

Prediksi Klasifikasi Royalti Batubara Menggunakan Algoritma *Fuzzy K-Nearest Neighbor* (Studi Kasus: CV.Fazar Utama)

Prediction Of Coal Royalty Classification Using Fuzzy K-Nearest Neighbor Algorithm (Case Study: CV. Fazar Utama)

Hanny Permata Lestari¹, Sri Wahyuningsih², dan Fidia Deny Tisna Amijaya³

¹Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman

²Laboratorium Statistika Terapan FMIPA Universitas Mulawarman

³Laboratorium Matematika Terapan FMIPA Universitas Mulawarman

E-mail: hannyplestari@gmail.com

Abstract

As one of energy commodity, coal has been explored and exploited in order to fulfil the energy needs of world community. The calculation of coal royalties in Indonesia is based on the value of the quality of exploited or sold coal and the benchmark coal price determined by the government. CV Fazar Utama Kota Samarinda East Kalimantan Province classifies coal royalties based on calculations that have been made on coal sample inspection such as low, medium, and high. The Fuzzy K-Nearest Neighbor (FK-NN) method is one of the lazy learner algorithm which is used to determine the classification prediction. The K-Fold Cross Validation technique is used to obtain the optimal K value on the Fuzzy K-Nearest Neighbor (FK-NN) algorithm in order to get more accurate prediction. In this research, the first step is to find the optimal K value by experimenting 1-Fold Cross Validation, 3-Fold Cross Validation and 5-Fold Cross Validation using the 80:20 proportion of training and testing data. The next step is find the percentage accuracy of coal Royalty classification prediction in CV Fazar Utama in 2017 using Fuzzy K-Nearest Neighbor (FK-NN) Algorithm with optimal K value. Based on this research, the optimal value of K used in FK-NN Algorithm is F3-NN obtained in 1-Fold Cross Validation experiment. Furthermore, the result shows that the percentage accuracy of coal Royalty classification prediction at CV Fazar Utama in 2017 using F3-NN with 1-Fold Cross Validation is 100%.

Keywords: Classification, Fuzzy K-Nearest Neighbor, K optimal

Pendahuluan

Klasifikasi adalah suatu pekerjaan yang melakukan penilaian terhadap suatu objek data untuk masuk ke dalam suatu kelas tertentu dari sejumlah kelas yang tersedia. Algoritma klasifikasi merupakan salah satu algoritma yang menjadi bagian dari *machine learning* yang merupakan proses penemuan model atau fungsi yang menjelaskan konsep kelas data, dengan tujuan dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui (Prasetyo, 2012).

Algoritma *Fuzzy K-Nearest Neighbor* (FK-NN) merupakan salah satu algoritma *lazy learner* yang digunakan untuk menentukan prediksi klasifikasi. Metode FK-NN melakukan prediksi secara tegas pada uji berdasarkan perbandingan K tetangga terdekat, dengan menetapkan nilai keanggotaan sebagai fungsi jarak vektor dari K-NN dan keanggotaan tetangga mereka di kelas-kelas yang memungkinkan. Pemilihan nilai K dalam metode FK-NN akan mempengaruhi kinerja metode tersebut. Teknik *K-Fold Cross Validation* digunakan untuk mendapatkan nilai K optimal pada Algoritma FK-NN agar penentuan prediksi klasifikasi lebih akurat (Prasetyo, 2014).

Batubara merupakan komoditas energi yang semakin banyak dieksplorasi dan dieksploitasi,

untuk pemenuhan kebutuhan energi masyarakat dunia. Perhitungan royalti batubara di Indonesia didasarkan atas nilai kualitas batubara yang dieksploitasi atau dijual serta Harga Patokan Batubara (HPB) yang ditentukan dari pemerintah. Nilai royalti yang dikenakan batubara tergantung dari kebijakan pemerintah tentang Penerimaan Negara Bukan Pajak (PNBP) jenis royalti batubara.

CV. Fazar Utama adalah perusahaan pemasaran batubara yang berlokasi di kota Samarinda. Royalti batubara yang digunakan oleh CV.Fazar Utama diklasifikasikan menjadi 3 yaitu rendah, sedang dan tinggi.

Penelitian terdahulu yang menjadi acuan dalam penelitian ini yaitu penelitian Nugraha dkk (2017) yang mencoba menentukan status gizi balita menggunakan FK-NN dan teknik *K-Fold Cross Validation* untuk mencari nilai K optimal. Dari penelitian tersebut didapatkan nilai K sebesar 4 dengan akurasi 84,37%. Nilai K tersebut digunakan pada algoritma FK-NN untuk memprediksi status gizi balita.

Berdasarkan uraian di atas, penulis tertarik untuk mengkaji analisis dengan mengambil studi kasus Prediksi Klasifikasi Royalti Batubara di

CV.Fazar Utama dengan menggunakan Algoritma Fuzzy K-Nearest Neighbor.

Analisis Multivariat

Analisis multivariat berasal dari kata *multy* dan *variate* yang berarti alat analisis lebih dari dua variabel. Analisis multivariat sering digunakan untuk memecahkan permasalahan dalam penelitian yang sifatnya sangat kompleks. Analisis ini dikelompokkan menjadi dua kelompok besar yaitu metode dependensi dan metode interdependensi (Widarjono, 2015).

Data Mining

Menurut Prasetyo (2012), ada tiga langkah utama dalam data *mining* yaitu eksplorasi atau pemrosesan awal data, membangun model dan melakukan validasi, dan penerapan. Data *mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. Model Prediksi

Model prediksi berkaitan dengan pembuatan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang didapat.

2. Analisis Kelompok

Analisis kelompok digunakan untuk pengelompokan (*clustering*) identifikasi data yang memiliki karakteristik tertentu.

3. Analisis asosiasi

Association (asosiasi) dinamakan juga analisis keranjang pasar, fungsi ini mengidentifikasi item-item produk yang kemungkinan dibeli konsumen bersamaan dengan produk lain.

4. Deteksi Anomali

Deteksi anomali (*anomaly detection*) berkaitan dengan pengamatan sebuah data dari sejumlah data yang secara signifikan mempunyai karakteristik yang berbeda dari sisa data yang lain.

Konsep Klasifikasi

Berdasarkan cara pelatihannya, algoritma klasifikasi dibagi menjadi dua macam, yaitu *eager learner* dan *lazy learner*.

Algoritma klasifikasi yang masuk kategori *eager learner* diantaranya adalah *Artificial Neural Network (ANN)*, *Support Vector Machine (SVM)*, dan *Bayesian*. Algoritma- algoritma klasifikasi yang masuk kategori ini diantaranya adalah *Rome classifier*, *K Nearest Neighbor (K-NN)*, *Fuzzy K-Nearest Neighbor (FK.NN)*, dan regresi linier (Prasetyo, 2014).

Konsep Kedekatan

Untuk mengukur ketidakmiripan dua data dengan beberapa atribut untuk setiap data

digunakan kuantitas jarak (*distance*). Ada banyak model pengukuran jarak dan yang paling sering digunakan adalah jarak *Euclidean*. Jarak Euclid dapat dihitung dengan rumus:

$$d(x_i, y_j) = \sqrt{\sum_{k=1}^p (x_{ik} - y_{jk})^2} \tag{1}$$

dimana:

d : Jarak Euclidean

x_{ik}: nilai variabel ke-*k* dari data *training* xke-*i*

y_{jk}: nilai variabel ke-*k* dari data *testing* y ke-*j*

p : jumlah variabel

Metode K-Fold Cross Validation

Menurut Pandie (2012), *K-Fold Cross Validation* merupakan salah satu metode pelatihan yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan melakukan randomisasi *dataset* sehingga sistem tersebut teruji untuk beberapa atribut input yang acak.

K-Fold Cross Validation dapat digunakan untuk memperkirakan tingkat kesalahan yang terjadi, karena data *training* pada setiap *fold* cukup berbeda dengan data *training* yang asli. Setiap perulangan disisakan satu subset untuk *testing* dan subset lainnya untuk data *training*. Jumlah data di dalam satu subset dapat dihitung menggunakan rumus:

$$b = \frac{n}{K_i} \tag{2}$$

dimana:

b = banyak data di dalam satu subset

n = banyak data yang digunakan

K_i = nilai *K-Fold Cross Validation*

Metode Fuzzy K-Nearest Neighbor (FK-NN)

Algoritma FK-NN merupakan pengembangan dari algoritma K-NN yang digabungkan dengan teori *fuzzy* dalam menyampaikan definisi pemberian label kelas pada data *testing* yang diprediksi. Dalam FK-NN dengan adanya konsep *fuzzy* maka sebuah data memiliki nilai keanggotaan pada setiap kelas, yang artinya sebuah data bisa dimiliki oleh kelas yang berbeda dengan nilai derajat keanggotaan dalam interval [0,1] (Prasetyo, 2012).

Proses FKNN

Menurut Prasetyo (2014), langkah-langkah pada algoritma *Fuzzy K-NN* yaitu:

1. Melakukan normalisasi data pada setiap variabel
2. Menentukan jumlah K tetangga terdekat
3. Mencari K tetangga terdekat dengan melakukan perhitungan jarak antar data *testing* dan semua data *training* menggunakan Persamaan (1)

4. Menghitung nilai keanggotaan $u(y_j, c_l)$ untuk setiap kelas, dimana $1 \leq l \leq s$ dengan rumus:

$$u(y_j, c_l) = \frac{\sum_{i=1}^K u(x_i, c_l) * d(x_i, y_j)^{\frac{-2}{(m-1)}}}{\sum_{i=1}^K d(x_i, y_j)^{\frac{-2}{(m-1)}}} \quad (3)$$

Keterangan:

$u(y_j, c_l)$: nilai keanggotaan data $testing x_j$ terhadap kelas c_l .

$j = 1, 2, 3 \dots q$, jumlah data $testing$

$l = 1, 2, 3 \dots s$, banyaknya kelas.

K : Jumlah tetangga terdekat

$u(x_i, c_l)$: nilai keanggotaan data tetangga dalam K tetangga pada kelas c_l , nilainya 1 jika data $training x_i$ milik kelas c_l atau 0 jika bukan milik kelas c_l .

$d(x_i, y_j)$: jarak data x_i ke data y_j dalam K tetangga terdekat.

m : bobot pangkat sama dengan banyaknya kelas yang besarnya $m > 1$.

5. Memilih nilai terbesar $u(y_j, c_l)$ untuk semua $1 \leq l \leq s$
 6. Menentukan kelas hasil prediksi dengan nilai keanggotaan terbesar.

(Prasetyo, 2012)

Normalisasi Data

Variabel dengan nilai yang besar memiliki pengaruh yang lebih besar dalam melakukan prediksi klasifikasi daripada variabel dengan nilai yang kecil. Untuk mengatasi masalah tersebut, digunakan teknik normalisasi sehingga semua variabel berada pada jangkauan yang sama dan tidak ada variabel yang memiliki pengaruh dominan terhadap variabel lainnya. Untuk menghitung normalisasi data dapat menggunakan rumus :

$$\bar{x}_k = \frac{1}{N} \sum_{l=1}^N x_{lk} \quad (4)$$

$$\sigma_k^2 = \frac{1}{N-1} \sum_{l=1}^N (x_{lk} - \bar{x}_k)^2 \quad (5)$$

$$\hat{x}_{lk} = \frac{x_{lk} - \bar{x}_k}{\sigma_k} \quad (6)$$

Dimana:

x_{lk} = data ke- l pada variabel ke- k dimana $k=1, 2, \dots, q$

\bar{x}_k = rata-rata pada variabel ke- k

σ_k = standar deviasi

\hat{x}_{lk} = normalisasi data ke- l variabel ke- k

Akurasi Prediksi

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua himpunan data dengan benar, tetapi tidak dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya (Rodiansyah, 2013). Untuk menghitung akurasi prediksi digunakan rumus:

$$a_{ij} = \frac{\text{Jumlah data yang diprediksi benar}}{b} \times 100\% \quad (7)$$

dimana:

a_{ij} = akurasi untuk subset ke- i dan K-NN ke- j

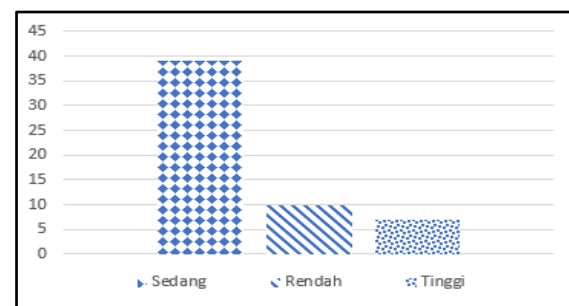
b = banyak data dalam satu subset data $testing$

Klasifikasi Royalti Batubara

Klasifikasi batubara digunakan untuk menggolongkan royalti batubara berdasarkan nilai jual dan kualitasnya. Secara luas, klasifikasi royalti batubara terdiri dari aspek komersial dan aspek ilmiah. Klasifikasi Royalti batubara untuk kepentingan ilmiah antara lain mencakup kualitas batubara, sedangkan untuk kebutuhan komersil antara lain nilai dagang dan royalti. Klasifikasi batubara menurut royalti di CV. Fazar Utama adalah rendah, sedang dan tinggi.

Hasil dan Pembahasan

1. Statistika Deskriptif



Gambar 1. Klasifikasi royalti batubara

Berdasarkan gambar CV. Fazar Utama melakukan *sampling* batubara sebanyak 56 kali setelah itu dilakukan pengujian harga royalti batubara. Didapatkan hasil klasifikasi tertinggi oleh royalti batubara kategori sedang sebanyak 39 data atau sebanyak 69% dari jumlah keseluruhan data yang ada.

2. Normalisasi data

Langkah Normalisasi data yaitu menghitung 56 data menggunakan persamaan (4), (5), dan (6).

3. Membagi data training dan testing

Langkah selanjutnya adalah membagi data ke dalam 2 bagian yaitu data *training* dan *testing*.

Tabel 1. Data Training

Sampel	TM	TS	Ash	GCV/GAR	HPB	Klasifikasi
A-01	-0,2587	-0,5840	-0,9933	0,8646	0,7297	3
A-02	0,3122	0,8076	1,7332	-0,4573	-0,8304	2
A-03	0,0431	-0,4680	-0,6102	0,2908	0,2599	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
A-43	0,0268	-0,2361	1,6318	-0,7749	-0,6652	2
A-44	-2,4770	-0,4680	-0,1257	2,2377	2,6254	3
A-45	-2,7217	-0,4680	-1,0271	2,4324	2,9807	3

Tabel 2. Data Testing

Sampel	TM	TS	Ash	GCV/GAR	HPB	Klasifikasi
A-46	-2,3955	-0,3520	-0,0131	2,9857	2,9192	3
A-47	-2,3955	0,4597	0,2123	2,9448	2,7154	3
A-48	-0,6012	-0,3520	0,0658	0,1166	0,2858	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
A-54	1,5845	-0,9319	0,6516	-1,5640	-1,2970	1
A-55	0,7527	-0,8159	1,4065	-1,3180	-1,0383	1
A-56	1,0381	-0,8159	0,1334	-0,8672	-0,6716	1

4. Penentuan Nilai K Optimal Dengan K-Fold Cross Validation

Pencarian nilai K optimal dengan teknik K-Fold Cross Validation akan dilakukan dengan tiga percobaan fold yaitu 1-Fold Cross Validation, 3-Fold Cross Validation dan 5-Fold Cross Validation untuk nilai K=1, 3, 5 dan 7. Algoritma K-Fold Cross Validation hanya menggunakan data training proporsi 80% sebanyak 45 data yang dijadikan sebagai data pelatihan. Langkah pertama yang dilakukan adalah randomisasi. Randomisasi dilakukan agar seluruh data mendapatkan peluang yang sama untuk menjadi data testing pada pelatihan. Berikut adalah hasil randomisasi 45 data dapat dilihat pada tabel 3.

Tabel 3. Data Hasil Randomisasi

Sampel	TM	TS	Ash	GCV/GAR	HPB	Klasifikasi
A-29	0,2144	0,2278	0,2461	0,4137	-0,0276	2
A-04	0,4509	-0,5840	-0,7229	0,0141	0,0191	2
A-13	1,4133	-0,4680	-0,0018	-1,4000	-1,1249	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
A-30	0,5977	0,8076	-0,6891	-0,5802	-0,5553	2
A-20	0,5569	-0,1201	0,3137	-0,3241	-0,4430	2
A-38	-0,2750	-0,1201	0,0658	0,1166	0,1195	2

Selanjutnya adalah mencari nilai K optimal dengan menggunakan 1-Fold Cross Validation, 3-Fold Cross Validation dan 5-Fold Cross Validation agar dapat diperoleh nilai K yang optimal pada klasifikasi FKNN. Langkah - langkah dalam menentukan nilai K optimal dengan 1-Fold Cross Validation ada 4 yaitu:

a. Menentukan subset data

Untuk 1-Fold Cross Validation(1-FCV), jumlah data dalam satu subset dapat dihitung dengan menggunakan Persamaan (2)

$$b = \frac{n}{Ki} = \frac{45}{1} = 45$$

Berdasarkan perhitungan, jumlah data dalam satu subset adalah 45 data. Masing-masing data mendapatkan giliran untuk menjadi data testing. Berdasarkan data Tabel 3, sampel A-29 digunakan terlebih dahulu sebagai data testing pertama dan 44 data lainnya sebagai data training pada perhitungan jarak Euclidean 1-FCV Misalkan sampel A-29 adalah data pertama, sampel A-04

adalah data kedua, dan seterusnya, maka data testing kedua adalah data kedua, data testing ketiga adalah data ketiga, dan seterusnya.

b. Menghitung Jarak Euclidean

Contoh perhitungan jarak Euclidean antara data testing pertama (sampel A-29) dengan data training pertama (sampel A-04) menggunakan Persamaan (1).

$$d(1,1) = \sqrt{(x_{11} - y_{11})^2 + \dots + (x_{15} - y_{15})^2}$$

$$= \sqrt{0,4509 - 0,2144)^2 + \dots + (0,0191 - (-0,0276))^2}$$

$$= 1,3474$$

Jarak Euclidean dihitung hingga data training sampel ke 44 (sampel A-38). Kemudian dilakukan perangkingan pada setiap perhitungan jarak Euclidean dan dipilih data dari rank 1 hingga 7 untuk perhitungan nilai keanggotaan.

c. Menghitung Nilai Keanggotaan dan Hasil Prediksi Klasifikasi

Untuk memprediksi klasifikasi, dapat dilakukan dengan menghitung nilai keanggotaan setiap kelas dengan bobot pangkat sama dengan banyaknya kelas atau m = 3. kelas dengan nilai keanggotaan terbesar akan menjadi kelas hasil prediksi. Contoh perhitungan nilai keanggotaan setiap kelas pada data testing pertama (sampel A-29) untuk K=1 menggunakan persamaan (3).

$$u(y_1, c_1) = \frac{u(x_1, c_1) \times d(x_1, y_1)^{\frac{-2}{(3-1)}}}{d(x_1, y_1)^{\frac{-2}{(3-1)}}} = \frac{0 \times 0,6132^{-1}}{0,6132^{-1}} = 0$$

$$u(y_1, c_2) = \frac{u(x_1, c_2) \times d(x_1, y_1)^{\frac{-2}{(3-1)}}}{d(x_1, y_1)^{\frac{-2}{(3-1)}}} = \frac{1 \times 0,6132^{-1}}{0,6132^{-1}} = 1$$

$$u(y_1, c_3) = \frac{u(x_1, c_3) \times d(x_1, y_1)^{\frac{-2}{(3-1)}}}{d(x_1, y_1)^{\frac{-2}{(3-1)}}} = \frac{0 \times 0,6132^{-1}}{0,6132^{-1}} = 0$$

Selanjutnya menentukan kelas hasil prediksi dengan memilih nilai keanggotaan terbesar dari u(y1, c1), u(y1, c2) dan u(y1, c3). Data testing pertama atau sampel A-29 dengan K=1 diprediksi masuk ke kelas 2 karena memiliki u(y1, c2) atau nilai keanggotaan terbesar.

Hasil perhitungan nilai keanggotaan setiap kelas dengan hasil prediksi pada data testing pertama (sampel A-29) untuk K=1, 3, 5 dan 7 dapat dilihat pada tabel 4.

Tabel 4. Nilai Keanggotaan dan Prediksi data testing pertama (sampel A-29)

Batas FK-NN	Nilai Keanggotaan			Hasil Prediksi Data Testing A-29
	Kelas 1	Kelas 2	Kelas 3	
1	0	1	0	2
3	0	1	0	2
5	0	1	0	2
7	0	1	0	2

Berdasarkan Tabel 4, data testing pertama sampel A-29 diprediksi masuk kelas 2.

Selanjutnya dilakukan hal yang sama untuk data *testing* kedua dan seterusnya hingga data *testing* ke 45 (Sampel A-38) sehingga mendapatkan hasil prediksi.

d. Perbandingan Klasifikasi Hasil Prediksi FK-NN dengan Data Asli pada 1-FCV

Hasil prediksi klasifikasi menggunakan FK-NN dengan nilai K 1, 3, 5 dan 7 pada 1-FCV dibandingkan dengan data asli untuk semua percobaan 45 data dapat dilihat pada Tabel 5.

Tabel 5.Perbandingan Hasil Prediksi Klasifikasi FK-NN dengan Klasifikasi Data Asli pada 1-FCV

Data Testing	FK-NN				Klasifikasi Pada Data Asli
	1	3	5	7	
A-29	2	2	2	2	2
A-04	2	2	2	2	2
A-13	1	1	1	1	1
A-23	2	2	2	2	2
A-06	3	3	2	2	3
⋮	⋮	⋮	⋮	⋮	⋮
A-30	2	2	2	2	2
A-20	2	2	2	2	2
A-38	2	2	2	2	2
Prediksi Benar	43	44	41	41	

Berdasarkan Tabel 5, diketahui bahwa angka yang bercetak tebal merupakan angka yang memiliki perbedaan dengan klasifikasi data asli. Semakin banyak prediksi yang sama dengan data asli maka nilai K tersebut menjadi lebih baik atau optimal untuk digunakan dalam prediksi klasifikasi.

e. Akurasi Hasil Prediksi Klasifikasi 1-FCV

Contoh perhitungan akurasi prediksi klasifikasi 1-FCV pada F1-NN untuk 45 data percobaan menggunakan persamaan (7).

$$a_{ij} = \frac{43}{45} \times 100\% = 95,56$$

Hasil perhitungan akurasi prediksi pada percobaan 1-FCV dapat dilihat pada tabel 6.

Tabel.6Akurasi Prediksi Klasifikasi 1-FCV

Nilai K	Prediksi Benar	Akurasi (%)
1	43	95,56
3	44	97,78
5	41	91,11
7	41	91,11

Berdasarkan Tabel 6 akurasi tertinggi untuk 1-FCV adalah F3-NN sebesar 97,78%.

Menentukan nilai K optimal dengan 3-Fold Cross Validation:

Langkah dalam menentukan nilai K optimal dengan 3-FCV sama seperti 1-FCV, perbedaannya hanya terletak pada jumlah subset.

a. Menentukan Subset Data

Jumlah data dalam satu subset pada 3-FCV adalah sebagai berikut:

$$b = \frac{45}{3} = 15$$

Berdasarkan perhitungan, jumlah data dalam satu subset adalah 15 data. Masing-masing subset mendapat giliran untuk menjadi data *testing*.

b. Menghitung Jarak Euclidean

Langkah perhitungan jarak *Euclidean* antara data *training* dan *testing* berdasarkan data Tabel 3, subset 1 yang berisi 15 data yaitu sampel A-29 sampai A-42 digunakan sebagai data *testing* terlebih dahulu dan 2 subset lainnya yang berjumlah 30 data menjadi data *training* menggunakan persamaan (1). Kemudian dilakukan perangkingan pada setiap perhitungan jarak *Euclidean* dan dipilih data dari *rank* 1 hingga 7 untuk perhitungan nilai keanggotaan.

c. Menghitung Nilai Keanggotaan dan Hasil Prediksi Klasifikasi

Perhitungan nilai keanggotaan setiap kelas pada data *testing* dilakukan untuk memprediksi klasifikasi. Dimulai dengan menghitung nilai keanggotaan sampel A-29 pada subset 1 menjadi data *testing* pertama dan 30 data pada subset 2 dan subset 3 menjadi data *training*. Perhitungan dilakukan sama seperti sebelumnya. Kelas dengan nilai keanggotaan terbesar akan menjadi kelas hasil prediksi. Nilai keanggotaan setiap kelas dihitung menggunakan Persamaan (2.14). Hasil perhitungan nilai keanggotaan data *testing* A-29 dapat dilihat pada Tabel 7.

Tabel 7. Hasil Perhitungan Nilai Keanggotaan Data *Testing* Pertama Pada Subset 1 (Sampel A-29)

Nilai K	Nilai Keanggotaan			Hasil Klasifikasi
	1	2	3	
1	0	1	0	2
3	0	1	0	2
5	0	1	0	2
7	0	1	0	2

Berdasarkan Tabel 7 data *testing* pertama sampel A-29 pada subset 1 untuk 3-FCV diprediksikan masuk kelas 2.

d. Prediksi Klasifikasi Perbandingan Klasifikasi Hasil Prediksi FK-NN dengan Data Asli pada 3-FCV Subset 1

Hasil prediksi klasifikasi menggunakan FK-NN dengan nilai K= 1, 3, 5 dan 7 pada 3-FCV dibandingkan dengan klasifikasi data asli untuk subset 1 dapat dilihat pada Tabel 8. Berdasarkan Tabel 8. dapat diketahui bahwa angka yang bercetak tebal merupakan angka yang memiliki perbedaan dengan klasifikasi data asli.

$$b = \frac{n}{K_i} = \frac{45}{5} = 9$$

Data Testing	FK-NN				Klasifikasi Pada Data Asli
	1	3	5	7	
A-29	2	2	2	2	2
A-04	2	2	2	2	2
A-13	1	1	1	1	1
A-23	2	2	2	2	2
A-06	3	2	2	2	3
⋮	⋮	⋮	⋮	⋮	⋮
A-34	2	2	2	2	2
A-27	3	3	3	3	3
A-42	2	2	2	2	2
Prediksi Benar	15	14	14	14	

Tabel 8.Perbandingan Hasil Prediksi Klasifikasi FK-NN dengan Klasifikasi Data Asli pada 3-FCV.

e. Akurasi Hasil Prediksi Klasifikasi 3-FCV

Contoh perhitungan akurasi prediksi klasifikasi 3-FCV pada F1-NN, F3-NN, F5-NN dan F7-NN untuk 15 data subset 1.

$$a_{11} = \frac{15}{15} = 1 \qquad a_{15} = \frac{15}{15} = 1$$

$$a_{13} = \frac{14}{15} = 0,9333 \qquad a_{17} = \frac{14}{15} = 0,9333$$

Setelah menyelesaikan perhitungan subset 1 selanjutnya subset 2 dan subset 3 yang digunakan menjadi data *testing*. Kemudian dicari nilai rata-rata akurasi hasil prediksi untuk 3 subset. Hasil rata-rata persentase akurasi hasil prediksi klasifikasi pada 3-FCV dapat dilihat pada tabel 9.

Tabel 9. Persentase Akurasi Hasil Prediksi Klasifikasi pada 3-FCV

Batas FK-NN	Subset			Rata-rata	Akurasi (%)
	1	2	3		
F1-NN	1	0,733	1	0,9111	91,11
F3-NN	0,933	0,667	1	0,8667	86,67
F5-NN	0,933	0,667	1	0,8667	86,67
F7-NN	0,933	0,667	1	0,8667	86,67

Berdasarkan Tabel 9 persentase akurasi prediksi klasifikasi pada 3-FCV untuk F1-NN sebesar 91,11% dilanjutkan dengan persentase F3-NN, F5-NN dan F7-NN dengan persentase 86,67%.

Menentukan nilai K optimal dengan 5-Fold Cross Validation

Langkah dalam menentukan nilai K optimal dengan 5-FCV sama seperti 3-FCV,perbedaannya hanya terletak pada jumlah subset.

a. Menentukan Subset Data

Jumlah data dalam satu subset pada5-FCVadalah sebagai berikut:

Berdasarkan perhitungan, jumlah data dalam satu subset adalah 9 data. Masing-masing subset mendapat giliran untuk menjadi data *testing*. Berdasarkan data Tabel 3, subset 1 yang berisi 9 data yaitu sampel A-29 sampai A-45 digunakan sebagai data *testing* terlebih dahulu dan 2 subset lainnya yang berjumlah 36 data menjadi data *training*.

Langkah selanjutnya, menghitung jarak *Euclidean*, nilai keanggotaan setiap kelas hingga akurasi hasil prediksi klasifikasi 5-FCV. Hasil persentase akurasi prediksi klasifikasi 5-FCV dapat dilihat pada tabel 10.

Tabel 10. Persentase Akurasi Hasil Prediksi Klasifikasi pada 5-FCV

Batas FK-NN	Subset					Rata-rata	Akurasi (%)
	1	2	3	4	5		
F1-NN	1	0,8889	0,8889	1	1	0,9556	95,56
F3-NN	1	0,7778	0,6667	1	1	0,8889	88,89
F5-NN	0,8889	0,7778	0,6667	1	1	0,8667	86,67
F7-NN	0,8889	0,7778	0,6667	0,8889	1	0,8444	84,44

5. Penentuan K Optimal dari percobaan 1-FCV, 3-FCV dan 5-FCV

Dari percobaan 1-FCV didapatkan persentase akurasi prediksi klasifikasi untuk 1-FCV, begitu juga dengan percobaan 3-FCV dan 5-FCV. Hasil persentase akurasi prediksi klasifikasi gabungan dari ketiga percobaan tersebut dapat dilihat pada Tabel 11.

Tabel 11. Persentase Akurasi Penentuan Nilai K Optimal

FK-NN	Persentase Akurasi (%)		
	1-FCV	3-FCV	5-FCV
F1-NN	95,56	91,11	95,56
F2-NN	97,78	86,67	88,89
F3-NN	91,11	86,67	86,67
F4-NN	91,11	86,67	84,44

Berdasarkan Tabel 11, dapat disimpulkan bahwa K optimal yang akan digunakan pada Algoritma *Fuzzy K-Nearest Neighbor* (FK-NN) untuk memprediksi klasifikasi royalti batubara di CV Fazar Utama adalah F3-NN karena memiliki persentase akurasi tertinggi dibandingkan dengan FK-NN lainnya.

6. Menentukan Akurasi Prediksi FK-NN Berdasarkan Nilai K Optimal

Nilai K optimal F3-NN akan digunakan dalam algoritma FK-NN untuk memprediksi klasifikasi royalti batubara di CV Fazar Utama menggunakan proporsi 80:20 menggunakan data *training* dan *testing* seperti pada Tabel 1 dan Tabel 2. Langkah untuk melakukan prediksi klasifikasi royalti

batubara di CV Fazar Utama dengan menggunakan algoritma FK-NN adalah sebagai berikut:

a. Menghitung Jarak Euclidean

Sebagai contoh, di bawah ini dijelaskan perhitungan jarak *Euclidean* untuk data *training* sampel A-01 terhadap data *testing* pertama sampel A-46.

$$d(1,1) = \sqrt{(x_{11} - y_{11})^2 + \dots + (x_{15} - y_{15})^2}$$

$$= \sqrt{((-0,2587) - (-2,3955))^2 + \dots + (0,7297 - 2,9192)^2}$$

$$= 3,8567$$

Jarak *Euclidean* dihitung hingga data *training* sampel A-45. Kemudian dilanjutkan dengan mencari K tetangga terdekat menggunakan K optimal yaitu 3, dengan hasil *rank* perhitungan jarak *Euclidean* (d) yang telah dihitung untuk menghitung nilai keanggotaan.

b. Menghitung Nilai Keanggotaan dan Hasil Prediksi

Sebagai contoh, di bawah ini dijelaskan perhitungan nilai keanggotaan setiap kelas pada data *testing* pertama yaitu sampel A-46.

$$u(y_1, c_1) = \frac{(u(x_1, c_1) \times d(x_1, y_1)^{\frac{-2}{3-1}}) + \dots + (u(x_3, c_1) \times d(x_3, y_1)^{\frac{-2}{3-1}})}{(d(x_1, y_1)^{\frac{-2}{3-1}}) + \dots + (d(x_3, y_1)^{\frac{-2}{3-1}})}$$

$$= \frac{(0 \times 0,8238^{-1}) + \dots + (0 \times 3,2563^{-1})}{0,8238^{-1} + \dots + 3,2563^{-1}} = 0$$

$$u(y_1, c_2) = \frac{(u(x_1, c_2) \times d(x_1, y_1)^{\frac{-2}{3-1}}) + \dots + (u(x_3, c_2) \times d(x_3, y_1)^{\frac{-2}{3-1}})}{(d(x_1, y_1)^{\frac{-2}{3-1}}) + \dots + (d(x_3, y_1)^{\frac{-2}{3-1}})}$$

$$= \frac{(0 \times 0,8238^{-1}) + \dots + (1 \times 3,2563^{-1})}{0,8238^{-1} + \dots + 3,2563^{-1}} = 0,1307$$

$$u(y_1, c_3) = \frac{(u(x_1, c_3) \times d(x_1, y_1)^{\frac{-2}{3-1}}) + \dots + (u(x_3, c_3) \times d(x_3, y_1)^{\frac{-2}{3-1}})}{(d(x_1, y_1)^{\frac{-2}{3-1}}) + \dots + (d(x_3, y_1)^{\frac{-2}{3-1}})}$$

$$= \frac{(1 \times 0,8238^{-1}) + \dots + (0 \times 3,2563^{-1})}{0,8238^{-1} + \dots + 3,2563^{-1}} = 0,8693$$

Karena nilai $u(y_1, c_3) > u(y_1, c_1)$ dan $u(y_1, c_2)$, maka Data *testing* pertama atau sampel A-46 diprediksi masuk ke kelas 3 karena memiliki $u(y_1, c_3)$ atau nilai keanggotaan terbesar sebesar 0,8693.

Selanjutnya dilakukan hal yang sama untuk data *testing* kedua, ketiga dan seterusnya hingga data *testing* ke 11 (Sampel A-56). Hasil perhitungan nilai keanggotaan setiap kelas dan perbandingan hasil prediksi klasifikasi pada data *testing* yang telah dilakukan dapat dilihat pada Tabel 12. Berdasarkan Tabel 12, terdapat kesamaan prediksi klasifikasi F3-NN dengan klasifikasi pada data asli untuk data *testing*.

Tabel 12. Hasil Perhitungan Nilai Keanggotaan Setiap Kelas dan Perbandingan Hasil Prediksi Klasifikasi F3-NN dengan Klasifikasi Asli Untuk Data *Testing*

No	Sampel	Nilai Keanggotaan			Hasil Prediksi	Keterangan	
		Klasifikasi Asli	Kelas 1	Kelas 2			Kelas 3
1	A-46	3	0	0,1307	0,8693	3	Sama
2	A-47	3	0	0,1839	0,8161	3	Sama
3	A-48	2	0	1	0	2	Sama
4	A-49	2	0	1	0	2	Sama
5	A-50	2	0	1	0	2	Sama
6	A-51	1	0,6713	0,3287	0	1	Sama
7	A-52	2	0,3679	0,6321	0	2	Sama
8	A-53	2	0	1	0	2	Sama
9	A-54	1	1	0	0	1	Sama
10	A-55	1	0,7313	0,2687	0	1	Sama
11	A-56	1	0,7056	0,2944	0	1	Sama

c. Akurasi Prediksi FK-NN

Hasil akurasi prediksi klasifikasi F3-NN dengan klasifikasi asli untuk data *testing* adalah sebagai berikut:

$$\text{Persentase Akurasi} = \frac{11}{11} \times 100\% = 100\%$$

Berdasarkan perhitungan, dapat disimpulkan bahwa persentase akurasi prediksi klasifikasi royalti batubara di CV Fazar Utama menggunakan Algoritma *Fuzzy K-Nearest Neighbor* (FK-NN) dengan nilai F3-NN adalah 100%.

Kesimpulan

Berdasarkan hasil analisis maka hasil penelitian ini dapat disimpulkan sebagai berikut:

1. Nilai K optimal yang digunakan pada Algoritma *Fuzzy K-Nearest Neighbor* (FK-NN) untuk memprediksi klasifikasi royalti batubara di CV Fazar Utama dengan melakukan percobaan *1-Fold Cross Validation*, *3-Fold Cross Validation* dan *5-Fold Cross Validation* adalah F3-NN.
2. Persentase akurasi prediksi klasifikasi royalti batubara di CV Fazar Utama menggunakan Algoritma *Fuzzy K-Nearest Neighbor* (FK-NN) dengan nilai K yang sudah optimal yaitu F3-NN sebesar 100%.

Daftar Pustaka

Andhika, S. P. A., Indriati, dan Candra, D. (2017). Implementasi Algoritma *Fuzzy K-Nearest Neighbor* untuk penentuan lulus tepat waktu. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X*.

CV. Fazar, Utama. (2017). *Royalti dan Patokan Harga Batubara Indonesia*. Samarinda: PT Grace Coal.

- Hermawati, F. A. (2013). *Data Mining*. Yogyakarta: Andi.
- Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X*.
- Pandie, E. S. Y. (2012). Implementasi Algoritma Data Mining K-Neighbor (K-NN) dalam Pengambilan Keputusan Pengajuan Kredit. *Jurnal Ilmu Komputer Universitas Nusa Cendana*.
- Nugraha, S. D., Putri R. R. M., dan Wihandika, R. C. (2013). Penerapan Fuzzy K-Nearest
- Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi.
- Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB*. Yogyakarta: Andi.
- Widarjono, A. (2015). *Analisis Multivariat Terapan dengan Program SPSS, AMOS dan SMARTPLS Edisi Kedua*. Yogyakarta: UPM STIM YKPN.