

## Penerapan Data Mining dengan Metode *K-Nearest Neighbor* untuk Memprediksi Penjualan Aksesoris Aquarium

Charles Chaywin<sup>1</sup>, Yudhistira Adhitya Pratama<sup>2\*</sup>

<sup>1,2,3</sup>Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pelita Harapan, Medan, Indonesia

Email: <sup>1</sup>03081190002@student.uph.edu, <sup>2\*</sup>yudhistira.adhitya@lecturer.uph.edu

### ABSTRACT

*PT. Surya Jaya Aquarium is an industry engaged in the sale of various types of aquarium accessories, such as aquarium machines, filter media, aquarium lights. Currently, PT. Surya Jaya Aquarium often lacks certain items when ordering goods from customers. Meanwhile, there is often an excess of other goods at the same time due to a lack of orders from customers. For this reason, it is necessary to carry out the process of predicting product sales at the company, so that the process of controlling product orders can be carried out. To carry out the prediction process, the K-Nearest Neighbor algorithm can be applied. The purpose of this algorithm is to classify new objects using features and training data samples. The data used in this study is sales data for aquarium accessories products sourced from sales for the last 3 years from 2020, 2021 and 2022 originating from PT. Surya Jaya Aquarium. After that the data is selected and will be used to be processed in predicting sales of aquarium accessories for the next period. The K-Nearest Neighbor technique is used in this study to model data that has been prepared using the Knowledge Discovery in Databases (KDD) stage. From the results of the tests carried out, information was obtained that the error rate (error) from the sales prediction results was 6,196%.*

*Keywords: Aquarium Accessories, Data Mining, K-Nearest Neighbor.*

### ABSTRAK

PT. Surya Jaya Aquarium adalah salah satu industri yang bergerak di bidang penjualan berbagai jenis aksesoris aquarium, seperti mesin-mesin aquarium, media filter, lampu aquarium. Saat ini, PT. Surya Jaya Aquarium sering kekurangan barang tertentu pada saat adanya pemesanan barang dari customer. Sementara itu sering kelebihan barang lainnya pada saat yang sama karena kurangnya pemesanan dari customer. Untuk itu, maka perlu dilakukan proses prediksi penjualan produk pada perusahaan, sehingga dapat dilakukan proses pengontrolan pemesanan produk. Untuk melakukan proses prediksi, maka dapat diterapkan algoritma K-Nearest Neighbor. Tujuan algoritma ini adalah untuk mengklasifikasikan objek baru menggunakan ciri dan sampel data training. Data yang digunakan dalam penelitian ini ialah data penjualan produk aksesoris aquarium bersumber pada penjualan 3 tahun terakhir dari tahun 2020, 2021, serta 2022 yang berasal dari PT. Surya Jaya Aquarium. Setelah itu data tersebut dipilih serta akan digunakan untuk diolah dalam memprediksi penjualan aksesoris aquarium untuk periode berikutnya. Teknik K-Nearest Neighbor digunakan dalam penelitian ini untuk memodelkan data yang telah disiapkan menggunakan tahap Knowledge Discovery in Databases (KDD). Dari hasil pengujian yang dilakukan, diperoleh informasi bahwa tingkat kesalahan (error) dari hasil prediksi penjualan adalah sebesar 6.196%.

Kata Kunci: Aksesoris Aquarium, Data Mining, K-Nearest Neighbor.

### 1. Pendahuluan

Seiring dengan kemajuan zaman, perkembangan ilmu pengetahuan dan teknologi semakin cepat. Setiap bagian dari kehidupan dipengaruhi oleh teknologi, yang tumbuh lebih maju dan signifikan. Teknologi yang ada diharapkan mampu memenuhi setiap kebutuhan informasi dan mendukung setiap tugas manusia serta mampu menyelesaikan setiap masalah yang mungkin

timbul, terutama untuk memperluas bisnis [1], [2]. Karena persaingan yang ketat dalam bisnis, khususnya di sektor penjualan, penting bagi pemilik bisnis untuk membuat rencana yang akan meningkatkan upaya penjualan dan pemasaran produk [3]. Salah satu strategi tersebut adalah dengan menggunakan data penjualan setiap hari, karena ini akan menghasilkan jumlah data yang terus meningkat. Informasi ini dapat digunakan dan diproses untuk memberikan informasi

yang berarti untuk meningkatkan penjualan dan iklan produk, bukan hanya berfungsi sebagai arsip bisnis.

Data mining adalah metode yang mengekstraksi dan mengidentifikasi informasi yang relevan dan pengetahuan terkait dari *data Base* berukuran besar menggunakan pendekatan statistik, teknik matematika, kecerdasan buatan, dan pembelajaran mesin [4]. Untuk mengumpulkan informasi yang dapat digunakan untuk mengantisipasi penjualan, data mining dapat digunakan untuk mengekstraksi informasi dari kumpulan data yang sangat besar [5]. Pendekatan data mining juga dapat digunakan untuk mendukung audit menyeluruh terhadap data penjualan dan proses bisnis terkait guna membantu memastikan integritas, keamanan, dan kualitas data yang digunakan dalam pengambilan keputusan [6]. Teknik melakukan data mining dapat menggunakan berbagai pendekatan, seperti klusterisasi, regresi, asosiasi, dan klasifikasi, untuk mengungkapkan pola atau informasi yang tersembunyi [7]. Cara kerja data mining melibatkan pencarian melalui kumpulan data besar untuk pola atau struktur yang segar dan praktis. Namun, tidak setiap upaya pencarian informasi termasuk dalam kategori data mining.

Metode-metode dalam melakukan data mining untuk prediksi terdapat beberapa jenis yaitu *Support Vector Machine* (SVM), *Apriori*, *Naïve Bayes*, *K-Nearest Neighbor* (KNN), *C4.5*. *K-Nearest Neighbor* merupakan teknik untuk mengategorikan sesuatu berdasarkan data sampel yang jaraknya paling dekat dengan barang tersebut. Hanya data sampel yang disimpan dan diklasifikasikan menggunakan teknik ini. Metode *K-Nearest Neighbor* memiliki kelebihan yaitu efektif bila diterapkan pada data yang cukup banyak dan menghasilkan data yang kuat atau jelas.

PT. Surya Jaya Aquarium adalah salah satu industri yang bergerak di bidang penjualan berbagai jenis aksesoris aquarium seperti : mesin-mesin aquarium, media filter, lampu aquarium. PT. Surya Jaya Aquarium merupakan distributor utama sehingga penjualan barang secara langsung kepada toko-toko aquarium di Indonesia bukan melalui toko online atau marketplace.

Berikut ini adalah tujuan dari penelitian ini berdasarkan bagaimana masalah tersebut dijelaskan di atas:

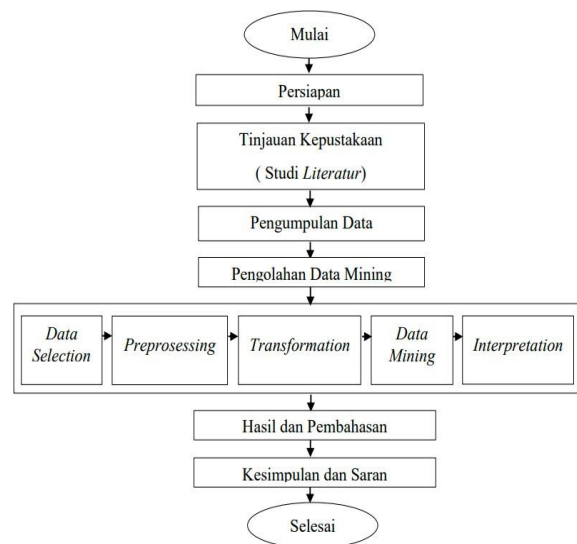
- Untuk mengetahui informasi prediksi hasil penjualan periode berikutnya pada PT. Surya Jaya Aquarium.
- Untuk mengetahui seberapa akurat metode *K-Nearest Neighbor* dalam memprediksi penjualan aksesoris aquarium pada PT. Surya Jaya Aquarium

## 2. Metode Penelitian

### 2.1 Tahapan Penelitian

Diagram penelitian pada gambar 1, menggambarkan langkah-langkah dalam tahapan penelitian. Yang akan dijelaskan sebagai berikut:

1. Persiapan. Proses berbincang dengan pemilik perusahaan dan menjelaskan rumusan masalah yang terjadi pada perusahaan. Data barang yang diambil ialah data penjualan aksesoris aquarium. Tetapkan batasan dan buat strategi penelitian.
2. Tinjauan literatur. Tinjauan dan penelitian literatur yang membahas prediksi penjualan dan masalah terkait dilakukan dalam kesadaran literatur.
3. Pengumpulan data. Dilakukan dengan cara mewawancarai bagian akuntansi pada PT. Surya Jaya Aquarium, observasi, dan dokumentasi digunakan untuk memperoleh data penjualan aksesoris aquarium.
4. Pengolahan data mining. Data yang terkumpul diolah sesuai dengan metode yang digunakan berdasarkan tahapan proses KDD.
5. Hasil dan pembahasan. Teknik *K-Nearest Neighbor* untuk data mining digunakan dalam diskusi ini untuk menekankan temuan.
6. Kesimpulan dan saran. Menarik kesimpulan dari temuan penelitian dan menawarkan saran untuk bisnis agar menjadi lebih baik.



Gambar 1 Tahapan Penelitian

### 2.2 Prediksi

Prediksi adalah cabang ilmu yang meramalkan kejadian di masa depan. Menerapkan data masa lalu ke masa depan diperlukan untuk peramalan (Chairis & Maulana, 2022). Prediksi melibatkan mempertimbangkan kuantitas, seperti permintaan untuk satu atau lebih hal di masa mendatang [9].

### 2.3 Penjualan

Penjualan ialah suatu kegiatan yang ditujukan untuk mencari pembeli, mempengaruhi dan memberikan petunjuk agar pembeli dapat menyesuaikan kebutuhannya dengan produk yang ditawarkan serta mengadakan perjanjian mengenai harga yang menguntungkan bagi kedua belah pihak [10].

## 2.4 Basis Data

Basis data adalah sekelompok objek data yang terhubung yang disusun menurut skema atau struktur tertentu, disimpan di perangkat keras komputer, dan diproses dengan perangkat lunak memiliki tujuan tertentu seperti memperbarui, mencari, memproses dengan perhitungan tertentu, dan memusnahkan [4]

## 2.5 Data Mining

Data mining adalah teknik untuk mengidentifikasi dan mengekstraksi informasi yang dapat digunakan dan mengumpulkan pengetahuan dari kumpulan data besar menggunakan pendekatan matematika, statistik, kecerdasan buatan, dan pembelajaran mesin [5].

Data mining adalah tindakan menganalisis kumpulan data besar yang disimpan dalam penyimpanan menggunakan teknik pengenalan pola termasuk pendekatan statistik dan matematika untuk menemukan tautan, pola, dan tren yang relevan [11]. Data mining adalah kumpulan prosedur untuk menemukan informasi baru dari kumpulan data yang tidak dapat ditemukan secara manual.

## 2.6 Klasifikasi

Klasifikasi merupakan proses pembelajaran suatu fungsi tujuan (target)  $f$  yang memetakan tiap himpunan atribut  $x$  ke satu dari label kelas yang didefinisikan sebelumnya. Fungsi target disebut juga model klasifikasi [12].

## 2.7 K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* adalah metode pembelajaran terawasi di mana mayoritas *K-Nearest Neighbor* digunakan untuk mengklasifikasikan hasil jarak baru. Tujuan algoritma ini adalah untuk mengklasifikasikan objek baru menggunakan ciri dan sampel data training.

Algoritma *K-Nearest Neighbor* (KNN) mengklasifikasikan objek berdasarkan data pembelajaran yang terdekat dengan objeknya. Ruang multidimensi diproyeksikan ke data pembelajaran, dengan setiap dimensi sesuai dengan properti data [13]. Tujuan algoritma KNN adalah mengategorikan objek baru berdasarkan properti dan data training-nya. di mana sebagian besar kategori KNN digunakan untuk mengklasifikasikan hasil sampel uji baru. Klasifikasi yang berdekatan digunakan oleh algoritma KNN untuk meramalkan nilai sampel data uji baru. *Euclidean Distance* adalah pengukuran yang digunakan. Jarak yang paling khas untuk data numerik adalah *Euclidean Distance*.

$$\sqrt{\sum_{i=1}^K (X_i - Y_i)^2}$$

Nilai dari set data training dan data testing masing-masing diwakili oleh nilai  $X_i$  dan  $Y_i$ . Nilai  $K$  digunakan untuk menunjukkan dimensi atribut.

Perhitungan untuk algoritma *K-Nearest Neighbor* :

1. Tetapkan nilai  $k$ .
2. Dengan menggunakan data training yang tersedia, tentukan jarak *Euclidean* kuadrat setiap objek.
3. Kelompokkan item-item ini ke dalam grup berdasarkan mana mereka memiliki jarak *Euclidean* terdekat
4. Kumpulkan label untuk kelas  $Y$  (klasifikasi *Nearest Neighbor*).
5. Dengan memanfaatkan kategori *Nearest Neighbor* dengan nilai *instance query* terhitung paling akurat.

## 2.8 Pengolahan Data

Pengolahan Data Mining dilakukan dengan cara mengikuti tahapan proses Knowledge Discovery in Database (KDD) untuk menghasilkan informasi sesuai dengan urutan yang telah ditentukan, yaitu :

- *Data selection*

*Data selection* dari sekumpulan data operasional yang perlu dilakukan sebelum langkah tinjauan informasi dalam KDD dimulai. Data hasil pencarian akan digunakan untuk data mining dan akan disimpan dalam file terpisah dari *database*. Penelitian ini mengandalkan data dari PT. Surya Jaya Aquarium yaitu data penjualan aksesoris aquarium tahun 2020, 2021, dan 2022.

Atribut yang digunakan dalam proses ini adalah :

1. Nama Barang adalah atribut pada tabel data penjualan yang merinci produk yang dibeli dari PT. Surya Jaya Aquarium.
2. Bulan adalah atribut pada tabel data penjualan memberikan rincian tentang kapan terjadinya transaksi penjualan.
3. Kuantitas adalah atribut pada tabel penjualan yang memberikan perincian tentang jumlah transaksi penjualan.

- *Pengolahan data*

Proses preprocessing harus diselesaikan sebelum proses data mining dapat dimulai. Pada tahap ini akan dilakukan integrasi data untuk menggabungkan data dari database yang berbeda, dan pembersihan data akan dilakukan untuk membuat dataset bersih yang dapat digunakan pada tahap berikutnya yaitu mining [14]. Kedua langkah tersebut dijelaskan secara rinci di bawah ini:

1. *Integrasi data*

Integrasi data adalah proses menyatukan kumpulan data yang berbeda untuk dianalisis. Basis data, kubus data, dan file adalah contoh sumber data. Integrasi data ini juga diperlukan untuk meningkatkan akurasi dan mempercepat proses data mining. Prosedur ini juga mencakup Pembersihan Data Ekstra, yang digunakan untuk menghilangkan file duplikat atau kesalahan dari data terintegrasi.

2. Data cleaning

Data cleaning adalah tindakan memodifikasi data untuk menilai kualitas data. Dapat juga menghapus data atau mengoreksi data. Data yang harus dibersihkan antara lain informasi yang salah, tidak lengkap, rusak, dan tidak sesuai format.

• Transformation

Pada tahap *Transformation*, data yang dipilih diubah agar cocok untuk data mining. Jenis atau pola informasi yang akan dicari dalam *database* memiliki pengaruh yang signifikan terhadap proses transformasi dalam KDD yang merupakan proses kreatif.

Pengelompokan data atribut, khususnya atribut bulan dan klasifikasi sebagai kriteria data yang menjadi target dalam proses mining, dikumpulkan pada seluruh tahapan data operasional yang digunakan untuk proses transformasi data mining.

• Data mining

Pada langkah ini digunakan teknik atau proses berdasarkan proses KDD yang lebih besar untuk mencari pola atau informasi yang menarik pada data yang dipilih. Teknik *K-Nearest Neighbor (KNN)*, yang karakteristiknya diinisialkan sebagai *k*, atau jumlah tetangga yang digunakan sebagai referensi *KNN*, digunakan dalam penelitian ini. *K* adalah bilangan bulat positif kecil yang ganjil.

Dalam penelitian ini, langkah-langkah pengerjaan pendekatan *KNN* adalah sebagai berikut:

1. Menentukan nilai *k*. Angka yang digunakan dalam penelitian ini untuk *k* yang digunakan dalam klasifikasi adalah 3, 5, 7, 9, 11, dan 13 karena tidak ada standar yang ditetapkan untuk perhitungan ini.
2. Penghitungan jarak antara data Training dan testing. *Euclidean Distance* adalah metodologi perhitungan jarak yang digunakan dalam algoritma *KNN* ini.
3. Urutan dari data yang dihitung. Jarak yang dihitung kemudian diatur secara menaik berdasarkan kedekatan dengan yang terjauh.
4. Mengelompokkan informasi hasil tes berdasarkan label mayoritas *K-Nearest Neighbor*.

• Evaluation

Pola informasi yang ditemukan melalui data mining harus disajikan dengan cara yang jelas bagi pihak yang berkepentingan. Pada fase ini, penting untuk menilai apakah pola atau informasi yang diamati bertentangan dengan pengetahuan atau teori sebelumnya. Pada tahap ini digunakan metode *K-Nearest Neighbor* untuk mengumpulkan pola penjualan aksesoris aquarium dari proses data mining, dan pola atau informasi yang dihasilkan dari proses data mining berupa aturan yang diperoleh dari perhitungan *K-Nearest Neighbor*.

3. Hasil dan Pembahasan

3.1 Hasil Pengumpulan Data

3.1.1 Data Selection

Data yang digunakan dalam penelitian ini ialah data penjualan produk aksesoris aquarium bersumber pada penjualan 3 tahun terakhir dari tahun 2020, 2021, serta 2022 yang berasal dari PT. Surya Jaya Aquarium.

Setelah itu data tersebut dipilih serta akan digunakan untuk diolah dalam memprediksi penjualan aksesoris aquarium terlaris. Adapun atribut yang digunakan dalam penentuan prediksi penjualan aksesoris terlaris ialah atribut Nama Barang, Kuantitas, dan Bulan. Bisa dilihat pada Tabel 1.

Tabel 1. Data Selection

No	Nama Barang	Bulan	Kuantitas
1	Yang YP-103	Januari	264
2	Yang YP-103	Februari	168
3	Yang YP-103	Maret	0
4	Yang YP-103	April	480
5	Yang YP-103	Mei	0
6	Yang YP-103	Juni	264
7	Yang YP-103	Juli	72
8	Yang YP-103	Agustus	96
9	Yang YP-103	September	480
10	Yang YP-103	Oktober	720
11	Yang YP-103	November	0
12	Yang YP-103	Desember	840
13	Yang YP-104	Januari	306
14	Yang YP-104	Februari	90
15	Yang YP-104	Maret	0
16	Yang YP-104	April	720
17	Yang YP-104	Mei	0
18	Yang YP-104	Juni	540
19	Yang YP-104	Juli	360
20	Yang YP-104	Agustus	0
21	Yang YP-104	September	90
22	Yang YP-104	Oktober	360
23	Yang YP-104	November	0
24	Yang YP-104	Desember	720
25	Yang YP-105	Januari	420
26	Yang YP-105	Februari	180
27	Yang YP-105	Maret	0
28	Yang YP-105	April	720
29	Yang YP-105	Mei	0
30	Yang YP-105	Juni	1032
31	Yang YP-105	Juli	396
32	Yang YP-105	Agustus	204
33	Yang YP-105	September	360
34	Yang YP-105	Oktober	360
35	Yang YP-105	November	0
36	Yang YP-105	Desember	840
37	Yang YP-106	Januari	48
38	Yang YP-106	Februari	16
39	Yang YP-106	Maret	0
40	Yang YP-106	April	40
41	Yang YP-106	Mei	0
42	Yang YP-106	Juni	56

43	Yang YP-106	Juli	0
44	Yang YP-106	Agustus	40
45	Yang YP-106	September	48
46	Yang YP-106	Oktober	80
47	Yang YP-106	November	0
48	Yang YP-106	Desember	120
49	Yang YP-107	Januari	56
50	Yang YP-107	Februari	88
51	Yang YP-107	Maret	0
52	Yang YP-107	April	40
53	Yang YP-107	Mei	0
54	Yang YP-107	Juni	104
...	...	...	...
864	Yang UV-45W	Desember	0

mendukung pemrosesan *data mining*. Karena ingin diprediksi penjualan pada periode berikutnya, maka data *training* terdiri dari data penjualan periode sebelumnya dan data target berupa data penjualan periode selanjutnya. Contohnya misalkan diambil data penjualan dari YP-103, data *training* yang digunakan untuk prediksi periode Januari 2023 serupa dengan yang ada di tabel 3

Tabel 3. Data Training Penjualan YP-103

$x_1$			$x_2$		
No	Periode	Total Penjualan	No	Periode	Total Penjualan
1	Jan 20	264	1	Feb 20	168
2	Feb 20	168	2	Mar 20	0
3	Mar 20	0	3	Apr 20	480
4	Apr 20	480	4	Mei 20	0
5	Mei 20	0	5	Juni 20	264
...	...	...	...	...	...
35	Nov 22	1680	35	Des 22	0

### 3.1.2 Preprocessing

Untuk memudahkan proses perhitungan prediksi, tahap preprocessing data dicoba untuk mengategorikan berbagai jenis penjualan barang aksesoris aquarium berdasarkan volume penjualan setiap bulan dan tahun. Setelah data dikelompokkan, langkah selanjutnya adalah melakukan penggabungan untuk memberikan informasi penjualan dari semua aksesoris aquarium.

Tahap preprocessing ini memiliki peran krusial dalam menyederhanakan data dan mempersiapkannya agar lebih mudah diolah oleh algoritma prediksi [15]. Pada tahap ini, kategorisasi penjualan dapat membantu dalam mengidentifikasi tren atau pola yang mungkin muncul dalam setiap jenis produk aksesoris. Setelah penggabungan data dilakukan, informasi lebih lanjut dapat diekstraksi untuk memberikan pandangan yang holistik terhadap kinerja penjualan secara keseluruhan.

Berikut adalah contoh tabel penjualan produk aksesoris aquarium setelah melalui tahap preprocessing:

Tabel 2. Sample Data YP-103

Bulan	2020	2021	2022
Jan	264	960	0
Feb	168	0	0
Mar	0	240	120
Apr	480	192	480
Mei	0	360	720
Jun	264	480	360
Jul	72	360	72
Agus	96	120	360
Sep	480	480	120
Okt	720	0	0
Nov	0	144	1680
Des	840	1440	0

### 3.1.3 Transformation

Hasil pengelompokan data *preprocessing* digunakan sebagai data training pada tahap *transformation* ini. Untuk membuat data *training* dari data yang ada, yang dikenal sebagai data target, perlu ditentukan terlebih dahulu atribut mana yang dapat mempengaruhi penjualan aksesoris aquarium terlaris. Data target terdiri dari data dengan atribut yang relevan yang akan

Berdasarkan Tabel 3, data *training* untuk penelitian ini hanya diperoleh dari tiga tahun sebelumnya, yaitu data penjualan pada tahun 2020, 2021 dan 2022. Data *training* dibagi menjadi dua kelompok yaitu data input ( $x_1$ ) dan target ( $x_2$ ).

### 3.2 Hasil Penelitian

Langkah – langkah dari algoritma *K-Nearest Neighbor* adalah sebagai berikut, berdasarkan tahapan data mining:

1. Menentukan nilai k. Meskipun belum ada pedoman yang pasti untuk menentukan nilai k yang digunakan, namun nilai k yang digunakan dalam penelitian ini adalah 3.
2. Dengan menggunakan perhitungan *Euclidean Distance*, tentukan jarak antara data *training* dan data *testing* selama langkah transformasi. Karena data pada periode Januari 2023 belum ada, maka dianggap ( $x_1, x_2$ ) dari periode Januari 2023 adalah (0, 0). Proses perhitungannya dapat dijabarkan sebagai berikut

Tabel 4. Hasil Perhitungan jarak Euclidean

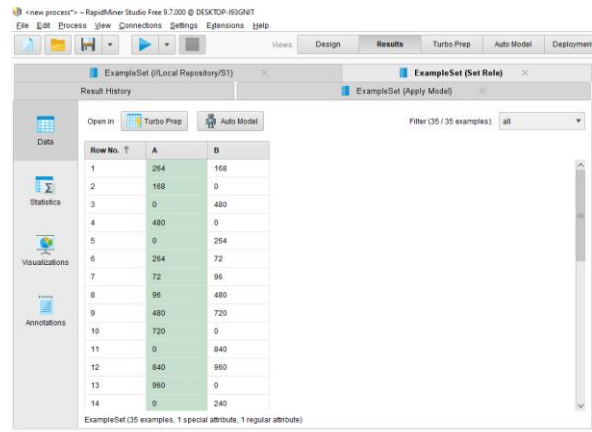
No	$x_1$	$x_2$	Perhitungan
1	264	168	$!(264 - 0)^2 + (168 - 0)^2 = 312.921$
2	168	0	$!(168 - 0)^2 + (0 - 0)^2 = 168$
3	0	480	$!(0 - 0)^2 + (480 - 0)^2 = 480$
4	480	0	$!(480 - 0)^2 + (0 - 0)^2 = 480$
...	...	...	...
35	1680	0	$!(1680 - 0)^2 + (0 - 0)^2 = 1680$

Daftar data yang dihitung. Jarak yang diperoleh kemudian diurutkan dari yang terdekat hingga terjauh setelah disortir. Setelah penyortiran, didapatkan :

Tabel 5. Hasil Pengurutan jarak Euclidean

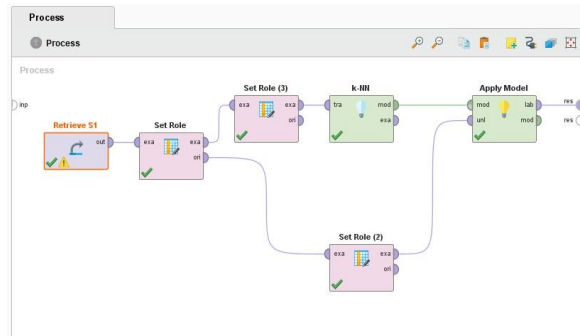
No	$x_1$	$x_2$	Hasil	Urutan
1	264	168	312.9217	12

2	168	0	168	7
3	0	480	480	18
4	480	0	480	19
5	0	264	264	9
6	264	72	273.6421	10
7	72	96	120	3
8	96	480	489.5059	21
9	480	720	865.3323	29
10	720	0	720	26
11	0	840	840	28
12	840	960	1275.617	32
13	960	0	960	31
14	0	240	240	8
15	240	192	307.35	11
16	192	360	408	17
17	360	480	600	24
18	480	360	600	25
19	360	120	379.4733	15
20	120	480	494.7727	22
21	480	0	480	20
22	0	144	144	6
23	144	1440	1447.182	34
24	1440	0	1440	33
25	0	0	0	1
26	0	120	120	4
27	120	480	494.7727	23
28	480	720	865.3323	30
29	720	360	804.9845	27
30	360	72	367.1294	13
31	72	360	367.1294	14
32	360	120	379.4733	16
33	120	0	120	5
34	0	1680	1680	34
35	1680	0	1680	35



Gambar 2. Importing data Training Pada Operator Read Excel

Setelah itu, masukkan nilai  $k = 3$ , dan drag and drop operator KNN. Kemudian gabungkan operator KNN dengan operator read Excel. tautkan KNN ke operator model yang berlaku. Terakhir, tautkan hasilnya ke operator model yang berlaku. Seperti yang dapat dilihat pada gambar 3



Gambar 3. Susunan Operator algoritma KNN

- Berdasarkan label mayoritas dari *K-Nearest Neighbor*, identifikasi kelompok data hasil pengujian. Karena  $k = 3$ , tiga jarak terdekat  $d7$ ,  $d15$ , dan  $d1$  dipilih.
- Kuantitas penjualan pada periode yang akan datang dapat diantisipasi dengan memanfaatkan kategori *Nearest Neighbor* yang paling mayoritas.

### 3.3 Pengujian Data

Aplikasi *Rapidminer* digunakan untuk pengujian data. *Rapidminer* adalah salah satu program data mining yang dapat digunakan untuk memeriksa kumpulan data dan menemukan pola data yang sesuai untuk tujuan pemrosesan data [16]. Modifikasi pola data diperlukan untuk mencapai tujuan pemrosesan data karena tidak semua algoritma yang ada dapat digunakan untuk memproses kumpulan data saat ini.

Setelah mengklasifikasikan, prosedur akan mengimpor tabel Microsoft Excel, setelah itu operator Read Excel akan ditempatkan dan data training akan dibaca dalam format Excel. Label kemudian dibuat dari atribut target. Gambar 2 menunjukkan hal ini.

Hasil pada gambar berikut merupakan langkah terakhir dari prosedur data mining dengan menggunakan *rapidminer* Setelah semua operator terhubung, tekan tombol ikon play atau F11. Tab hasil kemudian akan muncul, termasuk prediksi semua data yang memenuhi nilai  $k = 3$ .

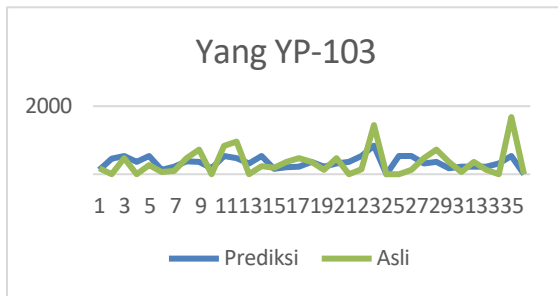
Row No.	prediction(B)	A	B
1	120	264	168
2	450	168	0
3	528	0	480
4	360	480	0
5	528	0	264
6	120	264	72
7	228	72	96
8	384	96	480
9	360	480	720
10	180	720	0
11	528	0	840
12	480	840	960
13	320	960	0
14	528	0	240

Gambar 4. Hasil prediksi penjualan Yang YP-103

Gambar 4 adalah hasil prediksi penjualan untuk Yang YP-103, untuk setiap bulan. Terlihat bahwa selisih antara hasil prediksi dengan nilai target tidak terlalu jauh, sehingga akurasi dari hasil prediksi metode K-NN cukup bagus. Berdasarkan hasil prediksi penjualan yang diperoleh, total penjualan untuk bulan Januari 2023 adalah sebesar 528 unit, seperti terlihat pada gambar berikut:

Row No.	prediction(B)	A	B
23	840	144	1440
24	0	1440	0
25	528	0	0
26	528	0	120
27	320	120	480
28	360	480	720
29	180	720	360
30	224	360	72
31	228	72	360
32	224	360	120
33	320	120	0
34	528	0	1680
35	0	1680	0
36	528	0	0

Gambar 5. Lanjutan Hasil prediksi penjualan Yang YP-103



Gambar 6. Grafik Perbandingan Hasil prediksi penjualan Yang YP-103

Berdasarkan hasil pengujian yang dilakukan, maka dapat dihitung nilai MAE (*Mean Absolute Error*) dari setiap hasil pengujian. Hasil yang diperoleh adalah sebagai berikut:

Tabel 6. Hasil Perhitungan Nilai MAE

No	Produk	Nilai MAE
1	YP-103	13.833
2	YP-104	-14.160
3	YP-105	107.705
4	YP-106	-3.176
5	YP-107	-10.074
6	YP-1200	-11.9544
7	YP-1600	-78.911
8	YP-1800	-56.5
9	YP-2600	-31.33
10	YP-100	4.259
11	YP-200	-8.33
12	YP-300	-2.5
13	YP-201	-12.5
14	YP-202	-4.44
15	YP-203	-3.556
16	UV-5W	-1.667
17	UV-7W	-1.667

18	UV-9W	-8.167
19	UV-11W	-10.833
<b>Rata-Rata</b>		6.196

Nilai MAE dari keseluruhan prediksi adalah sebesar 6.196, yang berarti bahwa persentase kesalahan (*error*) dari hasil prediksi adalah sebesar 6.196%.

#### 4. Kesimpulan

Teknik *K-Nearest Neighbor* digunakan dalam penelitian ini untuk memodelkan data yang telah disiapkan menggunakan tahap *Knowledge Discovery in Databases* (KDD). Hasil prediksi penjualan didasarkan pada perhitungan data mining yang menggunakan klasifikasi dan algoritma *K-Nearest Neighbor*. Dari hasil pengujian yang dilakukan, diperoleh informasi bahwa nilai MAE dari hasil prediksi adalah sebesar 6.196%, sehingga dapat disimpulkan bahwa hasil prediksi dari teknik *K-Nearest Neighbor* cukup akurat dalam memprediksi penjualan produk.

Beberapa saran dapat berguna untuk peneliti yang melakukan penelitian data mining pada masa yang akan datang berdasarkan hasil dari kesimpulan yang telah disampaikan, yaitu :

1. Membuat aplikasi untuk memprediksi penjualan aksesoris aquarium agar dapat dimanfaatkan oleh Perusahaan untuk menjadi lebih baik lagi.
2. Pembahasan pada perkiraan penjualan aksesoris aquarium terlaris ditambahkan sebanyak 5 tahun terakhir penjualan.

Menambahkan metode / algoritma lain pada prediksi sehingga hasil prediksi yang dilakukan memiliki tingkat akurasi yang lebih tinggi.

#### SUMBER RUJUKAN

##### Referensi

- [1] M. Muttaqin *et al.*, *Pengantar Teknologi Digital*. Yayasan Kita Menulis, 2023.
- [2] M. Muttaqin *et al.*, *Pengantar Internet*. Yayasan Kita Menulis, 2023.
- [3] A. Maulana *et al.*, *Manajemen Bisnis Digital dan E-Commerce*. Yayasan Kita Menulis, 2023.
- [4] A. Maulana *et al.*, *Pengantar Manajemen Basis Data dengan MySQL*. Yayasan Kita Menulis, 2023.
- [5] E. Turban, R. Sharda, J. E. Aronson, and D. King, *Business intelligence: A managerial approach*. Pearson Prentice Hall Upper Saddle River, NJ, 2008.
- [6] S. Wahyuddin *et al.*, *Audit Sistem Informasi*. Global Eksekutif Teknologi, 2023.
- [7] M. H. Siregar, "Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan)," *Jurnal Teknologi Dan Open Source*, vol. 1, no. 2, pp. 83–91, 2018.
- [8] G. Chairis and A. Maulana, "Analisis Perancangan dan Implementasi Sistem Informasi Stationary Berbasis Web pada PT. Indako Trading Coy," *Journal Information System Development (ISD)*, vol. 7, no. 2, pp. 78–90, 2022.
- [9] A. Syamil *et al.*, *Buku Ajar Manajemen Rantai Pasok*. PT. Sonpedia Publishing Indonesia, 2023.
- [10] W. P. Widharta, "Penyusunan strategi dan sistem penjualan dalam rangka meningkatkan penjualan toko damai," *Jurnal Strategi Pemasaran*, vol. 1, no. 2, pp. 1–15, 2013.

- [11] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques Second Edition," 2006. [Online]. Available: [www.mkp.com](http://www.mkp.com)
- [12] A. D. W. M. Sidik, I. H. Kusumah, A. Suryana, M. Artiyasa, A. P. Junfithrana, and others, "Gambaran Umum Metode Klasifikasi Data Mining," *FIDELITY: Jurnal Teknik Elektro*, vol. 2, no. 2, pp. 34–38, 2020.
- [13] A. Jaelani, "Deteksi dan klasifikasi tipe bangunan pada Citra Satelit menggunakan Metode K Nearest Neighbor," Universitas Islam Negeri Maulana Malik Ibrahim, 2020.
- [14] R. Takdirillah, "Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Penjualan Bisnis Ritel," *Edumatic: Jurnal Pendidikan Informatika*, vol. 4, no. 1, pp. 37–46, 2020.
- [15] I. Rahmayuni, "Perbandingan performansi algoritma c4.5 dan cart dalam klasifikasi data nilai mahasiswa prodi teknik komputer politeknik negeri padang," *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, vol. 2, no. 1, pp. 40– 46, 2014.
- [16] A. M. Siregar, S. Kom, M. K. D. A. N. A. Puspabhuana, S. Kom, and M. Kom, *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group, 2017.