



Exploration of Influential Variables and Accuracy of the Naive Bayes Classifier Algorithm for Classifying the Performance of Elementary School Students

Eksplorasi Variabel Berpengaruh dan Akurasi Algoritma Naive Bayes Classifier untuk Mengklasifikasikan Performa Siswa Sekolah Dasar

**Arini Aha Pekuwali^{1*}, Vidriana Oktoviana Bano²,
Alfred Domu D. Panja³, Fajar Indra Prasetyo⁴**

^{1,3,4}Program Studi Teknik Informatika, Universitas Kristen Wira Wacana Sumba, Indonesia

²Program Studi Pendidikan Biologi, Universitas Kristen Wira Wacana Sumba, Indonesia

E-Mail: ¹arini.pekuwali@unkriswina.ac.id, ²vidri.bano@unkriswina.ac.id,
³alfredpanja03@gmail.com, ⁴eternalsnow333@gmail.com,

Received Dec 07th 2024; Revised Apr 20th 2025; Accepted May 27th 2025; Available Online Jun 23th 2025, Published Jun 23th 2025

Corresponding Author: Arini Aha Pekuwali

Copyright © 2025 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Numeracy literacy enables individuals to use numbers and basic mathematical symbols to address practical challenges in everyday life. Data on mathematical performance among Indonesian students, as measured by the Program for International Student Assessment (PISA) in 2022, indicates that Indonesia ranked 71st out of 81 countries. This low PISA result aligns with the numeracy literacy score from the Minimum Competency Assessment (AKM) in Sumba Timur, which was recorded at 33.37 (on a scale of 0-100) in 2023. The extremely low PISA and AKM scores highlight the weak foundational mathematical skills among children, underscoring the need for early identification at the elementary school level. The growth of data in education and the evolution of modern education systems have driven the use of various data mining techniques to monitor student performance through diverse methods of investigation, analysis, and the uncovering of hidden information within educational systems. Data mining in the context of education is commonly referred to as educational data mining (EDM). This research involves collecting learning outcome data from 4th-grade elementary school students for mathematics subjects, along with several demographic variables. The findings reveal that the variables RT1, RT2, and PTS have a strong correlation with the dependent variable PAS. Using the Naive Bayes algorithm, the constructed model successfully classified student learning performance with an accuracy of 92%.

Keyword: Early Detection, Educational Data Mining, Mathematics, Numeracy Literacy, Student Performance

Abstrak

Literasi numerasi memungkinkan seseorang untuk menggunakan angka dan simbol matematika dasar guna menyelesaikan tantangan praktis dalam kehidupan sehari-hari. Data perkembangan kemampuan matematika di antara siswa Indonesia melalui penilaian *Program for International Student Assessment* (PISA) pada tahun 2022, menunjukkan bahwa Indonesia berada pada posisi ke-71 dari 81 negara. Hasil PISA yang rendah tersebut terkonfirmasi oleh hasil nilai Asesmen Kompetensi Minimum (AKM) literasi numerasi Sumba Timur yang berada di angka 33,37 (skala 0-100) pada tahun 2023. Nilai PISA dan AKM yang sangat rendah menunjukkan rendahnya pondasi kemampuan matematika anak, sehingga perlu adanya pengidentifikasian sejak dini kepada siswa Sekolah Dasar (SD). Perkembangan data dalam konteks pendidikan dan evolusi pendidikan modern telah mendorong penggunaan berbagai teknik data mining untuk memantau performa siswa dengan cara-cara penelusuran yang beragam untuk menganalisis dan menemukan informasi yang tersembunyi dalam sistem pendidikan. Data mining pada data pendidikan biasa disebut dengan *educational data mining* (EDM). Penelitian dilakukan dengan mengumpulkan data hasil belajar siswa SD kelas 4 untuk mata pelajaran matematika dan beberapa data demografis siswa. Melalui penelitian ini diketahui bahwa variabel RT1, RT2, dan PTS memiliki hubungan yang kuat dengan variabel terikat PAS. Model yang dibentuk oleh Algoritma Naive Bayes berhasil mengklasifikasikan performa belajar siswa dengan akurasi sebesar 92%.

Kata Kunci: Educational Data Mining, Literasi Numerasi, Matematika, Pendeteksian Dini, Performa Siswa



1. PENDAHULUAN

Kemampuan dasar dalam literasi numerasi adalah hal esensial untuk dimiliki oleh setiap anak. Literasi numerasi memungkinkan mereka untuk menggunakan angka dan simbol matematika dasar guna menyelesaikan tantangan praktis dalam kehidupan sehari-hari. Keterampilan numerasi yang kurang, dapat langsung terkait dengan risiko ekonomi yang lebih tinggi. Kemampuan matematika pada awal perkembangan anak memiliki korelasi yang signifikan secara statistik dengan status sosial ekonomi individu ketika mereka mencapai usia 42 tahun [1]. Literasi numerasi yang rendah berperan dalam menentukan tingkat pendapatan; individu dewasa yang memiliki tingkat literasi numerasi yang rendah cenderung memiliki peluang kerja yang terbatas, yang mengakibatkan pekerjaan dengan bayaran yang rendah dan berakhir dalam kemiskinan [2].

Tantangan yang berkelanjutan terkait dengan kemampuan matematika tetap menjadi masalah yang dapat menghalangi negara-negara berpendapatan rendah dan menengah dalam pencapaian tujuan pembangunan negara mereka [3]. Data perkembangan kemampuan matematika di antara siswa Indonesia selama tujuh siklus terakhir penilaian *Program for International Student Assessment* (PISA) menunjukkan adanya tren yang tidak stabil. Nilai rata-rata untuk kemampuan matematika dalam PISA adalah 500, sedangkan Indonesia mencatat skor 368 pada tahun 2000, yang naik menjadi 386 pada tahun 2015. Namun, skor tersebut mengalami penurunan menjadi 379 pada tahun 2018. Setelah pandemi, melalui pengukuran PISA 2022, skor literasi numerasi Indonesia mengalami penurunan 13 poin yaitu menjadi 366 [4]. Hal ini menempatkan Indonesia pada peringkat ke- 71 dari 81 negara.

Hasil PISA yang rendah tersebut terkonfirmasi oleh hasil nilai Asesmen Kompetensi Minimum (AKM) literasi numerasi Sumba Timur yang berada di angka 33,37 (skala 0-100) pada tahun 2023. Nilai ini diperoleh dari hasil wawancara sebagai data awal dengan Kepala Bidang Sekolah Dasar Dinas Pendidikan Kepemudaan dan Olahraga Kabupaten Sumba Timur.

Kabupaten Sumba Timur, sebagai salah satu daerah di Provinsi Nusa Tenggara Timur masuk dalam daftar daerah tertinggal. Menurut data Badan Pusat Statistik (BPS) tahun 2023, telah dilakukan pengukuran terhadap penduduk berusia 15 tahun ke atas menghabiskan rata-rata waktu lama sekolah yaitu 7,57 tahun. Artinya, rata-rata penduduk Kabupaten Sumba Timur hanya mengenyam bangku pendidikan sampai pada tingkat SMP kelas 7. Menyoroti hal ini, perlu adanya intervensi yang lebih intensif di tingkat lokal untuk meningkatkan akses dan kualitas pendidikan.

Nilai PISA dan AKM yang sangat rendah menunjukkan rendahnya pondasi kemampuan matematika anak, sehingga perlu adanya pengidentifikasian sejak dini kepada siswa Sekolah Dasar (SD). Perlu dilakukan analisis mendalam melalui data mining untuk melihat pengaruh faktor-faktor terhadap kemampuan matematika anak. Dengan menggunakan teknik *data mining*, nilai matematika siswa SD dapat diprediksi lebih akurat. Dengan pemahaman yang lebih baik tentang kinerja siswa di tingkat sekolah dasar, pemerintah dapat merancang kebijakan pendidikan yang lebih efektif dan berkelanjutan. Dengan demikian, manfaat pendeteksian dini kemampuan matematika siswa tidak hanya terbatas pada tingkat sekolah, tetapi juga memiliki dampak yang lebih luas dalam perumusan kebijakan pendidikan oleh pemerintah.

Perkembangan data dalam konteks pendidikan telah mendorong penggunaan berbagai teknik data mining untuk memantau performa dengan cara-cara penelusuran yang beragam untuk menganalisis dan menemukan informasi yang tersembunyi dalam sistem pendidikan [5]. Data mining pada data pendidikan biasa disebut dengan *Educational Data Mining* (EDM). Hasil yang diberikan oleh EDM dapat menjadi masukan untuk pembuat kebijakan pendidikan memberikan suatu keputusan berbasis data sehingga mendukung tujuan untuk meningkatkan efisiensi dan kualitas pengajaran dan pembelajaran. Kurikulum pendidikan saat ini hanya berorientasi pada kompetensi akhir siswa, kurikulum tidak memperhatikan perbaikan motivasi siswa selama kegiatan belajar mengajar. Oleh sebab itu, kontribusi riset ini memberikan model untuk evaluasi sistem pembelajaran siswa.

Permasalahan yang dapat diatasi melalui pendekatan EDM adalah evaluasi, penilaian, dan pemantauan terhadap proses pembelajaran siswa, termasuk di dalamnya adalah prediksi kelulusan siswa [6]. Prediksi kelulusan yang dilakukan akan menghasilkan prediksi status ‘lulus’ atau ‘tidak lulus’ untuk setiap siswa pada mata pelajaran. Prediksi kelulusan siswa diharapkan juga dapat menjadi peringatan dini untuk siswa yang terprediksi tidak lulus agar lebih mengantisipasi dan berhati-hati apabila proses belajar yang dilakukan belum maksimal. Prediksi siswa yang tidak lulus pun dapat menjadi dasar untuk guru dan sekolah dalam memberikan stimulus positif agar siswa tidak putus ada. Pemberian stimulus di tengah masa akademik diharapkan dapat mengintervensi performa siswa sehingga dapat mengubah hasil prediksi.

Penelitian terdahulu mengenai literasi numerasi matematika siswa SD di Sumba Timur [7]. Penelitian ini menunjukkan bahwa literasi numerasi siswa SD di Sumba Timur tergolong rendah. Prediksi kelulusan matematika siswa menggunakan EDM mencapai akurasi 90,89% [6]. Studi lain memprediksi hasil belajar mahasiswa dengan akurasi 94,24% menggunakan *Naive Bayes Classifier* (NBC) [8]. Langkah awal peningkatan literasi numerasi siswa SD di Sumba Timur adalah eksplorasi variabel dengan NBC sebagai metode EDM.

2. METODOLOGI PENELITIAN

Eksplorasi variabel yang mempengaruhi performa siswa dalam pembelajaran matematika melibatkan siswa kelas IV SD yang tersebar di tiga zona di Kabupaten Sumba Timur. SD yang berada di Zona 1 merupakan sekolah yang terdapat di Kecamatan Kota Waingapu dan Kecamatan Kambera. Sekolah-sekolah zona 1 mudah dijangkau dan terjadi persaingan antara sekolah yang berkualitas tinggi. Zona 2 merupakan daerah pinggiran kota, bisa merupakan daerah ibukota kecamatan. Daerah masih mudah dijangkau, namun taraf ekonomi orang tua siswa sudah mulai beragam. Sementara sekolah yang berada di zona 3 merupakan sekolah yang berada di daerah ibukota kecamatan, sekolah sulit dijangkau karena akses jalan yang tidak begitu baik.

2.1. Faktor-faktor yang mempengaruhi kemampuan matematika

Faktor-faktor yang mempengaruhi kemampuan matematika siswa yang digunakan dalam penelitian ini ditampilkan pada Table 1.

Table 1. Faktor-faktor yang mempengaruhi kemampuan matematika siswa

No	Jenis	Faktor	Referensi
1	Faktor Demografis	Jenis Kelamin	[9]
		Status sosial ekonomi	[10]
		Level Pendidikan Orang Tua	[11]
2	Faktor Individu	Bahasa yang digunakan di rumah	[11]
		Kemampuan aritmatika	[12]
3	Faktor Instruksional	Kompetensi guru	[13]
		Kurikulum sekolah	[14]
		Fasilitas sekolah	[15]

Pengambilan data menggunakan format data sebagai berikut:

1. Format data siswa mata pelajaran matematika ditampilkan pada Gambar 1, dalam format ini akan menampung data nama siswa, jenis kelamin, status sosial ekonomi orang tua siswa, level pendidikan orang tua, dan bahasa yang digunakan di rumah. Data siswa mata pelajaran matematika yang dikumpulkan yaitu data siswa kelas IV semester Gasal Tahun akademik 2022/2023, TA. 2023/2024, dan TA. 2024/2025.

DATA SISWA MATA PELAJARAN MATEMATIKA					
KELAS IV SEMESTER GASAL					
TAHUN AKADEMIK ...					
NAMA SEKOLAH:					
NO	NAMA SISWA	JENIS KELAMIN	STATUS SOSIAL EKONOMI	LEVEL PENDIDIKAN ORANG TUA	BAHASA DI RUMAH
1	SISWA 1	L / P	BAWAH/ MENENGAH/ ATAS	SD/ SMP/ SMA/ S1/ S2/ S3	BAHASA IBU/ BAHASA INDONESIA/ BAHASA INGGRIS
2	SISWA 2	L / P	BAWAH/ MENENGAH/ ATAS	SD/ SMP/ SMA/ S1/ S2/ S3	BAHASA IBU/ BAHASA INDONESIA/ BAHASA INGGRIS
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
DST	...	...	...	...	...

Gambar 1. Format Data Siswa Mata Pelajaran Matematika

2. Format data nilai siswa mata pelajaran matematika ditampilkan pada Gambar 2, dalam format ini akan menampung data nama siswa, nilai tugas sebelum penilaian tengah semester, penilaian tengah semester, nilai tugas sesudah penilaian akhir semester, dan penilaian akhir semester. Data nilai siswa mata pelajaran matematika yang dikumpulkan yaitu data siswa kelas IV semester Gasal Tahun akademik 2022/2023, TA. 2023/2024, dan TA. 2024/2025.

DATA NILAI SISWA MATA PELAJARAN MATEMATIKA

KELAS IV SEMESTER GASAL

TAHUN AKADEMIK ...

NAMA SEKOLAH :

NAMA GURU :

LATAR PENDIDIKAN GURU :

BERAPA LAMA MENGAJAR :

KURIKULUM SEKOLAH :

MEDIA PEMBELAJARAN :

NO	NAMA SISWA	TUGAS				PTS	TUGAS				PAS
		1	2	...	N		1	2	...	N	
1	SISWA 1	...	...	...	...	...	...	...	...	...	...
2	SISWA 2	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
DST	...	...	...	...	...	...	...	...	...	...	...

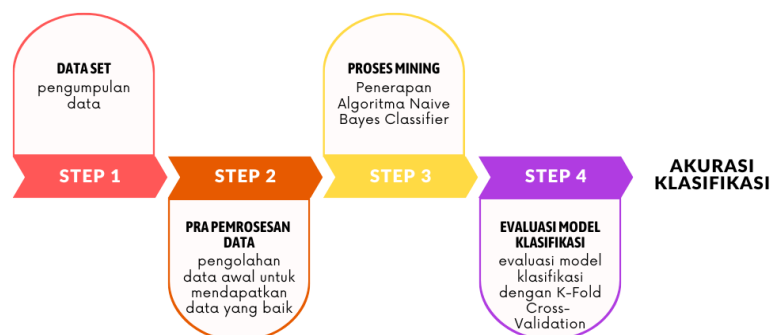
Gambar 2. Format Data Nilai Siswa Mata Pelajaran Matematika

Pengambilan sampel dilakukan di sepuluh Sekolah Dasar yang tersebar pada 3 Zona.

Table 2. Pengumpulan sampel data siswa

No	Zona	Nama Sekolah	Lokasi Sekolah	Jumlah Data Nilai Siswa (Tahun Akademik)		
				2022/ 2023	2023/ 2024	2024/ 2025
1	1	SD Negeri Waingapu 1	Kecamatan Kota Waingapu	20	20	20
2		SD Kristen Nasional Plus Kasih Agape	Kecamatan Kota Waingapu	20	20	20
3		SD Inpres Kalu	Kecamatan Kambera	20	20	20
4		SD Masehi Payeti 1	Kecamatan Kambera	20	20	20
5	2	SD Negeri Mondu	Kecamatan Kanatang	20	20	20
6		SD Negeri Taimanu	Kecamatan Kanatang	20	20	20
7		SD Masehi Melolo 1	Kecamatan Umalulu	20	20	20
8		SD Inpres Waimarang	Kecamatan Umalulu	20	20	20
9	3	SD Masehi Rakawatu	Kecamatan Lewa	20	20	20
10		SD Negeri Banggawatu	Kecamatan Tabundung	20	20	20

Setelah dilakukan pengumpulan data. Data yang diperoleh sebanyak 600 dataset siswa untuk mata pelajaran matematika. Tahap selanjutnya dilakukan pra-pemrosesan data atau pengolahan data awal untuk mendapatkan data yang baik sebelum data diolah menggunakan metode *Naive Bayes Classifier*. Setelah dilakukan pra-pemrosesan data, maka data set yang digunakan pada penelitian ini 600 data siswa dengan 16 atribut. Tahap ketiga dilakukan proses mining menggunakan metode *Naive Bayes Classifier* pada tools RapidMiner. Untuk memvalidasi nilai akurasi metode NBC diterapkan teknik *K-Fold Cross Validation* dan hasil akurasi dapat dilihat berdasarkan *Confusion Matrix*. Tahapan penelitian dapat dilihat pada Gambar 3.



Gambar 3. Tahapan penelitian

Atribut Pra Pemrosesan Data ditampilkan pada Table 3.

Table 3. Atribut pra pemrosesan data

No	Atribut	Jenis Data
1	Nama Siswa	Karakter
2	Jenis Kelamin	Kategorikal
3	Status Sosial Ekonomi	Kategorikal
4	Level Pendidikan Orang Tua	Kategorikal
5	Bahasa yang digunakan di rumah	Kategorikal
6	Rata-rata nilai Tugas sebelum PTS (RT1)	Numerik
7	Nilai PTS	Numerik
8	Rata-rata nilai Tugas setelah PTS (RT2)	Numerik
9	Nilai PAS	Numerik

Beberapa penelitian yang telah dilakukan menyatakan bahwa pra pemrosesan data perlu dilakukan untuk mendapatkan data set dengan kualitas yang baik. Seperti yang dilakukan [16] teknik preprocessing dilakukan untuk mendapatkan data dengan kualitas baik. Cara yang dilakukan antara lain *validation* data yaitu untuk menghilangkan pencilan, derau, data yang kosong dan yang inkonsisten. Pra pemrosesan data yang dilakukan oleh peneliti antara lain:

1. Pembersihan data, yaitu menghilangkan data yang kosong dan tidak lengkap. Misalnya, record siswa dengan status berhenti dan non aktif dihapuskan karena mengandung nilai tugas yang tidak lengkap.
2. Reduksi data, dilakukan guna mendapatkan data set dengan *record* dan jumlah atribut yang bersifat informatif saja. Sebagai contoh atribut nama tidak digunakan pada proses mining karena tidak relevan.
3. Transformasi data, digunakan untuk mengubah Nilai Tugas 1 – 6, Nilai PTS, Nilai PAS yaitu nilainya dibuatkan interval yang lebar dan kedalamannya sama.

Setelah dilakukan pra pemrosesan data, maka data set yang digunakan pada proses mining adalah 600 data siswa dengan 8 atribut yang telah dinormalisasi dan *missing value* tidak terdapat pada data set tersebut. Adapun rincian atribut yang digunakan pada proses mining ditampilkan pada Table 4.

Table 4. Atribut proses mining

No	Atribut	Jenis Data
1	Jenis Kelamin	Kategorikal
2	Status Sosial Ekonomi	Kategorikal
3	Level Pendidikan Orang Tua	Kategorikal
4	Bahasa yang digunakan di rumah	Kategorikal
5	Rata-rata nilai Tugas sebelum PTS (RT1)	Numerik
6	Nilai PTS	Numerik
7	Rata-rata nilai Tugas setelah PTS (RT2)	Numerik
8	Nilai PAS	Numerik

2.2. Analisis Data

Teknik analisis data menggunakan korelasional bivariat yang menjelaskan hubungan antara 2 variabel. Tujuan dari Teknik analisis ini untuk mengetahui kekuatan atau bentuk arah hubungan di antara 2 variabel dan besarnya yang disebabkan oleh variabel satu (variabel bebas) terhadap variabel lainnya (variabel terikat) [17]. Tabel Tingkat korelasi dan kekuatan hubungan antar variabel dapat dilihat pada table 5.

Table 5. Tingkat korelasi dan kekuatan hubungan antar variabel

Nilai r_{xy}	Interpretasi (Tingkat Hubungan)
0,00 – 0,20	Antara variabel x dan y terdapat korelasi, tetapi <i>sangat lemah</i> atau <i>sangat rendah</i> , sehingga korelasi tersebut seringkali diabaikan (dianggap tidak ada korelasi antara variabel x dan y)
0,20 – 0,40	Antara variabel x dan y terdapat korelasi yang <i>lemah atau rendah</i>
0,40 – 0,70	Antara variabel x dan y terdapat korelasi yang <i>sedang atau cukup</i>
0,70 – 0,90	Antara variabel x dan y terdapat korelasi yang <i>kuat atau tinggi</i>
0,90 – 1,00	Antara variabel x dan y terdapat korelasi yang <i>sangat kuat atau sangat tinggi</i>

2.3. Kategori hasil pembelajaran di Sekolah

Penilaian hasil pembelajaran di sekolah dapat dikategorikan dalam skala 4 [18]. Table 6 menunjukkan klasifikasi penilaian hasil belajar siswa.

Table 6. Klasifikasi penilaian hasil belajar siswa

Rentangan	Klasifikasi	Kelas
> 3.25 - 4	Sangat Baik (SB)	3
> 2.50 - 3.25	Baik (B)	2
> 1.75 - 2.50	Cukup (C)	1
0.00 - 1.75	Kurang (K)	0

3. HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

Eksplorasi variabel berpengaruh dan akurasi Algoritma NBC untuk mengklasifikasikan performa siswa sekolah dasar di Kabupaten Sumba Timur dilakukan pada 10 sekolah dasar yang tersebar pada 3 zona, disajikan pada Table 7.

Table 7. Pengumpulan sampel data siswa

No	Zona	Nama Sekolah	Lokasi Sekolah	Jumlah Data Nilai Siswa (Tahun Akademik)		
				2022/ 2023	2023/ 2024	2024/ 2025
1	1	SD Negeri Waingapu 1	Kecamatan Kota Waingapu	20	20	20
2		SD Kristen Nasional Plus Kasih Agape	Kecamatan Kota Waingapu	20	20	20
3		SD Inpres Kalu	Kecamatan Kambera	20	20	20
4		SD Masehi Payeti 1	Kecamatan Kambera	20	20	20
5	2	SD Negeri Mondu	Kecamatan Kanatang	20	20	20
6		SD Negeri Taimanu	Kecamatan Kanatang	20	20	20
7		SD Masehi Melolo 1	Kecamatan Umalulu	20	20	20
8	3	SD Inpres Waimarang	Kecamatan Umalulu	20	20	20
9		SD Masehi Rakawatu	Kecamatan Lewa	20	20	20
10		SD Negeri Banggawatu	Kecamatan Tabundung	20	20	20

Pada Table 7 menyebutkan jumlah sekolah dalam penelitian ini sebanyak 10 mewakili 3 zona yang ada pada Kabupaten Sumba Timur. Namun, dalam proses pengambilan data peneliti mengalami beberapa kendala diantaranya:

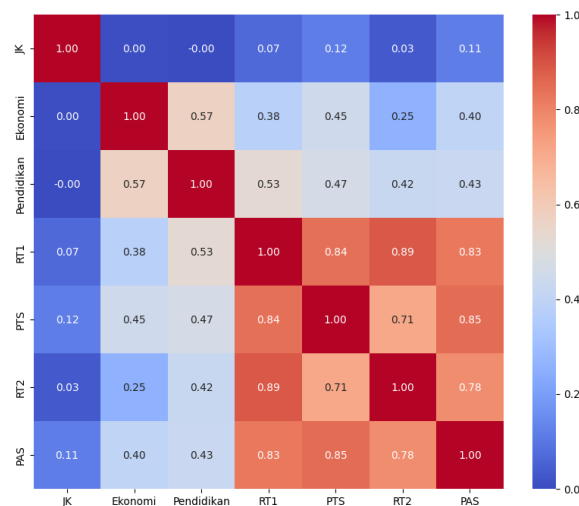
1. Ada bapak/ibu guru yang kehilangan data hasil belajar siswa karena laptop *error* terutama pada tahun ajaran 2022/2023.
2. Ada bapak/ibu guru yang telah berpindah tempat tugas ke sekolah lain, sehingga sekolah tidak mengarsipkan dengan baik data hasil belajar siswa.
3. Kurangnya pengarsipan oleh sekolah dalam merekap hasil belajar siswa dari semua mata pelajaran di setiap semester sepanjang tahun ajaran.

Berdasarkan kendala tersebut, membatasi peneliti untuk mendapatkan data hasil belajar siswa dari 10 sekolah tersebut. Kendala tersebut lebih banyak ditemui pada SD Negeri Mondu, SD Masehi Rakawatu dan SD Inpres Kalu. Peneliti mendapatkan alternatif sekolah lain yang dapat dipakai pada Zona yang sama. Tabel x menunjukkan data hasil belajar siswa yang telah didapatkan 8 dari 10 sekolah yang ditentukan semula.

Table 8. Identitas sekolah tempat pengumpulan data

No	Zona	Nama Sekolah	Lokasi Sekolah	Jumlah Data Nilai Siswa (Tahun Akademik)		
				2022/ 2023	2023/ 2024	2024/ 2025
1	1	SD Negeri Waingapu 1	Kecamatan Kota Waingapu	20	20	20
2		SD Kristen Nasional Plus Kasih Agape	Kecamatan Kota Waingapu	20	20	20
3		SD Masehi Payeti 1	Kecamatan Kambera	20	20	20
4		SD Negeri Poliwali	Kecamatan Nggaha Ori Angu	20	20	20
5	2	SD Negeri Taimanu	Kecamatan Kanatang	20	20	20
6		SD Masehi Melolo 1	Kecamatan Umalulu	20	20	20
7	3	SD Inpres Waimarang	Kecamatan Umalulu	20	20	20
8		SD Inpres Nangga Lindi Watu	Kecamatan Lewa	20	20	20

Informasi pada Gambar 4 menunjukkan rekapan dari 5 sekolah yang mewakili 3 zona.



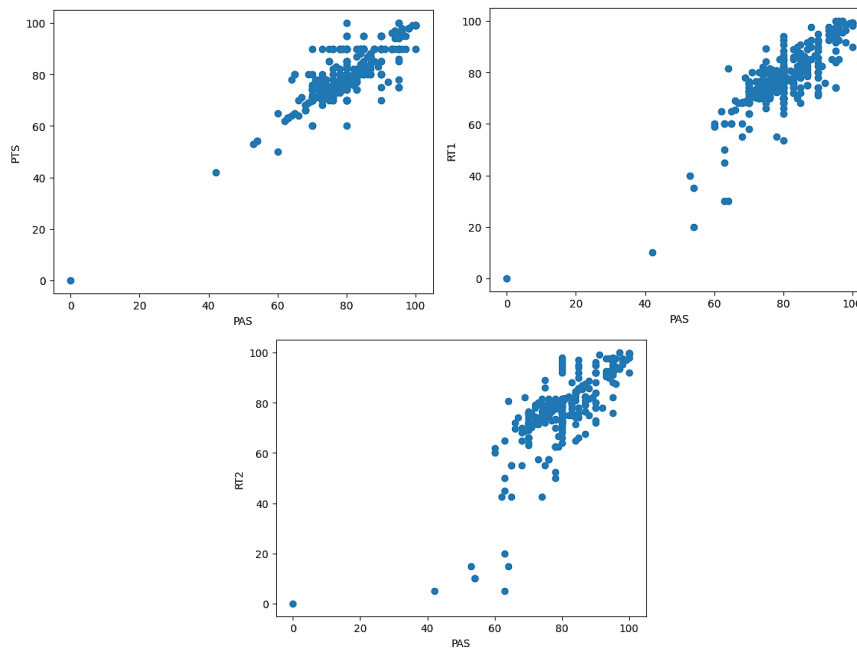
Gambar 4. Hasil data korelasional dari beberapa variabel

Gambar 4 menjelaskan korelasi atau hubungan beberapa variabel dalam penelitian ini yang dapat disimpulkan, sebagai berikut:

1. Pada rentang 0,70 – 0,90 menunjukkan variabel x dan y memiliki korelasi yang kuat atau tinggi. Data ini dapat dilihat pada variable RT1 dengan PAS dengan nilai koefisien korelasi sebesar 0,83. Nilai koefisien korelasi ini menunjukkan arah positif yang artinya terjadi hubungan searah antara variable RT1 dengan PAS, membuktikan bahwa jika nilai RT1 meningkat maka nilai PAS juga meningkat, begitupun sebaliknya. Selain itu, terdapat korelasi yang kuat atau tinggi pada variable PTS dengan PAS, dengan koefisien korelasi sebesar 0,85. Korelasi yang kuat atau tinggi juga terjadi pada variable PTS dengan RT1 dengan koefisien korelasi sebesar 0,84. Sama halnya dengan variabel RT2 dengan PAS yang menunjukkan nilai koefisien korelasi sebesar 0,78.
2. Pada rentang 0,40 – 0,70 menunjukkan variabel x dan y memiliki korelasi yang sedang atau cukup. Data ini terlihat pada variabel PAS dengan Ekonomi, variabel PAS dengan Pendidikan, Variabel PTS dengan Ekonomi, variabel PTS dengan pendidikan, Variabel RT1 dengan pendidikan, dan variabel pendidikan dengan ekonomi. Arah korelasi menunjukkan positif yang berarti menunjukkan hubungan yang searah antara variabel x dan y.
3. Rentang 0,20 – 0,40 menunjukkan variabel x dan y terdapat korelasi yang lemah atau rendah. Data ini terlihat pada variabel RT1 dengan Ekonomi dan variabel RT1 dengan RT2.
4. Rentang 0,00 – 0,20 menunjukkan variabel x dan y terdapat korelasi tetapi sangat lemah atau sangat rendah (bahkan dapat dikatakan tidak ada korelasi antara variabel x dan y). Data ini terlihat pada variabel PAS dengan RT2, RT2 dengan Pendidikan, RT 2 dengan Ekonomi, RT2 dengan PTS, JK dengan PAS, JK dengan PTS, JK dengan RT1, JK dengan RT2, JK dengan Pendidikan dan JK dengan Ekonomi.

Korelasi yang kuat atau tinggi pada rentang 0,70 - 0,90 di atas dapat juga disertai dengan diagram pencar (Scatter plot). Variabel terikat dalam penelitian ini adalah PAS. Scatter plot di bawah ini menunjukkan hubungan antara variabel terikat dengan variabel lainnya yang memiliki hubungan kuat atau tinggi. Variabel yang memiliki hubungan paling kuat dengan PAS adalah PTS, RT1 dan RT2. Hubungan tersebut terlihat pada Gambar scatter plot di bawah ini. Kekuatan hubungan tersebut ditunjukkan dengan plot atau titik-titik yang terkumpul lebih banyak pada 1 bagian tertentu.

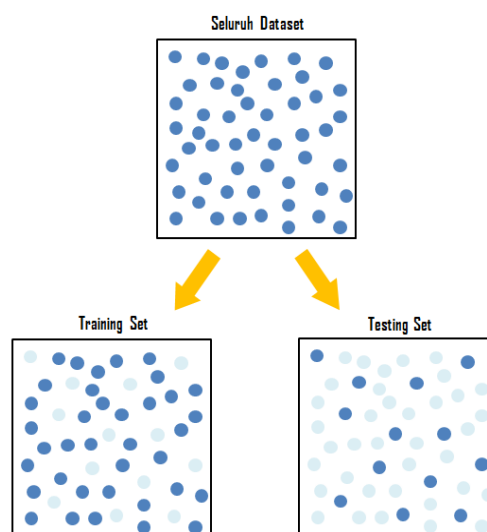
Scatter plot merupakan diagram yang berisi sekumpulan data yang terdiri atas dua variabel dengan nilai kuantitatif, diagramnya dapat dibuat dalam bentuk kumpulan titik yang terpengar pada sumbu koordinat [19]. Selain itu, *scatter plot* dapat juga digunakan untuk mengetahui apakah sebuah variabel memiliki hubungan yang berbanding lurus atau justru berbanding terbalik. Pada Gambar x di atas, menunjukkan hubungan kuat atau tinggi antara PAS dengan PTS, PAS dengan RT1, PAS dengan RT2 yang semuanya menunjukkan arah positif, arah plot atau titik-titik penciran bergerak kearah kanan dan ke arah atas. Jika nilai PAS semakin bertambah (semakin ke kanan) maka nilai PTS, RT1 dan RT2 juga semakin bertambah tinggi (semakin ke arah atas).



Gambar 5. Scatter plot kekuatan hubungan PAS dengan variabel RT1 , RT2 dan PTS.

Pada Gambar 5 menunjukkan variabel RT1 dengan PAS memiliki nilai koefisien korelasi sebesar 0,83 menunjukkan tingkat korelasi yang kuat atau tinggi. Selain itu, terdapat korelasi yang kuat atau tinggi juga pada variabel PTS dengan PAS, dengan koefisien korelasi sebesar 0,85. Korelasi yang kuat atau tinggi juga terjadi pada variabel PTS dengan RT1 dengan koefisien korelasi sebesar 0,84. Sama halnya dengan variabel RT2 dengan PAS yang menunjukkan nilai koefisien korelasi sebesar 0,78. Nilai koefisien korelasi ini menunjukkan arah positif yang menyimpulkan hubungan searah antara variabel RT1 dengan PAS, jika nilai RT1 meningkat maka nilai PAS juga meningkat, begitupun sebaliknya. Hal ini serupa dengan hasil penelitian Adanya hubungan yang signifikan antara *task value* (nilai tugas) siswa dengan nilai ulangan harian mata pelajaran Matematika siswa kelas VB MIN 1 Sintang tahun ajaran 2019/2020 [20]. Hal yang serupa yakni ada korelasi antara nilai tugas dengan nilai raport, ada korelasi antara nilai ulangan tengah semester dengan nilai raport, dan ada korelasi antara nilai akhir semester dengan nilai raport [21]. Hal ini dapat menjelaskan bahwa dengan adanya kenaikan atau penurunan nilai dari nilai ulangan harian, nilai ulangan tengah semester, dan nilai ulangan akhir semester akan diikuti pula dengan kenaikan atau penurunan nilai raport siswa.

Data yang terkumpul berjumlah 358 data. Data tersebut dibagi dalam 2 bagian, yaitu 75% data latih (*training set*) dan 25% data uji (*testing set*), simulasi ditunjukkan pada gambar 6.



Gambar 6. Pembagian data set

Penilaian Akhir Semester (PAS) sebagai variabel terikat dibagi dalam 4 kategori (Tabel Penilaian Hasil Belajar Siswa) yang ditunjukkan pada tabel 9.

Tabel 9. Klasifikasi nilai PAS

Rentangan	Klasifikasi	Kelas	Rentangan Nilai PAS
> 3.25 - 4	Sangat Baik (SB)	3	> 81.25 - 100
> 2.50 - 3.25	Baik (B)	2	> 62.50 - 81.25
> 1.75 - 2.50	Cukup (C)	1	> 43.75 - 62.50
0.00 - 1.75	Kurang (K)	0	0.00 - 43.75

Siswa yang mendapat nilai pada rentangan 81.25 - 100 diklasifikasikan Sangat Baik atau Kelas 3, yang berarti siswa pada kategori ini tidak perlu mendapatkan *treatment* khusus. Siswa dengan klasifikasi Sangat Baik dapat diarahkan untuk belajar mandiri, tanpa bantuan pendampingan khusus dari Guru.

Siswa yang mendapat nilai pada rentangan 62.50 - 81.25 mendapat klasifikasi Baik berada pada kelas 2. Nilai siswa pada rentangan 43.75 - 62.50 mendapat klasifikasi Cukup pada kelas 1. Nilai siswa pada rentangan 0.00 - 43.75 mendapat klasifikasi Kurang pada kelas 0. Siswa yang terklasifikasi Kurang (0), Cukup (1), dan Baik (2) perlu mendapatkan *treatment* khusus. Artinya, siswa pada klasifikasi ini perlu mendapat perhatian khusus dari Guru, contohnya dengan melakukan pendampingan yang lebih keras atau tekun atau kuat atau giat atau bersemangat lagi mendorong mengupayakan sesuatu sehingga memperoleh hasil belajar siswa yang optimal giat atau hebat. Salah satu cara yang dapat dilakukan Guru adalah mendorong peningkatan motivasi belajar siswa. Menumbuhkan motivasi belajar siswa merupakan salah satu teknik dalam mengembangkan kemampuan dan kemauan belajar [22]. Salah satu cara yang logis untuk memotivasi siswa dalam pembelajaran adalah mengaitkan pengalaman belajar dengan motivasi siswa. Guru sebagai orang yang membelajarkan siswa sangat berkepentingan dengan masalah ini, sehingga sebagai guru atau calon guru sebisa mungkin harus selalu berupaya untuk dapat meningkatkan motivasi belajar terutama bagi siswa yang mengalami kesulitan dalam belajar dengan menggunakan berbagai upaya yang dapat dilakukan oleh guru yaitu 1) Memperjelas tujuan yang ingin dicapai. 2) Membangkitkan motivasi siswa. 3) Ciptakan suasana yang menyenangkan dalam belajar. 4) Menggunakan variasi metode penyajian yang menarik. 5) Berilah pujian yang wajar setiap keberhasilan siswa. 6) Berikan penilaian. 7) Berilah komentar terhadap hasil pekerjaan siswa. 8) Ciptakan persaingan dan kerja sama.

Hasil klasifikasi menggunakan Algoritme NBC ditampilkan pada *confussion matrix*, ditunjukkan pada tabel 10.

Tabel 10. Confussion Matrix

		Prediksi			
		0	1	2	3
Aktual	0	0	0	0	0
	1	0	1	0	0
	2	0	1	54	5
	3	0	0	1	28

Data uji kelas 1 diklasifikasikan benar ke kelas 1 berjumlah 1 data. Data kelas 2 yang diklasifikasikan benar sejumlah 54 data dan 1 data diklasifikasikan salah ke kelas 1 serta 5 data diklasifikasikan salah ke kelas 3. Data kelas 3 yang diklasifikasikan benar berjumlah 28 data. Terdapat 1 data kelas 3 yang diklasifikasikan salah ke kelas 2. Sehingga, akurasi dari model pengklasifikasi yang telah terbentuk adalah:

$$akurasi = \frac{\text{Jumlah yang diklasifikasikan dengan benar}}{\text{Total sampel data yang diuji}} \times 100\%$$

$$akurasi = \frac{83}{90} \times 100\%$$

$$akurasi = 92\%$$

Jadi, akurasi hasil klasifikasi adalah 92%. Hal ini menunjukkan hasil uji model klasifikasi hasil belajar siswa di Kabupaten Sumba Timur sudah sangat akurat dengan persentase mendekati 100%. Dengan model ini menjelaskan bahwa NBC dapat digunakan untuk memprediksi performa siswa Sekolah Dasar untuk mata pelajaran Matematika Kabupaten Sumba Timur.

Hasil dari penelitian ini dilanjutkan dengan pengembangan aplikasi EDUMINE. Aplikasi EDUMINE merupakan sebuah sistem berbasis EDM yang mampu mengidentifikasi dan mengevaluasi kesulitan belajar siswa secara dini pada mata pelajaran matematika.

3.2. Diskusi

Akurasi Algoritma *Naive Bayes Classifier* untuk mengklasifikasikan performa siswa Sekolah Dasar di Kabupaten Sumba Timur mencapai 92%, menunjukkan sudah sangat akurat dengan persentase mendekati 100%. Untuk mengklasifikasikan performa siswa Sekolah Dasar selain menggunakan Algoritma *Naive Bayes Classifier*, dapat juga dibandingkan menggunakan metode-metode *Educational Data Mining* (EDM) lainnya.

Implikasi dari penelitian ini memberi informasi bahwa lebih banyak siswa SD di kabupaten ini yang masih perlu didampingi oleh guru secara khusus untuk mencapai KKM mata pelajaran Matematika dibandingkan siswa yang sudah mampu belajar mandiri. Selain itu, hasil penelitian ini juga memberi informasi bagi peneliti khususnya terkait materi-materi ajar Matematika yang masih menjadi “momok” bagi guru untuk menjelaskannya dengan baik kepada siswa SD.

Penelitian ini mempunyai keunggulan yaitu penelitian ini dilakukan di Kabupaten Sumba Timur, salah satu kabupaten yang tergolong daerah 3T (tertinggal, terdepan dan terluar) di Provinsi Nusa Tenggara Timur. Menurut data BPS tahun 2023, telah dilakukan pengukuran terhadap penduduk berusia 15 tahun ke atas menghabiskan rata-rata waktu lama sekolah yaitu 7,57 tahun. Dengan kata lain, rata-rata penduduk Kabupaten Sumba Timur hanya mengenyam bangku pendidikan sampai pada tingkat SMP kelas 7. Menyoroti hal ini, perlu adanya intervensi yang lebih intensif di tingkat lokal untuk meningkatkan akses dan kualitas pendidikan.

Penelitian ini juga dibuat dalam 3 zona tempat yang berbeda-beda sehingga dapat melihat perbedaan dan mengukur perbandingan yang secara signifikan terhadap hasil pencapaian belajar siswa di SD di masing-masing zona. Kebijakan pemerintah perlu dilakukan untuk mengadakan pemerataan jumlah guru di beberapa sekolah khususnya di Zona 2 dan 3 yang masih timpang.

Keterbatasan penelitian ini terletak pada minusnya data penelitian beberapa variabel yang diperoleh dari pihak sekolah, ada beberapa guru SD yang telah pindah tugas ke tempat baru dan sekolah tidak memiliki sistem penyimpanan data siswa dan nilai yang akurat dari setiap guru untuk semua mata pelajaran di sekolah. Hal ini membuat waktu pengumpulan data di lapangan semakin lama. Oleh karena itu, perlu adanya sistem yang dibangun oleh sekolah untuk penyimpanan semua data siswa dan nilai-nilai siswa untuk semua mata pelajaran di setiap semesternya.

4. KESIMPULAN

Melalui penelitian ini dapat disimpulkan bahwa pendidikan orang tua dan perekonomian keluarga cukup berperan dalam mendukung prestasi siswa di sekolah. Penelitian ini juga menghasilkan model berbasis *Naive Bayes Classifier* dengan akurasi 92% yang dapat digunakan untuk mengklasifikasikan performa hasil belajar siswa untuk mata pelajaran matematika. Model ini mempermudah guru dalam menentukan fokus dan langkah strategis untuk meningkatkan hasil belajar siswa secara efektif. Keterbatasan penelitian ini terletak pada minusnya data penelitian beberapa variabel yang diperoleh dari pihak sekolah, ada beberapa guru SD yang telah pindah tugas ke tempat baru dan sekolah tidak memiliki sistem penyimpanan data siswa dan nilai yang akurat dari setiap guru untuk semua mata pelajaran di sekolah. Hal ini membuat waktu pengumpulan data di lapangan semakin lama. Oleh karena itu, perlu adanya sistem yang dibangun oleh sekolah untuk penyimpanan semua data siswa dan nilai-nilai siswa untuk semua mata pelajaran di setiap semesternya.

REFERENSI

- [1] S. J. Ritchie and T. C. Bates, “Enduring Links From Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status,” *Psychol Sci*, vol. 24, no. 7, pp. 1301–1308, 2013, doi: 10.1177/0956797612466268.
- [2] W. Bruine de Bruin and P. Slovic, “Low numeracy is associated with poor financial well-being around the world,” *PLoS One*, vol. 16, no. 11, p. e0260378, Nov. 2021, doi: 10.1371/journal.pone.0260378.
- [3] N. Mabena, P. N. Mokgosi, and S. S. Ramapela, “Factors Contributing To Poor Learner Performance In Mathematics: A Case Of Selected Schools In Mpumalanga Province, South Africa,” *Problems of Education in the 21st Century*, vol. 79, no. 3, pp. 451–466, Jun. 2021, doi: 10.33225/pec/21.79.451.
- [4] E. L. NAPITUPULU, “Skor PISA 2022 Indonesia Turun, Peringkat Naik,” <https://www.kompas.id/baca/humaniora/2023/12/05/skor-pisa-2022-indonesia-turun-peringkat-naik>.
- [5] R. McLeod and G. P. Schell, “Management Information Systems,” *Pearson Prentice Hall*, 2007.
- [6] S. A. A. Kharis, E. Salsabila, and L. D. Haeruman, “Effect of Mathematical Concept Understanding and Mathematical Reasoning on Mathematical Literacy Abilities,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1747/1/012042.
- [7] S. Angu Bima, I. L. Ndakularak, M. E. Nggaba, E. Randjawali, and Y. M. Pakereng, “Evaluating Mathematical Literacy Among Pupils Using Early Grade Mathematics Assessment in East Sumba,” *Jurnal Paedagogy*, vol. 10, no. 2, p. 379, Apr. 2023, doi: 10.33394/jp.v10i2.7418.
- [8] A. A. Pekuwali, “Prediction of student learning outcomes using the Naive Bayesian Algorithm (Case Study of Tama Jagakarsa University),” in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, May 2020. doi: 10.1088/1757-899X/823/1/012056.

- [9] V. Pina, D. Martella, S. Chacón-Moscoso, M. Saracosti, and J. Fenollar-Cortés, "Gender-Based Performance in Mathematical Facts and Calculations in Two Elementary School Samples From Chile and Spain: An Exploratory Study," *Front Psychol*, vol. 12, Aug. 2021, doi: 10.3389/fpsyg.2021.703580.
- [10] J. N. White, "Socioeconomic, Demographic, Attitudinal and Involvement Socioeconomic, Demographic, Attitudinal and Involvement Factors Associated with Math Achievement in Elementary School. Factors Associated with Math Achievement in Elementary School," 2001. [Online]. Available: <https://dc.etsu.edu/etd/77>
- [11] T. R. Ferry, N. A. Fouad, and P. L. Smith, "The Role of Family Context in a Social Cognitive Model for Career-Related Choice Behavior: A Math and Science Perspective," *J Vocat Behav*, vol. 57, no. 3, pp. 348–364, Dec. 2000, doi: 10.1006/jvbe.1999.1743.
- [12] H. Halidin, C. Chairuddin, and I. Purnomo, "Analisis Kemampuan Penalaran Matematis Siswa Smp Pada Materi Aritmatika Sosial Ditinjau Dari Literasi Matematika Siswa," *EMTEKA: Jurnal Pendidikan Matematika*, vol. 4, no. 1, pp. 26–35, Mar. 2023, doi: 10.24127/emteka.v4i1.3448.
- [13] K. Putri, S. Dirgantoro, and U. P. Harapan, "Kompetensi Guru Matematika Dalam Mengembangkan Kompetensi Matematis Siswa." [Online]. Available: <http://www.pikiran-rakyat.com/pendidikan/2016/06/18/peringkat->
- [14] F. A. Wulandari, U. Meilina Kurniawati, and M. A. Rokhimawan, "Problematisasi Mata Pelajaran Matematika Dalam Pembelajaran Tematik Di Sekolah Dasar/ Madrasah Ibtidaiyah," 2020, [Online]. Available: <http://jurnal.umk.ac.id/index.php/RE>
- [15] S. Afsari, S. U. Siregar, and R. D. Harahap, "Pengaruh Manajemen Kelas dan Fasilitas Belajar terhadap Hasil Belajar Matematika Siswa," *Jurnal Basicedu*, vol. 7, no. 1, pp. 535–543, Feb. 2023, doi: 10.31004/basicedu.v7i1.4577.
- [16] M. Zainuddin, "Perbandingan 4 Algoritma Berbasis Particle Swarm Optimization (pso) Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 13, no. 1, 2019.
- [17] A. M. Nasir, "Statistik Pendidikan," Aug. 23, 2018. doi: 10.31227/osf.io/judwx.
- [18] M. Wahyu Suryandi Adam *et al.*, "Pengembangan Instrumen Penilaian Hasil Belajar Menggunakan Computer Based Test di SMA Negeri 10 Sidrap," *Journal on Education*, vol. 06, no. 01, pp. 5560–5573.
- [19] S. Nursalam and M. S. Pd, *Buku Daras Uin Alauddin*. 2015.
- [20] E. Satriya, "Korelasi Antara Task Value (Nilai Tugas) Siswa Dengan Nilai Ulangan Harian," *Jurnal Ilmiah Aquinas*, vol. 4, no. 1, pp. 133–140, Jan. 2021, [Online]. Available: <http://ejournal.ust.ac.id/index.php/Aquinas/index>
- [21] O. : Donny, A. Fahmi, and M. A. Hidayat, "Dengan Nilai Raport Semester Gasal Matematika Siswa Kelas Iv Sd Negeri 1 Bae Kudus," 2014.
- [22] S. Suprihatin, "Upaya Guru Dalam Meningkatkan Motivasi Belajar Siswa," *Jurnal Pendidikan Ekonomi UM Metro*, vol. 3, no. 1, pp. 73–82, Jan. 2015.