

PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK PREDIKSI HARGA TOMAT DI KOTA PAGAR ALAM

Penulis¹⁾ (Siti Muntari), Penulis²⁾ (Sasmita)

¹⁾ Program Studi Teknik Informatika, Institut Teknologi Pagar Alam
Jln.Masik Siagim No.75 Simpang Mbacang Kec.Dempo Tengah Kota Pagar Alam
Email: Muntariaza@gmail.com, sasmita@gmail.com.

Abstrak

Tujuan yang akan dicapai dalam penelitian ini yaitu memprediksi harga tomat di Kota Pagar Alam, dari permasalahan yang dialami oleh petani yaitu tidak stabilnya harga tomat di Kota Pagar Alam yang merugikan banyak petani. Serta prediksi harga tomat yang dilakukan oleh dinas pertanian dilakukan dengan cara melihat harga pasar yaitu jika permintaan banyak dan produksi sedikit maka harga mahal, sebaliknya jika permintaan sedikit dan produksi banyak maka harga murah. Metode pengumpulan data untuk memperoleh informasi yaitu dengan cara Observasi, Wawancara, dan Studi Pustaka. Penelitian ini menggunakan algoritma *K-Nearest Neighbor* dengan metode *CRISP-DM* dimana tahapannya meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, *Deployment*. Tools yang digunakan dalam pengolahan data yaitu *Rapid Miner*, metode pengujian menggunakan metode *Root Mean Square Error (RMSE)*. Hasil dari penelitian ini adalah didapatkan hasil akurasi modelling dengan rapid miner sebesar 96,30%, dan tingkat error pada pengujian *root mean square error (RMSE)* sebesar 0,05 yang dimana nilai tersebut dinilai bagus skala pengujian RMSE.

Kata kunci: Harga Tomat, *K-Nearest Neighbor*, *CRISP-DM*, *Rapid Miner*, *RMSE*

Abstract

The goal to be achieved in this study is to predict the price of tomatoes in Pagar Alam City, from the problems experienced by farmers, namely the unstable price of tomatoes in Pagar Alam City which harms many farmers. As well as the prediction of tomato prices carried out by the agricultural department is done by looking at market prices, namely if there is a lot of demand and little production then the price is expensive, otherwise if there is little demand and a lot of production, the price is cheap. The data collection method to obtain information is by means of Observation, Interview, and Literature Study. This study uses the K-Nearest Neighbor algorithm with the CRISP-DM method where the stages include Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Development. The tools used in data processing are Rapid Miner, the testing method uses the Root Mean Square Error (RMSE) method. The results of this study are obtained modeling accuracy with rapid miner of 96.30%, and the error rate in the root mean square error (RMSE) test is 0.05 which is considered good on the RMSE test scale.

Keywords: Tomato Price, *K-Nearest Neighbor*, *CRISP-DM*, *Rapid Miner*, *RMSE*.

1. PENDAHULUAN

Perkembangan teknologi dan kemajuannya semakin bertumbuh pesat seiring bertambahnya waktu, kehadiran teknologi dalam kehidupan manusia memberi banyak dampak dalam berbagai aspek kehidupan. Teknologi yang semakin canggih menawarkan kemudahan bagi kehidupan manusia, dengan adanya teknologi berbagi informasi menjadi lebih mudah, selain itu banyak pekerjaan manusia yang termudahkan dengan adanya teknologi. Secara sosiologis, teknologi merupakan salah satu aspek yang turut mempengaruhi setiap aktivitas, tindakan, serta perilaku manusia. Teknologi informasi dan komunikasi mampu mengubah pola hubungan dan pola interaksi antar manusia. Kehadiran teknologi ini merupakan sesuatu yang tidak dapat dipisahkan dari kehidupan manusia. Perkembangan teknologi yang cepat kadangkala membuat manusia tidak sempat untuk beradaptasi dengan kemajuan tersebut dan akibatnya terjadi anomie dalam masyarakat serta *cultural lag*. (Setiawan, 2018)

Definisi data mining adalah proses ekstraksi suatu data (sebelumnya tidak diketahui, bersifat implisit, dan dianggap tidak berguna) menjadi informasi atau pengetahuan atau pola dari data yang jumlahnya besar. (Suntoro, 2019)

Prediksi (peramalan) adalah usaha menduga atau memperkirakan sesuatu yang akan terjadi di waktu mendatang dengan memanfaatkan berbagai informasi yang relevan pada waktu-waktu sebelumnya (historis) melalui suatu metode ilmiah. Tujuan dari prediksi adalah mendapatkan informasi apa yang akan terjadi di masa datang dengan probabilitas kejadian terbesar. Metode prediksi dapat dilakukan secara kualitatif melalui pendapat para pakar atau secara kuantitatif dengan perhitungan secara matematis. (Wanto & Windarto, 2017) Salah satu algoritma yang dapat digunakan untuk memprediksi adalah algoritma *K-Nearest Neighbor (K-NN)*

Algoritma *K-Nearest Neighbor* (KNN) adalah merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan (K) tetangga terdekatnya. KNN termasuk algoritma supervised learning, dimana hasil dari *query instance* yang baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi algoritma. (Syukri Mustafa & Wayan Simpen, 2019)

Tomat (*Lycopersicon esculentum* Mill) merupakan tanaman hortikultura yang sangat dikenal masyarakat dan mempunyai banyak manfaat, baik sebagai sayuran, bahan baku industri obat-obatan dan kosmetik, maupun sebagai bahan baku pengolahan makanan. Rasa buah tomat yang manis-manis asam memberikan kesegaran pada tubuh dan cita rasa yang berbeda dibandingkan dengan buah-buah lain merupakan ciri khas yang digemari oleh hampir seluruh lapisan masyarakat. (Dewi Nazari et al., 2020)

Berdasarkan hasil observasi dan wawancara pada dinas pertanian Kota Pagar Alam, prediksi harga tomat di Kota Pagar Alam dilakukan dengan cara melihat keadaan pasar, jika permintaan banyak namun produksi tomat sedikit maka harga akan mahal, sedangkan jika permintaan sedikit dan produksi tomat banyak maka harga akan murah. Untuk harga tomat di Kota Pagar Alam tidak stabil, bisa sangat murah atau sangat mahal. Dengan harga tomat yang sangat murah membuat petani tomat merugi, keadaan itu merupakan masalah serius, mengingat tomat merupakan sayuran yang hampir wajib ada di setiap hidangan. Oleh karena itu, diperlukan suatu penelitian untuk mengatasi masalah tersebut, penelitian yang mengimplementasikan data mining dengan menggunakan metode algoritma *K-Nearest Neighbor* akan diterapkan untuk memprediksi harga tomat menggunakan data yang telah tersedia sebelumnya untuk mencari informasi dan rangkaian pola sehingga dapat memecahkan permasalahan yang ada. Penelitian ini diharapkan dapat bermanfaat bagi petani tomat di Kota Pagar Alam, sebagai acuan dalam penentuan waktu tanam.

Berdasarkan penelitian dari (Informasi et al., 2023) dengan judul “Perbandingan Performa Model *Machine Learning Support Vector Machine, Neural Network, Dan K-Nearest Neighbors* Dalam Prediksi Harga Saham” hasil penelitian ini menunjukkan bahwa algoritma *k-nearest neighbor* lebih akurat dibanding algoritma lainnya dalam memprediksi harga saham. Dimana algoritma *k-nearest neighbor* nilai *root mean*

square error (RMSE), paling rendah yaitu 0,037, dibandingkan dengan algoritma *neural network* yang memiliki nilai RMSE 0,048, dan algoritma *support vector machine* yang memiliki nilai RMSE 0,075. Membuktikan bahwa algoritma *k-nearest neighbor* memiliki nilai error terendah dalam prediksi harga saham.

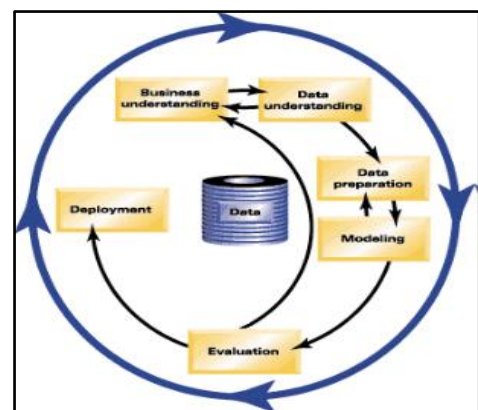
Berdasarkan penelitian dari (Lasulika & Bode, n.d.) dengan judul “Komparasi Algoritma Data Mining Menggunakan Forward Selection pada Prediksi Harga Jagung” hasil penelitian ini metode *k-nearest neighbor* merupakan metode yang sangat bagus atau baik dalam melakukan prediksi ataupun klasifikasi, hal ini dapat dilihat dari nilai RMSE yang dihasilkan yaitu 0,05.

Berdasarkan penelitian dari (Virdaus & Prasetyaningrum, n.d.) dengan judul “Penerapan Data Mining Untuk Memprediksi Harga Bawang Merah Di Yogyakarta Menggunakan Metode K-Nearest Neighbor” hasil penelitian ini yaitu penerapan nilai $k=5$ pada algoritma *K-Nearest Neighbor* menghasilkan tingkat akurasi tertinggi, terjadi pada pemodelan data dalam kondisi ke-2 dari empat kondisi yang diuji, yaitu sebesar 91,67%. Dengan demikian, hasil analisis data menunjukkan bahwa algoritma *K-Nearest Neighbor* berhasil diimplementasikan dengan baik untuk data bawang merah dan memberikan wawasan yang penting bagi petani yang ingin menanam bawang merah ini.

Dari ketiga penelitian diatas dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* memiliki keunggulan dalam memprediksi harga, sehingga dengan menggunakan algoritma ini dapat membantu pihak terkait dalam memprediksi harga dan mengambil langkah untuk kedepannya. Berdasarkan penelitian diatas dapat menjadi referensi bagi peneliti untuk membuat sebuah sistem prediksi harga tomat yang dapat membantu dinas pertanian dalam melakukan prediksi harga tomat secara tersistem.

2. METODE

Metode Pengembangan CRISP-DM



Gambar 1 Siklus CRISP-DM

Pada penelitian ini, digunakannya metodologi data mining CRISP-DM sebagai pemecah masalah yang umum untuk bisnis dan penelitian.

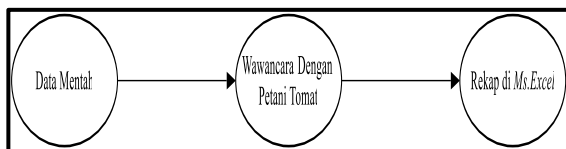
Metodologi ini terdiri dari enam tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Proses metodologi ini terdiri dari enam tahapan yang dapat dijelaskan sebagai berikut:

2.1 Business Understanding

Tahap awal metode CRISP-DM yaitu *business understanding*, yaitu mengacu pada data harga tomat di Kota Pagar Alam. Pada tahap ini diperlukan pemahaman tentang pentingnya pemanfaatan pada data tersebut, agar dapat digunakan untuk mengambil kebijakan yang lebih objektif untuk mempermudah dalam mengetahui hasil produksi harga tomat di Kota Pagar Alam.

2.2 Data Understanding

Pada tahap ini data yang diambil adalah data tomat Kota Pagar Alam, tahun 2023. Yang memiliki 5 atribut yaitu Nama, Bulan, Luas panen (Ha), Jumlah produksi (Ton), dan Harga jual tomat.



Gambar 2 Tahap Business Understanding

2.3 Data Preparation

Kemudian pada data dilakukan pemilihan atribut dan sebagian data dalam atribut akan dikonversikan agar proses data mining menjadi lebih mudah.

Persiapan data atau *data preparation* terbagi menjadi 2 bagian yaitu:

1. Data Selection

Memilih struktur basis data awal yang akan dijadikan data dan atribut untuk perhitungan data mining.

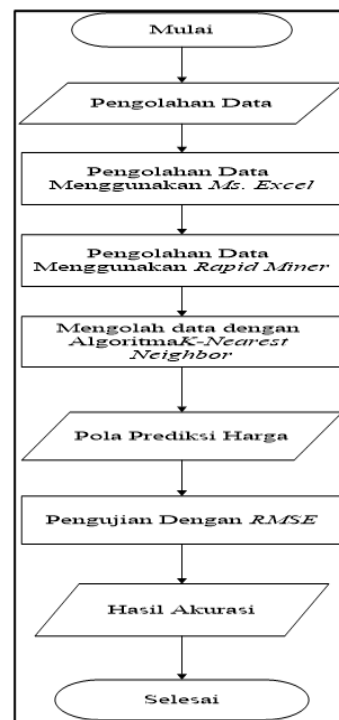
2. Data Processing

Data yang ada kemudian akan diproses terlebih dahulu sebelum diolah. Proses yang dilakukan adalah membersihkan data-data yang tidak lengkap (*noisy*).

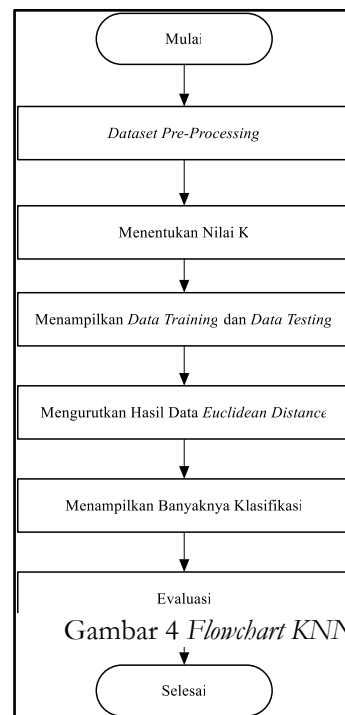
2.4 Modeling

Tahap ini adalah proses mencari informasi dalam sekumpulan data dengan metode atau algoritma tertentu mengikuti proses dari CRISP-DM. algoritma yang digunakan dalam penelitian ini adalah *K-Nearest Neighbor*.

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



Gambar 3 Flowchart alur penelitian



Gambar 4 Flowchart KNN

1. Menentukan nilai *K* untuk penentuan yang akan digunakan, dalam hal ini tidak ada aturan khusus dalam penentuan nilai *K*. Namun, untuk nilai *K* disarankan menggunakan angka ganjil, hal ini bertujuan agar terhindar dari data yang memiliki jumlah yang sama.
2. Menghitung jarak antara *data training* dan *data testing*. Perhitungan jarak yang digunakan

dalam penelitian ini yaitu *Euclidean Distance*, dengan rumus sebagai berikut:

dimana :

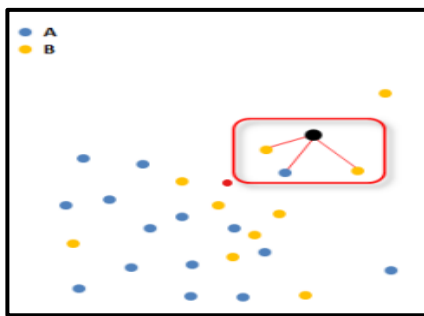
p_i = sample data / data training

q_i = data uji / data testing

i = variabel data n = dimensi data

$$euc = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

3. Mengurutkan jarak yang terbentuk dari hasil perhitungan dengan *Euclidean distance*.
4. Menentukan jarak terdekat sampai urutan nilai K
5. . Menentukan kelas yang sesuai



Gambar 5 Pola KNN

Keterangan:

Titik Biru : Kelas A

Titik Kuning : Kelas B

Gambaran pola diatas menunjukkan bahwa titik berwarna biru merupakan kelas A, dan titik berwarna kuning adalah kelas B.

2.5 Evaluation

Tahap ini adalah tahap dimana model sudah terbentuk lalu data kemudian dievaluasi keefektifan sebelum digunakan dan apakah siap untuk ditetapkan pada fase awal ditahap *business understanding*. Ditahap akhir ini juga ditentukan hasil proses data mining pada tahap evaluasi ini difokuskan pada model *K-Nearest Neighbor*.

Dalam penelitian ini peneliti menggunakan metode pengujian *Root Mean Square Error (RMSE)*, *RMSE* memiliki fungsi untuk mengumpulkan besarnya kesalahan dalam prediksi pada nilai data menjadi kesatuan keakurasian prediksi. *RMSE* memiliki akurasi yang tinggi dalam melakukan pengukuran dengan probabilitas hingga 50%, berikut persamaan rumus *RMSE*

Keterangan

Y' = Nilai Prediksi

Y = Nilai Sebenarnya

N = Jumlah Data

2.6 Devloymet

Pada tahap terakhir ini *CRISP-DM* adalah penyebaran pengetahuan atau informasi yang diperoleh dan diimplementasikan pada laporan dan melakukan analisis tiap pola yang diperoleh agar mudah dimengerti oleh pihak dinas pertanian dan petani.

3.1 Hasil

Setelah penelitian ini dilakukan menghasilkan suatu model klasifikasi perhitungan prediksi harga tomat di Kota Pagar Alam, didapatkanlah dengan algoritma, model yang dapat digunakan dalam mengklasifikasi harga tomat di Kota Pagar Alam dan indikator-indikator yang ada ialah yang menggambarkan performa seperti *accuracy*, *presisi*, *recall*, dan lain-lain. Data yang digunakan merupakan data harga tomat pada tahun 2023 yang diperoleh dengan teknik wawancara kepada petani. Data awal berjumlah 210 data yang terdiri dari 5 atribut yaitu Nama, Bulan, Luas Panen (Ha), Produksi (Kg), Harga (Rp) yang kemudian dilakukan proses data selection dan data cleaning, kemudian dilakukan proses modelling pada rapid miner dengan algoritma, pada tahap modelling ini data dibagi menjadi dua bagian yaitu data training dan data testing. Dengan perbandingan yaitu 90% data training dan 10% data testing. Sedangkan untuk nilai k yang digunakan yaitu $k=3$. Kemudian didapatkan hasil akurasi pada rapid

3. HASIL DAN PEMBAHASAN

miner sebesar 96.30%, dengan recall 100% dan presisi 89.71%

3.2.1 Business Understanding

Tahap pemahaman bisnis (*Business Understanding*) pada tahap ini didapatkan suatu masalah yang dialami banyak petani tomat di kota pagar alam, yaitu permasalahan harga tomat yang tidak stabil. Tujuan dari penelitian ini adalah untuk mengetahui algoritma yang tepat untuk memprediksi harga tomat di Kota Pagar Alam, sehingga menghasilkan suatu kebijakan untuk mengatasi permasalahan harga tomat bagi petani Kota Pagar Alam. Dimana sebelumnya tidak ada penelitian lebih lanjut mengenai harga tomat, yang dimana prediksi harga tomat dilakukan dengan melihat keadaan pasar saja.

3.2.2 Data Understanding

Pada tahap pemahaman data atau *data understanding* data yang digunakan adalah data produksi tomat pada tahun 2023, data diperoleh dengan teknik observasi yang dilakukan ke Dinas Ketahanan Pangan Kota Pagar Alam, studi pustaka dilakukan dengan cara mereview jurnal atau penelitian terdahulu yang memiliki objek atau metode yang sama, dan teknik wawancara yang dilakukan dengan petani tomat Kota Pagar Alam. Data yang diperoleh berjumlah 210 record dengan atribut sebanyak 5 yaitu, Nama, Bulan, Luas Panen (Ha), Produksi (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 Panen (Ha), Produksi (Kg), dan Harga (Rp). Atribut yang akan digunakan dalam proses data mining ini yaitu 4 atribut diantaranya Bulan, Luas Panen (Ha), Produksi (Kg), dan Harga (Rp).

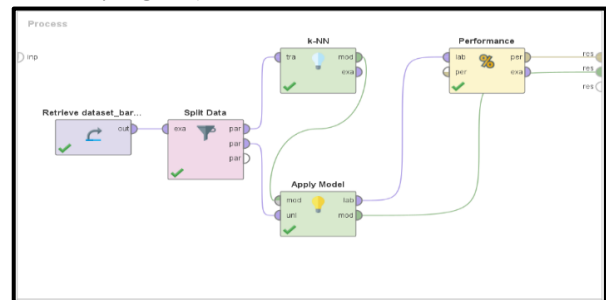
b. Data Cleaning

Tahap *processing* atau *cleaning* data merupakan proses pembersihan atau penghilangan data yang tidak lengkap (*noisy*),

data yang akan *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 tidak menemukan ada data yang *missing value*, data yang didapatkan telah sesuai dengan kriteria pengolahan data pada rapid miner, sehingga data dapat dilanjutkan pada tahap 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 210 *record* kemudian dilakukan



Gambar 6 Proses Modelling

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 gambar 7 dapat dilihat 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 regresi. 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*.

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

	precision	recall	f1-score	support
1.0	1.00	1.00	1.00	3
2.0	0.80	1.00	0.89	4
3.0	1.00	0.93	0.96	14
accuracy			0.95	21
macro avg	0.93	0.98	0.95	21
weighted avg	0.96	0.95	0.95	21

Gambar 7. Laporan hasil klasifikasi

Setelah tahapan *modeling* dilakukan pada *rapid miner* maka dapat dilihat pada Gambar 4.3 Dihasilkan 3 kelas harga yaitu Mahal, Murah, dan Sangat Mahal. Setelah tahapan *modeling* telah dilakukan dan kelas

KNNClassification	
Weighted 3-Nearest Neighbour model for classification.	
The model contains 21 examples with 3 dimensions of the following classes:	
Mahal	
Murah	
Sangat Mahal	

Gambar 8 Kelas Harga

harga sudah didapatkan maka pada Gambar 4.4 di atas dapat disimpulkan bahwa pengujian dengan menggunakan *Rapid miner* dengan menggunakan K senilai 3 maka di dapat hasil akurasi sebesar 96.30% dengan *Recall* 100% dan *Precession* 89.71%.

Akurasi : 96.30%				
	True mahal	Pred murah	Pred sangat mahal	Class precision
True mahal	61	0	7	89.71%
Pred murah	0	121	0	100.00%
Pred sangat mahal	0	0	0	0.00%
Class recal	100.00%	100.00%	0.00%	

Gambar 9 Akurasi *rapid miner*

3.2.5 Evaluation

Setelah didapatkan nilai akurasi sebesar 96.30% dengan *Recall* 100% dan *Precession* 89.71%, 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)

- Menampilkan laporan hasil klasifikasi

Proses ini untuk menghasilkan laporan klasifikasi. Laporan ini berisi

informasi tentang presisi, recall, nilai F1, dan dukungan untuk setiap kelas.

- Menampilkan hasil RMSE

Coding tersebut menampilkan nilai error dari hasil dengan pengujian *Root Means Square Error* (RMSE), seperti dibawah ini. Nilai error dari pengujian RMSE yaitu sebesar 0.05 dan *R-squared* yaitu 0.91, nilai tersebut didapatkan setelah melakukan proses pengujian pada *google colab* dengan metode *root mean square error*.

3.2.6 Deployment

```
print(f'Root Mean Squared Error (RMSE): {mse:.2f}')
print(f'R-Squared (R2): {r_squared:.2f}')
```

Root Mean Squared Error (RMSE): 0.05
R-Squared (R2): 0.91

Gambar 10 Nilai RMSE

Pada tahap ini akan dibuat laporan penelitian dan mempresentasikan dari model di atas yang telah melalui proses *data mining* dan *google collab*. Tahap pertama dari penelitian ini adalah *CRISP-DM*, dimana tahap awal dari *crisp-dm* adalah *business understanding* yaitu mencari permasalahan dan menentukan algoritma yang sesuai, tahap kedua yaitu *data understanding* adalah tahapan mengumpulkan data, data yang akan diolah adalah data tahun 2023 yang berjumlah 210 record dengan atribut nama, bulan, luas panen (Ha), produksi (Kg), dan harga (Rp) tahap ketiga yaitu *data preparation* tahap ini adalah penyempurnaan data yang akan diolah tahap awal dari *data preparation* yaitu *data selection* pada tahap ini dilakukan seleksi terhadap atribut yang akan digunakan, dimana dari 5 atribut data awal yang akan digunakan hanya 4 atribut yaitu bulan, luas panen (Ha), produksi (Kg), dan harga (Rp). Dan tahap kedua adalah *data cleaning*, tahap ini menghilangkan data yang tidak lengkap (*noisy*).

4. SIMPULAN

Berdasarkan hasil penelitian yang dilakukan maka didapatkan kesimpulan sebagai berikut:

- 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 tomat pada kota Pagar Alam. Setelah mengetahui tujuan bisnisnya, kemudian dilakukan tahap *modeling*.

2) Pada tahap *evaluation* dilakukan perhitungan *Root Mean Square Error* dan didapatkan nilai rata-rata dari *Accuracy* sebesar 0.05, dan untuk perhitungan menggunakan *Rapidminer* dengan algoritma *K-Nearest Neighbor* didapatkan yaitu Akurasi 96.30% *Precesion* 89,71% dan untuk *Recall* sebesar 100%.

5. DAFTAR PUSTAKA

- 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>
- Budiarto, E. H., Yuana, R. A., Maryono, D., & Maret, U. S. (2017). *Jurnal Ilmiah Pendidikan Teknik Kejuruan (JIPTTEK)*. 61–70.
- Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. *JURIKOM (Jurnal Riset Komputer)*, 8(6), 219.
<https://doi.org/10.30865/jurikom.v8i6.3655>
- Dewi, A. O. P. (2020). Big Data di Perpustakaan dengan Memanfaatkan Data Mining. *Anuva: Jurnal Kajian Budaya, Perpustakaan, Dan Informasi*, 4(2), 223–230.
<https://doi.org/10.14710/anuva.4.2.223-230>
- Dewi Nazari, A. P., Rusdiansyah, R., Meklin Siregar, A. P., & Rahmi, A. (2020). Pertumbuhan Dan Hasil Tanaman Tomat (*Lycopersicum Esculentum* Mill.) Pada Pemberian Pupuk Zn Dan Jarak Tanam Yang Berbeda. *Ziraa'Ah Majalah Ilmiah Pertanian*, 45(3), 241.
<https://doi.org/10.31602/zmip.v45i3.3482>
- Di, U., Tasikmalaya, K., & Web, B. (2022). *Jurnal teknik informatika*. 10(2).
- Drs.Cholid Narbuko, D. H. A. A. (2016). *Metodologi Penelitian*.
- Endah Wahyurini, SP, Msi, Ir Lagiman, M. (2020). *Teknik Budidaya Dan Pemuliaan Tanaman Tomat* (1st ed.). Lembaga Penelitian dan Pengabdian Kepada Masyarakat.
- Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi. *Teknika*, 8(1), 11–16.
<https://doi.org/10.34148/teknika.v8i1.95>
- 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>
- Febrianti, F., Hafiyusholeh, M., & Asyhar, A. H. (2016). Perbandingan Pengklusteran Data Iris Menggunakan Metode K-Means Dan Fuzzy C-Means. *Jurnal Matematika "Mantik,"* 2(1), 7.
<https://doi.org/10.15642/mantik.2016.2.1.7->