

# Pengelompokan Siswa Smk Diponegoro Berdasarkan Kedisiplinan Pembayaran Spp Menggunakan Algoritma K-Nearest Neighbors + Entropy

Umi Azizah<sup>1</sup>, Dwi Issadari Hastuti<sup>2</sup>, Sahri<sup>3</sup>  
Program Studi Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri  
Jl.Ahmad Yani No.10, Sukorejo, Bojonegoro  
[umi210797@gmail.com](mailto:umi210797@gmail.com)

**Abstract** — Discipline in the payment of tuition fees is an important aspect of school financial management, especially in private educational institutions. This research aims to optimize student grouping based on tuition payment discipline by applying the K-Nearest Neighbors (K-NN) algorithm integrated with the Entropy method as a feature weighting technique. The data used is historical data of tuition payments of students of SMK Diponegoro with attributes such as gender, number of siblings, profession and salary of parents, and status of aid recipients. After going through the preprocessing and normalization stages, the K-NN algorithm is applied with a value of  $K = 10$  and distance calculation using Euclidean Distance. Test results with confusion matrix show an accuracy of 75.86%, recall 80%, and precision 61.54%. This proves that the K-NN algorithm optimized with the Entropy method is quite effective in classifying students into the “Right” or “Late” category in paying tuition fees.

**Keywords**— Classification; Data Mining; Entropy; K-Nearest Neighbors; Tuition Payment.

## I. PENDAHULUAN

Pendidikan memiliki peran penting dalam pembangunan negara. Hal ini sesuai dengan Pasal 27 Undang-Undang Dasar 1945 yang menyatakan bahwa pendidikan bertujuan untuk menyiapkan warga negara yang baik, yaitu warga negara yang mengetahui hak dan kewajibannya [1]. Pendidikan merupakan upaya peningkatan ilmu pengetahuan yang diperoleh melalui lembaga formal dan informal untuk menghasilkan manusia yang berkualitas. Untuk mencapai mutu yang diharapkan, perlu ditetapkan tujuan pendidikan yang tepat. Tujuan pendidikan nasional sebagaimana tercantum dalam pembukaan Undang-Undang Dasar 1945 adalah mencerdaskan kehidupan negara. Kecerdasan yang dimaksud disini bukan hanya kecerdasan yang berorientasi pada kecerdasan intelektual saja, melainkan juga kecerdasan lengkap yang mengandung makna lebih luas [2].

Sekolah sebagai lembaga pendidikan formal berperan strategis dalam membentuk Sumber Daya Manusia (SDM) yang berkualitas. Melalui kurikulum terstruktur, sekolah meningkatkan kompetensi, keterampilan, dan karakter peserta didik agar siap menghadapi dunia kerja serta perkembangan ilmu dan teknologi. Dengan demikian, sekolah mencetak SDM yang berdaya saing dan berkontribusi dalam pembangunan. Sekolah diharapkan mampu memberikan kontribusi positif terhadap pengembangan dan peningkatan sumber daya manusia. Sekolah merupakan lembaga pendidikan yang

diselenggarakan oleh pemerintah dan swasta sebagai tempat terbaik untuk belajar, dengan harapan mampu menciptakan manusia seutuhnya melalui pengembangan kemampuannya. intelektual, potensial, spiritual, personal dan sosial serta terbentuknya manusia yang demokratis dan bertanggung jawab [3]. Sebagaimana lazimnya, dalam suatu lembaga pendidikan seperti sekolah tentu diperlukan administrasi. Administrasi memegang peranan krusial dalam berjalannya proses pendidikan, karena administrasi merupakan salah satu tumpuan guna meningkatkan kualitas pendidikan. Administrasi keuangan juga mempunyai peran yang finansial dan krusial. Dikarenakan administrasi keuangan memiliki tugas untuk mengelola keuangan sekolah agar berjalan dengan baik dan transparan dalam pelaksanaannya, serta memastikan dana digunakan dengan tepat.

SPP (Sumbangan Pembinaan Pendidikan) merupakan iuran rutin sekolah yang pembayarannya dilakukan satu bulan sekali. SPP merupakan salah satu bentuk kewajiban mahasiswa terhadap suatu lembaga pendidikan. Lembaga tersebut akan menggunakan sumbangan tersebut untuk mendanai berbagai kebutuhan sekolah sehingga kegiatan belajar mengajar di sekolah dapat berjalan lancar [4]. SPP (Sumbangan Pembinaan Pendidikan) merupakan proses di mana siswa dan administrasi keuangan sekolah melakukan transaksi dengan tujuan yang sama. SPP (Sumbangan Pembinaan Pendidikan) mempunyai peranan yang sangat penting di setiap sekolah terutama di sekolah swasta, karena

dengan SPP ini sekolah dapat memfasilitasi dan membayar kebutuhan sekolah tersebut [5].

Berdasarkan hasil wawancara dengan Kepala Sekolah dan Bendahara SMK Diponegoro, keterlambatan pembayaran SPP di SMK Diponegoro umumnya disebabkan oleh keterbatasan perencanaan keuangan keluarga, kurangnya pengingat formal, opsi pembayaran yang terbatas, dan sistem informasi keuangan sekolah yang belum terintegrasi, maka dapat disimpulkan bahwa penerapan algoritma K-Nearest Neighbors (K-NN) untuk mengelompokkan siswa berdasarkan kedisiplinan pembayaran SPP memiliki potensi besar untuk meningkatkan efisiensi pengelolaan keuangan sekolah. Dengan mengelompokkan siswa berdasarkan kedisiplinan mereka dalam pembayaran SPP, sekolah dapat memantau lebih baik siswa yang memiliki pola pembayaran tepat waktu serta memberikan perhatian lebih kepada siswa yang sering terlambat atau menunggak.

Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang tersimpan di dalam database besar. Data mining merupakan bagian dari proses KDD (Knowledge Discovery in Databases) yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan, transformasi, data mining, dan evaluasi hasil. KDD (Knowledge Discovery in Databases) secara umum juga dikenal sebagai basis data [6]. Menurut e-book Data Mining : An Overview from a Database Prespective Data mining, juga disebut penemuan pengetahuan dalam bidang basis data, mengacu pada proses non-sepele dalam mengekstraksi informasi tersembunyi, yang sebelumnya tidak diketahui, dan berpotensi berguna (seperti aturan keterbatasan, pengetahuan, keteraturan) dari basis data [7]. Adapun beberapa fungsi dan tugas data mining, Estimasi (Estimastion), Prediksi (Prediction), Klasifikasi (Classification), Klastering (Clustering), Asosiasi (Association) [8].

Klasifikasi merupakan salah satu jenis analisis data yang bertujuan untuk menentukan label kelas dari suatu sampel yang akan dikategorikan. Metode ini termasuk dalam supervised learning, di mana proses pembelajarannya berfokus pada menemukan hubungan antara atribut masukan dan atribut target. Tujuan utama dari klasifikasi adalah meningkatkan keandalan hasil yang diperoleh dari data sehingga dapat digunakan untuk pengambilan keputusan yang lebih akurat [9]. Klasifikasi merupakan salah satu permasalahan dalam data mining yang diselesaikan menggunakan metode supervised learning dan sangat bergantung pada data latih. Dalam data latih, distribusi jumlah data pada setiap kelas umumnya tidak seimbang. Dalam praktiknya, sering kali ditemukan bahwa jumlah sampel dalam masing-masing kelas memiliki perbedaan yang signifikan [10]. Beberapa algoritma klasifikasi yaitu K-Nearest Neighbors, Naive Bayes, Decision Tree, C45, dan sebagainya.

Algoritma K-Nearest Neighbors (K-NN) merupakan salah satu metode dalam data mining yang dikenal sebagai algoritma non-numerik dan dapat digunakan untuk tugas klasifikasi maupun regresi. Dalam proses klasifikasi menggunakan algoritma tertentu, diperlukan suatu dataset

yang terdiri dari data latih (training data) dan data uji (testing data). K-NN termasuk dalam kategori metode supervised learning, yang berarti bahwa algoritma ini memanfaatkan informasi dari data latih untuk mengklasifikasikan suatu objek berdasarkan kedekatan jaraknya. Prinsip utama dari K-NN adalah menentukan jarak antara data uji dan data latih menggunakan metode Euclidean atau metode lain yang relevan, kemudian mengklasifikasikan objek berdasarkan nilai k tetangga terdekatnya[11]. K-Nearest Neighbors (K-NN) merupakan salah satu metode dalam data mining yang banyak digunakan dalam berbagai aplikasi. Algoritma K-NN pertama kali diperkenalkan oleh Fix dan Hodges pada tahun 1951 dan dikenal sebagai metode yang sederhana namun efektif dalam klasifikasi data terawasi (supervised learning). Proses klasifikasi pada K-NN dilakukan dengan membandingkan objek yang akan diklasifikasikan terhadap data latih berdasarkan kedekatan jaraknya. Jarak antara objek dan data latih umumnya dihitung menggunakan metode Euclidean Distance atau metode perhitungan jarak lainnya yang sesuai [12].

Pada penelitian yang dilakukan oleh Maskuri dkk, [13] penelitian ini membahas tentang Penerapan Algoritma K-Nearest Neighbor (K-NN), untuk memprediksi penyakit stroke dengan tujuan untuk memodelkan dataset penyakit stroke menggunakan algoritma K-Nearest Neighbor (K-NN). Dataset yang digunakan dalam penelitian ini berjumlah 100 data, dengan pembagian data latih dan data uji menggunakan metode split validation dengan rasio 80% untuk data latih dan 20% untuk data uji. Pengujian kinerja algoritma K-NN dalam memprediksi penyakit stroke dilakukan dengan menggunakan nilai k sebesar 9, yang menghasilkan tingkat akurasi sebesar 95%.

Pada penelitian yang juga dilakukan oleh Muhammad Rhoisyid dkk, [14] penelitian ini juga membahas tentang Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah di Perguruan Tinggi, Berdasarkan lima kali pengujian algoritma K-Nearest Neighbor (K-NN) dengan variasi nilai k sebesar 3, 5, 7, 9, dan 11, diperoleh hasil bahwa kinerja terbaik dicapai pada k = 3, dengan tingkat akurasi sebesar 52,82%.

Penelitian yang dilakukan oleh Meita Dwi dkk, [15] membahas tentang Algoritma K-Nearest Neighbor Untuk Klasifikasi Ketidaktepatan Waktu Pembayaran Uang SPP, Berdasarkan hasil penelitian mengenai penerapan algoritma K-Nearest Neighbor (K-NN) dalam klasifikasi ketidaktepatan waktu pembayaran SPP di SMK Ma'arif Salam, dapat disimpulkan bahwa metode ini menghasilkan tingkat akurasi sebesar 82,35%. Penelitian ini menggunakan pembagian data latih dan data uji dengan rasio 80% : 20%. Hasil akurasi tersebut menunjukkan bahwa algoritma K-NN dapat diterapkan secara efektif untuk mengidentifikasi ketidaktepatan pembayaran SPP di SMK Ma'arif Salam.

Berdasarkan latar belakang yang telah di jelaskan diatas maka, bagaimana Implementasi Algoritma K-Nearest Neighbors dengan menambahkan metode entropy dalam melakukan klasifikasi terhadap kedisiplinan pembayaran SPP di SMK Diponegoro.

## II. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah klasifikasi menggunakan algoritma K Nearest Neighbors (KNN) dengan pembobotan fitur menggunakan Entropy untuk mengelompokkan pembayaran SPP siswa SMK Diponegoro. Studi kasus meliputi 145 data siswa sebagai data utama.

#### A. Klasifikasi

Klasifikasi merupakan suatu proses dalam analisis data yang bertujuan untuk menemukan pola dan fungsi yang dapat mendeskripsikan serta membedakan satu kelas data dengan kelas lainnya. Proses ini digunakan untuk mengelompokkan data ke dalam kategori yang telah ditentukan sebelumnya. Klasifikasi termasuk ke dalam supervised learning, karena melibatkan penggunaan data yang telah dianalisis sebelumnya sebagai data latih (training data) untuk membangun model yang dapat menguji serta mengklasifikasikan data baru. Proses klasifikasi terdiri dari dua tahap utama, yaitu pembelajaran (training) dan pengujian (testing). Pada tahap pembelajaran, model dibangun menggunakan data latih, sedangkan pada tahap pengujian, model diuji menggunakan data uji guna mengevaluasi tingkat keakuratannya [12].

Menurut Yuli Mardi [8] Dalam klasifikasi, terdapat variabel target kategori, yaitu variabel yang digunakan untuk menentukan kelompok atau kelas dalam suatu dataset. Proses ini bertujuan untuk mengelompokkan data berdasarkan karakteristik tertentu sesuai dengan kategori yang telah ditentukan. Sebagai contoh, dalam analisis pendapatan, klasifikasi dapat digunakan untuk membagi individu ke dalam tiga kategori, yaitu pendapatan tinggi, pendapatan menengah, dan pendapatan rendah, berdasarkan faktor-faktor seperti gaji, pengeluaran, dan aset yang dimiliki.

Selain itu Mardi juga menjelaskan [8], klasifikasi banyak diterapkan dalam berbagai bidang bisnis dan penelitian. Dalam dunia bisnis, metode ini digunakan untuk segmentasi pelanggan, seperti mengklasifikasikan pelanggan berdasarkan tingkat loyalitas atau kebiasaan belanja. Dalam bidang kesehatan, klasifikasi diterapkan untuk mendiagnosis penyakit, misalnya mengelompokkan pasien berdasarkan tingkat risiko suatu penyakit. Sementara itu, dalam bidang keuangan, metode ini dapat digunakan untuk menilai risiko kredit, seperti menentukan apakah seorang pemohon pinjaman termasuk dalam kategori risiko rendah, menengah, atau tinggi. Dengan demikian, klasifikasi merupakan teknik penting dalam analisis data yang memungkinkan pemetaan objek ke dalam kategori yang terdefinisi, sehingga membantu dalam pengambilan keputusan yang lebih terstruktur dan berbasis data.

#### B. Metode Entropy

Metode Entropy merupakan salah satu pendekatan dalam pengambilan keputusan multi-kriteria. Entropy berfungsi untuk mengukur tingkat ketidakpastian atau ketidakaturan dalam suatu data. Dalam konteks pembobotan kriteria, metode ini digunakan untuk menilai sejauh mana suatu kriteria mampu membedakan atau mengklasifikasikan data secara efektif [16]. Tahapan dalam metode entropy merupakan sebagai berikut:

##### 1) Menentukan data awal

Setiap proses pengambilan keputusan memberikan nilai berdasarkan preferensi yang mencerminkan tingkat kepentingan suatu kriteria tertentu.

##### 2) Normalisasi data

Setiap nilai kriteria dinormalisasi menggunakan metode Fuzzy yang telah ditetapkan, di mana hasil normalisasi tersebut dinyatakan sebagai nilai normalisasi  $K_{ij}$ .

Menentukan nilai matriks  $A_{ij}$

$$a_{ij} = \frac{K_{ij}}{\sum_{i=1}^m \sum_{j=1}^n K_{ij}} \quad (1)$$

Keterangan:

- $a_{ij}$  = hasil perhitungan matrix data kriteria
- $K_{ij}$  = Nilai setia kriteria dari Normalisasi data awal
- $i$  = responden ke 1,2,...,i
- $j$  = kriteria ke 1,2,...,j
- $m$  = jumlah pengambilan keputusan
- $n$  = jumlah kriteria

Perhitungan nilai entropy setiap kriteria

$$E_j = \left[ \frac{-1}{\ln m} \sum_{i=1}^m [a_{ij} \ln(a_{ij})] \right] \quad (2)$$

Keterangan:

- $E_j$  = nilai Bobot Entropy
- $\ln$  = nilai log dari total pengambilan keputusan

##### 3) Perhitungan dispersi setiap kriteria

$$D_j = 1 - E_j \quad (3)$$

Keterangan:

- $D_j$  = Nilai dispersi Entropy

##### 4) Normalisasi nilai dispersi

$$W_j = \frac{D_j}{\sum D_j} \quad (4)$$

Keterangan:

- $W_j$  = nilai Normalisasi dispersi (bobot prioritas kriteria)

#### C. Algoritma K-Nearest Neighbors

Algoritma K-Nearest Neighbors (K-NN) merupakan metode klasifikasi yang menentukan kategori suatu data berdasarkan kedekatan jarak terhadap objek data lainnya. Pemilihan nilai K yang optimal bergantung pada karakteristik data yang digunakan. Nilai K yang lebih besar dapat mengurangi pengaruh noise dalam proses klasifikasi, namun di sisi lain, dapat menyebabkan batas antar kelas menjadi kurang tegas. Algoritma K-NN dikategorikan sebagai metode berbasis instance atau non-parametrik dan dianggap sebagai salah satu pendekatan paling sederhana dalam proses data mining [17].

K-Nearest Neighbors (K-NN) merupakan salah satu algoritma dalam kategori lazy learning, yang digunakan untuk mengidentifikasi K objek terdekat atau memiliki kemiripan dengan data uji berdasarkan data baru yang diinputkan [12]. Proses pencarian kedekatan ini dilakukan menggunakan Euclidean Distance, yang dapat dihitung menggunakan persamaan berikut:

$$Euc = \sqrt{((a_1 - b_1)^2 + \dots + (a_n - b_n)^2)} \quad (5)$$

Untuk  $a = a_1, a_2, \dots, a_n$  dan  $b = b_1, b_2, \dots, b_n$  merupakan perumpamaan untuk n nilai atribut dari dua record untuk atribut dengan nilai kategori

Gambar dibawah ini merupakan flowchart pemrosesan algoritma K-NN



Gambar 1. Flowchart Algoritma K-NN

Dimana :

- Memasukkan dataset yang akan di klasifikasikan menggunakan algoritma K-NN
- Menentukan nilai K
- Menghitung Euclidiance Distance dengan rumus sebagai berikut :

$$Euc = \sqrt{\sum_{i=1}^n (pi - qi)^2} \quad (6)$$

Dimana :

- pi : Data training
- qi : Data testing
- n : Dimensi data
- i : Variabel data

- Menyusun jarak yang terbentuk
- Mengelompokkan jarak terdekat
- Menentukan kelas atau label berdasarkan jarak (distance)

Algoritma K-Nearest Neighbor (K-NN) merupakan salah satu algoritma klasifikasi yang bersifat non-parametrik dan berbasis instance. Meskipun memiliki keunggulan dalam hal kesederhanaan dan efektivitas pada berbagai jenis data, algoritma ini juga memiliki keterbatasan yang perlu diperhatikan dalam penerapannya [18]. Berikut merupakan kelebihan dan kekurangan dari algoritma K-NN :

1) *Kelebihan K-Nearest Neighbor (K-NN)*

- Sederhana dan mudah diimplementasikan, karena tidak memerlukan pelatihan model(model free).
- Cocok untuk data non-linear, karena keputusan bergantung pada tetangga terdekat, atau bukan pada fungsi tertentu.
- Fleksibel dalam menangani fitur dengan berbagai jenis nilai.
- Adaptif terhadap data baru, karena data latih dapat secara langsung membangun ulang model.

2) *Kekurangan K-Nearest Neighbor (K-NN)*

- Sensitif terhadap nilai K.

- Tidak efisien untuk data berdimensi tinggi.
- Bergantung pada skala atribut.
- Terkena dampak curse of dimensionality, pada data dimensi tinggi jarak antar titik menjadi kurang bermakna sehingga algoritma kurang efektif.

D. *Confusion Matrix*

Terdapat pendekatan yang lebih komprehensif dan efektif dalam mengevaluasi kinerja suatu sistem klasifikasi, yaitu melalui penggunaan confusion matrix. Confusion matrix merupakan sebuah tabel yang menyajikan informasi mengenai jumlah data uji yang berhasil diklasifikasikan dengan tepat maupun yang mengalami kesalahan klasifikasi [19]. Melalui representasi ini, evaluasi terhadap akurasi dan kinerja sistem klasifikasi dapat dilakukan secara lebih rinci dan menyeluruh. Teknik ini memungkinkan identifikasi titik-titik kelemahan sistem dalam membedakan antar kelas, sehingga memberikan wawasan yang lebih mendalam dibandingkan pengukuran akurasi secara umum. Oleh karena itu, confusion matrix dipandang sebagai alat yang sederhana namun sangat efektif dalam mengukur dan menilai performa suatu algoritma klasifikasi terhadap data uji [20].

Secara umum, evaluasi kinerja suatu algoritma dapat dilakukan dengan menggunakan confusion matrix. Evaluasi ini bertujuan untuk mengukur tingkat akurasi prediksi suatu model dengan membandingkan hasil prediksi kelas terhadap kelas aktual, sehingga dapat diketahui jumlah prediksi yang tepat maupun yang terlambat.

TABEL 1  
CONFUSION MATRIX

Confusion matrix	Prediksi	
	Positif	Negatif
Aktual	TP	FP
	FN	TN

Dengan penjelasan sebagai berikut :

- TP (True Positif), jumlah yang kelas aktual dan prediksi termasuk kelas positif.
- TN (True Negatif), jumlah kelas aktual dan prediksi termasuk kelas negatif.
- FP (False Positif), jumlah kelas aktual termasuk negatif dan kelas prediksi termasuk positif.
- FN (False Negatif), jumlah kelas aktual termasuk positif dan kelas prediksi termasuk negatif

III. HASIL DAN PEMBAHASAN

A. *Pengumpulan Data*

Pengumpulan data pada penelitian ini melibatkan beberapa pihak, seperti murid SMK Diponegoro Purwosari Bojonegoro dan pihak sekolah. Hal ini bertujuan agar mendapatkan data yang relevan dan perizinan dari pihak sekolah terkait. Setelah dilakukannya pengumpulan data, maka diperoleh dataset sebagai berikut:

TABEL 2  
DATASET SISWA

No	Nama	Jenis Kelamin	Saudara	Profesi Ortu	Ketepatan
1	Abi Pradityo Alfiantri	Laki - Laki	3	Petani	Terlambat
2	Achmad Imam Muchlis Bachtiar	Laki - Laki	2	Wiraswasta	Tepat
3	Achmad Junaidi	Laki - Laki	0	Petani	Tepat
4	Ahmad Aditya Khoiri	Laki - Laki	0	Petani	Tepat
5	Ahmad Ainur Roziqin	Laki - Laki	2	Petani	Terlambat
6	Ahmad Fathoni Arzaq	Laki - Laki	0	Petani	Terlambat
7	Ahmad Ferdianto	Laki - Laki	7	Petani	Terlambat
8	Ahmad Huffron Azhari	Laki - Laki	0	Wiraswasta	Tepat
...	...	...	...	...	...
14	Zulul Tristanto	Laki - Laki	0	Pedagang	Tepat
15					

**B. Pre-processing**

Sebelum melakukan pelatihan data, proses pre-processing sangat penting untuk memastikan bahwa data dalam kondisi optimal dan siap digunakan dalam analisis serta perhitungan berdasarkan variabel yang telah ditentukan. Tahapan ini meliputi berbagai langkah, seperti pembersihan data untuk menghapus nilai yang tidak valid atau tidak relevan, normalisasi data agar skala setiap fitur seragam, serta transformasi data jika diperlukan, seperti mengubah data kategorikal menjadi format numerik. Dengan melakukan pre-processing, kualitas data dapat ditingkatkan sehingga menghasilkan model yang lebih akurat dan dapat diandalkan dalam proses klasifikasi atau prediksi. Berikut merupakan tabel pre-processing dan normalisasi setiap fitur.

1) *Jenis Kelamin*

TABEL 3  
NORMALISASI JENIS KELAMIN

No	Jenis Kelamin	Kategori
1	Laki - Laki	1
2	Perempuan	2

2) *Jumlah Saudara*

TABEL 4  
NORMALISASI JUMLAH SAUDARA

No	Jumlah Saudara	Kategori
1	Saudara 0	1

2	Saudara 1 atau 2	2
3	Saudara lebih dari 2	3

3) *Profesi Orang Tua*

TABEL 5  
NORMALISASI PROFESI ORANG TUA

No	Profesi Orang tua	Kategori
1	Petani	1
2	Wiraswasta	2
3	Pedagang	3
4	Buruh Tani	4
5	Lain-lain	5

4) *Gaji Orang Tua*

TABEL 6  
NORMALISASI GAJI ORANG TUA

No	Gaji	Kategori
1	Kurang dari 1 juta	1
2	Lebih dari atau sama dengan 3 juta	3

5) *Penerima Bantuan*

TABEL 7  
NORMALISASI PENERIMA BANTUAN

No	Penerima Bantuan	Kategori
1	Ya	1
2	Tidak	2

6) *Hasil Pre-processing*

TABEL 8  
HASIL PRE-PROCESSING

No	Nama	Jenis Kelamin	Jumlah Saudara	Profesi Ortu	Gaji Ortu	Penerima Bantuan	Tunggal
1	Abi Pradityo Alfiantri	1	2	1	2	1	Terlambat
2	Achmad Imam Muchlis Bachtiar	1	2	2	2	2	Tepat
3	Achmad Junaidi	1	2	1	2	2	Tepat
4	Ahmad Aditya Khoiri	1	2	1	2	1	Tepat
5	Ahmad Ainur Roziqin	1	2	1	3	1	Terlambat
6	Ahmad Fathoni Arzaq	1	2	1	2	1	Terlambat
7	Ahmad Ferdianto	1	3	1	1	1	Terlambat
8	Ahmad Huffron Azhari	1	2	2	2	2	Tepat
..	.....	.....	.....	.....	.....	.....	.....
14	Zulul						Tepat
15	Tristanto	1	2	3	3	2	

**C. Perhitungan Metode Entropy**

Penelitian ini melibatkan sebanyak 145 data sebagai objek penelitian. Tahapan penyelesaian dengan menggunakan metode entropy dijelaskan pada langkah-langkah sebagai berikut.

**1) Menentukan matriks normalisasi (Kij)**

Pada sub bab ini menentukan matriks normalisasi (Kij) dengan rumus sebagai berikut:

$$k_{ij} = \frac{N_{kriteria}}{N_{max}}$$

TABEL 9  
Matriks Normalisasi (Kij)

N o	Nama	Jenis Kelamin	Jumlah Saudara	Profesi Ortu	Gaji Ortu	Penerima Bantuan
1	Abi Pradityo Alfiandri	0,5	0,6666 66667	0,2	0,666 6666 67	0,5
2	Achmad Imam Muchlis Bachtiar	0,5	0,6666 66667	0,4	0,666 6666 67	1
3	Achmad Junaidi	0,5	0,6666 66667	0,2	0,666 6666 67	1
4	Ahmad Aditya Khoiri	0,5	0,6666 66667	0,2	0,666 6666 67	0,5
5	Ahmad Ainur Roziqin	0,5	0,6666 66667	0,2	0,666 6666 67	0,5
6	Ahmad Fathoni Arzaq	0,5	0,6666 66667	0,2	0,666 6666 67	0,5
7	Ahmad Ferdianto	0,5	0,6666 66667	0,2	0,333 3333 33	0,5
8	Ahmad Huffron Azhari	0,5	0,6666 66667	0,4	0,666 6666 67	1
1	ZULUL	0,5	0,6666 33333	0,6	1	1
4	TRISTANTO		66667			
5						
	Total	100,5	97,333 33333	47	92,33 3333 33	101,5

Tabel di atas menyajikan data yang telah dinormalisasi ke dalam matriks (Kij). Proses normalisasi dilakukan dengan membagi setiap nilai pada kriteria tertentu menuju nilai maksimum dari kriteria tersebut. Setelah memperoleh nilai-nilai hasil normalisasi untuk setiap alternatif, langkah selanjutnya adalah menjumlahkan seluruh nilai pada masing-masing kriteria, contoh salah satu perhitungan pada tabel diatas nilai kriteria (Jumlah Saudara)/nilai max (2/3=0,666666667), kriteria (Jenis Kelamin)/nilai max (1/2=0,5), kriteria (Profesi Ortu)/nilai max (1/5= 0,2), kriteria (Gaji Ortu)/nilai max (2/3= 0,666666667).

**2) Menghitung nilai entropy kriteria (Ej)**

Selanjutnya menghitung nilai entropy kriteria (Ej), dengan rumus sebagai berikut:

$$E_j = \left[ \frac{-1}{\ln m} \sum_{i=1}^m [a_{ij} \ln(a_{ij})] \right]$$

TABEL 10  
ENTROPY KRITERIA (Ej)

N o	Nama	Jenis Kelamin	Jumlah Saudara	Profesi Ortu	Gaji Ortu	Penerima Bantuan
1	Abi Pradityo Alfiandri	-	-	-	-	-
		0,026	0,0341	0,023	0,035	0,02617
		38460	34292	23227	60195	3428
		2		9	2	
2	Achmad Imam Muchlis Bachtiar	-	-	-	-	-
		0,026	0,0341	0,040	0,035	0,04551
		38460	34292	56543	60195	7821
		2		3	2	
3	Achmad Junaidi	-	-	-	-	-
		0,026	0,0341	0,023	0,035	0,04551
		38460	34292	23227	60195	7821
		2		9	2	
4	Ahmad Aditya Khoiri	-	-	-	-	-
		0,026	0,0341	0,023	0,035	0,02617
		38460	34292	23227	60195	3428
		2		9	2	
5	Ahmad Ainur Roziqin	-	-	-	-	-
		0,026	0,0341	0,023	0,035	0,02617
		38460	34292	23227	60195	3428
		2		9	2	
6	Ahmad Fathoni Arzaq	-	-	-	-	-
		0,026	0,0341	0,023	0,035	0,02617
		38460	34292	23227	60195	3428
		2		9	2	
7	Ahmad Ferdianto	-	-	-	-	-
		0,026	0,0470	0,023	0,020	0,02617
		38460	357	23227	30331	3428
		2		9	2	
8	Ahmad Huffron Azhari	-	-	-	-	-
		0,026	0,0341	0,040	0,035	0,04551
		38460	34292	56543	60195	7821
		2		3	2	
..	.....	.....	.....	.....	.....	.....
1	Zulul Tristanto	-	-	-	-	-
4		0,026	0,0341	0,055	0,049	0,04551
5		38460	34292	67199	01160	7821
		2		9	9	
	TOTAL	-	-	-	-	-
		4,917	4,9483	4,813	4,946	4,91712
		07364	95961	41920	39985	1876
		3		5	7	
	ENTROPY	0,988	0,9943	0,967	0,993	0,98802
		01219	05948	18439	90486	189
		8		3	1	

Pada tabel diatas menjelaskan hasil penghitungan yang telah didapat pada tabel sebelumnya. Contoh perhitungan pada tabel diatas (Jenis Kelamin)  $0,004975124 * (\ln(0,004975124)) = 0,026384602$ , (JumlahSaudara)  $0,006849315 * (\ln(0,006849315)) = -0,034134292$ . Setelah semua hasil di dapatkan lalu menjumlahkan semua kriteria contoh kriteria Jenis Kelamin =  $-4,917073643$  lalu menghitung Entropy dengan cara  $(-1/\ln(\text{banyaknya kriteria}=145)) \square$  (nilai total= $-4,917073643$ ) =  $0,988012198$ .

**3) Menghitung nilai dispersi setiap kriteria (Dij)**

Berikutnya menghitung nilai dispersi (Dij) dengan rumus sebagai berikut:

$$D_{ij} = 1 - E_j$$

TABEL 11  
Matriks Dispersi (Dij)

Jenis Kelamin	Jumlah Saudara	Profesi Ortu	Gaji Ortu	Penerima Bantuan
---------------	----------------	--------------	-----------	------------------

0,0119878	0,00569405	0,032815	0,006095	
02	2	607	139	0,01197811
0,068570711				

Pada tabel diatas ini menjelaskan hasil dari perhitungan Dij dengan contoh pada kriteria (Jenis Kelamin) 1-(nilai Entropy=0,988012198) = 0,011987802, pada kriteria (Jumlah Saudara) 1-(nilai Entropy=0,994305948) = 0,005694052 kemudian, hasil dari perhitungan nilai Dij dijumlahkan dengan hasil 0,068570711.

4) *Menghitung nilai dispersi (Wj)*

Langkah terakhir yakni menghitung nilai dispersi (Wj), dengan rumus sebagai berikut.

$$IV. W_j = \frac{D_j}{\sum D_j}$$

TABEL 12  
NILAI DISPERSI (Wj)

Jenis Kelamin	Jumlah Saudara	Profesi Ortu	Gaji Ortu	Penerima Bantuan
0,1748239	0,08303912	0,478565	0,088888	
44	8	941	38	0,174682606

Pada tabel diatas merupakan hasil akhir dari perhitungan entropy dengan contoh kriteria (JenisKelamin)/Total(0,011987802/0,068570711=0,174823944)(JumlahSaudara)/Total(0,005694052/0,068570711=0,083039128). Setelah perhitungan nilai dispersi (Wj) selesai, maka perhitungan tersebut akan digunakan sebagai input data baru (data testing) dalam tahap selanjutnya yaitu menggunakan algoritma K-NN.

D. *Perhitungan Algoritma K-Nearest Neighbors*

Pada sub bab ini akan dijabarkan tentang rincian perhitungan yang telah dilakukan pada bab sebelumnya. Pada sub bab ini, setiap tahapan yang sebelumnya dijelaskan secara umum akan dijelaskan berdasarkan hasil yang telah diperoleh dari proses perhitungan manual menggunakan microsoft excel yang telah dilakukan. Berikut disajikan tabel yang memuat hasil perhitungan euclidean distance dan peringkat (rank) masing-masing data (record), dengan jumlah data keseluruhan yaitu 145 data, perhitungan ini menggunakan parameter nilai K yakni K=10

TABEL 13  
HASIL PERHITUNGAN ALGORITMA K-NN

Euclidian Distance	Rank	Termasuk K-NN	Status
0,127755094	14	Tidak	-
0,093174621	7	Ya	Tepat
0,154757447	65	Tidak	-
0,127755094	14	Tidak	-
0,127755094	14	Tidak	-
0,127755094	14	Tidak	-
0,131121619	63	Tidak	-
0,093174621	7	Ya	Tepat
0,210493888	118	Tidak	-
0,211741399	124	Tidak	-
0,130719281	54	Tidak	-

0,032450065	1	Ya	Terlambat
0,097199158	10	Ya	Tepat
.....	.....	.....	.....
.....	.....	.....	.....
0,128166738	42	Tidak	-
0,15758722	96	Tidak	-
0,361719115	138	Tidak	-
0,210493888	118	Tidak	-
0,211741399	124	Tidak	-

E. *Confussion Matrix*

Tabel dibawah merupakan hasil perhitungan berdasarkan perhitungan antara kelas aktual dan kelas prediksi. Tabel ini menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas, sehingga dapat digunakan untuk memicu kinerja klasifikasi model.

TABEL 14  
HASIL PERHITUNGAN CONFUSION MATRIX

Confussion Matrix		
Aktual	Prediksi	
	Tepat	Terlambat
Tepat	8	5
Terlambat	2	14
Total	29	

Setelah memperoleh hasil perhitungan berdasarkan perhitungan antara kelas aktual dan kelas prediksi, langkah selanjutnya adalah menghitung metrik evaluasi kinerja model. Metrik-metrik tersebut meliputi akurasi, presisi, dan recall sejauh mana model mampu melakukan klasifikasi secara tepat, secara keseluruhan maupun pada. Perkiraan metrik ini bertujuan untuk sejauh mana model mampu melakukan klasifikasi secara tepat, baik secara keseluruhan maupun pada masing-masing kelas.

TABEL 15  
HASIL AKURASI

akurasi	0,75862	75,86%
recall	0,8	80,00%
presisi	0,61538	61,54%

$$Recall = \frac{TP}{TP+FN} = \frac{8}{8+2} = 0.8$$

$$Precision = \frac{TP}{TP+FP} = \frac{8}{8+5} = 0.6154$$

$$Accuracy = \frac{TP+TN}{Total} = \frac{8+14}{29} = 0.7586$$

V. KESIMPULAN

Berdasarkan penjabaran yang telah dijelaskan sebelumnya mengenai implementasi algoritma K-Nearest Neighbors (K-NN) dengan penambahan metode Entropy dalam

mengklasifikasikan kedisiplinan pembayaran SPP di SMK Diponegoro, maka dapat disimpulkan bahwa metode tersebut dapat diterapkan dengan baik untuk menyelesaikan permasalahan klasifikasi tepat waktu atau terlambatnya pembayaran SPP siswa. Penerapan tambahan metode Entropy berfungsi sebagai pembobotan fitur untuk mengoptimalkan pengaruh masing-masing atribut dalam proses klasifikasi. Dengan menggunakan nilai  $K = 10$  dan perhitungan jarak, sistem klasifikasi yang dibangun mampu memberikan hasil akurasi sebesar 75,86%, recall sebesar 80%, dan presisi sebesar 61,54%, yang menunjukkan bahwa algoritma K-NN + Entropy cukup efektif dan layak digunakan dalam mengelompokkan siswa berdasarkan pola pembayaran SPP.

#### DAFTAR PUSTAKA

- [1] A. Nazarpour, G. Rostami Paydar, F. Mehregan, S. J. Hejazi, and M. A. Jafari, "Application of geographically weighted regression (GWR) and singularity analysis to identify stream sediment geochemical anomalies, case study, Takab Area, NW Iran," *J. Geochemical Explor.*, vol. 235, no. 1, pp. 103–115, 2022, doi: 10.1016/j.gexplo.2022.106953.
- [2] B. Y. A. Aziizu, "Tujuan Besar Pendidikan Adalah Tindakan," *Pros. Penelit. dan Pengabd. Kpd. Masy.*, vol. 2, no. 2, pp. 295–300, 2015, doi: 10.24198/jppm.v2i2.13540.
- [3] R. Zakhroh, "Pengaruh Kinerja Tenaga Administrasi Sekolah Terhadap Kualitas Layanan Administrasi Non Akademik," *Didaktika*, vol. 19, no. 2, pp. 59–70, 2013.
- [4] Denih, R. G. Wendasmoro, and S. Ramos, "Rancang Bangun Aplikasi Pembayaran SPP Berbasis Web (Studi Kasus: SMK Tri Patria Nusantara Kabupaten Bogor)," *J. Manaj. Inform. Jayakarta*, vol. 2, no. 1, pp. 125–131, 2022.
- [5] D. Oscar, Y. I. Maulana, A. Haidir, and Abdul Ghani Alhaq, "Sistem Informasi SPP Dan Pembayaran Sekolah Berbasis Web Pada Mts AL-Ihsan Pondok Gede Bekasi," *J. Speed - Sentra Penelit. Eng. dan Edukasi*, vol. 11, no. 3, pp. 7–12, 2019, [Online]. Available: <http://ejurnal.net/portal/index.php/speed/article/view/799/730>
- [6] C. Zai, "Implementasi Data Mining Sebagai Pengolahan Data," *J. Teknoinfo*, vol. 16, no. 1, p. 46, 2022, [Online]. Available: <http://portaldata.org/index.php/portaldata/article/view/107>
- [7] M. S. Chen, J. Han, and P. S. Yu, "Data mining: An overview from a database perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 8, no. 6, pp. 866–883, 1996, doi: 10.1109/69.553155.
- [8] Y. Mardi, "Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database ( KDD ) . Jurnal Edik Informatika," *J. Edik Inform.*, 2020.
- [9] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v11i3.2777.
- [10] A. A. Arifiyanti and E. D. Wahyuni, "Smote: Metode Penyeimbang Kelas Pada Klasifikasi Data Mining," *SCAN - J. Teknol. Inf. dan Komun.*, vol. 15, no. 1, pp. 34–39, 2020, doi: 10.33005/scan.v15i1.1850.
- [11] Q. A. A'yuniyah and M. Reza, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru," *Indones. J. Inform. Res. Softw. Eng.*, vol. 3, no. 1, pp. 39–45, 2023, doi: 10.57152/ijirse.v3i1.484.
- [12] F. I. Silfana and M. A. Barata, "Using K-NN Algorithm for Evaluating Feature Selection on High Dimensional Dataset," *J. Tek. Inform.*, vol. 17, no. 2, 2024, doi: <https://doi.org/10.15408/jti.v17i2.40866>.
- [13] M. N. Maskuri, K. Sukerti, and R. M. Herdian Bhakti, "Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke Stroke Disease Predict Using KNN Algorithm," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 1, pp. 130–140, 2022.
- [14] M. R. Akhmad and T. A. Y. Siswa, "Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 185, 2022, doi: 10.35889/progresif.v18i2.921.
- [15] M. D. Anggraeni, K. Kusri, and M. R. Arief, "K-Nearest Neighbor Algorithm for Classification Inaccuracy in Payment of SPP Money," *JTECS J. Sist. Telekomun. Elektron. Sist. Kontrol Power Sist. dan Komput.*, vol. 4, no. 1, p. 67, 2024, doi: 10.32503/jtecs.v4i1.4891.
- [16] Setiawansyah, "Penerapan Metode Entropy dan Grey Relational Analysis dalam Evaluasi Kinerja Karyawan," *J. Data Sci. Inf. Syst.*, vol. 2, no. 1, pp. 29–39, 2024.
- [17] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 6, no. 2, pp. 118–127, 2021, doi: 10.31294/ijcit.v6i2.10438.
- [18] J. Han, M. Kamber, and J. Pei, *Techniques to Improve Classification Accuracy*. 2012.
- [19] W. I. Rahayu, C. Prianto, and E. A. Novia, "Perbandingan Algoritma K-Means dan Naive Bayes untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan pada PT. Pertamina (Persero)," *J. Tek. Inform.*, vol. 13, no. 2, pp. 1–8, 2021.
- [20] R. Nurhidayat and K. E. Dewi, "Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek," *Komputa J. Ilm. Komput. dan Inform.*, vol. 12, no. 1, pp. 91–100, 2023, doi: 10.34010/komputa.v12i1.9458.