

Penerapan Data Mining Untuk Memprediksi Penjualan Buah Dan Sayur Menggunakan Metode K-Nearest Neighbor (Studi Kasus : PT. Central Brastagi Utama)

Ayu Azlina Putri^{1*},

¹Fakultas Ilmu Komputer dan Teknologi Informasi, Program Studi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia
Email: ^{1*}payu9050@gmail.com

Abstrak—PT. Central Brastagi Utama merupakan supermarket yang menjual banyak produk konsumsi diantaranya buah dan sayur. Terdapat banyak jenis buah dan sayur segar berkualitas yang berasal dari dalam maupun luar negeri. Sayangnya selama ini belum ada sistem yang mengatur prediksi atau peramalan untuk penjualan buah dan sayur pada PT. Central Brastagi Utama. Sehingga sering terjadi penumpukan barang, barang rusak dan busuk, atau bahkan kekurangan stok barang yang mengakibatkan kerugian bagi perusahaan. Untuk mengatasi masalah tersebut diperlukan suatu peramalan atau Forecasting. Maka digunakan teknik Data Mining dengan metode K-Nearest Neighbor. Diharapkan dengan menggunakan teknik tersebut dapat mengolah data 3 tahun terakhir menjadi suatu informasi yang dapat membantu pihak perusahaan dalam penyediaan stok barang. Hasil dari penelitian ini adalah prediksi penjualan sejumlah produk, dan menentukan kategori penjualan produk tersebut laris, sedang atau sedikit.

Kata Kunci: Peramalan, Data Mining, Forecasting, K-Nearest Neighbor

Abstract—PT. Central Brastagi Utama is a supermarket that sells many consumer products including fruits and vegetables. There are many types of quality fresh fruit and vegetables that come from within and outside the country. Unfortunately, so far there is no system that regulates predictions or forecasts for fruit and vegetable sales at PT. Central Brastagi Utama. So that there is often an accumulation of goods, damaged and rotten goods, or even a shortage of goods that results in losses for the company. To solve this problem, a forecasting or Forecasting is needed. Then the Data Mining technique is used with the K-Nearest Neighbor method. It is hoped that using this technique can process the last 3 years of data into information that can help the company in providing stock of goods. The results of this research are the prediction of sales of a number of products, and determine which category of product sales are in demand, moderate or little.

Keywords: Forecasting, Data Mining, Forecasting, K-Nearest Neighbor.

1. PENDAHULUAN

Penjualan merupakan faktor terpenting bagi sebuah perusahaan. Karena dengan adanya penjualan, maka suatu perusahaan akan mendapat keuntungan yang lebih supaya bisa melanjutkan usaha tersebut. Namun Persaingan bisnis di era perdagangan bebas seperti sekarang ini sangat ketat, setiap perusahaan dituntut untuk mempersiapkan dirinya secara profesional dan fleksibel sehingga perusahaan tidak hanya mampu bertahan, tetapi juga mampu tumbuh dan berkembang. Untuk itu perusahaan harus memiliki strategi yang tepat agar bisa memenuhi kebutuhan pasar. Maka dari itu persaingan terletak pada bagaimana sebuah perusahaan dapat menghadirkan produk yang lebih murah, lebih baik namun dengan stok barang yang selalu terpenuhi.

PT. Central Brastagi Utama merupakan supermarket yang menjual banyak produk konsumsi diantaranya buah dan sayur. Terdapat banyak jenis buah dan sayur segar berkualitas yang berasal dari dalam maupun luar negeri. Maka untuk menjaga kepuasan pelanggan, pihak PT. Central Brastagi Utama harus memiliki satu strategi pemasaran yang baik. Sayangnya strategi yang masih menggunakan insting dan naluri dari seorang pimpinan perusahaan terkadang sering kali salah dan tidak efektif.

Selama ini belum ada sistem yang mengatur prediksi atau peramalan untuk penjualan buah dan sayur pada PT. Central Brastagi Utama. Sehingga sering terjadi penumpukan barang, barang rusak dan busuk, atau bahkan kekurangan stok barang yang mengakibatkan kerugian bagi perusahaan. Maka dari itu diperlukan suatu sistem prediksi atau peramalan penjualan yang bersifat terkomputerisasi.

Prediksi atau peramalan (forecasting) adalah suatu perhitungan untuk meramalkan masa depan melalui pengujian keadaan dimasa lalu. Peramalan ini berarti memperkirakan besar penjualan di masa depan, bahkan menentukan potensi penjualan dan luas pasar yang dikuasai di masa yang akan datang[1]. Peramalan ini juga berguna membantu pemilik perusahaan dalam mengambil keputusan terhadap penyediaan stok barang. Dalam peramalan atau prediksi diperlukan data mining agar dapat mencari informasi dari jumlah data yang sangat besar.

Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengolahan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar[2]. Data mining digunakan untuk menggali informasi dari data yang besar sehingga dapat digunakan dalam memprediksi penjualan. Dalam peramalan digunakan banyak metode yang bisa digunakan salah satunya metode K-Nearest Neighbor (K-NN). Metode ini merupakan metode yang digunakan untuk mengklasifikasi suatu objek berdasarkan data pembelajaran yang jaraknya paling dekat dari data objek tersebut[3]. Metode K-NN ini sangat efektif apabila training datanya mendekati tak terhingga. Diharapkan dengan menggunakan metode K-Nearest Neighbor dalam memprediksi penjualan dalam membantu dalam merencanakan penyediaan stok barang.

Pada penelitian sebelumnya peramalan dilakukan dengan metode K-Nearest Neighbor yang diimplementasikan untuk peramalan harga saham. penelitian tersebut melakukan prediksi harga saham dengan teknologi data mining untuk

menganalisis volume data bisnis dan keuangan. Algoritma K-Nearest Neighbor digunakan karena memiliki akurasi yang tinggi dengan rasio kesalahan kecil. Hasil dari prediksi atau peramalan bermanfaat untuk membantu investor dan manajemen dalam pengambilan keputusan investasi. Hasil dari penelitian tersebut menunjukkan bahwa hasil prediksi dengan metode K-Nearest Neighbor mempunyai tingkat akurasi yang cukup tinggi dengan data harga saham sebenarnya[4].

2. METODOLOGI PENELITIAN

2.1 Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi tersebut dihasilkan dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut Knowledge Discovery Databases (KDD)[5].

2.2 Prediksi Penjualan

Prediksi adalah Suatu kegiatan yang memperkirakan apa yang terjadi pada masa yang akan datang. Sedangkan Penjualan merupakan pembelian sesuatu (barang atau jasa) dari suatu pihak kepada pihak lainnya dengan mendapatkan ganti uang dari pihak tersebut. Penjualan juga merupakan suatu sumber pendapatan perusahaan, semakin besar penjualan maka semakin besar pula pendapatan yang diperoleh perusahaan [10].

2.3 Algoritma K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbor (K-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek baru berdasarkan (K) tetangga terdekatnya. K-Nearest Neighbor (K-NN) termasuk algoritma Supervised Learning, dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari label class atau kategori pada K-NN. Kelas yang paling banyak muncul, yang akan menjadi kelas hasil klasifikasi[11].

Salah satu cara untuk menghitung jarak dekat atau jauhnya tetangga menggunakan metode Euclidian Distance. Euclidian Distance sering digunakan untuk menghitung jarak. Euclidian Distance berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua objek, di bawah ini merupakan rumus Euclidian Distance:

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

Dimana:

$d(x,y)$ = jarak antara data x ke data y

x_i = nilai X pada training data

y_i = nilai X pada testing data

n = batas jumlah banyaknya data

Jika hasil nilai dari rumus di atas besar maka akan semakin jauh tingkat keserupaan antara kedua objek dan sebaliknya jika hasil nilainya semakin kecil maka akan semakin dekat tingkat keserupaan antar objek tersebut. Objek yang dimaksud adalah training data dan testing data. Dalam algoritma ini, nilai K yang terbaik itu tergantung pada jumlah data. Ukuran nilai K yang besar belum tentu menjadi nilai K yang terbaik begitupun juga sebaliknya. Langkah-langkah untuk menghitung algoritma K-NN:

1. Menentukan nilai K (nilai K dipilih secara manual).
2. Menghitung jarak antara data training dan data testing dengan menggunakan rumus Euclidian Distance.
3. Kemudian mengurutkan objek-objek (data training) tersebut ke dalam kelompok yang berdasarkan jarak terkecil.
4. Menetapkan kelas, dimana kelas yang dipilih adalah kelas dengan jumlah nilai K terbanyak pada data testing.

3. HASIL DAN PEMBAHASAN

3.1 Analisa Masalah

PT. Central Brastagi Utama dalam penyediaan produk hanya berdasarkan insting dari seorang manager. Manager hanya melakukan uji rasa dan melihat bagus tidaknya suatu produk dari tampilannya tanpa dilakukan perhitungan secara mendalam. Jika menurut manager suatu produk itu bisa dijual, maka manager akan memesan produk tersebut dalam jumlah tertentu. Hal ini sering sekali mengakibatkan penumpukan barang, barang rusak dan busuk, atau bahkan kekurangan stok. Masalah yang sering terjadi adalah penumpukan barang, barang rusak dan busuk, atau bahkan kekurangan stok. Untuk itu diperlukan suatu sistem yang dapat digunakan dalam memprediksi penjualan buah dan sayur pada PT. Central Brastagi Utama. Sehingga diharapkan dapat meminimalisir terjadinya kerugian.

Adapun data yang digunakan dalam penelitian ini adalah data pilihan yang diberikan oleh pihak PT. Central Brastagi Utama. Data yang digunakan adalah data 3 tahun terakhir, yaitu mulai dari tahun 2018. Data tersebut akan dijadikan sebagai bahan dalam perbandingan dalam memprediksi penjualan buah dan sayur.

Tahap analisa dan pembahasan ini, algoritma yang dipakai dalam memprediksi penjualan buah dan sayur adalah K-Nearest Neighbor dan RapidMiner sebagai alat bantu penerapan data mining. Dalam menganalisa digunakan data yang real dari pihak PT. Central Brastagi Utama.

3.1.1. Penerapan Algoritma K-Nearest Neighbor (KNN)

Dalam pengumpulan data terdapat sumber data, sumber data yang dihimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder. Sebagai metode dalam pengumpulan data produk yang menjadi landasan utama dalam proses prediksi ini adalah menggunakan wawancara. Wawancara dilakukan dengan pihak Accounting dan diperoleh data produk buah dan sayur sebanyak 300 data.

Selanjutnya melakukan pengambilan sample data. Sample data ini merupakan 10% dari jumlah data keseluruhan yang didapat. Kemudian dilakukan pembersihan data untuk mengurangi efek noise pada saat proses perhitungan dan menghilangkan atribut yang tidak digunakan. Adapun data yang telah di cleaning adalah sebagai berikut:

Tabel 1. Tabel Data Teratribut

N O	Nama barang	KUANTUM						KUANTUM						KATEGORI	
		Januari	Februari	Mar et	Apr il	Mei	Jun i	Juli	Agust us	Septemb er	Oktob er	Novemb er	Desemb er		Tota l
1	Apel Merah	643	890	853	612	783	547	512	932	1343	978	567	765	9425	SEDAN G
2	Apel Hijau	243	231	197	340	271	167	280	340	255	345	275	265	3209	SEDAN G
3	Apel Fuji USA	1030	1430	1728	909	902	924	1020	1123	1874	1290	892	976	14098	BANYAK
4	Apel Fuji USA	568	874	679	738	567	537	476	783	860	876	476	647	8081	SEDAN G
5	P.Jepang, Sinseiki	2859	2671	2981	1987	1781	1890	1233	2902	3450	3891	1841	1220	28706	BANYAK
6	Jeruk HM, Murc, Aust, 2PH, Spy	687	456	533	234	282	567	235	566	235	234	566	254	4849	SEDAN G
7	P. Jambu Hijau	602	662	325	664	632	573	822	560	635	562	355	352	6744	SEDAN G
8	Jeruk Ponkam Mandarin	324	1024	2342	2402	1240	650	342	345	123	234	431	345	9802	SEDAN G
9	P. Korea	1249	1455	1902	1448	2402	2235	2535	3455	2345	2596	2350	4535	28507	BANYAK
10	P. Xiang Lie	867	596	968	868	850	672	685	958	857	950	869	852	9992	SEDAN G
11	Kiwi	455	359	654	874	765	445	557	654	586	56	747	455	6607	SEDAN G
12	Plum	350	234	145	456	352	203	420	345	235	330	124	421	3615	SEDAN G
13	Sunskist RRC	2410	2394	4341	2324	5952	1204	3910	2240	4432	3450	2342	4536	39535	BANYAK
14	Delima	325	455	324	532	355	343	350	290	243	230	143	231	3821	SEDAN G
15	Longan	1244	2052	2034	2350	3459	2395	2305	1958	1858	2395	2592	2591	27233	BANYAK
16	Manggis, Rambutan	604	758	956	675	958	586	867	955	859	954	869	858	9899	SEDAN G

Setelah memberi nilai pada atribut penentu, selanjutnya adalah menyediakan data testing dan data training. Data testing bisa langsung diperoleh dari data selection, hanya saja yang menjadi patokan perbandingan adalah field kuantum yaitu antara bulan januari sampai desember. Sedangkan data training merupakan data yang baru muncul yang kemudian akan dilakukan prediksi. Adapun data training pada penelitian kali ini yaitu terhadap produk Chaisem, Sunkist Import dan Anggur. Dimana data penjualan tersebut diambil secara acak.

1. Penghitungan data testing ke-1, yaitu:

Tabel 2. Tabel Data Testing Chaisem

No	Nama barang	Jan	Feb	Mar	Apr	Mei	Juni	Jujujj juli	Agust	Sept	Okt	Nop	des
1	CHAISEM	23	45	53	25	44	49	41	28	47	41	24	21
2	SUNKIST IMPORT	431	536	338	732	941	636	824	138	534	838	732	841
3	ANGGUR	838	1032	982	765	867	526	2802	736	878	827	982	2027

$$d1 = \sqrt{(643-23)^2 + (890-45)^2 + (853-53)^2 + (612-25)^2 + (783-44)^2 + (547-49)^2 + (512-41)^2 + (932-28)^2 + (1343-47)^2 + (978-41)^2 + (567-24)^2 + (765-21)^2} = 2,705.95$$

$$d2 = \sqrt{(243-23)^2 + (231-45)^2 + (197-53)^2 + (340-25)^2 + (271-44)^2 + (167-49)^2 + (280-41)^2 + (340-28)^2 + (255-47)^2 + (345-41)^2 + (275-24)^2 + (265-21)^2} = 825.28$$

$$d3 = \sqrt{(1030-23)^2 + (1430-45)^2 + (1728-53)^2 + (909-25)^2 + (902-44)^2 + (924-49)^2 + (1020-41)^2 + (1123-28)^2 + (1874-47)^2 + (1290-41)^2 + (892-24)^2 + (976-21)^2} = 4,091.70$$

$$d4 = \sqrt{(568-23)^2 + (874-45)^2 + (679-53)^2 + (738-25)^2 + (567-44)^2 + (537-49)^2 + (476-41)^2 + (783-28)^2 + (860-47)^2 + (876-41)^2 + (476-24)^2 + (647-21)^2} = 2,260.49$$

$$d5 = \sqrt{(2859-23)^2 + (2671-45)^2 + (2981-53)^2 + (1987-25)^2 + (1781-44)^2 + (1890-49)^2 + (1233-41)^2 + (2902-28)^2 + (3450-47)^2 + (3891-41)^2 + (1841-24)^2 + (1220-21)^2} = 8,636.58$$

Pengurutan dari hasil perhitungan. Jarak diurutkan dari yang paling dekat jaraknya ke yang paling jauh. Setelah diurutkan diperoleh sebagai berikut:

$d_{29} = 53.37$, $d_{26} = 55.43$, $d_{27} = 65.07$, $d_{24} = 70.51$, $d_{28} = 73.55$, $d_{23} = 77.56$, $d_{30} = 86.74$, $d_{25} = 88.34$, $d_{22} = 96.42$, $d_{21} = 224.70$, $d_{17} = 623.80$, $d_2 = 825.28$, $d_{19} = 930.66$, $d_{12} = 994.21$, $d_{14} = 1,034.15$, $d_6 = 1,398.78$, $d_7 = 1,883.57$, $d_{11} = 1,924.54$, $d_4 = 2,260.49$, $d_1 = 2,705.95$, $d_{16} = 2,767.08$, $d_{10} = 2,787.15$, $d_8 = 3,772.94$, $d_3 = 4,091.70$, $d_{20} = 4,435.96$, $d_{18} = 5,269.92$, $d_{15} = 7,930.07$, $d_5 = 8,636.58$, $d_9 = 8,650.94$, $d_{13} = 12,148.92$.

Menentukan kelompok data hasil uji berdasarkan label mayoritas dari K tetangga terdekat. Karena nilai $K = 3$, maka diambil 3 jarak terkecil yaitu: d_{29} , d_{26} , dan d_{27} .

Tabel 3. Tabel Hasil Prediksi Chaisem

No	Jarak	Kategori
1	d_{29}	SEDIKIT
2	d_{26}	SEDIKIT
3	d_{27}	SEDIKIT

Dari tabel diatas terdapat 3 kategori SEDIKIT, maka untuk data testing dengan nama barang CHAISEM diklasifikasikan kedalam class kategori SEDIKIT, dengan prediksi total penjualan 441 Kg.

2. Perhitungan untuk data testing ke-2, yaitu:

Tabel 4. Data Testing Sunkist Import

No	Nama barang	Jan	Feb	Mar	Apr	Mei	Juni	Jujujj juli	Agust	Sept	Okt	Nop	des
1	CHAISEM	23	45	53	25	44	49	41	28	47	41	24	21
2	SUNKIST IMPORT	431	536	338	732	941	636	824	138	534	838	732	841
3	ANGGUR	838	1032	982	765	867	526	2802	736	878	827	982	2027

$$d_1 = \sqrt{(643-431)^2 + (890-536)^2 + (853-338)^2 + (612-732)^2 + (783-941)^2 + (547-636)^2 + (512-824)^2 + (932-138)^2 + (1343-534)^2 + (978-838)^2 + (567-732)^2 + (765-841)^2} = 1,384.79$$

$$d_2 = \sqrt{(243-431)^2 + (231-536)^2 + (197-338)^2 + (340-732)^2 + (271-941)^2 + (167-636)^2 + (280-824)^2 + (340-138)^2 + (255-534)^2 + (345-838)^2 + (275-732)^2 + (265-841)^2} = 1,472.76$$

$$d_3 = \sqrt{(1030-431)^2 + (1430-536)^2 + (1728-338)^2 + (909-732)^2 + (902-941)^2 + (924-636)^2 + (1020-824)^2 + (1123-138)^2 + (1874-534)^2 + (1290-838)^2 + (892-732)^2 + (976-841)^2} = 2,501.66$$

$$d_4 = \sqrt{(568-431)^2 + (874-536)^2 + (679-338)^2 + (738-732)^2 + (567-941)^2 + (537-636)^2 + (467-824)^2 + (783-138)^2 + (860-534)^2 + (876-838)^2 + (476-732)^2 + (647-841)^2} = 1,070.99$$

$$d_5 = \sqrt{(2859-431)^2 + (2671-536)^2 + (2981-338)^2 + (1987-732)^2 + (1781-941)^2 + (1890-636)^2 + (1233-824)^2 + (2902-138)^2 + (3450-534)^2 + (3891-838)^2 + (1841-732)^2 + (1220-841)^2} = 6,949.56$$

Pengurutan jarak mulai dari yang paling dekat jaraknya ke yang paling jauh. Maka pengurutannya sebagai berikut:
 $d_7 = 891.67$, $d_4 = 1,070.99$, $d_{16} = 1,129.00$, $d_{11} = 1,149.08$, $d_{10} = 1,202.56$, $d_{12} = 1,358.38$, $d_{14} = 1,381.44$, $d_1 = 1,384.79$, $d_6 = 1,467.47$, $d_2 = 1,472.76$, $d_{19} = 1,555.37$, $d_{17} = 1,760.87$, $d_{21} = 2,039.63$, $d_{22} = 2,148.68$, $d_{23} = 2,165.18$, $d_{25} = 2,166.22$, $d_{24} = 2,189.34$, $d_{26} = 2,200.36$, $d_{29} = 2,210.63$, $d_{27} = 2,230.28$, $d_{28} = 2,240.70$, $d_{30} = 2,263.18$, $d_3 = 2,501.66$, $d_{20} = 2,538.30$, $d_8 = 2,879.17$, $d_{18} = 3,357.24$, $d_{15} = 5,838.93$, $d_9 = 6,762.22$, $d_5 = 6,949.56$, $d_{13} = 10,176.62$.

Berdasarkan label mayoritas dari K, maka diambil 3 jarak terkecil, yaitu:

Tabel 5. Tabel Hasil Prediksi Sunkist Import

No	Jarak	Kategori
1	d_7	SEDANG
2	d_4	SEDANG
3	d_{16}	SEDANG

Dari tabel diatas terdapat 3 kategori SEDANG, maka untuk data testing dengan nama barang SUNKIST IMPORT diklasifikasikan kedalam class kategori SEDANG, dengan prediksi total penjualan 7,521 Kg.

3. Perhitungan untuk data testing ke-3, yaitu:

Tabel 6. Data Testing Anggur

No	Nama barang	Jan	Feb	Mar	Apr	Mei	Juni	juli	Agust	Sept	Okt	Nop	des
1	CHAISEM	23	45	53	25	44	49	41	28	47	41	24	21
2	SUNKIST IMPORT	431	536	338	732	941	636	824	138	534	838	732	841
3	ANGGUR	838	1032	982	765	867	526	2802	736	878	827	982	2027

$$d_1 = \sqrt{(643-838)^2 + (890-1032)^2 + (853-982)^2 + (612-765)^2 + (783-867)^2 + (547-526)^2 + (512-2802)^2 + (932-736)^2 + (1343-878)^2 + (978-827)^2 + (567-982)^2 + (765-2027)^2} = 2,718.85$$

$$d2 = \sqrt{(243-838)^2 + (231-1032)^2 + (197-982)^2 + (340-765)^2 + (271-867)^2 + (167-526)^2 + (280-2802)^2 + (340-736)^2 + (255-878)^2 + (345-827)^2 + (275-982)^2 + (265-2027)^2} = 3,608.16$$

$$d3 = \sqrt{(1030-838)^2 + (1430-1032)^2 + (1728-982)^2 + (909-765)^2 + (902-867)^2 + (924-526)^2 + (1020-2802)^2 + (1123-736)^2 + (1874-878)^2 + (1290-827)^2 + (892-982)^2 + (976-2027)^2} = 2,564.47$$

$$d4 = \sqrt{(568-838)^2 + (874-1032)^2 + (679-982)^2 + (738-765)^2 + (567-867)^2 + (537-526)^2 + (467-2802)^2 + (783-736)^2 + (860-878)^2 + (876-827)^2 + (476-982)^2 + (647-2027)^2} = 2,802.89$$

$$d5 = \sqrt{(2859-838)^2 + (2671-1032)^2 + (2981-982)^2 + (1987-765)^2 + (1781-867)^2 + (1890-526)^2 + (1233-2802)^2 + (2902-736)^2 + (3450-878)^2 + (3891-827)^2 + (1841-982)^2 + (1220-2027)^2} = 6,285.04$$

Pengurutan jarak mulai dari yang paling dekat jaraknya ke yang paling jauh. Maka pengurutannya sebagai berikut:
 d16 = 2,310.52, d10 = 2,482.60, d3 = 2,564.47, d20 = 2,606.70, d1 = 2,718.85, d4 = 2,802.89, d7 = 2,823.35, d11 = 2,998.15, d12 = 3,441.40, d6 = 3,447.04, d14 = 3,504.19, d19 = 3,562.52, d18 = 3,588.72, d2 = 3,608.16, d17 = 3,864.16, d8 = 3,899.62, d21 = 4,127.03, d23 = 4,247.58, d22 = 4,250.85, d25 = 4,253.39, d24 = 4,274.11, d26 = 4,289.82, d27 = 4,316.72, d28 = 4,323.78, d30 = 4,347.89, d15 = 4,806.80, d9 = 5,276.82, d5 = 6,285.04, d13 = 8,689.18.

Berdasarkan label mayoritas dari K, maka diambil 3 jarak terkecil, yaitu:

Tabel 7. Tabel Hasil Prediksi Anggur

No	Jarak	Kategori
1	d16	BANYAK
2	d10	SEDANG
3	d3	BANYAK

Dari tabel diatas terdapat 2 kategori BANYAK dan 1 kategori SEDANG, maka untuk data testing dengan nama barang ANGGUR diklasifikasikan kedalam class kategori BANYAK, dengan prediksi total penjualan 13,262 Kg.

3.2. Implementasi

Kebutuhan sistem yang digunakan dalam implementasi program ini, terbagi menjadi dua perangkat. Perangkat tersebut, yaitu perangkat lunak dan perangkat keras. adapun perangkat lunak dan perangkat keras yang digunakan adalah sebagai berikut:

1. Perangkat Lunak (Software)

Software yang digunakan untuk mengimplementasikan program adalah:

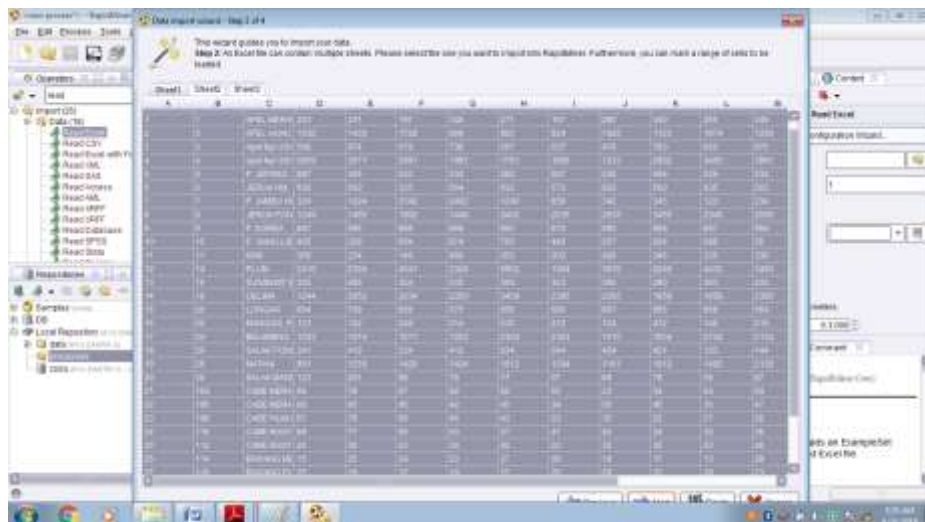
- a. Rapidminer5
- b. Microsoft excel 2007

2. Perangkat Keras (Hardware)

Hardware yang digunakan dalam mengimplementasikan program adalah:

- a. Processor : Intel(R) Celeron(R) CPU N2840 @ 2.16GHz 2.16GHz
- b. Hardisk : 234 GB
- c. RAM : 2.00 GB

Dalam implementasi data mining untuk memprediksi penjualan buah dan sayur, data yang digunakan adalah data penjualan pada satu tahun terakhir yaitu tahun 2018. Dimana data tersebut terbagi menjadi data testing dan data training. Implementasi kali ini menggunakan software aplikasi rapidminer5. Tahapan awal dalam penyipian data yang diperlukan adalah drag and drop operator read excel kemudian lakukan importing tabel microsoft excel kedalam proses.



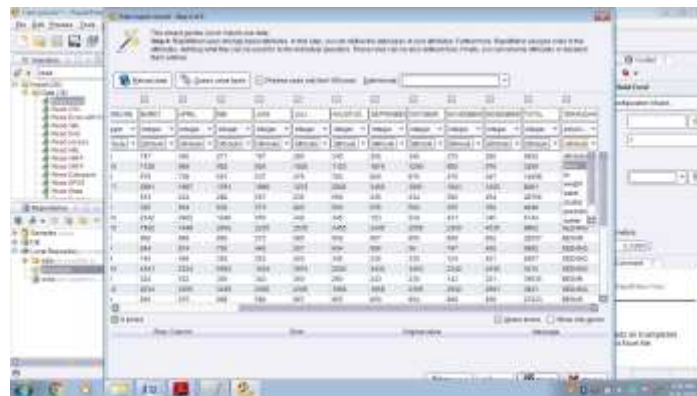
Gambar 1. Tampilan Select Data Tabel Excel

Gambar diatas merupakan data awal yang akan diproses, kemudian klik next.



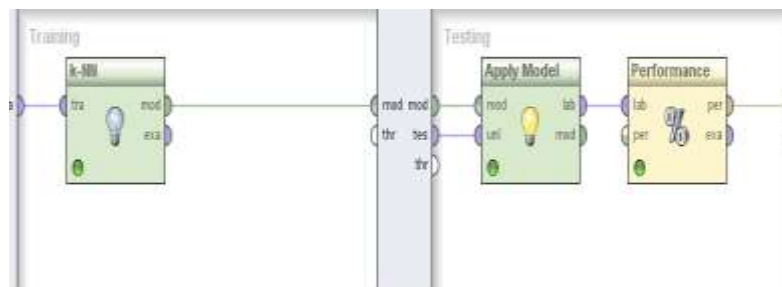
Gambar 2. Tampilan Tabel Cleaning

Pada tahapan ini data akan di cleaning dari data yang noise. Kemudian klik next.



Gambar 3. Tampilan Importing Data Training Pada Operator Read Excel

Tahapan ini kita harus mengganti atribut tabel menjadi label untuk dataset yang menjadi target, kemudian klik finish. Karena tidak semua algoritma yang ada sesuai dan dapat mengolah dataset yang ada, maka harus dilakukan penyesuaian pola data dan sesuai dengan tujuan dari pengolahan data tersebut.



Gambar 4. Model Sub Proses Clasification K-Nearest Neighbor

Adapun hasil Accuracy dari performance vector adalah sebesar 88.89 % ditunjukan seperti berikut:

	trk BEKANG	trk BEKIH	trk KEDL	trk KADON
pred BEKANG	4	0	0	0
pred BEKIH	0	7	0	0
pred KEDL	1	0	3	0
pred KADON	0	0	0	1

Accuracy: 88.89%

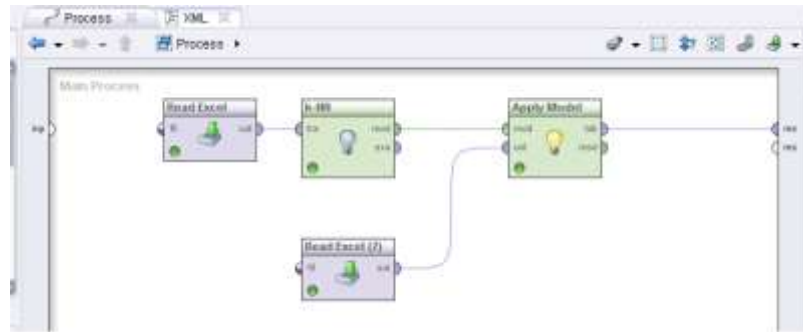
Gambar 5. Hasil perhitungan Nilai Akurasi

Dari data diatas didapat bentuk model klasifikasi K-Nearest Neighbor.



Gambar 6. Model Klasifikasi K-Nearest Neighbor

Setelah melakukan klasifikasi, tahap selanjutnya adalah drag and drop operator read excel untuk importing tabel data training. Selanjutnya drag and drop operator K-NN dengan memasukkan nilai k = 3. Kemudian hubungan operator read excel dengan operator K-NN, setelah itu hubungkan operator K-NN dengan operator Apply Model. Selanjutnya ulangi drag and drop read excel(2) untuk importing tabel data testing, sambungkan pada operator apply model setelah itu hubungkan pada result seperti gambar berikut:



Gambar 7. Susunan operator Algoritma K-Nearest Neighbor

Tahap terakhir dari proses data mining dengan rapidminer adalah dimana setelah semua operator terhubung kemudian klik icon play atau tombol F11, maka akan muncul tab result, yang isinya adalah hasil prediksi dari semua data testing yang memenuhi nilai k = 3.

Row No	TOTAL	confidence_0	confidence_1	confidence_2	predictionK	NO	KODE	NAMA BARA	JANUARI	FEBRUARI	MARET	APRIL	MEI
1	441	0	0	1	KECIL	1	128	CHAISEM	23	45	53	25	44
2	7521	1	0	0	SEDANG	2	130	SUNKIST M	431	536	338	732	941
3	13262	0.333	0.667	0	BESAR	3	132	ANGGUR	638	1032	982	765	867

Gambar 8. Tampilan Hasil Prediksi

Gambar diatas menunjukkan hasil data pengklasifikasian data testing dimana nama barang Chaisem masuk ke prediction class Kecil dengan total penjualan sebanyak 441 kg, Sunkist Import masuk ke prediction class kategori Sedang dengan total penjualan 7.521 kg, dan Anggur masuk ke prediction class kategori Besar dengan total penjualan 13.262 kg.

4. KESIMPULAN

Pada penelitian ini dilakukan pemodelan menggunakan algoritma k-nearest neighbor dengan menggunakan data yang diolah berdasarkan tahapan knowledge discovery in database (KDD). Berdasarkan hasil perhitungan data mining dengan menggunakan teknik klasifikasi dan algoritma K-Nearest Neighbor. Berdasarkan uraian tersebut, maka dapat ditarik kesimpulan Prediksi penjualan buah dan sayur pada PT. Central Brastagi Utama dilakukan dengan memanfaatkan penerapan data mining dan menggunakan algoritma K-Nearest Neighbor. Data yang digunakan adalah data penjualan

tahun 2018. Data yang didapat sebanyak 300 sampel dan diekstrak menjadi 10% dari data keseluruhan. Sehingga dalam perhitungannya digunakan 30 data sampel yang dihitung dengan menerapkan algoritma K-Nearest Neighbor. Hasil prediksi penjualan Chaisem masuk ke prediction class Kecil dengan total penjualan sebanyak 441 kg, Sunkist Import masuk ke prediction class kategori Sedang dengan total penjualan 7.521 kg, dan Anggur masuk ke prediction class kategori Besar dengan total penjualan 13.262 kg dengan nilai akurasi prediksi penjualan sebesar 88,89 %.

REFERENCES

- [1] P. Penjualan, F. Pada, and C. V Octo, "IMPLEMENTASI METODE K-NEAREST NEIGHBOR UNTUK PREDIKSI PENJUALAN FURNITURE PADA CV. OCTO AGUNG JEPARA," 2016.
- [2] J. SUNTORO, DATA MINING: Algoritma dan Implementasi dengan Pemrograman php, 1st ed. jakarta: pt elax media komputindo kelompok gramedia, 2019.
- [3] M. M. K. Neighbor, "Penerapan data mining untuk prediksi penjualan produk elektronik terlaris menggunakan metode k-nearest neighbor," 2018.
- [4] K. M. Desa, K. Rt, R. W. Dengan, and M. Time, "IMPLEMENTASI DATA MINING UNTUK MERAMALKAN PENJUALAN DI MINIMARKET IDOLA JL PATI-TAMBAKROMO SERIES," pp. 29–37.
- [5] M. S. retno tri vulandari, S.Si., data mining teori dan aplikasi rapidminer, II. yogyakarta: gava media, 2018.
- [6] E. Buulolo, "Implementasi Algoritma Apriori Pada Sistem Persediaan Obat (Studi Kasus : Apotik Rumah Sakit Estomihi Medan)," pp. 71–83, 2013.
- [7] D. Aprilla, C. D. A. Baskoro, L. Ambarwati, and I. W. S. Wicaksana, "Data Mining dengan Rapid Miner," p. 139, 2013.
- [8] Raddini Gusti Rahayu, "e-book Microsoft Excel 2007," in Working with Microsoft Excel 2007, MUGI (Microsoft User group Indonesia).
- [9] N. S. Pinem and D. P. Utomo, "Implementasi Fuzzy Logic Dengan Infrensi Tsukamoto Untuk Prediksi Jumlah Kemasan Produksi (Studi Kasus: PT. Sinar Sosro Medan)," *Pelita Inform. Inf. dan Inform.*, vol. 9, no. 1, pp. 56–60, 2020.
- [10] D. P. Utomo and Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 2, pp. 437–444, 2020.
- [11] D. P. Utomo, P. Sirait, and R. Yunis, "Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5. 0," *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 4, pp. 994–1006, 2020.
- [12] B. S. Pranata and D. P. Utomo, "Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel Motor (Study Kasus Bengkel Sinar Service)," *Bull. Inf. Technol.*, vol. 1, no. 2, pp. 83–91, 2020.
- [13] D. P. Utomo and B. Purba, "Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. 1, pp. 846–853, 2019.
- [14] R. Amelia and D. P. Utomo, "ANALISA POLA PEMESANAN PRODUK MODERN TRADE INDEPENDENT DENGAN MENEREPAKAN ALGORITMA FP. GROWTH (STUDI KASUS: PT. ADAM DANI LESTARI)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 416–423, 2019.