

IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI DATA STOK FUKUBI OUTFIT MENGGUNAKAN METODE K- NEAREST NEIGHBOR

Firdaus Bintang Putra, Mukh Taufik Chulkamdi, Filda Febrinita
 Program Studi Teknik Informatika S1, Fakultas Teknologi Informasi
 Universitas Islam Balitar; Jl. Majapahit No.2-4, Sananwetan, Kec. Sananwetan,
 Kota Blitar, Jawa Timur, (0342) 813145
 bintangputra48984@gmail.com

ABSTRAK

Fukubi Outfit adalah perusahaan yang menjual kaos, hoddie, jaket, dan sweater dalam berbagai ukuran. Namun, perusahaan sering mengalami penumpukan atau kekurangan stok karena belum ada sistem prediksi pengelolaan stok. Masalah ini juga mempengaruhi gambaran laba rugi perusahaan. Untuk mengatasi masalah tersebut, digunakan teknik data mining dengan metode K-Nearest Neighbors Regression. Dengan metode ini, diharapkan data penjualan satu tahun terakhir bisadiolah menjadi informasi yang membantu pengelolaan stok. Implementasi metode ini dilakukan menggunakan Google Colab dan Python. Hasil pengujian dengan 2930 data dan pembagian 80% data latih dan 20% data uji menunjukkan nilai RMSE terbaik sebesar 0,75 pada K=8. Hasil ini menunjukkan model K-Nearest Neighbors baik dalam menyelesaikan kasus ini.

Kata kunci : Regresi, K – Nearest Neighbors, Data Mining, Root Mean Square Error, Google Collaborative, Console Python

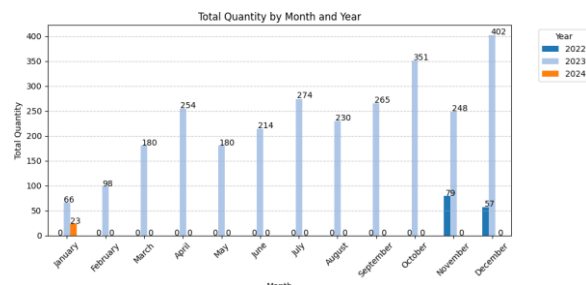
1. PENDAHULUAN

Pendataan dan ketersediaan data pun semakin cepat adanya sehingga banyak data dan kebutuhan informasi sebagai pendukung pengambilan keputusan untuk membuat bisnis dan dukungan infrastruktur di bidang teknik informatika. Di Fukubi Outfit proses pengelolaan persediaan stok masih bersifat standar, yaitu ketika sudah terjadi kekosongan stok produk diketahui, barulah dilakukan restock pengelolaan terhadap produk yang kosong tersebut. Selain itu, tidak adanya proses update stok setelah ada produk yang terjual, membuat pihak owner maupun karyawan Fukubi Outfit kesulitan dalam mengetahui detail stok yang harus diperbarui maupun stok yang masih tersedia di Fukubi Outfit itu sendiri. Padahal, perihal mengenai detail produk dan jumlah pengambilan adalah pertanyaan yang selalu diajukan supplier dalam proses restock produk selaku pemasok stok di Fukubi Outfit.

Masalah selanjutnya dihadapkan pada pengambilan keputusan dalam proses restock produk. Data penjualan pada Fukubi Outfit hanya menampilkan data detail produk yang data tersebut ditampilkan dengan penginputan satu – persatu dan diurutkan berdasarkan periode waktu penjualan tersebut terjadi. Pada bagian jumlah terjual menggunakan input data numerik, yaitu angka satu. Sehingga apabila produk tersebut terjual lebih dari satu, maka inputan data detail produk tersebut akan diulang sebanyak produk itu terjual.

Berdasarkan permasalahan di atas, harus dilakukan penerapan model *machine learning* sebagai pengolahan terhadap data penjualan *outfit* berdasarkan tipe *outfit*, ukuran *outfit*, warna *outfit*, nama *brand outfit* dan menghitung kemunculan data tersebut. Tujuan dari model *machine learning* tersebut adalah untuk mempercepat proses perhitungan penjualan

sehingga detail produk dengan penjualan tinggi dapat diketahui secara cepat. Karena banyaknya data dan kategori data yang dikelola serta representasi dari data tersebut sangat bermacam-macam, maka pemanfaatan informasi dan pengetahuan menggunakan data mining diharapkan dapat membantu memberikan solusi atau menjadi dasar pengambilan keputusan dalam pengelolaan stok dan prediksi stok untuk masa mendatang.



Gambar 1. Grafik penjualan fukubi Outfit

Penelitian yang berjudul “Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor” dengan studi kasus pada Toko Rizky Barokah Nganjuk memperoleh hasil akurasi terbaik pada 86,66%. Tujuan penelitian ini adalah untuk mengidentifikasi produk Unilever yang paling sering terjual dan memberikan prediksi stok yang perlu disiapkan untuk masa mendatang [1].

Penelitian selanjutnya mengenai algoritma K-NN juga diterapkan untuk prediksi penjualan kartu paket kuota internet. Untuk membantu PT. Akses Lintas Nusantara yang bergerak pada penjualan kartu paket kouta internet dalam perencanaan persediaan dan identifikasi terhadap produk mana yang paling laris, maka diperlukan mekanisme perhitungan algoritma untuk proyeksi penjualan di masa mendatang karena

perusahaan tidak dapat memproyeksikan jumlah penjualan. Dari hasil perhitungan K-NN, diperoleh persentase akurasi sebesar 71,43% untuk prediksi penjualan produk tahun 2020 [2].

Pada penelitian ini metrik yang digunakan untuk mengukur perbedaan atau perbandingan antara nilai prediksi yang dihasilkan oleh suatu model dengan nilai aktual yang diamati adalah *Root Mean Square Error* (RMSE) Selain itu, terdapat *Mean Absolute Percentage Error* (MAPE) yang dihitung sebagai rata-rata diferensiasi absolut antara nilai peramalan dan aktual, yang dinyatakan sebagai presentase nilai aktual.

2. TINJAUAN PUSTAKA

Penggunaan algoritma K-NN juga diterapkan pada penelitian oleh Bahtiar (2023) dengan studi kasus pada usaha bisnis yang bergerak di bidang penjualan aneka rangka dari berbagai jenis kayu yaitu Toko Kusen Kembar Jaya. Penelitian ini menghasilkan nilai perhitungan prediksi terhadap data penjualan senilai 88,89% dan 80,01% terhadap data kayu terpakai. Berdasarkan pemaparan tersebut, metode ini sangat cocok digunakan untuk prediksi penjualan masa depan karena dapat mengolah data penjualan produk secara tepat guna serta memberi tahu pemilik usaha bagaimana melakukan manajemen stok bahan baku untuk mencegah kekosongan. [10]

2.1. E-Commerce

Data mining merupakan suatu proses dimana informasi yang bernilai diperoleh dari suatu basis data yang besar. Proses ini melibatkan ekstraksi data secara cermat untuk menghasilkan informasi baru yang dapat memberikan kontribusi pada pengambilan keputusan [3].

Dalam penelitian yang berjudul “Penerapan *Data Mining* Untuk Klasifikasi Pandemi COVID-19 di Indonesia dengan *Algoritma Naive Bayes*”, mendefinisikan data mining merupakan wujud penerapan untuk menemukan format dan representasi prediktif mengacu pada data masa lalu dalam suatu masa [4].

2.2. Machine Learning

Machine Learning merupakan suatu bentuk pemrograman komputer yang menggunakan data historis untuk membentuk model pembelajaran yang pada akhirnya mencapai kinerja optimal saat mengekstraksi informasi dari kumpulan data [5].

2.3. K-Nearest Neighbor

Algoritma KNN merupakan suatu metode pengklasifikasian data berdasarkan jarak terpendek terhadap objek data [6]. Penentuan nilai K optimal untuk algoritma ini bergantung pada karakteristik data yang ada. Nilai K yang tinggi dapat membantu mengurangi pengaruh *noise* pada proses klasifikasi, namun juga dapat mengaburkan batas antar klasifikasi.

2.4. Knowledge Discovery in Database (KDD)

Knowledge Discovery In Database (KDD) adalah proses untuk menentukan informasi yang berguna serta pola-pola yang terdapat dalam data yang besar dan sebelumnya tidak diketahui. Tahapan-tahapan dalam proses KDD meliputi: pertama, Data Selection, di mana data dipilih berdasarkan variabel yang relevan untuk penelitian; kedua, Pre-processing/Cleaning, yang melibatkan pembersihan data untuk memastikan kebersihan data sebelum analisis, dengan mengurangi jumlah data dari 2930 menjadi 50 record yang bersih; ketiga, Transformation, untuk menghasilkan satu set data yang siap untuk analisis; dan terakhir, Analisis data.

2.5. Regresi

Metode regresi statistik digunakan untuk mengukur dan mensimulasikan hubungan antara satu variabel (variabel dependen) dan satu atau lebih variabel independen. Regresi juga dapat digunakan untuk mengevaluasi kekuatan hubungan antara variabel dan memperkirakan masa depan berdasarkan pola hubungan yang ditemukan sebelumnya [5].

2.6. Prediksi

Prediksi adalah suatu proses untuk mengestimasi atau meramalkan nilai suatu variabel di masa mendatang dengan mempertimbangkan data historis. Dalam konteks ini, data kuantitatif, yang mencakup angka atau nilai numerik, sering digunakan untuk melakukan prediksi. tujuan prediksi adalah untuk menemukan perkiraan atau estimasi seakurat mungkin berdasarkan informasi yang telah tersedia [7].

2.7. Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE), yang dihasilkan dari perhitungan metode tertentu, digunakan untuk membandingkan nilai yang diprediksi oleh model hipotetis dengan nilai dari hasil pengamatan. Dengan kata lain, RMSE menilai tingkat kesesuaian antara model prediksi dan data aktual. Persamaan RMSE dapat dilihat pada persamaan berikut

$$RMSE = \sqrt{\sum \frac{(y^1 - y^2)^2}{n}} \dots \dots \dots (1)$$

Keterangan

Y = Nilai aktual

Y' = Nilai prediksi

n = jumlah data

2.8. Python

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. *Python* diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. *Python* bisa dibilang bahasa pemrograman dengan tujuan umum yang

dikembangkan secara khusus untuk membuat source code mudah dibaca [8].

2.9. Matplotlib

Matplotlib merupakan *library* yang digunakan dalam Bahasa pemrograman *python* untuk membuat visualisasi data. *Library* yang dirilis pada tahun 2003 ini merupakan *library* visualisasi data yang paling populer dan banyak digunakan oleh *python users*.

2.10. Numpy

Numpy merupakan salah satu *library* pada Bahasa pemrograman *python* yang digunakan untuk melakukan perhitungan matematis seperti: aljabar, statistik, maupun perhitungan kompleks yang lain.

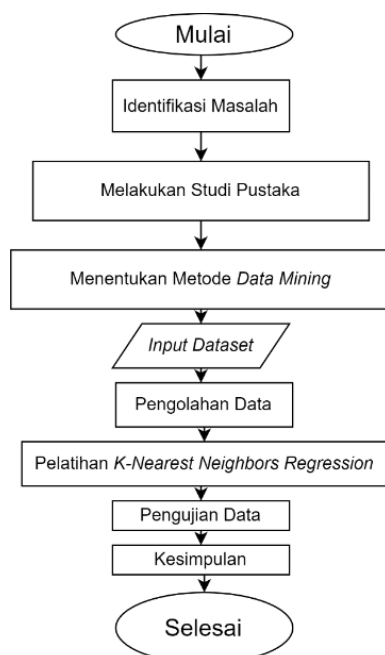
2.11. Manajemen Barang

Manajemen adalah kumpulan disiplin ilmu dan seni yang berkaitan dengan bagaimana menggunakan sumber daya, termasuk sumber daya manusia, dengan cara yang efisien dan efektif untuk mencapai tujuan tertentu [9].

Sedangkan Manajemen adalah disiplin ilmu dan seni yang mencakup perencanaan, pengorganisasian, penyusunan, pengarahan, dan pengawasan sumber daya manusia untuk mencapai tujuan tertentu.

3. METODE PENELITIAN

Jenis penelitian yang akan digunakan adalah penelitian deskriptif. Penelitian deskriptif adalah metode penelitian untuk menggambarkan subjek atau objek yang ditelitinya secara lebih mendalam, terperinci, dan luas. Metode ini biasanya digunakan guna memecahkan atau menjawab suatu persoalan dengan mengumpulkan data-data, melakukan analisis, klasifikasi, membuat kesimpulan, dan laporan.



Gambar 2. alur tahapan penelitian

Pada penelitian ini masalah yang muncul adalah Fukubi Outfit yang menghadapi isu terkait fluktuasi data penjualan, serta kurangnya prediksi untuk menentukan penjualan. Kondisi ini berdampak pada perputaran modal yang ada didalam Fukubi Outfit, kemudian melakukan studi pustaka penelitian dilakukan dengan mencari informasi mengenai penerapan algoritma *K-Nearest Neighbor Regression* melalui teori yang terdapat dalam buku, jurnal, dan artikel-artikel, Setelah menyelesaikan Studi Pustaka, langkah berikutnya adalah memilih metode yang paling sesuai untuk menerapkan teknik regresi.

Pada tahap pengumpulan data peneliti mengambil data dari history arsip penjualan selama setahun lebih yang ada di Fukubi Outfit, data tersebut meliputi Nama, No *Whatsapp*, Keterangan dan Qty atau barang yang terjual pada hari itu. Kemudian langkah selanjutnya adalah pengolahan data Dalam tahap ini, dilaksanakan proses pembersihan data yang akan digunakan untuk prediksi menggunakan metode *K-Nearest Neighbors Regression*. Sebagai contoh, data penjualan dan kuantitasnya masih bersatu, sehingga diperlukan tindakan pembersihan seperti pemisahan data yang terkandung dalam informasi keterangan, selanjutnya melatih KNN Regression Dalam melatih model *K-Nearest Neighbors* menggunakan *Google Colab* dan menggunakan *library Scikit-Learn* dalam bahasa pemrograman *Python*, satu aspek yang perlu ditekankan adalah penentuan nilai *k* yang optimal.

Penting untuk memilih nilai *k* yang menghasilkan hasil maksimal dalam perhitungan menggunakan metode *K-Nearest Neighbors*. Proses *K-Nearest Neighbors* ini melibatkan perhitungan berdasarkan tetangga terdekat dari data yang ada, terakhir adalah Pengujian Data Pada tahap evaluasi hasil prediksi, dilakukan pengukuran kualitas dengan menggunakan *Root Mean Square Error (RMSE)*.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Tabel 1. Variabel Data Fukubi Outfit

No	Nama Variabel	Keterangan
1.	Tanggal Pembelian	Variabel ini berisi waktu terjadinya transaksi
2.	Nama Pembeli	Variabel ini berisi nama pembeli di Fukubi Outfit
3.	No Whatsapp	Variabel ini berisi nomor whatsapp dari setiap pembeli di toko fukubi outfit
4.	Alamat	Variabel ini berisi alamat dari setiap pembeli di toko fukubi outfit
5.	Keterangan	Variabel ini berisi keterangan detail produk yang dibeli konsumen fukubi outfit
6.	Kuantiti	Variabel ini berisi jumlah kuantiti produk terjual setiap produk

Tahapan pertama yang dilakukan adalah pengambilan data sekunder dari riwayat data penjualan pengumpulan data dilakukan melalui metode *Fukubi Outfit* tahun 2022 –2023.

Tabel 2. Dataset Fukubi Outfit

No	Tanggal/ Bulan	Nama Pembeli	No. Whatsapp	Alamat	Keterangan Produk	QTY	Total Harian	Total Bulanan
1	1/12	Dela Aprillia	6.28135E+12	Kalimantan Tengah	ZP RRQ VOL 2 Hitam Lengan Ungu - XL	1		
2	1/12	Riyan JR	6.28578E+12	DKI Jakarta	HD ZP RRQ Vol 1 – Hitam Lengan Putih - L	1		
3	1/12	Herlina	6.29E+11	Jawa Barat	Varsity Onic Vol 2 – Bordir - M	1		
4	1/12	Evan Setiono	6.8383E+12	Jawa Tengah	Varsity Onic Vol 2 – Bordir - M	1		
5	1/12	Reza	6.28516E+12	Jawa Barat	Varsity Onic Vol 2 – Sablon - XL	1	5	
.....								
2926	30/11	Mochammad Satria Bakti	6.2831E+12	Jawa Barat	Varsity Echo Vol 1 – Ungu Terong – 2XL	1		
2927	30/11	Muhammad Fitra	6.28194E	DKI Jakarta	Varsity Echo Vol 1 – Ungu Terong - L	1		
2928	30/11	Theo Berasa	6.28133E+12	DIY	Varsity Onic Vol 2 – Bordir - L	1		
2929	30/11	Enrique Lawrence Tandy	6.28198E+12	Sumatra Utara	Varsity Onic Vol 2 – Bordir - XL	1		
2930	30/11	Adib Kripto	6.01115E+11	Malaysia	Varsity Echo Vol 1 – Ungu Terong - L	1	10	249

Tabel 2 berisi sample penjualan pada toko fukubi dengan jumlah total penjualan sebanyak 2930 transaksi. Setelah dilakukan pengambilan data dalam bentuk file csv. Langkah selanjutnya adalah melakukan proses *pre-processing* data. . Dalam penelitian ini penulis menggunakan *Google colab* untuk proses data mining dan pemrograman python. *Google Colab* merupakan platform berbasis *cloud* yang memungkinkan penulis untuk menjalankan *code* secara virtual

```
file = "dataset_fukubi.csv"
```

```
df = pd.read_csv(file)
```

```
df
```

df ini merupakan variabel yang dibuat oleh penulis untuk melakukan pembacaan pada file csv. Sehingga proses *running* dari *source code* tersebut menghasilkan *dataset* fukubi sebagai berikut

Tabel 3. Data Import di Python

Index	Tanggal	Nama Pembeli	Whatsapp	Alamat	Keterangan	Qty
0	12/1/2023	Dela Aprillia	6.28E+12	Kab.Barito Utara, Teweh Tengah, Kalimantan Tengah	ZP RRQ VOL 2 – Hitam Lengan Ungu - XL	1.0
1	NaN	Riyan JR	6.29E+12	Kota Jakarta Utara, Penjaringan, Dki Jakarta	HD ZP RRQ Vol 1 – Hitam Lengan Putih – L	1.0
2	NaN	Herlina	6.29E+11	Kab.Bogor, Cileungsi, Jawa Barat	Varsity Echo Vol 1 – Ungu Terong – Bordir – L	1.0
3	NaN	Evan Setiono	6.28E+12	Kab.Banyumas, Karanglewas, Jawa Tengah	Varsity Onic Vol 2 – Bordir – M	1.0
4	NaN	Reza	6.29E+12	Kota Bekasi	Varsity Onic Vol 2 – Sablon – XL	1.0

Setelah melakukan proses impor data CSV, langkah selanjutnya adalah mengisi data tanggal yang kosong menggunakan *library* pandas. Data tanggal

yang kosong akan diisi dengan data dari baris di atasnya, sesuai dengan tabel 3 yang menunjukkan

bahwa data tanggal memiliki kekosongan yang perlu diisi dari data yang ada di baris sebelumnya.

```
df['TANGGAL'].fillna(method='ffill', inplace=True)
df.TANGGAL
```

```
0      12/1/2023
1      12/1/2023
2      12/1/2023
3      12/1/2023
4      12/1/2023
...
2925   11/30/2023
2926   11/30/2023
2927   11/30/2023
2928   11/30/2023
2929   11/30/2023
Name: TANGGAL, Length: 2930, dtype: object
```

Gambar 3. Hasil Pengisian Data Tanggal Untuk Data Kosong

Setelah pengisian data tanggal yang kosong selesai dilakukan, langkah berikutnya adalah mengubah data tanggal menjadi variabel yang terdiri dari bulan, hari, dan tahun. Hal ini diperlukan karena dalam perhitungan *K-Nearest Neighbors*, data tanggal harus diubah ke dalam format numerik seperti pada gambar 4.

```
df[['bulan', 'hari', 'tahun']] = df['TANGGAL'].str.split('/', expand=True)
df
```

	TANGGAL	NAMA PEMBELI	NO. WHATSAPP	ALAMAT	KETERANGAN	QTY	bulan	hari	tahun
0	12/1/2023	Dela Aprilia	6.28E+12	KAB. BARITO UTARA, TEWEH TENGAH, KALIMANTAN TE...	ZP RRG VOL 2 Hitam Lengan Ungu - XL	1.0	12	1	2023
1	12/1/2023	Riyan JR	6.29E+12	KOTA JAKARTA UTARA, PENJARINGAN DKO JAKARTA,	HD ZP RRG Vol 1 - Hitam Lengan Putih - L	1.0	12	1	2023
2	12/1/2023	Herlina	6.29E+11	KAB. BOGOR, CILEUNGSI, JAWA BARAT	Varsity Echo Vol 1 Ungu Terong - Border - L	1.0	12	1	2023
3	12/1/2023	Evan Setono	6.28E+12	KAB. BANYUMAS, KARANGSEWU, JAWA TENGAH	Varsity Onic Vol 2 - Border - M	1.0	12	1	2023
4	12/1/2023	Reza	6.29E+12	KOTA BEKASI, BEKASI UTARA, JAWA BARAT	Varsity Onic Vol 2 - Sablon - XL	1.0	12	1	2023

Gambar 4. contoh hasil pemisahan tanggal

Tabel 4. Pengolahan Data Fukubi Outfit

No	Keterangan Produk	Bulan	Tahun	QTY
0	Bm alter ego vol 1 – merah putih – 2xl	11	2022	2.0
1	Bm alter ego vol 1 – merah putih – 2xl	2	2023	1.0
2	Bm alter ego vol 1 – merah putih – 2xl	3	2023	1.0
3	Bm alter ego vol 1 – merah putih – 2xl	4	2023	1.0
4	Bm alter ego vol 1 – merah putih – 2xl	5	2023	1.0
....				
1132	Zp varsity evos vol 2 – hitam - xl	5	2023	2.0
1133	Zp varsity evos vol 2 – hitam – xl	6	2023	1.0
1134	Zp varsity evos vol 2 – hitam – xl	8	2023	1.0
1135	Zp varsity evos vol 2 – hitam lengan navy - 5xl	1	2024	1.0
1136	Zp varsity evos vol 2 – hitam lengan navy - xl	1	2023	1.0

Pada tabel 4 di atas, terlihat data sudah disesuaikan dengan pengolahan data pada penelitian ini. Dengan tersisa 3 kolom atribut data yang berisi keterangan produk, bulan, tahun dan qty. Pengolahan data tersebut menghitung jumlah penjualan untuk setiap detail produk dalam kurun tahun 2022 dan tahun 2023. Tabel di atas menampilkan salah satu proses data mining dalam penelitian ini yaitu mengklasifikasikan setiap detail produk sehingga jumlah penjualan setiap detail produk dapat diketahui sebagai proses awal regresi data. Data diklasifikasikan dalam bentuk array.

Setelah proses ini selesai, langkah berikutnya adalah menghilangkan data yang kosong dan menggabungkan data berdasarkan kata kunci dari keterangan produk.

```
[ ] df = df.dropna(subset=['KETERANGAN'])
[ ] df_clean_ket = df[(df['KETERANGAN'] != ' ') & (df['KETERANGAN'] != '\n')]
[ ] df_ket_bul = df_clean_ket.groupby(['KETERANGAN', 'bulan', 'tahun'])['QTY'].sum().reset_index()
df_ket_bul
```

	KETERANGAN	bulan	tahun	QTY
0	bm alter ego vol 1 - merah putih - 2xl	11	2022	2.0
1	bm alter ego vol 1 - merah putih - 2xl	2	2023	1.0
2	bm alter ego vol 1 - merah putih - 2xl	3	2023	1.0
3	bm alter ego vol 1 - merah putih - 2xl	4	2023	1.0
4	bm alter ego vol 1 - merah putih - 2xl	5	2023	1.0
...	...	...	...	...
1132	zp varsity evos vol 2 - hitam - xl	5	2023	2.0
1133	zp varsity evos vol 2 - hitam - xl	6	2023	1.0
1134	zp varsity evos vol 2 - hitam - xl	8	2023	1.0
1135	zp varsity evos vol 2 - hitam lengan navy - 5xl	1	2024	1.0
1136	zp varsity evos vol 2 - hitam lengan navy - xl	1	2023	1.0

1137 rows x 4 columns

Gambar 5. Hasil Pengelompokan Data

Pada Gambar 5 terlihat bahwa data telah disesuaikan untuk memenuhi kebutuhan prediksi penjualan barang di toko Fukubi. Dari yang awalnya memiliki 6 kolom, sekarang hanya tersisa 3 kolom: keterangan produk, bulan, tahun, dan kuantitas (QTY). Penyederhanaan ini dilakukan agar data lebih mudah dianalisis dibandingkan dengan tabel asli. Berikut adalah tabel yang menampilkan hasil pengolahan data Fukubi Outfit yang relevan dengan penelitian ini.

Adapun detail produk yang diklasifikasikan pada tabel di atas adalah sebagai berikut

- 1) Bm Alter ego vol 1 – merah putih – 2xl
- 2) Zp varsity evos vol 2 – hitam – xl
- 3) Zp varsity evos vol 2 – hitam lengan navy – 5xl

Pengolahan data terhadap detail – detail produk tersebut menghasilkan data jumlah penjualan setiap detail produk yang nantinya data tersebut akan digunakan dalam proses regresi data dan implementasi *K – Nearest Neighbors*.

4.2. Proses Transformasi Data

Proses transformasi data ini melibatkan pengubahan format atau struktur data. Tujuan dari proses transformasi ini adalah untuk membuat data penjualan yang telah melalui proses *pre-processing* menjadi lebih siap untuk proses pelatihan K- *Nearest Neighbors*. Berikut adalah sampel yang terdapat dalam variabel keterangan.

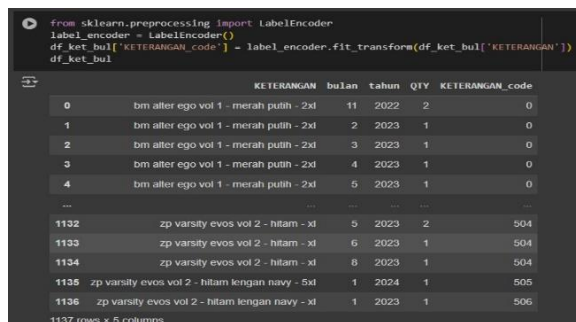
```
Array(['bm alter ego vol 1 - merah putih - 2xl',
      'bm alter ego vol 1 - merah putih - 3xl',
      'bm alter ego vol 1 - merah putih - 4xl',
      'bm alter ego vol 1 - merah putih - l',
      'bm alter ego vol 1 - merah putih - m',
      'bm alter ego vol 1 - merah putih - s',
      'bm alter ego vol 1 - merah putih - xl',
      ], dtype=object)
```

Dari kategori diatas akan dirubah menjadi

Tabel 5. Sampel Konversi Label Keterangan

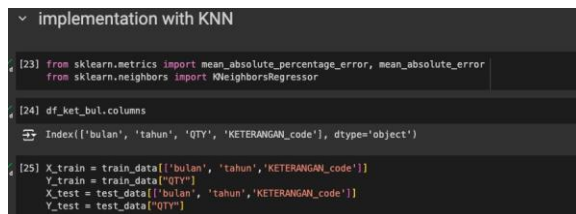
Keterangan	Keterangan_Code
'bm alter ego vol 1 - merah putih - 2xl',	1
'bm alter ego vol 1 - merah putih - 3xl',	2
'bm alter ego vol 1 - merah putih - 4xl',	3
'bm alter ego vol 1 - merah putih - l',	4
'bm alter ego vol 1 - merah putih - m',	5
'bm alter ego vol 1 - merah putih - s',	6
'bm alter ego vol 1 - merah putih - xl',	7

Untuk menerapkan cara kerja tersebut di Python, perlu mengimpor library *sklearn* dengan menggunakan *LabelEncoder*. Hasilnya dapat dilihat pada gambar 6 berikut ini.



Gambar 6. transform keterangan ke numerik

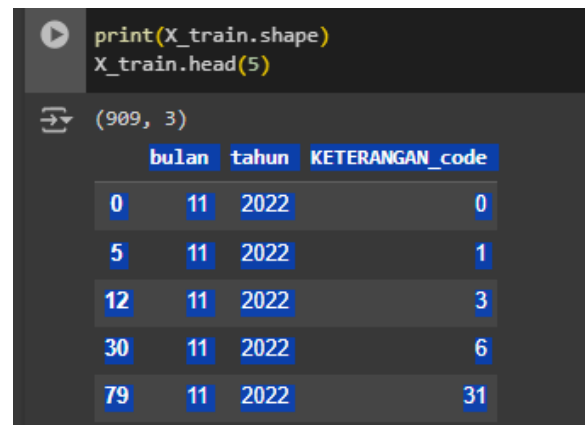
Dari proses pengolahan tersebut, data kini siap untuk pelatihan menggunakan algoritma K-Nearest Neighbors. Data tersebut sudah dalam bentuk numerik, sehingga variabel keterangan dapat dihilangkan karena sudah dikonversi ke dalam bentuk numerik. Kode untuk menggunakan library K-Nearest Neighbors dapat dilihat pada Gambar 6. Pada implementasi ini menggunakan library *Sklearn*. Untuk tahap awal import library *SKLearn* yang berisi fungsi untuk menentukan rata-rata persentase *error*, dan fungsi untuk KNN.



Gambar 7. Kode Awal KNN

Selanjutnya, dilakukan pengumpulan kolom yang diperlukan untuk keperluan data mining. Dalam

penelitian ini, kolom yang dibutuhkan adalah bulan, tahun, dan keterangan. Berikut adalah contoh dari variabel x yang digunakan sebagai data yang akan diolah.



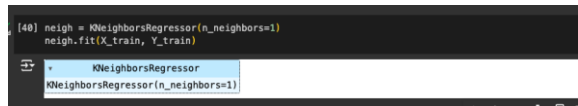
Gambar 8. Data Variabel X

Selanjutnya, hasil dari pemisahan data untuk variabel yang akan dipakai sebagai tujuan (target), yaitu variabel y, adalah QTY atau kuantitas. Dalam kode tersebut diketahui bahwa panjang data yang digunakan untuk pelatihan adalah 909, dan terdapat satu kolom saja yang merupakan variabel QTY.



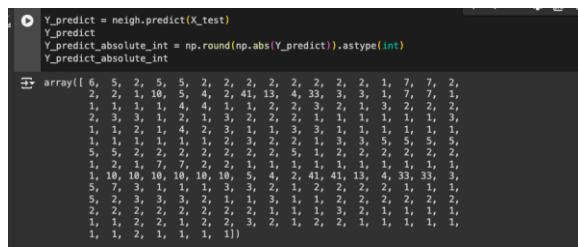
Gambar 9. Data Variabel Y

Setelah itu, dilakukan proses KNN *Regression* untuk melakukan prediksi. Berikut adalah kode program untuk melakukan proses prediksi dengan KNN *Regression*, di mana nilai dihitung menggunakan rumus menggunakan *euclidean distance* seperti yang ada didalam gambar 10.



Gambar 10. Penggunaan Regresi KNN

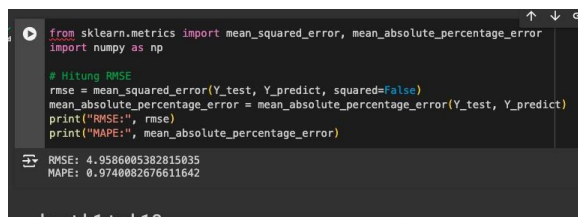
Setelah pelatihan dilakukan, proses selanjutnya adalah melakukan prediksi menggunakan data uji dengan pembagian 80% untuk pelatihan dan 20% untuk pengujian. Hal ini bertujuan untuk mengevaluasi kinerja dari algoritma *K-Nearest Neighbors*. Berikut ini adalah contoh hasil prediksi yang dapat dilihat pada Gambar 11.



Gambar 11. Hasil Prediksi KNN

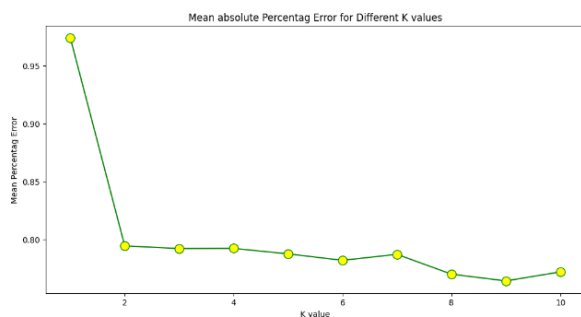
4.3. Pengujian

Pengujian yang dilakukan dalam penelitian ini menggunakan metode MAPE. Pengujian dilakukan dari setiap nilai uji coba K dari rentang 1 – 10.



Gambar 12. Hasil Perhitungan RMSE Dan MAPE

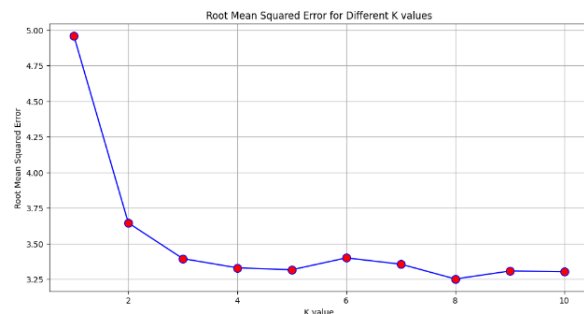
Gambar 12 diatas merupakan visualisasi hasil pengujian MAPE yang telah dilakukan Kode program awal untuk perhitungan RMSE dan MAPE ditunjukkan pada gambar 12.



Gambar 13. komparasi K values MAPE

Berdasarkan gambar 13 dapat dilihat nilai k = 9 memiliki hasil yang lebih baik dengan nilai pengujian k yang lain hal ini dikarenakan nilai persentase error semakin mengecil pada angka 76%. Sehingga untuk

pengujian nilai k = 9 merupakan nilai yang optimal untuk hasil eror yang rendah pada pengujian *MAPE*. Selanjutnya adalah Pengujian dengan menghitung nilai RMSE dari setiap nilai uji coba k dari rentang 1 – 10. Berikut hasil dari Nilai RMSE yang muncul pada setiap nilai K. berikut merupakan bar line visualisasi perbandingan dari metode pengujian RMSE dari setiap kluster.



Gambar 14. hasil visualisasi rmse

Berdasarkan gambar 14, dapat dilihat bahwa nilai K pada data memiliki variasi dalam nilai RMSE (Root Mean Square Error), dengan beberapa nilai K1 sampai K2 menunjukkan nilai RMSE yang cukup tinggi namun K8 dan 9 mengalami penurunan persentase error. Pengujian RMSE ini bertujuan untuk menentukan nilai K yang optimal untuk kasus prediksi stock di fukubi. Dalam analisis ini, nilai RMSE untuk berbagai nilai K adalah sebagai berikut.

Tabel 6. Hasil RMSE

Nilai K	RMSE
K 1	4.9586
K 2	3.6449
K 3	3.3944
K 4	3.3295
K 5	3.3161
K 6	3.3995
K 7	3.3553
K 8	3.2516
K 9	3.3074
K 10	3.3024

Dari Tabel 6 di atas, terlihat bahwa nilai RMSE menurun secara signifikan dari K= 1 ke K= 2 dan terus menurun hingga mencapai K= 5. Setelah itu, tingkat naik dan turun nilai RMSE menjadi lebih kecil dan cenderung stabil pada K6 dan seterusnya. Nilai RMSE terendah ditemukan pada K= 8, yaitu sebesar 3.2516, yang menunjukkan bahwa K= 8 menghasilkan prediksi yang paling optimal dalam penelitian ini.

Jadi dapat disimpulkan, nilai RMSE yang lebih rendah dapat menunjukkan bahwa model *K Nearest Neighbors* memiliki performa yang lebih baik dalam meminimalisir kesalahan. Oleh karena itu, K= 8 dapat dianggap yang optimal berdasarkan hasil RMSE yang dihasilkan

5. KESIMPULAN DAN SARAN

Implementasi metode K-Nearest Neighbors (KNN) untuk memprediksi penjualan Fukubi melibatkan konversi data menggunakan Label Encoder, pembagian data menjadi data latih dan uji, serta perhitungan jarak Euclidean untuk menemukan K tetangga terdekat. Model KNN kemudian melakukan regresi dengan menghitung rata-rata nilai target dari tetangga terdekat.

Pengujian algoritma menggunakan 2930 data penjualan dengan 80% data latih dan 20% data uji serta nilai $k = 7$ menghasilkan Root Mean Square Error (RMSE) sebesar 0,75. Ini menunjukkan bahwa model KNN efektif dalam kasus regresi ini. Untuk penelitian selanjutnya, disarankan menambah variabel x , meningkatkan jumlah data, mengeksplorasi variasi pengujian, dan mempertimbangkan teknik ensemble untuk meningkatkan akurasi prediksi.

DAFTAR PUSTAKA

- [1] Alfani W.P.R., A., Rozi, F., & Sukmana, F. (2021). Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 6(1), 155–160. <https://doi.org/10.29100/jupi.v6i1.1910>
- [2] Handoko, D., Tambunan, H. S., & Hardinata, J. T. (2021). Analisis Penjualan Produk Paket Kuota Internet Dengan Metode K-Nearest Neighbor. *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, 6(1), 111. <https://doi.org/10.30645/jurasik.v6i1.275>
- [3] Suntoro, J. (2019). *DATA MINING: Algoritma dan Implementasi dengan Pemrograman php*. Elex Media Komputindo
- [4] Damanik, A. R., Hartama, D., & Damanik, I. S. (2022). Prediksi Tingkat Kematian di Indonesia Akibat Covid-19 Menggunakan Algoritma Naïve Bayes. *SISTEMASI: Jurnal Sistem Informasi*, 11(1), 1–11.
- [5] Sidik, A. D., & Ansawarman, A. (2022). Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning. *Formosa Journal of Multidisciplinary Research*, 1(3), 559–568. <https://doi.org/10.55927/fjmr.v1i3.745>
- [6] Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 6(2), 118–127. <https://doi.org/10.31294/ijcit.v6i2.10438>
- [7] Adiguno, S., Syahra, Y., & Yetri, M. (2022). Prediksi Peningkatan Omset Penjualan Menggunakan Metode Regresi Linier Berganda. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 1(4), 275. <https://doi.org/10.53513/jursi.v1i4.5331>
- [8] Perkovic, L. (2012). Introduction to Computing Using Phyton : An. Binus University, 33-34.
- [9] Astuti, R., Akbar, L. Y., & Hasibuan, L. S. (2021). Pengaruh Motivasi Dan Pelatihan Terhadap Kinerja Karyawan PT. Deltamas Surya Indah Mulia Medan. *Seminar Nasional Teknologi Edukasi dan Humaniora 2021*, 8–13.
- [10] Bahtiar, R. (2023). Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor. *Jurnal Informatika MULTI*, 1(3), 203–214