



ISSN : 2339 - 1871

# JURNAL ILMIAH BETRIK

Besemah Teknologi Informasi dan Komputer

Editor Office : LPPM Sekolah Tinggi Teknologi Pagar Alam, Jln. Masik Siagim No. 75  
Simpang Mbacang, Pagar Alam, SUM-SEL, Indonesia  
Phone : +62 852-7901-1390.  
Email : [betrik@sttpagaralam.ac.id](mailto:betrik@sttpagaralam.ac.id) | [admin.jurnal@sttpagaralam.ac.id](mailto:admin.jurnal@sttpagaralam.ac.id)  
Website : <https://ejournal.sttpagaralam.ac.id/index.php/betrik/index>

## IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI HARGA CABAI MERAH KOTA PAGAR ALAM MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Sasmita, Siti muntari, Wike Agustina

Program Studi Teknik Informatika  
Institut Teknologi Pagar Alam

Jalan masuk siagim No 75 Simpang mbacang Kota Pagar Alam  
Sur-el : [sasmitha661@gmail.com](mailto:sasmitha661@gmail.com)<sup>1</sup> [muntariaza@gmail.com](mailto:muntariaza@gmail.com)<sup>2</sup>

**Abstrak:** Penelitian ini bertujuan untuk implementasi data mining dalam Memprediksi Harga Tanaman Cabai Merah Kota Pagar Alam menggunakan algoritma *K-Nearest Neighbor*. Permasalahan saat ini dimana para petani serta dinas ketahanan pangan dan perikanan masih memprediksi harga cabai dengan melakukan metode survey langsung ke pasar dan data tersebut tidak dilakukan pengolahan lanjutan pada data apalagi jika harga cabai tidak menentu dan berubah-ubah maka tidak adanya hasil simpulan kategori untuk pengklasifikasian dari historis harga, tentu saja hal ini menjadi bahan pertimbangan bagaimana cara mengantisipasi masalah ini. Metode pengumpulan data untuk memperoleh informasi yang dibutuhkan yaitu dengan cara Observasi, Wawancara, Studi Pustaka, dan Dokumentasi. Penelitian ini menggunakan Algoritma *K-Nearest Neighbor* dengan metode *CRISP-DM*. Dimana tahapan meliputi Pemahaman bisnis (*Business Understanding*), Pemahaman Data (*Data Understanding*), Persiapan Data (*Data Preparation*), Pemodelan (*Modelling*), Pengujian (*Evaluation*), dan Penyebaran (*Deployment*). Data di olah menggunakan aplikasi *Rapid Miner*, metode pengujian menggunakan *Root Mean Square Error (RMSE)*. Didapatkan bahwa hasil dari penelitian ini yaitu untuk perhitungan *Root Means Square Error* dan didapatkan nilai rata-rata 0.09% dan untuk perhitungan menggunakan *Rapidminer* dengan algoritma *K-Nearest Neighbor* didapatkan yaitu *Accuracy* sebesar 94.03% dengan *Recall* 83.56% dan *Precession* 88.73%.

**Kata kunci :** Cabai Merah, Prediksi, *K-Nearest Neighbor*, *CRISP-DM*, *RMSE*.

**Abstract:** This research aims to implement data mining in predicting the price of red chili plants in Pagar Alam City using the *K-Nearest Neighbor* algorithm. The current problem is that farmers and the food security and fisheries service still predict chili prices by conducting direct survey methods to the market and the data is not further processed on the data, especially if the price of chili is uncertain and changes, there are no conclusions for the category for classifying historical prices, of course this is a consideration on how to anticipate this problem. The method of collecting data to obtain the information needed is by Observation, Interview, Literature Study, and Documentation. This study uses the *K-Nearest Neighbor* Algorithm with the *CRISP-DM* method. Where the stages include *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, and *Deployment*. Data is processed using the *Rapid Miner* application, the testing method uses *Root Mean Square Error (RMSE)*. It was found that the results of this study were for the calculation of *Root Means Square Error* and obtained an average value of 0.09% and for the calculation using *RapidMiller* with the *K-Nearest Neighbor* algorithm, the *Accuracy* was 94.03% with *Recall* 83.56% and *Precession* 88.73%

**Keywords:** Red Chili, Prediction, *K-Nearest Neighbor*, *CRISP-DM*, *RMSE*.

### 1. PENDAHULUAN (Font 12)

Perkembangan teknologi dan informasi berawal dari kemajuan dibidang komputerisasi, penggunaan komputer pada masa awal untuk sekedar menulis,

membuat grafik dan gambar serta alat menyimpan data yang luar biasa telah berubah menjadi alat komunikasi dengan jaringan lunak dan bisa mencakup seluruh dunia (Setiawan, 2018). Sehingga dalam

teknologi-teknologi tersebut perlu adanya penerapan teknologi analisis data agar data-data tersebut menjadi sebuah informasi yang berguna bagi semua kalangan.

*Big Data* adalah konsep yang mengacu pada *volume* data yang sangat besar, kecepatan penghasilan data yang tinggi, keragaman jenis data, dan kompleksitas yang tinggi. Data dalam kategori *Big Data* tidak dapat dengan mudah diolah atau dikelola dengan menggunakan alat dan metode tradisional. Konsep ini juga mencakup teknik-teknik analisis data yang inovatif untuk mengungkap wawasan yang mendalam, tren, dan pola yang tersembunyi dalam data tersebut (Wijoyo et al., 2023). Penambangan data (*data mining*) juga dikenal sebagai menemukan pengetahuan dari data (*knowledge discovery from data*). *Data mining* dapat memiliki dengan tidak sah dinamai “*knowledge mining from data*” atau “*knowledge mining*” yaitu pada penambangan dari data dengan jumlah sangat besar. *Data mining* dapat dilihat sebagai sebuah hasil dari evolusi natural dari teknologi informasi. *Data mining* juga dapat diartikan sebagai pengestrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan (Zai, 2022). *Prediksi (prediction)* adalah memperkirakan nilai-nilai data bertipe apa saja dan kapan saja (masa lalu, sekarang, dan masa depan). Terdapat satu istilah yang mirip dengan *prediksi*, yaitu *peramalan (forecasting)* adalah memperkirakan nilai-nilai data *time series* dimasa yang akan datang, *prediksi* yaitu suatu kegiatan untuk meramalkan harga penjualan dimasa yang akan datang berarti menentukan perkiraan besarnya *volume* harga jual (Damuri et al., 2021).

Algoritma *K-Nearest Neighbor (K-NN)* adalah sebuah metode untuk melakukan klasifikasi terhadap objek kedalam kelas yang sudah ditentukan berdasarkan kedekatan objek tersebut dengan data pembelajaran atau *data mining*. Algoritma ini bekerja dengan teknik yang sederhana

yaitu mengklasifikasikan suatu kelompok data yang belum diketahui kelasnya ke kelompok data tertentu berdasarkan kedekatan (jarak) data tersebut ke data pembelajaran (Virdaus & Prasetyaningrum, n.d.).

Dikota Pagar Alam Cabai merah termasuk bahan pangan yang dibudidayakan sampai sekarang dimana *varietas* cabai yang menjadi hal yang wajib diadakan pada setiap menu hidangan, apalagi cabai juga menambah keberagaman dan keceriaan dalam dunia kuliner global. Asal usul cabai secara keseluruhan dapat ditelusuri kembali ke Amerika Selatan, di mana buah ini telah menjadi bagian integral dari makanan sejak diperkenalkan oleh penjelajah Eropa pada abad ke-15. Seiring dengan sejarah cabai yang panjang, cabai merah keriting telah menjadi bahan pokok dalam dapur banyak budaya. Di berbagai masakan Asia, buah ini sering digunakan untuk memberikan sentuhan pedas pada tumisan, sup, atau hidangan lainnya (Labolo et al., 2022).

Berdasarkan studi pendahuluan wawancara serta observasi pada dinas ketahanan pangan dan petani kota pagar alam cara memprediksi harga cabai dilakukan dengan metode survey langsung ke pasar dan data tersebut tidak dilakukan pengolahan lanjutan apalagi jika harga cabai tidak menentu atau berubah-ubah, dimana jika saat pasokan cabai meningkat dan permintaan sedikit maka harga akan mengalami penurunan begitupun jika pasokan cabai sedikit tetapi permintaan banyak maka harga akan melonjak, hal ini karena ada beberapa faktor yang mempengaruhi harga cabai merah tersebut diantaranya tingkat produksi didaerah, luas panen yang tidak menentu. Karena pada dasarnya penelitian ini diharapkan dapat membantu para petani untuk menentukan waktu tanam cabai, dalam hal ini jika tidak dilakukan penelitian yang mendalam maka dapat berdampak kerugian bagi petani maupun masyarakat akan masalah harga cabai 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 cabai tersebut. Karena apabila analisis data ini dikaji dan digali dengan mengimplementasikan metode yang akurat maka dapat mengetahui pola serta pengetahuan untuk kebijakan.

Sandy Tri Wijaya (2023) dengan judul Penerapan metode *k-nearest neighbor* untuk prediksi harga jagung dengan pengujian RMSE dan Penelitian ini melakukan penerapan metode *K-Nearest Neighbor* untuk memprediksi harga jagung di Kabupaten Tulungagung dapat membantu para peternak maupun petani dalam mengambil keputusan untuk melakukan transaksi dengan bahan dasar jagung di bulan depan. Hasil prediksi dari penerapan tersebut yaitu 6400, 6416, 6333, 6333 dan 0 data diambil setiap minggunya dengan *error* sebesar 185,1497 dan persentase sebesar 95% yang menjadi tingkat keakuratan dari metode *K-Nearest Neighbor* untuk memprediksi harga jagung. Korelasi dengan penelitian yaitu peneliti menggunakan algoritma *K-Nearest Neighbor* dan sama-sama melakukan prediksi harga serta pada pengujian RMSE. Dapit Virdaus (2021) dengan judul Penerapan data mining untuk memprediksi harga bawang merah di Yogyakarta menggunakan metode *k-nearest neighbor*, Penelitian ini mengimplementasikan data mining dengan menggunakan metode algoritma *K-Nearest Neighbor* akan diterapkan untuk memprediksi harga bawang merah menggunakan data yang telah tersedia sebelumnya. Pada penelitian ini dilakukan empat skenario atau kondisi yang dijadikan sasaran data bawang merah dalam pengolahan datanya agar diperoleh hasil yang lebih *variatif* dan lebih presisi dengan tingkat akurasi prediksi tertinggi. Setelah melakukan berbagai macam percobaan pada data yang diadakan dalam penelitian ini, atau lebih tepatnya penerapan empat kondisi pada data, diperoleh hasil akurasi tertinggi dengan nilai 91,67%. Korelasi dengan penelitian yaitu peneliti menggunakan algoritma *K-Nearest Neighbor* dan sama-sama melakukan

prediksi harga. Yuwantoro Mukhlisin (2020) dengan judul Prediksi Harga Beras Premium dengan Metode Algoritma *K-Nearest Neighbor*, Penelitian ini melakukan penerapan salah satu metode data mining dalam proses prediksi harga beras dengan membandingkan harga beras pada tahun 2014 - 2019 dari BPS Kota Bandung. Dataset yang digunakan berasal dari BPS Kota Bandung dari tahun 2014 hingga 2019 dan BMKG Kota Bandung dengan tahun yang sama. Adapun metode yang digunakan adalah algoritma regresi *K-Nearest Neighbor* (KNN) serta untuk pengujiannya menggunakan RMSE. Hasil dari penelitian ini, metode *K-Nearest Neighbor* dengan model regresi dapat melakukan prediksi terhadap harga beras pada tahun 2014 - 2019 dengan nilai RMSE 0,125 dan parameter  $K = 2$  yang sudah dinormalisasi mengklasifikasikan data ke dalam kelas tertentu berdasarkan kedekatannya dengan tetangga terdekat dalam hal atribut atau variabel. Korelasi dengan penelitian yaitu peneliti menggunakan algoritma *K-Nearest Neighbor* dan sama-sama melakukan prediksi harga serta untuk pengujian dengan RMSE. Herdi Budiman (2023) dengan judul Penerapan Data Mining untuk Prediksi Pergerakan Harga Saham Menggunakan Algoritma *K-Nearest Neighbor*, Penelitian ini bertujuan untuk mengetahui implementasi metode *K-Nearest Neighbor* dalam memprediksi pergerakan harga saham. Penelitian ini menggunakan data historis harga saham sebagai data latih untuk melatih model *K-Nearest Neighbor*. Proses pelatihan melibatkan identifikasi dan perangkan saham-saham yang memiliki karakteristik serupa berdasarkan data historis. Penelitian ini menambahkan beberapa variabel-variabel yang memiliki pengaruh terhadap pergerakan harga saham yang membedakan dari penelitian terdahulu. Pada penelitian ini, ditemukan model prediksi dengan tingkat *accuracy* tertinggi sebesar 62,54%, *precision* sebesar 64,14%, dan *recall* sebesar 92,08%. Hasil ini diperoleh dengan menggunakan nilai  $k=53$  pada random data ke-6 dalam data set ke-2. Korelasi dengan penelitian yaitu peneliti menggunakan algoritma *K-Nearest Neighbor*. Muis Nanja

(2015) dengan judul Prediksi Harga Komoditi Lada, Hasil penelitian menunjukkan bahwa algoritma K-Nearest Neighbor berbasis forward selection memberikan kinerja yang terbaik dibandingkan dengan KNN berbasis backward elimination dan SVM berbasis seleksi atribut, khususnya forward selection untuk memprediksi komoditi lada. Model yang diusulkan dievaluasi dengan data time series lada hitam dan lada putih. Korelasi dengan penelitian ialah peneliti menggunakan algoritma K-Nearest Neighbor.

Tujuan dari penelitian ini adalah implementasi data mining Untuk Memprediksi Harga Tanaman Cabai Merah Kota Pagar Alam menggunakan algoritma K-NN.

## 2. METODE PENELITIAN (Font 12)

### 2.1 Metode Pengumpulan Data

Teknik pengumpulan data yang dilakukan penelitian ini antara lain:

#### a. Observasi

Observasi adalah mencari data dengan cara survey terhadap metode penyampaian informasi, yang sebelumnya dilakukan. Dalam metode ini peneliti mengamati secara langsung terhadap objek penelitian.

#### b. Wawancara

Wawancara adalah mencari dan mengumpulkan data dengan cara mengajukan pertanyaan terhadap narasumber. Dalam metode ini peneliti mengajukan beberapa pertanyaan kepada pihak dinas ketahanan pangan untuk mendapatkan informasi yang dibutuhkan.

#### c. Dokumentasi

Penelitian ini dilakukan dengan cara menggunakan data yang diperoleh dari dinas pertanian baik itu arsip ataupun lainnya.

#### d. Studi Pustaka

Dalam penulisan, peneliti menggunakan berbagai referensi seperti buku pedoman, buku dan jurnal yang berkaitan dengan penelitian ini seperti buku data mining metode klasifikasi sebagai landasan teori untuk

menyelesaikan masalah yang dihadapi.

### 2.2 Pengujian Root Means Square Error

*Root Mean Square Error* atau yang sering disingkat RMSE adalah metrik evaluasi yang sering digunakan untuk mengukur perbedaan atau perbandingan antara nilai prediksi yang dihasilkan oleh suatu model dengan nilai aktual yang diamati. RMSE memiliki fungsi untuk mengumpulkan besarnya kesalahan dalam prediksi pada nilai data menjadi kesatuan keakuratan prediksi (Wijaya et al., 2023). RMSE memiliki akurasi yang tinggi dalam melakukan pengukuran dengan probabilitas hingga 50% pada algoritma yang digunakan untuk memprediksi data, berikut persamaan rumus :

$$RMSE = \sqrt{\sum \frac{(Y' - Y)^2}{n}}$$

Keterangan:

$Y^1$  = Nilai Prediksi

$Y^2$  = Nilai Sebenarnya

$n$  = Jumlah Data

### 2.3 Metode Pengembangan

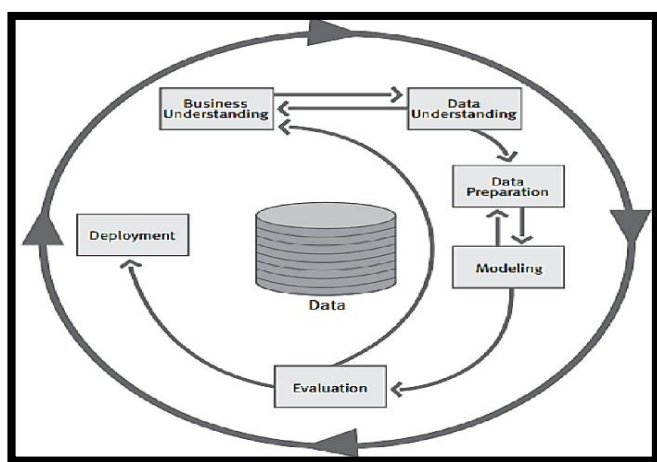
#### a. Metode CRISP-DM

*CRISP-DM* merupakan sebuah standarisasi data mining yang digunakan sebagai strategi pemecahan masalah yang secara umum merupakan dari segi bisnis atau unit penelitian (Algoritma, 2022). Proses metodologi terdiri dari 6 tahapan yang dapat dijelaskan sebagai berikut.

1. *Business Understanding* (Pemahaman Bisnis) Berbagai hal dilakukan dalam fase ini, misal Memahami kebutuhan dan tujuan prospek bisnis, kemudian menginterpretasikan pengetahuan tersebut dalam bentuk mengidentifikasi tantangan data mining dan mengembangkan rencana dan strategi untuk mencapai tujuan data mining.
2. *Data Understanding* (Pemahaman Data) Tahap ini diawali dengan pengumpulan data, deskripsi data, dan penilaian kualitas data.
3. *Data Preparation* (Persiapan Data) Pada tahap ini, yaitu pembuatan kumpulan data mentah terakhir. Beberapa tugas dilakukan, termasuk pembersihan data, pemilihan data, pencatatan dan atribut, dan transformasi data, yang digunakan sebagai masukan dalam fase pemodelan.

4. *Modelling* (Pemodelan) Pada fase ini, pembelajaran mesin terlibat langsung dalam mendefinisikan teknik penambangan data, alat penambangan data, dan algoritma data mining.
5. *Evaluation* (Pengujian) Langkah ini dilakukan dengan mempertimbangkan tingkat performa dari pola yang dihasilkan oleh algoritma.
6. *Deployment* (Penyebaran) Tahap ini dilakukan dengan membuat laporan dan artikel jurnal menggunakan model yang dihasilkan.

Dari pengertian diatas peneliti menyimpulkan metode *Crisps-Dm* merupakan salah satu metode pengembangan data yang sering digunakan dalam data mining.



Gambar 1 : *CRISP-DM*

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil

Penelitian ini menghasilkan model klasifikasi perhitungan prediksi harga cabai merah kota Pagar Alam dengan data pada tahun 2023 yaitu 224 data, yang dilakukan berdasarkan metode *CRISP-DM* yang memiliki beberapa fase yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. *business understanding* merumuskan masalah prediksi harga cabai sebagai masalah klasifikasi untuk memprediksi kategori dari harga tersebut, *data understanding* menghasilkan data dengan cara pengumpulan data serta deskripsi data seperti dataset dan atribut data, *data preparation* melakukan proses data cleaning atau pembersihan data, *modeling* melakukan pembuktian dari pengujian, *evaluation* menghasilkan akurasi dari model K-NN,

*Jurnal Ilmiah Betrik, Volume.15, No.03, Desember 2024*

*deployment* pengimplementasian model.

#### 3.2 Pembahasan

##### 3.2.1 Business Understanding

Tujuan utama dari penelitian ini adalah prediksi untuk mengklasifikasi harga cabai merah ke dalam bentuk kategori atau class seperti murah, mahal, sangat mahal dengan menilai situasi sumber data yang tersedia seperti data historis harga cabai serta data pasokan untuk dinas ketahanan pangan, untuk klasifikasi data harga cabai dilakukan pada periode perbulan tahun 2023, dengan ini kita dapat mengembangkan rencana untuk mengumpulkan dan memproses data, membangun model klasifikasi menggunakan algoritma *K-Nearest Neighbor* serta menguji dan memvalidasi model.

##### 3.2.2 Data Understanding

Pada tahap pemahaman data atau *data understanding* dimana dalam proses ini melakukan pengumpulan data atau observasi langsung dengan cara mengajukan pertanyaan terhadap narasumber atau petani cabai merah dikota pagar alam kebebeberapa desa untuk data harga cabai merah pada tahun 2023, dan ada juga beberapa hal yang dilakukan seperti studi pustaka dilakukan dengan cara mereview jurnal atau penelitian terdahulu yang memiliki objek atau metode yang sama. Data yang diperoleh dan digunakan untuk penelitian ini ada berjumlah 224 record data dengan 5 atribut yaitu, Nama, Bulan, Luas Panen (Ha), Hasil Panen (Kg) dan Harga (Rp). Kemudian data diolah dalam bentuk suatu dataset pada *Microsoft excel*

##### 3.2.3 Data Preparation

Pada tahap ini dilakukan penyempurnaan data, seperti menghapus data yang kosong atau *noisy*, tujuan dari proses ini adalah agar data dapat diolah pada tahap selanjutnya yaitu *modelling*, agar diperoleh hasil yang maksimal dan memudahkan peneliti dalam pengolahan data dengan *software rapid miner*.

##### a. Data Selection

Pemilihan atau seleksi data merupakan suatu proses yang harus dilakukan sebelum melakukan pengolahan data pada tahap selanjutnya, pada data selection dilakukan pemilihan terhadap atribut yang akan digunakan dan dilakukan proses konversi pada beberapa

atribut dalam suatu dataset, yang bertujuan agar proses data mining menjadi lebih mudah. Dari 5 atribut pada dataset yaitu Nama, Bulan, Luas Lahan (Ha), Hasil Panen (Kg), dan Harga (Rp). Atribut yang akan digunakan dalam proses data mining ini yaitu 4 atribut diantaranya Luas Lahan (Ha), Hasil Panen (Kg), dan Harga (Rp).

Peneliti juga melakukan konversi terhadap atribut Bulan yang dikonversi kedalam bentuk integer, dan atribut Harga (Rp) yang dikonversi kedalam bentuk *polynomial* dan dijadikan 3 kategori dengan rentang Murah, Mahal, dan Sangat Mahal.

### b. Data Cleaning

Tahap *processing* atau *cleaning* data merupakan proses pembersihan atau penghilangan data yang tidak lengkap (*noisy*), data yang akan di *cleaning* atau dibersihkan yaitu data yang tidak dapat diolah pada tahap selanjutnya, misalnya dalam suatu atribut terdapat data bernilai 0, maka data tersebut harus dihilangkan. Dalam dataset yang akan diolah peneliti memastikan tidak ada data yang *missing value*, sehingga data dapat dilanjutkan pada tahap selanjutnya. Untuk data yang diproses pada tahap *cleaning*, data sudah dalam kondisi yang baik tanpa adanya nilai yang hilang atau tidak lengkap maka oleh karena itu disimpulkan bahwa tidak ada tindakan pembersihan data yang diperlukan, dan data langsung siap digunakan dalam tahap analisis dan pemodelan selanjutnya

### 3.2.4 Modelling

Pada tahap ini dilakukan permodelan dengan algoritma yang ditentukan yaitu *K-Nearest Neighbor* sebagai perhitungannya proses modelling ini didukung oleh *tools* pengolahan data yaitu *software rapid miner*. Dataset yang berjumlah 224 record kemudian dilakukan pembagian data training dan data *testing*, dengan perbandingan 90% data *training* dan 10% data *testing*. Proses *modelling* pada *rapid miner* dapat dilihat pada gambar dibawah ini: Pada tahap ini ialah tahapan proses *modelling* pada *tools rapid miner*, dataset yang telah dilakukan proses *selection* dan *cleaning* data kemudian ditambahkan label pada atribut yaitu harga, lalu diproses menggunakan *tools rapid miner*. Kemudian menggunakan operator *split* data untuk membagi dataset menjadi data

*training* dan data *testing*, perbandingan keduanya yaitu 90% data *training* dan 10% data *testing*, lalu memasukkan operator *K-Nearest Neighbor* (KNN) yang dapat digunakan untuk prediksi maupun regressi. Algoritma K-NN didasarkan pada jarak tetangga terdekat sebagai nilai prediksi dari *instance* yang baru, pada penelitian ini nilai K yang sudah ditentukan adalah K=3 lalu dihubungkan dengan Operator *Apply Model* digunakan untuk menerapkan model yang telah diperlihatkan sebelumnya menggunakan data training pada *unlabeled* (data *Testing*) yang belum memiliki label. Kemudian operator *performance* ditambahkan untuk pengujian akurasi pada *rapid miner*.

Langkah selanjutnya menghasilkan Klasifikasi dari K-NN dimana dapat dilihat bahwa akurasi algoritmanya ialah 94.03%, serta terdapat 3 cluster yang telah terbagi diantaranya class Murah, Mahal, dan Sangat Mahal kemudian untuk hasil data pada class Murah yaitu 61 data, selanjutnya untuk hasil class Mahal yaitu 107 data, terakhir pada class Sangat Mahal ialah 33 data.

PerformanceVector			
PerformanceVector:			
accuracy: 94.03%			
ConfusionMatrix:			
True:	Murah	Mahal	Sangat Mahal
Murah:	61	0	0
Mahal:	12	95	0
Sangat Mahal:	0	0	33

accuracy: 94.03%				
	true Murah	true Mahal	true Sangat Mahal	class precision
pred. Murah	61	0	0	100.00%
pred. Mahal	12	95	0	88.79%
pred. Sangat Mahal	0	0	33	100.00%
class recall	83.56%	100.00%	100.00%	

Terlihat dari gambar di atas bahwa pengujian dengan menggunakan *Rapid miner* didapatkan hasil dengan akurasi sebesar 94.03% untuk *Recall* 83.56% dan *Precession* 88.73%.

### 3.2.5 Evaluation

Setelah didapatkan nilai akurasi sebesar 94.03% dengan *Recall* 83.56% dan *Precession* 88.73%, dari hasil *modelling* yang sesuai dengan standar *K-Nearest Neighbor*, langkah selanjutnya akan dilakukan tahap *evaluation*

atau pengujian menggunakan metode *Root Mean Square Error (RMSE)* yang sudah diprogram kedalam *python* menggunakan *Google Collaboratory (colab)*. *Root mean square error (RMSE)* merupakan parameter yang digunakan untuk mengevaluasi nilai hasil dari pengukuran terhadap nilai sebenarnya atau nilai dianggap benar. Semakin kecil nilai *RMSE*, maka pengklasteran data semakin mendekati benar. (Febrianti et al., 2016).

### 1. Import Library

Langkah pertama dari pengujian dengan *python* pada *google collab* adalah *import library*, proses ini wajib dilakukan karena akan

```
#import Library
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from sklearn.metrics import mean_squared_error, r2_score
```

berpengaruh terhadap hasil dan cara kerja program berikutnya

Gambar 2 : Import library

### 2. Memasukkan Dataset

Dataset yang telah siap kemudian diunggah kedalam *Google Drive*, agar dapat diolah pada tahap selanjutnya.

```
#memasukkan data
df = pd.read_excel("MILOW.xlsx")
df
```

Gambar 3 : memasukkan dataset

### 3. Menghapus kolom yang tidak digunakan pada proses K-NN

```
#menghapus atribut bulan yang tidak digunakan pada proses knn
df.drop('bulan', axis=1, inplace=True)
df.info()
```

Gambar 4 : menghapus kolom

Coding ini berfungsi untuk menghapus atribut bulan karena tidak digunakan untuk proses k-nn dan atribut dataset yang akan digunakan adalah Luas Panen (Ha), Produksi (Kg), dan

Harga (Rp) serta untuk mengetahui informasi dari dataset yang akan digunakan seperti

```
#Menampilkan Visualisasi Korelasi antar data
plt.figure(figsize=(10,8))
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()
```

variable, tipe data dan jumlah.

### 4. Menampilkan data tiap kategori

Coding ini digunakan untuk menghitung jumlah kemunculan setiap nilai pada kolom 'Class'.

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

Kategori	count
3.0	90
2.0	73
1.0	61

Name: count, dtype: int64

Gambar 5: Menampilkan data tiap kategori

### 5. Menampilkan nilai tiap korelasi antar variabel

Tahap selanjutnya yaitu menampilkan

```
correlation = df.corr()
correlation['Harga (Rp)'].sort_values(ascending=False)
```

Variable	Correlation with Harga (Rp)
Harga (Rp)	1.000000
Luas Lahan (Ha)	-0.317574
Hasil Panen (Kg)	-0.340158
Kategori	-0.931960

Name: Harga (Rp), dtype: float64

korelasi atau keterkaitan antar tiap data.

Proses dibawah ini untuk menampilkan atribut yang akan di proses pada algoritma yang dipakai.

Gambar :

```

from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
X = pd.DataFrame(scaler.fit_transform(X), columns=X.columns)
X
    
```

	Luas Lahan (Ha)	Hasil Panen (Kg)	Harga (Rp)
0	1.235402	1.604430	-1.158646
1	1.235402	1.537777	0.300222
2	1.235402	0.932347	-0.510260
3	1.235402	1.376699	-0.388688
4	1.235402	1.604430	-1.442315
...	...	...	...
219	-2.596245	-1.561581	2.083284
220	-2.596245	-1.339405	0.259698
221	-2.596245	-1.544918	0.664939
222	-2.596245	-1.156110	1.678043
223	-2.596245	-1.550472	1.880663

224 rows x 3 columns

Gambar 6 : nilai tiap kolerasi antar variabel

6. Menampilkan visualisasi korelasi antar data

Plot visualisasi korelasi yaitu menampilkan korelasi data yang berpengaruh besar dan kecil, seperti korelasi dari ketegori 0.19, harga - 0.32, hasil panen 0.68 dapat dibilang korelasi yang berpengaruh kecil sedangkan luas lahan 1.00 dapat dikatakan korelasinya berpengaruh besar.

7. Menampilkan atribut yang akan di proses pada algoritma K-NN

	Luas Lahan (Ha)	Hasil Panen (Kg)	Harga (Rp)
96	0.383925	0.043642	-0.368426
203	-1.319029	-1.222763	-0.550784
169	-0.467552	0.743497	-0.550784
136	-0.041814	-0.622887	-1.239695
182	-0.467552	-0.722866	-0.927659
150	0.383925	0.049196	-0.935764
44	-0.041814	0.215828	-0.935764
198	-1.319029	-1.444939	-0.591309
64	1.235402	1.882150	1.394374
8	1.235402	1.043435	-1.361267
107	0.383925	0.121403	1.495684
37	0.383925	0.576865	-0.125281
5	1.235402	-0.061892	-1.361267
145	-0.041814	-1.078348	1.272801
63	-0.041814	1.110087	-0.327902
184	-0.467552	-1.328296	0.786512
139	-0.041814	0.310253	-0.753405
206	-1.319029	-1.422721	-0.939816
207	-1.319029	-1.339405	1.491632
90	0.383925	-0.945042	1.455160
110	-0.041814	2.043227	-1.118122
74	1.235402	0.932347	1.799615
125	-0.041814	-0.739529	-0.753405

8. Melakukan pembagian data training dan data testing

Tahap ini melakukan pembagian data dimana pada data training 90% dengan jumlah data 201 yang dipakai sedangkan untuk data testing yaitu 10% dan jumlah data yang dipakai ialah 23 data

```
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
```

Out: ((201, 3), (23, 3))

Gambar :

### 9. Menampilkan data training

Proses ini untuk menampilkan data training dimana ada 201 data pada X\_Training dari pembagian dengan data testing

	Luas Lahan (Ha)	Hasil Panen (Kg)	Harga (Rp)
75	1.235402	1.443352	-0.348164
155	0.383925	0.932347	-0.935764
12	1.235402	0.899020	-0.449474
158	0.383925	-0.883944	1.110705
160	-0.467552	0.493549	-0.550784
...	...	...	...
67	1.235402	0.560201	-0.348164
192	-1.319029	-1.172773	0.300222
117	-0.041814	0.987891	0.016553
47	-0.041814	0.621300	-0.408950
172	-0.467552	0.104740	-0.753405

201 rows x 3 columns

Gambar :

### 10. Menampilkan data testing

Proses ini untuk menampilkan data testing dimana ada 23 data pada X\_Testing dari pembagian dengan data training

Gambar :

### 11. Menampilkan hasil prediksi

Proses ini digunakan untuk menampilkan hasil prediksi data testing

```
y_pred = knn.predict(X_test)
y_pred
```

Out: array([2., 3., 3., 3., 3., 3., 3., 3., 1., 3., 1., 2., 3., 1., 2., 1., 2., 3., 1., 1., 3., 1., 3.])

### 12. Menampilkan hasil akurasi dari data testing

Proses selanjutnya yaitu menampilkan hasil akurasi dari data testing yaitu 0.9130

```
from sklearn.metrics import accuracy_score
print('Model accuracy score: {0:0.4f}'.format(accuracy_score(y_test, y_pred)))
```

Out: Model accuracy score: 0.9130

Gambar :

### 13. Menampilkan hasil akurasi dari data training

Proses selanjutnya yaitu menampilkan hasil akurasi dari data testing yaitu 1.0000

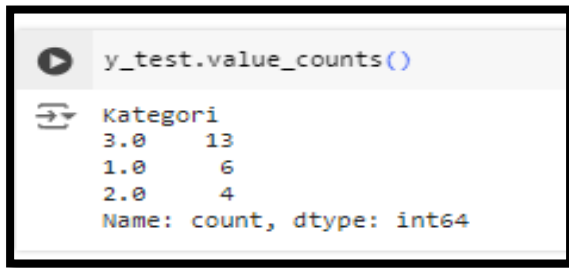
```
y_pred_train = knn.predict(X_train)
print('Training-set accuracy score: {0:0.4f}'.format(accuracy_score(y_train, y_pred_train)))
```

Out: Training-set accuracy score: 1.0000

Gambar :

### 14. Menampilkan hasil kategori dari data testing

Proses ini menampilkan hasil kategori kemunculan data testing yang akan masuk ke kelas murah, mahal, sangat mahal (3,2,1)



```

y_test.value_counts()
kategori
3.0    13
1.0     6
2.0     4
Name: count, dtype: int64

```

Gambar :

## 5. SIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan sebagai berikut:

- 1) Dengan metode *Cross Industry Standard for Data Mining (CRISPDM)*, Dilakukannya proses *business Understanding* terlebih dahulu untuk mengetahui tujuan bisnis penelitian yaitu sebuah model prediksi harga cabai merah pada kota Pagar Alam. Setelah mengetahui tujuan bisnisnya, kemudian dilakukan tahap *modeling*.
- 2) Terdapat 224 record data untuk data training ada 201 data dan data testing ada 23 data dimana untuk hasilnya ada 3 class harga diantaranya class murah, class mahal dan class sangat mahal, untuk class murah sendiri itu ada 61 data selanjutnya untuk class mahal 107 data dan terakhir untuk class mahal ada 33 data. Pada tahap *evaluation* dilakukan perhitungan *Root Means Square Error* dan didapatkan nilai rata-rata 0.09% dan untuk perhitungan menggunakan *Rapidminer* dengan algoritma *K-Nearest Neighbor* didapatkan yaitu *Accuracy* sebesar 94.03% dengan *Recall* 83.56% dan *Precession* 88.73%. dimana hasil tersebut didapat pada nilai  $K=3$  dengan bantuan *Google collab* dan *RapidMiner* Nilai  $K$  merupakan nilai jumlah tetangga terdekat pada algoritma KNN, sehingga nilai tersebut merupakan nilai terbaik yang

didapatkan untuk mengklasifikasi prediksi harga tanaman cabai merah kota pagar alam.

- 3) Berdasarkan hasil kualifikasi diatas, maka penerapan algoritma KNN dapat dinyatakan baik dalam mengklasifikasi suatu data, dan juga dapat diterapkan Untuk membantu dalam mengklasifikasi prediksi harga tanaman cabai merah kota pagar alam.
- 4) Dasdgsgh

## DAFTAR RUJUKAN

- Ademariana, K., Aristoteles, A., Lumbanraja, F. R., & Andrian, R. (2021). Clustering K-Means Jenis Kata Pada Laporan Kegiatan Kuliah Kerja Nyata (Kkn) Universitas Lampung Menggunakan Word2Vec. *Jurnal Pepadun*, 2(2), 221–228. <https://doi.org/10.23960/pepadun.v2i2.64>
- Algoritma, C. (2022). *Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada SD Maria Mediatrix Menggunakan. 2.*
- Bidang, P., Sains, K., Mardi, Y., Gajah, J., No, M., & Barat, S. (n.d.). *Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database ( KDD ) . Jurnal Edik Informatika.*
- Billah, M. B., Jangcik, I., Sasmita, & Yuhelmi. (2023). Penerapan Metode K- Nearest Neighbor Klasifikasi Jenis Buah Semangka Dengan Image Processing. *Zonasi: Jurnal Sistem Informasi*, 5(3), 535–548. <https://doi.org/10.31849/zn.v5i3.16301>
- Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). *Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako.* 8(6), 219–225. <https://doi.org/10.30865/jurikom.v8i6.3655>
- Fatma Ayu Rahman, A., Wartulas, S., Raya Pagojengan, J. K., & Brebes, P. (2020). Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus Di Universitas Peradaban). *Ade Fatma Ayu Rahman IJIR*, 1(2), 70–77.
- Febriansyah, F., & Muntari, S. (2023). Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 8(1), 66–77. <https://doi.org/10.14421/jiska.2023.8.1.66-77>
- Guntara, R. G. (2023). *Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab.* 2(6), 2091–2100.
- Henri Slat, A., Harga Pokok, A., & Henri Slat Fakultas Ekonomi Jurusan Akuntansi Universitas Sam ratulangi Manado, A. (2013). Analisis Harga Pokok Produk Dengan Metode Full Costing Dan Penentuan Harga Jual. *110 Jurnal EMBA*, 1(3), 110–117.
- Irfon, G., & Soen, E. (2022). *Implementasi Cloud Computing dengan Google Colaboratory Pada Aplikasi Pengolah Data Zoom Participants.* 6(1),

- 24–30.
- Khotijah, S., Driyani, D., & Juliana. (2020). Rancang Bangun Edukasi Hardware Komputer Berbasis Android Menggunakan App Inventor Untuk Siswa Sekolah Dasar. *Jurnal Informatika SIMANTIK*, 5(1), 16–21.
- Labolo, A. Y., Bode, A., Colanus, I., Drajana, R., Karim, J., Ichsan, U., & Pohuwato, U. (2022). *Comparasi Algoritma Forecasting Svm , K-Nn Dan Nn*. 4307(2), 289–299.
- Muhamad, Z. (2018). Jurnal Teknik Informatika Atmaluhur. *Jurnal Teknik Informatika Atmaluhur*, 6(1), 40.
- Nindian Puspa Dewi. (2020). Implementasi Holt-Winters Exponential Smoothing untuk Peramalan Harga Bahan Pangan di Kabupaten Pamekasan. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 11(2), 223–236. <https://doi.org/10.31849/digitalzone.v11i2.4797>
- Pelangi, K. C. (2021). *Pangan Di Provinsi Gorontalo Menggunakan Metode K-Nn ( K- Nearest Neighbor )*. 6(2), 2–6.
- Prestasi, D. A. N., & Lalu, M. (n.d.). *Data Mining Untuk Memprediksi Prestasi Siswa Data Mining To Predict Student ' S Achievement Based On Socio-Economic , Motivation , Discipline And*. 4, 222–231.
- Puspita, D., & Aminah, S. (2022). Implementasi Naive Bayes Untuk Sistem Prediksi. *Jurnal Ilmiah Teknosains*, 8(2), 1–6.
- Puspita, D., Syahri, R., Studi, P., Informatika, T., Alam, K. P., & Mining, D. (2023). *Implementasi Algoritma Apriori Untuk Menentukan Strategi*. 8(2).
- Rahmadayanti, F., Lovita, A., & Muntari, S. (2023). *Klasifikasi Kelayakan Penerima Bantuan Sosial Dengan*. 1002–1007.
- Saputri, dkk, F. I. (2022). Perhitungan harga pokok produk dan penerapan cost plus pricing method dalam rangka penerapan harga jual pempek dos. *Society : Jurnal Pengabdian Masyarakat*, 1(1), 51–58. <https://doi.org/10.55824/jpm.v1i1.21>
- Sekolah, L. (2019). *Implementasi Pendidikan Karakter melalui Kegiatan Pembelajaran di Lingkungan Sekolah*. 5, 173–190.
- Setiawan, D. (2018). Dampak Perkembangan Teknologi Informasi dan Komunikasi Terhadap Budaya. *JURNAL SIMBOLIKA: Research and Learning in Communication Study*, 4(1), 62. <https://doi.org/10.31289/simbollika.v4i1.1474>
- Studi, P., Informatika, T., & Alam, P. (2022). *Penerapan Metode Simple Additive Weighing Untuk Penentuan Level Kondisi Penyandang Disabilitas*. 3(4), 559–565. <https://doi.org/10.47065/bits.v3i4.1391>
- Sujarwo, A., & Si, M. (n.d.). *No Title*.
- Syaputra, A. (2022). Implementasi Metode Random Sampling Pada Animasi Motion Grapich Herbisida Dan Fungsida. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 11(2), 142–147. <https://doi.org/10.32736/sisfokom.v11i2.1370>
- Virdaus, D., & Prasetyaningrum, P. T. (n.d.). *Penerapan Data Mining Untuk Memprediksi Harga Bawang Merah Di Yogyakarta Menggunakan Metode K-Nearest Neighbor Implementation Of Data Mining To Predict The Price Of Shallots In Yogyakarta Using The K-Nearest Neighbor Method*. 84, 1–8.
- Wijaya, S. T., Santi, I. H., Wulansari, Z., Informatika, T., Islam, U., Blitar, B., Blitar, K., Timur, J., Neighbor, K., Jagung, H., Mean, R., Error, S., & Tulungagung, K. (2023). *Penerapan Metode K-Nearest Neighbor Untuk Prediksi Harga*. 7(2).
- Wijoyo, A., Nurdiansah, A., Prasajo, D. S., & Ardiana, R. (2023). *Manajemen Data Besar ( Big Data ) Dalam Konteks Sistem Informasi Manajemen*. 1(2), 1–7.
- Wiratmaja, G. H., Wijaya, W. S., Pramana, D. M. A., & Aditya, K. G. R. (2021). Program Menghitung Banyak Bata pada Ruangan Menggunakan Bahasa Python. *TIERS Information Technology Journal*, 2(1), 12–22.
- Zai, C. (2022). *Implementasi Data Mining Sebagai Pengolahan Data*. 2(3), 1–12.

