM) bagi pembangunan bangsa dan negara. Mengingat begitu pentingnya pendidikan, negara sangat mendukung

warga negaranya untuk meraih pendidikan setinggi-tingginya. Namun, tidak setiap orang memiliki kesempatan untuk mengenyam pendidikan karena permasalahan. Program Beasiswa dapat membantu siswa yang memiliki masalah dalam hal pembiayaan. Beasiswa adalah pemberian bantuan keuangan yang diberikan kepada perorangan yang bertujuan untuk digunakan demi keberlangsungan pendidikan yang ditempuh dan diberikan oleh lembaga pemerintah, perusahaan ataupun yayasan.

Demikian juga di SMK YPM 14 Sumobito masih mengalami kendala pada proses pengambilan keputusan untuk menentukan mahasiswa mana saja yang berhak mendapatkan beasiswa karena proses penilaiannya tidak selalu diputuskan berdasarkan perhitungan yang pasti tetapi kebijakan dari pembuat keputusan yang akhirnya menentukan penerima beasiswa. Hal ini dikarenakan belum ada sebuah metode yang dapat memprediksi calon penerima beasiswa.[2] Proses seleksi penerimaan beasiswa merupakan hal yang paling penting, agar tidak salah sasaran. Seleksi penerimaan beasiswa ini melibatkan banyak kriteria yang akan dinilai. Semakin banyak siswa yang mengajukan, maka akan semakin menyulitkan sekolah dalam melakukan perhitungan yang akan menentukan siapa yang layak menerima beasiswa tersebut.

Proses penerimaan beasiswa di SMK YPM 14 Sumobito akan lebih efektif dan efisien jika menggunakan sebuah system pendukung keputusan yang dapat membantu mempercepat dan mempermudah pengambilan keputusan.

Maka dari itu diperlukan suatu sistem atau aplikasi yang dapat membantu pihak sekolah dalam mengambil keputusan penentuan penerima beasiswa.

Penelitian ini akan menerapkan algoritma *Naïve Bayes* dalam penentuan calon penerima beasiswa. *Naïve Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma *Naïve Bayes* adalah salah satu metode *data mining* yang termasuk kedalam sepuluh klasifikasi data mining yang paling populer diantara algoritma-algoritma lainnya[1]. Metode *Naïve Bayes* juga dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan metode pengklasifikasian yang lain dalam hal akurasi dan efisiensi komputasi.

## METODE PENELITIAN

Metode Penelitian Populasi dan Sampel

### a. Populasi

Populasi adalah keseluruhan obyek penelitian dengan batas-batas persoalan yang sudah cukup jelas. Populasi adalah wilayah generalisasi yang terdiri atas obyek/subyek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya. Jadi populasi adalah keseluruhan objek penelitian dengan batas persoalan yang telah ditentukan sebelumnya untuk dipelajari dan

kemudian menghasilkan suatu kesimpulan. Populasi dalam penelitian ini merupakan siswa yang dinyatakan layak mendapatkan bantuan beasiswa dan siswa yang tidak layak mendapatkan bantuan beasiswa tahun 2022-2023 yang berasal dari data internal di salah satu Sekolah yang ada di SMK YPM 14 Sumobito.

b. Sampel

Sampel adalah bagian dari jumlah dan karakteristik yang dimiliki oleh populasi tersebut. Penelitian menggunakan sampel yang diambil dari populasi. Apa yang dipelajari dari sampel itu, kesimpulannya akan dapat diberlakukan untuk populasi. Untuk itu sampel yang diambil dari populasi harus betul-betul representatif (mewakili). Sampel dari penelitian ini adalah data siswa yang layak mendapatkan bantuan dan siswa yang tidak layak mendapatkan bantuan beasiswa. Data tersebut bersifat intern yang belum dipublikasikan oleh pihak sekolah dapat dilihat pada Table 3.1 jumlah sampel yang diambil :

**Tabel 3.1 Sampel Dataset**

Dataset	Layak	Tidak Layak	Total Sampel
Data Training	29	155	184
Data Testing	9	33	42

Metode Pengumpulan Data

Metode pengumpulan data dibagi menjadi dua sumber data yaitu data primer dan data sekunder. Data Primer diperoleh melalui pertanyaan lisan dengan menggunakan metode

wawancara, observasi, kuesioner dll. Data sekunder adalah data dari sumber data lain atau data yang diperoleh dari obyek penelitian yaitu berupa dokumentasi[2]. Data sekunder diperoleh dari instansi dan bacaan terkait penelitian. Jadi data primer adalah data yang langsung diperoleh dari sumber dan data sekunder adalah data yang diperoleh dari sumber lain baik dokumentasi, literatur, buku, jurnal, dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti.

Dalam penelitian ini metode pengumpulan data untuk mendapatkan sumber data yang digunakan adalah metode pengumpulan data sekunder. Data utama diperoleh dari data siswa yang layak mendapatkan bantuan beasiswa dan siswa yang tidak layak mendapatkan bantuan beasiswa sedangkan data pendukung lainnya didapat dari buku, jurnal dan publikasi lainnya.

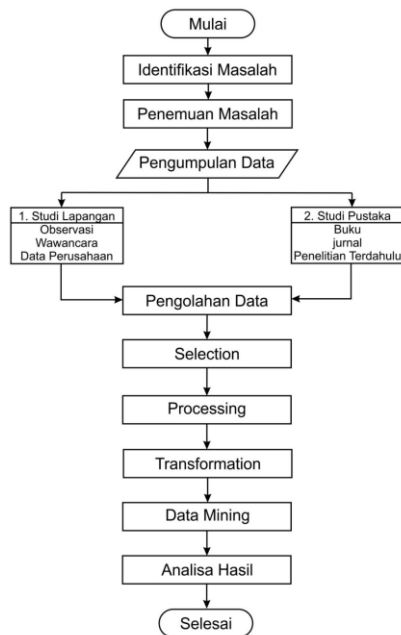
Diagram Alir Penelitian

Langkah – langkah penelitian pada tugas akhir Implementasi Algoritma Naïve Bayes untuk Memprediksi Kelayakan Penerima Bantuan Beasiswa di SMK YPM 14 Sumobito di ilustrasikan seperti pada Gambar 3.1.

Pada diagram alir akan terdapat tahapan proses pengolahan data dari data awal yang belum diolah hingga proses data mining dan menghasilkan suatu kesimpulan yang diperoleh setelah data diolah. Dari kesimpulan tersebut diharapkan dapat membantu SMK YPM 14 Sumobito dalam memprediksi kelayakan penerima bantuan siswa dan sebagai bahan seleksi dalam memilih atau menentukan

penerima manfaat tersebut kepada sekolah tersebut.

Tahapan penelitian pada Tugas Akhir Implementasi Algoritma Naïve Bayes untuk memprediksi kelayakan penerima bantuan beasiswa di SMK YPM 14 Sumobito dapat diilustrasikan seperti gambar 3.2 dibawah ini :



### Metode Analisis dan Pengujian Data

Teknik Analisis data menggunakan berupa matematika terhadap angka atau numerik dan nominal. Pada penelitian ini, analisis data dilakukan melalui data siswa salah satu sekolah yang ada di SMK YPM 14 Sumobito Jombang dengan nilai rata-rata siswa yang mendapatkan beasiswa dan siswa yang tidak layak mendapatkan bantuan. Data diolah dan di uji dalam pengujian pada algoritma Naïve Bayes. Kemudian pengujian *Rule* yang diperoleh Naïve Bayes tersebut kemudian diuji dengan *confusion matrix* untuk mengukur tingkat akurasi yang akan dihasilkan dari metode tersebut.

Metode penelitian yang digunakan pada eksperimen ini menggunakan metodologi standar dalam penelitian *data mining* adalah model *Cross-Standard Industry for Data Mining* (CRISP-DM). Model CRISP-DM (Cross – Industry Standard Process for Data Mining) yang terdiri dari 6 tahap proses yaitu : *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment*.

#### a. Business Understanding

Berdasarkan data masyarakat yang menerima beasiswa menunjukkan adanya tidak tepat sasaran penerima beasiswa tersebut. Pada penelitian ini dilakukan pengembangan dengan komparasi algoritma Naïve Bayes dengan tujuan untuk mengetahui kelayakan siswa dalam penerima bantuan beasiswa dan meningkatkan akurasi dari perhitungan algoritma klasifikasi tersebut.

#### b. Data Understanding

Pada tahap *Data Understanding*, dilakukan pengumpulan data, melakukan analisis penyelidikan data (data siswa penerima bantuan beasiswa) untuk mengenali lebih lanjut data dan pencarian pengetahuan awal kemudian mengevaluasi kualitas dari data tersebut. Adapun sumber data utama yang digunakan dalam penelitian ini menggunakan data siswa di SMK YPM 14 Sumobito dengan beberapa atribut tersebut. Data tersebut dianalisis misalnya jumlah data yang akan diambil, dan jumlah data dengan keterangan layak atau tidak layak.

#### c. Data Preparation

Pada tahapan ini data siswa yang

terdiri dari siswa yang layak dan yang tidak layak, terdapat beberapa atribut, akan dilakukan beberapa penyeleksian untuk menghasilkan data yang dibutuhkan.

d. Modelling

Dilakukan pemrosesan *data training* sehingga akan menghasilkan beberapa aturan dan akan membentuk sebuah pohon keputusan.

e. Evaluation

Melihat hasil akurasi pada proses

NO	ATRIBUT	KATEGORI	
1	Pekerjaan Orang Tua	Swasta	
		Wiraswasta	
		Petani	
		Pedagang	
		Penjahit	
		PNS/Polisi/TNI/Polri	
2	Penghasilan Orang Tua	<1.000.000	Rendah
		1.100.000-1.500.000	Sedang
		1.600.000-2.500.000	Tinggi
		>2.600.000	Sangat Tinggi
3	Tanggungjawab Orang Tua	< 3 Orang	Cukup
		>3 Orang	Banyak
4	Jarak	< 1 KM	Dekat
		1-5 KM	Jauh
		> 5 KM	Sangat Jauh
5	Nilai	<=80	Baik
		>80	Sangat Baik
6	Status Siswa	Yatim/Piatu	
		Tidak	
7	Sikap	Sangat Baik	
		Baik	

klasifikasi komparasi algoritma

Naïve Bayes serta evaluasi dengan *confision matrixe*. Evaluasi bertujuan untuk menentukan nilai kegunaan dari model yang telah berhasil kita buat pada langkah sebelumnya.

f. Develoyment

Berdasarkan penelitian yang dilakukan dengan penerapan model komparasi algoritma naïve bayes untuk menentukan calon penerima beasiswa.

**Tabel Kategori Kriteria Data Beasiswa**

Adapun Kriteria yang digunakan untuk menentukan calon penerima beasiswa SMMK YPM 14 Sumobito dapat dikategorikan seperti diatas. Data yang digunakan adalah siswa di tahun pelajaran 2021/2022.

**HASIL DAN PEMBAHASAN**

Dalam melakukan penerapan Algoritma Naïve Bayes, memerlukan lebih dari satu data. Adapun data yang dibutuhkan adalah sebagai berikut :

- a. Daftar nama calon penerima beasiswa
- b. Nilai Siswa dan Sikap kepribadian
- c. Dokumen Pribadi calon penerima beasiswa

Data mentah diatas akan diproses menjadi data-set untuk memudahkan penerapan Algoritma data mining.

**Pengolahan Awal Data**

Sebelumnya diterapkan algoritma data mining terhadap sebuah data-set, perlu dilakukan pengolahan awal yang bertujuan untuk mendapatkan data-set

yang dapat diolah dengan cepat dan menghasilkan kesimpulan yang tepat. Hasil dari pre processing dari data sensus menjadi data set, data yang akan digunakan didalam laporan ini berupa data sampling sebanyak 42 data (20%) dari data testing,

**Penghitungan Manual Naïve Bayes**

Data Training Uji Coba Berjumlah 42 buah data (20%) dan Data Training berjumlah 184 data (80%). Variabel yang menjadi acuan adalah Tanggungan Orang Tua, Jarak Rumah, Pekerjaan Orang Tua, Penghasilan Orang Tua, Status (Yatim/Piatu atau Tidak, Nilai, Sikap.

Berikut contoh penghitungan data training

- Tanggungan Orang Tua |
  - Cukup
  - Banyak
- Jarak Rumah |
  - Dekat
  - Jauh
  - Sangat Jauh
- Pekerjaan Orang Tua |
  - Pedagang
  - Penjahit
  - Petani
  - Swasta
  - Wiraswasta
  - PNS
  - Polisi
- Penghasilan Orang Tua |
  - Rendah
  - Sedang
  - Tinggi
  - Sangat Tinggi
- Status |
  - Tidak
  - Yatim/Piatu
- Nilai |
  - Baik
  - Sangat Baik
- Sikap |

Baik  
Sangat Baik

Probabilitas Class Layak :

$$P(\text{Layak}) = 29/184 =$$

$$0,157608696$$

Probabilitas Class Tidak

Layak :

$$P(\text{Tidak Layak}) = 155/184 =$$

$$0,842391$$

Menghitung kemungkinan

kategori Layak :

P (Class. Layak | Tanggungan Orang Tua. Cukup)

$$= 18/29 = 0,620689655$$

P (Class. Layak | Tanggungan Orang Tua. Banyak)

$$= 11/29 = 0,379310345$$

P (Class. Layak | Jarak Rumah. Dekat)

$$= 2/29 = 0,068965517$$

P (Class. Layak | Jarak Rumah. Jauh)

$$= 22/29 = 0,75862069$$

P (Class. Layak | Jarak Rumah. Sangat Jauh)

$$= 5/29 = 0,172413793$$

P (Class. Layak | Pekerjaan Orang Tua. Pedagang)

$$= 4/29 = 0,137931034$$

P (Class. Layak | Pekerjaan Orang Tua. Penjahit)

$$= 1/29 = 0,034482759$$

P (Class. Layak | Pekerjaan Orang Tua. Petani)

$$= 14/29 = 0,482758621$$

P (Class. Layak | Pekerjaan Orang Tua. Swasta)

$$= 10/29 = 0,344827586$$

P (Class. Layak | Pekerjaan Orang Tua. Wiraswasta)

$$= 0/29 = 0$$

P (Class. Layak | Pekerjaan Orang Tua. PNS)

$$= 0/29 = 0$$

P (Class. Layak | Pekerjaan Orang Tua. Polisi)

$$= 0/29 = 0$$

P (Class. Layak | Penghasilan Orang Tua. Rendah)  
 =  $25/29 = 0,862068966$   
 P (Class. Layak | Penghasilan Orang Tua. Sedang)  
 =  $4/29 = 0,137931034$   
 P (Class. Layak | Penghasilan Orang Tua. Tinggi)  
 =  $0/29 = 0$   
 P (Class. Layak | Penghasilan Orang Tua. Sangat Tinggi)  
 =  $0/29 = 0$   
 P (Class. Layak | Status. Tidak)  
 =  $20/29 = 0,689655172$   
 P (Class. Layak | Status. Yatim/Piatu)  
 =  $9/29 = 0,310344828$   
 P (Class. Layak | Nilai. Baik)  
 =  $8/29 = 0,275862069$   
 P (Class. Layak | Nilai. Sangat Baik)  
 =  $21/29 = 0,724137931$   
 P (Class. Layak | Sikap. Baik)  
 =  $16/29 = 0,551724138$   
 P (Class. Layak | Sikap. Sangat Baik)  
 =  $13/29 = 0,448275862$

Menghitung kemungkinan kategori Tidak Layak :  
 P (Class. Tidak Layak | Tanggungan Orang Tua. Cukup)  
 =  $111/155 = 0,716129$   
 P (Class. Tidak Layak | Tanggungan Orang Tua. Banyak)  
 =  $44/155 = 0,283871$   
 P (Class. Tidak Layak | Jarak Rumah. Dekat)  
 =  $9/155 = 0,058065$   
 P (Class. Tidak Layak | Jarak Rumah. Jauh)

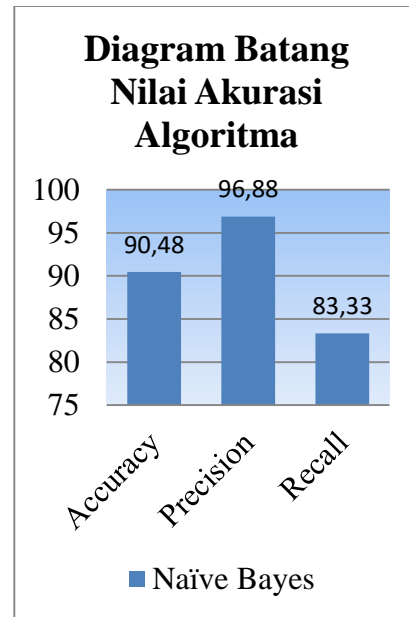
=  $92/155 = 0,593548$   
 P (Class. Tidak Layak | Jarak Rumah. Sangat Jauh)  
 =  $54/155 = 0,348387$   
 P (Class. Tidak Layak | Pekerjaan Orang Tua. Pedagang)  
 =  $24/155 = 0,154839$   
 P (Class. Tidak Layak | Pekerjaan Orang Tua. Penjahit)  
 =  $1/155 = 0,006452$   
 P (Class. Tidak Layak | Pekerjaan Orang Tua. Petani)  
 =  $32/155 = 0,206452$   
 P (Class. Tidak Layak | Pekerjaan Orang Tua. Swasta)  
 =  $60/155 = 0,387097$   
 P (Class. Tidak Layak | Pekerjaan Orang Tua. Wiraswasta)  
 =  $35/155 = 0,225806$   
 P (Class. Tidak Layak | Pekerjaan Orang Tua. PNS)  
 =  $2/155 = 0,012903$   
 P (Class. Tidak Layak | Pekerjaan Orang Tua. Polisi)  
 =  $1/155 = 0,225806$   
 P (Class. Tidak Layak | Penghasilan Orang Tua. Rendah)  
 =  $87/155 = 0,56129$   
 P (Class. Tidak Layak | Penghasilan Orang Tua. Sedang)  
 =  $46/155 = 0,296774$   
 P (Class. Tidak Layak | Penghasilan Orang Tua. Tinggi)  
 =  $16/155 = 0,103226$   
 P (Class. Tidak Layak | Penghasilan Orang Tua. Sangat Tinggi) =  $6/155 = 0,03871$   
 P (Class. Tidak Layak | Status. Tidak)  
 =  $151/155 = 0,974194$   
 P (Class. Tidak Layak | Status. Yatim/Piatu)

$= 4/155 = 0,025806$   
 P (Class. Tidak Layak | Nilai. Baik)  
 $= 42/155 = 0,270968$   
 P (Class. Tidak Layak | Nilai. Sangat Baik)  
 $= 113/155 = 0,729032$   
 P (Class. Tidak Layak | Sikap. Baik)  
 $= 126/155 = 0,812903$   
 P (Class. Tidak Layak | Sikap. Sangat Baik)  
 $= 29/155 = 0,187097$

Karena Probabilitas Class Tidak Layak lebih besar dari Class Layak, maka dapat disimpulkan bahwa data diatas termasuk dalam kategori Tidak layak.

#### Analisis Evaluasi hasil dan Validasi Model

Dari hasil pengujian diatas komparasi dari data mining yaitu algoritma Naïve bayes, bahwa pengukuran akurasi menggunakan confusion matrix terbukti bahwa diperoleh hasil dengan Accuracy sebesar 90%, precision . Hasil pengujian terbaik dalam metode algoritma pada data mining dapat dilihat di tabel 4.10 berikut ini :



Gambar Diagram Batang Nilai Akurasi Algoritma

#### 5.1 Kesimpulan

Kesimpulan dari penelitian ini adalah data calon penerima beasiswa SMK YPM 14 Sumobito Jombang yang memiliki beberapa atribut seperti Jumlah Tanggungan Anak, Jarak Rumah, Pekerjaan Orang Tua, Penghasilan Orang Tua, Status Anak, Nilai dan Sikap. Hal tersebut berpengaruh terhadap keTidak Layakan penerima beasiswa, karena tidak semua calon penerima beasiswa bisa mendapatkannya. Dalam proses data mining, menerapkan sebuah metode untuk mengetahui tingkat keTidak Layakan dan tidak Tidak Layaknya siswa untuk mendapatkan beasiswa, namun tidak bisa dilakukan hanya sekali saja, karena penerima manfaat beasiswa dapat berganti sewaktu waktu seiring semakin berkembangnya tingkat jumlah peserta didik yang ada di SMK YPM 14 Sumobito tersebut. Dari proses tersebut



dapat disimpulkan bahwa jenis Tanggungan orang tua dan Pekerjaan orang tua sangat mempengaruhi tingkat keTidak Layakan sebuah bantuan beasiswa tersebut. Dengan menggunakan tools RapidMiner ditemukan tingkat akurasi 90,48%.

### DAFTAR PUSTAKA

- [1] R. H. Dewi and N. Indrawati, "Rancang Bangun Sistem Pendukung Keputusan Penerimaan Beasiswa Menggunakan Metode Naive Bayes Classifier," *Semin. Nas. ke - 9 Rekayasa Teknol. Ind. dan Inf. Ranc.*, vol. 17, no. 3, pp. 79–84, 2014.
- [2] Titis Eka Setyawati," Perancangan Sistem Pendukung Keputusan Untuk Prediksi Penerima Beasiswa Menggunakan Algoritma Naive Bayes," *Jurnal Siskomti*, Vol. 3 No. 2 Juli 2020
- [3] Joy Nashar Utamajaya and Andi Mentari A.P and Siti Masnunah," Penerapan Algoritma Naive Bayes Untuk Penentuan Calon Penerima Beasiswa PIP Pada SDN 023 Penajam", *Jurnal Sistem Informasi (2019) 3(1) 11-17*