

Analisis Prediksi Harga Beras Berbasis Kualitas Menggunakan Algoritma *K-Nearest Neighbor*

Dedi Setiadi¹⁾, Efan²⁾

^{1), 2)} Program Studi Teknik Informatika, Institut Teknologi Pagar Alam
Jalan Masik Siagim No.75 Kota Pagar Alam, Sumatera Selatan
Email : dedisetiadi1212@gmail.com¹⁾, efan@itpa.ac.id²⁾

ABSTRACT

Rice is an important element in improving national food security. There are several factors that affect the price of rice, including uncertain production. This study aims to predict the price of rice in the city of Pagar Alam using the *K-Nearest Neighbor (K-NN)* algorithm. The data used is rice price data from 2019 to 2023 obtained from the Pagar Alam City Trade Office. The resulting classification model is evaluated for its performance using various indicators such as accuracy, precision, and recall. The research process begins with a business understanding to identify factors that affect rice prices. Furthermore, data understanding was carried out on 240 rice data samples including the attributes of year, month, production, price, and type of rice. This data is processed using the *K-NN* method, which involves calculating the Euclidean distance between training data and testing data and testing accuracy through split data that the *K-Nearest Neighbor* method is effective in predicting rice prices with an accuracy of 95%, which shows that this model has a good ability to classify rice prices accurately. This research has been successful and can help the Trade Service in making decisions regarding rice prices and be a reference for similar research in the future.

Keywords : Beras, CRISP-DM, Data Mining, Prediksi, K-NN

ABSTRAK

Beras merupakan unsur yang penting dalam meningkatkan ketahanan pangan nasional. Terdapat beberapa faktor yang mempengaruhi harga beras tersebut diantaranya yaitu produksi yang tidak menentu. Penelitian ini bertujuan untuk memprediksi harga beras di kota Pagar Alam dengan menggunakan algoritma *K-Nearest Neighbor (K-NN)*. Data yang digunakan adalah data harga beras dari tahun 2019 hingga 2023 yang diperoleh dari Dinas Perdagangan Kota Pagar Alam. Model klasifikasi yang dihasilkan dievaluasi performanya menggunakan berbagai indikator seperti akurasi, presisi, dan recall. Proses penelitian dimulai dengan pemahaman bisnis untuk mengidentifikasi faktor-faktor yang mempengaruhi harga beras. Selanjutnya, dilakukan pemahaman data terhadap 240 sampel data beras yang mencakup atribut tahun, bulan, produksi, harga, dan jenis beras. Data ini diolah menggunakan metode *K-NN*, yang melibatkan perhitungan jarak Euclidean antara data training dan data testing serta pengujian akurasi melalui split data bahwa metode *K-Nearest Neighbor* efektif dalam memprediksi harga beras dengan akurasi sebesar 95%, yang menunjukkan bahwa model ini memiliki kemampuan yang baik dalam mengklasifikasikan harga beras secara akurat. Penelitian ini telah berhasil dan dapat membantu Dinas Perdagangan dalam pengambilan keputusan terkait harga beras dan menjadi referensi bagi penelitian serupa di masa mendatang.

Kata Kunci : Beras, CRISP-DM, Data Mining, Prediksi, K-NN

1. Pendahuluan

Seiring berjalannya waktu, perkembangan teknologi informasi dan komunikasi sangat pesat, dan peranannya dalam kehidupan sehari-hari dapat dilihat di berbagai bidang kehidupan manusia, termasuk pertanian. (Sari, 2023). Di era globalisasi, pembangunan pertanian berkelanjutan tidak lepas dari dampak pesatnya perkembangan ilmu pengetahuan dan teknologi, (Arie, 2019) pada sektor pertanian yang disebut dengan *smart farming* (Mukti & Saputri, 2023) dengan memanfaatkan *machine learning*, yang dapat menjadi solusi untuk mencapai ketahanan pangan. (Satria et al., 2023) *Machine learning* adalah cabang kecerdasan buatan (AI) yang berfokus pada pengembangan algoritma dan model yang memungkinkan komputer belajar dari data dan membuat prediksi atau keputusan tanpa diprogram secara eksplisit. *Machine learning* (Fathurohman, 2021) sering digunakan dalam proses data mining untuk membantu menemukan pola yang kompleks dan membuat prediksi.

Data mining adalah (Firmansyah & Nurdiawan, 2023) proses pengumpulan data penting dalam jumlah besar atau *big data*. Proses ini seringkali menggunakan sejumlah metode, seperti penggunaan matematika, statistik, dan teknologi kecerdasan buatan. (Ikhwan & Aslami, 2020) Data mining, sering juga disebut *Knowledge Discovery In Database* (KDD) (Takdirillah, 2020) adalah aktivitas yang melibatkan pengumpulan data historis dan menggunakannya untuk menemukan keteraturan, pola, atau hubungan dalam data besar. Hasil dari data mining ini dapat digunakan untuk pengambilan keputusan di masa depan atau prediksi.

Prediksi atau peramalan (Setiadi & Syahri, 2022) merupakan proses perkiraan yang berupa pengukuran besarnya jumlah sesuatu pada waktu yang akan datang berdasarkan data masa lampau yang di analisis secara ilmiah dengan metode tertentu. Dalam melakukan peramalan atau prediksi yang menjadi faktor utama adalah pemilihan metode karena berpengaruh terhadap hasil peramalan. (Seruni et al., 2020) Dengan demikian, diperlukan suatu metode komputasi cerdas yang dapat melakukan prediksi tersebut yaitu menggunakan algoritma *K-Nearest Neighbor* (K-NN). (Hasan & Wati, 2021) Karena metode *K-Nearest Neighbor* memiliki beberapa kelebihan yaitu ketangguhan terhadap training data yang memiliki banyak noise dan efektif apabila training datanya besar. (Rahmadini et al., 2023)

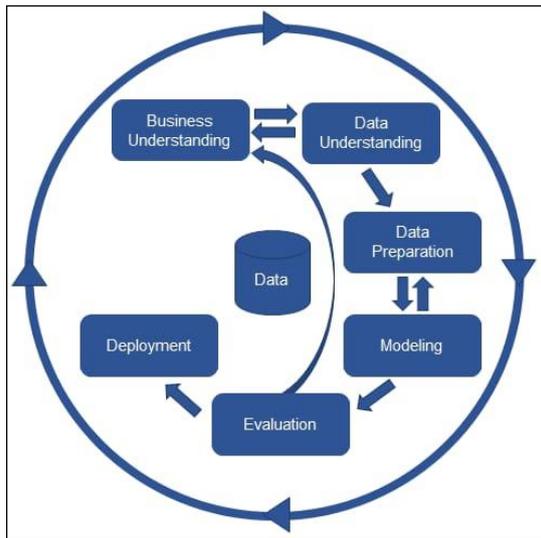
Penerapan algoritma yang digunakan pada data mining menggunakan algoritma *K-Nearest Neighbor*, (Virdaus & Prasetyaningrum, 2020) yang mana alasan pemilihan algoritma ini adalah karena algoritma ini memiliki cara hitung yang sederhana dan simpel namun memiliki tingkat akurasi yang tinggi dan termasuk salah satu dari sepuluh algoritma atau metode paling populer pada perhitungan data mining, serta juga tangguh terhadap data *training sample* yang noisy. (Asminah et al., 2023)

Salah satu makanan pokok yang dikonsumsi oleh penduduk di dunia adalah beras. Menurut FAO dalam, disebutkan bahwa beras adalah makanan pokok bagi lebih dari setengah populasi dunia, termasuk Indonesia. Mayoritas masyarakat Indonesia menjadikan beras yang telah diolah menjadi nasi sebagai makanan pokok sehari-hari. Oleh karena itu, stabilisasi pasokan dan harga beras merupakan unsur yang penting dalam meningkatkan ketahanan pangan dan mendorong pembangunan nasional

Berdasarkan wawancara dan observasi pada dinas perdagangan kota Pagar Alam cara memprediksi harga beras dilakukan dengan metode *survey* langsung ke pasar, dan data tersebut tidak dilakukan pengolahan lanjutan apalagi jika harga beras tidak menentu atau berubah-ubah, dimana jika saat pasokan beras meningkat dan permintaan sedikit maka harga akan mengalami penurunan begitupun jika pasokan beras sedikit tetapi permintaan banyak maka harga akan melonjak, hal ini karena ada beberapa faktor yang mempengaruhi harga beras tersebut diantaranya tingkat produksi yang tidak menentu. Maka tentu saja hal ini menjadi bahan pertimbangan bagaimana cara mengantisipasi masalah ini. Untuk dapat mengantisipasi masalah yang dihadapi maka dibutuhkan suatu metode untuk memprediksi serta dapat mengkaji keadaan atau permasalahan yang sedang dialami sebenarnya, seperti halnya harga beras tersebut. Karena apabila analisis data ini dikaji dan digali dengan mengimplementasikan metode yang akurat maka dapat mengetahui pola serta pengetahuan untuk mengambil keputusan.

2. Metode

Cross Industry Standard Process For Data Mining (CRISP-DM) (Setiadi et al., 2024) merupakan model dari proses yang menyediakan kerangka kerja untuk melaksanakan proyek data mining yang independent dari sektor industri dan teknologi yang digunakan. CRISP-DM bertujuan untuk membuat proyek data mining yang besar, lebih cepat dan mudah untuk dikelola (Muallif et al., 2023). CRISP-DM merupakan salah satu metode yang digunakan untuk menyelesaikan permasalahan yang dalam suatu penelitian. Metodologi ini terdiri dari enam tahapan yaitu *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment*. Gambar 1 menunjukkan model CRIPS-DM. (Hasanah et al., 2021)



Gambar 1. Metode CRISP-DM

a. Pemahaman Bisnis

Pada tahap pertama ini diperlukan pemahaman tentang faktor-faktor bisnis yang mempengaruhi harga pasar beras. Dengan pemahaman bisnis yang baik prediksi harga beras dapat lebih akurat. Oleh karena itu, pemahaman bisnis menjadi dasar penting dalam membangun model prediksi harga beras yang efektif. (Muallif et al., 2023)

b. Pemahaman Data

Pemahaman terhadap kebutuhan data terkait dengan Prediksi harga beras, data diambil dari Dinas Perdagangan pemahaman data mengacu pada pengelompokkan harga beras dengan data yang dikumpulkan yaitu data harga beras selama 5 tahun pada tahun 2019 Sampai dengan 2023 dengan jumlah 240 data.

c. Persiapan Data

Setelah tahap pemahaman data partisipatif selesai, tahap pengolahan data ini menghasilkan kumpulan data akhir yang diterapkan pada pemodelan data. Dalam tahapan ini yaitu membangun dataset akhir dari berupa data mentah. Ada beberapa hal yang akan dilakukan mencakup melakukan pembersihan data (*Data Cleaning*), dataset yang sudah dilakukan proses cleaning kemudian akan dilakukan pemilihan data (*Data Selection*), record dan atribut-atribut. (Pambudi, 2023)

d. Pemodelan

Pada tahapan ini, bermacam teknik pemodelan yang dipilih dan diaplikasikan pada kumpulan data yang disiapkan. untuk penentuan teknik data mining, alat bantu data mining serta algoritma data mining. Adapun pemodelan yang dilakukan pada penelitian ini adalah untuk prediksi harga beras berdasarkan kualitas dengan metode K-NN. Setelah dataset melewati tahap processing atau pengolahan data awal didapatkan berupa dataset yang sudah siap diolah menggunakan *K-Neares Neighbor*. Agar mendapatkan akurasi yang lebih optimal digunakan aplikasi *Rapid Miner* (Nahjan et al., 2023) sebagai *tools* pengolahan dataset.

e. Evaluasi

Pada tahap ini telah difokuskan pada model yang dihasilkan sudah sesuai standar K-NN dan tidak ada yang dilewatkan saat melakukan tahap awal hingga tahap pemodelan selesai dengan pengujian menggunakan *Root Mans Square Error* (RMSE). (Irnanda et al., 2022) Metode ini salah satu cara standar untuk mengukur kesalahan suatu model dalam memprediksi. RMSE merupakan metode yang cukup sering digunakan dalam mengevaluasi kinerja prediksi. Dengan menggunakan RMSE, error yang ada menunjukkan seberapa besar perbedaan hasil estimasi dengan hasil yang akan diestimasi. Hal yang membuat berbeda karena adanya keacakan pada data karena tidak mengandung estimasi yang lebih akurat. RMSE digunakan untuk mencari nilai keakurasian hasil peramalan dengan data history dengan menggunakan rumus yang ditunjukkan dengan nilai kesalahan. Berikut ini merupakan rumus untuk perhitungan RMSE :

$$dst(X_1 - X_2) = \sqrt{\sum_{i=1}^n (X_{1i} - X_{2i})^2} \tag{1}$$

f. Penyebaran

Pada tahap terakhir ini CRIPS-DM adalah (Mirantika, 2021) penyebaran pengetahuan atau informasi yang telah diperoleh selanjutnya akan di implementasikan pada laporan dan melakukan analisis tiap analisis prediksi harga beras yang diperoleh supaya dapat mudah dimengerti oleh pihak Dinas Perdagangan Pagur Alam.

3. Hasil dan Pembahasan

Setelah penelitian ini menghasilkan model klasifikasi prediksi harga beras didapatkan dengan algoritma dan indikator-indikator yang ada ialah yang menggambarkan performa seperti *accuracy*, *presisi* dan *recal*. Data yang digunakan adalah data harga beras tahun 2019 sampai 2023 dengan jenis beras premim dan medium dari harga 8500 sampai dengan 15000 di Dinas Perdagangan. Hasil dari indikator-indikator tersebut merupakan rangkaian dari proses panjang perhitungan algoritma *K-Nearest Neighbor* mulai dari perhitungan mencari nilai K kemudian menghitung jarak antara *data testing* dan *data training*. Mengurutkan jarak dari yang terkecil ke terbesar, lalu menentukan jumlah label mayoritas berdasarkan nilai K, selain itu juga dilakukan pengujian menggunakan *split* data untuk menghitung nilai akurasi.

Hasil dari perhitungan tersebut dilakukan hingga masuk pada perhitungan pengukur jarak. Pengurutan jarak dilakukan menggunakan *euclidean distance*, dengan merubah, jenis beras akan menjadi nilai numerik dimana beras premium adalah 1 dan beras medium adalah 2, berikut adalah langkah-langkah perhitungan *euclidean distance* :

$$\begin{aligned}
 &2019, 1, 2, 11000, 1 \\
 &= \sqrt{(2 - 2)^2 + (14000 - 11000)^2 + (1 - 1)^2} \\
 &= \sqrt{0^2 + 3000^2 + 0^2} \\
 &= \sqrt{0 + 9000000 + 0} \\
 &= \sqrt{9000000} \\
 &= 3000 \dots\dots\dots(2)
 \end{aligned}$$

$$\begin{aligned}
 &2019, 1, 1.3, 11000, 1 \\
 &= \sqrt{(2 - 1.3)^2 + (14000 - 11000)^2 + (1 - 1)^2} \\
 &= \sqrt{0.7^2 + 3000^2 + 0^2} \\
 &= \sqrt{0.49 + 9000000 + 0} \\
 &= \sqrt{9000000.49} \\
 &= 30000081\dots\dots\dots(3)
 \end{aligned}$$

$$\begin{aligned}
 &2019, 1, 3, 10000, 2 \\
 &= \sqrt{(2 - 3)^2 + (14000 - 10000)^2 + (1 - 2)^2} \\
 &= \sqrt{(-1)^2 + 4000^2 + 1^2} \\
 &= \sqrt{1 + 16000000 + 1} \\
 &= \sqrt{16000002} \\
 &= \dots\dots\dots 4000.01 \\
 &\dots\dots\dots(4)
 \end{aligned}$$

3.1. Business Understanding

Pada tahap pertama ini diperlukan pemahaman tentang faktor-faktor bisnis yang mempengaruhi harga pasar beras. Dengan pemahaman bisnis yang baik prediksi harga beras dapat lebih akurat. Oleh karena itu, pemahaman bisnis menjadi dasar penting dalam membangun model prediksi harga beras yang efektif. Tujuan bisnis dari pengolahan data ini dapat dijadikan sebagai salah satu dasar pengambilan keputusan penelitian ini menerapkan metode KNN untuk mengklasifikasi data. (Suhanda et al., 2020) Dalam pengklasifikasian dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan agar mengurangi ke tidak tepatan dalam menentukan harga beras berdasarkan kualitas. Strategi awal yang diterapkan adalah observasi ke Dinas Perdagangan kota Pagar Alam lalu meminta data harga beras dengan cara melakukan memberikan surat pengajuan penelitian kepada Dinas Perdagangan kota Pagar Alam.

3.2. Data Understanding

Pada fase pemahaman data ini, data yang telah didapatkan dari harga beras pada tahun 2019-2023 maka hasil data yang didapat 240 dengan atribut 6 yaitu tahun, bulan, produksi, harga beras, jenis beras dan kategori. Data yang diterima dalam bentuk *microsoft excel* dan data yang didapat semuanya terisi dan tidak ada data yang kosong jadi peneliti tidak perlu melakukan *missing data* dan atributnya semua berguna dalam penelitian ini. Atribut yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Dataset Tahun 2019

Bln	Prod (Ton)	Harga (Rp)	Jenis Beras	Kategori
1	2	11000	Premium	Mahal
1	1,3	11000	Premium	Mahal
1	3	10000	Medium	Mahal
1	3,5	10000	Medium	Mahal
2	2	11000	Premium	Mahal
2	1,4	11000	Premium	Mahal
2	3,3	10000	Medium	Mahal
2	3	10000	Medium	Mahal
3	2	10000	Premium	Mahal
3	2	10000	Premium	Mahal
3	5	9500	Medium	Murah
3	4,5	9500	Medium	Murah
4	2,2	10000	Premium	Mahal
4	2,3	10000	Premium	Mahal
4	5	9000	Medium	Murah
4	4	9000	Medium	Murah
5	2,5	10000	Premium	Mahal
5	2	10000	Premium	Mahal
5	5,1	9500	Medium	Murah
5	6	9500	Medium	Murah
6	5	9500	Premium	Murah
6	4,5	9500	Premium	Murah
6	6	9000	Medium	Murah
6	5	9000	Medium	Murah

3.3. Data Preparation

Tahapan prediksi harga beras data asli yang didapatkan peneliti 5 atribut diantaranya yaitu (1) tahun, (2) bulan, (3) produksi, (4) harga (5) jenis beras (6) kategori. Data yang diterima dalam bentuk *microsoft*

excel semua datanya terisi dan tidak ada data yang kosong jadi peneliti tidak perlu melakukan *missing value* dan atributnya semuanya berguna dalam peneliti ini.

1. Data Cleaning

Data cleaning bertujuan untuk proses identifikasi perbaikan di dalam data, yang pemrosesannya dilakukan dalam penelitian ini tidak terdapat data kosong tidak ada data yang duplikat dan juga tidak ada data yang salah input atau data yang hilang dari keenam atribut tersebut.

2. Data Integration

Data integration bertujuan untuk menggabungkan data dari berbagai sumber. Dalam proses yang dilakukan tidak terdapat ketidaksesuaian pada data yang berbeda dan juga tidak ada penggabungan data.

3. Data Trasformation

Data transformation bertujuan untuk mengubah struktur data, format data, atau nilai data sedemikian rupa sehingga menghasilkan dataset yang sesuai untuk proses mining atau sesuai dengan struktur, format dan nilai agar bisa dilakukan perhitungan K-NN.

4. Data Reduction

Data reduction bertujuan untuk proses pengurangan data untuk mengurangi volume data tanpa menghilangkan informasi penting.

3.4. Modelling

Modelling adalah deskripsi atau *knowledge* berkualitas yang dibangun oleh sistem atau proses dari kalkulasi dan prediksi yang berterima. Proses *modelling* dalam proses pengolahan data karena *modelling* sangat berperan pada tahap ini. Satu dataset yang telah dipreparasi dengan baik dapat diolah dala satu jenis algoritma dengan dataset yang digunakan.

1. Penerapan K-NN

Langkah pertama yang harus dilakukan adalah memanggil *dataset* dari *library python*. Setelah file data harga beras selama 5 tahun diimport kan kedalam *library*, selanjutnya melakukan pemanggilan data., dengan *coding* pada Gambar 2, yang berfungsi untuk memanggil data.

```
[2] df = pd.read_excel("dataset.xlsx")
df
```

Gambar 2. Coding Import Data

Kemudian akan tampil *dataset* harga beras selama 5 tahun, yang dapat dilihat pada Gambar 3.

	Tahun	Bulan	Produksi(Ton)	Harga(Rp)	Jenisberas	Kategori
0	2019	1	2.0	11000	Premium	mahal
1	2019	1	1.3	11000	Premium	mahal
2	2019	1	3.0	10000	Medium	mahal
3	2019	1	3.5	10000	Medium	mahal
4	2019	2	2.0	11000	Premium	mahal
...
235	2023	11	2.0	13500	Medium	sangat mahal
236	2023	12	1.7	15000	Premium	mahal
237	2023	12	1.0	15000	Premium	sangat mahal
238	2023	12	2.0	13500	Medium	sangat mahal
239	2023	12	3.0	13500	Medium	sangat mahal

Gambar 3. Tampilan Data

2. Memanggil label data

Selanjutnya melakukan pemanggilan data yang menjadi label yaitu kategori dengan kode program pada Gambar 4.

```
[6] y = df["Kategori"]
y.head(11)
```

Gambar 4. Coding Menampilkan Label Data

Setelah melakukan pemanggilan data yang dilabelkan akan ditampilkan isi dari label data seperti pada Gambar 5.

```
0 2
1 2
2 2
3 2
4 2
5 2
6 2
7 2
8 2
9 2
10 3
Name: Kategori, dtype: int64
```

Gambar 5. Tampilan Label Data

3. Melakukan transformasi data jenis beras

Melakukan transformasi data jenis beras dengan mengubah jenis beras premium itu adalah angka 1 sedangkan jenis beras premium adalah angka 2, Gambar 6 menunjukkan *coding* transformasi data.

```
[3] Jenisberas = {'Premium': 1, 'Medium': 2}
df.Jenisberas = [Jenisberas[item] for item in df.Jenisberas]
df.head()
```

Gambar 6. Coding Transformasi Data

Setelah itu jenis beras akan berubah menjadi angka yang dapat dilihat pada Gambar 7.

	Tahun	Bulan	Produksi(Ton)	Harga(Rp)	Jenisberas	Kategori
0	2019	1	2.0	11000	1	mahal
1	2019	1	1.3	11000	1	mahal
2	2019	1	3.0	10000	2	mahal
3	2019	1	3.5	10000	2	mahal
4	2019	2	2.0	11000	1	mahal

Gambar 7. Hasil Transformasi Data

4. Menghitung jumlah kemunculan kategori

Digunakan untuk menghitung berapa kali setiap nilai muncul dalam kolom "kategori" dari data yang dapat dilihat pada Gambar 8.

```
[7] df['Kategori'].value_counts()

Kategori
2    117
1     72
3     51
Name: count, dtype: int64
```

Gambar 8. Data Kategori

5. Korelasi data

Digunakan untuk menghitung korelasi antara kolom-kolom dalam data dan kemudian mengurutkan nilai korelasi terhadap kolom "Jenis beras" dalam urutan menurun yang dapat dilihat pada Gambar 9.

```
[8] correlation = df.corr()
correlation['Kategori'].sort_values(ascending=False)
```

Gambar 9. Coding Menampilkan Nilai Korelasi

Setelah itu akan Mengurutkan nilai korelasi dalam urutan menurun. Dengan cara ini, nilai korelasi yang paling tinggi (positif) terhadap "kategori" akan berada di bagian atas, sedangkan nilai korelasi yang paling rendah (negatif) akan berada di bagian bawah yang dapat dilihat pada Gambar 10.

```
Kategori      1.000000
Produksi(Ton)  0.674135
Jenisberas    0.345989
Bulan         -0.005946
Harga(Rp)    -0.870976
Name: Kategori, dtype: float64
```

Gambar 10. Nilai Korelasi Menurun

Berikut penjelasannya :

Kategori : menunjukkan bahwa "Kategori" memiliki korelasi positif sempurna dengan dirinya sendiri, yang memang diharapkan karena setiap variabel pasti berkorelasi sempurna dengan dirinya sendiri.

Produksi (Ton) : menunjukkan adanya korelasi positif yang kuat antara "Kategori" dan "Produksi(Ton)", artinya, ketika nilai "Kategori" meningkat, "Produksi (Ton)" cenderung ikut meningkat.

Jenisberas : menunjukkan adanya korelasi positif sedang antara "Kategori" dan "Jenis beras". artinya, ada kecenderungan "Jenis beras" untuk meningkat seiring dengan peningkatan nilai "Kategori".

Bulan : menunjukkan adanya korelasi negatif yang sangat lemah antara "Kategori" dan "Bulan". Variabel "Bulan" tidak banyak berubah seiring dengan perubahan "Kategori".

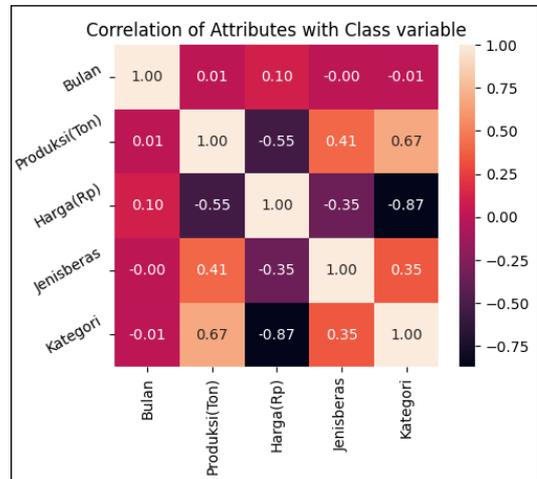
Harga (Rp) : menunjukkan adanya korelasi negatif yang kuat antara "Kategori" dan "Harga (Rp)", artinya, ketika nilai "Kategori" meningkat, "Harga (Rp)" cenderung menurun

6. Menampilkan visualisasi korelasi

Menampilkan visualisasi atau tampilan dari korelasi yang dapat dilihat pada Gambar 11.

```
plt.figure(figsize=(5,4))
plt.title('Correlation of Attributes with Class variable')
a = sns.heatmap(correlation, square=True, annot=True, fmt='.2f', linecolor='white')
a.set_xticklabels(a.get_xticklabels(), rotation=90)
a.set_yticklabels(a.get_yticklabels(), rotation=30)
plt.show()
```

Gambar 11. Coding Menampilkan Visualisasi Korelasi Setelah itu akan tampil seperti pada Gambar 11.



Gambar 12. Visualisasi Korelasi

Berikut Penjelasan dari Gambar 12.

a. Bulan :

Tidak memiliki korelasi signifikan dengan variabel lainnya. Korelasinya berkisar antara -0.01 hingga 0.10, yang menunjukkan hubungan yang sangat lemah atau hampir tidak ada.

b. Produksi (Ton) :

Dengan Harga (Rp) : -0.55, Korelasi negatif sedang, artinya ketika produksi meningkat, harga cenderung menurun.

Dengan Jenis beras : 0.41, Korelasi positif sedang, artinya ketika produksi meningkat, jenis beras tertentu juga cenderung meningkat.

Dengan Kategori : 0.67, Korelasi positif kuat, artinya ketika kategori meningkat, produksi juga cenderung meningkat.

c. Harga (Rp) :

Dengan Produksi (Ton) : -0.55, Korelasi negatif sedang, artinya ketika produksi meningkat, harga cenderung menurun.

Dengan Jenis beras : -0.35, Korelasi negatif sedang, artinya ketika harga meningkat, jenis beras tertentu cenderung menurun.

Dengan Kategori: -0.87, Korelasi negatif kuat, artinya ketika kategori meningkat, harga cenderung menurun.

d. Jenis beras :

Dengan Produksi (Ton) : 0.41, Korelasi positif sedang, artinya ketika harga meningkat, jenis beras tertentu cenderung menurun.

Dengan Harga (Rp) : -0.35, Korelasi negatif sedang, artinya ketika harga meningkat, jenis beras tertentu cenderung menurun.

Dengan Kategori : 0.35, Korelasi positif sedang, artinya ketika kategori meningkat, jenis beras tertentu juga cenderung meningkat.

d. Kategori :

Dengan Produksi(Ton) : 0.67, Korelasi positif kuat, artinya ketika kategori meningkat, produksi juga cenderung meningkat.

Dengan Harga(Rp) : -0.87, Korelasi negatif kuat, artinya ketika kategori meningkat, harga cenderung menurun.

Dengan Jenisberas: 0.35, Korelasi positif sedang, artinya ketika harga meningkat, jenis beras tertentu cenderung menurun.

7. Memanggil data training dan data testing

Data training merupakan dataset yang kita latih atau membuat menjalankan fungsi pada algoritma sesuai dengan tujuannya. Testing berfungsi untuk bagian dataset yang kita testing untuk melihat keakuratan atau performance pada sebuah algoritma X_train sebagai data X yang dilatih X_test sebagai data X yang di testing parameter selanjutnya adalah test_size sebesar 20%. Biasanya sebagai aturan umum 20% proposri testing dan training 80% selanjutnya parameter random state 0 untuk kedata normal, dengan menggunakan coding pada Gambar 13.

```
[ ] from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.1, random_state = 0)
X_train.shape, X_test.shape
((216, 4), (24, 4))
```

Gambar 13. Coding Memanggil Data Training dan Data Testing

Hasil data training 216 data dan hasil data testing 24 data yang bisa di lihat pada Gambar 14 dan Gambar 15.

	Bulan	Produksi(Ton)	Harga(Rp)	Jenisberas
5	2	1.4	11000	1
22	6	6.0	9000	2
199	2	1.5	12000	2
97	1	2.0	10500	1
12	4	2.2	10000	1
...
67	5	5.5	9000	2
192	1	2.0	13000	1
117	6	2.0	11000	1
47	12	2.2	10500	2
172	8	2.0	12500	1

216 rows x 4 columns

Gambar 14. Data Training

	Bulan	Produksi(Ton)	Harga(Rp)	Jenisberas
109	4	1.5	10500	1
71	6	5.4	9000	2
37	10	2.0	10000	1
74	7	6.0	8500	2
108	4	3.3	10500	1
227	9	2.4	13000	2
156	4	1.5	12000	1
220	8	2.5	14000	1
152	3	1.5	12500	1
194	1	2.7	12000	2
76	8	2.5	10500	1
202	3	1.0	12000	2
83	9	2.0	9500	2

Gambar 15. Data Testing

8. Prediksi K-NN

Untuk melihat seberapa akurat data pada algoritma K-NN, maka dilakukan prediksi dengan menggunakan coding pada Gambar 16.

```
[ ] y_pred = knn.predict(X_test)
y_pred
array([2, 3, 2, 3, 2, 1, 1, 1, 1, 1, 2, 1, 3, 1, 1, 3, 1, 2, 1, 2, 2, 2, 2, 3])
```

Gambar 16. Coding Prediksi K-NN

9. Menampilkan hasil data training dan testing

Digunakan untuk menghitung dan menampilkan skor akurasi dari model K-Nearest Neighbors (K-NN), dengan menampilkan Data training dan Data Testing, seperti coding pada Gambar 17.

```
[ ] print('Training set score: {:.4f}'.format(knn.score(X_train, y_train)))
print('Test set score: {:.4f}'.format(knn.score(X_test, y_test)))
Training set score: 0.9954
Test set score: 0.9583
```

Gambar 17. Coding Menampilkan Data Training dan Data Testing

Nilai Training: 0.9954 : Model KNN memiliki akurasi 99.54% pada data, menunjukkan bahwa model tersebut sangat baik dalam mempelajari pola pada data training. Artinya, 99.54% dari data training berhasil diprediksi dengan benar oleh model.

Nilai Testing: 0.9583 : Model K-NN memiliki akurasi 95.83% pada data uji, ini adalah indikasi bagaimana model bekerja pada data baru yang tidak dilihat selama pelatihan.

Akurasi testing sebesar 95.83% sedikit lebih rendah dibandingkan akurasi training (99.54%), selisih yang kecil ini menunjukkan bahwa model tidak mengalami overfitting yang signifikan. Model mampu bekerja dengan baik pada data yang belum pernah dilihat sebelumnya, sehingga memiliki generalisasi yang baik. Nilai testing accuracy ini mengindikasikan bahwa model Anda memiliki performa yang baik dan dapat diandalkan untuk tugas prediksi dengan tingkat kesalahan sekitar 4.17%.

10. Hasil Akurasi

Setelah perintah kode program yang semuanya telah berhasil baru menampilkan akurasi dari algoritma K-NN dengan mendapatkan nilai akurasi yang dapat dilihat pada Gambar 18.

```
[ ] from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
1	0.90	1.00	0.95	9
2	1.00	0.90	0.95	10
3	1.00	1.00	1.00	5
accuracy			0.96	24
macro avg	0.97	0.97	0.96	24
weighted avg	0.96	0.96	0.96	24

Gambar 18. Hasil Akurasi dari K-NN

Dari tampilan google colab tersebut, disajikan dalam Tabel 2. sebagai berikut :

Tabel 2. Hasil Pengujian

Kelas	Presisi	Recall	F1-score	Support
1 (sangat mahal)	0,90	1,00	0,95	9
2 (mahal)	1,00	0,90	0,95	10
3 (murah)	1,00	1,00	1,00	5

Model ini menunjukkan performa yang sangat baik dengan akurasi 96%, serta presisi, recall, dan F1-score yang tinggi untuk setiap kelas, model ini memiliki kemampuan generalisasi yang sangat baik. Semua metrik menunjukkan performa yang seimbang, sehingga model dapat digunakan secara andal untuk prediksi kelas, bahkan jika dataset memiliki kelas yang berbeda atau tidak seimbang. Dalam konteks prediksi harga beras berdasarkan kualitas, model ini cukup kuat untuk diterapkan dalam pengambilan keputusan berbasis data.

13. Pengujian

Pada pengujian menggunakan Mean Squared Error (MSE) dan R-Squared (R²), dengan menggunakan coding pada Gambar 19.

```
[ ] print(f'Mean Squared Error (MSE): {mse:.2f}')
print(f'R-Squared (R2): {r_squared:.2f}')
```

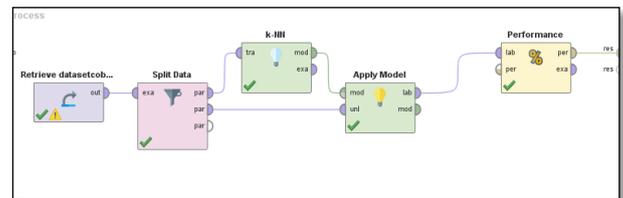
Mean Squared Error (MSE): 0.04
R-Squared (R2): 0.93

Gambar 19. Coding Pengujian

Mean Squared Error (MSE) adalah rata-rata kuadrat dari perbedaan antara nilai yang diprediksi dan nilai aktual. Nilai 0.04 menunjukkan bahwa model memiliki kesalahan rata-rata yang kecil. R-Squared (R²) adalah ukuran seberapa baik data cocok dengan model regresi. Nilai 0.93 menunjukkan bahwa 93% variasi dalam data dijelaskan oleh model, yang menunjukkan kecocokan yang sangat baik.

3.5. Evaluation

Dari hasil pengujian yang dilakukan pada tahapan modelling akan diperoleh model yang selanjutnya model tersebut akan dievaluasi untuk mengukur kinerja dari algoritma KNN. Untuk membuktikan hasil dari klasifikasi di atas, maka penulis menguji model di atas dengan menggunakan Rappid minner agar mendapatkan model yang akurat, seperti pada Gambar 20.



Gambar 20. Tampilan Rapid Miner

Pada gambar tersebut di dalam proses Cross Validation lalu memasukkan Operator K-NN untuk menghasilkan model berdasarkan algoritma K-Nearest Neighbor yang dapat digunakan untuk klasifikasi maupun regresi. Algoritma K-NN didasarkan pada jarak tetangga terdekat sebagai nilai prediksi dari instance yang baru. lalu dihubungkan dengan Operator split data digunakan untuk menerapkan model yang telah memperlihatkan sebelumnya menggunakan data training pada unlabeled (Data Testing) yang belum memiliki label, yang dapat dilihat pada Gambar 21.

accuracy: 95.85%

	true mahal	true murah	true sangat mahal	class precision
pred. mahal	90	1	3	95.74%
pred. murah	0	40	0	100.00%
pred. sangat mahal	4	0	55	93.22%
class recall	95.74%	97.56%	94.83%	

Gambar 21. Prediksi dari Instance

Terlihat dari Gambar 21. Bahwa pengujian dengan menggunakan Rappid miner maka di dapat hasil akurasi sebesar 95.85%.

3.6. Deployment

Pada tahap ini akan dibuat laporan penelitian dan mempresentasikan dari model di atas yang telah melalui proses data mining. Hal-hal yang akan dipresentasikan mulai dari pengetahuan yang diperoleh selama proses tersebut berjalan sehingga dapat dipahami oleh pengguna. Tahapan pembuatan laporan dapat dilakukan setelah melalui proses evaluation dari penelitian terhadap algoritma K-Nearest Neighbor untuk memprediksi harga beras.

4. Kesimpulan

Penelitian ini berhasil menerapkan algoritma K-Nearest Neighbor (K-NN) untuk memprediksi harga beras berdasarkan kualitasnya. Dengan menggunakan metode CRISP-DM (Cross Industry Standard Process for Data Mining), penelitian ini melibatkan tahapan

pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran. Hasil penelitian menunjukkan bahwa algoritma K-NN dapat memberikan prediksi harga beras dengan akurasi yang cukup baik, ditunjukkan oleh nilai *Mean Squared Error* (MSE) sebesar 0,04 dan nilai *R-Squared* (R^2) sebesar 0,93. Hal ini menunjukkan bahwa model prediksi yang dibangun mampu menjelaskan variabilitas harga beras berdasarkan kualitasnya dengan baik. Perhitungan menggunakan *Rapid Miner* dengan algoritma *K-Nearest Neighbor* didapatkan yaitu di dapat hasil akurasi sebesar 95,85% sedangkan menggunakan *google colab* didapatkan 96%. Berdasarkan hasil kualifikasi diatas, maka penerapan algoritma K-NN dapat dinyatakan baik dalam mengklasifikasi suatu data, dan juga dapat diterapkan di Dinas Perdagangan Untuk membantu dalam memprediksi harga beras.

Daftar Pustaka

- Arie, J. S. (2019). Implementasi Algoritma KNN Dalam Memprediksi Curah Hujan dan Temperatur Untuk Tanaman Padi. *SISITI: Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, 8(1).
- Asminah, A., Setiadi, D., & Susanti, T. (2023). Klasifikasi Jenis Buah Durian dengan Metode K-Nearest Neighbor. *BINA INSANI ICT JOURNAL*, 10(2), 176–187.
- Fathurohman, A. (2021). Machine Learning Untuk Pendidikan: Mengapa dan Bagaimana. *Jurnal Informatika Dan Teknologi Komputer (JITEK)*, 1(3), 57–62.
- Firmansyah, F., & Nurdiawan, O. (2023). Penerapan Data Mining Menggunakan Algoritma Frequent Pattern-Growth Untuk Menentukan Pola Pembelian Produk Chemicals. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 547–551.
- Hasan, N. F., & Wati, V. (2021). Deteksi Cyberbullying pada Facebook Menggunakan Algoritma K-Nearest Neighbor. *Journal of Smart System*, 1(1), 35–44.
- Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, 5(2), 103–108.
- Ikhwan, A., & Aslami, N. (2020). Implementasi Data Mining untuk Manajemen Bantuan Sosial Menggunakan Algoritma K-Means. (*JurTI*) *Jurnal Teknologi Informasi*, 4(2), 208–217.
- Irnanda, K. F., Windarto, A. P., & Damanik, I. S. (2022). Optimasi Particle Swarm Optimization Pada Peningkatan Prediksi dengan Metode Backpropagation Menggunakan Software RapidMiner. *JURIKOM (Jurnal Riset Komputer)*, 9(1), 122–130.
- Mirantika, N. (2021). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat. *Nuansa Informatika*, 15(2), 92–98.
- Muallif, I. S., Budiman, H., & Ransi, N. (2023). Penerapan Data Mining untuk Prediksi Pergerakan Harga Saham Menggunakan Algoritma K-Nearest Neighbor. *Prosiding Seminar Nasional Pemanfaatan Sains Dan Teknologi Informasi*, 1(1), 297–306.
- Mukti, H., & Saputri, A. R. (2023). Smart Farming Asistance (S-Farm): Strategi Pengembangan Fintech Berbasis Bumdes Bagi Komoditas Pertanian Kabupaten Grobogan dalam Menyongsong SDGS 2030: Smart Farming Asistance (S-Farm): Strategi Pengembangan Fintech Berbasis Bumdes Bagi Komoditas Pertani. *Journal Science Innovation and Technology (SINTECH)*, 3(2), 32–42.
- Nahjan, M. R., Heryana, N., & Voutama, A. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101–104.
- Pambudi, A. (2023). Penerapan Crisp-Dm Menggunakan Mlr K-Fold Pada Data Saham Pt. Telkom Indonesia (Persero) Tbk (Tlkm)(Studi Kasus: Bursa Efek Indonesia Tahun 2015-2022). *Jurnal Data Mining Dan Sistem Informasi*, 4(1), 1–14.
- Rahmadini, R., LorencisLubis, E. E., Priansyah, A., Yolanda, R. W. N., & Meutia, T. (2023). Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Mahasiswa Akuntansi Samudra*, 4(4), 223–235.
- Sari, D. A. (2023). Pemanfaatan Information and Communication Technology (ICT) pada Usahatani Salak Pondoh. *INTEGRAL: Jurnal Inovasi, Teknologi Terapan, Dan Litbang*, 2(2).
- Satria, A., Badri, R. M., & Safitri, I. (2023). Prediksi Hasil Panen Tanaman Pangan Sumatera dengan Metode Machine Learning. *Digital Transformation Technology*, 3(2), 389–398.
- Seruni, D. S., Furqon, M. T., & Wihandika, R. C. (2020). Sistem Prediksi Pertumbuhan Jumlah Penduduk Kota Malang menggunakan Metode K-Nearest Neighbor Regression. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(4), 1075–1082.
- Setiadi, D., Sasmita, S., & Yolanda, M. (2024). Penerapan Algoritma Regresi Linier Berganda Untuk Memprediksi Hasil panen Padi Di Kota Pagar Alam. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer Dan Manajemen)*, 5(2), 337–438.
- Setiadi, D., & Syahri, R. (2022). Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Pengguna

- Narkoba di Kota Pagar Alam. *JUTIM (Jurnal Teknik Informatika Musirawas)*, 7(1), 1–10.
- Suhanda, Y., Kurniati, I., & Norma, S. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika Dan Komputer*, 6(2), 12–20.
- Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Penjualan Bisnis Ritel. *Edumatic: Jurnal Pendidikan Informatika*, 4(1), 37–46.
- Virdaus, D., & Prasetyaningrum, P. T. (2020). Penerapan Data Mining Untuk Memprediksi Harga Bawang Merah Di Yogyakarta Menggunakan Metode K-Nearest Neighbor. *Journal Of Information System And Artificial Intelligence*, 1(1), 1–8.