

Prediksi Harga Tandan Buah Segar dengan Algoritma K-Nearest Neighbor

Silvi Joya Arditna Br Bukit*, Rakhmat Kurniawan R.

Fakultas Sains Dan Teknologi, Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ^{1,2}Joyasilvi2001@gmail.com, ²rakhmat.kr@uinsu.ac.id

Email Penulis Korespondensi: Joyasilvi2001@gmail.com

Submitted: 21/09/2023; Accepted: 28/09/2023; Published: 29/09/2023

Abstrak—Kelapa sawit dan produk turunannya merupakan sumber devisa bagi negara ini, karena perlu adanya upaya untuk memelihara dan mengembangkan kesinambungan peningkatan kelapa sawit sebagai sumber daya alam yang potensial. perusahaan melakukan analisis statistik pada faktor penghambat hasil panen bulan sebelumnya dengan nilai koreksi 5% – 12%. Akan tetapi analisis seperti ini masih menghasilkan hasil prediksi yang kurang tepat, hal ini dikarenakan proses perhitungan masih melibatkan teknik perkiraan dari pengalaman pribadi, melihat dari pola produksi di waktu sebelumnya dan faktor penentu lain seperti luas lahan, jumlah pokok dan umur tanam. Akibatnya target prediksi sering kali mengalami kesalahan dan hasil produksi berlebih atau kurang dari target. Oleh sebab itu diperlukannya perhitungan prediksi yang lebih baik dalam menentukan target produksi kelapa sawit. Prediksi yang akurat dapat membantu perusahaan dalam mengambil keputusan untuk meningkatkan hasil produksi. Untuk melakukan peramalan diperlukan penerapan Algoritma K-Nearest Neighbor dapat digunakan untuk memprediksi harga sawit pada waktu yang akan datang. Berdasarkan hasil perhitungan data mining menggunakan harga TBS kelapa sawit dari tahun 2018 sampai tahun 2023 (Mei 2023) diperoleh kesimpulan bahwa prediksi harga TBS kelapa sawit pada bulan ke-67 (Juli 2023) memiliki tingkat akurasi sebesar 10,667 dengan $k=3$ dan 19,200 dengan $k=5$.

Kata Kunci: Prediksi; Harga; TBS; Data Mining; K-Nearest Neighbor

Abstract—Palm oil and its derivative products are a source of foreign exchange for this country, because efforts are needed to maintain and develop the sustainability of palm oil as a potential natural resource. The company carries out statistical analysis on the factors inhibiting the previous month's harvest with a correction value of 5% – 12%. However, this kind of analysis still produces inaccurate prediction results, this is because the calculation process still involves estimation techniques from personal experience, looking at previous production patterns and other determining factors such as land area, principal amount and planting age. As a result, prediction targets often experience errors and production results are excessive or less than the target. Therefore, better predictive calculations are needed in determining palm oil production targets. Accurate predictions can help companies make decisions to increase production output. To carry out forecasting, it is necessary to apply the K-Nearest Neighbor Algorithm which can be used to predict palm oil prices in the future. Based on the results of data mining calculations using palm oil FFB prices from 2018 to 2023 (May 2023), it was concluded that the prediction of palm oil FFB prices in the 67th month (July 2023) had an accuracy level of 10,667 with $k=3$ and 19,200 with $k=5$.

Keywords: Predictions; Prices; TBS; Data Mining; K-Nearest Neighbor

1. PENDAHULUAN

Sektor pertanian di Indonesia dibedakan menjadi tiga jenis yaitu perkebunan, sawah dan ladang. Dari ketiga jenis sektor pertanian, sektor perkebunan yang lebih banyak diminati dikarenakan pertanian jenis perkebunancenderung memiliki nilai jual yang tinggi, pembudidayaan dalam skala besar, serta daya tariknya yang terus meningkat. Sektor tanaman perkebunan di Indonesia banyak didominasi oleh tanaman kelapa sawit, kakao, karet, tebu dan kopi. Dari kelima tanaman ini kelapa sawit yang paling menguntungkan. Dilansir dari <https://www.ajnn.net/news/harga-tbs-kelapa-sawit-naik-di-aceh-barat/index.html> Harga beli Tandan Buah Segar (TBS) di Pabrik Kelapa Sawit (PKS) Aceh Barat mengalami kenaikan pada pekan pertama Juli 2023. Kenaikan ini dipicu oleh nilai jual Crude Palm Oil (CPO) dunia yang bergerak naik. Namun, untuk pekan selanjutnya nilai serupa belum dapat dipastikan akan bertahan atau tidak. Community Development Officer perusahaan PKS PT Karya Tanah Subur, Azra Husaini mengatakan, untuk harga beli sawit pada pekan pertama Juli naik menjadi Rp50 dibandingkan akhir Juni atau minggulu yang bertahan pada harga Rp1.750. "Harga sawit minggu ini mengalami kenaikan dibandingkan minggu lalu. Pekan lalu Rp1.750 sekarang Rp1.800 atau naik Rp50. Kenaikan ini dipicu naiknya harga CPO," Azra Husaini, Sabtu, 8 Juli 2023.

Dalam situs <https://www.suara.com/bisnis/2022/01/26/143508/penyumbang-terbesar-devisa-negara-indonesia-adalah-kelapa-sawit-berapa-nominalnya> diinformasikan bahwa tidak semua alat tukar atau mata uang bisa menjadi devisa suatu negara. Namun, sejauh ini kelapa sawit menyumbang devisanegara terbesar sepanjang 2020 yakni sebesar 25,60 miliar dolar Amerika atau sekitar Rp358 triliun. Dengan besar devisa itu, industri sawit juga telah membuat neraca perdagangan Indonesia mengalami surplus 21,70 miliar dolar pada 2020. Kelapa sawit bahkan jadi penyumbang devisa negara terbesar dalam 20 tahun.

Kelapa sawit banyak ditanam di wilayah Jawa, Sumatra, Nusa Tenggara, Kalimantan, Sulawesi, Maluku dan Papua. Namun perkebunan kelapa sawit paling luas berada di wilayah Sumatra dan Kalimantan. Hal ini disebabkan di wilayah Indonesia Timur belum banyak diminati oleh para investor dikarenakan permasalahan infrastruktur dan keamanan. Selain itu diperlukan biaya yang lebih besar karena perlunya pembukaan lahan, proses produksi dan pemasaran hasil. Salah satu perusahaan kelapa sawit yang ada di wilayah Sumatra adalah

PT.Salapian Indo Sawit yang beralamat di Desa Ujung Teran Dusun Kec.Salapian Kabupaten Langkat. Dalam sebuah wawancara dengan KTU perusahaan yang tergolong dalam jenis perkebunan besar swasta (PBS) ini mulai melakukan proses penanaman kelapa sawit pada tahun hingga sekarang. Hasil produksi kelapa sawit sangat berpengaruh pada jumlah target permintaan olahan, target hasil olahan, anggaran upah pekerja panen, dan anggaran produksi. Oleh sebab itu dalam menargetkan jumlah produksi berikutnya perusahaan melakukan perhitungan produksi terlebih dahulu atau melakukan prediksi, tujuannya untuk mengetahui kebijakan apa saja yang akan diambil.

Prediksi atau forecasting adalah suatu upaya untuk memperoleh gambaran mengenai apa yang akan terjadi di masa yang akan datang. Untuk melakukan peramalan diperlukan metode tertentu dan metode mana yang digunakan tergantung dari data dan informasi yang akan diramal serta tujuan yang hendak dicapai. Dalam hal ini gambaran yang didapat tersebut akan menjadi acuan untuk membuat suatu keputusan. Pada kondisi yang tidak menentu sulit bagi kita untuk menentukan suatu perencanaan yang efektif. Peramalan dapat membantu para pemimpin untuk mengurangi ketidakpastian dalam melakukan perencanaan[1].

Penerapan Algoritma K-Nearest Neighbor dapat digunakan untuk memprediksi harga sawit pada waktu yang akan datang. Ketepatan algoritma KNN sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi[2]. Algoritma KNN memiliki beberapa kelebihan yaitu ketangguhan terhadap training data yang memiliki banyak noise dan efektif apabila training data-nya besar[3]. Sedangkan, kelemahan KNN adalah KNN perlu menentukan nilai dari parameter K (jumlah dari tetangga terdekat), training berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan dan atribut mana yang harus digunakan untuk mendapatkan hasil terbaik, dan biaya komputasi cukup tinggi karena diperlukan perhitungan jarak dari tiap query instance pada keseluruhan training sample[4]. Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan training sample[5]. Classifier tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik query, akan ditemukan sejumlah K obyek atau (titik training) yang paling dekat dengan titik query. Klasifikasi menggunakan voting terbanyak diantara klasifikasi dari K obyek[6].

Allah SWT berfirman dalam QS. Luqman ayat 34 tentang prediksi :

نَّ اللَّهُ عِنْدَهُ عِلْمُ السَّاعَةِ وَيُنزِلُ الْغَيْثَ وَيَعْلَمُ مَا فِي الْأَرْحَامِ وَمَا تَدْرِي نَفْسٌ مَّاذَا تَكْسِبُ غَدًا وَمَا تَدْرِي نَفْسٌ بِأَيِّ أَرْضٍ تَمُوتُ إِنَّ اللَّهَ عَلِيمٌ خَبِيرٌ

Artinya : Sesungguhnya hanya di sisi Allah ilmu tentang hari Kiamat; dan Dia yang menurunkan hujan, dan mengetahui apa yang ada dalam rahim. Dan tidak ada seorang pun yang dapat mengetahui (dengan pasti) apa yang akan dikerjakannya besok. Dan tidak ada seorang pun yang dapat mengetahui di bumi mana dia akan mati. Sungguh, Allah Maha Mengetahui, Maha Mengetahui.

Dalam penelitian ini, penulis mengacu kepada penelitian lain sebagai referensi, salah satu penelitian yang sejenis yang dilakukan oleh Resti Hutami dkk(2016) membahas tentang metode K-Nearest Neighbor yang diimplementasikan untuk prediksi penjualan furniture. Penelitian tersebut melakukan prediksi penjualan furniture dengan teknologi data mining untuk menganalisis volume data penjualan. Metode K-Nearest Neighbor digunakan karena memiliki akurasi yang tinggi dengan rasio kesalahan yang minim. Hasil dari prediksi tersebut menunjukkan bahwa metode K- Nearest Neighbor berhasil diimplementasikan untuk menyelesaikan kasus prediksi penjualan dengan tingkat eror sebesar 6% dan akurasi sebesar 94%[7]. Penelitian selanjutnya yang dilakukan oleh Andi Gita Novianti dkk, dengan judul “Penerapan Algoritma K-Nearest Neighbor (KNN) Untuk Prediksi Waktu Kelulusan Mahasiswa”. Pada penelitian ini menerapkan Algoritma K-Nearest Neighbor (KNN) dan Fungsi Similarity untuk menghitung kemiripan data dalam sebuah perangkat lunak yang dapat memberikan prediksi waktu kelulusan mahasiswa. Hasil pengujian menggunakan aplikasi prediksi waktu kelulusan dengan 7 (tujuh) kriteria yaitu IPS1-IPS4, jumlah SKS lulus sampai semester 4, jurusan SLTA, program studi, asal suku, penghasilan orang tua dan jenis kelamin di dapat akurasi untuk Program Studi Teknik Informatika S1 sebesar 84% sedangkan Program Studi Sistem Informasi S1 sebesar 87% [8]. Penelitian selanjutnya yang dilakukan oleh Risman Hendri dkk membahas tentang penentuan calon penerima beasiswa dengan menggunakan metode K-Nearest Neighbor. Dari pengujian yang dilakukan terhadap 22 data sampel yang dijadikan acuan dalam perhitungan K-Nearest Neighbor dalam menghasilkan keputusan diperoleh nilai keakuratan sebesar 90,90%[9]. Penelitian selanjutnya yang dilakukan oleh Ferry Hermawan & Halim Agung membahas tentang memprediksi penjualan berdasarkan kategori barang. Hasil dari penelitian ini adalah K-Nearest Neighbor dapat memprediksi penjualan di tahun 2015 berdasarkan data penjualan barang dari tahun 2012-2014 dengan menggunakan Euclidean Distance, dengan tingkat keberhasilan metode 58,33% pada nilai toleransi error 10% dan rata-rata keakuratan prediksi 88,54% yang tergolong memiliki kinerja bagus dan memprediksi penjualan berdasarkan kategori barang dengan tingkat keberhasilan algoritma 70% pada nilai toleransi error 10% dan rata-rata keakuratan prediksi 85,91% yang tergolong memiliki kinerja bagus[10]. Penelitian selanjutnya yang dilakukannya oleh Tundo membahas tentang bagaimana metode fuzzy tsukamoto melakukan prediksi produksi minyak kelapa sawit. Berdasarkan perhitungan tersebut didapatkan nilai AFER dengan metode Tsukamoto menggunakan rule yang terbentuk dari REPTree sebesar 23,17% sehingga akurasi kebenaran yang dihasilkan sebesar 76,83%[11]. Penelitian selanjutnya Amalia Aldina Thoha dan Sudradjat (2019) dalam penelitian yang berjudul “Pengolahan panen kelapa sawit (Elaeis Guineensis Jacq) di kebun Adolina, Sumatera Utara”. Tujuan penelitian ini adalah untuk meningkatkan pengetahuan tentang budidaya tanaman kelapa sawit dan secara khusus 8 menganalisis faktor-faktor yang

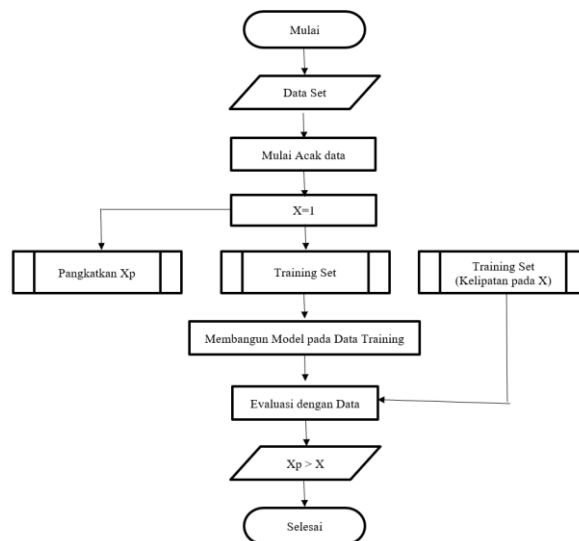
mempengaruhi produksi kelapa sawit yaitu pemanenan. Teknik pengumpulan data dengan metode observasi, wawancara, dan dokumentasi. Jenis data yang digunakan yaitu menggunakan data primer dan data sekunder. Hasil penelitian adalah tingginya kehilangan panen yang mempengaruhi produksi tanaman kelapa sawit. Angka kerapatan panen yang didapat 61,36% bahwa kebutuhan tenaga kerja panen sekitar 2/3 ha per orang. Kapasitas panen adalah 145,75 TBS dan kualitas mutu buah yang dipanen mencapai hampir 100% di atas setandar yang ditetapkan perusahaan [12].

Berdasarkan hasil wawancara dengan kltu, selama ini dalam melakukan perhitungan target, perusahaan melakukan analisis statistik pada faktor penghambat hasil panen bulan sebelumnya dengan nilai koreksi 5% – 12%. Akan tetapi analisis seperti ini masih menghasilkan hasil prediksi yang kurang tepat, hal ini dikarenakan proses perhitungan masih melibatkan teknik perkiraan dari pengalaman pribadi, melihat dari pola produksi di waktu sebelumnya dan faktor penentu lain seperti luas lahan, jumlah pokok dan umur tanam. Selain itu perusahaan juga melihat dari kondisi pasar. Teknik prediksi seperti ini masih dilakukan karena pihak perusahaan merasa teknik ini mudah dilakukan dibandingkan dengan menggunakan tools peramalan atau teori peramalan. Akibatnya target prediksi sering kali mengalami kesalahan dan hasil produksi berlebih atau kurang dari target. Oleh sebab itu diperlukannya perhitungan prediksi yang lebih baik dalam menentukan target produksi kelapa sawit. Prediksi yang akurat dapat membantu perusahaan dalam mengambil keputusan untuk meningkatkan hasil Produksi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada penelitian diperlukan tahapan-tahapan yang bermanfaat dalam menyelesaikan permasalahan sehingga tujuan dalam penelitian tercapai. Adapun kerangka kerja dalam penelitian ini dapat dilihat pada gambar berikut:



Gambar 1. Flowchart Perhitungan Algoritma K-Nearest Neighbor

2.2 Prediksi

Prediksi merupakan seni dalam melakukan perhitungan untuk menentukan kejadian dimasa depan dengan menggunakan data-data di masa lalu. Prediksi pada implementasinya membentuk sebuah aturan dengan mengimplementasikan metode perhitungan yang kemudian dilakukan analisis [13]. Teknik prediksi dimulai dengan melakukan eksplorasi data guna untuk mencari pola data pada suatu kurun waktu. Prediksi/forecasting adalah menentukan jumlah kebutuhan bulan mendatang terkait dengan dukungan data historis (historical data) atau serangkaian waktu/periode yang dianalisis sehingga dapat diperhitungkan untuk memprediksi jumlah kebutuhan pada bulan mendatang. Prediksi juga dapat digunakan dalam pengklasifikasian, tidak hanya untuk memprediksi time series, karena sifatnya yang bisa menghasilkan class berdasarkan atribut yang ada.

2.3 Algoritma K-Nearest Neighbor (KNN)

Algoritma k-Nearest Neighbor merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan jumlah k yang telah ditentukan dan mengklasifikasikan ke dalam kelas baru [14]. K-NN termasuk algoritma supervised learning dimana hasil dari query instance yang baru diklasifikasi berdasarkan mayoritas dari kategori pada k-NN. Kelas yang paling banyak muncul itu yang akan menjadi kelas hasil klasifikasi [15]. Tujuan dari algoritma ini adalah mengklasifikasikan objek baru berdasarkan atribut dan training sample. Algoritma k-Nearest Neighbor menggunakan klasifikasi ketetanggaan (neighbor) sebagai nilai prediksi dari query instance yang baru [16]–[18]. Teknik pencarian tetangga

terdekat pada umumnya dilakukan dengan menggunakan formula jarak euclidean. Langkah-langkah untuk menghitung algoritma k-NN:

- a. Menentukan nilai k.
- b. Menghitung kuadrat jarak euclid (query instance) masing-masing objekterhadap training data yang diberikan.
- c. Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil.
- d. Mengumpulkan label class Y (klasifikasi Nearest Neighbor).
- e. Dengan menggunakan kategori Nearest Neighbor yang paling mayoritas maka dapat diprediksikan nilai query instance yang telah dihitung.

3. HASIL DAN PEMBAHASAN

3.1 Data Transformation

Pada penelitian ini akan digunakan analisis data terhadap harga TBS kelapa sawit. Tahap pertama adalah melakukan penumpulan data. Data yang digunakan dalam penelitian ini adalah data harga TBS kelapa sawit berdasarkan 5 tahun terakhir, yaitu pada tahun 2018 hingga tahun 2022. Kemudian data tersebut diseleksi dan akan digunakan untuk diolah untuk memprediksi harga TBS di tahun selanjutnya. Pada tahap transformation, proses pembentukan data training berdasarkan data yang ada, data harus di seleksi terlebih dahulu untuk menentukan atribut mana yang dapat mempengaruhi kenaikan harga TBS. Adapun data training digunakan adalah seperti yang ditunjukkan oleh tabel 1.

Tabel 1. Data Training Harga TBS Kelapa Sawit

2117	2145	2118	2300	2228	2157	2256	2432	2530	2600	2890	3200	3250	3330	3125
2117	2145	2118	2300	2228	2157	2256	2432	2530	2600	2890	3200	3250	3330	3125
2117	2145	2118	2300	2228	2157	2256	2432	2530	2600	2890	3200	3250	3330	3125
2117	2145	2118	2300	2228	2157	2256	2432	2530	2600	2890	3200	3250	3330	3125
2117	2145	2118	2300	2228	2157	2256	2432	2530	2600	2890	3200	3250	3330	3125
2117	2145	2118	2300	2228	2157	2256	2432	2530	2600	2890	3200	3250	3330	3125
2117	2145	2118	2300	2228	2157	2256	2432	2530	2600	2890	3200	3250	3330	3125
2117	2145	2118	2300	2228	2157	2256	2432	2530	2600	2890	3200	3250	3330	3125
2117	2100	2118	2300	2228	2157	2256	2432	2530	2600	2890	3200	3250	3330	3200
2117	2100	2118	2300	2228	2157	2256	2432	2530	2600	2890	3200	3250	3330	3200
2117	2100	2118	2102	2228	2157	2256	2432	2530	2600	2890	3174	3250	3330	3200
2117	2100	2118	2102	2228	2157	2256	2432	2513	2600	2890	3174	3250	3220	3200
2117	2100	2118	2102	2228	2157	2256	2432	2513	2600	2991	3174	3250	3220	3200
2117	2100	2118	2102	2228	2157	2256	2432	2513	2559	2991	3174	3250	3220	3200
2117	2100	2118	2150	2228	2157	2256	2432	2513	2559	2991	3174	3250	3220	3200
2001	2100	2118	2150	2228	2143	2256	2432	2519	2559	2991	3174	3230	3220	3200
2001	2100	2118	2150	2228	2143	2256	2432	2519	2559	2991	3174	3230	3220	3200
2001	2118	2118	2150	2228	2143	2173	2432	2519	2559	2991	3174	3230	3220	3200
2001	2118	2118	2150	2228	2143	2173	2432	2519	2559	2991	3174	3230	3220	3200
2001	2118	2118	2150	2228	2143	2173	2432	2519	2559	2991	3174	3230	3220	3200
2001	2118	2102	2150	2313	2143	2173	2415	2519	2559	2991	3174	3230	3250	3200
2001	2118	2102	2150	2313	2143	2173	2415	2519	2742	2991	3100	3230	3250	3210
2001	2118	2102	2150	2313	2143	2173	2415	2519	2742	2991	3100	3230	3250	3210
2001	2118	2102	2150	2313	2143	2173	2415	2519	2742	2991	3100	3050	3250	3210
3450	3425	3220	3150	2950	2875	2800	2800	2700	2230	2680	2690	2650	2540	2400
3450	3425	3220	3150	2950	2875	2800	2800	2700	2274	2680	2690	2650	2540	2400
3450	3500	3220	3150	2950	2875	2800	2800	2700	2274	2680	2690	2650	2540	2400
3500	3500	3220	3150	2950	2850	2800	2800	2775	2274	2680	2720	2650	2540	2400
3500	3500	3220	3150	2950	2850	2800	2800	2775	2274	2680	2720	2650	2540	2400
3500	3500	3220	3150	2950	2850	2800	2800	2775	2274	2630	2720	2760	2540	2480
3500	3500	3220	3150	2950	2850	2800	2800	2775	2274	2630	2720	2760	2540	2480
3500	3500	3220	3150	2950	2850	2800	2800	2775	2274	2630	2720	2760	2530	2480
3500	3500	3220	3000	2950	2850	2800	2800	2775	2274	2630	2720	2760	2530	2480
3500	3500	3220	3000	2950	2850	2800	2800	2775	2285	2630	2720	2760	2530	2480
3500	3500	3210	3000	2950	2850	2745	2800	2775	2285	2630	2720	2760	2530	2480
3500	3500	3210	3000	3900	2850	2745	2800	2775	2285	2630	2720	2760	2530	2480
3500	3500	3210	3000	3900	2850	2745	2825	2775	2285	2630	2720	2760	2530	2480
3500	3500	3210	3000	3900	2850	2745	2825	2775	2285	2630	2720	2760	2530	2429
3400	3500	3210	3000	3900	2850	2745	2825	2775	2285	2630	2650	2760	2530	2429
3400	3500	3210	3000	3900	2850	2745	2825	2775	2285	2630	2650	2760	2510	2429
3400	3500	3210	3000	3900	2825	2745	2825	2775	2285	2700	2650	2760	2510	2429
3400	3500	3210	3000	3900	2825	2745	2825	2700	2438	2700	2650	2760	2510	2429

2400	2530	2560	2410	2400	2200	2001	1998	2045	2102	2228	2332	1863	2411	2513
2400	2530	2560	2410	2400	2200	2001	1998	2045	2102	2228	2332	1863	2411	2513
2400	2530	2560	2410	2400	2200	2001	1998	2045	2102	2228	2332	1863	2411	2513
2400	2530	2560	2410	2400	2200	2001	1998	2045	2102	2228	2332	1863	2411	2513
2400	2530	2560	2410	2400	2200	2001	1998	2045	2102	2228	2332	1863	2411	2513
2400	2530	2560	2410	2400	2200	2001	1998	2045	2102	2228	2332	1863	2411	2513
2400	2530	2560	2410	2400	2200	2001	1998	2118	2102	2228	2332	1863	2411	2513
2400	2450	2560	2410	2400	2200	2001	1998	2118	2102	2228	2332	1863	2411	2513
2400	2450	2430	2410	2400	2180	2001	1998	2118	2102	2228	2332	1863	2411	2513
2400	2450	2430	2410	2400	2180	2001	1998	2118	2102	2228	2332	1863	2411	2513
2400	2450	2430	2410	2375	2180	2001	1998	2118	2102	2228	2332	1863	2411	2513
2480	2450	2430	2430	2375	2180	2001	1998	2118	2102	2228	2332	1863	2411	2513
2480	2450	2430	2430	2375	2180	2001	1998	2118	2102	2228	2332	1863	2411	2513
2480	2450	2430	2430	2375	2180	2001	1998	2102	2125	2228	2332	1863	2411	2513
2480	2450	2430	2430	2375	2180	2001	1998	2102	2125	2313	2146	1863	2415	2639
2480	2450	2430	2430	2375	2180	2117	1990	2102	2125	2313	2146	2173	2415	2639
2480	2450	2430	2430	2375	2180	2117	1990	2102	2125	2313	2146	2173	2415	2639
2480	2450	2430	2430	2375	2180	2117	1990	2102	2125	2313	2146	2173	2415	2639
2429	2450	2430	2430	2300	2170	2117	1990	2102	2125	2313	2146	2173	2415	2639
2429	2500	2430	2430	2300	2170	2117	1990	2102	2125	2313	2146	2173	2415	2639
2429	2500	2430	2430	2300	2170	2117	1990	2102	2125	2313	2146	2173	2415	2639
2429	2500	2460	2430	2300	2170	2117	1990	2102	2125	2313	2146	2173	2415	2639
2429	2500	2460	2430	2300	2170	2117	1990	2102	2125	2313	2146	2173	2415	2639

2559	2991	3174	2840	2860	3300	3345	2160	2220	1400	1800	2100	1930	2310	2290
2559	2991	3174	2840	2860	3380	3345	2160	2190	1350	1750	2100	1930	2280	2290
2559	2991	3174	2840	2920	3380	3345	2160	2160	1300	1690	2080	1950	2280	2290
2559	2991	3174	2840	2970	3380	3345	2260	2010	1300	1770	2060	1950	2280	2290
2559	2991	3174	2840	3070	3380	3345	2290	2010	1300	1800	2080	1985	2330	2290
2559	2991	3174	2920	3070	3380	3345	2290	2070	1200	1830	2080	2060	2330	2290
2559	2991	3174	2920	3070	3350	3380	2240	2110	1100	1830	2100	2060	2330	2270
2559	2991	3174	2920	3080	3350	3380	2240	2110	1100	1830	2025	2060	2330	2290
2559	2991	3174	2970	3080	3350	3380	2240	2110	1150	1830	2055	2060	2330	2290
2559	2991	3174	3020	3050	3350	3380	2240	2080	1150	1930	2085	2090	2330	2290
2559	2991	3174	3020	3100	3350	3380	2240	2080	1150	1930	2105	2160	2280	2290
2559	2991	3174	3070	3100	3350	3380	2240	2080	1230	1950	2105	2160	2280	2290
2559	2991	3174	3070	3085	3350	3430	2240	2080	1260	2050	2090	2160	2280	2240
2559	2991	3174	3070	3070	3350	3430	2290	2000	1260	2050	2090	2160	2280	2240
2559	2991	3174	3070	3070	3250	3430	2320	2000	1350	2050	2120	2210	2250	2270
2559	2991	3174	3070	3150	3250	3430	2320	1970	1380	2050	2120	2210	2250	2240
2742	3126	3100	3170	3150	3250	3410	2320	1920	1360	2050	2120	2210	2190	2240
2742	3126	3100	3170	3150	3250	3410	2320	1960	1400	2000	2120	2210	2190	2240
2742	3126	3100	3170	3150	3250	3430	2420	1960	1480	1950	2120	2270	2190	2240
2742	3126	3100	3200	3150	3320	3430	2420	1960	1480	1950	2120	2300	2190	2240
2742	3126	3100	3200	3130	3320	3430	2420	1610	1480	1920	2120	2320	2190	2220
2742	3126	3100	3200	3180	3290	3380	2380	1510	1480	1920	2120	2320	2190	2220
2742	3126	3100	3200	3200	3210	3340	2380	1430	1530	2020	2160	2320	2190	2220
2742	3126	3100	3180	3300	3210	3340	2380	1350	1530	2050	2160	2320	2220	2220
2742	3126	3100	3180	3300	3230	3340	2360	1300	1550	2130	2060	2320	2220	2220

2282	2230	2483	2377	2142	1959	2051
2282	2230	2483	2377	2142	1959	2051
2282	2230	2483	2377	2142	1959	2051
2282	2230	2483	2377	2142	1959	2051
2282	2230	2483	2377	2142	1959	2051
2282	2230	2483	2377	2142	1959	2051
2282	2230	2483	2377	2142	1959	2051
2282	2230	2483	2377	2033	1959	2051
2282	2230	2538	2442	2033	1931	2100
2282	2274	2538	2442	2033	1931	2100
2282	2274	2538	2442	2033	1931	2100
2282	2274	2538	2442	2033	1931	2100
2282	2274	2538	2442	2033	1931	2100
2282	2274	2538	2442	2033	1931	2100
2282	2274	2538	2442	2088	1931	2100

2282	2274	2538	2442	2088	1931	2100
2282	2274	2525	2318	2088	1936	2100
2258	2396	2525	2318	2088	1936	2100
2258	2396	2525	2318	2088	1936	2100
2258	2396	2525	2318	2088	1936	2100
2258	2396	2525	2318	2088	1936	2100
2258	2396	2525	2318	2016	1936	2100
2258	2396	2525	2256	2016	1936	2100
2258	2396	2525	2256	2016	1936	2100
2258	2396	2478	2256	2016	1936	2132
2258	2438	2478	2256	2016	2018	2132

Data training yang digunakan pada penelitian ini adalah diambil dari bulan januari 2018 pada data ke 25 sampai dengan Juni 2023 pada data ke 25. Sedangkan untuk data testing digunakan untuk data testing (uji) yang digunakan untuk prediksi untuk bulan Juni 2023 dengan memasukkan data bulan januari 2018 pada data ke 26-31 sebelumnya sampai dengan Juni 2023 pada data ke26-31.

Tabel 2. Data Testing

2001	2118	2102	2122	2313	2147	2173	2415	2539	2742	3126	3100	3050	3250	3210
2001	2118	2102	2122	2313	2147	2173	2415	2539	2742	3126	3000	3050	3250	3210
2001	2118	2102	2122	2313	2147	2173	2415	2539	2742	3126	3000	3050	3250	3210
2001	2118	2102	2122	2313	2147	2173	2415	2539	2742	3126	3000	3050	3250	3210
2001	2118	2102	2122	2313	2147	2173	2415	2539	2742	3126	3000	3050	3250	3210
3400	3400	3500	3200	3000	3900	2825	2745	2825	2700	2438	2700	2650	2720	2500
3400	3400	3500	3200	3000	3900	2825	2745	2825	2700	2438	2700	2650	2720	2500
3400	3400	3500	3200	3000	3900	2825	2745	2825	2700	2438	2700	2650	2720	2500
3400	3400	3500	3300	2130	3900	2825	2745	2825	2700	2438	2700	2650	2720	2500
3400	3400	3950	3300	2130	3900	2825	2745	2220	2700	2438	2700	2650	2700	2500
3400	2320	1400	3300	2100	3900	2825	2745	2220	2700	2438	2700	2650	2700	2500
2500	2500	2460	2430	2300	2170	2117	1990	2102	2122	2313	2147	2173	2415	2539
2500	2500	2460	2430	2300	2170	2117	1990	2102	2122	2313	2147	2173	2415	2539
2500	2500	2460	2500	2300	2170	2117	1990	2102	2122	2313	2147	2173	2415	2539
2500	2500	2460	2500	2300	2170	2117	1990	2102	2122	2313	2147	2173	2415	2539
2500	2500	2460	2500	2300	2170	2117	1990	2102	2122	2313	2147	2173	2415	2539
2500	2000	2460	2500	2300	2170	2117	1990	2102	2122	2313	2147	2173	2415	2539
2742	3126	3100	3160	3300	3260	2630	2320	1300	1590	2130	2060	2320	2220	2220
2742	3126	3100	2760	3300	3260	2410	2320	1300	1660	2130	2060	2320	2220	2220
2742	3126	3100	2760	3300	3260	2460	2320	1330	1660	2130	2030	2320	2220	2220
2742	3126	3100	2860	3300	3290	2160	2320	1400	1760	2130	1840	2310	2220	2220
2742	3126	3100	2860	3300	3310	2160	2320	1400	1800	2130	1840	2310	2220	2220
2742	3126	3100	2860	3300	3345	2160	2320	1400	1800	2100	1840	2310	2220	2220
2258	2438	2478	2256	2016	2018	2132								
2258	2438	2478	2256	2016	2018	2132								
2258	2438	2478	2256	2016	2018	2132								
2258	2438	2478	2256	2016	2018	2132								
2258	2438	2478	2256	2016	2018	2132								
2258	2438	2478	2256	2016	2018	2132								

3.2 K-Nearest Neighbor

Berdasarkan dari tahapan data mining untuk algoritma K-nearest Neighbor, adapun langkah-langkah K-nearest Neighbor sebagai berikut :

- a. Penentuan nilai k. Penentuan nilai k yang digunakan tidak memiliki aturan yang baku, sehingga pada penelitian ini nilai k yang digunakan adalah 3.
- b. Hitung jarak data training dan data uji (testing) yang ada pada tahap transformation dengan menggunakan perhitungan Euclidean Distance sebagai berikut :
 Penentuan distance (jarak) data training dan data testing

$$dist = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + \dots} \tag{1}$$

Sehingga didapatkan jarak sebagai berikut pada tabel 3.

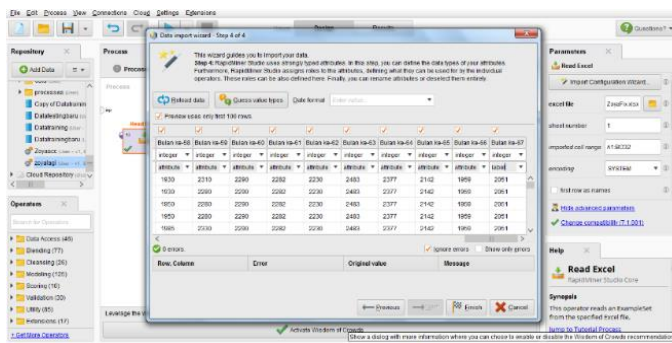
Tabel 3. Distance Data Training dan Data Testing

Distance
2820,38
2842,87
2846,16
2885,98
2972,48
2981,08
2936,92
2957,39
2983,55
2763,37
2579,1
2579,17
2610,95
2618,72
2638,21
2645,32

- c. Pengurutan data hasil perhitungan. Jarak yang telah didapatkan kemudian diurutkan dari yang paling dekat jaraknya sampai dengan yang paling jauh (ascending) adalah 2579,1; 2579,17; 2610,95 ; 2818,72; 2638,21; 2645,32; 2763,37; 2820,38; 2842,87; 2846,16; 2885,98; 2936,92; 2972,48; 2981,08; 2983,55; 2957,39. Menentukan kelompok hasil uji berdasarkan label mayoritas dari k-nearest terdekat. Karena nilai k=3 maka diambil 3 jarak terkecil, yaitu jarak yang dipilih hanya 3, yaitu 2610,95 ; 2818,72; dan 2638,21.
- d. Langkah selanjutnya dengan menggunakan kategori Nearest Neighbor yang paling mayoritas maka dapat diprediksikan harga TBS kelapa sawit periode berikutnya menggunakan Matlab Rapid Miner 10.0.

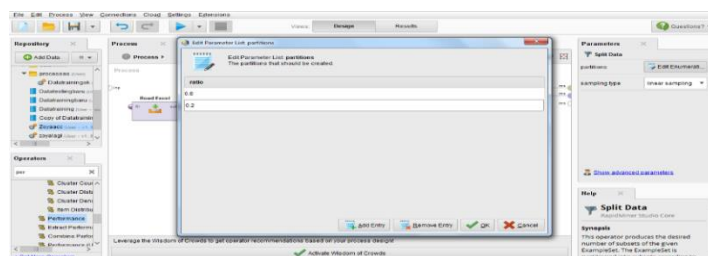
3.3 Implementasi Rapid Miner

Rapid miner merupakan salah satu software data mining pengolahan data set untuk mencari pola data sesuai dengan tujuan dari pengolahan data tersebut. Dilakukan adalah drag and drop operator read excel kemudian lakukan importing tabel microsoft excel ke dalam proses untuk membaca data training dalam format excel, karena yang akan di prediksi pada bulan ke-67 atau bulan Juli 2023 maka pada atribut bulan ke-67 diganti menjadi label.



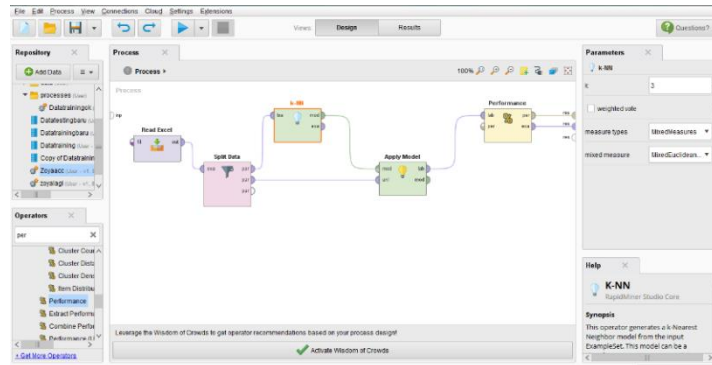
Gambar 1. Importing data Training pada Operator Read Excel

Pada gambar 2., data dimasukkan ke dalam sistem berdasarkan harga TBS setiap bulannya, data harga yang asli dijadikan sebagai integer, sedangkan data yang akan diprediksi dijadikan label. Selanjutnya drag and drop operator Split Data untuk membagi data testing dan data training. Pada Split Data, masukkan entry data 80% atau 0,8 sebagai data training, dan 20% atau 0,2 sebagai data testing seperti pada gambar 3.



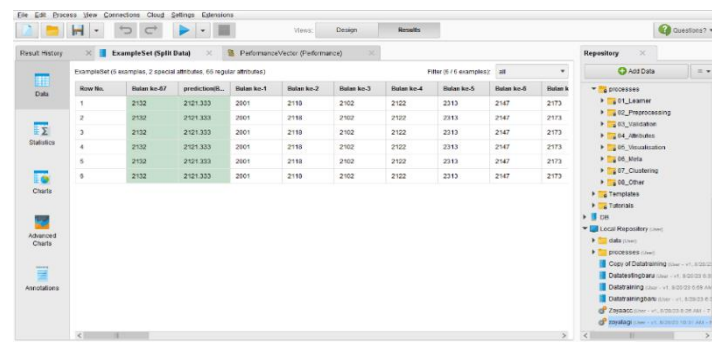
Gambar 2. Tahap Split Data

Selanjutnya, tambahkan operator KNN, K-NN dengan memasukkan nilai k=3 seperti pada gambar 4.



Gambar 3. Susunan Operator Algoritma KNN dengan K=3

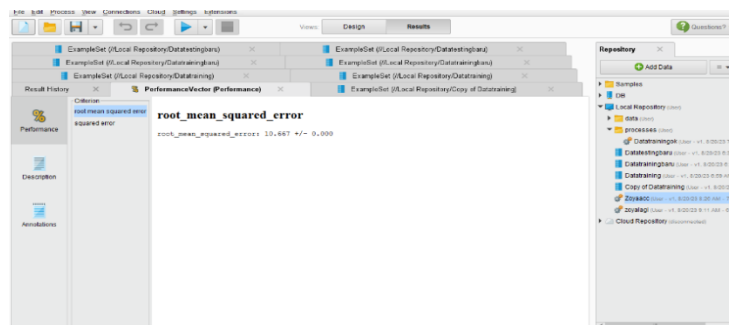
Kemudian tambahkan operator Apply Model dan Performance, Apply Model digunakan untuk mengetahui ramalan bulan ke-67. Hasil ramalan tersebut dapat dilihat pada gambar 5.



Row No.	Bulan ke-67	prediction	Bulan ke-1	Bulan ke-2	Bulan ke-3	Bulan ke-4	Bulan ke-5	Bulan ke-6	Bulan ke-7
1	2132	2121.333	2001	2119	2102	2122	2313	2147	2173
2	2132	2121.333	2001	2118	2102	2122	2313	2147	2173
3	2132	2121.333	2001	2119	2102	2122	2313	2147	2173
4	2132	2121.333	2001	2118	2102	2122	2313	2147	2173
5	2132	2121.333	2001	2118	2102	2122	2313	2147	2173
6	2132	2121.333	2001	2119	2102	2122	2313	2147	2173

Gambar 4. Hasil ramalan bulan ke-67

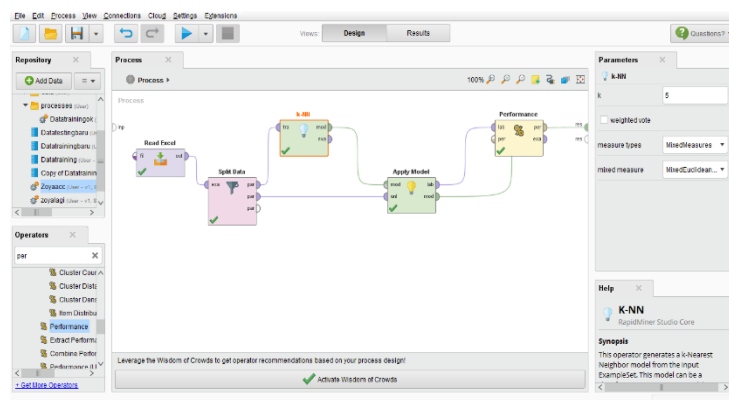
Kemudian operator Performance digunakan untuk mengetahui tingkat akurasi penggunaan metode KNN dengan k=3. Hasil akurasi digunakan dengan metode RMSE (Root Mean Square Error) sebesar 10,667 atau 0,106%.



Operator	Performance
root_mean_squared_error	root_mean_squared_error: 10.667 +/- 0.000

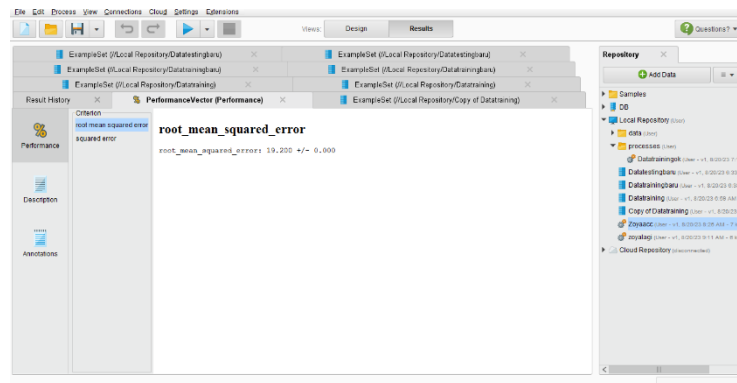
Gambar 5. Tingkat Akurasi Menggunakan RMSE

Untuk melihat keefektifan penggunaan nilai k, penggunaan KNN dengan k=3 dibandingkan dengan uji model lagi menggunakan k=5. Pengujian model dapat dilihat pada gambar 7.



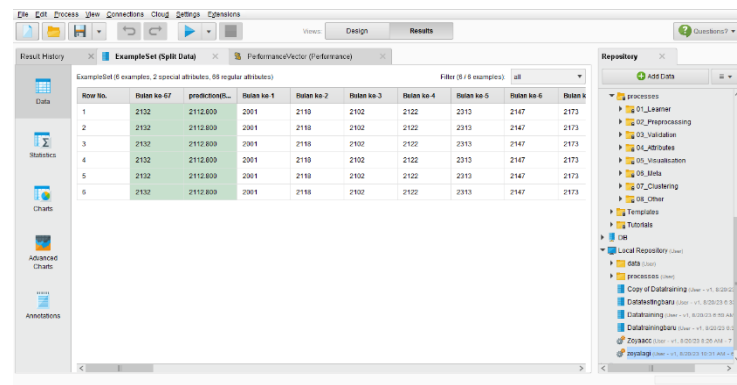
Gambar 6. Input k=5

Berdasarkan gambar 7, nilai akurasi KNN menggunakan $k=5$ adalah 19,200 atau 0,19%.



Gambar 7. Hasil Akurasi RMSE dengan $k=5$

Selanjutnya untuk menentukan ramalan bulan ke-31 dengan $k=5$, didapatkan hasil seperti pada gambar 9.



Gambar 8. Ramalan Bulan 31 dengan $K=5$.

Berdasarkan data testing dan data training yang telah diinput pada Rapidminer, dapat dilihat bahwa pada $k=3$, harga TBS pada bulan ke-67 (Juli 2023) adalah Rp. 2132 sedangkan prediksi harga TBS adalah Rp. 2121,333 dengan hasil akurasi ramalan ini adalah sebesar 10,667. Sedangkan pada $k=5$, harga TBS pada bulan ke-67 (Juli 2023) adalah Rp. 2132 dengan prediksi sebesar Rp. 2112,800 dengan hasil akurasi ramalan ini adalah sebesar 19,200. Hasil ini menunjukkan terdapat perbedaan akurasi berdasarkan nilai k , namun nilai $k=3$ memiliki harga ramalan yang lebih mendekati sehingga dinyatakan dalam kasus ini, nilai $k=3$ lebih efektif digunakan.

4. KESIMPULAN

Pada penelitian ini dilakukan pemodelan dengan menggunakan algoritma K-nearest Neighbor dengan menggunakan data yang diolah menggunakan tahapan Knowledge Discovery in Database (KDD). Berdasarkan hasil perhitungan data mining menggunakan harga TBS kelapa sawit dari tahun 2018 sampai tahun 2023 (Mei 2023) diperoleh kesimpulan bahwa prediksi harga TBS kelapa sawit pada bulan ke-67 (Juli 2023) memiliki tingkat akurasi sebesar 10,667 dengan $k=3$ dan 19,200 dengan $k=5$. Sehingga dapat disimpulkan bahwa penggunaan KNN pada kasus ini lebih efektif menggunakan $k=3$ karena RMSE paling akurat yang mendekati 0.

REFERENCES

- [1] I. Watson, Applying case-based reasoning: techniques for enterprise systems. Morgan Kaufmann Publishers Inc., 1998.
- [2] M. Kuhn and K. Johnson, Feature engineering and selection: A practical approach for predictive models. Chapman and Hall/CRC, 2019.
- [3] Kusriani and E. T.L, Algoritma dan Data Mining. Yogyakarta: Andi, 2009.
- [4] K. Liu, Z. Li, C. Yao, J. Chen, K. Zhang, and M. Saifullah, "Coupling the k-nearest neighbor procedure with the Kalman filter for real-time updating of the hydraulic model in flood forecasting," Int. J. Sediment Res., vol. 31, no. 2, pp. 149–158, 2016.
- [5] R. I. Ndaumanu and M. R. Arief, "Analisis prediksi tingkat pengunduran diri mahasiswa dengan metode K-Nearest Neighbor," Yogyakarta: STMIK AMIKOM Yogyakarta, 2014.
- [6] H.-I. Eum, S. P. Simonovic, and Y.-O. Kim, "Climate change impact assessment using k-nearest neighbor weather generator: case study of the Nakdong River basin in Korea," J. Hydrol. Eng., vol. 15, no. 10, pp. 772–785, 2010.
- [7] R. Hutami and E. Z. Astuti, "Implementasi Metode K-Nearest Neighbor Untuk Prediksi Penjualan Furniture Pada CV. Octo Agung Jepara," Univ. Dian Nuswantoro Semarang, vol. 3, no. 2, pp. 40–51, 2016.
- [8] A. G. Novianti and D. Prasetyo, "Penerapan algoritma k-nearest neighbor (k-nn) untuk prediksi waktu kelulusan

- mahasiswa,” in Seminar Nasional APTIKOM (SEMNASTIKOM), 2017, pp. 108–113.
- [9] H. Risman, D. Nugroho, and Y. R. W. Utami, “Penerapan Metode K-Nearest Neighbor Pada Aplikasi Penentu Penerima Beasiswa Mahasiswa Di STMIK Sinar Nusantara Surakarta,” *J. Teknol. Inf. dan Komun.*, vol. 3, no. 2, 2015.
- [10] F. Hermawan and H. Agung, “Implementasi Metode K-Nearest Neighbor Pada Aplikasi Data Penjualan PT. Multitek Mitra Sejati,” *J. Sains dan Teknol.*, vol. 4, no. 2, 2017.
- [11] T. Tundo, “Prediksi Produksi Minyak Kelapa Sawit Menggunakan Metode Fuzzy Tsukamoto Dengan Rule Yang Terbentuk Menggunakan Decision Tree Reptree,” *J. Nas. Pendidik. Tek. Inform. JANAPATI*, vol. 9, no. 2, pp. 253–265, 2020.
- [12] A. A. Thoha, “Pengelolaan Panen Kelapa Sawit (*Elaeis guineensis* Jacq.) di Kebun Adolina, Sumatera Utara,” *Bul. Agrohorti*, vol. 5, no. 2, pp. 157–166, 2017.
- [13] S. K. Pal and S. C. K. Shiu, *Foundations of soft case-based reasoning*. John Wiley & Sons, 2004.
- [14] M. Furqan and H. L. S. Sriani, “Klasifikasi daun bugenvil menggunakan gray level co-occurrence matrix dan K-Nearest Neighbor,” *J. CoreIT*, vol. 6, no. 1, pp. 22–29, 2013.
- [15] E. Elnahrawy, “Log-based chat room monitoring using text categorization: A comparative study,” in *The International Conference on Information and Knowledge Sharing*, US Virgin Islands, 2002.
- [16] W. F. Ardianti, S. Sriani, and A. H. Hasugian, “Application of color extraction methods and k-nearest neighbor to determine maturity avocado butter,” *J. Tek. Inform. CIT Medicom*, vol. 15, no. 1, pp. 9–20, 2023.
- [17] C. Romero, S. Ventura, P. G. Espejo, and C. Hervás, “Data mining algorithms to classify students,” in *Educational data mining 2008*, 2008.
- [18] M. Sarkar and T.-Y. Leong, “Application of K-nearest neighbors algorithm on breast cancer diagnosis problem,” in *Proceedings of the AMIA Symposium*, 2000, p. 759.