

PENERAPAN DATA MINING UNTUK MEMPREDIKSI HARGA BAHAN PANGAN DI INDONESIA MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

Rahmadini^{a*}, Enjel Erika LorencisLubis^b, Aji Priansyah^c, Yolanda R.W.N.^d, Tuti Meutia^e

^{a,e}Fakultas Ekonomi, Universitas Samudra

^bFakultas Ilmu Komputer, Universitas Methodist Indonesia

^cFakultas Teknik, Universitas Malikussaleh

^dFakultas Ekonomi dan Bisnis, Universitas Muhammadiyah Sumatera Utara

^{a*}Email: niniedini@gmail.com

Abstract

Food is one of the important sectors for human life, because the basic human need is food. Over time, food prices in Indonesia are often unstable. This of course has a big impact on the community and farmers. The cause of this food price volatility can be caused by several factors, such as environmental factors, pest and planthopper attacks, and dry land. The application of one of the data mining methods in the process of predicting food prices, especially predicting the price of rice in Indonesia with this prediction, people throughout Indonesia can be more vigilant in the future to prevent hunger. The benefit of predicting food prices for farmers is to prevent large losses caused by the drop in food prices. Meanwhile, the Government's goal is to maintain food security in the territory of the Indonesian Government to prevent hunger. The dataset used comes from the National PIHPS Website. The method used is the K-Nearest Neighbor (KNN) regression algorithm and for testing it uses RMSE and MAE. The results of this study, the K-Nearest Neighbor method with the regression model can predict the price of rice in January 2019 - December 2021 which is recorded every month with the addition of variable data Harvested Area (ha), Production Yield (tons) of the total experiment using range $k = (2,10)$ then the best prediction results are by using $k = 2$ with MAE and RMSE for training data 52.77 and 96.40 and for testing data 55.55 and 81.64 and $K = 2$ parameters which have been normalized.

Keywords : *rice price prediction, data mining, K-Nearest Neighbor (KNN) regression algorithm*

Abstrak

Pangan merupakan salah satu sektor yang penting untuk kehidupan manusia, karena kebutuhan pokok manusia yaitu kebutuhan pangan. Seiring berjalannya waktu, harga pangan di Indonesia sering kali tidak stabil. Hal ini tentu saja berpengaruh yang besar bagi masyarakat dan petani. Penyebab dari ketidakstabilan harga Bahan Pangan ini bisa disebabkan oleh beberapa faktor, seperti faktor lingkungan, serangan hama dan wereng, dan lahan kekeringan. Penerapan salah satu metode data mining dalam proses prediksi harga bahan pangan khususnya prediksi harga beras di Indonesia dengan prediksi ini masyarakat di Seluruh Indonesia bisa lebih berjaga-jaga di waktu yang akan datang guna mencegah terjadinya kelaparan. Manfaat memprediksi harga bahan pangan ini bagi para petani untuk mencegah terjadinya kerugian besar yang disebabkan Anjloknya harga bahan pangan tersebut. Sedangkan Pemerintah yaitu untuk agar tetap terjaganya ketahanan pangan di wilayah Pemerintah Indonesia guna mencegah terjadinya kelaparan. Dataset yang digunakan berasal dari Website PIHPS Nasional. Adapun metode yang digunakan adalah algoritma regresi K-Nearest Neighbor (KNN) serta untuk pengujiannya menggunakan RMSE dan MAE. Hasil dari penelitian ini, metode K-Nearest

Neighbor dengan model regresi dapat melakukan prediksi terhadap harga Beras pada bulan Januari 2019 – Desember 2021 yang didata pertiap bulannya dengan diambahkan data variabel Luas Panen(ha), Hasil Produksi(ton) dari total percobaan dengan menggunakan range $k=(2,10)$ maka hasil prediksi terbaik dengan menggunakan $k = 2$ dengan MAE dan RMSE untuk data training 52.77 dan 96.40 dan untuk data testing 55.55 dan 81.64 dan parameter $K = 2$ yang sudah dinormalisasi.

Kata kunci : prediksi harga beras, data mining, algoritma regresi K-Nearest Neighbor (KNN)

PENDAHULUAN

Tiga isu dunia saat ini sedang marak maraknya di dunia ini, seperti *Food, Water and Energy*. Salah satu dari tiga isu tersebut, *food* (pangan) adalah isu yang paling sering disorot oleh pemerintah. Seperti sektor pertanian merupakan salah satu sektor yang penting dalam kehidupan masyarakat. Indonesia merupakan negara kepulauan yang terbesar di dunia. Total luas daratan di negara ini sekitar 190 juta hektar, yang sekitar 55 juta hektarnya adalah lahan pertanian 129 juta hektarnya adalah perhutanan. Pada lahan pertanian, sekitar 24 juta hektar terdiri dari lahan yang subur. Dengan ketersediaan lahan yang luas, pertanian menjadi salah satu penghasil ekonomi negara dengan menyediakan lapangan pekerjaan dan sebagai penghasil devisa negara. Pertanian juga sebagai sumber ekonomi utama bagi masyarakat pedesaan. Sebagai salah satu sumber ekonomi masyarakat, penentuan harga yang tepat sudah menjadi hal yang sangat penting. Dan pastinya petani tentu saja mempunyai keinginan untuk mendapat keuntungan dari hasil bertani. Tetapi akhir akhir ini sering terjadi perubahan harga komoditas pertanian, seperti komoditas beras yang dimana harga bahan pangan tersebut bisa berubah tidak menentu secara signifikan. Hal ini tentu mempengaruhi bagi petani dan masyarakat .Banyak faktor yang mempengaruhi harga beras tersebut, seperti faktor lingkungan/ cuaca, serangan hama dan wereng dan kekeringan di berbagai daerah. Beras merupakan kebutuhan pokok manusia ,tanpa makan manusia tidak bertenaga,bisa sakit bahkan bisa berujung kematian. Selain faktor lingkungan dan serangan hama, perubahan *supply* dan *demand* juga menyebabkan harga pangan tidak stabil. Misalnya, permintaan yang tinggi dan kurangnya pasokan sangat berpengaruh pada harga. Jika pasokan sedikit dan permintaan banyak, maka harga pangan akan mengalami kenaikan, demikian juga untuk sebaliknya.Ketidakpastian harga produk pertanian meningkatkan risiko yang dihadapi oleh para pemangku kepentingan yang terlibat dalam rantai pasokan produk pertanian.

Risiko akan merugikan kepentingan produsen pertanian, hal tersebut bertentangan dengan pembangunan pertanian yang sehat tetapi juga pembangunan ekonomi nasional yang sehat. Oleh karena itu, menjadi acuan penting bagi produsen, konsumen, pengolah, perencana pembangunan pedesaan dan orang lain serta instansi/lembaga yang terlibat di pasar untuk menggunakan metode dalam melakukan suatu prediksi untuk mendapatkan harga produk pertanian dan kemudian memprediksi fluktuasi harga produk pertanian. Maka dari itu, prediksi harga sangat penting untuk dilakukan. Dengan dilakukannya ini, bisa memudahkan petani untuk mengambil keputusan melakukan penanaman waktu yang tepat dan memilih tanaman apa yang ditanam yang pada saat panen petani mendapatkan

harga yang tinggi guna mencegah kerugian dan memperoreh hasil yang maksimal. Dalam penelitian sebelumnya yang dilakukan oleh Mohamad Efendi Lasulika menggunakan *algoritma KNearest Neighbor (KNN)* tentang prediksi harga komoditi jagung diperoleh akurasi sebesar 98.7% [4]. Dalam penelitian ini algoritma yang digunakan adalah algoritma KNN berbasis *Particle Swarm Optimazation (PSO)*. Sedangkan hasil dari penelitian ini algoritma yang digunakan adalah algoritma K-Nearest Neighbor Regresor. Implementasi algoritma *K-Nearest Neighbor (KNN)* bertujuan untuk memprediksi harga beras tahun 2019-2021 data dari pasar tradisional di seluruh Indonesia yang diambil dari website PIHPS Nasional bulan Januari 2019-Desember 2021, proses algoritma regresi *K-Nearest Neighbor (KNN)* dan dari total percobaan dengan menggunakan range $k=(2,10)$ maka hasil prediksi terbaik dengan menggunakan $k = 2$ dengan MAE dan RMSE untuk data training 52.77 dan 96.40 dan untuk data testing 55.55 dan 81.64 dan parameter $K = 2$ yang sudah dinormalisasi.

Permasalahan yang dibahas dalam penelitian ini adalah bagaimana mengimplementasikan algoritma regresi *K-Nearest Neighbor (KNN)* dalam memprediksi harga Komodias Beras di Indonesia. Adapun batasan masalah dalam penelitian ini adalah data yang digunakan adalah data set bersumber dari *website PIHPS Nasional* bulan Januari 2019-Desember 2021. Data harga bahan pangan yang diprediksi yaitu harga beras ,Yang didata perbulan dari *website PIHPS Nasional* dari proses algoritma regresi *K-Nearest Neighbor (KNN)* dan menghitung RMSE (*Root Mean Square Error*) dan MAE (*Mean Absolute Error*). penelitian sebelumnya yang dilakukan oleh Lasulika, Implementasi algoritma *K-Nearest Neighbor (KNN)* bertujuan untuk memprediksi harga beras premium tahun 2019 dari BPS Kota Bandung dengan harga beras tahun 2019 dari proses algoritma regresi *K-Nearest Neighbor (KNN)* dan dibandingkan dengan K data tetangga terdekat untuk memperoleh nilai RMSE (*Root Mean Square Error*) yang kecil, karena nilai RMSE semakin kecil semakin akurat.

Adapun tujuan dari penelitian adalah mengimplementasikan algoritma regresi *K-Nearest Neighbor (KNN)* dalam memprediksi harga Komodias beras dan menganalisis algoritma *regresi K-Nearest Neighbor* dengan metode lain untuk memprediksi harga Komodias beras.

KERANGKA TEORITIS

BMKG

Badan Meteorologi, Klimatologi, dan Geofisika (disingkat BMKG), sebelumnya bernama Badan Meteorologi dan Geofisika (disingkat BMG) adalah Lembaga Pemerintah Nonkementerian Indonesia (LPNK) yang mempunyai tugas melaksanakan tugas pemerintahan di bidang meteorologi, klimatologi, dan geofisika. Badan Meteorologi Klimatologi dan Geofisika (BMKG) merupakan satu-satunya lembaga di Indonesia yang telah ditunjuk secara resmi untuk mengeluarkan informasi mengenai keadaan cuaca dan iklim. Penyebaran informasi cuaca yang cepat dari dan ke Unit Pelaksana Teknis (UPT) BMKG merupakan salah satu hal pokok yang mendukung penyelenggaraan BMKG. Salah satu metoda penyebaran data yang digunakan BMKG adalah menggunakan sistem multicast yang memberikan solusi alternatif model penyebaran informasi cuaca dan

diseminasi cuaca buruk. Sistem multicast menawarkan pilihan untuk dapat mengirimkan file dari satu server ke banyak client secara bersamaan.

Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi. Pengertian Prediksi sama dengan ramalan atau perkiraan. Menurut kamus besar bahasa Indonesia, prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan merupakan input bagi proses perencanaan dan pengambilan keputusan. Tujuan dari pada diadakannya peramalan kebijakan adalah untuk memperoleh informasi mengenai perubahan dimasa yang akan datang yang akan mempengaruhi terhadap implementasi kebijakan serta konsekuensinya. Oleh karenanya, sebelum rekomendasi diformulasikan perlu adanya peramalan kebijakan sehingga akan diperoleh hasil rekomendasi yang benar-benar akurat untuk diberlakukan pada masa yang akan datang. Didalam memprediksi kebutuhan yang akan datang dengan berpijak pada masa lalu, dibutuhkan seseorang yang memiliki daya sensitifitas tinggi dan mampu membaca kemungkinan-kemungkinan dimasa yang akan datang. Peramalan kebijakan juga diperlukan untuk mengontrol, dalam artian, berusaha merencanakan dan menetapkan kebijakan sehingga dapat memberikan alternatif-alternatif tindakan yang terbaik yang dapat dipilih diantara berbagai kemungkinan yang ditawarkan oleh masa depan .

Machine Learning

Machine learning adalah cabang kecerdasan buatan (*Artificial Intelligence*) adalah disiplin ilmu yang mencakup perancangan dan pengembangan algoritme yang memungkinkan komputer untuk mengembangkan perilaku berdasarkan data empiris, seperti dari sensor data basis data. Sistem pembelajar dapat memanfaatkan contoh (data) untuk menangkap ciri yang diperlukan dari probabilitas yang mendasarinya (yang tidak diketahui). Data dapat dilihat sebagai contoh yang menggambarkan hubungan antara variabel yang diamati. Fokus besar penelitian pemelajaran mesin adalah bagaimana mengenali secara otomatis pola kompleks dan membuat keputusan cerdas berdasarkan data. Kesukarannya terjadi karena himpunan semua perilaku yang mungkin, dari semua masukan yang dimungkinkan, terlalu besar untuk diliput oleh himpunan contoh pengamatan (data pelatihan). Karena itu pembelajar harus merampatkan (generalisasi) perilaku dari contoh yang ada untuk menghasilkan keluaran yang berguna dalam kasus-kasus baru. Mesin yang dimaksud di sini adalah mesin dalam pengertian lebih mendekati kepada 'sistem', bukan mesin 'mekanik'. Istilah pemelajaran pertama kali muncul dalam disiplin ilmu kecerdasan buatan. Pemelajaran berarti menambah pengetahuan, memahami dengan belajar, dan mengikuti perintah. Pemelajaran mesin merupakan salah satu cabang

dari kecerdasan buatan yang membahas mengenai pembangunan sistem yang didapat berdasarkan pada pembelajaran data, atau sebuah studi yang mempelajari cara untuk memprogram sebuah komputer untuk belajar.

Algoritma K-Nearest Neighbor

Algoritma *k* tetangga terdekat (bahasa Inggris: *k-nearest neighbour algorithm*, disingkat *k-NN*) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran digambarkan ke ruang berdimensi banyak dengan tiap-tiap dimensi mewakili tiap ciri/fitur dari data. Klasifikasi data baru dilakukan dengan mencari label *k* tetangga terdekat. Label terbanyak yang muncul menjadi label data baru. Bila $k = 1$, data baru dilabeli dengan label tetangga terdekat. Jarak yang biasa dipakai adalah jarak Euklides. Algoritma KNN Regressor adalah algoritma dengan mengelompokkan data berdasarkan letak ketetanggaannya, pengelompokan data tersebut bergantung pada jumlah nilai *k*. Nilai RMSE dihasilkan untuk mengukur tingkat error dari model yang sudah dibuat. Untuk lebih jelasnya seperti berikut :

1. Menentukan *K* itu sebagai bilangan bulat positif berdasarkan ketersediaan data pembelajaran.
2. Memilih tetangga terdekat dari data baru sebanyak *K*.
3. Menghasilkan nilai prediksi dari data sampel baru.
4. Menentukan RMSE yang bergantung pada pada nilai sebenarnya dengan nilai prediksi

Berikut ini adalah kelebihan atau keunggulan dari algoritma *K- Nearest Neighbor*:

1. Sangat sederhana dan mudah dipahami
2. Sangat mudah diterapkan
3. Dapat digunakan dalam proses klasifikasi maupun regresi.
4. Dapat digunakan dalam jumlah kelas yang berbeda-beda
5. Tidak memerlukan proses training dan proses pembangunan model, karena data baru langsung dikelaskan.
6. Sangat mudah jika akan dilakukan penambahan data
7. Parameter yang diperlukan sedikit, yaitu hanya jumlah tetangga yang dipertimbangkan (*K*), dan metode perhitungan jaraknya (distance metrik)
8. Hasil pemodelan tidak linear, sehingga cocok untuk klasifikasi data yang batasannya tidak linear.

Sedangkan beberapa kelemahan atau kekurangan dari algoritma *k-nearest neighbor* antara lain:

1. Perlu menentukan nilai *K* yang tepat.
2. Computation cost yang tinggi
3. Waktu pemrosesan yang lama jika datasetnya sangat besar.
4. Tidak cukup bagus jika diterapkan pada high dimensional data
5. Sangat sensitif pada data yang memiliki banyak noise (*noisy data*), banyak data yang hilang (*missing data*), dan pencilan (*outliers*) .

Data Mining

Data mining sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Salah satu teknik yang dibuat dalam *data mining* adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. *Data mining* memiliki banyak sekali fungsi, Untuk fungsi utamanya sendiri yaitu ada dua; Yaitu fungsi *descriptive* dan fungsi *predictive*. fungsi deskripsi dalam *data mining* adalah sebuah fungsi untuk memahami lebih jauh tentang data yang diamati. Dengan melakukan sebuah proses diharap bisa mengetahui perilaku dari sebuah data tersebut. Data tersebut itulah yang nantinya dapat digunakan untuk mengetahui karakteristik dari data yang dimaksud. Dengan menggunakan Fungsi *descriptive Data mining*, Maka nantinya bisa menemukan pola tertentu yang tersembunyi dalam sebuah data. Dengan kata lain jika pola yang berulang dan bernilai itulah karakteristik sebuah data bisa diketahui. Sedangkan Fungsi prediksi merupakan sebuah fungsi bagaimana sebuah proses nantinya akan menemukan pola tertentu dari suatu data. Pola-pola tersebut dapat diketahui dari berbagai variabel-variabel yang ada pada data. Ketika sudah menemukan pola, Maka pola yang didapat tersebut bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya. Dalam melakukan pengumpulan informasi tentunya terdapat metode, Metode tersebut akan membantu dalam proses menemukan data. *Data mining* akan menyediakan perencanaan dari ide hingga implementasi akhir.

A. Proses pengambilan Data

1. **Data Cleansing**, Proses dimana data-data yang tidak lengkap, mengandung error dan tidak konsisten dibuang dari koleksi data. Ketahui juga data lifecycle management untuk mengetahui tentang pengolahan data.
2. **Data Integration**, Proses integrasi data dimana yang berulang akan dikombinasikan.
3. **Selection**, Proses seleksi atau pemilihan data yang relevan terhadap analisis untuk diterima dari koleksi data yang ada.
4. **Data Transformation**, Proses transformasi data yang sudah dipilih ke dalam bentuk mining procedure melalui cara dan agresi data.
5. **Data Mining**, Proses yang paling penting dimana akan dilakukan berbagai teknik yang diaplikasikan untuk mengekstrak berbagai pola-pola potensial untuk mendapatkan data yang berguna.
6. **Pattern Evolution**, Sebuah proses dimana pola-pola menarik yang sebelumnya sudah ditemukan dengan identifikasi berdasarkan measure yang telah diberikan
7. **Knowledge Presentation**, Merupakan proses tahap terakhir, Dalam hal ini digunakan teknik visualisasi yang bertujuan membantu user dalam mengerti dan menginterpretasikan hasil dari penambahan data.

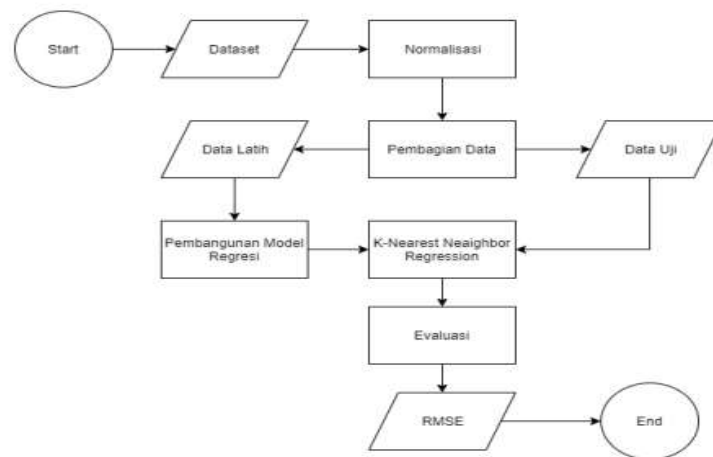
B. Teknik dalam Proses Penambangan Data

1. **Predictive Modeling**, Terdapat dua teknik yaitu *Classification* dan *Value Prediction*

2. **Database Segmentation**, Melakukan partisi database menjadi sejumlah *segmen*, *cluster*, atau *record* yang sama
3. **Link analysis**, Sebuah teknik untuk membuat hubungan antara *record* yang individu atau sekumpulan record dalam database.
4. **Deviation detection**, Sebuah teknik untuk mengidentifikasi *outlier* yang mengekspresikan sebuah deviasi dari ekspektasi yang sudah diketahui sebelumnya.
5. **Nearest Neighbour**, Yaitu teknik yang memprediksi pengelompokan, Teknik ini sendiri merupakan teknik yang tertua yang digunakan dalam *data mining*.
6. **Clustering**, merupakan teknik untuk mengklasifikasikan data berdasarkan kriteria masing-masing data.
7. **Decision Tree**, Merupakan teknik generasi selanjutnya, dimana teknik ini adalah sebuah model prediktif yang dapat digambarkan seperti pohon. Setiap node yang terdapat dalam struktur pohon tersebut mewakili sebuah pertanyaan yang digunakan untuk menggolongkan data .

Sistem yang Dibangun

Proses untuk memprediksi harga beras mulai dari pengambilan data, preprocessing, proses, dan evaluasi terhadap hasil. Diagram alur sistem dapat dilihat pada gambar 1.



Gambar 1. Alur sistem program

Dimulai dari Proyek akhir distart ,kemudian mengumpulkan dataset dari [website https://hargapangan.id](https://hargapangan.id) atau PIHPS Nasional ,Setelah dataset terkumpul kemudian data tersebut dinormalisasi , Tujuan normalisasi data yaitu untuk Menghilangkan dan mengurangi penginputan data yang sama secara berulang dan Memastikan data yang bersifat depedensi.Setelah data sudah dinormalisasi kemudian data dibagi menjadi 2 bagian yaitu *Data Training* dan *Data Testing*. Kemudian dilakukan pengujian terhadap data dengan menggunakan Algoritma *K-Nearest Neighbor Regression*, Setelah dilakukan pengujian kemudian hasil tersebut dievaluasi , kemudian diambil nilai RMSE yang paling kecil yang akan dibandingkan dengan nilai MAE terkecil, kemudian ditariklah sebuah kesimpulan dan yang terakhir proyek selesai (*End*) ,

Dataset

Sumber data harga beras pada penelitian ini diperoleh dari data set Website PIHPS Nasional atau <https://hargapangan.id> di Indonesia. dataset luas panen(ha) dan hasil Produksi (ton) diperoleh dari dataset BPS (Badan Pusat Statistik). Data harga beras yang diambil dari data pasar tradisional perharinya yang ada diseluruh Indonesia kemudian data tersebut dikumpulkan dan diperoleh berupa data harga rata - rata setiap bulannya, data harga beras yang di ambil dalam penelitian ini dari Januari tahun 2019 - Desember 2021 dengan total jumlah *record* sebanyak 36 . Preprocessing data yang dilakukan pada penelitian ini menggunakan normalisasi data. Normalisasi data ini digunakan untuk mengelompokkan data ke dalam skala atau jangkauan tertentu sehingga mempermudah dalam mengolah data. Data yang ada masih berupa data univariate, sehingga perlu diubah lagi menjadi multivariate, sehingga menjadi beberapa variabel independen. Pada tahap evaluasi ini, nilai yang dihasilkan dari pengujian berupa nilai *Root Mean Square Error* (RMSE). Nilai RMSE yang kecil menunjukkan hasil akurasi prediksi yang terbaik.

Normalisasi

Pada tahap ini normalisasi data diperlukan untuk membuat data yang nilainya menyebar bisa dinormalkan dengan normalisasi dengan cara membuat rentang nilainya menjadi lebih kecil. Normalisasi yang dilakukan menggunakan min-max normalization yang membuat rentang 0-1. Persamaan min-max *normalization* dapat dilihat sebagai berikut :

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (1)$$

Dimana X adalah rata-rata dari suatu data, min(X) adalah nilai minimal dari data X, dan max(X) adalah nilai maksimal dari data X.

Pembagian Data

Memprediksi suatu model yang dibuat harus terdapat dataset yang akan dipecah menjadi dua bagian yaitu data latih dan data uji. Dataset terdiri dari empat parameter yaitu bulan, luas panen(ha), Hasil Produksi (ton) sebagai variabel x dan label harga beras sebagai parameter y. Data latih digunakan untuk membuat model regresi yang diujikan terhadap data uji, data latih yang alokasikan kurun waktu tahun 2019 - 2021 dengan *record* sebanyak 12 data untuk masing-masing tahun, dan penulis mengambil data uji pada tahun 2019-2021.

Pembangunan Model Regresi KNN

Regresi KNN menggunakan jarak kedekatan antar tetangga yang direpresentasikan nilainya dengan variabel “k”. Jarak kedekatan antar tetangga dihitung menggunakan persamaan *euclidean distance* sebagai berikut.

$$\text{dist}((x_1, y_1, \dots), (x_2, y_2, \dots)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + \dots} \quad (2)$$

Dimana, x_1 dan x_2 merupakan suatu titik yang ditinjau dari sumbu x dan y. y_1 dan y_2 merupakan titik acuan yang ditinjau dari sumbu x dan y. X_1 merupakan parameter dataset dari bulan, x_2 merupakan parameter dataset dari luas Panen, y_1 merupakan parameter dataset dari jumlah hasil Produksi (ton) dan y_2 merupakan parameter dataset dari . Setelah harga beras titik-titik dibandingkan, pilih jarak yang terkecil, ambil sebanyak “k” sehingga mendapatkan nilai jarak terkecil sebanyak “k”, setelah itu hitung rata-rata nilai euclidean distance. Misalkan euclidean distance titik A ke titik B,C,D,E,F masing-masing berjarak 5 cm, 8 cm, 9 cm, 4 cm, dan 7 cm dengan parameter $k = 3$ maka ambil jarak yang terkecil sebanyak 3 nilai yaitu titik B,E,F dengan nilai 5 cm, 4 cm, dan 6 cm, setelah itu hitung nilai rata-rata dari ketiga titik tersebut yaitu jumlahkan ketiga titik kemudian dibagi banyaknya titik. Perhitungan tersebut menghasilkan nilai 5 dari perhitungan $(5+4+6)/3$.

Evaluasi Model

Evaluasi merupakan penilaian dari model regresi yang sudah dibuat. Metode regresi dikatakan baik jika memiliki *Root Mean Square Error* (RMSE) yang artinya hasil prediksi regresi yang dibuat oleh model regresi lebih mendekati nilai sebenarnya. Cara menghitung RMSE dapat dilihat pada persamaan.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3)$$

Dimana n adalah banyaknya data test, $\hat{y}$ merupakan nilai prediksi harga beras dari model yang sudah dibuat, sedangkan y adalah nilai harga beras sebenarnya.

Evaluasi Hasil

Pengujian yang dilakukan penulis di penelitian ini dilakukan 2 skenario untuk menemukan RMSE paling rendah untuk menghasilkan model yang baik menggunakan KNN regression dan membandingkan RMSE KNN regression dengan MAE KNN regression.

Preprocessing Data

Pada *Preprocessing* data penulis terlebih dahulu melakukan normalisasi data kemudian melakukan pembagian data menjadi 2 bagian yaitu *Data Training* dan *Data Testing*

	Luas Panen (ha)	Hasil Produksi (ton)	Harga (Rp)
31	-0.074364	-0.162279	11700.0
20	0.327415	0.413390	11850.0
16	0.986548	0.740846	11950.0
30	0.449887	0.437297	11700.0
22	-0.719578	-0.591094	11800.0

Gambar 2. Hasil Data Training

	Luas Panen (ha)	Hasil Produksi (ton)	Harga (Rp)
23	-1.374803	-1.303152	11800.0
4	0.080414	-0.082419	11700.0
18	0.024278	-0.035720	11900.0
21	-0.236442	-0.130905	11850.0
19	0.754358	0.555140	11850.0

Gambar 3 . Hasil Data Testing

Hasil Pengujian

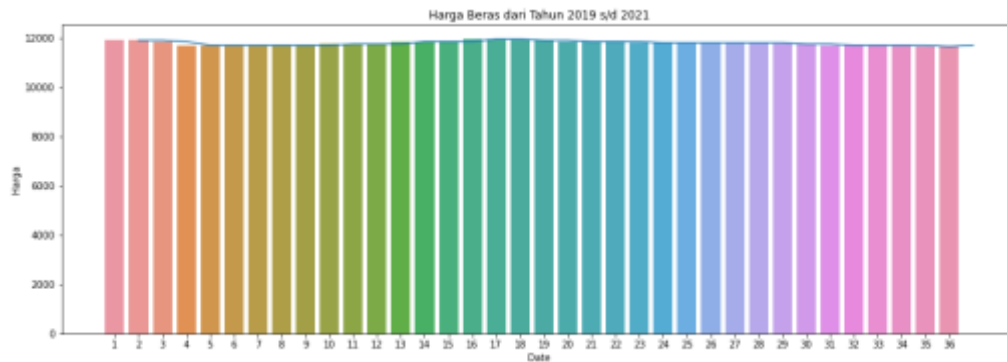
Skenario pertama bertujuan untuk menentukan nilai k yang paling optimal yang digunakan untuk memprediksi harga beras dengan RMSE yang rendah sehingga model yang dihasilkan menjadi baik. Penulis mendapatkan **nilai k = 2** merupakan nilai terbaik untuk data yang ternormalisasi. Dari total percobaan dengan menggunakan range k=(2,10) maka hasil prediksi terbaik dengan menggunakan k = 2 dengan MAE dan RMSE untuk data training 52.77 dan 96.40 dan untuk data testing 55.55 dan 81.64.

	K	MAE Training	RSME Training	MAE Testing	RSME Testing
0	2	52.777778	96.407699	55.555556	81.649658
1	3	59.876543	96.407699	70.370370	77.777778
2	4	58.333333	96.407699	86.111111	94.648472
3	5	61.481481	96.407699	90.000000	96.666667
4	6	62.037037	96.407699	89.814815	97.301555
5	7	63.756614	96.407699	91.269841	101.323442
6	8	66.8981438	96.407699	90.972222	99.978296
7	9	66.255144	96.407699	87.654321	96.438641
8	10	69.444444	96.407699	85.555556	96.407699

Gambar 4 . Tabel Hasil Pengujian

Visualisasi Model

Dari gambar dapat kita lihat secara jelas grafik harga beras pertiap bulannya dari Januari 2019- Desember 2021 sebanyak 36 record. Dari Visualisasi model yang berbentuk histogram dapat kita lihat perbandingan harga komoditas beras tidak terlalu jauh mengalami penerunanan maupun kenaikan bahkan bias dikaakan harga beras relatif stabil dari tiap bulan selama 3 tahun yaitu pada Januari 2019- Desember 2021.



Gambar 5. Visualisasi Model

Prediksi Harga Menggunakan Algoritma K-Nearest Neighbor

Hasil Dari Pengujian Prediksi Harga Menggunakan *Algoritma K-Nearest Neighbor* diperoleh prediksi harga yang tidak jauh beda dari harga sebelumnya.

	Harga (Rp)	Prediksi Harga (Rp)
23	11800.0	11800.0
4	11700.0	11740.0
18	11900.0	11740.0
21	11850.0	11760.0
19	11850.0	11765.0
9	11750.0	11790.0
34	11650.0	11800.0
3	11700.0	11815.0
0	11900.0	11800.0

Gambar 6. Hasil Pengujian Prediksi Harga

KESIMPULAN, KETERBATASAN DAN SARAN

Berdasarkan Hasil pengujian , diperoleh kesimpulan dari total percobaan dengan menggunakan range $k=(2,10)$ maka hasil prediksi terbaik dengan menggunakan $k = 2$ dengan MAE dan RMSE untuk data training 52.77 dan 96.40 dan untuk data testing 55.55 dan 81.64 dan parameter $K = 2$ yang sudah dinormalisasi, dan pada visualisasi model, diperoleh kesimpulan grafik harga beras tidak terlalu signifikan mengalami penurunan serta kenaikan. Selanjutnya, pada Preprocessing data penulis membuat 2 bagian yaitu Data

Training Dan Data Testing hasilnya masing-masing mendekati 0 dan metode KNN Regressor lebih unggul karena prediksi harga tidak terlalu jauh dari harga sebelumnya. Dan berdasarkan total percobaan dengan menggunakan Metode KNN Regressor diperoleh range $k=(2,10)$ maka hasil prediksi terbaik dengan menggunakan $k = 2$ dengan MAE dan RMSE untuk data training 52.77 dan 96.40 dan untuk data testing 55.55 dan 81.64 dan parameter $K = 2$ yang sudah dinormalisasi.

Penelitian ini memiliki keterbatasan yang dapat diperbaiki oleh peneliti berikutnya, yaitu dataset yang di gunakan hanya bersumber dari website PIHPS Nasional bulan Januari 2019-Desember 2021. Data harga bahan pangan yang diprediksi yaitu hanya harga beras ,Yang didata perbulan dari website PIHPS Nasional dari proses algoritma regresi K-Nearest Neighbor (KNN) dan menghitung RMSE (*Root Mean Square Error*) dan MAE (*Mean Absolute Error*). Jadi, saran untuk peneliti selanjutnya untuk dapat dikembangkan adalah dengan menambahkan beberapa algoritma machine learning lainnya seperti *Decision Tree*, *Line Regression*, *Support Vector Machine* dan lainnya sebagai perbandingan dalam melakukan prediksi. Dapat menggunakan dan menambahkan dataset yang akan digunakan machine learning dalam mempelajari sehingga dapat mendapatkan hasil yang lebih baik.

REFERENSI

- 1) E. Quincieu, "Asian Development Bank Papers on Indonesia," Summary of Indonesia's Agriculture, Natural Resources, and Environment Sector Assesment, 2015.
- 2) K. Pertanian, "Rencana Strategis Kementrian Pertanian tahun 2015 - 2019," Kementrian Pertanian Republik Indonesia, 2015.
- 3) P. H. S. Indonesia, "Tehnik dan Budidaya Penanaman Padi System of Rice Intensification (SRI)," Pusat Pelatihan Kewirausahaan Sampoerna , Pasuruan, 2018.
- 4) M. E. Lasulika, "Prediksi Harga Komoditi Jagung Menggunakan K-NN dan Particle Swarm Optimazation sebagai Fitur Seleksi," ILKOM Jurnal Ilmiah, vol. 9, no. 3, p. 233, 2017.
- 5) W. W. C. Sri, "Komparasi Beberapa Metode Estimasi Kesalahan Pengukuran," Jurnal Penelitian dan Evaluasi Pendidikan, vol. 13, no. 2, pp. 182-197, 2013.
- 6) I. Student, "4 Pengertian Beras Menurut Para Ahli," [Online]. Available: <https://www.indonesiastudents.com/pengertian-beras-menurut-para-ahli/>.
- 7) P. Bulog, "Sekilas Perum Bulog," [Online]. Available: <http://www.bulog.co.id/sekilas.php>. [Diakses January 2020].
- 8) Y. N. N. Bikan, "Kajian Implementasi Sistem Multicast dalam Sistem Jaringan BMKG," 2015.
- 9) Herdianto, "Prediksi Kerusakan Motor Induksi Menggunakan Metode Jaringan Saraf Tiruan Backpropagation," dalam Tesis, Universitas Sumatera Utara, Medan, 2013.
- 10) U. D. Arni, "Definisi dan Konsep Utama Machine Learning," Kamis Agustus 2018. [Online]. Available: <https://garudacyber.co.id/artikel/915-definisi-dan-konsep-utama-machine-learning>.

- 11) M. B. S. Imandoust, "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events : Theoretical Background," *Int. Journal of Engineering Research and Applications*, vol. 3, no. 5, pp. 605-610, 2013.
- 12) E. Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*, Yogyakarta: CV. Andi Offset, 2012.
- 13) Wang, L., Feng, J., Sui, X., Chu, X., & Mu, W. "Agricultural product price forecasting methods: research advances and trend". *British Food Journal*, Vol. 122 Issue. 7, pp 2121–2138. 2020
- 14) Guan, H., Dai, Z., Guan, S., & Zhao, A. "A forecasting model based on high-order fluctuation trends and information entropy". *Entropy*, Vol. 20 Issue 9, pp 1–15, 2018.
- 15) Supranto, J. *Metode Ramalan Kuantitatif Untuk Perencanaan*. Jakarta: Erlangga, 1984
- 16) Bray, S. R., & Wang, B. "Forecasting unprecedented ecological fluctuations". *PLoS Computational Biology*, Vol 16 Issue 6 , pp 1–17, 2020
- 17) Subagyo, Pangestu. *Forecasting Konsep dan aplikasi*. Yogyakarta: BPPE UGM, 1986.
- 18) Girond, F., Randrianasolo, L., Randriamampionona, L., Rakotomanana, F., Randrianarivelosia, M., Ratsitorahina, M., Brou, T. Y., Herbreteau, V., Mangeas, M., Zigiumugabe, S., Hedje, J., Rogier, C., & Piola, P. "Analysing trends and forecasting malaria epidemics in Madagascar using a sentinel surveillance network: a webbased application". *Malaria Journal*, Vol. 16 Issue 1, pp 1–11, 2017.
- 19) P. I. Sijabat et al., "Algoritma Backpropagation Prediksi Harga Komoditi terhadap Karakteristik Konsumen Produk Kopi Lokal Nasional," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 11, no. 1, pp. 96–107, May 2020, doi: 10.31849/DIGITALZONE.V11I1.3880.
- 20) A. A. Fardhani, D. I. N. Simanjuntak, and A. Wanto, "Prediksi Harga Eceran Beras Di Pasar Tradisional Di 33 Kota Di Indonesia Menggunakan Algoritma Backpropagation," *J. InfomediaTeknik Inform. Multimed. Jar.*, vol. 3, no. 1, pp. 25–30, Aug. 2018, doi: 10.30811/JIM.V3I1.625.
- 21) "Pusat Informasi Harga Pangan Strategis Nasional.," May 15, 2022. <https://hargapangan.id> (accessed May 15, 2022).