

## Klasifikasi Batubara Berdasarkan Jenis Kalori dengan Menggunakan Perbandingan Jarak *Euclid*, Jarak *Manhattan*, dan Jarak *Minkowski* pada Algoritma *Modified K-Nearest Neighbor* (Studi Kasus: PT. XYZ Samarinda)

Neni Rahayu<sup>1</sup>, Wasono<sup>2,\*</sup>, Fidia Deny Tisna Amijaya<sup>3</sup>

<sup>1,2,3</sup> *Laboratorium Matematika Komputasi Jurusan Matematika FMIPA Universitas Mulawarman*

Dikirim: September 2022; Diterima: Maret 2023; Dipublikasi: Maret 2023

Alamat Email Korespondensi: [wasono@fmipa.unmul.ac.id](mailto:wasono@fmipa.unmul.ac.id)

### Abstrak

Metode *Modified K-Nearest Neighbor* (MK-NN) merupakan perbaikan dari metode K-NN yang dapat mengatasi permasalahan *outlier*. Perhitungan nilai akurasi pada MK-NN dengan menambahkan perhitungan nilai fungsi validitas dan *weight voting* digunakan untuk meningkatkan nilai akurasi rendah pada K-NN. Penelitian ini membandingkan tiga jarak yang sering digunakan dalam K-NN, yaitu *Euclid*, *Manhattan* dan *Minkowski*. Penelitian ini bertujuan untuk mengetahui hasil pengklasifikasian batubara berdasarkan jenis kalori dan tingkat akurasi jarak terbaik dari klasifikasi pada metode MK-NN. Data yang digunakan adalah data batubara PT. XYZ Tahun 2017 dengan menggunakan 7 variabel bebas yaitu *Total Moisture* (TM), *Moisture* (M), *Ash*, *Volatile Matter* (VM), *Fixed Carbon* (FC), *Total Sulphur* (TS) dan *Gross Calorific Value* (GCV). Hasil penelitian menunjukkan bahwa dari ketiga jarak yang dibandingkan jarak *Manhattan* memiliki tingkat akurasi yang paling baik yaitu sebesar 100%, sedangkan jarak *Euclid* sebesar 92% dan jarak *Minkowski* sebesar 52%, hasil akurasi terbaik pada MK-NN tersebut diperoleh ketika *K* bernilai 1 pada proporsi 80%:20%.

### Kata Kunci:

Batubara, *Euclid*, *Manhattan*, *Minkowski*, *Modified K-Nearest Neighbor*

### PENDAHULUAN

*Data mining* merupakan proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. *Data mining* berisi pencarian pola yang diinginkan dalam *database* besar untuk membantu pengambilan keputusan di waktu yang akan datang [1]. Klasifikasi adalah salah satu tugas dari *data mining* yang bertujuan untuk memprediksi label kategori benda yang tidak diketahui sebelumnya, dalam membedakan antara objek yang satu dengan yang lainnya berdasarkan atribut atau variabel.

Klasifikasi merupakan penggolongan atau pengelompokan fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya belum diketahui atau pembagian sesuatu menurut kelas-kelasnya. Metode *K-Nearest Neighbor* (K-NN) melakukan

klasifikasi terhadap objek berdasarkan data latih yang menggunakan jarak terdekat atau kemiripan terhadap objek tersebut. Pada fase pembelajaran, algoritma ini melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran [2]. Metode *Modified K- Nearest Neighbor* (MK-NN) merupakan perbaikan dari metode K-NN yang dapat mengatasi permasalahan *outlier* pada saat perhitungan nilai bobot pada K-NN dengan menambahkan perhitungan nilai validitas. Tujuan dari algoritma MK-NN adalah untuk meningkatkan akurasi dari K-NN, dengan menambahkan fungsi Validitas dan *Weight Voting* [3].

Batubara merupakan bahan galian strategis dan menempati posisi yang sangat penting dalam pembangunan nasional, maka posisi batubara sebagai bahan bakar alternatif yang sangat diharapkan dapat mengantisipasi krisis energi dengan meningkatkan pemanfaatannya untuk keperluan domestik sebagai bahan bakar pada pembangkit tenaga listrik, industri maupun untuk kepentingan ekspor [4].

Berdasarkan latar belakang di atas, penelitian ini mengklasifikasikan batubara dengan mengambil studi kasus klasifikasi batubara di PT XYZ Samarinda yang mana sebelumnya penelitian serupa telah dilakukan oleh [5] menggunakan metode *Genetic Modified K- Nearest Neighbor*. Pengembangan metode dilakukan penulis dengan menambahkan 3 jarak sehingga didapatkan judul "Klasifikasi Batubara Berdasarkan Jenis Kalori dengan Menggunakan Perbandingan Jarak *Euclid*, Jarak *Manhattan*, dan Jarak *Minkowski* pada Algoritma *Modified K-Nearest Neighbor* (Studi Kasus: PT. XYZ Samarinda)".

## **METODE PENELITIAN**

Pada penelitian ini, populasi yang digunakan yaitu seluruh data hasil pemeriksaan laboratorium batubara pada PT. XYZ Samarinda tahun 2017, sedangkan sampel pada penelitian ini adalah 60 data hasil pemeriksaan laboratorium batubara yang setiap klasifikasinya masing-masing memiliki 20 data di PT. XYZ Samarinda tahun 2017. Teknik sampling yang digunakan adalah *Total Sampling* atau *Sampling Jenuh*. Variabel pada penelitian ini adalah nomor sampel batubara, hasil pemeriksaan laboratorium batubara PT. XYZ Samarinda, dan klasifikasi batubara berdasarkan jenis kalori.

Adapun langkah-langkah analisis algoritma MK-NN sebagai berikut :

1. Analisis statistika deskriptif
2. Normalisasi data
3. Pengacakan data
4. Pembagian data *training* dan data *testing*
5. Perhitungan jarak *Euclid*, *Manhattan* dan *Minkowski*

## **HASIL DAN PEMBAHASAN**

1. Analisis Statistika Deskriptif

Pada analisis deskriptif ini, akan dilakukan gambaran umum dari hasil pengumpulan data batubara berdasarkan jenis kalori di PT. XYZ Samarinda. Berikut adalah hasil analisis deskriptif yang dapat dilihat pada Tabel 1.

Tabel 1. Perbandingan Statistika Deskriptif

Kategori		TM (%)	M (%)	ASH (%)	VM (%)	FC (%)	TS (%)	GCV (Kcal/Kg)
Lignite	Rata-Rata	34,784	18,962	13,775	35,859	31,455	0,908	4442
	Variansi	107,958	27,775	54,032	4,963	2,594	1,749	15968,316
	Nilai Minimum	10,39	5,45	5,27	30,09	28,7	0,15	4178
	Nilai Maksimum	51,99	25,57	33,94	39,03	34,06	5,7	4598
Sub-Bituminous	Rata-Rata	26,356	14,694	5,162	40,711	39,435	0,818	5525,35
	Variansi	1,271	1,605	3,833	1,019	2,644	0,155	23297,713
	Nilai Minimum	24,37	12,59	3,31	38,59	35,42	0,44	5235
	Nilai Maksimum	28,98	17,06	10,73	42,31	42,24	2,06	5817
Bituminous	Rata-Rata	16,773	10,804	3,355	40,455	45,389	0,667	6303,55
	Variansi	20,755	5,467	2,667	3,340	3,894	0,621	86895,839
	Nilai Minimum	6,63	5,69	0,75	37,54	41,39	0,16	5974
	Nilai Maksimum	21,9	13,34	7,75	43,78	49,27	2,87	6974

Berdasarkan Tabel 1 dapat dilihat bahwa nilai rata-rata dan variansi sangat bervariasi, beberapa dari data di atas juga memiliki nilai rata-rata lebih besar daripada nilai variansi.

## 2. Normalisasi Data

Adapun perhitungan untuk normalisasi data adalah sebagai berikut.

Menghitung rata-rata:

$$\begin{aligned}\bar{x}_1 &= \frac{1}{n} \times \sum_{i=1}^n x_{i1} \\ &= \frac{1}{60} \times (20,73 + 29,63 + 44,85 + \dots + 20,67) \\ &= \frac{1}{60} \times 1558,25 \\ &= 25,97\end{aligned}$$

Menghitung varian dan simpangan baku:

$$\begin{aligned}\sigma_1^2 &= \frac{1}{n-1} \times \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 \\ &= \frac{1}{60-1} \times [(20,73 - 25,97)^2 + (29,63 - 25,97)^2 + \dots + (20,67 - 25,97)^2] \\ &= \frac{1}{59} \times 5717,91 \\ &= 96,914 \\ \sigma_1 &= \sqrt{96,914} \\ &= 9,8445\end{aligned}$$

Adapun nilai normalisasi data pada variabel *Total Moisture* (TM) untuk data ke-1, 2 dan 3 adalah:

$$\begin{aligned}\hat{x}_{11} &= \frac{x_{11} - \bar{x}_1}{\sigma_1} \\ &= \frac{20,73 - 25,97}{9,8445} \\ &= -0,5323 \\ \hat{x}_{21} &= \frac{x_{21} - \bar{x}_1}{\sigma_1} \\ &= \frac{29,63 - 25,97}{9,8445} \\ &= 0,3716 \\ \hat{x}_{31} &= \frac{x_{31} - \bar{x}_1}{\sigma_1} \\ &= \frac{44,85 - 25,97}{9,8445} \\ &= 1,9177\end{aligned}$$

Perhitungan yang sama dilakukan seterusnya dari data ke-4 sampai dengan data ke-60 dan dilanjutkan untuk semua variabel. Berikut adalah hasil normalisasi data yang diperoleh dapat dilihat pada Tabel 2.

Tabel 2. Data Normalisasi

NO	Kelas	TM (%)	M (%)	ASH (%)	VM (%)	FC (%)	TS (%)	GCV (Kcal/Kg)
1	1	-0,5323	-1,0748	3,3228	-3,1419	-1,1902	-0,3608	-1,5063
2	1	0,3717	0,3289	1,7266	-1,0139	-1,6147	0,5099	-1,5653
3	1	1,9177	2,2363	0,0817	-1,3979	-1,1952	-0,5371	-1,3253
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
58	3	-0,766	-0,4489	-0,7589	0,1029	1,1072	2,2843	0,7444
59	3	-1,2962	-0,6786	-0,7133	0,5257	1,0406	-0,4930	1,3162
60	3	-0,5385	-0,3119	-0,7369	1,2409	0,4379	1,9647	1,0058

Selanjutnya akan dilakukan pengacakan data untuk dijadikan sebagai data *training* dan data *testing*.

### 3. Pengacakan Data

Pengacakan data pada penelitian ini bertujuan agar semua data mendapat kesempatan yang sama menjadi data *training* dan data *testing*. Berikut adalah hasil pengacakan data yang diperoleh dapat dilihat pada Tabel 3.

Tabel 3. Pengacakan Data

NO	Kelas	TM (%)	M (%)	ASH (%)	VM (%)	FC (%)	TS (%)	GCV (Kcal/Kg)
1	1	-1,5827	-1,9748	4,16496	-2,9975	-1,4399	1,90957	-1,1054
2	2	-0,1626	0,0085	-0,6473	1,08942	0,16163	-0,1844	0,22038
3	2	0,18479	-0,3414	0,10524	0,50105	-0,0814	-0,3938	-0,0586
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
58	2	-0,1108	0,05276	0,04082	-0,1472	-0,0182	-0,1954	-0,1101
59	3	-0,4328	-0,3329	-1,0495	0,49753	1,14053	-0,504	0,9329
60	1	-0,3434	0,52487	1,42967	-1,1372	-1,395	5,40323	-1,2273

Selanjutnya akan dilakukan pembagian data *training* dan data *testing*.

### 4. Pembagian Data *Training* dan Data *Testing*

Adapun perhitungan pembagian data *training* dan data *testing* sebagai berikut.

Jumlah data *training* = Proporsi data *training* × n

$$= \frac{80}{80+20} \times 60$$

$$= 48$$

Jumlah data *testing* = n – Jumlah data *training*

$$= 60 - 48$$

$$= 12$$

Berdasarkan hasil perhitungan dengan proporsi 80%:20% diperoleh bahwa 48 data pertama hasil randomisasi data untuk masing-masing variabel penelitian berfungsi sebagai data *training* yang akan digunakan pada metode MK-NN. Berikut adalah data *training* yang diperoleh dapat dilihat pada Tabel 4.

Tabel 4. Data *Training*

NO	Kelas	TM (%)	M (%)	ASH (%)	VM (%)	FC (%)	TS (%)	GCV (Kcal/Kg)
1	1	-1,5827	-1,9748	4,16496	-2,9975	-1,4399	1,90957	-1,1054
2	2	-0,1626	0,0085	-0,6473	1,08942	0,16163	-0,1844	0,22038
3	2	0,18479	-0,3414	0,10524	0,50105	-0,0814	-0,3938	-0,0586
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
46	2	0,09134	0,3668	-0,3865	0,57856	-0,1564	0,05804	-0,0687
47	2	-0,0184	-0,2866	-0,4808	0,53981	0,47794	-0,0081	0,28698
48	1	0,3717	0,32886	1,72661	-1,0139	-1,6147	0,50991	-1,5653

Berikut adalah data *testing* yang diperoleh dapat dilihat pada Tabel 5.

Tabel 5. Data *Testing*

NO	Kelas	TM (%)	M (%)	ASH (%)	VM (%)	FC (%)	TS (%)	GCV (Kcal/Kg)
1	1	0,86436	0,40895	0,92692	-0,5876	-1,0304	-0,7024	-1,2122
2	3	-1,7158	-1,9242	0,05025	-0,5171	1,70823	-0,3718	1,68815
3	2	0,30567	-0,0927	-0,4211	1,16341	-0,0332	0,43276	0,25054
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10	2	-0,1108	0,05276	0,04082	-0,1472	-0,0182	-0,1954	-0,1101
11	3	-0,4328	-0,3329	-1,0495	0,49753	1,14053	-0,504	0,9329
12	1	-0,3434	0,52487	1,42967	-1,1372	-1,395	5,40323	-1,2273

Selanjutnya dilakukan perhitungan dari ketiga jarak yang digunakan.

## 5. Perhitungan Jarak sebagai berikut.

### a. Jarak *Euclid*

#### 1) Perhitungan nilai validitas

Adapun perhitungan validitas adalah sebagai berikut.

Perhitungan validitas pada data *training* ke-1

$$\text{validitas}(a) = \frac{1}{K} \sum_{i=1}^k s(\text{lbl}(x_a))$$

$$\begin{aligned} \text{validitas}(1) &= \frac{1}{1} [s((x_1). (n_1(x_{36})))] \\ &= \frac{1}{1} \times 1 \\ &= 1 \end{aligned}$$

Perhitungan validitas pada data *training* ke-2

$$\text{validitas}(a) = \frac{1}{K} \sum_{i=1}^k s(\text{lbl}(x_a))$$

$$\begin{aligned} \text{validitas}(2) &= \frac{1}{1} [s((x_1). (n_1(x_5)))] \\ &= \frac{1}{1} \times 1 \\ &= 1 \end{aligned}$$

Perhitungan di atas dilakukan sampai dengan data *training* ke-48.

#### 2) Perhitungan *weighted voting*

Adapun perhitungan *weight voting* adalah sebagai berikut.

- Perhitungan *weight voting* data *testing* 1

$$\begin{aligned} w(x_{27}, y_1) &= \text{validitas (27)} \times \frac{1}{d(x_{27}, y_1) + 0,5} \\ &= 1 \times \frac{1}{0,6537+0,5} \\ &= 0,86675 \end{aligned}$$

- Perhitungan *weight voting* data *testing* 2

$$\begin{aligned} w(x_8, y_2) &= \text{validitas (8)} \times \frac{1}{d(x_8, y_2) + 0,5} \\ &= 1 \times \frac{1}{1,6469+0,5} \\ &= 0,46579 \end{aligned}$$

⋮

- Perhitungan *weight voting* data *testing* 11

$$\begin{aligned} w(x_{15}, y_{11}) &= \text{validitas (15)} \times \frac{1}{d(x_{15}, y_{11}) + 0,5} \\ &= 1 \times \frac{1}{0,3384+0,5} \\ &= 1,19274 \end{aligned}$$

- Perhitungan *weight voting* data *testing* 12

$$\begin{aligned} w(x_{19}, y_{12}) &= \text{validitas (19)} \times \frac{1}{d(x_{19}, y_{12}) + 0,5} \\ &= 1 \times \frac{1}{4,6048+0,5} \\ &= 0,19589 \end{aligned}$$

Adapun perhitungan *voting* diperoleh perbandingan kelas prediksi data *testing* dan kelas data *testing* asli dapat dilihat pada Tabel 6.

Tabel 6. Hasil Perbandingan Kelas Prediksi Data *Testing* dan Kelas Aslinya Jarak *Euclid*

Data <i>Testing</i>	Kelas Prediksi	Kelas Data Asli	Keterangan
1	1	1	Benar
2	3	3	Benar
3	2	2	Benar
4	1	1	Benar
5	1	1	Benar
6	3	3	Benar
7	1	1	Benar
8	2	2	Benar
9	3	3	Benar
10	2	2	Benar
11	2	3	Salah
12	2	2	Benar

Selanjutnya dilakukan perhitungan untuk menentukan nilai akurasi.

### 3) Perhitungan Nilai Akurasi

Berikut perhitungan akurasi prediksi tepat dari metode MK-NN yaitu sebagai berikut:

$$\begin{aligned}
 \text{akurasi} &= \frac{\text{jumlah data yang diprediksi benar}}{\text{jumlah prediksi yang dilakukan}} \times 100\% \\
 &= \frac{11}{12} \times 100\% \\
 &= 92\%
 \end{aligned}$$

b. Jarak *Manhattan*

1) Perhitungan nilai validitas

Adapun perhitungan validitas adalah sebagai berikut.

Perhitungan validitas pada data *training* ke-1

$$\begin{aligned}
 \text{validitas}(a) &= \frac{1}{K} \sum_{i=1}^k s(\text{lbl}(x_a)) \\
 \text{validitas}(1) &= \frac{1}{1} [s((x_1). (n_1(x_{36})))]) \\
 &= \frac{1}{1} \times 1 \\
 &= 1
 \end{aligned}$$

Perhitungan validitas pada data *training* ke-2

$$\begin{aligned}
 \text{validitas}(a) &= \frac{1}{K} \sum_{i=1}^k s(\text{lbl}(x_a)) \\
 \text{validitas}(2) &= \frac{1}{1} [s((x_1). (n_1(x_5)))] \\
 &= \frac{1}{1} \times 1 \\
 &= 1
 \end{aligned}$$

Perhitungan di atas dilakukan sampai dengan data *training* ke-48.

2) Perhitungan *weighted voting*

Adapun perhitungan *weight voting* adalah sebagai berikut.

• Perhitungan *weight voting* data *testing* 1

$$\begin{aligned}
 w(x_{27}, y_1) &= \text{validitas}(27) \times \frac{1}{d(x_{27}, y_1) + 0,5} \\
 &= 1 \times \frac{1}{1,3+0,5} \\
 &= 0,55435
 \end{aligned}$$

• Perhitungan *weight voting* data *testing* 2

$$\begin{aligned}
 w(x_{21}, y_2) &= \text{validitas}(21) \times \frac{1}{d(x_{21}, y_2) + 0,5} \\
 &= 1 \times \frac{1}{3,55+0,5} \\
 &= 0,24697
 \end{aligned}$$

⋮

• Perhitungan *weight voting* data *testing* 11

$$\begin{aligned}
 w(x_{15}, y_{11}) &= \text{validitas}(15) \times \frac{1}{d(x_{15}, y_{11}) + 0,5} \\
 &= 1 \times \frac{1}{0,77+0,5} \\
 &= 0,78703
 \end{aligned}$$

• Perhitungan *weight voting* data *testing* 12

$$\begin{aligned}
 w(x_{48}, y_{12}) &= \text{validitas}(48) \times \frac{1}{d(x_{19}, y_{12}) + 0,5} \\
 &= 1 \times \frac{1}{6,78+0,5} \\
 &= 0,13732
 \end{aligned}$$

Adapun perhitungan *voting* diperoleh perbandingan kelas prediksi data *testing* dan kelas data *testing* asli dapat dilihat pada Tabel 7.

Tabel 7. Hasil Perbandingan Kelas Prediksi Data *Testing* dan Kelas Aslinya Jarak *Manhattan*

Data <i>Testing</i>	Kelas Prediksi	Kelas Data Asli	Keterangan
1	1	1	Benar
2	3	3	Benar
3	2	2	Benar
4	1	1	Benar
5	1	1	Benar
6	3	3	Benar
7	1	1	Benar
8	2	2	Benar
9	3	3	Benar
10	2	2	Benar
11	3	3	Benar
12	1	1	Benar

Selanjutnya dilakukan perhitungan untuk menentukan nilai akurasi.

### 3) Perhitungan Nilai Akurasi

Berikut perhitungan akurasi prediksi tepat dari metode MK-NN yaitu sebagai berikut:

$$\begin{aligned}
 \text{akurasi} &= \frac{\text{jumlah data yang diprediksi benar}}{\text{jumlah prediksi yang dilakukan}} \times 100\% \\
 &= \frac{12}{12} \times 100\% \\
 &= 100\%
 \end{aligned}$$

#### c. Jarak *Minkowski*

##### 1) Perhitungan nilai validitas

Adapun perhitungan validitas adalah sebagai berikut.

Perhitungan validitas pada data *training* ke-1

$$\begin{aligned}
 \text{validitas}(a) &= \frac{1}{K} \sum_{i=1}^k s(\text{lbl}(x_a)) \\
 \text{validitas}(1) &= \frac{1}{1} [s((x_1). (n_1(x_{36})))] \\
 &= \frac{1}{1} \times 1 \\
 &= 1
 \end{aligned}$$

Perhitungan validitas pada data *training* ke-2

$$\begin{aligned}
 \text{validitas}(a) &= \frac{1}{K} \sum_{i=1}^k s(\text{lbl}(x_a)) \\
 \text{validitas}(2) &= \frac{1}{1} [s((x_1). (n_1(x_5)))] \\
 &= \frac{1}{1} \times 1 \\
 &= 1
 \end{aligned}$$

Perhitungan di atas dilakukan sampai dengan data *training* ke-48.

##### 2) Perhitungan *weighted voting*

Adapun perhitungan *weight voting* adalah sebagai berikut.

- Perhitungan *weight voting* data *testing* 1

$$\begin{aligned} w(x_{27}, y_1) &= \text{validitas (27)} \times \frac{1}{d(x_{27}, y_1) + 0,5} \\ &= 1 \times \frac{1}{0,0537+0,5} \\ &= 1,80597 \end{aligned}$$

- Perhitungan *weight voting* data *testing* 2

$$\begin{aligned} w(x_8, y_2) &= \text{validitas (8)} \times \frac{1}{d(x_8, y_2) + 0,5} \\ &= 1 \times \frac{1}{0,666+0,5} \\ &= 0,85796 \end{aligned}$$

⋮

- Perhitungan *weight voting* data *testing* 11

$$\begin{aligned} w(x_{15}, y_{11}) &= \text{validitas (15)} \times \frac{1}{d(x_{15}, y_{11}) + 0,5} \\ &= 1 \times \frac{1}{0,006+0,5} \\ &= 1,975 \end{aligned}$$

- Perhitungan *weight voting* data *testing* 12

$$\begin{aligned} w(x_{19}, y_{12}) &= \text{validitas (19)} \times \frac{1}{d(x_{19}, y_{12}) + 0,5} \\ &= 1 \times \frac{1}{23,37+0,5} \\ &= 0,0419 \end{aligned}$$

Adapun perhitungan *voting* diperoleh perbandingan kelas prediksi data *testing* dan kelas data *testing* asli dapat dilihat pada Tabel 8.

Tabel 8. Hasil Perbandingan Kelas Prediksi Data *Testing* dan Kelas Aslinya Jarak *Minkowski*

Data Testing	Kelas Prediksi	Kelas Data Asli	Keterangan
1	1	1	Benar
2	3	3	Benar
3	3	2	Salah
4	1	1	Benar
5	1	1	Benar
6	3	3	Benar
7	1	1	Benar
8	3	2	Salah
9	2	3	Salah
10	3	2	Salah
11	2	3	Salah
12	2	2	Benar

Selanjutnya dilakukan perhitungan untuk menentukan nilai akurasi.

### 3) Perhitungan Nilai Akurasi

Berikut perhitungan akurasi prediksi tepat dari metode MK-NN yaitu sebagai berikut:

$$\begin{aligned} \text{akurasi} &= \frac{\text{jumlah data yang diprediksi benar}}{\text{jumlah prediksi yang dilakukan}} \times 100\% \\ &= \frac{8}{12} \times 100\% \\ &= 58\% \end{aligned}$$

## 6. Perbandingan Nilai Akurasi

Perbandingan nilai akurasi untuk mengetahui nilai akurasi terbaik diantara jarak *Euclid*, jarak *Manhattan* dan jarak *Minkowski*. Adapun hasil perbandingan nilai akurasi dapat dilihat pada Tabel 9.

Tabel 9. Hasil Perbandingan akurasi antar tiga jarak

Proporsi	Jarak <i>Euclid</i>	Jarak <i>Manhattan</i>	Jarak <i>Minkowski</i>
80%:20%	92%	100%	52%

Pada perhitungan menggunakan proporsi 80%:20% didapatkan nilai akurasi jarak *Euclid* sebanyak 92%, jarak *Manhattan* sebanyak 100%, dan jarak *Minkowski* sebanyak 52%.

**PENUTUP**

Berdasarkan hasil penelitian dan pembahasan, maka kesimpulan yang dapat diambil dalam pengklasifikasian dengan metode MK-NN dalam klasifikasi batubara di PT. XYZ Samarinda berdasarkan kelas prediksi dengan menggunakan Jarak *Euclid* yang termasuk klasifikasi *Lignite* adalah data *testing* 1, 4, 5 dan 7, klasifikasi *Sub-Bituminous* adalah data *testing* 3, 8, 10, 11 dan 12, klasifikasi *Bituminous* adalah data *testing* 2, 6 dan 9. Jarak *Manhattan* yang termasuk klasifikasi *Lignite* adalah data *testing* 1, 4, 5, 7 dan 12, klasifikasi *Sub-Bituminous* adalah data *testing* 3, 8, dan 10, klasifikasi *Bituminous* adalah data *testing* 2, 6, 9 dan 11. Jarak *Minkowski* yang termasuk klasifikasi *Lignite* adalah data *testing* 1, 4, 5 dan 7, klasifikasi *Sub-Bituminous* adalah data *testing* 9, 11 dan 12, klasifikasi *Bituminous* adalah data *testing* 2, 3, 6, 8 dan 10.

Perbandingan akurasi metode pengukuran jarak *Euclid*, jarak *manhattan*, dan jarak *minkowski* untuk pengklasifikasian batubara di PT. XYZ telah dilakukan dan memberikan nilai atau tingkat akurasi yang berbeda-beda, yaitu untuk jarak *Euclid* sebanyak 92%, jarak *manhattan* sebanyak 100%, dan jarak *minkowski* sebanyak 52%. Dengan demikian, dapat disimpulkan bahwa jarak *manhattan* merupakan jarak dengan persentasi akurasi terbaik yang dapat diterapkan dalam algoritma *Modified K-Nearest Neighbor* untuk menghitung k lasifikasi batubara berdasarkan jenis kalori dengan  $K=1$  dan proporsi 80%:20%.

Berdasarkan penelitian yang telah dilakukan, pada penelitian selanjutnya dapat melakukan perbandingan jarak dengan metode klasifikasi lainnya seperti *Support Vector Machine (SVM)*.

**DAFTAR PUSTAKA**

- [1] Hermawati, F. A. (2013). *Data Mining*. Yogyakarta: Andi.
- [2] Feldman, R., dan Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press.
- [3] Parvin, H., Alizadeh, H dan Minae-Bidgoli, B. (2008). *MKNN: Modification on K-Nearest Neighbor Classification*. San Francisco. USA.
- [4] Sukandarrumidi. (2017). *Batubara dan Pemanfaatannya: Pengantar Teknologi Batubara Menuju Lingkungan Bersih*. Yogyakarta: UGM Press.
- [5] Wahyudi, N., Wahyuningsih, S., dan Amijaya, F. (2020). Optimasi Klasifikasi Batubara Berdasarkan Jenis Kalori dengan menggunakan *Genetic Modified K-Nearest Neighbor (GMK-NN)*. *EKSPONENSIAL*, 10(2), 103-112