

Prediksi Penjualan Pada PT. Sewu Segar Nusantara Dengan Algoritma K-Nearest Neighbor (KNN)

Muhamad Fahrurrozi ¹, Taufik Hidayat ², Asri Nurhafsari ³

^{1,2,3} Universitas Islam Syekh Yusuf, Tangerang, Indonesia

¹ Ozirozi41@gmail.com, ² thidayat@unis.ac.id, ³ anurhafsari@unis.ac.id

ABSTRACT	Keywords / Kata Kunci
<p>Dalam dunia bisnis, analisis dan prediksi penjualan memiliki peran krusial dalam mengoptimalkan operasi perusahaan. distributor buah yang merupakan salah satu entitas bisnis yang memiliki tantangan unik dalam mengelola persediaan dan memenuhi permintaan pelanggan. Pengambilan keputusan yang tepat dalam mengelola stok buah yang tepat waktu dapat membantu perusahaan menghindari biaya penyimpanan berlebihan dan kekurangan persediaan yang dapat merugikan bisnis. <i>K-Nearest Neighbor (KNN)</i> adalah salah satu algoritma pembelajaran mesin yang telah digunakan secara luas dalam berbagai bidang, termasuk analisis bisnis. Algoritma ini berfokus pada pengelompokan data berdasarkan kemiripan fitur. Dalam konteks prediksi penjualan, KNN dapat digunakan untuk memprediksi permintaan buah berdasarkan pola-pola pembelian pelanggan sebelumnya dan fitur-fitur lainnya. dari hasil pengujian dapat dilihat bahwa terdapat nilai rata-rata akurasi sebesar 86%, recall sebesar 87%, presisi sebesar 81%, dan f1-score sebesar 84%. Hasil akurasi tersebut menunjukkan bahwa model yang telah dibuat sangat baik dalam memprediksi penjualan buah di PT. Sewu Segar Nusantara</p>	<p>Data Mining, Forecasting, K-Nearest Neighbor</p>
<p><i>In the business world, sales analysis and prediction play a crucial role in optimizing company operations. As a fruit distributor, which is one of the business entities facing unique challenges in inventory management and meeting customer demands. Making informed decisions in managing timely fruit stock can help the company avoid excessive storage costs and inventory shortages that could adversely affect the business. K-Nearest Neighbor (KNN) is one of the widely used machine learning algorithms across various fields, including business analysis. This algorithm focuses on grouping data based on feature similarity. In the context of sales prediction, KNN can be employed to forecast fruit demand by considering patterns of previous customer purchases and other features. From the test results, it can be observed that there is an average accuracy value of 86%, recall of 87%, precision of 81%, and an f1-score of 84%. These accuracy results indicate that the created model is highly effective in predicting fruit sales at PT. Sewu Segar Nusantara.</i></p>	<p>Data Mining, Forecasting, K-Nearest Neighbor</p>

1. PENDAHULUAN

Penjualan memiliki peran yang sangat penting dalam operasi sebuah perusahaan. Melalui penjualan, sebuah perusahaan dapat menghasilkan keuntungan yang memungkinkannya untuk menjalankan usahanya. Namun, dalam lanskap persaingan era perdagangan bebas saat ini, bisnis ditantang untuk lebih mempersiapkan diri secara profesional dan juga fleksibel. Hal ini bukan hanya untuk bertahan, namun juga untuk tumbuh dan berkembang. Untuk mencapai hal ini, perusahaan harus menggunakan strategi yang efektif untuk memenuhi permintaan pasar.

Oleh karena itu, persaingan terletak pada kemampuan suatu perusahaan untuk menyediakan produk yang hemat biaya, unggul, dan selalu memiliki persediaan yang memadai. PT. Sewu Segar Nusantara, sebuah supermarket yang menawarkan berbagai produk konsumen termasuk buah dan sayur, memperoleh beragam produk segar berkualitas dari pemasok lokal maupun internasional. Untuk memastikan kepuasan dari pelanggan, perusahaan harus mempunyai strategi pemasaran yang kuat dan kreatif. Namun demikian, strategi yang hanya bergantung pada insting dan intuisi kepemimpinan terkadang dapat menjadi tidak efektif dan tidak akurat.

Hingga saat ini, belum ada pendekatan sistematis untuk meramalkan atau memprediksi penjualan buah dan sayur di PT. Sewu Segar Nusantara. Akibatnya, terjadi masalah seperti persediaan berlebih, kerusakan, dan bahkan kekurangan stok yang menyebabkan kerugian finansial. Untuk mengatasi hal ini, diperlukan sistem prediksi penjualan yang terkomputerisasi. Peramalan melibatkan prediksi masa depan berdasarkan data masa lalu. Ini meliputi estimasi penjualan di masa depan, penentuan potensi penjualan, dan proyeksi pangsa pasar (Amalia YR, 2018). Hal ini sangat penting untuk pengambilan keputusan dalam manajemen persediaan.

Penambangan data (data mining) sangat penting untuk peramalan yang akurat, karena dapat mengambil informasi dari kumpulan data yang besar. Penambangan data adalah bidang lintas ilmu yang menggabungkan pembelajaran mesin, pengolahan pola, statistik, manajemen basis data, dan visualisasi untuk mengekstrak informasi dari basis data besar (R, Rozi and Sukmana, 2021). Ini membantu mengungkap wawasan dari data yang luas, membantu dalam meramalkan penjualan. Di antara berbagai metode peramalan, pendekatan K-Nearest Neighbor (K-NN) patut diperhatikan. Metode ini mengklasifikasikan objek berdasarkan data pelatihan terdekat (Handoko, Tambunan and Hardinata, 2021). K-NN sangat efektif jika dilatih dengan dataset yang luas. Implementasi metode K-NN untuk peramalan diharapkan dapat membantu dalam perencanaan persediaan stok.

Pada penelitian sebelumnya, metode K-Nearest Neighbor telah digunakan untuk meramalkan harga saham. Ini melibatkan prediksi harga saham melalui analisis penambangan data dari dataset keuangan dan bisnis. Algoritma K-NN dipilih karena akurasi yang tinggi dan tingkat kesalahan yang rendah. Hasil prediksi ini bermanfaat bagi investor dan manajemen dalam membuat keputusan investasi. Penelitian ini menunjukkan bahwa prediksi berbasis K-NN menghasilkan tingkat akurasi yang lebih tinggi dibandingkan dengan harga saham yang sebenarnya (Dewi, Sri Puspita; dkk, 2022).

2. METODE

2.1 Klasifikasi

Klasifikasi merupakan metode pada data mining yang dapat beroperasi dengan mengenali karakteristik atribut dari himpunan data yang telah dilakukan identifikasi dan juga ditentukan sebelumnya. Salah satu aspek utama dari klasifikasi adalah keberadaan kelas atau tujuan tertentu. Pendekatan ini berfokus pada mengelompokkan suatu data ke dalam kelas atau tujuan tertentu untuk membentuk suatu model atau aturan. (Mandar, 2021).

2.2 Algoritma K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (K-NN) merupakan sebuah metode dalam *machine learning* yang digunakan untuk mengklasifikasi atau memprediksi terhadap suatu objek berdasarkan mayoritas label atau kategori dari tetangga terdekatnya. Dalam K-NN, objek yang akan diklasifikasikan diberikan label berdasarkan mayoritas dari label objek-objek terdekatnya dalam dataset pelatihan. Metode ini dapat digunakan untuk masalah klasifikasi di mana kita ingin mengkategorikan suatu objek ke dalam kelas atau kategori tertentu berdasarkan atribut-atribut yang dimiliki oleh objek tersebut. (Irsyad, Winanjaya and Rizki, 2022). Berikut persamaan *naive bayes* (Surbakti, 2021):

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

$d(x, y)$: jarak *euclidean* antara data x dan y
 x & y : dua titik dalam ruang *euclidean*
 n : adalah jumlah dimensi dalam ruang Euclidean

2.3 Data Mining

Data mining atau penambangan data adalah suatu proses mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan terkait berbagai basis data yang berhubungan dengan memanfaatkan metode dari statistik, matematika, kecerdasan buatan, dan juga pembelajaran mesin (Aqmila, D 2022).

2.4 Python

Python merupakan suatu bahasa pemrograman yang berorientasi pada objek interaktif yang menyediakan struktur data tingkat tinggi. Python sebagai bahasa pemrograman interpretatif yang multiguna dengan perancangan yang berfokus pada keterbacaan suatu kode agar sintaks yang digunakan lebih mudah untuk dipahami. Python juga dikatakan sebagai bahasa yang dapat menggabungkan kemampuan, kapabilitas, dengan sintaks kode yang sangat jelas. Bahasa pemrograman dari python dirancang khusus untuk memberikan kemudahan bagi programmer dalam membuat program, baik dari segi efisiensi waktu, mudah dalam pengembangan dan kompatibilitasnya dengan sistem. Python juga dapat digunakan untuk menciptakan aplikasi yang *stand alone* (berdiri sendiri) maupun *scripting programming* (pemrograman script). (Yufita and Fatayat, 2022).

2.5 Confusion Matrix

Confusion Matrix ialah teknik yang dipakai agar dapat menghitung kinerja dari suatu model dan terdapat informasi yang membandingkan dari hasil yang telah diproses oleh sistem dengan hasil yang sebenarnya (Nugroho, KS 2019). *Confusion matrix* dapat digunakan untuk merepresentasikan kinerja suatu model dalam empat parameter. Empat parameter tersebut antara lain true negative (TN), true positive (TP), false negative (FN) dan false positive (FP). Keempat parameter ini dimanfaatkan sebagai acuan dalam rumus untuk menghitung *performance metrics* seperti *accuracy*, *precision*, *recall* dan *f1-score*. *Accuracy* merupakan nilai prediksi benar atas keseluruhan data, sedangkan *precision* merupakan nilai prediksi benar positif dibandingkan keseluruhan hasil yang diprediksi positif. *Recall* yaitu nilai prediksi benar positif dibandingkan keseluruhan data yang benar positif, *f1-score* merupakan rata-rata perbandingan dari nilai *precision* dan *recall* (R Arthana , 2019) .

$$Akurasi = \frac{TP+TN}{TP+TN+FN+FP} \quad (2)$$

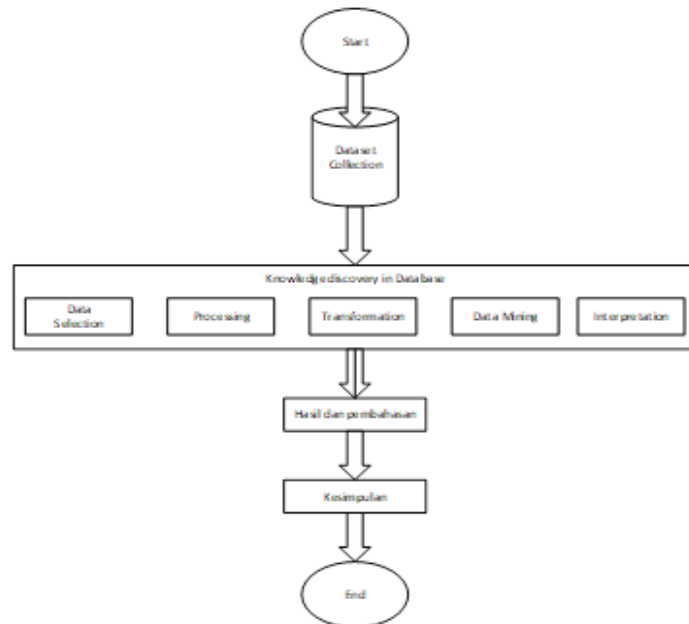
$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1 - Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (5)$$

2.6 Tahap Analisis Data

Model yang digunakan pada riset kali ini meliputi berbagai alur di dalam proses KDD. Pertama metode KDD akan menetapkan visi dan evaluasi akan menjadi yang terakhir. Tahapan proses KDD seperti gambar berikut :



Gambar 1. Flowchart Tahapan Analisis Data KDD

1. Seleksi Data
Tahap ini melibatkan pemilihan data yang relevan dan signifikan untuk analisis lebih lanjut.
2. Preprocessing Data
Tahap ini mencakup proses untuk pembersihan data, penghapusan noise, penghapusan data yang tidak relevan, dan penggabungan atau pemisahan data jika diperlukan.
3. Transformasi Data
Pada tahapan ini, data yang telah dipilih diubah menjadi format yang sesuai untuk analisis, seperti format numerik atau kategori *Data Mining*. Tahapan ini yang akan dilaksanakan adalah metode penggalian data agar mengolah data yang tersedia. Metode yang akan dipakai dalam proses ini yaitu *naive bayes classifier*.
4. Data mining
Tahap ini melibatkan penggunaan teknik data mining untuk mengidentifikasi pola atau hubungan yang tersembunyi dalam data.
5. Evaluasi
Tahap ini melibatkan evaluasi hasil data mining dan penilaian kualitas informasi yang dihasilkan.

3. HASIL DAN PEMBAHASAN

3.1 Analisa Masalah

PT. Sewu Segar Nusantara mengandalkan insting seorang manajer dalam pemilihan produk. Manajer melakukan uji rasa dan menilai daya tarik visual suatu produk tanpa melakukan perhitungan mendalam. Jika manajer percaya suatu produk dapat laku di pasaran, mereka akan memesan produk tersebut dalam jumlah tertentu. Praktik ini sering kali mengakibatkan adanya penumpukan barang,

kerusakan, atau bahkan terjadi kekurangan stok. Permasalahan yang sering muncul termasuk penumpukan barang rusak atau basi dan kekurangan stok. Untuk mengatasi hal ini, diperlukan sistem yang dapat memprediksi penjualan buah dan sayur di PT. Sewu Segar Nusantara, dengan tujuan meminimalkan kerugian. Penelitian ini menggunakan data terpilih yang disediakan oleh PT. Sewu Segar Nusantara, berfokus pada tiga tahun terakhir, dimulai dari 2020. Data ini menjadi dasar perbandingan dalam memprediksi penjualan buah. Dalam tahap analisis dan diskusi, algoritma yang dipilih untuk prediksi penjualan adalah K-Nearest Neighbor, diimplementasikan dengan Python. Analisis ini menggunakan data nyata dari PT. Sewu Segar Nusantara untuk mencapai hasil yang akurat. Setelah memberi nilai pada atribut penentu, tahapan selanjutnya yaitu menyediakan data testing dan juga data training.

3.2 Metode Pengumpulan Data

Dalam pengumpulan data terdapat sumber data, dimana sumber data yang didapatkan langsung oleh peneliti disebut sebagai sumber primer. Sedangkan apabila data yang didapatkan berasal dari tangan kedua dapat disebut sebagai sumber sekunder. Sebagai metode dalam pengumpulan data produk yang menjadi landasan utama dalam proses prediksi ini adalah menggunakan wawancara.

Pada penelitian ini wawancara dilakukan kepada bagian Accounting perusahaan. Dari hasil wawancara tersebut diperoleh data terkait produk buah dan sayur sebanyak 360 data. Tahapan selanjutnya adalah melakukan pengambilan sample data. Sample data ini merupakan 10% dari jumlah data keseluruhan yang diperoleh. Kemudian dilakukan *cleaning* atau pembersihan data yang bermanfaat untuk mengurangi efek noise pada saat proses perhitungan dan juga menghilangkan atribut yang tidak digunakan. Adapun data-data yang telah di *cleaning* adalah sebagai berikut:

Tabel 1. Tabel Data Teratribut

No	Item No	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Aug	Sep	Okt	Nov	Des	Total	Status
1	101010104	22911	23636	27094	25432	21061	26595	28295	28174	27518	28120	27253	28797	314886	Sedang
2	101010106	158373	163381	187293	175789	145580	183827	195588	194746	190210	194378	188377	199053	2176595	Banyak
3	101010112	2536	2618	3001	2817	2332	2947	3134	3121	3048	3116	3020	3191	34881	Sedikit
4	101010206	153794	158658	181878	170708	141372	178514	189935	189115	184711	188760	182932	193300	2113677	Banyak
5	101010208	38046	39249	44993	42229	34973	44161	46986	46784	45693	46696	45253	47819	522882	Sedang
6	101010210	1466	1512	1733	1627	1348	1701	1811	1802	1760	1799	1744	1842	20145	Sedikit
7	101010214	56308	58089	66591	62502	51759	65359	69541	69242	67628	69111	66976	70772	773878	Sedang
8	101010215	10624	10960	12563	11793	9765	12332	13120	13064	12759	13038	12637	13354	146009	Sedang
9	101010503	18126	18700	21435	20119	16661	21039	22384	22289	21770	22247	21560	22781	249111	Sedang
...
36	334010101	2116	1374	1528	1103	478	775	1667	1459	1859	2022	1561	1371	17313	Sedikit

3.3 Data Mining

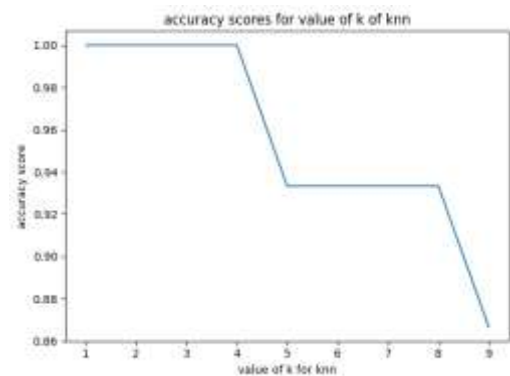
Data mining merupakan proses analisis data dari berbagai sumber yang kemudian diolah menjadi informasi, pengetahuan, atau pola penting. Tujuan dari proses ini adalah untuk meningkatkan keuntungan dan mengurangi biaya pengeluaran, atau bahkan keduanya. Penulis melakukan percobaan implementasi algoritma *K-Nearest Neighbor* terhadap dataset dengan data training 60%, data testing 40%, dan jarak $k=10$. Berikut adalah modeling *K-Nearest Neighbor*

```
[37] from sklearn.neighbors import KNeighborsClassifier

[36] k_range = list(range(1,10))
scores = []
for k in k_range:
    model_knn = KNeighborsClassifier(n_neighbors=k)
    model_knn.fit(X_train, y_train)
    y_pred = model_knn.predict(X_test)
    scores.append(accuracy_score(y_test, y_pred))

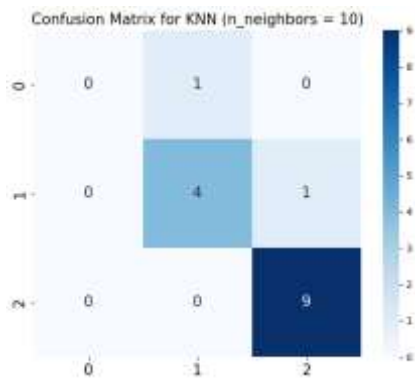
plt.plot(k_range, scores)
plt.xlabel('value of k for knn')
plt.ylabel('accuracy score')
plt.title('accuracy scores for value of k of knn')
plt.tight_layout()
plt.show()
```

Gambar 2. Code Perhitungan *K-Nearest Neighbor*



Gambar 3. Hasil Akurasi Score *K-Nearest Neighbor*

Dari hasil tersebut maka di dapatkan hasil akurasi score 86%, dan confusion matrix yang di dapat dari perhitungan tersebut adalah sebagai berikut :



Gambar 4. Hasil confusion matriks KNN

The classification report for KNN (n_neighbors = 10) is:

	precision	recall	f1-score	support
Banyak	0.00	0.00	0.00	1
Sedang	0.80	0.80	0.80	5
Sedikit	0.90	1.00	0.95	9
accuracy			0.87	15
macro avg	0.57	0.60	0.58	15
weighted avg	0.81	0.87	0.84	15

Gambar 5. Hasil klasifikasi

Didapatkan hasil *Weight Average* untuk precision 81%, recall 87%, dan f1-score 84%. Hasil akurasi tersebut menunjukkan bahwa model yang telah dibuat sangat baik dalam memprediksi penjualan buah mahasiswa di PT. Sewu Segar Nusantara. Untuk membuktikan hasil dari klasifikasi di atas, maka penulis menguji model di atas dengan sebuah data baru untuk diprediksi. Data tersebut akan dihitung dengan dataset yang ada pada Tabel 1. Berikut merupakan data yang akan diprediksi :

Tabel 2. Data Prediksi Baru

No	Item No	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Aug	Sep	Okt	Nov	Des	Total	Status
1	737020201	15400	15400	15400	30200	13600	26600	42600	42600	42600	46800	46800	46800	384800	Sedang
2	636020103	74399	69486	99446	136299	119634	73081	72799	70745	70842	82059	103562	99744	1072098	Banyak
3	636020104	4002	3543	5281	7376	6629	3985	3951	3787	3790	4477	5660	5505	57984	Sedikit
4	636020105	2737	2608	3732	5861	5178	2443	2421	2338	2360	2796	3661	3450	39584	Sedikit

1) Perhitungan data testing ke-1, dengan Item No 737020201

$$d1 = \sqrt{\frac{(22911 - 15400)^2 + (23636 - 15400)^2 + (27094 - 15400)^2 + (25432 - 30200)^2 + (21061 - 13600)^2 + (26595 - 26600)^2 + (28295 - 42600)^2 + (28174 - 42600)^2 + (27518 - 42600)^2 + (28120 - 46800)^2 + (27253 - 46800)^2 + (28797 - 46800)^2}{12}}$$

Total perhitungan data testing Item No 737020201.

$d1 = 60858$ $d2 = 529250$ $d3 = 111852$ $d4 = 511133$ $d5 = 75423$,
 $d6 = 115497$ $d7 = 133273$ $d8 = 86250$ $d9 = 67742$ $d10 = 1260787$
 $d11 = 11923$ $d12 = 217897$ $d13 = 66484$ $d14 = 966079$ $d15 = 116560$
 $d16 = 115118$ $d17 = 116560$ $d18 = 117689$ $d19 = 117910$ $d20 = 111510$
 $d21 = 117873$ $d22 = 113284$ $d23 = 117550$ $d24 = 101424$ $d25 = 107714$
 $d26 = 102300$ $d27 = 114133$ $d28 = 112529$ $d29 = 107637$ $d30 = 60393$
 $d31 = 66910$ $d32 = 82909$ $d33 = 60858$ $d34 = 148322$ $d35 = 89085$
 $d36 = 116142$

Tabel 3. Hasil Data Prediksi 737020201

Data	Jumlah data	Status
d1	60858	sedikit
d2	529250	sedang
d3	111852	sedang
d4	511133	sedang
d5	75423	sedikit
d6	115497	sedang
d7	133273	sedang
d8	86250	sedikit
d9	67742	sedikit
d10	1260787	banyak

Hasil data prediksi menunjukkan status “sedang” berjumlah 5, status “sedikit” berjumlah 4, dan status “banyak” berjumlah 1. Maka hasil prediksi adalah benar karena status “sedang” memiliki jumlah yang lebih banyak.

2) Perhitungan data testing ke-2, dengan Item No 636020103

Tabel 4. Hasil Data Prediksi 636020103

Data	Jumlah data	Status
d1	253714	sedang
d2	432386	sedang
d3	310343	sedang
d4	416301	sedang
d5	221746	sedang
d6	313622	sedang
d7	201263	sedang
d8	286459	sedang
d9	265877	sedang
d10	1134616	banyak

Hasil data prediksi menunjukkan status “sedang” berjumlah 9, status “sedikit” berjumlah 0, dan status “banyak” berjumlah 1. Maka hasil prediksi adalah salah karena status “sedang” memiliki jumlah yang lebih banyak.

3) Perhitungan data testing ke-3, dengan Item No 636020103

Tabel 5. Hasil Data Prediksi 636020103

Data	Jumlah data	Status
d1	78569	sedikit
d2	617518	sedang
d3	11303	sedikit
d4	599284	sedang
d5	138531	sedang
d6	13208	sedikit
d7	211135	sedang
d8	30829	sedikit
d9	59743	sedikit
d10	1350940	banyak

Hasil data prediksi menunjukkan status “sedikit” berjumlah 5, status “sedang” berjumlah 4, dan status “banyak” berjumlah 1. Maka hasil prediksi adalah benar karena status “sedikit” memiliki jumlah yang lebih banyak.

4) Perhitungan data testing ke-4, dengan Item No 636020105

Tabel 6. Hasil Data Prediksi 636020105

Data	Jumlah data	Status
d1	82633	sedikit
d2	621897	sedang
d3	8073	sedikit
d4	8608	sedikit
d5	215429	sedang
d6	34245	sedikit
d7	63685	sedikit
d8	1355343	banyak
d9	10973	sedikit
d10	310861	sedang

Hasil data prediksi menunjukkan status “sedikit” berjumlah 6, status “sedang” berjumlah 3, dan status “banyak” berjumlah 1. Maka hasil prediksi adalah benar karena status “sedikit” memiliki jumlah yang lebih banyak.

3.4 Evaluasi

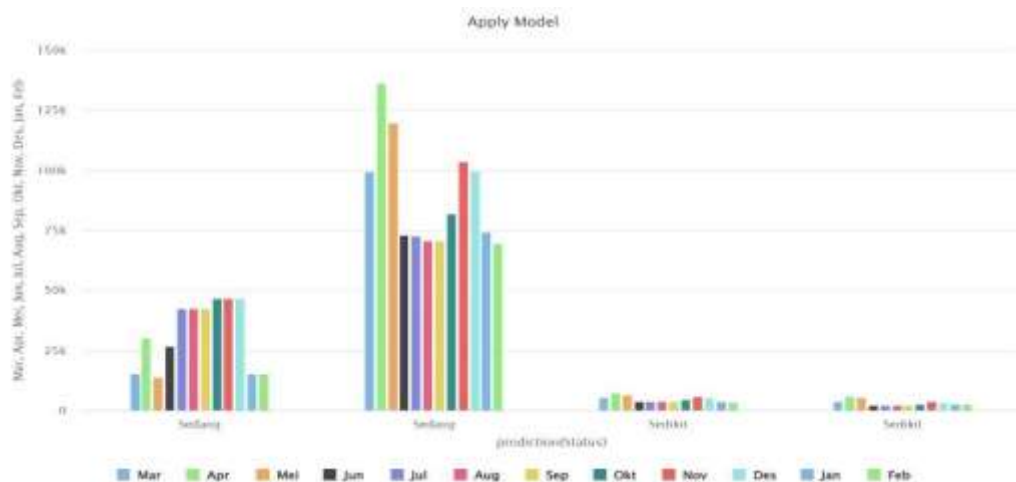
Berdasarkan hasil perhitungan K-Nearest Neighbor dengan dataset yang di bagi 60:40 , maka di hasilkan akurasi score 86% dan confussion matrix dengan nilai precision 81%, recall 87%, dan f1-score 84%. Dan juga dengan perhitungan data prediksi baru menghasilkan akurasi score 75%.

The accuracy for KNN is: 86.66666666666667 %

The classification report for KNN (n_neighbors = 10) is:

	precision	recall	f1-score	support
Banyak	0.00	0.00	0.00	1
Sedang	0.80	0.80	0.80	5
Sedikit	0.90	1.00	0.95	9
accuracy			0.87	15
macro avg	0.57	0.60	0.58	15
weighted avg	0.81	0.87	0.84	15

Gambar 6. Hasil Akurasi prediksi dan confussion matrix



Gambar 7. Hasil Akurasi data prediksi

4. KESIMPULAN

Dalam penelitian ini, dilakukan pembuatan model menggunakan algoritma k-nearest neighbor dengan memproses data berdasarkan langkah-langkah knowledge discovery in database (KDD). Hasil perhitungan dari data mining dengan menggunakan teknik klasifikasi dan algoritma K-Nearest Neighbor menunjukkan bahwa prediksi penjualan buah dan sayur di PT. Sewu Segar Nusantara dilakukan melalui penerapan data mining dan pemanfaatan algoritma K-Nearest Neighbor. Data yang digunakan dalam penelitian ini adalah data penjualan dari tahun 202. Dari dataset tersebut, terdapat 360 sampel data yang diekstrak menjadi 10% dari total data. Sebagai hasilnya, terdapat 36 data sampel yang dianalisis menggunakan algoritma K-Nearest Neighbor. Hasil prediksi menunjukkan tingkat akurasi sebesar 86%, dengan nilai presisi sebesar 81%, recall sebesar 87%, dan f1-score sebesar 84%. Hasil akurasi yang tinggi menunjukkan bahwa model tersebut sangat baik dalam memprediksi penjualan buah . Model ini dapat

memberikan informasi yang berharga dan berguna bagi PT. Sewu Segar Nusantara dalam mengoptimalkan penjualan.

DAFTAR PUSTAKA

- Amalia, Yulia Rizki. *Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus: PT. Bintang Multi Sarana Palembang)*. Diss. UIN RADEN FATAH PALEMBANG, 2018.
- Aqmila, D. 2022. "Perancangan Media Pembelajaran Bahasa Pemograman Python Menggunakan Aplikasi SCRATCH Untuk Siswa Sekolah Menengah Pertama (SMP)," *Skripsi*, 2022.
- Dewi, Sri Puspita, Nurwati Nurwati, and Elly Rahayu. "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor." *Building of Informatics, Technology and Science (BITS)* 3.4 (2022): 639-648.
- Handoko, Dedi, Heru Satria Tambunan, and Jaya Tata Hardinata. "Analisis Penjualan Produk Paket Kuota Internet Dengan Metode K-Nearest Neighbor." *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)* 6.1 (2021): 111-119
- Irsyad, Z., Winanjaya, R. and Rizki, F. (2022) 'Klasifikasi Kepuasan Siswa Terhadap Fasilitas Bengkel Automotif Pada SMK HKBP Pematangsiantar dengan Algoritma Naive Bayes', *zahra jurnal*, 1(2), pp. 106–111.
- K. P. Mandar, "Peguruang: Conference Series," vol. 3, no. 1, 2021.
- Nugroho, K S. 2019. "Confusion Matrix untuk Evaluasi Model pada Supervised Learning," *ksnugroho.medium.com*, 2019. <https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-pada-unsupervised-machine-learning-bc4b1ae9ae3f> (accessed Jul. 26, 2023).
- R. Arthana, "Mengenal Accuracy, Precision, Recall dan Specificity serta yang diprioritaskan dalam Machine Learning," *rey1024.medium.com*, 2019. <https://rey1024.medium.com/mengenal-accuracy-precision-recall-dan-specificity-septa-yang-diprioritaskan-b79ff4d77de8> / 2023-07-26 (accessed Jul. 26, 2023).
- R, A.A.W.P., Rozi, F. and Sukmana, F. (2021) 'Prediksi penjualan produk unilever menggunakan metode k-nearest neighbor', *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 06, pp. 155–160.
- Surbakti, NK (2021), "Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering (Studi Kasus : RSUD.Bangkalan)," *J. Inf. Technol.*, vol. 1, no. 2, pp. 47–53, 2021, doi: 10.32938/jitu.v1i2.1470.
- Yufita and Fatayat. 2022. "Prediksi Jumlah Anak Berisiko Stunting Menggunakan Algoritma Naive Bayes," 2022, [Online]. Available: <http://journal.um-surabaya.ac.id/index.php/JKM/article/view/2203>