

**Penerapan K-Nearst Neighbor Regression Dalam Memprediksi
Penjualan Produk Di Berdikari Kopi**

Alif Okta Febrian

Universitas Bhayangkara

alif.okta.febrian19@mhs.ubharajaya.ac.id

ABSTRACT

Sales is an activity or concrete business carried out by a seller to transfer the goods or services he owns to consumers with the aim of making a profit or profit. forecasting or predicting the goods they sell. The application of the K-Nearst Neighbor Regression Algorithm is widely used for data mining processes with the aim of obtaining prediction or forecasting results for products sold in the future. The process of extracting information from data is applied using the KDD (Knowledge Discovery Data) research method and implementing a data mining evaluation model by applying K-Fold Crossvalidation to determine the K value that will be used for data mining. In this research, it was found that the K-Nearst Neighbor Regression algorithm was successful in predicting future sales data for a café called Berdikari Kopi in order to streamline the stock of goods purchased.

Keywords : Prediction, K-Nearest Neighbor Regression, KDD (Knowledge Discovery Data, K-Fold Cross Validation.

ABSTRAK

Penjualan merupakan suatu kegiatan ataupun usaha konkret yang dilakukan oleh seorang penjual untuk memindahkan barang atau jasa yang dimilikinya kepada konsumen dengan tujuan mendapatkan laba ataupun keuntungan, dalam melakukan penjualan para pengiat pun harus mengikuti perkembangan zaman agar usaha yang dilakukannya dapat bertahan dengan salah satu caranya yaitu melakukan peramalan ataupun prediksi terhadap barang yang dijualnya. Penerapan Algoritma K-Nearst Neihbor Regression banyak digunakan untuk proses data mining dengan tujuan untuk mendapatkan hasil prediksi ataupun peramalan terhadap produk yang dijual kedepannya. Proses dalam menggali informasi dari suatu data diterapkan dengan metode penelitian KDD (Knowledge Discovery Data) dan penerapan evaluation model data mining dengan menerapkan K-Fold Crossvalidation untuk menentukan nilai K yang akan digunakan untuk data mining. Dalam penelitian ini didapatkan keberhasilan algoritma K-Nearst Neighbor Regression dalam memprediksi data penjualan sebuah café yang bernama Berdikari Kopi kedepannya guna mengefisiensikan stock barang yang dibelinya.

Kata kunci : Prediksi, K-Nearst Neighbor Regression, KDD (Knowledge Discovery Data, K-Fold Crossvalidation.

PENDAHULUAN

Penjualan merupakan suatu kegiatan ataupun usaha konkret yang dilakukan oleh seorang penjual untuk memindahkan barang atau jasa yang dimilikinya kepada konsumen dengan tujuan mendapatkan laba ataupun keuntungan[1]. Dalam melakukan penjualan para pengiat pun harus mengikuti perkembangan zaman agar usaha yang dilakukannya dapat bertahan dengan salah satu caranya yaitu melakukan peramalan ataupun prediksi terhadap barang yang dijualnya guna mengetahui

perkembangan pembelian barang tersebut. Peramalan atau prediksi merupakan perhitungan yang dilakukan untuk meramalkan keadaan prospektif khususnya penjualan yang akan terjadi dimasa mendatang terkait dengan hal yang sudah terjadi sebagai tolak ukur perhitungan yang selanjutnya hasil dari pada perhitungan tersebut dapat dijadikan bahan pertimbangan untuk volume penjualan kedepannya[2]. Proses peramalan ataupun prediksi dapat dilakukan dengan cara memproses sebuah data penjualan atau sering dikenal dengan proses data mining. Penerapan data mining sudah banyak digunakan diberbagai bidang salah satunya di bidang perdagangan, yang dimana proses tersebut digunakan untuk membantu dalam pengambilan keputusan terkait dengan peramalan ataupun prediksi. Dalam penerapannya data mining dapat digunakan dengan berbagai macam algoritma, salah satu algoritma yang cocok digunakan untuk melakukan peramalan ataupun prediksi yaitu algoritma K-Nearest Neighbor Regression yang merupakan algoritma untuk supervised learning atau unsupervised learning dengan jumlah data yang besar. Café di era saat ini sudah menjadi pemandangan sehari-sehari dengan banyaknya jenis café yang ada di Indonesia. Coffe shop merupakan salah satu café yang sangat banyak dan sering dikunjungi oleh berbagai kalangan di Indonesia. Berdasarkan artinya café merupakan bentuk daripada restoran informal yang dibentuk dengan tujuan membentuk rasa nyaman untuk bersantai para penikmatnya[3]. Berdikari Kopi merupakan salah satu dari coffe shop yang ada di Indonesia yang terletak di Tambun, Kabupaten Bekasi. Berdikari Kopi menyuguhkan tempat yang estetik dengan berbagai menu minuman serta makanan yang dapat dinikmati untuk para pelanggannya. Menu minuman yang disajikan terbuat dari bahan baku yang berbeda-beda sehingga memiliki kebutuhan yang berbeda setiap menunya. Berdikari Kopi menjadi 3 jenis kategori minuman yang dijualnya yaitu Espressobased, Milkbased, dan Mocktail Adapun data penjualan produk minuman Espressobasedberdikari kopi dari tahun 2020-2022 seperti yang diperlihatkan pada **Gambar 1.1**.

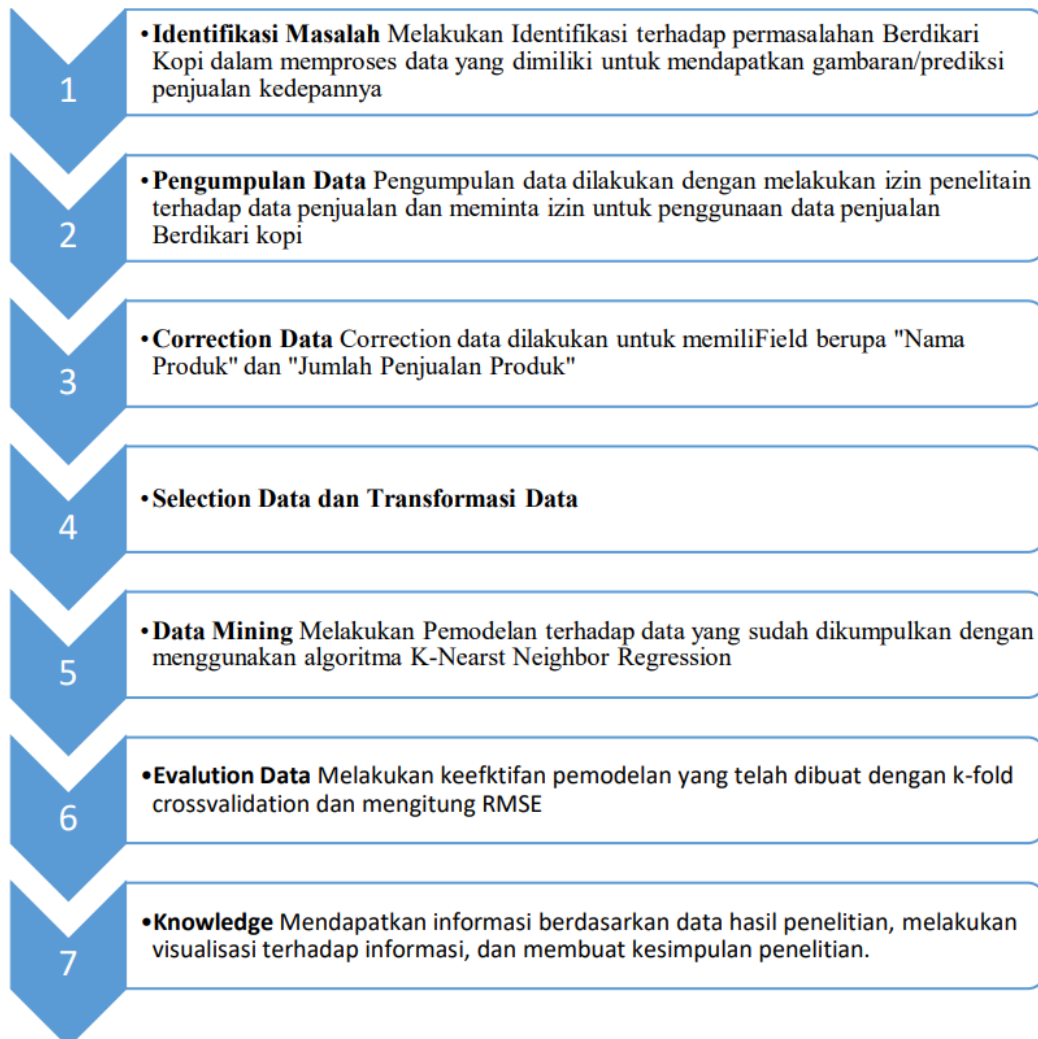
Gambar 1. Data Penjualan Berdikari Kopi

Produk	Tahun			Kategori
	2020	2021	2022	
Kopi Merdeka	795	716	829	Espressobased
Caramel Latte	539	500	482	
Hazelnate Latte	462	322	318	
Coffe Latte	612	612	631	
Vanila Latte	430	466	501	
Cappucino	920	867	800	
Manual Brew	675	628	710	

METODE PENELITIAN

Metode penelitian yang digunakan pada penelitian ini adalah metode penelitian KDD untuk penerapan data mining dalam menggali informasi data penjualan Berdikari Kopi. Adapun kerangka berfikir yang digunakan pada penelitian ini seperti diperlihatkan pada **Gambar 2.1**.

Gambar 2.1 Kerangka Penelitian Penerapan K-Nearest Neighbor Regression Dalam Memprediksi Penjualan Produk Di Berdikari Kopi



Keterangan:

1. **Identifikasi Masalah**
Identifikasi merupakan proses dimana dilakukannya observasi terhadap objek penelitian yaitu berdikari kopi dan mencari permasalahan yang ada untuk dijadikan bahan riset pada penelitian.
2. **Pengumpulan Data**
Pengumpulan data ini merupakan tahapan setelah dilakukannya observasi terhadap permasalahan yang ada, berdasarkan hasil observasi yang telah dilakukan di Berdikari kopi maka data yang dibutuhkan untuk riset penelitian ini yaitu data riwayat penjualan produk minuman berdikari kopi kategori espressobased mulai bulan Januari 2020 – Desember 2022.
3. **Correction Data**

Correction data merupakan tahapan dimana melakukan penyeleksianmelakukan pemilihan field dalam data yang dibutuhkan untuk penelitian. Data yang digunakan pada penelitian ini hanya data nama produk, jumlah produk terjual perbulan, dan kategori minuman pada bulan Januari 2020 – Desember 2022.

4. Selection dan Transformation

Data Tahapan ini dilakukan untuk melakukan penyeleksian terhadap data yang akan digunakan seperti duplikasi data dan data yang tidak konsisten, selanjutnya data tersebut dilakukan transformasi agregasi untuk agar dapat digunakan pada tahapan machine learning. MinMaxScaler pada penelitian ini digunakan untuk mentransformasikan data menjadi default [0,1].

5. Data mining

Tahapan ini merupakan tahapan pengimplementasian algoritma K-NN untuk mendapatkan ekstraksi informasi mengenai riwayat penjualan produk berdikari kopi.

6. Evaluation Data

Evaluation data digunakan untuk mengevaluasi hasil dari kinerja algoritma KNN yang digunakan. Evaluation juga bisa digunakan untuk mencari kemungkinan percobaan terbaik dengan metode seperti evaluasi K-fold crossvalidation terhadap data yang dilakukan pada pengujian model machine learning untuk mendapatkan iterasi K terbaik dengan harapan dapat men ekstrak data dengan baik, serta menerapkan RMSE untuk menentukan galat eror dari hasil pengetesan iterasi K-Fold Crossvalidation untuk mendapatkan nilai K terbaik dengan nilai RMSE yang rendah.

7. Knowledge

Tahapan ini merupakan tahapan dimana akan mendapatkan hasil dari tahapan tahapan sebelumnya untuk menghasilkan informasi yang baik serta melakukan visualisasi terhadap data informasi yang didapatkan.

2.1 Sumber Data

Sumber data dalam penelitian ini diperoleh dengan menggali data dari dua sumber yaitu:

1. Data Sekunder

Data sekunder yang diperoleh dalam penelitian ini didapatkan dari pemilik usaha dengan rincian data penjualan produk berdikari kopi.

2.2 Landasan Teori

2.2.1 Prediksi atau Peramalan

Prediksi merupakan suatu proses memperkirakan suatu yangkemungkinan yang paling mungkin akan terjadi dimasa depan secara sistematis berdasarkan informasi masa lalu yang sudah terjadi dengan tujuan mendapatkan jawaban terdekat yang akan terjadi dimasa depan. Peramalan atau biasa yang dikenal dengan forecasting dalam data mining merupakan perhitungan terhadap nilai – nilai sebuah

objek/variable yang saling berhubungan berdasarkan historitis dari objek ataupun variable tersebut[4]. Berdasarkan defisini mengenai peramalan (forcasting) maka dapat disimpulkan bahwa peramalan adalah suatu proses untuk memprediksi kejadian dimasa yang akan datang yang dilakukan secara sistematis demgam didasari dengan inforamsi yang didapatkan dimasa sebelumnya

2.2.2 Data Mining

Data Mining merupakan suatu teknik untuk mengekstraksi suatu informasi ataupun data yang berjumlah data besar dari database yang belum terpola menjadi data yang terstruktur dan dapat dijadikan informasi yang berguna untuk digunakan salah satunya untuk kebutuhan bisnis[5]. Data Mining juga dapat didefinisikan proses pengumpulan dan pengolahan data (sebelumnya tidak diketahui, bersifat implisit, dianggap tidak berguna) diubah menjadi informasi yang berpola dari jumlah data yang sangat besar. Dalam proses data mining terdapat metode yang dapat digunakan seperti KDD (Knowledge Discovery Database), SEMMA, CRISP-DM, dll, selain dari pada metode yang digunakan data mining memerlukan algoritma – algoritma untuk melakukan ekstraksi data menjadi informaasi, adapun algoritma – algoritma yang dapat digunakan seperti, Linear Regression, Support vector Machine, K-Nearest Neighbor, Naïve Bayes, dan lain-lain[6]. Berdasarkan definisi mengenai data mining maka dapat disimpulkan bahwa data mining merupakan pengekstrasian data berjumlah besar yang tidak terstruktur ataupun berpola diubah menjadi data yang dapat dijadikan informasi yang berguna sesuai dengan kebutuhan serta menerapkan metode seperti KDD, SEMMA, CRISP-DM dan algortima pendukungnya

2.2.3 Knowledge Discovery Database

Knowledge Discovey Database (KDD) merupakan proses ekstraksi data untuk mengidentifikasi sekumpulan data yang besar menjadi informasi potensial yang bermanfaat. Adapun tahapan – tahapan yang dilakukan seperti berikut:

1. Data Cleaning, Tahapan ini dilakukan pembersihan terhadap data yang tidak konsisten.
2. Data Integration, Tahapan ini merupakan tahapan untuk menggabungkan data dari berbagai sumber informasi data.
3. Data Selection, Tahapan ini digunakan untuk melakukan penyeleksian terhadap data yang sudah dikumpulkan dari berbagai sumber dengan memilih field data yang berguna untuk digunakan sebagai data penelitian.
4. Data Transformation, Tahapan ini dilakukan untuk mengubah/mengkosilidasikan data kedalam bentuk operasi yang siap digunakan untuk pemodelan machine learning.
5. Data Mining, Tahapan ini dilakukan untuk menggali informasi dari data yang sudah dikumpulkan guna mendapatkan informasi yang berguna.
6. Evalution, Tahapan ini merupakan tahapan yang digunakan untuk melakukan evaluasi terhadap pemodelan yang telah dilakukan sehingga dapat mencari tahu model terbaik untuk penambangan informasi dari data yang digunakan.

7. Knowledge, Tahapan ini merupakan tahapan terakhir untuk memperoleh informasi dari data yang diproses dan melakukan visualisasi informasi agar mudah untuk di pahami [7].

2.2.4 K-Nearest Neighbor Regression

KNN-Regression merupakan algoritma supervised learning yang dimana hasil dari query instance diklasifikasikan berdasarkan mayoritas kategori pada KNN-Regression dengan menggunakan atribut dan training data. Algoritma KNN Regression bekerja berdasarkan jarak terpendek dari query instance ke training data. Salah satu cara untuk menghitung jarak KNN-Regression menggunakan metode Euclidean Distance yang berfungsi untuk menguji ukuran yang diinterpretasikan kedekatan antara kedua objek. Adapun rumus persamaan dari Euclidean Distance seperti pada persamaan (2.1).

$$euc = \sqrt{\sum_k^d (a_k - b_k)^2} \quad (2.1)$$

Keterangan:

a = data latih

b = data uji

k = variable data

d = dimensi data

Algoritma KNN-Regression dalam penerapannya memerlukan pengetahuan tentang data latih, data uji, dan nilai k . Berikut penjelasan alur penggunaan Algoritma KNN-Regression sebagai berikut:

- a. Menentukan parameter k
- b. Menentukan jarak antara data latih dan data uji
- c. Menentukan jarak yang telah diperhitungkan
- d. Menentukan jarak yang terdekat sampai dengan nilai k [8].

Selanjutnya K-Nearest Neighbor Regression memberikan petunjuk ke pada nilai k terdekat regresi yang dimana tujuan dari pada pencarian tersebut mendapatkan output dari nilai yang di prediksi. K-Nearest Neighbor Regression mengamsumsikan lokalitas di dalam sekumpulan data dengan harapan x yang dimiliki memiliki nilai dan frekuensi yang sama terhadap y untuk $f(x)$ dan membuat x' dapat dimodelkan dengan rata - rata nilai output dari sample terdekat x . Maka dari itu rumus yang dapat di imlementasikan untuk K-Nearest Neighbor Regression untuk mendapatkan nilai prediksi dapat menggunakan persamaan rumus (2.2):

$$f(x') = \frac{1}{k} \sum_{i \in N_k(x')} y_i \quad ((2.2))$$

Keterangan:

x' = prediksi atau estimasi

$Nk(x')$ = Tetangga Terdekat

K = Jumlah Tetangga Terdekat

y_i = Output tetangga terdekat[9].

2.2.5 K-Fold Crossvalidation

Validasi silang k-fold, data awal dipartisi secara acak menjadi k secara mutual himpunan bagian eksklusif seperti K_1, K_2, \dots, K_n , masing-masing berukuran kira-kira sama. Pelatihan dan pengujian dilakukan sebanyak K_n kali. Pada iterasi i , partisi K_i dicadangkan sebagai set pengujian, dan partisi lainnya digunakan secara kolektif untuk melatih model. Artinya, diiterasi pertama, himpunan bagian K_2, \dots, K_n secara kolektif berfungsi sebagai set pelatihan untuk mendapatkan yang pertama model yang diuji pada K_1 . Berbeda dengan metode ketidaksepakatan dan subsampling acak, di sini setiap sampel digunakan dalam jumlah yang sama untuk pelatihan dan satu kali untuk pengujian[10].

HASIL DAN PEMBAHASAN

3.1 Identifikasi Masalah

Permasalahan yang ditemukan saat ini terjadi di Berdikari Kopi yaitu sering terjadinya kelebihan ataupun kekurangan stock biji kopi untuk pelaksanaan bisnis yang dijalani. Permasalahan yang terjadi ini membuat beberapa dampak kepada Berdikari Kopi salah satunya adalah kerugian dikarenakan kebelihan stock yang membuat kualitas kopi menurun dan tidak dapat digunakan. Maka dari itu Berdikari Kopi memerlukan prediksi penjualan kedepannya agar dapat mengantisipasi permasalahan ini. Data mining untuk menyelesaikan ini dapat menerapkan pedekaran Machine Learning menggunakan Algoritma K-Nearest Neighbor Regression untuk memprediksi penjualan kedepannya agar Berdikari Kopi dapat menentukan jumlah stock biji kopi yang akan digunakan dalam pelaksanaan bisnisnya.

3.2 Pengumpulan dan Pengolahan Data

Data yang digunakan pada penelitian ini merupakan data penjualan Berdikari Kopi tahun 2020, 2021, dan 2022 dengan produk espresobased (produk yang membutuhkan biji kopi sebagai bahan utamanya). Data yang didapatkan merupakan rekap data penjualan Berdikari Kopi perbulannya dengan sub menu yang lainnya, selanjutnya dilakukan pemilihan (Corectiion) data yang akan digunakan pada penelitian.

3.3 Implementasi

3.3.1 Pembentukan Dataset

Data penjualan Berdikari Kopi yang sudah didapatkan akan digunakan untuk proses prediksi menggunakan pendekatan machine learning, data yang akan digunakan sebagai data sampel untuk pemodelan yaitu nama produk dan jumlah penjualannya setiap bulannya seperti contoh **Gambar 3.1** merupakan data penjualan produk kopi merdeka pada tahun 2020, 2021, dan 2022.

Gambar 3.1 Data Penjualan Kopi Merdeka

Tahun	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
2020	79	66	45	40	38	80	80	76	75	43	75	98
2021	70	60	55	50	47	60	55	68	68	43	60	80
2022	77	70	63	60	58	73	64	76	77	58	65	88

Dataset yang digunakan pada proses data mining merupakan data penjualan 2020, 2021, dan 2022 dengan field tabel yang digunakan nama produk dan jumlah penjualan perbulannya. Dataset akan dibagi menjadi 2 bagian yaitu data training dan data test. Data training merupakan data penjualan produk selama 12 bulan dan bulan 13 merupakan target pada tahun tersebut, sebagai contoh data training yang akan digunakan yaitu data penjualan kopi merdeka tahun 2020 bukan ke-1 sampai bulan ke-12 dan bulan ke-1 tahun 2021 merupakan kolom target pada tahun 2020. Data testing merupakan data yang digunakan untuk memprediksi penjualan ditahun berikut dengan menggunakan data penjualan 12 bulan sebelumnya, sebagai contoh data penjualan bulan ke-2 2020 sampai dengan bulan ke-1 2021 merupakan data yang digunakan untuk memprediksi penjualan di bulan ke-2 tahun 2022. Adapun salah satu tabel pembagian dataset untuk menjadi data training dan data test setiap produknya diperlihatkan pada **Gambar 3.2 dan Gambar 3.3**

Gambar 3.2 Data Training

B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	Target
79	66	45	40	38	80	80	76	75	43	75	98	70
66	45	40	38	80	80	76	75	43	75	98	70	60
45	40	38	80	80	76	75	43	75	98	70	60	55
40	38	80	80	76	75	43	75	98	70	60	55	50
38	80	80	76	75	43	75	98	70	60	55	50	47
80	80	76	75	43	75	98	70	60	55	50	47	60
80	76	75	43	75	98	70	60	55	50	47	60	55
76	75	43	75	98	70	60	55	50	47	60	55	68
75	43	75	98	70	60	55	50	47	60	55	68	68
43	75	98	70	60	55	50	47	60	55	68	68	43
75	98	70	60	55	50	47	60	55	68	68	43	60
98	70	60	55	50	47	60	55	68	68	43	60	80
70	60	55	50	47	60	55	68	68	43	60	80	77
60	55	50	47	60	55	68	68	43	60	80	77	70
55	50	47	60	55	68	68	43	60	80	77	70	63
50	47	60	55	68	68	43	60	80	77	70	63	60
47	60	55	68	68	43	60	80	77	70	63	60	58
60	55	68	68	43	60	80	77	70	63	60	58	73
55	68	68	43	60	80	77	70	63	60	58	73	64
68	68	43	60	80	77	70	63	60	58	73	64	76

68	43	60	80	77	70	63	60	58	73	64	76	77
43	60	80	77	70	63	60	58	73	64	76	77	58
60	80	77	70	63	60	58	73	64	76	77	58	65
80	77	70	63	60	58	73	64	76	77	58	65	88

Gambar 3.3 Data Test

B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
60	55	50	47	60	55	68	68	43	60	80	77
55	50	47	60	55	68	68	43	60	80	77	70
50	47	60	55	68	68	43	60	80	77	70	63
47	60	55	68	68	43	60	80	77	70	63	60
60	55	68	68	43	60	80	77	70	63	60	58
55	68	68	43	60	80	77	70	63	60	58	73
68	68	43	60	80	77	70	63	60	58	73	64
68	43	60	80	77	70	63	60	58	73	64	76
43	60	80	77	70	63	60	58	73	64	76	77
60	80	77	70	63	60	58	73	64	76	77	58
80	77	70	63	60	58	73	64	76	77	58	65
77	70	63	60	58	73	64	76	77	58	65	88

3.3.2 Normalisasi Dataset

Normalisasi dataset digunakan untuk mengatur skala atas data satu dengan data yang lainnya. MinMaxScaler digunakan pada tahapan ini untuk mengatur skala antar data dengan harapan skala antara data [0,1] dengan artian transformation data paling rendah adalah 0 dan data paling besar adalah 1. Adapun penggunaan MinMaxScaler pada Machine Learning dijelaskan dalam bentuk source code sebagai berikut :

1. `from sklearn.preprocessing import MinMaxScaler` merupakan source code yang digunakan untuk mengimport fitur Sklearn pada preprocessing data.
2. `minmaxscaler = MinMaxScaler ()` merupakan pembuatan kelas/variable minmaxscaler dengan mengaktifkan fitur `MinMaxScaler()` dengan default skala antar data [0,1].
3. `Datatrain_norm = minmaxscaler.fit_transform(data)` merupakan source code yang berfungsi untuk membuat variable data yang merupakan dataset training ditransformasikan menggunakan minmaxscaler dan dimasukkan kedalam variable `Datatrain_norm`.
4. `Pd.DataFrame(Datatrain_norm)` merupakan source code yang digunakan untuk menampilkan `Datatrain_norm` kedalam bentuk tabel. Berikut hasil dari penggunaan MinMaxScaler terhadap dataset training kopi merdeka yang diperlihatkan pada **Gambar 3.4**.

Gambar 3.4 Normalisasi Dataset

0	0.683333	0.466667	0.116667	0.033333	0.000000	0.672727	0.672727	0.600000	0.581818	0.000000	0.581818	1.000000	0.600000
1	0.466667	0.116667	0.033333	0.000000	0.700000	0.672727	0.600000	0.581818	0.000000	0.581818	1.000000	0.490909	0.377778
2	0.116667	0.033333	0.000000	0.700000	0.700000	0.600000	0.581818	0.000000	0.581818	1.000000	0.490909	0.309091	0.266667
3	0.033333	0.000000	0.700000	0.700000	0.633333	0.581818	0.000000	0.581818	1.000000	0.490909	0.309091	0.218182	0.155556
4	0.000000	0.700000	0.700000	0.633333	0.616667	0.000000	0.581818	1.000000	0.490909	0.309091	0.218182	0.127273	0.088889
5	0.700000	0.700000	0.633333	0.616667	0.083333	0.581818	1.000000	0.490909	0.309091	0.218182	0.127273	0.072727	0.377778
6	0.700000	0.633333	0.616667	0.083333	0.616667	1.000000	0.490909	0.309091	0.218182	0.127273	0.072727	0.309091	0.266667
7	0.633333	0.616667	0.083333	0.616667	1.000000	0.490909	0.309091	0.218182	0.127273	0.072727	0.309091	0.218182	0.555556
8	0.616667	0.083333	0.616667	1.000000	0.533333	0.309091	0.218182	0.127273	0.072727	0.309091	0.218182	0.454545	0.555556
9	0.083333	0.616667	1.000000	0.533333	0.366667	0.218182	0.127273	0.072727	0.309091	0.218182	0.454545	0.454545	0.000000
10	0.616667	1.000000	0.533333	0.366667	0.283333	0.127273	0.072727	0.309091	0.218182	0.454545	0.454545	0.000000	0.377778
11	1.000000	0.533333	0.366667	0.283333	0.200000	0.072727	0.309091	0.218182	0.454545	0.454545	0.000000	0.309091	0.822222
12	0.533333	0.366667	0.283333	0.200000	0.150000	0.309091	0.218182	0.454545	0.454545	0.000000	0.309091	0.672727	0.755556
13	0.366667	0.283333	0.200000	0.150000	0.366667	0.218182	0.454545	0.454545	0.000000	0.309091	0.672727	0.618182	0.600000
14	0.283333	0.200000	0.150000	0.366667	0.283333	0.454545	0.454545	0.000000	0.309091	0.672727	0.618182	0.490909	0.444444
15	0.200000	0.150000	0.366667	0.283333	0.500000	0.454545	0.000000	0.309091	0.672727	0.618182	0.490909	0.363636	0.377778
16	0.150000	0.366667	0.283333	0.500000	0.500000	0.000000	0.309091	0.672727	0.618182	0.490909	0.363636	0.309091	0.333333
17	0.366667	0.283333	0.500000	0.500000	0.083333	0.309091	0.672727	0.618182	0.490909	0.363636	0.309091	0.272727	0.666667
18	0.283333	0.500000	0.500000	0.083333	0.366667	0.672727	0.618182	0.490909	0.363636	0.309091	0.272727	0.545455	0.466667
19	0.500000	0.500000	0.083333	0.366667	0.700000	0.618182	0.490909	0.363636	0.309091	0.272727	0.545455	0.381818	0.733333
20	0.500000	0.083333	0.366667	0.700000	0.650000	0.490909	0.363636	0.309091	0.272727	0.545455	0.381818	0.600000	0.755556
21	0.083333	0.366667	0.700000	0.650000	0.533333	0.363636	0.309091	0.272727	0.545455	0.381818	0.600000	0.618182	0.333333
22	0.366667	0.700000	0.650000	0.533333	0.416667	0.309091	0.272727	0.545455	0.381818	0.600000	0.618182	0.272727	0.488889
23	0.700000	0.650000	0.533333	0.416667	0.366667	0.272727	0.545455	0.381818	0.600000	0.618182	0.272727	0.400000	1.000000

3.3.3 Pembagian Variabel Data

Tahapan ini melakukan pembagian variable data yang akan digunakan dengan menggunakan variable x dan variable y. Adapun script yang digunakan pada tahapan ini seperti yang diperlihatkan pada **Gambar 3.5**.

Gambar 3.5 Pembagian Variabel Data

```
X = datatrain_norm[:, [0,1,2,3,4,5,6,7,8,9,10,11]]
y_norm = datatrain_norm[:, [-1]]
y = y_norm.reshape(24)
```

3.3.4 K-Fold Crossvalidation

Tahapan ini merupakan tahapan validasi pembagian data yang akan digunakan dengan menerapkan beberapa kemungkinan data percobaan. Dengan menerapkan percobaan validasi sebanyak 10 kali. Adapun script yang digunakan seperti yang diperlihatkan pada **Gambar 3.5** dan hasil penerapan dari k-fold crossvalidation seperti pada **Gambar 3.6**

Gambar 3.6 Hasil K-Fold Crossvalidation

```
TRAIN: [ 0 1 2 4 5 6 7 8 9 10 11 12 14 15 16 17 19 20 21 22 23] TEST: [ 3 13 18]
TRAIN: [ 0 1 2 3 4 5 6 7 8 9 10 11 12 13 15 16 18 19 21 22 23] TEST: [14 17 20]
TRAIN: [ 0 1 3 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20 21 22 23] TEST: [ 2 4 10]
TRAIN: [ 0 1 2 3 4 5 8 9 10 11 12 13 14 15 16 17 18 20 21 22 23] TEST: [ 6 7 19]
TRAIN: [ 0 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 22 23] TEST: [ 1 21]
TRAIN: [ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 17 18 19 20 21 22 23] TEST: [ 0 16]
TRAIN: [ 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 16 17 18 19 20 21 22] TEST: [15 23]
TRAIN: [ 0 1 2 3 4 5 6 7 8 10 11 12 13 14 15 16 17 18 19 20 21 23] TEST: [ 9 22]
TRAIN: [ 0 1 2 3 4 5 6 7 9 10 11 13 14 15 16 17 18 19 20 21 22 23] TEST: [ 8 12]
TRAIN: [ 0 1 2 3 4 6 7 8 9 10 12 13 14 15 16 17 18 19 20 21 22 23] TEST: [ 5 11]
```

3.3.5 Splitting Data

Tahapan berikut merupakan pembagian data training dan data testing yang akan digunakan untuk pemodelan pada penerapan data mining. Adapun script dari pembagian data training dan data test diperlihatkan pada **Gambar 3.7**

Gambar 3.7 Script Splitting Data

```
for train_index, test_index in kf.split(X,y):
    print("-----")
    print("Index TRAIN:", train_index, "Index TEST:", test_index)
    X_train, X_test = X[train_index], X[test_index]
    y_train, y_test = y[train_index], y[test_index]
    print("X_TRAIN: \n",X_train,"\n X_Test: \n",X_test,"\n Y_Train: \n", y_train,"\nY_Test: ", y_test)
    print("-----\n\n")
```

3.3.6 K-Nearst Neighbor Regression

Tahapan selanjutnya yaitu menerapkan algoritma K-Nearst Neighbor Regression yang digunakan untuk pemodelan pada machine learning. Penerapan K-fold Crossvalidation digunakan sebelum dilakukan pemodelan K-Nearst Neighbor Regression untuk dapat menentukan nilai k terbaik yang akan digunakan untuk pemodelan. Adapun script yang digunakan seperti yang diperlihatkan pada **Gambar 3.8**.

Gambar 3.8 Script K-Nearst Neighbor

```
from sklearn.neighbors import KNeighborsRegressor
model = KNeighborsRegressor(metric='euclidean')
model

KNeighborsRegressor
KNeighborsRegressor(metric='euclidean')

from matplotlib import pyplot as plt
k_range = range(1, 15)
rmse_err = []
for k in k_range:
    model = KNeighborsRegressor(n_neighbors=k, metric='euclidean')
    scores = cross_val_score(model, X, y, scoring='neg_mean_absolute_error', cv=kf)
    RMSE = sqrt(mean(abs(scores)))
    rmse_err.append(RMSE)
    print('RMSE value for k= ', k, 'is:', RMSE)
plt.plot(k_range, rmse_err, '-ok')
plt.xlabel('Value of K For KNN')
plt.ylabel('RMSE')
plt.grid(True)
plt.show()
```

Setelah mendapatkan formula terbaik untuk pemodelan ini selanjutnya membuat variable/fungsi kelas prediksi terhadap data test yang dimiliki untuk mendapatkan hasil prediksi penjualan kedepannya. Adapun script yang digunakan untuk prediksi seperti di perlihatkan pada **Gambar 3.9** dan hasil dari script yang digunakan diperlihatkan pada **Gambar 3.10**.

Gambar 3.9 Script Prediksi K-Nearst Neighbor Regression

```
y_pred = modelfix.predict(x_tes)
y_pred
data_test['prediksi'] = y_pred
data_test
```

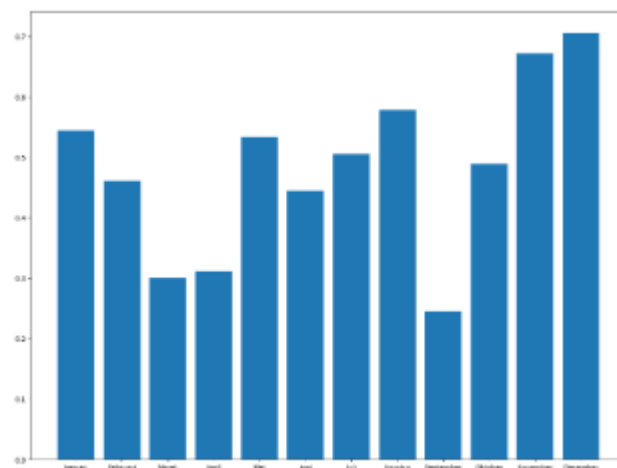
Gambar 3.10 Hasil Prediksi K-Nearst Neighbor Regression

	Bln1	Bln2	Bln3	Bln4	Bln5	Bln6	Bln7	Bln8	Bln9	Bln10	Bln11	Bln12	prediksi
0	60	55	50	47	60	55	68	68	43	60	80	77	0.544444
1	55	50	47	60	55	68	68	43	60	80	77	70	0.461111
2	50	47	60	55	68	68	43	60	80	77	70	63	0.300000
3	47	60	55	68	68	43	60	80	77	70	63	60	0.311111
4	60	55	68	68	43	60	80	77	70	63	60	58	0.533333
5	55	68	68	43	60	80	77	70	63	60	58	73	0.444444
6	68	68	43	60	80	77	70	63	60	58	73	64	0.505556
7	68	43	60	80	77	70	63	60	58	73	64	76	0.577778
8	43	60	80	77	70	63	60	58	73	64	76	77	0.244444
9	60	80	77	70	63	60	58	73	64	76	77	58	0.488889
10	80	77	70	63	60	58	73	64	76	77	58	65	0.672222
11	77	70	63	60	58	73	64	76	77	58	65	88	0.705556

3.3.7 Visualisasi

Tahapan ini digunakan untuk memvisualisasikan hasil prediksi agar mudah di baca dan dipahami. **Gambar 3.11** merupakan visualisasi dari hasil prediksi penjualan kedepannya berdasarkan hasil data training dan data test kopi merdeka

Gambar 3.11 Visualisasi Hasil Prediksi Kopi Merdeka



KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai prediksi penjualan produk Berdikari Kopi dengan penerapan algoritma K-Nearest Neighbors Regression maka dapat disimpulkan sebagai berikut:

1. Algoritma K-Nearest Neighbors Regression berhasil memprediksi penjualan produk Berdikari Kopi kedepannya dengan mampu memprediksi penjualan terbanyak pada produk kopi merdeka terjadi pada bulan Desember, Caramel Latte pada bulan Desember, Hazelnut Latte pada bulan Desember, Coffe Latte memiliki prediksi yang hampir rata disetiap bulannya namun pada bulan April diprediksi mejadi bulan dengan penjualan terbanyak, Vanilla Latte diprediksi memiliki penjualan terbanyak pada bulan Desember, Cappucino diprediksi memiliki penjualan terbanyak pada bulan November, Manual Brew diprediksi memiliki penjualan terbanyak pada bulan Februari.
2. Nilai akurasi Error atau RMSE yang didapatkan pada setiap data yang training dan test memiliki nilai yang berbeda-beda sebagai berikut:
 - a. Kopi Merdeka Nilai K = 4, RMSE = 0,42042747205583425
 - b. Caramel Latte Nilai K = 1, RMSE = 0,37578201289061347
 - c. Hazelnute Latte Nilai K = 1, RMSE = 0,5287564026405102
 - d. Coffe Latte Nilai K = 12, RMSE = 0,4364682389658536
 - e. Vanilla Latte Nilai K= 1, RMSE = 0,4055535528269063
 - f. Cappucino Nilai K = 7, RMSE = 0,4897406848275423
 - g. Manual Brew Nilai K = 5, RMSE = 0,46095139708671073
3. Analisis yang dilakukan selanjutnya untuk menentukan stock biji kopi dilihat dari hasil prediksi yang didapatkan bahwa produk Berdikari Kopi akan menagalami paling banyak pesanan yaitu pada bulan Desember.

DAFTAR PUSTAKA

- A. M. Siregar and A. Puspabhuana, DATA MINING. CV Kekata Group, 2017
- A. W. Alfani P R et al., "Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor," pp. 155–160, 2021.
- C. Beaumont, S. Makridakis, S. C. Wheelwright, and V. E. McGee, "Forecasting: Methods and Applications," in *Forecasting: Methods and Applications*, vol. 35, no. 1, JSTOR, 1984, p. 151. doi: 10.2307/2581936.
- D. Nurfauzan and T. Fatimah, "Implementasi Algoritma K-Nearest Neighbors Regression Dalam Memprediksi Harga Saham," Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta Indonesia, 2022, [Online]. Available: <https://senafti.budiluhur.ac.id/index.php/>
- E. Supramono, "Analisis Pengaruh Price dan Promotion Terhadap Penjualan Rokok Pada Swalayan Bintang di Sampit," *Jurnal Terapan Manajemen dan Bisnis*, pp. 1–1, 2015.
- G. Triyandana, L. A. Putri, and Y. Umaidah, "Penerapan Data Mining Pengelompokan

Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means,” *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, no. 1, p. 40, 2022, [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>

- J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, 3rd ed. USA: Morgan kaufmann, 2012.
- J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, 3rd ed. USA: Morgan kaufmann, 2012.
- J. Suntoro, *Data Mining : Algoritma dan Implementasi dengan pemograman PHP*. PT Elex Media Komputindo, 2019
- O. Kramer, “Unsupervised K-Nearest Neighbor Regression,” Jul. 18, 2011. [Online]. Available: <http://arxiv.org/abs/1107.360>