



Penerapan Data Mining Untuk Memprediksi Penjualan Di Erigo Store Dengan K-Nearest Neighbor

Rini Hartati Saragih¹, William Aldo Pamungkas², Farhan Yumna³, Delima Sitanggang⁴, Sumita Wardani⁵

^{1,2,3,4,5}Jurusan Sistem Informasi Dan Sistem Informatika, Fakultas Sains Dan Teknologi, Universitas Prima Indonesia, Indonesia

Email: rinisaragih69@gmail.com¹, williamaldo137@gmail.com², farhanyumna2001@gmail.com³, delimasitanggang@unprimdn.ac.id⁴, sumitawardani@unprimdn.ac.id⁵*

Abstract

Erigo Store is a company that offers fashion retail with high quality and design goods and is also in great demand by consumers. The Erigo Store company also has many categories of fashion products it offers. Based on the results of this study, modeling is carried out by applying the K-Nearest Neighbor algorithm where the data is processed based on the Knowledge Discovery in Database (KDD) stage. As for the results of research with the application of data mining, the results obtained from sales predictions for 5 product categories. The dataset used from 2021 to 2023 on the Erigo Store from Marketplace Shopee. The data categories sold are T-Shirt, Pants, Jacket, Outwear, and Shirt. The results of each category have an accuracy of 100% for the Shirt category, 95% for the Jacket category, 92.31% for the Pants category, 89.47% for the Outwear category, and 60% for the T-Shirt category. From the total prediction of sales categories, the accuracy of the overall product sales classification is 83.62% with testing using rapid miner tools to determine the sales of each product of the year.

Keywords: Data Mining, K-Nearest Neighbor (KNN), Forecasting, Rapid Miner

Abstrak

Erigo Store adalah perusahaan yang menawarkan fashion retail dengan barang yang berkualitas dan desain yang tinggi dan juga sangat diminati oleh para konsumen. Perusahaan Erigo Store juga mempunyai banyak kategori produk fashion yang ditawarkannya. Berdasarkan hasil penelitian ini, pemodelan yang dilakukan dengan menerapkan algoritma K-Nearest Neighbor dimana data yang diolah berdasarkan tahapan Knowledge Discovery in Database (KDD). Adapun hasil dari penelitian dengan penerapan data mining maka hasil yang diperoleh dari prediksi penjualan untuk 5 kategori produk. Dataset yang digunakan mulai tahun 2021 sampai 2023 pada Erigo Store dari Marketplace Shopee. Kategori data yang terjual yaitu T-Shirt, Pants, Jacket, Outwear, dan Shirt. Adapun hasil dari setiap kategori yang memiliki akurasi kategori Shirt sebesar 100%, kategori Jacket 95%, kategori Pants 92,31%, kategori Outwear 89,47% dan kategori T-Shirt 60%. Dari total keseluruhan prediksi kategori penjualan diperoleh hasil akurasi klasifikasi penjualan keseluruhan produk sebesar 83,62% dengan pengujian menggunakan tools rapid miner untuk menentukan penjualan setiap produk per tahunnya.

Kata Kunci: Data Mining, K-Nearest Neighbor (KNN), Forecasting, Rapid Miner

1. PENDAHULUAN

Penjualan adalah aktivasi atau perjanjian yang menjual produk atau layanan. Dalam proses penjualan, penjual atau jasa mengalihkan kepemilikan barang kepada pembeli dengan harga tertentu. Oleh karena itu, konsep penjualan merupakan salah satu cara untuk memotivasi konsumen agar membeli produk yang ditawarkan. Banyak faktor yang dapat menjadi pengaruh tingkat penjualan, termasuk produk itu sendiri, seperti harga, distribusi, promosi, dan layanan purna jual. Estimasi tingkat penjualan dapat dijadikan acuan untuk menentukan kelangsungan usaha dan tingkat keuntungan yang diinginkan[1].

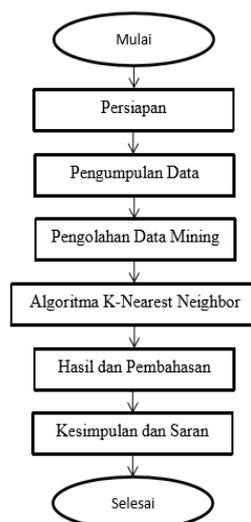
Prediksi merupakan suatu model berkaitan dengan kemampuannya untuk memprediksi hasil terbaik yang paling mendekati kenyataan, berdasarkan data yang dimasukkan. Salah satu fungsi prediksi adalah untuk membantu perusahaan memutuskan jumlah barang yang harus disediakan perusahaan. Selain itu, prediksi dapat membantu perusahaan dalam memastikan persediaan, karena peramalan dapat memberikan pengaruh yang baik bagi perusahaan, sehingga dapat diasumsikan bahwa risiko kesalahan akibat kesalahan perencanaan persediaan berkurang. Biasanya prediksi digunakan untuk mencari data yang akurat, sehingga diperlukan data mining [2].

Data mining adalah teknik pembelajaran mesin, pengenalan pola, database, statistik, dan teknik visualisasi yang dapat memecahkan masalah pengambilan informasi dari penyimpanan database yang besar. Adapun metode Algoritma K-Neighbor Neighbor adalah suatu algoritma yang dapat digunakan untuk memprediksi atau mengklasifikasikan data berdasarkan jenis data pada suatu dataset yang ada. Algoritma *K-Nearest Neighbor* dapat mengklasifikasikan data berdasarkan nilai k yang diberikan. Nilai k pada K-Neighbor sebaiknya menggunakan nilai ganjil saat menggunakan proses klasifikasi. Situasinya berbeda ketika digunakan untuk memprediksi nilai k dalam *K-Nearest Neighbor* di mana ada bilangan ganjil atau genap. Pada penelitian ini menggunakan metode Algoritma *K-Nearest Neighbor* yang digunakan untuk mengklasifikasikan penjualan Erigo Store berdasarkan data terdekat terhadap data yang diolah [3-5]. Peneliti berharap hasil prediksi dilakukan dapat digunakan untuk melakukan analisis prediktif pada periode penjualan *Erigo Store* di masa yang akan datang sehingga memberikan pengaruh besar dalam proses penjualan produk di *marketplace Shopee* atau *marketplace* lain yang sudah melakukan *partnership* dengan *Erigo Store*.

2. METODOLOGI PENELITIAN

2.1. Tahapan Penelitian

Dalam penelitian ini, terlebih dahulu harus dilakukan tahapan penelitian. Hal ini untuk kenyamanan peneliti melakukan proses penelitiannya. Tahapan penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan tahapan penelitian yang telah disebutkan di atas, yaitu tahapan penelitian yang dilakukan dalam proses penelitian ini. Tahapan penelitian di atas menunjukkan proses dari awal kegiatan penelitian hingga akhir penelitian. Tahapan penelitian dapat dijelaskan sebagai berikut:

a) Persiapan

Tahapan ini merupakan tahapan pengolahan data penjualan di Erigo Store.

b) Pengumpulan Data

Pengumpulan data dalam penelitian ini adalah metode dokumentasi yaitu cara pengumpulan data yang dilakukan melalui pencarian *review* pelanggan di *Erigo Store* pada *marketplace Shopee* antara peneliti dengan subjek penelitian dan sumber data. Pada penelitian kali ini data dicari melalui *review* pelanggan tentang penjualan produk *Erigo Store*. Data yang digunakan sebagai proses data mining merupakan data penjualan *Erigo Store* dari bulan Januari 2021 sampai bulan juni 2023.

c) Pengolahan Data Mining

Setelah data dianalisa kemudian akan diolah menggunakan metode data mining dengan tools *Rapid miner*. Data yang diolah meliputi dua bagian yaitu kategori produk dan total penjualan produk.

d) Hasil dan Pembahasan

Pada tahap ini dideskripsikan hasil dari proses data mining menggunakan algoritma *K-Nearest Neighbor* dengan *rapid miner*.

e) Kesimpulan dan Saran

Menarik kesimpulan dari hasil penelitian dan memberikan rekomendasi kepada perusahaan untuk perbaikan selanjutnya serta kepada penelitian yang selanjutnya agar dapat memanfaatkan penelitian ini sebagai tinjauan penelitian berikutnya.

2.2. Objek Penelitian

Objek penelitian dilakukan terhadap usaha *marketplace shopee* yaitu *Erigo Store* yang berlokasi di Legok, Kecamatan Legok, Kabupaten Tangerang, Banten. Perusahaan ini memiliki banyak kategori produk yang dijual. Berdasarkan banyaknya jumlah konsumen maka diperlukan prediksi penjualan produk yang akan dijual untuk membantu perusahaan merencanakan persediaan produk yang dijual yang akan diprediksi berdasarkan data penjualan dua tahun terakhir.

2.3. Waktu Penelitian

Penelitian ini dilakukan selama satu bulan yaitu pada bulan Agustus 2022 sampai bulan September 2022.

2.4. Jenis Penelitian

Dalam penelitian ini digunakan jenis penelitian kuantitatif. Metode penelitian kuantitatif adalah metode untuk menjawab pertanyaan penelitian yang berkaitan dengan data berupa angka dan program statistik [10]. Dengan metode ini, diharapkan dapat mempermudah proses penelitian.

2.5. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam membuat laporan penelitian ini adalah :

a) Observasi Review Produk



Metode observasi *review* produk adalah metode atau pendekatan untuk memperoleh data dengan cara melihat ulasan singkat berdasarkan fakta penilaian dari pelanggan yang telah membeli produk [11]. Dalam pengumpulan data ini dilakukan melalui *review* pelanggan *Erigo Store* pada *marketplace shopee*.

b) Dokumentasi

Dokumentasi adalah cara peneliti kualitatif untuk memperoleh informasi umum dari sudut pandang subjek dengan menggunakan sumber tertulis dan dokumen lain yang ditulis atau dibuat langsung oleh subjek yang bersangkutan [12]. Yang dimaksud data dokumentasi disini adalah sejumlah data penjualan produk yaitu dari Januari tahun 2021 sampai dengan Juni 2023.

2.6. Analisis Data

Analisis data adalah proses pengumpulan data secara sistematis untuk membantu peneliti menarik kesimpulan [13]. Adapun tahapan analisis data adalah sebagai berikut :

1. Data Selection

Data selection adalah pemilihan (selection) data dari sekumpulan data operasional. Data yang dipilih digunakan untuk proses data mining dan disimpan dalam file terpisah dari basis data operasional [14]. Data yang digunakan dalam penelitian ini adalah data penjualan berdasarkan penjualan bulan Januari 2021 sampai dengan Juni 2023 yang berasal dari *marketplace shopee Erigo Store*. Data tersebut kemudian dipilih dan diolah untuk memprediksi penjualan produk. Karakteristik yang digunakan untuk menentukan ramalan penjualan produk adalah kategori produk, kuantitas, serta bulan dan tahun.

2. Preprocessing

Tujuan preprocessing dalam data mining adalah mengubah data menjadi format yang lebih sederhana dan lebih efisien untuk kebutuhan pengguna[15]. Tahap preprocessing difungsikan untuk memudahkan proses perhitungan prediksi, digunakan pengelompokan jenis penjualan produk berdasarkan kategori dan jumlah penjualan per bulan dan tahun. Setelah mengkategorikan data, semua data dikumpulkan untuk dijumlahkan total penjualan untuk semua kategori produk.

3. HASIL DAN PEMBAHASAN

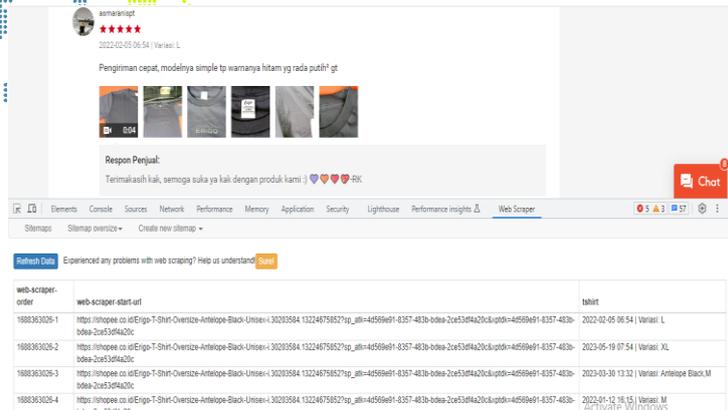
3.1. Hasil

Setelah menganalisis rancangan penelitian menggunakan data mining untuk menghitung hasil penjualan yang diharapkan untuk produk *Erigo Store* di *marketplace shopee* menggunakan algoritma *K-Nearest Neighbor*, penelitian ini diakhiri dengan proses data mining, maka hasil yang dicapai oleh peneliti adalah mengetahui prediksi penjualan produk *Erigo Store* pada *marketplace shopee* berdasarkan data penjualan bulan Januari tahun 2021 sampai dengan Juni 2023. Peneliti juga dibantu oleh *tools Rapid Miner* untuk mempermudah dalam proses data mining tersebut.

3.2. Pembahasan

3.2.1. Metode Web Scraping

Web scraping adalah metode yang digunakan dalam penelitian dalam memperoleh informasi data dari website secara otomatis tanpa harus menyalin secara manual sehingga memudahkan dalam melakukan pencarian sesuatu. Tujuan web scraper yaitu untuk mendapatkan informasi data tertentu dengan cara mengambil dan mengekstraksi data dengan ukuran yang bervariasi.



Gambar 2. Hasil Web Scraper

3.2.2. Data Selection

Data yang dipakai untuk diseleksi dalam penelitian adalah data penjualan produk *Erigo Store* pada *Marketplace Shopee* berdasarkan penjualan dari bulan Januari tahun 2021 sampai dengan Juni 2023. Kemudian data tersebut akan diolah dalam memprediksi penjualan *Erigo Store*.

Tabel 1. Data Selection Penjualan Erigo Store

No.	Nama Produk	Kuant	Bulan
1	Erigo T Shirt Project Summer	159	JANUARI
2	Erigo T Shirt Longboard Green Unisex	98	JANUARI
3	Erigo Chino Short Pants Light Color Unisex	50	JANUARI
.....
1167	Erigo Flannel Vilga Brown Unisex	26	JUNI

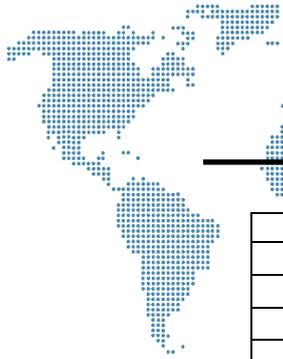
3.2.3. Preprocessing

Tahapan preprocessing untuk mengkategorikan penjualan *Erigo Store* berdasarkan jumlah penjualan per tahun untuk membantu peneliti menyelesaikan proses prediksi. Setelah dilakukan pengkategorian data, semua kategori data dijumlahkan menjadi data penjualan untuk setiap kategori produk *Erigo Store*.

Tabel 2. Sampel Data Penjualan T-Shirt

Bulan	Tahun 2021	Tahun 2022	Tahun 2023
Januari	257	2078	523
Februari	160	794	529
Maret	405	1388	675
April	886	1793	1222
Mei	2534	484	484
Juni	1093	254	607
Juli	837	325	
Agustus	660	318	
September	1099	488	
Oktober	1196	465	
November	1330	293	
Desember	3094	440	

Berdasarkan Tabel 2 hasil dari jumlah penjualan kategori T-Shirt setiap bulan dan tahun dengan total sebanyak 22671 produk.



Tabel 3. Sampel Data Penjualan Pants

Bulan	Tahun 2021	Tahun 2022	Tahun 2023
Januari	914	512	419
Februari	998	409	365
Maret	1355	1026	208
April	2866	964	437
Mei	4565	109	123
Juni	1333	295	117
Juli	975	294	
Agustus	857	157	
September	1454	210	
Oktober	1192	421	
November	1468	352	
Desember	1648	437	

Berdasarkan Tabel 3 hasil dari jumlah penjualan kategori Pants setiap bulan dan tahun dengan total sebanyak 24811 produk.

Tabel 4. Sampel Data Penjualan Jacket

Bulan	Tahun 2021	Tahun 2022	Tahun 2023
Januari	2207	569	222
Februari	2129	379	163
Maret	1821	617	80
April	1882	520	85
Mei	3205	89	62
Juni	1146	129	219
Juli	576	508	
Agustus	667	293	
September	1332	173	
Oktober	420	450	
November	926	383	
Desember	1353	296	

Berdasarkan Tabel 4 hasil dari jumlah penjualan kategori Jacket setiap bulan dan tahun dengan total sebanyak 22901 produk.

Tabel 5. Sampel Data Penjualan Outwear

Bulan	Tahun 2021	Tahun 2022	Tahun 2023
Januari	102	1556	518
Februari	609	545	596
Maret	1291	699	633
April	1861	701	900
Mei	2746	426	310
Juni	1248	457	386
Juli	1159	621	
Agustus	1091	351	
September	2681	222	
Oktober	1704	287	
November	1415	308	
Desember	3385	354	

Berdasarkan Tabel 5 hasil dari jumlah penjualan kategori Outwear setiap bulan dan tahun dengan total sebanyak 29162 produk.

Tabel 6. Sampel Data Penjualan Shirt

Bulan	Tahun 2021	Tahun 2022	Tahun 2023
Januari	162	353	1256
Februari	82	410	346
Maret	24	1436	48
April	13	3455	67
Mei	1959	1568	789
Juni	1248	1198	764
Juli	881	1602	
Agustus	531	1439	
September	1439	1303	
Oktober	835	752	
November	300	632	
Desember	721	1561	

Berdasarkan Tabel 3.6 hasil dari jumlah penjualan kategori Pants setiap bulan dan tahun dengan total sebanyak 27174 produk.

3.2.4. Transformation

Pada tahapan *transformation* hasil preprocessing data kemudian digunakan untuk data training. Proses ini juga dapat membentuk data training berdasarkan data yang harus diseleksi, sebelumnya untuk mengidentifikasi atribut yang dapat mempengaruhi penjualan produk *Erigo Store*, yaitu data target, dimana data target adalah data yang memiliki atribut yang relevan untuk mendukung dalam proses *data mining*.

Tabel 7. Data Training Penjualan T-Shirt

No.	Bulan-1	Bulan-2	Bulan-3	Bulan-4	Bulan-5	Target
1.	257	160	405	886	2534	1093
2.	160	405	886	2534	1093	837
3.	405	886	2534	1093	837	660
4.	886	2534	1093	837	660	1099
5.	2534	1093	837	660	1099	1196
6.	1093	837	660	1099	1196	1330
7.	837	660	1099	1196	1330	3094
8.	660	1099	1196	1330	3094	2078
9.	1099	1196	1330	3094	2078	794
10.	1196	1330	3094	2078	794	1388
11.	1330	3094	2078	794	1388	1793
12.	3094	2078	794	1388	1793	484
13.	2078	794	1388	1793	484	254
14.	794	1388	1793	484	254	325
15.	1388	1793	484	254	325	318
16.	1793	484	254	325	318	488
17.	484	254	325	318	488	465
18.	254	325	318	488	465	293
19.	325	318	488	465	293	440

Berdasarkan Tabel 7 *data training* yang digunakan pada penelitian ini diambil dari dua tahun terakhir yaitu data penjualan tahun 2021 dan 2022. *Data training* dibagi menjadi 2 bagian yaitu data input dan target. Pada Tabel di atas data input merupakan data penjualan dari bulan ke-1 sampai bulan ke-5, sedangkan dengan data target menggunakan data bulan ke-6. Dan untuk bulan ke-2 sampai bulan ke-6 digunakan sebagai data input dengan target atau keluarannya bulan ke-7 dan seterusnya sampai batas data yang ada. Pada *data testing* yang digunakan untuk prediksi ditunjukkan pada Tabel berikut :

Tabel 8. Data Testing Penjualan T-Shirt

Bulan-1	Bulan-2	Bulan-3	Bulan-4	Bulan-5
1.388	1.793	484	254	325
1.793	484	254	325	318
484	254	325	318	488
254	325	318	488	465
325	318	488	465	293
318	488	465	293	440
488	465	293	440	523
465	293	440	523	529
293	440	523	529	675
440	523	529	675	1.222
523	529	675	1.222	484
529	675	1.222	484	607

Berdasarkan Tabel 8 hasil data yang diuji adalah data penjualan dari bulan Maret 2022 - Juni 2023 yang akan menghasilkan prediksi penjualan.

3.2.5. K-Nearest Neighbor

Adapun langkah-langkah dari tahap data mining untuk perhitungan algoritma *K-Nearest Neighbor* dalam penelitian ini yaitu :

1. Penentuan nilai k yang digunakan pada penelitian ini adalah 3.
2. Menghitung jarak antara *data training* dan data uji yang ada pada tahapan transformation menggunakan *Euclidean Distance* :

$$d_1 = \sqrt{(257-318)^2 + (160-488)^2 + (405-465)^2 + (886-293)^2 + (2534-440)^2} = 2.202,59773$$

$$d_2 = \sqrt{(160-318)^2 + (405-488)^2 + (886-465)^2 + (2534-293)^2 + (1093-440)^2} = 2.375,10926$$

$$d_3 = \sqrt{(405-318)^2 + (886-488)^2 + (2534-465)^2 + (1093-293)^2 + (837-440)^2} = 2.290,05305$$

$$d_4 = \sqrt{(886-318)^2 + (2534-488)^2 + (1093-465)^2 + (837-293)^2 + (660-440)^2} = 2.290,73350$$

$$d_5 = \sqrt{(2534-318)^2 + (1093-488)^2 + (837-465)^2 + (660-293)^2 + (1099-440)^2} = 2.446,16741$$

$$d_6 = \sqrt{(1093-318)^2 + (837-488)^2 + (660-465)^2 + (1099-293)^2 + (1196-440)^2} = 1.047,70131$$

$$d_7 = \sqrt{(837-318)^2 + (660-488)^2 + (1099-465)^2 + (1196-293)^2 + (1330-440)^2} = 1.519,34525$$

$$d_8 = \sqrt{(660-318)^2 + (1099-488)^2 + (1196-465)^2 + (1330-293)^2 + (3094-440)^2} = 3.023.86028$$

$$d_9 = \sqrt{(1099-318)^2 + (1196-488)^2 + (1330-465)^2 + (3094-293)^2 + (2078-440)^2} = 3.519,67214$$

$$d_{10} = \sqrt{(1196-318)^2 + (1330-488)^2 + (3094-465)^2 + (2078-293)^2 + (794-440)^2} = 3.420,96916$$

$$d_{11} = \sqrt{(1330-318)^2 + (3094-488)^2 + (2078-465)^2 + (794-293)^2 + (1388-440)^2} = 3.401,00779$$

$$d_{12} = \sqrt{(3094-318)^2 + (2078-488)^2 + (794-465)^2 + (1388-293)^2 + (1793-440)^2} = 3.656,79518$$

$$d_{13} = \sqrt{(2078-318)^2 + (794-488)^2 + (1388-465)^2 + (1793-293)^2 + (484-440)^2} = 2.509,00398$$

$$d_{14} = \sqrt{(794-318)^2 + (1388-488)^2 + (1793-465)^2 + (484-293)^2 + (254-440)^2} = 1.694,47248$$

$$d_{15} = \sqrt{(1388-318)^2 + (1793-488)^2 + (484-465)^2 + (254-293)^2 + (325-440)^2} = 1.692,04964$$

$$d_{16} = \sqrt{(1793-318)^2 + (484-488)^2 + (254-465)^2 + (325-293)^2 + (318-440)^2} = 1.495,32805$$

$$d_{17} = \sqrt{(484-318)^2 + (254-488)^2 + (325-465)^2 + (318-293)^2 + (488-440)^2} = 323,79159$$

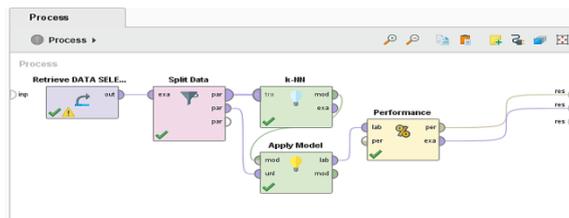
$$d_{18} = \sqrt{(254-318)^2 + (325-488)^2 + (318-465)^2 + (488-293)^2 + (465-440)^2} = 301,53606$$

$$d_{19} = \sqrt{(325-318)^2 + (318-488)^2 + (488-465)^2 + (465-293)^2 + (293-440)^2} = 284,02640$$

- Jarak yang dihasilkan kemudian diurutkan dari terdekat ke yang terjauh jaraknya (*ascending*). Setelah dilakukan pengurutan didapatkan :
 $d_{19} = 284,02640$, $d_{18} = 301,53606$, $d_{17} = 323,79159$, $d_6 = 1.047,70131$, $d_{16} = 1.495,32805$, $d_7 = 1.519,34525$, $d_{15} = 1.692,04964$, $d_{14} = 1.694,47248$, $d_1 = 2.202,59773$, $d_3 = 2.290,05305$, $d_4 = 2.290,73350$, $d_2 = 2.375,10926$, $d_5 = 2.446,16741$, $d_{13} = 2.509,00398$, $d_8 = 3.023.86028$, $d_{11} = 3.401,00779$, $d_{10} = 3.420,96916$, $d_9 = 3.519,67214$, $d_{12} = 3.656,79518$.
- Tentukan hasil kelompok data uji berdasarkan mayoritas data dari k tetangga terdekat. Tiga jarak terkecil yaitu d_{19} , d_{18} dan d_{17} diperoleh dari nilai $k = 3$.

3.2.6. Implementasi Rapid Miner

Setelah data selection diperoleh, kemudian data di masukkan ke *tools rapid miner* dan dilakukan *split* data dengan ratio pengujian 10 % dari jumlah data yang ada lalu dihubungkan dengan *operator K-Nearest Neighbor* dengan nilai $k = 3$, setelah itu sambungkan dengan *operator apply model* dan hubungkan ke *operator performance* untuk mendapatkan nilai akurasi.



Gambar 3. Subproses Pada Operator

Untuk nilai akurasi diterapkan pada model diatas dapat dilihat pada Gambar 3:

	true Kategori	true T Shirt	true Pants	true Jacket	true Outwear	true Shirt	class preci...
pred Kategori	0	0	0	0	0	0	0.00%
pred T Shirt	0	21	1	2	6	5	60.00%
pred Pants	0	0	24	0	0	2	92.31%
pred Jacket	0	0	0	19	0	1	95.00%
pred Outwear	0	0	1	0	17	1	89.47%
pred Shirt	0	0	0	0	0	16	100.00%
class recall	0.00%	100.00%	92.31%	90.48%	73.91%	64.00%	

Gambar 4. Performance Matriks

Berdasarkan Gambar 4 hasil dari model yang ditetapkan oleh *Rapid Miner* untuk membentuk pola pemodelan dengan keakuratan datanya yaitu sebesar 83,62 %.

	True: Kategori	T Shirt	Pants	Jacket	Outwear	Shirt
Kategori:	0	0	0	0	0	0
T Shirt :	0	21	1	2	6	5
Pants:	0	0	24	0	0	2
Jacket:	0	0	0	19	0	1
Outwear:	0	0	1	0	17	1
Shirt:	0	0	0	0	0	16

Gambar 5. Model Klasifikasi KNN

Berdasarkan Gambar 5 memaparkan pemodelan produk berdasarkan klasifikasi dalam metode *K-Nearest Neighbor* menggunakan *Rapid Miner*.

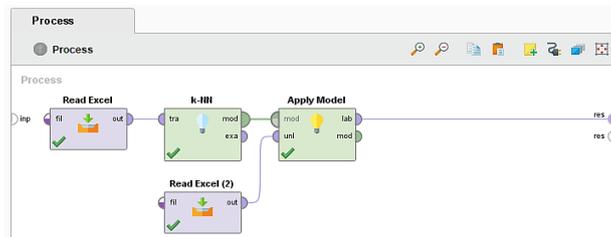
3.2.7. Susunan Operetor Algoritma *K-Nearest Neighbor*

Setelah pemodelan klasifikasi, selanjutnya yaitu *drag and drop operator read excel* kemudian lakukan *Importing Tabel Microsoft Excel* kedalam proses untuk membaca *data training didalam format excel*, ganti atribut target menjadi label.

	Bln-1 integer	Bln-2 integer	Bln-3 integer	Bln-4 integer	Bln-5 integer	Target integer label
1	257	160	405	886	2534	1093
2	160	405	886	2534	1093	837
3	405	886	2534	1093	837	660
4	886	2534	1093	837	660	1099
5	2534	1093	837	660	1099	1196
6	1093	837	660	1099	1196	1330
7	837	660	1099	1196	1330	3094
8	660	1099	1196	1330	3094	2078
9	1099	1196	1330	3094	2078	794
10	1196	1330	3094	2078	794	1388
11	1330	3094	2078	794	1388	1793
12	3094	2078	794	1388	1793	484

Gambar 6. Mengganti Atribut Target Menjadi Label

Lakukan *drag and drop operator K-NN* dengan memasukkan nilai $k=3$. Kemudian hubungkan *operator read excel* pada *operator K-NN*. Setelah itu hubungkan *K-NN* dengan *operator apply model*, kemudian *drag operator read excel (2)* dan *importing Tabel data testing*, hubungkan pada *operator apply model* setelah itu hubungkan pada *result*. Ditunjukkan pada Gambar 7.



Gambar 7. Subproses Operator Algoritma K-NN

3.2.8. Hasil Algoritma K-NN

Tahap terakhir dari *rapid miner* data mining adalah setelah semua terhubung dengan *operator*, lalu klik *icon play* di bagian atas *rapid miner* untuk menjalankan program, maka muncul hasil prediksi penjualan yang sesuai dengan nilai $k = 3$, yaitu :

Row No.	prediction(T...	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5
1	562.445	1388	1793	484	254	325
2	440.415	1793	484	254	325	318
3	413.189	484	254	325	318	488
4	372.266	254	325	318	488	465
5	403.209	325	318	488	465	293
6	398.921	318	488	465	293	440
7	398.610	488	465	293	440	523
8	399.158	465	293	440	523	529
9	394.534	293	440	523	529	675
10	697.702	440	523	529	675	1222
11	675.955	523	529	675	1222	484
12	410.613	529	675	1222	484	607

Gambar 8. Hasil Prediksi Penjualan T-Shirt



Berdasarkan Gambar 8 adalah hasil dari prediksi penjualan kategori t-shirt dan prediksi penjualan tertinggi terdapat pada bulan Oktober dengan jumlah sebanyak 697.

Row No.	prediction(T...	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5
1	217.585	1026	964	109	295	294
2	304.230	964	109	295	294	157
3	405.633	109	295	294	157	210
4	390.850	295	294	157	210	421
5	408.548	294	157	210	421	352
6	404.425	157	210	421	352	437
7	403.160	210	421	352	437	419
8	403.500	421	352	437	419	365
9	403.794	352	437	419	365	208
10	400.291	437	419	365	208	437
11	404.059	419	365	208	437	123
12	403.978	365	208	437	123	117

Gambar 9. Hasil Prediksi Penjualan Pants

Berdasarkan Gambar 9 adalah hasil dari prediksi penjualan kategori pants dan prediksi penjualan tertinggi terdapat pada bulan Mei dengan jumlah sebanyak 408.

Row No.	prediction(T...	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5
1	291.446	617	520	89	129	508
2	212.507	508	89	129	508	293
3	369.573	89	129	508	293	173
4	381.653	129	508	293	173	450
5	252.612	508	293	173	450	383
6	316.675	293	173	450	383	296
7	408.109	173	450	383	296	222
8	364.005	450	383	296	222	163
9	414.376	383	296	222	163	80
10	343.968	296	222	163	80	85
11	309.148	222	163	80	85	62
12	345.085	163	80	85	62	219

Gambar 10. Hasil Prediksi Penjualan Jacket

Berdasarkan Gambar 10 merupakan hasil prediksi penjualan kategori jacket dan prediksi penjualan tertinggi terdapat pada bulan September dengan jumlah sebanyak 414.

Row No.	prediction(T...	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5
1	395.449	699	701	426	457	621
2	288.009	701	426	457	621	351
3	373.702	426	457	621	351	222
4	314.247	457	621	351	222	287
5	311.923	621	351	222	287	308
6	318.409	351	222	287	308	354
7	316.818	222	287	308	354	518
8	287.702	287	308	354	518	596
9	375.551	308	354	518	596	633
10	403.270	354	518	596	633	900
11	373.237	518	596	633	900	310
12	436.138	596	633	900	310	386

Gambar 11. Hasil Prediksi Penjualan Outwear

Berdasarkan Gambar 11 merupakan hasil prediksi penjualan kategori outwear dan prediksi penjualan tertinggi terdapat pada bulan Desember dengan jumlah sebanyak 436.



Row No.	prediction(T...	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5
1	1063.985	1436	3455	1568	1198	1602
2	1088.670	3455	1568	1198	1602	1439
3	733.653	1568	1198	1602	1439	1383
4	606.679	1198	1602	1439	1383	752
5	622.572	1602	1439	1383	752	632
6	1620.998	1439	1383	752	632	1561
7	634.522	1383	752	632	1561	1256
8	637.258	752	632	1561	1256	364
9	1054.070	632	1561	1256	364	48
10	746.508	1561	1256	364	48	67
11	1787.471	1256	364	48	67	789
12	1955.026	364	48	67	789	764

Gambar 12. Hasil Prediksi penjualan Shirt

Berdasarkan Gambar 12 merupakan hasil prediksi penjualan kategori shirt dan prediksi penjualan tertinggi terdapat pada bulan September dengan jumlah sebanyak 1955.

4. SIMPULAN

Berdasarkan hasil penelitian ini, pemodelan yang dilakukan dengan menerapkan algoritma *K-Nearest Neighbor* dimana data yang diolah berdasarkan tahapan *Knowledge Discovery in Database* (KDD). Adapun hasil dari penelitian dengan penerapan data mining maka hasil yang diperoleh dari prediksi penjualan untuk 5 kategori produk. Dataset yang digunakan mulai tahun 2021 sampai 2023 pada *Erigo Store* dari *Marketplace Shopee*. Kategori data yang terjual yaitu T-Shirt, Pants, Jacket, Outwear, dan Shirt. Adapun hasil dari setiap kategori yang memiliki akurasi kategori Shirt sebesar 100%, kategori Jacket 95%, kategori Pants 92,31%, kategori Outwear 89,47% dan kategori T-Shirt 60%. Dari total keseluruhan prediksi kategori penjualan diperoleh hasil akurasi klasifikasi penjualan keseluruhan produk sebesar 83,62% dengan pengujian menggunakan *tools rapid miner* untuk menentukan penjualan setiap produk per tahunnya.

Berdasarkan kesimpulan yang telah dibuat, maka ada beberapa saran yang dapat disampaikan yaitu pada penelitian selanjutnya akurasi penjualan dapat diprediksi dengan melakukan pengujian menggunakan algoritma klasifikasi lainnya seperti *neural network*, *random forest*, menambah kategori yang belum diprediksi dan lain-lain pada produk *Erigo Store* untuk mendapatkan hasil yang lebih beragam dan tentunya memberikan informasi yang bernilai. Pada penelitian selanjutnya juga dapat mengimplementasikan dalam bentuk sistem yang dapat membantu dalam perhitungan prediksi penjualan produk *Erigo Store* sehingga perusahaan lebih mudah mengetahui tingkat penurunan dan kenaikan penjualan yang dimanfaatkan dalam mendukung penyediaan produk *Erigo Store*.

DAFTAR PUSTAKA

- [1] Rahayu Elly, Nurwati, Dewi Puspita Sri. 2022. "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor ". Building of Informatics, Technology and Science (BITS). Volume 3, No 4, Maret 2022 Page : 639 – 648.
- [2] Paruntu, Stacia. 2018. "Analisis Ramalan Penjualan Dan Persediaan Produk Sepeda Motor Suzuki Pada PT. Sinar Galesong Mandiri Malalayang". Jurnal EMBA Vol 6. 4 September 2018. ISSN 2303-1174.

-
- [3] Alghifari, Fauzan, Didi Juardi. 2021. "*Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes*". Jurnal Ilmiah Informatika (JIF). ISSN (Print) 2337-8379. ISSN (Online) 2615-1049.
 - [4] Bahri, Syaiful, Akhyar Lubis. 2020. "*Metode Klasifikasi Decision Tree Untuk Memprediksi Juara English Premier League*". Jurnal Sintaksis : Pendidikan Guru Sekolah Dasar, IPA, IPS, Dan Bahasa Inggris. Vol 2, No.1, April 2020.
 - [5] Khairi, Ahmad, Dkk. 2021. "*Implementasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Masyarakat Pra Sejahtera Desa Sarikerep Kecamatan Sukapura*". Tritogi : Jurnal Ilmu Teknologi, Kesehatan, Dan Humaniora, 2 (3), September-Desember 2021.
 - [6] Hardiyanto, Bagus, Fahrur Rozi. 2020. "*Prediksi Penjualan Sepatu Menggunakan Metode K-Nearest Neighbor*". JOEICT (Jurnal Of Education And Information Communication Technology). Volume 04, Nomor 02, Desember 2020 : 13-18. ISSN 2459-993X.
 - [7] Aisha, Alfani, dkk. 2021. "*Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor*". JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika). Volume 06, Nomor 01, Juni 2021 : 155-160.
 - [8] Hermawan, Feny. 2017. "*Implementasi Metode K-Nearest Neighbor Pada Aplikasi Data Penjualan Pt.Multitek Mitra Sejati*". Universitas Bunda Mulia Kalbiscentia, ISSN 2356-4393. Volume 4, Nomor 2. Agustus 2017.
 - [9] shopee.co.id/erigostore#product_link.
 - [10] Aryani, Deviacita, dkk. 2019. "*Implementasi Web Scraping Untuk Pengambilan Data Pada Situs Marketplace*". Jurnal Sistem Dan Teknologi Informasi. Vol.7, No.4, Oktober 2019. p-ISSN : 2460-3562/e-ISSN : 2620-8989.
 - [11] Wahidmurni, 2017. "Pemaparan Metode Penelitian Kuantitatif". UIN Maulana Malik Ibrahim Malang Juni 2017.
 - [12] Ramadhani, Tiara. 2021. "Pengaruh Rating Dan Online Consumer Review Terhadap Minat Beli Dengan Kepercayaan Sebagai Pemediator Pada Marketplace Shopee Dalam Perspektif Ekonomi Islam". Fakultas Ekonomi Dan Bisnis Islam Universitas Islam Negeri Raden Intan Lampung 2021.
 - [13] Herdiansyah, Haris. 2010. "Metodologi Penelitian Kualitatif". Fakultas Tarbiyah Iain Walisongo, (Jakarta: Salemba Humanika, 2010).
 - [14] Moleong, Lexy J. 2021. "Metodologi Penelitian Kualitatif". Jakarta: Universitas Indonesia Press 2021.
 - [15] Dan, Bi Thi Thant, dkk. 2015. "Implementasi Iterative Dichotomiser 3 Pada Data Kelulusan Mahasiswa S1 Di Universitas Sebelas Maret". Jurnal ITSMART. Vol 4. No.2. Desember 2015. ISSN :2301 – 7201.
 - [16] Meilina, Popy. 2014. "*Penerapan Data Mining Dengan Metode Kalsifikasi Menggunakan Decision Tree Dan Regresi*". Jurnal Teknologi Universtias Muhammadiyah Jakarta. Volume 7 No.1 Januari 2015 ISSN : 2085 – 1669 e-ISSN : 2460 – 028.