

## KLASIFIKASI SEKOLAH MENENGAH PERTAMA/SEDERAJAT WILAYAH BIREUEN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS BERBASIS *WEB*

Rozzi Kesuma Dinata<sup>1</sup>, Fajriana<sup>2</sup>, Zulfa<sup>3</sup>, Novia Hasdyna<sup>4</sup>

<sup>1,2,3</sup> Program Studi Teknik Informatika Universitas Malikussaleh

Jl. Kampus Unimal Bukit Indah, Blang Pulo, Lhokseumawe, Aceh, Indonesia

<sup>4</sup> Program Studi Teknik Informatika Universitas Islam Kebangsaan Indonesia

Jl. Medan – Banda Aceh Sp. Alue Awe, Lhokseumawe, Aceh, Indonesia

<sup>1</sup>rozzi@unimal.ac.id, <sup>2</sup>fajriana@unimal.ac.id, <sup>3</sup>umiulfa5@gmail.com, <sup>4</sup>noviahasdyna@gmail.com

Page | 33

**Abstrak**— Pada penelitian ini diimplementasikan algoritma *K-Nearest Neighbor* dalam pengklasifikasian Sekolah Menengah Pertama/Sederajat berdasarkan peminatan calon siswa. Tujuan penelitian ini adalah untuk memudahkan pengguna dalam menemukan sekolah SMP/ sederajat berdasarkan 8 kriteria sekolah yaitu akreditasi, fasilitas ruangan, fasilitas olah raga, laboratorium, ekstrakurikuler, biaya, tingkatan kelas dan waktu belajar. Adapun data yang digunakan dalam penelitian ini didapatkan dari Dinas Pendidikan Pemuda dan Olahraga Kabupaten Bireuen. Hasil penelitian dengan menggunakan K-NN dan pendekatan *Euclidean Distance* dengan  $k=3$ , diperoleh nilai precision sebesar 63,67%, recall 68,95% dan accuracy sebesar 79,33% .

**Kata Kunci**— Klasifikasi, K-Nearest Neighbor, Euclidean Distance, Akurasi

**Abstract**— In this research, the K-Nearest Neighbor algorithm was implemented in the classification of junior high schools/equivalent based on the interest of prospective students. The purpose of this study is to facilitate users in finding junior high schools/equivalent based on 8 school criteria, namely accreditation, room facilities, sports facilities, laboratories, extracurricular activities, fees, grade levels and study time. The data used in this study were obtained from the Bireuen District Youth and Sports Education Office. The results of the study using K-NN and Euclidean Distance approach with  $k=3$ , obtained precision values of 63.67%, recall 68.95% and accuracy of 79.33%.

**Keywords**— Classification, K-Nearest Neighbor, Euclidean Distance, Accuracy

### I. PENDAHULUAN

Wilayah Bireuen terdiri dari 17 Kecamatan. Setiap kecamatan terdapat 5 sampai 8 SMP/MTS yang keseluruhannya mencapai 111 Sekolah, 86 SMP dan 25 MTS. Setiap sekolah memiliki kriteria atau fasilitas yang berbeda-beda, baik fasilitas ruangan, olah raga, ekstrakurikuler, akreditasi, biaya dan lain sebagainya.

Pada penelitian ini penulis mengimplementasikan algoritma k-nearest neighbor dalam membangun sebuah sistem yang akan memudahkan masyarakat dalam mengakses dan mendapatkan informasi terkait sekolah-sekolah yang memiliki keunggulan baik dibidang fasilitas pendidikan, ekstrakurikuler dan yang lainnya.

*K-Nearest Neighbor* (KNN) adalah salah satu algoritma data mining yang bisa digunakan dalam klasifikasi data [1].

### II. TINJAUAN PUSTAKA

#### A. *K-Nearest Neighbor*

K-Nearest Neighbor merupakan salah satu metode yang digunakan dalam pengklasifikasian [2]. Adapun

langkah-langkah penggunaan metode K-NN ini dijelaskan sebagai berikut[3]:

1. Tentukan parameter K
2. Hitung jarak antara data yang akan dievaluasi dengan semua pelatihan
3. Urutkan jarak yang terbentuk (urut naik)
4. Tentukan jarak terdekat sampai urutan K
5. Pasangkan kelas yang bersesuaian
6. Cari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi [4].

Ada beberapa Metode Pendekatan K-NN. Dalam penelitian ini Metode Pendekatan yang digunakan adalah Metode Pendekatan Euclidean Distance. Berikut adalah rumus untuk mencari nilai akar dari kuadrat dua vector[5].

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (1)$$

Keterangan:

$x_1$  = Sampel data

x2 = Data uji atau data testing  
i = Variabel data  
d = Jarak  
p = Jumlah data training

B. Confusion Matrix

Confusion matrix melakukan pengujian untuk memperkirakan obyek yang benar dan salah. [6].

TABEL I  
CONFUSION MATRIX

Nilai prediksi	Nilai Aktual	
	TP	TN
	FP	TN

Keterangan :

TP = True positif  
TN = True negatif  
FP = False positif.  
FN = False negatif

Rumus untuk perhitungan confusion matrix jika di tuliskan seperti dibawah ini:

- a. Precision digunakan untuk mengukur tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

- b. Recall digunakan untuk mengukur tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi, pada persamaan.

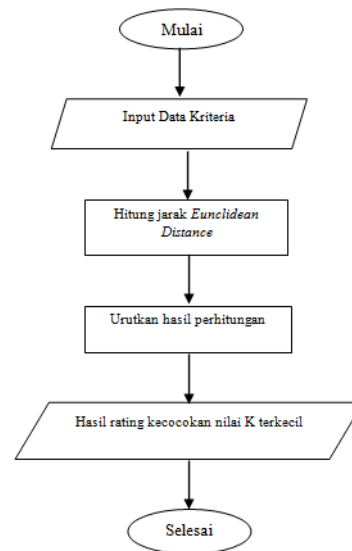
$$Recall = \frac{TP}{TP + FN} \quad (3)$$

- c. Accuracy digunakan untuk mengukur kinerja sebuah metode [8].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

III. METODOLOGI PENELITIAN

Adapun alur penelitian ini dapat dilihat pada diagram alir penelitian berikut.



Gbr. 1 Skema Sistem Klasifikasi Sekolah SMP dan MTS Wilayah Bireuen

IV. HASIL DAN PEMBAHASAN

A. Manajemen Basis Data

1. Tabel\_Admin

TABEL II  
ADMIN

No	Nama	Type	Width	Keterangan
1	admin_id	Int	11	Primary key
2	admin_username	Varchar	50	Username admin
3	admin_password	Varchar	50	Password admin

2. Tabel\_DataSekolah

TABEL III  
DATA SEKOLAH

No	Nama	Type	Width	Keterangan
1	id	Int	5	Primary key
2	nama_sekolah	varchar	50	Nama sekolah
3	latitude	float	-	Titik koordinat sekolah
4	longitude	float	-	Titik koordinat sekolah

3. Tabel\_DataTraining

TABEL IV  
DATA TRAINING

No	Nama	Type	Width	Keterangan
1	id	Int	5	Primary key
2	nama_sekolah	varchar	100	Nama sekolah
3	X1	Int	5	Akreditasi
4	X2	Int	5	Fasilitas ruangan
5	X3	Int	5	Fasilitas olahraga
6	X4	Int	5	Laboratorium
7	X5	Int	5	Biaya perbulan
8	X6	Int	5	Ekstrakurikuler

9	X7	Int	5	Tingkatan kelas
10	X8	Int	5	Waktu Belajar
11	klasifikasi	varchar	50	Rekomendasi sekolah
12	value	Int	5	Value

Jumlah	49	11	11	109	180
--------	----	----	----	-----	-----

Page | 35 4. Tabel\_Data Kriteria

TABEL V  
DATA KRITERIA

No	Nama	Type	Width	Keterangan
1	id	Int	11	Primary key
2	kode	Varchar	10	Kode criteria
3	nama	Varchar	50	Nama criteria

TABEL X  
PREDICTED CLASS K=2

Class Prediksi	Sangat Rekomendasi	Rekomendasi	Kurang Rekomendasi	Jumlah
Sangat Rekomendasi	5	6	0	11
Rekomendasi	8	28	4	40
Kurang Rekomendasi	0	2	7	9
Jumlah	13	36	11	60

5. Tabel\_Data Sub Kriteria

TABEL VI  
DATA SUB KRITERIA

No	Nama	Type	Width	Keterangan
1	sub_id	Int	5	Primary key
2	kode_kriteria	Varchar	10	Kode criteria
3	Value	Varchar	100	Value
4	bobot	Int	5	Bobot

TABEL XI  
CONFUSION MATRIX K=2

Confusion Matrix	TP	FP	FN	TN	Jumlah
Sangat Rekomendasi	5	8	6	41	60
Rekomendasi	28	8	12	12	60
Kurang Rekomendasi	7	4	2	47	60
Jumlah	40	20	20	100	180

B. Hasil Perhitungan Jarak dengan Euclidean Distance

TABEL VII  
HASIL PERHITUNGAN JARAK 1

No	Jarak	Class	k
1	1	Sangat Rekomendasi	1
27	1	Sangat Rekomendasi	2
30	2	Sangat Rekomendasi	3

Berdasarkan nilai jarak yang telah diperoleh dari hasil perhitungan *data testing* 1 terhadap 140 *data training*, nilai jarak terkecil didapatkan pada data ke-1, ke-27 dan data ke-30.

C. Hasil Perhitungan Akurasi K-NN

Perhitungan akurasi didapatkan dengan *confusion matrix*.

TABEL VIII  
PREDICTED CLASS K=1

Class Prediksi	Sangat Rekomendasi	Rekomendasi	Kurang Rekomendasi	Jumlah
Sangat Rekomendasi	7	4	0	11
Rekomendasi	3	34	3	40
Kurang Rekomendasi	0	1	8	9
Jumlah	10	39	11	60

TABEL XII  
PREDICTED CLASS K=3

Class Prediksi	Sangat Rekomendasi	Rekomendasi	Kurang Rekomendasi	Jumlah
Sangat Rekomendasi	6	5	0	11
Rekomendasi	8	23	9	40
Kurang Rekomendasi	0	2	7	9
Jumlah	14	30	16	60

TABEL XIII  
CONFUSION MATRIX K=3

Confusion Matrix	TP	FP	FN	TN	Jumlah
Sangat Rekomendasi	6	8	5	41	60
Rekomendasi	23	7	17	13	60
Kurang Rekomendasi	7	9	2	42	60
Jumlah	36	24	24	96	180

Adapun hasil perhitungan akurasi dapat dilihat pada tabel 14.

TABEL XIV  
HASIL PERHITUNGAN AKURASI

k	Precision	Recall	Akurasi
k=1	0,766355866	0,791750842	0,877777778
k=2	0,5995856	0,644107744	0,777777778
k=3	0,544246032	0,632744108	0,733333333
Jumlah	0,636729166	0,689534231	0,796296296
Persentase	63,67%	68,95%	79,62%

TABEL IX  
CONFUSION MATRIX K=1

Confusion Matrix	TP	FP	FN	TN	Jumlah
Sangat Rekomendasi	7	3	4	46	60
Rekomendasi	34	5	6	15	60
Kurang Rekomendasi	8	3	1	48	60

Berdasarkan tabel 14, dapat dilihat grafik berikut.



Persentase tersebut menunjukkan bahwa algoritma *K-Nearest Neighbor* memiliki kinerja yang baik dan akurat dalam mengklasifikasikan data sekolah SMP dan MTS di wilayah Bireuen.

#### REFERENSI

Page | 37

- [1] Luthfi Anshori, Rekyan Regasari Mardi Putri, Tibyani, "Implementasi Metode K-Nearest Neighbor untuk Rekomendasi Keminatan Studi (Studi Kasus: Jurusan Teknik Informatika Universitas Brawijaya)", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2, No. 7, Juli 2018, hlm. 2745-2753
- [2] Sumarlin, "Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM," *Jurnal Sistem Informasi Bisnis*, 13 April 2015.
- [3] H. Risman, D. Nugroho, and Y. R. WU, "Penerapan Metode K-Nearest Neighbor Pada Aplikasi," *Jurnal TIKomSiN*, vol. 3, no. 2, pp. 19–25, 2013.
- [4] Yeni Kustiyahningsih and N. Syafa'ah, "Sistem Pendukung Keputusan Untuk Menentukan Jurusan Pada Siswa Sma Menggunakan Metode Knn Dan Smart," *Jsii*, vol. 1, no. 1, pp. 19–28, 2014.
- [5] M. M. K. Neighbor, "Seleksi Pegawai dan Dosen UMRI Berbasis E-Recruitment," pp. 71–80.
- [6] J. T. Informatika, F. I. Komputer, and U. D. Nuswantoro, "Berdasarkan Enam Tipe Pattern Menggunakan Metode Euclidean," pp. 1–5.
- [7] I. Menarianti, "Klasifikasi data mining dalam menentukan pemberian kredit bagi nasabah koperasi," *J. Ilm. Teknosains*, vol. 1, no. 1, pp. 1–10, 2015.
- [8] J. Riany, M. P. Lukman, and M. Fajar, "Penerapan Deep Sentiment Analysis pada Angket Penilaian Terbuka Menggunakan K-Nearest Neighbor," *Sisfo*, vol. 06, no. 01, pp. 147–156, 2017.
- [9] Dyah Ayu Wulandari, "Implementasi Data Mining Untuk Menentukan Penjurusan Sekolah Menengah Atas Menggunakan Algoritma *K-Nearest Neighbor* Klasifikasi Pada SMAN 1 PACE", *Simki-Techsain vol. 01 No . 10 tahun 2017*.