

Implementasi K-Nearest Neighbor untuk Prediksi Harga Bibit Kelapa Sawit dengan Metode CRISP-DM

Reni Nursyanti¹⁾, Graha Prakarsa²⁾, Siti Nuraini Rahmawati³⁾, Agung Andriana⁴⁾

¹⁾ Informatika, Fakultas Teknologi Dan Informatika, Universitas Informatika Dan Bisnis Indonesia

^{2,3)} Sistem Informasi, Fakultas Teknologi Dan Informatika, Universitas Informatika Dan Bisnis Indonesia

⁴⁾ Sistem Informasi, Fakultas Sain Dan Teknologi, Universitas Terbuka

Email: reninursyanti@unibi.ac.id; grahaprakarsa@unbi.ac.id; sitinuraini@studenunibi.ac.id;
053316533@ecampus.ut.ac.id.

Diterima:
10 Februari 2025

Diterima Setelah Revisi:
26 Februari 2025

Dipublikasikan:
27 Februari 2025

Abstrak

Kelapa sawit merupakan tanaman penghasil minyak nabati paling produktif dan memiliki peran penting dalam perekonomian Indonesia. Namun, fluktuasi harga bibit kelapa sawit menjadi tantangan bagi Pusat Penelitian Kelapa Sawit (PPKS) Kebun Sungai Lilin akibat ketidakseimbangan antara permintaan dan produksi. Penelitian ini bertujuan untuk memprediksi harga bibit kelapa sawit menggunakan algoritma K-Nearest Neighbor (KNN) dengan pendekatan Cross-Industry Standard Process for Data Mining (CRISP-DM). Dataset yang digunakan terdiri dari 2.049 catatan, dengan pembagian 80% data pelatihan dan 20% data uji. Dengan nilai $k = 5$, model KNN menghasilkan akurasi 99,21%, dengan 3 data salah klasifikasi dan 1.626 data benar. Hasil prediksi menunjukkan kategori harga "MAHAL". Meskipun KNN terbukti efektif dalam memprediksi harga bibit kelapa sawit dengan akurasi tinggi, metode ini memiliki beberapa keterbatasan, seperti sensitivitas terhadap pemilihan nilai k dan ketergantungan pada distribusi data dalam ruang multidimensi. Selain itu, KNN dapat menjadi kurang efisien ketika diterapkan pada dataset yang lebih besar karena kompleksitas komputasi yang meningkat dalam proses pencarian tetangga terdekat. Dibandingkan dengan metode lain, seperti Support Vector Machine (SVM) atau Random Forest, KNN lebih sederhana dalam implementasi namun kurang robust terhadap data dengan dimensi tinggi atau distribusi yang tidak merata. Hasil studi ini menunjukkan bahwa KNN dapat menjadi alat yang berguna bagi pemangku kepentingan dalam pengambilan keputusan dan strategi bisnis terkait harga bibit kelapa sawit. Namun, untuk aplikasi skala besar atau lingkungan dengan variabilitas data yang lebih tinggi, eksplorasi metode lain yang lebih adaptif dapat menjadi pilihan yang lebih optimal.

Kata Kunci: Bibit Kelapa Sawit, Harga, Kategorisasi, *K-Nearest Neighbor*

Abstract

Oil palm is the most productive vegetable oil-producing crop and plays a crucial role in Indonesia's economy. However, fluctuations in oil palm seedling prices pose a challenge for the Pusat Penelitian Kelapa Sawit (PPKS) Kebun Sungai Lilin due to the imbalance between demand and production. This study aims to predict oil palm seedling prices using the K-Nearest Neighbor (KNN) algorithm with the Cross-Industry Standard Process for Data Mining (CRISP-DM) approach. The dataset consists of 2,049 records, divided into 80% training data and 20% testing data. With $k = 5$, the KNN model achieved an accuracy of 99.21%, with 3 misclassified data points and 1,626 correctly classified instances. The prediction results indicate a "HIGH" price category. Although KNN has proven effective in predicting oil palm seedling prices with high accuracy, the method has several

limitations, such as sensitivity to the choice of k value and dependence on data distribution in a multidimensional space. Additionally, KNN may become inefficient when applied to larger datasets due to the increased computational complexity in searching for the nearest neighbors. Compared to other methods such as Support Vector Machine (SVM) or Random Forest, KNN is simpler to implement but less robust for high-dimensional data or imbalanced distributions. This study demonstrates that KNN can be a useful tool for stakeholders in decision-making and business strategy related to oil palm seedling prices. However, for large-scale applications or environments with higher data variability, exploring alternative methods that are more adaptive may be a more optimal choice..

Keywords: Oil Palm Seedlings, Price, Categorization, K-Nearest Neighbor

1 PENDAHULUAN

Badan Pusat Statistik mendata peningkatan hasil produksi kelapa sawit terjadi pada tahun 2017-2021 rata-rata sebesar 9,88% per tahun. Faktor yang mendasari kenaikan harga pada bibit kelapa sawit yaitu banyaknya permintaan dan meningkatnya hasil produksi kelapa sawit. Banyaknya pesaing dalam produksi bibit kelapa sawit membuat Perusahaan harus dapat mengantisipasi terjadinya fluktuasi harga dan juga memprediksi harga yang akan dipasarkan di masa depan. Salah satu cara untuk mengantisipasi fluktuasi harga adalah dengan melakukan prediksi harga. Prediksi harga ditujukan untuk melakukan peramalan atau perkiraan harga dimasa depan guna meminimalisir resiko dan memaksimalkan keuntungan yang akan didapatkan. Salah satu cara menemukan informasi dari sejumlah data yang besar yaitu dengan menggunakan teknik data mining.

Algoritma K-Nearest Neighbor merupakan suatu metode yang memakai algoritma supervised dimana output berdasarkan sampel uji yang baru diklasifikasikan menurut domain berdasarkan kategori dalam KNN (Yahya et al,2020). KNN membandingkan data uji yang diberikan dengan data latih yang sama. (Okta Danik N et all, 2017). Prediksi harga telah dilakukan dalam penelitian oleh (Mirdayanti Hiku et al., 2019), di mana Metode K-Nearest Neighbor digunakan untuk memprediksi harga eceran beras di pasar tradisional Gorontalo. Hasil penelitian ini menunjukkan tingkat akurasi Confusion Matrix sebesar 80%. Pemilihan algoritma *K-Nearest Neighbor* (KNN) dengan pendekatan *Cross-Industry Standard Process for Data Mining* (CRISP-DM) didasarkan pada beberapa keunggulan utama. KNN merupakan metode yang sederhana namun efektif dalam menangani masalah klasifikasi. Berbeda dengan *Artificial Neural Network* (ANN) atau *Support Vector Machine* (SVM) yang memerlukan tuning parameter yang kompleks dan sulit diinterpretasikan, KNN menghasilkan prediksi yang mudah dipahami oleh pemangku kepentingan di industri kelapa sawit. Selain itu, KNN memiliki kemampuan menangani data dengan pola yang tidak selalu linear, yang sering ditemukan dalam analisis harga bibit kelapa sawit yang dipengaruhi oleh faktor ekonomi dan produksi. Dibandingkan dengan metode regresi yang lebih cocok untuk hubungan linear, KNN lebih fleksibel dalam mengenali pola dalam dataset. Keunggulan lain dari KNN adalah ketahanannya terhadap *outlier*, karena metode ini mengklasifikasikan data berdasarkan kedekatan dengan data historis tanpa bergantung pada distribusi tertentu. Hal ini membuatnya lebih andal dalam menghadapi fluktuasi harga yang disebabkan oleh faktor eksternal. Dengan menggunakan pendekatan CRISP-DM, proses analisis data menjadi lebih terstruktur dan sistematis, mulai dari pemahaman bisnis, eksplorasi data, persiapan data, pemodelan, evaluasi, hingga implementasi hasil, sehingga model yang dihasilkan dapat lebih mudah diterapkan dalam pengambilan keputusan bisnis.

Penelitian selanjutnya dilakukan oleh Dapit & Putri 2020 yaitu Prediksi Harga Bawang Merah Dengan Menggunakan Metode K-Nearest Neighbor (KNN) diperoleh hasil akurasi sebesar 91.67% dengan menggunakan nilai $K=5$ (Virdaus & Prasetyaningrum, 2020). Penelitian selanjutnya

dilakukan oleh Edward Ridley Tauran yaitu Prediksi Harga Saham Pt Bank Central Asia Tbk Berdasarkan Data Dari Bursa Efek Indonesia Menggunakan Metode K-Nearest Neighbors (KNN) Hasil penelitian ini didapatkan akurasi sebesar 61,79% nilai K yang menghasilkan akurasi tinggi yaitu K=37 (2021). Penelitian selanjutnya dilakukan oleh Wiyli Yustanti Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah Hasil akurasi dari prediksi data testing adalah sebesar 80% dengan menggunakan nilai K sebesar K=5 (2012).

Pusat Penelitian Kelapa Sawit (PPKS) Kebun Sungai Lilin adalah salah satu perusahaan penyedia bibit tanaman kelapa sawit yang menyediakan bibit kelapa sawit unggul dan telah lolos sertifikasi UPTD-BPSBTP Provinsi Sumatera Selatan dengan label biru dan sertifikat mutu benih. Dengan jaminan kualitas bibit diharapkan dapat mampu memenuhi permintaan pasar yang ada. Berdasarkan data dari Badan Pusat Statistik Kabupaten Musi Banyuasin Kecamatan Sungai Lilin memiliki luas perkebunan kelapa sawit yang belum menghasilkan pada tahun 2020 yaitu sebesar 1512,0 hektar dan luas perkebunan kelapa sawit yang menghasilkan pada tahun 2020 yaitu sebesar 1450,0 hektar.

Permasalahan yang ada pada Pusat Penelitian Kelapa Sawit (PPKS) Kebun Sungai Lilin yaitu banyaknya permintaan dan sedikitnya produksi yang dihasilkan dengan begitu harga yang dipasarkan pun mengalami fluktuasi. Adapun tujuan dari penelitian ini adalah Untuk mengetahui bagaimana algoritma k-nearest neighbor di terapkan dalam memprediksi harga bibit kelapa sawit dan berapa akurasi yang dihasilkan untuk prediksi harga bibit kelapa sawit menggunakan metode k-nearest neighbor.

2 KAJIAN PUSTAKA

K-Nearest Neighbor (KNN) adalah salah satu algoritma machine learning yang digunakan untuk klasifikasi dan regresi. KNN merupakan metode berbasis instance (instance-based learning) dan termasuk dalam kategori lazy learning, karena tidak membangun model eksplisit saat pelatihan, tetapi melakukan prediksi berdasarkan kedekatan dengan data yang sudah ada (Virdaus, D, 2020).

Metode KNN bekerja dengan mencari sejumlah k tetangga terdekat (nearest neighbors) dari suatu data uji berdasarkan jarak tertentu. Kelas dari data uji kemudian ditentukan berdasarkan mayoritas kelas dari tetangga-tetangga terdekatnya.

Langkah-langkah utama dalam KNN adalah:

1. Menentukan nilai k (jumlah tetangga terdekat yang digunakan untuk prediksi).
2. Menghitung jarak antara data uji dan semua data pelatihan, biasanya menggunakan: Jarak Euclidean (paling umum digunakan):

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

Jarak Manhattan

Jarak Minkowski

3. Menentukan k tetangga terdekat dengan jarak terkecil.
4. Melakukan klasifikasi atau regresi berdasarkan mayoritas kelas atau rata-rata nilai dari tetangga terdekat.

Pemilihan Nilai k, Pemilihan nilai k yang tepat sangat penting dalam performa KNN, k terlalu kecil yaitu model menjadi terlalu sensitif terhadap outlier (high variance). Sedangkan k terlalu besar yaitu model menjadi terlalu general dan kehilangan detail lokal (high bias). Biasanya, nilai k ganjil digunakan untuk menghindari hasil prediksi yang ambigu pada klasifikasi biner.

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah kerangka kerja standar yang digunakan dalam proyek data mining dan machine learning. Metode ini dikembangkan untuk

memberikan pendekatan sistematis dalam eksplorasi dan analisis data, sehingga proses pengolahan data menjadi lebih terstruktur, efisien, dan dapat direplikasi di berbagai industri.

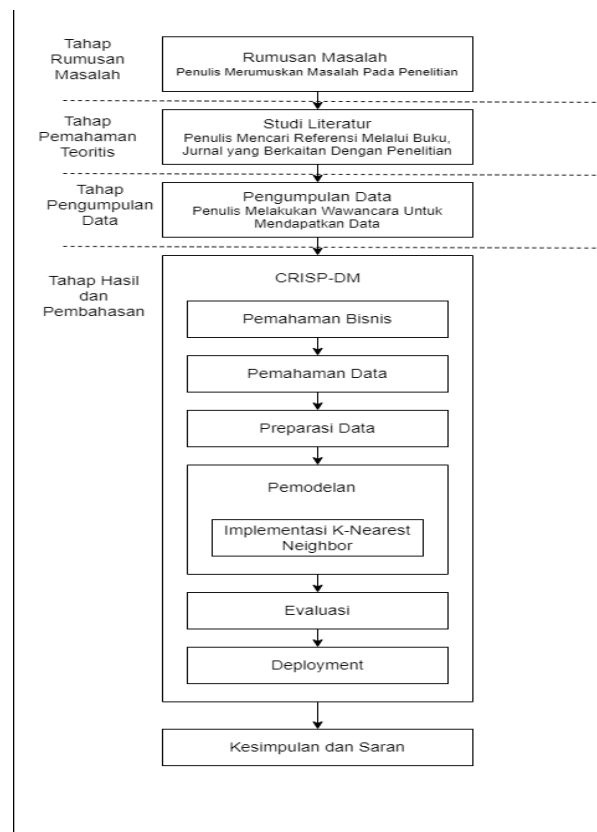
Metode CRISP-DM terdiri dari enam tahapan utama, yang bersifat iteratif dan fleksibel, yaitu:

1. Business Understanding (Pemahaman Bisnis), Pada tahap ini, tujuan utama proyek didefinisikan berdasarkan kebutuhan bisnis. Pemahaman mendalam terhadap permasalahan yang ingin diselesaikan sangat penting untuk menentukan arah analisis data.
2. Data Understanding (Pemahaman Data), Tahap ini mencakup eksplorasi awal dataset, termasuk pengumpulan data, analisis deskriptif, dan identifikasi potensi masalah seperti missing values atau outlier.
3. Data Preparation (Persiapan Data), Pada tahap ini, data yang telah dikumpulkan dibersihkan dan disiapkan agar bisa digunakan dalam model machine learning. Langkah-langkah yang dilakukan meliputi: Pembersihan data (mengatasi missing values dan outlier), Transformasi data (normalisasi atau standarisasi), Feature selection (pemilihan variabel yang relevan).
4. Modeling (Pembuatan Model), Tahap ini melibatkan pemilihan dan penerapan algoritma machine learning yang sesuai untuk menyelesaikan masalah yang telah dirumuskan. Model dapat diuji dengan berbagai parameter untuk mendapatkan hasil terbaik.
5. Evaluation (Evaluasi Model), Model yang telah dibangun dievaluasi menggunakan metrik performa seperti akurasi, presisi, recall, dan confusion matrix. Hasil evaluasi ini menentukan apakah model sudah cukup baik atau perlu dilakukan penyesuaian..
6. Deployment (Implementasi Model), Tahap terakhir adalah mengintegrasikan model ke dalam sistem yang dapat digunakan untuk pengambilan keputusan. Jika model hanya untuk penelitian, dokumentasi hasil dan kesimpulan juga bisa menjadi bagian dari tahap ini.

3 METODE PENELITIAN

Pada penelitian ini menggunakan metode yang mengacu pada metodologi Cross Industry Standard Process for Data Mining (CRISP-DM) dimana pada tahap penelitiannya memiliki enam fase. Dalam pengimplementasiannya terdiri dari enam fase yaitu: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment yang digambarkan pada Tahapan Penelitian sebagai berikut:

1. Tahap Perumusan Masalah. Rumusan masalah yang diangkat pada penelitian ini yaitu mengenai bagaimana memprediksi harga bibit kelapa sawit menggunakan algoritma K-Nearest Neighbor yang nantinya akan di kaji dan didapatkan informasi yang berkaitan dengan masalah tersebut.
2. Tahap Pemahaman Teoritis. Tahap ini dilakukan pembelajaran studi literatur mengenai prediksi dan klasifikasi harga dan yang berhubungan dengan keduanya.
3. Tahap Pengumpulan Data. Pada tahap ini melakukan wawancara kepada bagian administrasi kantor PPKS Kebun Sungai Lilin berdasarkan pertanyaan yang ajukan mengenai harga bibit kelapa sawit kepada bagian administrasi yaitu jika harga Tandan Buah Segar (TBS) mengalami kenaikan maka kemungkinan besar permintaan bibit kelapa sawit akan meningkat atau karena banyak petani yang berpindah dari tanaman karet dan lebih memilih untuk menanam sawit karena harga Tandan Buah Segar (TBS) mengalami kenaikan.
4. Tahap Hasil dan Pembahasan. Penelitian ini menggunakan metode CRISP-DM (Cross Industry Standard Process for Data Mining) dengan menggunakan *tools RapidMiner*
5. Kesimpulan. Tahap ini menarik kesimpulan yang telah di lakukan untuk menentukan bahwa penerapan metode yang digunakan dapat membantu memecahkan masalah yang ada.



Gambar 1. Tahapan Penelitian

4 HASIL DAN PEMBAHASAN

Pembahasan penelitian dengan Model *Cross Industry Standard Process for Data Mining* (CRISP-DM) dengan menerapkan empat fase:

1. Business Understanding (Fase Pemahaman Bisnis)

Dimana pada tahap ini terdapat empat fase yaitu [1] **Determine Business Objective**, berdasarkan pengolahan data penjualan bibit kelapa sawit tahun 2017, 2018, 2019, 2020, 2021 antara lain untuk meningkatkan kualitas bibit kelapa sawit dan juga memberikan harga yang dapat dijangkau oleh para petani kelapa sawit. [2] **Assess Situation**, berdasarkan terjadinya fluktuasi harga pada bibit kelapa sawit kemudian akan dilakukan prediksi harga bibit kelapa sawit dengan menggunakan metode k-nearest neighbor [3] **Determine Data Mining Goals**, yaitu memprediksi harga bibit kelapa sawit dengan menggunakan metode k-nearest neighbor dan mengevaluasi kinerja dari model. [4] **Produce Project Plan**, yaitu pengumpulan data penjualan bibit kelapa sawit PPKS kebun Sungai Lilin tahun 2017, 2018, 2019, 2020, 2021 dengan atribut yang digunakan adalah tanggal, nomor penjualan, stock awal, jenis bibit (Dura Psifera, Dura Yangambi Psifera, Main Nursery, PreNursery), umur, harga satuan, harga total, penambahan bibit, akhir bibit, dan stock akhir.

2. Data Understanding (Fase Pemahaman Data)

Tahapan yang di lakukan [1] **Collect Initial Data**, pengumpulan data awal dimana data didapat dari kantor PPKS kebun sungai lilin merupakan data penjualan bibit kelapa sawit PPKS kebun sungai lilin. Data yang diperoleh adalah data penjualan periode 5 tahun terakhir yaitu 2017, 2018, 2019, 2020, 2021. Jumlah data yang diperoleh sebanyak 2.049 dengan 13 atribut terdapat pada Tabel 1.

Tabel 1. Data Penjualan

Tanggal	Data Penjualan	Stok Awal	DP	DYP	MN	PN	Umur Bibit	Harga Satuan	Total Harga	Penambahan Bibit	Seleksi Bibit/Afkir Bibit	Stok Akhir
24/11/2017	243/BKS/SL/XI/2017	262,226	1	0	250	0	30	32000	8000000	0	0	261,976
24/11/2017	244/BKS/SL/XI/2017	261,976	1	0	250	0	30	32000	8000000	0	0	261,726
24/11/2017	245/BKS/SL/XI/2017	261,726	1	0	250	0	30	32000	8000000	0	0	261,476
25/11/2017	246/BKS/SL/XI/2017	261,476	1	0	250	0	30	32000	8000000	0	0	261,226
25/11/2017	247/BKS/SL/XI/2017	261,226	1	0	250	0	30	32000	8000000	0	0	260,976
25/11/2017	248/BKS/SL/XI/2017	260,976	0	1	250	0	30	33000	8250000	0	0	260,726
25/11/2017	249/BKS/SL/XI/2017	260,726	0	1	250	0	30	33000	8250000	0	0	260,476
25/11/2017	250/BKS/SL/XI/2017	260,426	0	1	250	0	30	33000	8250000	0	0	260,226
25/11/2017	251/BKS/SL/XI/2017	260,226	0	1	250	0	30	33000	8250000	0	0	259,976
25/11/2017	252/BKS/SL/XI/2017	259,976	0	1	250	0	30	33000	8250000	0	0	259,726
25/11/2017	253/BKS/SL/XI/2017	259,226	0	1	250	0	30	33000	8250000	0	0	259,476
20/01/2017	21/BKS/SL/I/2017	231,277	1	0	400	0	21	30000	12000000	0	0	230,877
20/01/2017	22/BKS/SL/I/2017	230,877	1	0	150	0	21	31000	4650000	0	0	230,727
16/11/2019	140/BKS/SL/XI/2019	230,727	1	0	500	0	19	38000	19000000	0	0	230,227
16/11/2019	02/BKS/SL/XI/2019	230,227	1	0	0	75	8	15500	1162500	0	0	230,152
17/11/2019	03/BKS/SL/XI/2019	230,152	1	0	0	2100	8	15500	32550000	0	0	228,052
18/11/2019	04/BKS/SL/XI/2019	228,052	1	0	0	310	8	15500	4805000	0	0	227,742
18/11/2019	141/BKS/SL/XI/2019	227,742	1	0	500	0	19	38000	19000000	0	0	227,242
18/11/2019	142/BKS/SL/XI/2019	227,242	1	0	500	0	19	38000	19000000	0	0	226,742
18/11/2019	143/BKS/SL/XI/2019	226,742	1	0	388	0	19	38000	14744000	0	0	226,354
19/11/2019	144/BKS/SL/XI/2019	226,354	1	0	250	0	19	38000	9500000	0	0	226,104
4/10/2021	393/BKS/SL/X/2021	226,104	1	0	105	0	16	40000	4200000	0	0	225,999
5/10/2021	394/BKS/SL/X/2021	225,999	1	0	0	300	4	16500	4950000	0	0	225,699
6/10/2021	395/BKS/SL/X/2021	225,699	1	0	0	1000	4	16500	16500000	0	0	224,699
6/10/2021	396/BKS/SL/X/2021	225,699	1	0	100	0	16	40000	4000000	0	0	224,599
7/10/2021	397/BKS/SL/X/2021	224,599	1	0	80	0	16	40000	3200000	0	0	224,519
22/07/2021	308/BKS/SL/X/2021	249,704	1	0	250	0	13	28000	7000000	0	0	349,454

[2] *Describe Data*, mendeskripsikan data yang telah didapatkan sebelumnya mengenai atribut dari data yang telah didapat dan memahami data yang akan diolah dalam melakukan prediksi harga bibit kelapa sawit terdapat pada Tabel 2. [3] *Explore Data*, fase ini bertujuan untuk menentukan sumber data yang digunakan dalam penelitian ini adalah data penjualan bibit kelapa sawit tahun 2017-2021 pada bagian administrasi PPKS kebun Sungai Lilin, terdapat pada Tabel 3. [4] *Verify Data Quality*, Fase verifikasi kualitas data akan dilakukan proses pengecekan seperti missing value, noise, data yang tidak konsisten, atau data outlier. Dimana verifikasi kualitas data ini dilakukan dengan tujuan untuk mengetahui seberapa baik kualitas data yang telah didapatkan sebelumnya

Tabel 2. Deskripsi Data

No Atribut	Jenis	Keterangan
1 Nomor Penjualan	Polynomial	Nomor penjualan bibit kelapa sawit
2 Stock Awal	Integer	Stock awal bibit kelapa sawit yang akan di jual
3 DP	Integer	Jenis bibit kelapa sawit Dura Psifera
4 DYP	Integer	Jenis bibit kelapa sawit Dura Yangambi Psifera
5 MN	Integer	Main Nursery, bibit kelapa sawit berumur 10-12 bulan dan sudah siap ditanam di lahan terbuka
6 PN	Integer	PreNursery, bibit kelapa sawit tahap pengembangbiakan kecambah sampai berumur 2-3 bulan
7 Umur Bibit	Integer	Umur bibit kelapa sawit
8 Harga Satuan	Integer	Harga satuan bibit kelapa sawit
9 Total Harga	Integer	Total harga pembelian bibit kelapa sawit
10 Penambahan Bibit	Integer	Penambahan bibit kelapa sawit
11 Seleksi Bibit/Afkir Bibit	Integer	Pemusnahan bibit kelapa sawit yang tidak layak dipasarkan
12 Stock Akhir	Integer	Stock akhir bibit kelapa sawit

3. Data Preparation (Fase Persiapan Data)

Persiapan data dilakukan sebelum melakukan teknik data mining. Persiapan data dilakukan untuk mengubah data mentah menjadi data yang siap digunakan dan di proses menggunakan teknik data mining. Pada fase pengolahan data dilakukan pada data set penjualan bibit kelapa sawit sebanyak 2049 record dengan 13 atribut yaitu: [1] **Select Data**, pemilihan data dilakukan untuk memilih atribut yang akan digunakan pada pembentukan model algoritma klasifikasi. Adapun atribut yang digunakan dalam penentuan prediksi harga bibit kelapa sawit adalah jenis bibit (Dura Psifera, Dura Yangambi Psifera, Main Nursery, Pre Nursery), harga satuan, dan umur bibit. [2] **Clean Data**, memastikan bahwa data yang digunakan sudah benar-benar telah memiliki kualitas yang baik yaitu memiliki missing value diantaranya jenis bibit PreNursery (PN) terdapat 1 missing value, penambahan bibit terdapat 1 missing value, seleksi bibit/afkir bibit terdapat 1 missing value, dan jenis bibit Dura Yangambi Psifera (DYP). [3] **Construct Data**, yaitu membangun atribut turunan dari atribut yang ada. Pada tahapan ini menambahkan satu buah atribut baru yang dijadikan sebagai kelas/label target. Atribut ini berisikan dua buah nilai yaitu “Murah”, “Mahal”.

[4] **Transformation Data**, Pada fase transformasi data dilakukan penyederhanaan data yang ada, dimana atribut yang tidak berpengaruh terhadap prediksi harga akan dihapus/dihilangkan sehingga hanya akan menyisakan beberapa atribut yang akan digunakan dalam perhitungan k-NN.

Tabel 3. *Explore Data*

No Atribut	Keterangan	Eksplorasi Data
1 Nomor Penjualan	Nomor penjualan bibit kelapa sawit tahun 2017, 2018, 2019, 2020, 2021	Jumlah nomor penjualan keseluruhan adalah 2,049
2 Stock Awal	Stock awal bibit kelapa sawit dari tahun 2017, 2018, 2019, 2020, 2021	Jumlah stock awal bibit kelapa sawit 1. Tahun 2017 = 250,319 2. Tahun 2018 = 204,342 3. Tahun 2019 = 288,324 4. Tahun 2020 = 274,961 5. Tahun 2021 = 177,667
3 Jenis Bibit	Jenis bibit kelapa sawit terdiri dari Dura Psifera dan Dura Yangambi Psifera dan dibagi menjadi PreNursery dan Main dengan 2021 Nursery	Jenis bibit kelapa sawit tahun 2017 sampai 2021
4 Umur Bibit	Umur bibit merupakan umur dari kedua jenis bibit kelapa sawit	Umur bibit yang ada adalah mulai dari umur 11 bulan – 20 bulan
5 Harga Satuan	Harga satuan bibit kelapa sawit	Harga bibit kelapa sawit mulai dari
6 Total Harga	Total harga penjualan bibit kelapa sawit	Total harga bibit dihitung dari banyaknya penjualan dan harga satuan bibit kelapa sawit
7 Penambahan Bibit	Penambahan bibit kelapa sawit	Penambahan bibit dilakukan tidak menentu
8 Seleksi Bibit/Afkir Bibit	Seleksi bibit/afkir bibit kelapa sawit	Seleksi bibit dilakukan saat terdapat bibit kelapa sawit yang tidak lolos sertifikasi
9 Stock Akhir	Stock akhir penjualan bibit kelapa sawit	Stock akhir bibit dari tahun 2017 sampai 2021 adalah 448,824

Tabel 4. *Data Transformasi*

No	DP	DYP	MN	PN	Umur	Harga
1	1	0	400	0	21	30000
2	0	1	150	0	21	31000
3	1	0	0	300	3	14500
4	0	1	47	0	21	33000
5	1	0	0	140	5	14500
6	1	0	300	0	15	32000
7	1	0	0	2000	8	15500
8	1	0	0	100	3	16,500
9	0	1	100	0	14	38000
10	1	0	500	0	20	38000
11	1	0	0	100	11	15500
12	1	0	100	0	15	25,500
13	0	1	250	0	16	38000
14	0	1	500	0	16	38000
15	1	0	0	55	12	15500
16	1	0	500	0	17	38000
17	0	1	500	0	17	40000
18	1	0	250	0	13	28,000
19	0	1	350	0	17	40000
20	1	0	0	260	4	16,500

4. Modeling (Fase Pemodelan)

Pada fase pemodelan prediksi harga bibit kelapa sawit ini menggunakan metode k-nearest neighbor. Data yang telah melalui 3 tahap awal proses CRISP-DM akan di olah menggunakan metode k-nearest neighbor dan dibangun menggunakan tools RapidMiner. Data yang digunakan sebanyak 2049 yang terdiri dari tiga kelas yaitu murah, mahal, dan untuk mengetahui pola harga apakah harga bibit kelapa sawit murah, mahal atau berdasarkan variabel-variabel yang ada menggunakan metode k-nearest neighbor

Hasil menunjukkan bahwa akurasi prediksi mencapai 99,21% dengan nilai $k=5$. Hal ini berarti bahwa model KNN mampu memprediksi harga bibit kelapa sawit dengan tingkat akurasi yang tinggi. Analisis lebih lanjut menunjukkan bahwa:

1. Presisi untuk kelas Murah sebesar 98,93%, artinya hampir semua bibit yang diprediksi murah memang benar-benar murah.
2. Presisi untuk kelas Mahal sebesar 99,29%, artinya hampir semua bibit yang diprediksi mahal memang benar-benar mahal.
3. Recall untuk kelas Murah sebesar 97,63%, artinya hampir semua bibit yang murah teridentifikasi dengan benar oleh model.
4. Recall untuk kelas Mahal sebesar 99,68%, artinya hampir semua bibit yang mahal teridentifikasi dengan benar oleh model.
5. Kesimpulannya, metode KNN terbukti efektif dalam memprediksi harga bibit kelapa sawit dengan tingkat akurasi yang tinggi.

5 SIMPULAN

Hasil penelitian menunjukkan bahwa metode *K-Nearest Neighbor* (KNN) dengan pendekatan CRISP-DM mampu memprediksi harga bibit kelapa sawit dengan tingkat akurasi 99,21%, sehingga memiliki beberapa implikasi praktis yang signifikan bagi industri kelapa sawit, terutama dalam aspek perencanaan bisnis, pengelolaan rantai pasok, dan stabilitas harga. Penelitian ini membuktikan bahwa metode KNN dengan CRISP-DM dapat menjadi alat prediksi yang efektif dan andal dalam industri kelapa sawit. Dengan implementasi yang tepat, hasil penelitian ini dapat membantu petani, distributor, investor, dan regulator dalam membuat keputusan yang lebih cerdas, strategis, dan berbasis data, sehingga mendorong pertumbuhan industri kelapa sawit yang lebih stabil dan berkelanjutan. Namun, penelitian ini memiliki beberapa keterbatasan. Dataset yang digunakan mungkin belum cukup representatif untuk menangkap semua kondisi pasar, terutama dalam menghadapi faktor eksternal seperti kebijakan pemerintah, perubahan iklim, atau volatilitas harga global. Selain itu, akurasi yang sangat tinggi perlu dikaji lebih lanjut untuk memastikan tidak terjadi *overfitting*, sehingga model tetap dapat digunakan secara luas dalam berbagai skenario. Ke depan, eksplorasi metode lain seperti *Random Forest*, *Support Vector Machine* (SVM), atau *Neural Networks* dapat dilakukan untuk membandingkan kinerja dengan KNN. Penggunaan teknik *ensemble learning* atau *deep learning* juga dapat menjadi alternatif untuk meningkatkan ketepatan prediksi, terutama jika jumlah data yang tersedia semakin besar dan kompleks.

DAFTAR PUSTAKA

- Setiawan, A., & Sari, R. E. (2018). Penerapan Metode Profile Matching Sebagai Penyortiran Bibit Kelapa Sawit. *Jurnal Informatika Kaputama (JIK)*, 2(2), 71-81.
- Ikhlas, M. (2019). Penerapan metode mfep (multifactor evaluation process) dalam pengambilan keputusan pemilihan bibit kelapa sawit terbaik. *Jurnal Sains dan Teknologi*, 19(1).
- Budiman dkk, (2020). Data mining implementation using naïve bayes algorithm and decision tree J48 in determining concentration selection, *International Journal of Quantitative Research and Modeling* 1 (3), 123-134
- Virdaus, D., & Prasetyaningrum, P. T. (2020). Penerapan Data Mining Untuk Memprediksi Harga Bawang Merah Di Yogyakarta Menggunakan Metode K-Nearest Neighbor. *Journal Of ...*, 84, 1-8.
- Yahya, Y., & Puspita Hidayanti, W. (2020). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada "Lombok Vape On." *Infotek : Jurnal Informatika Dan Teknologi*, 3(2), 104-114.

- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-Sakti)*, 5(2), 697–711.
- Tyo Anggin Virnand, (2021) Penerapan Algoritma K-Nearest Neighbor (K-Nn) Pada Klasifikasi Kualitas Hasil Pengeringan Bunga Cengkih
- Munawwaroh, D. A., & Primandari, A. H. (2022). Implementasi Crisp-Dm Model Menggunakan Metode Decision Tree Dengan Algoritma Cart Untuk Prediksi Lila Ibu Hamil Berpotensi Gizi Kurang. *Delta: Jurnal Ilmiah Pendidikan Matematika*, 10(2), 367-380.
- Suryadi, L., Pratiwi, N. E., Ardhy, F., & Riswanto, P. (2022). Penerapan Data Mining Prediksi Penjualan Mebel Terlaris Menggunakan Metode K-Nearest Neighbor (K-Nn)(Studi Kasus: Toko Zerita Meubel). *JUSIM (Jurnal Sistem Informasi Musirawas)*, 7(2), 174-184.
- Zaien Bin Umar Alaydrus, Arief Andy Soebroto, (2023) Klasifikasi Masa Panen Varietas Unggul Kedelai menggunakan K-Nearest Neighbor
- Yudiana, Y., Agustina, A. Y., & Khofifah, N. (2023). Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan. *Indonesian Journal of Islamic Economics and Business*, 8(1), 1-20
- Kurniawan, I., Rahaningsih, N., & Suprapti, T. (2024). IMPLEMENTASI ALGORITMA REGRESI LINIER DAN K-NEAREST NEIGHBOR UNTUK PREDIKSI HARGA RUMAH. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 1187-1193.
- Lalapa, N. M. (2023). Implementasi Metode Regresi Linear Sederhana Untuk Prediksi Harga Cabai Rawit. *Jurnal Ilmiah Ilmu Komputer Banthayo Lo Komputer*, 2(2), 96-103.
- Lo, A. A. P., & Tjioe, V. J. E. (2024, September). Penerapan Model CRISP-DM untuk Prediksi Penyakit Diabetes Menggunakan Metode K-Nearest Neighbor dan Logistic regression. In *Prosiding Seminar Nasional Universitas Ma Chung (Informatika & Sistem Informasi; Bahasa dan Seni; Farmasi)* (Vol. 4, pp. 48-57).
- Yulianto, K. E., & Christy, V. M. (2024, September). Prediksi Kualitas Air di Jawa Timur Menggunakan Metode CRISP-DM dengan Algoritma K-NN dan Regresi Logistik Multinomial. In *Prosiding Seminar Nasional Universitas Ma Chung (Informatika & Sistem Informasi; Bahasa dan Seni; Farmasi)* (Vol. 4, pp. 16-24).
- Dewi, S. P., Nurwati, N., & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, 3(4), 639-648.
- Bahtiar, R. (2023). Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor. *Jurnal Informatika MULTI*, 1(3), 203-214.