

PENERAPAN ALGORITMA *K-NEAREST NEIGHBOR* UNTUK PREDIKSI KELULUSAN SISWA DI SD NEGERI 1 KEDUNGSARI

Application of K-Nearest Neighbor Algorithm to Predict Student Graduation at SD Negeri 1 Kedungsari

Muhammad Affikri^{1*}, Meri Chrismes Aruan²,
Muhamad Irsan³

Program Studi Teknik Informatika
Fakultas Teknik dan Ilmu Komputer
Universitas Indraprasta PGRI

Author Email: muhammadaffikri@gmail.com

Author Email: muhammadaffikri@gmail.com¹,
meriprincess08aruan@gmail.com², atstairway@gmail.com³

Received: September 17, 2025. **Revised:** October 10, 2025. **Accepted:**
October 13, 2025. **Issue Period:** Vol.9 No.4 (2025), Pp: 1388-1402

Abstrak: Kelulusan siswa menjadi indikator penting dalam mengukur efektivitas proses pembelajaran dan keberhasilan lembaga pendidikan. Rendahnya tingkat kelulusan dapat dipengaruhi oleh berbagai faktor, seperti nilai akademik, kehadiran, dan faktor lainnya. Melalui metode algoritma *K-Nearest Neighbor* dengan cara menghitung jarak antara atribut data latih dan atribut data masukan baru yang akan diprediksi menggunakan perhitungan *Euclidean Distance*. Faktor kelulusan yang dipertimbangkan antara itu nilai, kehadiran, ekstrakurikuler, sikap, prestasi, dan pendidikan orang tua. Hasil penelitian menunjukkan bahwa program menghasilkan sistem prediksi dengan keakuratan yang maksimal. Namun masih terdapat beberapa kekurangan program yang perlu diperbaiki seperti penambahan jumlah data, dan juga mempertimbangkan pengaruh faktor lainnya.

Kata kunci: Program Data Mining, *K-Nearest Neighbor*, Prediksi Kelulusan Siswa

Abstract: Student graduation is an important indicator in measuring the effectiveness of the learning process and the success of educational institutions. Low graduation rates can be influenced by various factors, such as academic grades, attendance, and other factors. Using the *K-Nearest Neighbor* algorithm method, the distance between training data attributes and new input data attributes will be predicted using the *Euclidean Distance* calculation. Graduation factors considered include grades, attendance, extracurricular activities, attitudes, achievements, and parental education. The results of the study show that the program produces a prediction system with maximum accuracy. However, there are still several program shortcomings that need to be improved, such as increasing the amount of data and also considering the influence of other factors.

Keywords: Data Mining Program, *K-Nearest Neighbor*, Student Graduation Prediction



DOI: 10.52362/jisamar.v9i4.2065

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

I. PENDAHULUAN

Pendidikan secara umum adalah upaya yang bisa direncanakan dalam mempengaruhi orang lain baik itu secara individu, kelompok ataupun masyarakat sehingga mereka melakukan apa yang telah diharapkan oleh pelaku pendidikan. Pendidikan merupakan salah satu pilar utama dalam pembangunan suatu bangsa. Di Indonesia, upaya peningkatan mutu pendidikan terus dilakukan melalui berbagai kebijakan dan program, mulai dari peningkatan kualitas pengajar hingga pemanfaatan teknologi informasi dalam dunia pendidikan. Tujuan dari pendidikan adalah untuk mendapatkan pengetahuan untuk mengembangkan bakat ataupun kemampuan seseorang serta taat kepada Tuhan Yang Maha Esa untuk dapat meningkatkan pola pikir yang berkualitas yang unggul dalam segala hal[1].

Upaya peningkatan mutu pendidikan sendiri memiliki tujuan yaitu tidak hanya meningkatkan kualitas pendidik tetapi secara langsung juga meningkatkan kualitas lulusan siswa atau peserta didik. Kelulusan dapat diartikan sebagai poin terpenting yang harus dipenuhi siswa dalam proses pembelajarannya. Standar kompetensi lulusan memegang peranan yang sangat penting dalam pengembangan kurikulum tahunan Kompetensi lulusan diukur berdasarkan kemampuan berpikir, bertindak, atau keterampilan untuk menilai pengetahuan dan keterampilan setiap siswa. Standar Kompetensi Lulusan bermanfaat sebagai dasar penilaian dan pemantauan proses kemajuan dan hasil belajar peserta didik. Untuk mencapai standar kompetensi lulusan yang sesuai, maka setiap lulusan harus meningkatkan kompetensinya sesuai dengan perubahan zaman dan kebutuhan[2]. Kelulusan berperan sangat penting untuk menyusun kurikulum setiap tahunnya. Kompetensi lulusan diukur dari kemampuan berfikir, bersikap dan tindakan atau keterampilan setiap peserta didik untuk menilai pengetahuan dan keterampilannya. Standar kompetensi lulusan sering di kenal di kalangan sekolah sebagai Kurikulum. Kurikulum merupakan jalan yang wajib di lalui bagi setiap peserta didik, apabila peserta didik tidak dapat menyelesaikan kurikulum dalam suatu pendidikan maka tidak dinyatakan lulus. Kelulusan pun di tandai dengan pemberian ijazah dari sekolah yang bersangkutan[3].

Kelulusan siswa sendiri menjadi indikator penting dalam mengukur efektivitas proses pembelajaran dan keberhasilan lembaga pendidikan. Rendahnya tingkat kelulusan dapat dipengaruhi oleh berbagai faktor, seperti nilai akademik, kehadiran, kondisi ekonomi keluarga, dan faktor lainnya. Oleh karena itu, diperlukan metode yang dapat membantu pihak sekolah dalam mengidentifikasi siswa yang berpotensi tidak lulus sedini mungkin. Namun, salah satu tantangan yang masih sering dihadapi adalah tingkat kelulusan siswa yang belum merata di berbagai jenjang pendidikan, khususnya pada tingkat sekolah dasar. Pendidikan di sekolah dasar merupakan jenjang pendidikan formal pertama yang akan menentukan arah pengembangan potensi peserta didik [4]. Sekolah dasar yaitu lembaga pendidikan yang menyelenggarakan program pendidikan sebagai dasar untuk mempersiapkan siswanya agar dapat dan mampu melanjutkan pendidikannya ke lembaga pendidikan yang lebih tinggi.

Dalam penelitian ini dilakukan implementasi sistem dengan menerapkan Algoritma *K-Nearest Neighbor* guna untuk memprediksi kelulusan siswa. Algoritma *K-Nearest Neighbor* merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

II. METODE DAN MATERI

2.1 Perumusan Masalah

Penelitian ini diangkat dari kebutuhan akan sistem prediksi kelulusan siswa yang dapat membantu pihak sekolah dan orang tua siswa dalam mengidentifikasi peserta didik yang berisiko tidak lulus lebih awal. Berdasarkan observasi awal dan diskusi dengan pihak sekolah di SD Negeri 1 Kedungsari, ditemukan bahwa penentuan kelulusan masih dilakukan secara manual berdasarkan penilaian guru terhadap nilai akademik dan kehadiran siswa. Proses ini dinilai memerlukan waktu yang lama yaitu sampai selesainya proses belajar mengajar dan cenderung.

Dalam pendidikan di sekolah dasar, kelulusan siswa tidak hanya ditentukan oleh nilai akademik, tetapi juga melibatkan aspek lain seperti sikap, perilaku, dan lingkungan pendukung. Namun, di SD Negeri 1 Kedungsari, proses penentuan kelulusan masih dilakukan secara manual dan cenderung hanya mempertimbangkan nilai akademik, sehingga berisiko menimbulkan ketidaksesuaian kualitas siswa kejenjang yang lebih tinggi.



Indikator penting lainnya seperti kedisiplinan, keaktifan dalam kegiatan ekstrakurikuler, perilaku siswa, prestasi non-akademik, hingga latar belakang pendidikan orang tua belum dianalisis secara sistematis sebagai faktor pendukung kelulusan. Hal ini menunjukkan perlunya pendekatan berbasis data untuk memprediksi kelulusan secara lebih objektif dan menyeluruh. Sebagai dasar untuk mendukung pemecahan masalah yang telah diidentifikasi, dilakukan kajian pustaka terhadap penelitian terdahulu dan teori-teori yang relevan. Penelitian yang dilakukan oleh Supriadi[5]. menunjukkan bahwa metode data mining, khususnya algoritma *K-Nearest Neighbor* (KNN), dapat digunakan untuk memprediksi kelulusan siswa dengan akurasi yang tinggi. Dalam konteks tersebut, data seperti nilai akademik, kehadiran, dan sikap siswa dijadikan sebagai atribut untuk klasifikasi kelulusan.

Selain itu, Wanto & Widiarto[5] menyatakan bahwa prediksi berbasis machine learning menjadi semakin penting dalam dunia pendidikan karena mampu mengolah data historis dan memberikan hasil prediktif yang objektif. Metode KNN bekerja dengan cara membandingkan data baru dengan sejumlah data tetangga terdekat untuk menentukan kelasnya, sehingga sangat cocok diterapkan dalam permasalahan klasifikasi seperti kelulusan siswa. Beberapa faktor lain yaitu variabel-variabel non-akademik seperti perilaku, ekstrakurikuler, serta latar belakang orang tua dapat dijadikan indikator tambahan dalam memodelkan kelulusan. Oleh karena itu, pemanfaatan pendekatan data mining dalam penelitian ini tidak hanya didasarkan pada data nilai semata, melainkan juga pada kombinasi data multidimensional untuk memperoleh hasil prediksi yang lebih akurat dan komprehensif.

2.2 Pengumpulan Data

1. Teknik Pengumpulan Data

Tahap pengumpulan data bertujuan untuk memperoleh informasi yang relevan sebagai dasar dalam membangun sistem prediksi kelulusan siswa menggunakan algoritma *K-Nearest Neighbor* (KNN). Data yang dikumpulkan berupa data akademik dan non-akademik siswa yang berperan sebagai atribut dalam proses klasifikasi.

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah sebagai berikut:

- Observasi: Penulis melakukan observasi secara langsung ke SD Negeri 1 Kedungsari untuk mendapat informasi dan data yang diperlukan dalam penelitian
- Wawancara: Penulis melakukan wawancara secara langsung dengan Ibu Rini Supriyanti selaku operator SD Negeri 1 Kedungsari dengan tujuan memperoleh data tambahan yang didapatkan melalui dokumen resmi sekolah
- Studi Pustaka: Penulis mempelajari referensi-referensi jurnal, buku, maupun penelitian terdahulu yang dinilai relevan dengan penelitian yang sedang dilakukan

2. Teknik Analisis Data

Setelah data diperoleh, dilakukan proses analisis data dengan tahapan sebagai berikut :

- Preprocessing Data
Data cleaning: Menghapus data duplikat, tidak lengkap, atau tidak relevan.
Normalisasi data: Mengubah data ke skala yang seragam agar dapat diproses oleh algoritma KNN.
Transformasi data: Mengonversi data kategorikal menjadi numerik jika diperlukan.
- Pembagian Data
Data latih: Digunakan untuk membentuk model prediksi.
Data uji: Digunakan untuk menguji akurasi model.

3. Evaluasi Model

- Menggunakan confusion matrix untuk membandingkan hasil prediksi dengan data actual
- Mengukur kinerja model menggunakan matriks akurasi, presisi, dan recall

2.3 Penyelesaian Masalah

Dalam penelitian ini, permasalahan utama yang diangkat adalah bagaimana memanfaatkan data siswa untuk memprediksi kelulusan secara lebih objektif dan sistematis. Selama ini, kelulusan siswa lebih banyak ditentukan berdasarkan penilaian akademis dan secara manual oleh guru, sementara aspek lain seperti kehadiran, kedisiplinan, dan latar belakang siswa belum dianalisis secara menyeluruh. Oleh karena itu, dibutuhkan sebuah



pendekatan berbasis data yang mampu mengolah berbagai variabel tersebut untuk menghasilkan prediksi yang akurat.

Untuk menyelesaikan permasalahan tersebut, penelitian ini menggunakan algoritma *K-Nearest Neighbor* (KNN). KNN merupakan salah satu algoritma dalam data mining yang berfungsi untuk melakukan klasifikasi berdasarkan kemiripan atau kedekatan antar data. Algoritma ini bekerja dengan cara mencari sejumlah k data tetangga terdekat dari data baru yang akan diklasifikasikan, lalu menentukan kelas mayoritas dari tetangga tersebut sebagai hasil prediksi.

Algoritma *K-Nearest Neighbor* dipilih penulis sebagai metode dalam penelitian ini karena beberapa pertimbangan, yaitu:

1. Sempel dan efisien: KNN mudah dipahami dan diterapkan, cocok untuk kasus klasifikasi seperti kelulusan (Lulus / Tidak Lulus)
2. Tidak memerlukan model pelatihan yang kompleks: Algoritma ini bersifat *lazy learning*, sehingga proses pelatihan (training) tidak memakan banyak waktu
3. Mampu menangani data multidimensi: KNN dapat mengolah berbagai atribut siswa, baik numerik maupun kategorikal, dengan metode pengukuran jarak seperti *Euclidean Distance*
4. Sudah terbukti dalam penelitian sejenis: Banyak penelitian terdahulu yang berhasil menggunakan KNN untuk klasifikasi pendidikan, seperti prediksi kelulusan dan penjurusan siswa

2.4 Implementasi Algoritma

Sebagai solusi atas permasalahan yang telah dianalisis, penelitian ini mengimplementasikan algoritma *K-Nearest Neighbor* (KNN) untuk membangun sistem prediksi kelulusan siswa di SD Negeri 1 Kedungsari. Data yang digunakan dalam penelitian ini terdiri dari berbagai atribut yang relevan dengan proses kelulusan, di antaranya: nilai akademik siswa, kehadiran, kedisiplinan, keaktifan dalam kegiatan ekstrakurikuler, perilaku selama di sekolah, prestasi, serta latar belakang pendidikan orang tua.

Data ini terlebih dahulu melalui proses pra-pemrosesan, yang mencakup pembersihan data dari nilai kosong atau duplikat, normalisasi agar setiap atribut berada pada skala yang sama, serta konversi data kategorikal menjadi bentuk numerik jika diperlukan. Setelah proses klasifikasi dijalankan, hasil prediksi kemudian dievaluasi menggunakan data uji. Evaluasi dilakukan melalui confusion matrix untuk mengetahui jumlah prediksi benar dan salah.

2.5 Simpulan Penelitian

Penelitian ini telah melalui sejumlah tahapan dimana untuk membangun sistem prediksi kelulusan siswa berbasis algoritma *K-Nearest Neighbor* (KNN). Tahap awal dimulai dengan perumusan masalah, yang didasari oleh kondisi penentuan kelulusan siswa yang masih bersifat manual dan belum memanfaatkan data secara menyeluruh. Selanjutnya, dilakukan pengumpulan data melalui dokumentasi dari sekolah, yang mencakup berbagai atribut seperti nilai akademik, kehadiran, sikap, keaktifan ekstrakurikuler, dan latar belakang keluarga. Data tersebut dianalisis dan dipersiapkan melalui proses pra-pemrosesan, seperti pembersihan data dan normalisasi, guna menghasilkan data yang siap digunakan untuk proses klasifikasi.

Tahapan dilanjutkan dengan analisis penyelesaian masalah, di mana dipilih algoritma KNN sebagai metode klasifikasi yang sesuai berdasarkan karakteristik data dan kebutuhan penelitian. Algoritma ini dipilih karena sederhana, fleksibel, dan dapat bekerja dengan baik meskipun data pelatihan terbatas. Dari keseluruhan tahapan tersebut, dapat disimpulkan bahwa metode penelitian ini telah dirancang secara sistematis untuk mendukung pengembangan sistem prediksi kelulusan siswa secara terarah, mulai dari identifikasi masalah hingga implementasi model prediksi.

2.6 Data Mining

Data Mining merupakan proses iteratif dan interaktif untuk menemukan pola-pola atau model baru yang shahih (sempurna), bermanfaat dan dapat dimengerti dalam suatu database yang besar (massive database). Data mining berisi pencarian trend atau pola yang diinginkan dalam basis data besar dalam membantu pengambilan keputusan di waktu yang akan datang. Data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang dimiliki[6]. Data mining sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar, yang dapat diartikan sebagai pengestrakan informasi baru yang diambil dari bongkahan data besar yang



membantu pengambilan keputusan. Data mining dapat menemukan tren dan pola tersembunyi yang tidak muncul dalam analisis query sederhana sehingga dapat memiliki bagian penting dalam hal menemukan pengetahuan dan membuat keputusan[7]. Berdasarkan kedua pendapat di atas, maka dapat disimpulkan bahwa data mining merupakan suatu proses yang digunakan dalam menemukan hubungan yang berarti, pola dan tren dengan cara mengekstrak pengetahuan dari data dalam jumlah yang besar yang tersimpan dalam database.

Adapun tahap-tahap dalam data mining antara lain:

1. Data Cleaning

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak relevan. Pada umumnya data yang diperoleh, baik dari database memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik.

2. Transformasi Data

Data diubah atau digabung ke dalam format yang sesuai untuk diproses. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data. Pada penelitian ini metode transformasi yang digunakan adalah label encoding dan normalisasi data.

Fungsi dari label encoding adalah mengubah data kategorikal menjadi numerikal dengan setiap kategori memiliki tingkatan nilai. Sedangkan untuk normalisasi dilakukan karena data yang telah dikumpulkan biasanya mempunyai nilai yang tinggi. Maka perlu melakukan normalisasi data guna untuk mengurangi redundansi dan meningkatkan integritas data. Rumus dari normalisasi data yang digunakan penulis sendiri adalah rumus normalisasi min-max sebagai berikut:

$$x_{new} = \frac{x_{old} - x_{min}}{x_{max} - x_{min}}$$

- Keterangan:
1. X new = Nilai X baru
 2. X old = Nilai X lama
 3. X min = Nilai X terkecil
 4. X max = Nilai X terbesar

3. Proses Mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Metode yang dipilih dalam penelitian ini yaitu algoritma *K-Nearest Neighbor*.

4. Evaluasi dan Validasi Pola

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Dalam tahap ini dilakukan korelasi model Pearson dengan rumus:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Keterangan:

- r = Koefisien korelasi
- x_i = Nilai variabel x
- \bar{x} = Rata-rata nilai variabel x
- y = Nilai variabel y (target)
- \bar{y} = Rata-rata nilai variabel y (target)

Dimana korelasi ini digunakan untuk menemukan seberapa kuat hubungan antara atribut dengan target, lalu dilakukan validasi yang hasilnya berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

5. Presentasi Pengetahuan



Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat.

2.7 Algoritma *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* merupakan suatu algoritma klasifikasi dan termasuk ke dalam algoritma supervised learning. Algoritma ini menggunakan klasifikasi terhadap suatu objek berdasarkan data jarak tetangga (neighbor) terdekatnya. Kemiripan data dapat lebih dari satu, maka dari itu KNN dapat mengklasifikasi sejumlah data k yang mirip juga data yang memiliki banyak kemiripan. Jarak yang digunakan pada KNN disebut dengan Jarak Eucliden. Selain sederhana dalam mencari jarak terdekat antar data, Kelebihan algoritma ini dapat menggeneralisasi himpunan data training yang relatif kecil[8].

Algoritma *K-Nearest Neighbor* adalah pendekatan untuk proses penyelesaian kasus dengan menghitung pembobotan pada sejumlah fitur yang telah ditentukan. Pendekatan ini ditentukan dengan menghitung kedekatan antara kasus yang baru dengan kasus yang lama[9]. KNN adalah algoritma nonparametric. Nonparametrik berarti tidak ada parameter atau jumlah parameter tetap terlepas dari ukuran data. Sebaliknya, parameter akan ditentukan oleh ukuran kumpulan data pelatihan, meskipun tidak ada asumsi yang perlu dibuat untuk distribusi data yang mendasarinya. Dengan demikian, KNN bisa menjadi pilihan terbaik untuk setiap studi klasifikasi yang melibatkan sedikit atau tanpa pengetahuan sebelumnya tentang distribusi data[10].

Sehingga dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* (KNN) adalah metode klasifikasi dalam supervised learning yang bekerja dengan mengukur kedekatan antara data baru dan data pelatihan menggunakan jarak *Euclidean*. Sebagai algoritma nonparametrik, KNN tidak memerlukan asumsi terhadap distribusi data, sehingga fleksibel digunakan meski pengetahuan awal terhadap data terbatas. Kelebihannya terletak pada kesederhanaan dan kemampuannya untuk mengklasifikasikan data berdasarkan kemiripan tanpa membutuhkan model pelatihan yang kompleks.

Algoritma *K-Nearest Neighbor* (KNN) memiliki Langkah-langkah dalam penerapannya, yaitu sebagai berikut:

- Tentukan parameter K = jumlah banyaknya tetangga terdekat
- Hitung jarak antara data baru dan semua data yang ada di data training dengan rumus *Euclidean Distance*:

$$d = |x_i - y_i| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Dimana:

d = Jarak

x_i = Data uji

y_i = Data latih

i = Variabel data

n = Banyaknya data

- Urutkan jarak tersebut dan tentukan tetangga mana yang terdekat berdasarkan jarak minimum ke K
- Tentukan kategori dari tetangga terdekat
- Gunakan kategori mayoritas yang sederhana dari tetangga yang terdekat tersebut sebagai nilai prediksi dari data yang baru.

III. PEMBAHASAN DAN HASIL

Pada penelitian yang dilakukan penulis ini, algoritma yang digunakan adalah algoritma *K-Nearest Neighbor* dengan menerapkan perhitungan *Euclidean Distance* untuk memprediksi kelulusan siswa di SD Negeri 1 Kedungsari. Perhitungan algoritma pada penelitian kali ini pun memiliki tahapan berikut :

1. Pengumpulan dan Pengolahan Data

Pada tahapan ini, pengumpulan data dilakukan langsung pada SD Negeri 1 Kedungsari dengan cara observasi langsung dan wawancara dengan operator dari pihak sekolah. Sehingga dari tahapan pengumpulan data ini diperoleh 50 data siswa dengan atribut-atribut sebagai berikut:



DOI: 10.52362/jisamar.v9i4.2065

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

No.	Nama	Nilai	Sekolah	Ekstrakurikuler	Sikap	Prestasi	Pendidikan Orang Tua	Tingkat Kelulusan
1	ANDRONEDA RAQ JAWANIKER	86,77	67,83	Mengikuti	Sangat Baik	Berprestasi	SNP / sederajat	lulus
2	AJHA WILDMAN KHARISMANSER	79,03	67,81	Tidak	Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
3	EGOP PUTRA BESEMAS	81,10	70,30	Mengikuti	Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
4	SEFANUR CRISTOPHER JUNIARSA	89,90	89,82	Mengikuti	Sangat Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
5	Wani Lintang Kusuma	85,17	86,34	Mengikuti	Sangat Baik	Berprestasi	SD	lulus
6	RAMUL ZAZA AHSANI	81,82	88,68	Mengikuti	Baik	Tidak Memenuhi Prestasi	SD	lulus
7	MALINDIA PUTRI MAGANTI	83,88	86,59	Mengikuti	Sangat Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
8	Muhamad Aswadi Nabil	83,81	88,17	Mengikuti	Sangat Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus
9	MUHAMMAD RIZKI ALOTTA	78,50	70,12	Tidak	Kurang Baik	Tidak Memenuhi Prestasi	AAA / sederajat	Tidak Lulus
10	MUHAMMAD ABULLAH SAMEN EL HARI	86,04	80,34	Tidak	Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
11	MUHAMMAD RIZKI SETAWAN	78,11	84,27	Tidak	Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus
12	MARHACHUL ANTO ULYA	81,05	86,17	Mengikuti	Sangat Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
13	MALVA ALIHA SETYAWAN	81,90	88,99	Mengikuti	Sangat Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
14	MARHA ABHITAMA AL AYE	81,60	80,00	Mengikuti	Sangat Baik	Berprestasi	SD	lulus
15	RAVULA	77,04	86,76	Tidak	Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus
16	WANDUK NISAH KHELINA	80,96	88,79	Mengikuti	Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
17	MARVELITA PUTRI WULFA	86,20	86,34	Mengikuti	Sangat Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
18	BAHANI ASA KAHARABA MUMTIZ	80,87	86,76	Mengikuti	Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
19	INDA ALTHAFUNNILA	77,28	86,88	Tidak	Baik	Tidak Memenuhi Prestasi	SD	lulus
20	SADWA	78,20	86,34	Tidak	Baik	Tidak Memenuhi Prestasi	SD	lulus
21	INDA FARIDATI	80,12	87,81	Mengikuti	Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus
22	SH Rafismawati	81,50	80,30	Tidak	Kurang Baik	Tidak Memenuhi Prestasi	SNP / sederajat	Tidak Lulus
23	FANDI RAHMAT SAPUTRA	75,48	80,82	Tidak	Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus
24	WAFU SAM'ANI	80,85	88,79	Mengikuti	Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
25	YUDHA NURYANSA	81,00	80,00	Mengikuti	Sangat Baik	Tidak Memenuhi Prestasi	SD / sederajat	lulus
26	TARISA FARIDA PUTRI	81,84	86,17	Mengikuti	Sangat Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
27	Ahmad Faisal	89,3	92,54	Tidak	Kurang Baik	Tidak Memenuhi Prestasi	SNP / sederajat	Tidak Lulus
28	ADITHA RAHOGA KHARITSA	78,78	80,80	Mengikuti	Baik	Tidak Memenuhi Prestasi	SD / sederajat	lulus
29	MUHAMMAD NUR ROG SEPRAWAN	79,22	80,30	Mengikuti	Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus
30	RAMU LUCIO BANTARO SUPRAWAN	78,00	86,44	Tidak	Baik	Tidak Memenuhi Prestasi	SD	lulus
31	ADITHA	78,22	82,76	Mengikuti	Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus
32	ADITHA KURNIAWAN	78,21	84,28	Mengikuti	Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus
33	Akasya Nadira Ananty Putri	81,22	85,70	Mengikuti	Sangat Baik	Berprestasi	SD / sederajat	lulus
34	Amra Yusuf Nurwanan	80,11	88,38	Tidak	Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
35	BELLA ALIHA	80,67	83,78	Mengikuti	Baik	Tidak Memenuhi Prestasi	SD / sederajat	lulus
36	BELLA GRACIA ALONIA SYHESALU	81,67	81,12	Mengikuti	Baik	Tidak Memenuhi Prestasi	SD	lulus
37	DEWI ALIHA KHORRAMIYA	80,80	87,76	Mengikuti	Baik	Berprestasi	AAA / sederajat	lulus
38	HA SARISA	80,00	80,78	Tidak	Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
39	GRASEDA HEMPA ANTANER	80,11	80,55	Mengikuti	Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
40	MAUSA ARUNA MAHYA	81,80	83,78	Mengikuti	Baik	Tidak Memenuhi Prestasi	SD	lulus
41	MARISA SALSILA	81,20	80,90	Mengikuti	Baik	Tidak Memenuhi Prestasi	SD / sederajat	lulus
42	MUHAMMAD DIMASRA SULZAN HADAR	79,78	86,27	Mengikuti	Baik	Berprestasi	SNP / sederajat	lulus
43	MUHAMMAD RIZI	80,67	87,80	Mengikuti	Sangat Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
44	Nadira Anisa Mahayani	77,89	80,55	Mengikuti	Baik	Tidak Memenuhi Prestasi	AAA / sederajat	lulus
45	NADIA AGANI PUTRI	86,89	88,88	Mengikuti	Baik	Tidak Memenuhi Prestasi	SD / sederajat	lulus
46	NATYAL NIDA PRAMANSARI	81,22	86,88	Tidak	Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus
47	RENAN MELATI	79,84	84,81	Mengikuti	Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus
48	EDWAN PRAMBODO	76,21	80,70	Tidak	Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus
49	SAMA LUTHIANA HOSMAN	80,90	80,76	Tidak	Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus
50	Daniella Anggrani	87,78	82,90	Tidak	Sangat Baik	Tidak Memenuhi Prestasi	SNP / sederajat	lulus

Gambar 1. Data Siswa

2. Transformasi Data

Pada tahapan ini dilakukan normalisasi pada atribut nilai dan kehadiran untuk menyamakan skala fitur dalam rentang antara 0-1 agar nilai yang terlalu besar tidak mendominasi dalam sebuah fitur atau atribut. Sedangkan pada atribut ekstrakurikuler, sikap, prestasi dan pendidikan orang tua akan dilakukan transformasi data untuk mengubah nilai kategorikal menjadi numerical sehingga dataset menjadi seperti gambar berikut:



No	Nama	Nilai	Presensi	Ekstrakurikuler	Sikap	Perilaku	Pendidikan Orang Tua	Status Kelulusan
1	ANDRIANITA FANG ANANVA	0,87	0,79	1	2	1	1	1
2	AGILA WILDAH KHARIMASINDIA	0,66	0,75	0	1	0	1	1
3	EAGY PUTRA OESMAN	0,67	0,33	1	1	0	1	1
4	BENKUN LUYKOTON JAWAJARDA	0,76	0,88	1	2	0	1	1
5	IRINA LITANG FIKUMTA	0,85	0,83	1	2	1	1	1
6	DESSA ZULIA ANJANI	0,68	0,67	1	1	0	1	1
7	MAULIDENIA PUTRI ANADARA	0,72	0,66	1	2	0	1	1
8	MUKHAMMAD NAUFAL NABIR	0,78	0,92	1	2	0	1	1
9	MUHAMMAD RIZKI ADITYA	0,69	0,11	0	0	0	1	1
10	MUHAMMAD ABDULLAH SALIMAN AL FARRIS	0,52	0,33	0	1	0	1	1
11	SALAMAHENDU RIZQI OTTAMAH	0,78	0,65	0	1	0	1	1
12	NANDHETUL ADEL LAYLA	0,70	0,92	1	2	1	1	1
13	NANDA ARIYO SETIAWAN	0,75	0,66	1	2	0	1	1
14	NABILA ABRIYAMA AL ARIF	0,78	1,00	1	2	1	1	1
15	NABILA	0,50	0,88	0	1	0	1	1
16	NANDIA NURUL KHUSRA	0,64	0,66	1	1	0	1	1
17	NOVILZA PUTRI BELAYIA	0,70	0,63	1	2	0	1	1
18	BABAHATA KHARABA MUMTAZ	0,64	0,64	1	1	0	1	1
19	FUDA ALFATHALUNNISA	0,55	0,67	0	1	0	1	1
20	SADENKA	0,58	0,63	0	1	0	1	1
21	DIKKA FARIDISSA	0,61	0,70	1	1	0	1	1
22	SP Rahmawati	0,67	0,13	0	0	0	1	1
23	FANDY RAMHAN SAFURNA	0,68	0,50	0	1	0	1	1
24	WALYO SYA'BANI	0,65	0,66	1	1	0	1	1
25	TUDHA NOVANDHA	0,67	1,00	1	2	0	1	1
26	ZAINNA FARUKA PUTRI	0,62	0,52	1	2	0	0	1
27	Amel Fani	0	0	0	0	0	1	1
28	ADYTA RANGGA SUKUTRA	0,52	0,39	1	1	0	1	1
29	MUKHAMMAD NUFUS SEPTAWAN	0,59	0,34	1	1	0	0	1
30	RANBU LUCI BERTANG SURNAMAN	0,49	0,64	0	1	0	1	1
31	ARIFIN	0,38	0,39	1	1	0	1	1
32	ANINDA KURNIAWAN	0,58	0,63	1	1	0	1	1
33	Almaga Nadia Anantya Putri	1,00	0,17	1	2	0	1	1
34	Anna Yusuf Kurnawan	0,61	0,63	0	1	0	0	1
35	BELLA AGILA	0,78	0,88	1	1	0	1	1
36	BELLA GRACEVIA ANINDA SYLIVIANA	0,60	0,21	1	1	0	0	1
37	DEVA AGILA ANGGRAENIYA	0,67	0,78	1	1	0	1	1
38	FA SAFHA	0,52	0,43	0	1	2	1	1
39	GRISIELA HARIFA ARTANIS	0,78	0,25	1	1	0	1	1
40	MALISA ALVIN MAFIA	0,67	0,88	1	1	0	1	1
41	MARUF SALABILA	0,65	0,29	1	1	0	1	1
42	MUHAMMAD GABRIELA ELLEN HANJAR	0,49	0,63	1	1	0	0	1
43	SALAMAHENDU RIZKI	0,67	0,71	1	0	0	1	1
44	NADIA ANNA MARWAN	0,55	0,25	1	1	0	1	1
45	NANDA ANISA PUTRI	0,61	0,00	1	1	0	1	1
46	NATYLA RIZKA HERNANDEZ	0,78	0,00	0	1	0	0	1
47	BENARA HUSNANA	0,49	0,40	1	1	0	1	1
48	ADYAN PRANAPODI	0,50	0,17	0	1	0	1	1
49	SANIA LUTHIANA FATMAM	0,71	0,66	0	1	0	1	1
50	Saniya Anisa	0,63	0,39	0	0	0	1	1

Gambar 2. Transformasi Data Siswa

3. Perhitungan Algoritma

Pengklasifikasian dengan algoritma *K-Nearest Neighbor* dibagi menjadi dua proses, yaitu proses data latih dan data uji. Pada proses latih, KNN menggunakan data sampel yang terdiri dari variabel-variabel dan kelas target yang diambil dari banyaknya kelas klasifikasi sebagai input. Sedangkan pada proses uji, K-NN menggunakan nilai perhitungan jarak untuk atribut-atribut dari setiap data uji terhadap seluruh atribut pada data latih dengan rumus *Euclidean Distance*.

Hasil pengklasifikasian berdasarkan nilai parameter K akan dibandingkan untuk mendapatkan hasil akurasi terbaik. Selanjutnya dihasilkan sejumlah nilai K tetangga terdekat, dimana hasil dari data uji yang baru diklasifikasikan berdasarkan mayoritas dari kelas kategori pada K tetangga terdekat. Contoh implementasi algoritma KNN dengan menggunakan 50 data siswa yang ada pada Gambar 2.

Berikan label kepada tiap atribut:

- X1 = Rata-rata nilai siswa
- X2 = Presensi siswa
- X3 = Ekstrakurikuler
- X4 = Sikap dan Perilaku
- X5 = Prestasi
- X6 = Latar pendidikan orang tua
- Kelas = Status kelulusan

Selanjutnya masukan data uji yang akan dicari kelasnya yaitu siswa baru dengan nilai atribut yang sudah dinormalisasi dan ditransformasikan datanya, yaitu: X1 = 0,52; X2 = 0,79; X3 = 0; X4 = 1; X5 = 0; X6 = 1.

Sebelum melakukan perhitungan jarak, perlu menentukan nilai K tetangga terdekat yang akan digunakan. Pada permasalahan ini, nilai parameter K yang akan digunakan yaitu K = 3, K = 5, dan K = 7. Tahap berikutnya yaitu menghitung jarak berdasarkan *Euclidean Distance*. Berikut adalah contoh perhitungan data uji dengan data siswa z:

Perhitungan data latih 1 dengan data uji:

$$d = \sqrt{(0,87 - 0,52)^2 + (0,79 - 0,79)^2 + (1 - 0)^2 + (2 - 1)^2 + (1 - 0)^2 + (3 - 1)^2} = 2,668711527$$



Perhitungan data latihan 2 dengan data uji:

$$d = \sqrt{(0,46 - 0,52)^2 + (0,75 - 0,79)^2 + (0 - 0)^2 + (1 - 1)^2 + (0 - 0)^2 + (1 - 1)^2} \\ = 0,068052897$$

Perhitungan data latihan 3 dengan data uji:

$$d = \sqrt{(0,67 - 0,52)^2 + (0,33 - 0,79)^2 + (1 - 0)^2 + (1 - 1)^2 + (0 - 0)^2 + (2 - 1)^2} \\ = 1,493963856$$

Perhitungan data latihan 4 dengan data uji:

$$d = \sqrt{(0,76 - 0,52)^2 + (0,58 - 0,79)^2 + (1 - 0)^2 + (2 - 1)^2 + (0 - 0)^2 + (1 - 1)^2} \\ = 1,448974247$$

Perhitungan data latihan 5 dengan data uji:

$$d = \sqrt{(0,85 - 0,52)^2 + (0,83 - 0,79)^2 + (1 - 0)^2 + (2 - 1)^2 + (1 - 0)^2 + (2 - 1)^2} \\ = 2,027750291$$

Perhitungan data latihan 6 dengan data uji:

$$d = \sqrt{(0,66 - 0,52)^2 + (0,67 - 0,79)^2 + (1 - 0)^2 + (1 - 1)^2 + (0 - 0)^2 + (3 - 1)^2} \\ = 2,244146201$$

Perhitungan data latihan 7 dengan data uji:

$$d = \sqrt{(0,72 - 0,52)^2 + (0,96 - 0,79)^2 + (1 - 0)^2 + (2 - 1)^2 + (0 - 0)^2 + (3 - 1)^2} \\ = 2,463797989$$

Perhitungan data latihan 8 dengan data uji:

$$d = \sqrt{(0,78 - 0,52)^2 + (0,92 - 0,79)^2 + (1 - 0)^2 + (2 - 1)^2 + (0 - 0)^2 + (2 - 1)^2} \\ = 1,756442084$$

Perhitungan data latihan 9 dengan data uji:

$$d = \sqrt{(0,5 - 0,52)^2 + (0,21 - 0,79)^2 + (0)^2 + (0 - 1)^2 + (0 - 0)^2 + (1 - 1)^2} \\ = 1,156952419$$

Perhitungan data latihan 10 dengan data uji:

$$d = \sqrt{(0,52 - 0,52)^2 + (0,33 - 0,79)^2 + (0 - 0)^2 + (1 - 1)^2 + (0 - 0)^2 + (2 - 1)^2} \\ = 1,099325273$$

Perhitungan jarak *Euclidean Distance* dilakukan sampai dengan data latihan ke 50. Setelah jarak data uji ke setiap data latihan sudah diperoleh, tahap selanjutnya adalah mengurutkan setiap nilai jarak tersebut dari yang terkecil (terdekat). Kemudian dilakukan pemeriksaan setiap kelas untuk $K = 3$, $K = 5$ dan $K = 7$ tetangga terdekat. Hasil pengurutan jarak dapat dilihat pada tabel dibawah berikut:

Tabel I. Hasil Pengujian Data

Data	<i>Euclidean Distance</i>	<i>Ranking</i>	Kelas
------	---------------------------	----------------	-------



DOI: 10.52362/jisamar.v9i4.2065

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

1	2,668711527	50	1
2	0,068052897	1	1
3	1,493963856	30	1
4	1,448974247	26	1
5	2,027750291	44	1
6	2,244146201	45	1
7	2,463797989	48	1
8	1,756442084	39	1
9	1,156952419	16	1
10	1,099325273	13	1
11	0,168834833	2	1
12	1,431370511	23	1
13	1,752354027	38	1
14	2,027405295	43	1
15	2,001998266	42	1
16	1,010704434	10	1
17	1,441100229	25	1
18	1,421591136	22	1
19	1,007638503	8	1
20	2,000932262	41	1
21	1,417953961	21	1
22	1,188769709	18	0
23	0,292917695	4	1
24	1,009794049	9	1



25	1,437767001	24	1
26	1,742765672	37	1
27	1,249783918	19	0
28	1,117278868	15	1
29	1,487686915	29	1
30	0,249686155	3	1
31	2,274583888	46	1
32	1,067652052	12	1
33	1,90235529	40	1
34	1,005405944	7	1
35	1,496245211	31	1
36	1,538544508	34	1
37	2,5093005	49	1
38	1,102666384	14	1
39	1,532955124	33	1
40	1,48114385	28	1
41	2,294452975	47	1
42	1,740197827	36	1
43	1,458249589	27	1
44	1,514074393	32	1
45	1,646328966	35	1
46	1,297691158	20	1
47	1,056008629	11	0
48	0,623601392	6	1



49	0,464621518	5	1
50	1,160422876	17	1

Pada tabel I menunjukkan perolehan jarak data uji ke setiap data latih, tahap selanjutnya adalah mengurutkan setiap nilai jarak tersebut dari yang terkecil (terdekat) ke jarak terbesar. Kemudian dilakukan pemeriksaan ke setiap kelas berdasarkan inputan nilai K (tetangga terdekat). Pada penelitian ini nilai parameter K yang digunakan yaitu K = 3, K = 5 dan K = 7, Berikut disajikan hasil klasifikasi berdasarkan nilai K = 3 pada tabel di bawah ini:

Tabel II. Hasil Pengujian Data

<i>Euclidean Distance</i>	<i>Ranking</i>	<i>Kelas</i>	<i>K = 3</i>	<i>K = 5</i>	<i>K = 7</i>
0,068052897	1	1	1	1	1
0,168834833	2	1	1	1	1
0,249686155	3	1	1	1	1
0,292917695	4	1		1	1
0,464621518	5	1		1	1
0,623601392	6	1			1
1,005405944	7	1			1
1,007638503	8	1			
1,009794049	9	1			
1,010704434	10	1			
1,056008629	11	0			
...			
2,668711527	50	1			

Berdasarkan Tabel II kelas yang mayoritas muncul saat K = 3 yaitu kelas 1, pada K = 5 yaitu kelas 1, dan pada K = 7 yaitu kelas 1. Sehingga diperoleh hasil klasifikasi untuk data uji siswa baru tergolong dalam kategori klasifikasi “Status Kelulusan” kelas 1 yang artinya siswa diprediksi “Lulus”.

4. Evaluasi dan Validasi

Pada tahapan ini, akan dilakukan pengevaluasian dan validasi kinerja dari sistem prediksi. Untuk yang pertama akan dilakukan evaluasi dari korelasi atribut dengan target, dimana akan dicari seberapa besar kekuatan atau hubungan antara atribut dengan target. Dengan dataset pada yang ada pada gambar 2 dan diketahui nilai dari rata-rata dari tiap atribut X yaitu: { 0.64, 0.55, 0.66, 1.22, 0.12, 1.5 } dan rata-rata dari atribut Y yaitu : { 0.94 }. Sehingga didapat nilai perhitungan pada gambar berikut:



No.	$x_i - \bar{x}$	$y_i - \bar{y}$								
1	0,23	0,238	0,34	0,78	0,88	1,3	0,06	0,06	0,06	0,06
2	-0,18	0,2	-0,66	-0,22	-0,12	-0,5	0,06	0,06	0,06	0,06
3	0,07	-0,22	0,34	-0,22	-0,12	0,5	0,06	0,06	0,06	0,06
4	0,12	0,03	0,34	0,78	-0,12	-0,5	0,06	0,06	0,06	0,06
5	0,21	0,26	0,34	0,78	0,88	0,5	0,06	0,06	0,06	0,06
6	0,02	0,12	0,34	-0,22	-0,12	1,5	0,06	0,06	0,06	0,06
7	0,08	0,41	0,34	0,78	-0,12	1,5	0,06	0,06	0,06	0,06
8	0,14	0,37	0,34	0,78	-0,12	0,5	0,06	0,06	0,06	0,06
9	-0,14	-0,34	-0,66	-1,22	-0,12	-0,5	0,06	0,06	0,06	0,06
10	-0,12	-0,22	-0,66	-0,22	-0,12	0,5	0,06	0,06	0,06	0,06
11	-0,08	0,08	-0,66	-0,22	-0,12	-0,5	0,06	0,06	0,06	0,06
12	0,06	0,37	0,34	0,78	-0,12	-0,5	0,06	0,06	0,06	0,06
13	0,09	0,41	0,34	0,78	-0,12	0,5	0,06	0,06	0,06	0,06
14	0,14	0,45	0,34	0,78	0,88	0,5	0,06	0,06	0,06	0,06
15	-0,09	0,33	-0,66	-0,22	-0,12	1,5	0,06	0,06	0,06	0,06
16	0	0,55	0,34	-0,22	-0,12	-0,5	0,06	0,06	0,06	0,06
17	0,15	0,28	0,34	0,78	-0,12	-0,5	0,06	0,06	0,06	0,06
18	0	0,55	0,34	-0,22	-0,12	0,5	0,06	0,06	0,06	0,06
19	-0,11	0,12	-0,66	-0,22	-0,12	0,5	0,06	0,06	0,06	0,06
20	-0,08	0,28	-0,66	-0,22	-0,12	1,5	0,06	0,06	0,06	0,06
21	-0,02	0,2	0,34	-0,22	-0,12	0,5	0,06	0,06	0,06	0,06
22	-0,57	-0,22	-0,66	-1,22	-0,12	-0,5	-0,94	-0,94	-0,94	-0,94
23	-0,16	-0,05	-0,66	-0,22	-0,12	-0,5	0,06	0,06	0,06	0,06
24	-0,01	0,55	0,34	-0,22	-0,12	-0,5	0,06	0,06	0,06	0,06
25	0,03	0,45	0,34	0,78	-0,12	-0,5	0,06	0,06	0,06	0,06
26	0,03	0,37	0,34	0,78	-0,12	-1,5	0,06	0,06	0,06	0,06
27	-0,64	-0,3	-0,66	-1,22	-0,12	-0,5	-0,94	-0,94	-0,94	-0,94
28	-0,12	-0,26	0,34	-0,22	-0,12	-0,5	0,06	0,06	0,06	0,06
29	-0,05	-0,22	0,34	-0,22	-0,12	-1,5	0,06	0,06	0,06	0,06
30	-0,15	-0,01	-0,66	-0,22	-0,12	-0,5	0,06	0,06	0,06	0,06
31	-0,08	-0,17	0,34	-0,22	-0,12	1,5	0,06	0,06	0,06	0,06
32	-0,14	-0,13	0,34	-0,22	-0,12	-0,5	0,06	0,06	0,06	0,06
33	0,88	-0,58	0,34	0,78	0,88	-0,5	0,06	0,06	0,06	0,06
34	-0,03	0,28	-0,66	-0,22	-0,12	-1,5	0,06	0,06	0,06	0,06
35	0,14	-0,17	0,34	-0,22	-0,12	0,5	0,06	0,06	0,06	0,06
36	0,05	-0,34	0,34	-0,22	-0,12	-1,5	0,06	0,06	0,06	0,06
37	0,23	-0,17	0,34	-0,22	0,88	1,5	0,06	0,06	0,06	0,06
38	0,09	-0,17	-0,66	-0,22	-0,12	0,5	0,06	0,06	0,06	0,06
39	0,12	-0,3	0,34	-0,22	-0,12	0,5	0,06	0,06	0,06	0,06
40	0,03	-0,17	0,34	-0,22	-0,12	0,5	0,06	0,06	0,06	0,06
41	0,01	-0,26	0,34	-0,22	-0,12	1,5	0,06	0,06	0,06	0,06
42	-0,15	0,08	0,34	-0,22	0,88	-1,5	0,06	0,06	0,06	0,06
43	0,25	0,16	0,34	0,78	-0,12	-0,5	0,06	0,06	0,06	0,06
44	-0,09	-0,3	0,34	-0,22	-0,12	0,5	0,06	0,06	0,06	0,06
45	0,17	-0,55	0,34	-0,22	-0,12	0,5	0,06	0,06	0,06	0,06
46	0,12	-0,55	-0,66	-0,22	-0,12	-1,5	0,06	0,06	0,06	0,06
47	-0,19	-0,09	0,34	-0,22	-0,12	-0,5	-0,94	-0,94	-0,94	-0,94
48	-0,14	-0,38	-0,66	-0,22	-0,12	-0,5	0,06	0,06	0,06	0,06
49	0,09	-0,17	-0,66	-0,22	-0,12	-0,5	0,06	0,06	0,06	0,06
50	0,19	-0,26	-0,66	0,78	-0,12	-0,5	0,06	0,06	0,06	0,06

Gambar 3. Perhitungan Korelasi

Setelah dilakukan perhitungan dengan menggunakan korelasi pearson didapatkan hasil akhir perhitungan sebagai berikut:

Tabel III. Hasil Korelasi pearson

Atribut	$\sum(x_i - \bar{x})(y_i - \bar{y})$	$\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}$	Hasil	Rank
Nilai	1,3886	2,108089	0,659	1
Kehadiran	0,61588	3,399566	0,181	3
Ekstrakurikuler	0,98	5,62498	0,174	4
Sikap	2,66	6,412145	0,415	2
Prestasi	0,36	3,858704	0,093	6
Pendidikan Orang Tua	1,5	10,68691	0,140	5

Sehingga hasil akhir dari perhitungan korelasi adalah membuktikan bahwa atribut dengan hubungan signifikan dengan kelulusan siswa diurutkan sebagai berikut:

- Nilai: Dengan nilai korelasi positif kuat sebesar 0,659 menjadi atribut yang berpengaruh paling besar terhadap kelulusan siswa.



- Sikap: Dengan nilai korelasi positif sedang sebesar 0,415 sehingga disimpulkan sikap semakin sangat baik maka berpengaruh baik terhadap kelulusan.
- Kehadiran: Dengan nilai korelasi positif lemah sebesar 0,181 sehingga disimpulkan presentasi kehadiran punya pengaruh kecil terhadap kelulusan siswa.
- Ekstrakurikuler: Dengan nilai korelasi positif lemah sebesar 0,174 sehingga disimpulkan ekstrakurikuler punya pengaruh kecil terhadap kelulusan siswa.
- Pendidikan Orang Tua: Dengan nilai korelasi positif lemah sebesar 0,140 sehingga disimpulkan Pendidikan orang tua punya pengaruh kecil terhadap kelulusan siswa.
- Prestasi: Dengan nilai korelasi positif sangat lemah sebesar 0,093 sehingga disimpulkan prestasi kurang berpengaruh terhadap kelulusan siswa.

IV. KESIMPULAN

Dari hasil penelitian yang telah dilakukan penulis, dapat diambil kesimpulan sebagai berikut :

1. Sistem prediksi kelulusan siswa SD N 1 Kedungsari dengan menggunakan algoritma *K-Nearest Neighbor* dapat berjalan dengan baik dan menerapkan algoritma dengan maksimal
2. Algoritma *K-Nearest Neighbor* adalah sebuah metode Data Mining yang mengklasifikasikan sebuah data dengan cara mencari tetangga terdekatnya dengan nilai K
3. Semakin besar data yang ada, maka klasifikasi dan prediksi yang dihasilkan akan semakin baik, optimal dan akurat
4. Atribut dengan pengaruh terbesar terhadap kelulusan siswa yaitu nilai dengan koefisien korelasi sebesar = 0.659, diikuti dengan sikap/perilaku dengan koefisien korelasi sebesar = 0.415, absensi dengan koefisien korelasi sebesar = 0,181, ekstrakurikuler dengan koefisien korelasi sebesar = 0.174, Pendidikan orang tua dengan koefisien korelasi sebesar = 0.140, dan dengan nilai koefisien korelasi terendah yaitu prestasi sebagai atribut dengan hubungan terlemah terhadap kelulusan yaitu dengan nilai sebesar = 0.093
5. Dari sistem yang dibuat, yaitu menerapkan algoritma KNN dengan dataset siswa SD N 1 Kedungsari yang memiliki kriteria nilai, kehadiran, ekstrakurikuler, sikap, prestasi dan pendidikan orang tua dengan value yang telah dinormalisasikan dan ditransformasikan. Lalu penentuan nilai tetangga (K) sebesar 3, 5, dan 7 yang dimana mempengaruhi nilai precision, recall dan accuracy. Serta menghitung jarak dengan *Euclidean Distance*. Didapatkan kesimpulan bahwa sistem memiliki keakurasian sebesar 90%. Sehingga sistem dinilai memiliki kemampuan prediksi yang baik dengan akurasi yang hampir sempurna

REFERENSI

- [1] A. S. Arwildayanto and W. T. Sumar, "Analisis kebijakan pendidikan kajian teoretis, eksploratif, dan aplikatif," *Education Policy Analysis: Theoretical, Exploratory, and Application*. Bandung, Indonesia: Cendekia Press, 2018.
- [2] Sani Susanti, Desy Natalia Manurung, Lio Je Cio Ginting, Mildani Ulfa Nazha, and Romauli Siregar, "Kualifikasi Penentuan Kelulusan dan Analisis Penilaian Pendidikan Melalui Kemampuan Peserta Didik," *Semantik : Jurnal Riset Ilmu Pendidikan, Bahasa dan Budaya*, vol. 2, no. 3, pp. 43–50, Jun. 2024, doi: 10.61132/semantik.v2i3.756.
- [3] N. Nurmamaryam and M. Musyarapah, "Standar Kompetensi Lulusan (SKL) Dan Implementasinya Di Madrasah Aliyah Negeri Kapuas (Mata Pelajaran Al-Qur'an Hadis)," *Al Qalam: Jurnal Ilmiah Keagamaan dan Kemasyarakatan*, vol. 16, no. 6, p. 2094, Dec. 2022, doi: 10.35931/aq.v16i6.1390.
- [4] W. Wuryandani, B. Maftuh, and dan Dasim Budimansyah, "PENDIDIKAN KARAKTER DISIPLIN DI SEKOLAH DASAR," 2015.
- [5] D. Supriadi, E. Suryana, J. Meranti Raya Nomor, and S. Lebar Bengkulu, "2023 Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa Sma Negeri 11 Kota Bengkulu," 2023.
- [6] M. Wahyudi, L. Pujiastuti, and S. Solikhun, "Penerapan Data Mining Dalam Mengelompokkan Data Pengangguran Terbuka Menurut Provinsi Menggunakan Algoritma K-Means," in *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 2020, pp. 432–440.
- [7] A. Wanto et al., *Data Mining: Algoritma dan Implementasi*. Yayasan kita menulis, 2020.



DOI: 10.52362/jisamar.v9i4.2065

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

- [8] S. Puspa, “Ahmad Yusuf 3) 1), 3) Sistem Informasi, Fakultas Sains dan Teknologi, UIN Sunan Ampel, Surabaya 2) Matematika, Fakultas Sains dan Teknologi,” 2021.
- [9] M. Kholil, “Seminar Nasional Sistem Informasi dan Teknologi Informasi 2018 SENSITEK 2018 STMIK Pontianak,” 2018.
- [10] R. S. Daulay, “Analisis Kritis dan Pengembangan Algoritma K-Nearest Neighbor (KNN): Sebuah Tinjauan Literatur,” *Jurnal Pendidikan Sains dan Komputer*, vol. 4, no. 02, pp. 131–141, Dec. 2024, doi: 10.47709/jpsk.v4i02.5055.



DOI: 10.52362/jisamar.v9i4.2065

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).