IMPLEMENTASI ALGORITMA REGRESI LINEAR DALAM PREDIKSI PERSEDIAAN *VOUCHER* DI RAFFA *CELL* SUKABUMI

Alfi Nur Afiati, Prajoko, Fathia Frazna Az-Zahra
Teknik Informatika, Universitas Muhammadiyah Sukabumi
Jl. R. Syamsudin, S.H. No. 50, Cikole, Kec. Cikole, Kota Sukabumi, Jawa Barat 43113
Alfi072@ummi.ac.id

ABSTRAK

Prediksi merupakan sebuah upaya untuk mengantisipasi kejadian di masa depan dengan menggunakan data atau informasi dari kejadian sebelumnya. Perubahan dalam dunia bisnis, terutama dipicu oleh kemajuan teknologi informasi khususnya di sektor penjualan, tentu memberi dampak signifikan dalam hal manajemen data dan analisis, menjadikannya aset berharga di berbagai sektor termasuk industri penjualan seperti Raffa Cell. Pemahaman mendalam terhadap latar belakang persediaan stok *voucher* menjadi kunci dalam menjaga keberlanjutan dan pertumbuhan bisnis. Penelitian ini didasarkan pada situasi di Raffa Cell Sukabumi, di mana ketidakpastian terkait ketersediaan stok voucher menjadi tantangan utama yang dapat memengaruhi kelancaran operasi bisnis. Fokus penelitian ini yaitu penerapan algoritma regresi linear untuk memprediksi penjualan di Raffa Cell Sukabumi. Algoritma regresi linear memodelkan hubungan linier antara variabel independen dengan variabel dependen yang diprediksi untuk menghasilkan estimasi akurat berdasarkan pola data historis. Penelitian ini dapat digunakan untuk mengatasi ketidakpastian ketersediaan stok *voucher* dan mengoptimalkan proses transaksi penjualan, sehingga memperkuat daya saing bisnis Raffa Cell di pasar yang semakin kompetitif. Setelah atribut data yang akan menjadi bagian dari prediksi ditentukan, didapatkan hasil pengujian performa model dengan nilai RMSE 29%, MAE 25%, MSE 84%, dan R2 90%.

Kata kunci : *Prediksi, Efisiensi Operasional, Pengelolaan, Persediaan, Regresi Linear*

1. PENDAHULUAN

Sistem persediaan barang adalah mekanisme yang digunakan untuk mengatur stok barang di gudang. Saat ini, sistem persediaan barang sudah banyak diterapkan oleh perusahaan-perusahaan yang sedang berkembang, terutama dalam pengelolaan data barang. Persediaan barang adalah komponen utama yang sangat penting dalam suatu perusahaan, karena persediaan tersebut dijual secara terus-menerus demi kelancaran operasi bisnis [1].

Manajemen persediaan yang efisien sangat penting, tidak hanya untuk memenuhi kebutuhan pelanggan, tetapi juga untuk menjaga kesehatan finansial toko dan keunggulan kompetitif. Dengan memahami permintaan pelanggan, siklus hidup produk, dan tren pasar, pemilik toko dapat mengembangkan strategi cerdas untuk pengadaan dan pemeliharaan persediaan, serta beradaptasi dengan perubahan di masa depan.

Selain itu, manajemen persediaan mencerminkan seberapa baik sebuah toko mengelola risiko dan meminimalkan kerugian, sehingga penting untuk mengatasi potensi ketidakseimbangan stok. Integrasi teknologi untuk pemantauan stok secara real-time meningkatkan pengambilan keputusan dan efisiensi operasional. Pemahaman yang mendalam tentang manajemen persediaan merupakan investasi strategis yang penting untuk mempertahankan dan mengembangkan bisnis toko di pasar yang kompetitif.

Dalam kasus Raffa Cell, sebuah toko ponsel dan aksesori, tantangan yang dihadapi dan menyebabkan ketidakefisienan operasional dan ketidakpastian dalam perencanaan bisnis yaitu tidak adanya sistem peramalan transaksi yang efektif dan persediaan *voucher*.

Persediaan *voucher* bukan hanya merupakan aspek administratif dari operasional toko. Oleh karena itu, pemahaman mendalam tentang manajemen stok *voucher* tidak hanya merupakan kebutuhan, tetapi juga kunci bagi keberlanjutan dan pertumbuhan bisnis. Dengan kata lain, persediaan tidak hanya berkaitan dengan kebutuhan langsung pelanggan, tetapi juga penting untuk menjaga kesehatan finansial toko. Manajemen stok yang efisien berkontribusi pada peningkatan kinerja finansial dan pertumbuhan toko. Dengan merinci aspek-aspek seperti tingkat permintaan pelanggan, siklus hidup produk, dan tren pasar, pemilik atau manajer toko dapat merancang strategi cerdas untuk pengadaan dan pemeliharaan stok. Ini tidak hanya bertujuan untuk memenuhi kebutuhan pasar saat ini tetapi juga untuk mengantisipasi dan beradaptasi dengan perubahan di masa depan.

Manajemen stok *voucher* dapat menjadi aset strategis yang dapat meningkatkan kinerja bisnis dan daya saing toko jika dipahami secara mendalam. Keputusan terkait persediaan memengaruhi baik operasional toko maupun citranya, menjadikan investasi dalam manajemen stok sangat penting untuk mencapai kesuksesan jangka panjang.

Catatan transaksi di Raffa Cell saat ini belum optimal, yang menyebabkan terjadinya fluktuasi atau ketidakpastian tersedianya stok *voucher* setiap bulan. Untuk mengatasi masalah ini, memprediksi fluktuasi penjualan menjadi sangat penting agar tantangan dalam transaksi dapat diminimalkan. Prediksi ini, yang

didasarkan pada data historis dan tren terkini, membantu memperkirakan kejadian di masa depan dengan tingkat akurasi yang tinggi, meskipun hasilnya mungkin tidak selalu tepat [2].

Dalam penelitian ini yang berjudul "IMPLEMENTASI ALGORITMA REGRESI LINEAR DALAM PREDIKSI PERSEDIAAN *VOUCHER* DI RAFFA CELL SUKABUMI", digunakan algoritma regresi linear untuk menghasilkan sebuah model *machine learning* yang dapat membantu dalam hal manajemen persediaan *voucher* dan juga diharapkan dapat menjadi suatu panduan bagi instansi terkait.

2. TINJAUAN PUSTAKA

2.1. Persediaan *Voucher*

Persediaan *voucher* merujuk pada jumlah *voucher* dagang yang masih tersedia pada akhir periode tertentu. Sebagai alternatif, ini juga dapat diartikan sebagai *voucher* dagang yang diperoleh dengan tujuan untuk dijual kembali. Berdasarkan definisi ini, persediaan *voucher* mencakup *voucher* yang dimaksudkan untuk dibeli dan dijual kembali, dengan jumlah total yang ditentukan pada akhir periode tertentu [3].

2.2. Machine Learning

Pembelajaran mesin adalah metode pemrograman komputer yang menggunakan data historis untuk melatih sebuah model, sehingga memungkinkan model tersebut berfungsi secara optimal dalam mengekstraksi informasi dari suatu set data [4].

Tujuan utama pembelajaran mesin adalah untuk mengembangkan model yang dapat menangkap pola data. Menurut Tom M. Mitchell, pembelajaran mesin didefinisikan sebagai program komputer yang meningkatkan kinerjanya dalam suatu tugas dengan belajar dari pengalaman.

Dari penjelasan tersebut, dapat disimpulkan bahwa pembelajaran mesin adalah teknik pemrograman komputer yang memanfaatkan data historis untuk melatih model. Tujuan utamanya adalah agar model tersebut mencapai kinerja tinggi dalam mengidentifikasi pola data dan mengekstraksi wawasan yang bermakna dari suatu set data.

2.3. Prediksi

Prediksi adalah usaha untuk memperkirakan atau mengantisipasi peristiwa di masa depan dengan menggunakan informasi yang relevan dari periode sebelumnya (historis) melalui pendekatan ilmiah [5].

Dalam prediksi, akurasi dihitung dengan membandingkan jumlah data positif benar dan negatif benar terhadap total data yang ada. Secara sederhana, akurasi menunjukkan seberapa dekat hasil prediksi dengan nilai sebenarnya. Rumus akurasi mencerminkan hubungan ini [6].

2.4. Regresi Linear

Regresi Linear adalah metode statistik yang digunakan untuk menganalisis dan memahami hubungan linier antara variabel dependen (output) dan satu atau lebih variabel independen (input). Tujuan utamanya adalah untuk mengukur dan memodelkan hubungan ini sehingga dapat digunakan untuk membuat prediksi atau estimasi. Metode ini didasarkan pada asumsi bahwa hubungan antara variabel dependen dan independen dapat diwakili oleh garis lurus yang dinyatakan dalam persamaan matematika. Dalam persamaan ini, Y mewakili kelas; X1, X2, ..., Xn mewakili nilai atribut; dan a, b1, ..., bn adalah bobot yang diperoleh dari data sampel.

$$\begin{gathered}\sum y = a_n + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum x_1 y = a \sum x_1 + b_1 \sum x_1 x_2 + b_2 \sum x_1 x_2 \quad (1) \\ \sum x_2 y = a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2\end{gathered}$$

Keterangan:
Y = Variabel tidak bebas (nilai yang diprediksikan)
X = Variabel bebas
a = Konstanta (nilai Y apabila X1, X2… Xn = 0)
b = Koefisien regresi (nilai peningkatan ataupun penurunan).

Regresi linear dapat diterapkan di berbagai bidang seperti ekonomi, ilmu sosial, ilmu alam, dan machine learning. Namun, jika hubungan antara variabel-variabel tersebut tidak linear, maka metode regresi non-linear mungkin lebih tepat untuk menganalisis dan memodelkan hubungan tersebut. Regresi linear sederhana pertama kali diperkenalkan oleh Francis Galton pada tahun 1886 untuk mempermudah prediksi keuntungan atau kerugian di masa mendatang. Dengan menggunakan metode ini, kita dapat menghitung data dan membuat prediksi yang lebih akurat untuk mengurangi risiko kerugian [7]

2.5. *Knowledge Discovery in Databases* (KDD)

Knowledge Discovery in Databases (KDD) adalah proses komprehensif dan berulang yang bertujuan untuk menemukan pengetahuan yang berguna dan bermakna dari sejumlah besar data yang kompleks. Proses KDD terdiri dari lima tahapan sebagai berikut:

a. Seleksi Data
   Tidak semua data dalam *database* digunakan; hanya data yang relevan yang dipilih untuk dianalisis atau diekstraksi dari *database* [8]
b. Pra-processing
   Pra-processing data adalah langkah awal dalam pengolahan data. Pada tahap ini, data yang akan diproses disiapkan untuk menghindari adanya data yang tidak konsisten atau mengganggu [8].
c. Transformation
d. Transformasi adalah tahap di mana data diubah atau disesuaikan agar sesuai dengan model atau

algoritma yang akan digunakan dalam proses analisis data [8].

e. *Data Mining*
Ini adalah proses pencarian dan penggalian pengetahuan hingga menghasilkan model yang dapat digunakan untuk mendapatkan informasi yang penting dan berguna [8].

f. Interpretasi/Evaluasi
Tahap ini melibatkan representasi hasil model yang telah dihasilkan serta pengujian akurasi dan kesesuaiannya terhadap data terkait [8].

2.6. *Mean Square Error* (MSE) & *Root Mean Square Error* (RMSE)

Mean Square Error (MSE) adalah salah satu metode yang digunakan untuk mengevaluasi rata-rata kesalahan kuadrat dalam peramalan, yaitu perbedaan rata-rata kuadrat antara nilai yang diprediksi dan nilai aktual. Secara umum, nilai MSE selalu positif (tidak nol), yang mengindikasikan adanya ketidaksempurnaan dalam hasil peramalan. Persamaan 11 menggambarkan bentuk umum dari *Mean Square Error*.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (3)$$

Keterangan:
n = jumlah total data atau observasi
y_i = nilai sebenarnya dari observasi ke-i
$\hat{y}_i$ = nilai prediksi dari observasi ke-i
$(y_i - \hat{y}_i)^2$= selisih kuadrat dari nilai sebenarnya dan nilai prediksi

Root Mean Square Error (RMSE) atau *Root* Mean Square Deviation adalah metode yang digunakan untuk menghitung tingkat kesalahan dalam hasil estimasi. RMSE menunjukkan seberapa besar perbedaan atau deviasi antara hasil estimasi dan nilai sebenarnya. Metode ini digunakan untuk mengukur tingkat kesalahan dari hasil analisis yang dilakukan dengan metode tertentu, seperti penggunaan data pelatihan dan data pengujian. Persamaan 12 menggambarkan bentuk umum dari *Root Mean Square Error* [8].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (4)$$

[9]

Keterangan:
n = jumlah total data atau observasi
y_i = nilai sebenarnya dari observasi ke-i
$\hat{y}_i$ = nilai prediksi dari observasi ke-i
$(y_i - \hat{y}_i)^2$= selisih kuadrat dari nilai sebenarnya dan nilai prediksi

2.7. Python

Python adalah bahasa pemrograman berbasis objek yang mendukung interaksi interaktif. Bahasa ini dikenal karena struktur data tingkat tinggi dan sifat interpretatifnya, dengan fokus pada kejelasan dan kemudahan pemahaman kode. Keunggulan Python terletak pada kombinasi kemampuannya dan sintaksis yang jelas. Bahasa ini dirancang untuk mempermudah *programmer* dalam mengembangkan program dengan efisiensi waktu, kemudahan pengembangan, dan kompatibilitas dengan berbagai sistem. Python bisa digunakan untuk membuat aplikasi mandiri atau skrip pemrograman [10].

2.8. Google Collaboratory

Google Colab, atau Google Colaboratory, adalah dokumen yang dapat dijalankan yang memungkinkan pengguna untuk menyimpan, menulis, dan berbagi kode melalui Google Drive. Platform ini mirip dengan Jupyter Notebook tetapi berbasis cloud, sehingga dapat diakses melalui browser seperti Mozilla Firefox dan Google Chrome. Google Colab memungkinkan eksekusi kode Python tanpa perlu instalasi atau pengaturan tambahan, karena konfigurasi dilakukan otomatis di cloud. Dengan demikian, pengguna dapat dengan mudah menjalankan dan berkolaborasi dalam proyek pemrograman tanpa perlu mengatur konfigurasi lokal [11].

2.9. Visual Studio Code

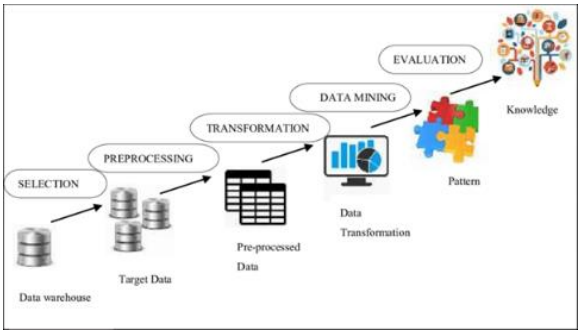
Microsoft telah mengembangkan Visual Studio Code, sebuah editor kode open-source yang tersedia untuk Windows, Linux, dan macOS. Editor ini menawarkan berbagai fitur seperti debugging, integrasi dengan Git dan GitHub, analisis statis, snippet, dan *refactoring* kode. Visual Studio Code sangat fleksibel, memungkinkan pengguna untuk menyesuaikan tema, pintasan keyboard, preferensi, serta memperluas fungsionalitasnya melalui ekstensi untuk meningkatkan kemampuannya [12].

2.10. Flask

Flask adalah *microframework* yang dikembangkan oleh Armin Ronacher. *Framework* ini dikenal karena kecepatan dan ringannya, dirancang untuk mengurangi kompleksitas sebanyak mungkin. Dengan motto "pengembangan web, setetes demi setetes", Flask memungkinkan pembuatan situs web dengan cepat meskipun hanya menggunakan perpustakaan yang terkelola dengan baik [12]

3. METODE PENELITIAN

Metode penelitian bertujuan untuk merinci perencanaan dan pelaksanaan penelitian, termasuk pengumpulan data dan kebutuhan yang diperlukan. Rencana jadwal kegiatan akan dijelaskan untuk memastikan kelancaran dan keberhasilan penelitian. Metode yang digunakan dalam penelitian ini adalah *Knowledge Discovery in Databases* (KDD), yang dirancang untuk memberikan panduan terstruktur pada setiap langkah penelitian.



Gambar 1. Alur Metode KDD

3.1. Seleksi Data

Pada tahap ini, pemilihan data dilakukan dengan teliti, hanya menggunakan data penjualan *voucher* dari tahun lalu. Tujuannya adalah untuk memastikan proses pengolahan data berlangsung lebih lancar dan efisien sesuai dengan tujuan penelitian.

3.2. Pra-processing

Pada tahap ini, data *voucher* diperiksa untuk menghapus duplikat dan mengisi nilai yang kosong. Langkah ini bertujuan untuk mengoptimalkan input data dan meminimalkan kesalahan, sehingga hasil perhitungan algoritma menjadi akurat.

3.3. Transformation

Beberapa metode data mining memerlukan struktur data tertentu sebelum diterapkan. Oleh karena itu, data harus diubah untuk memenuhi persyaratan metode data mining yang akan digunakan, mempengaruhi hasil akhir dari proses tersebut.

3.4. Data Mining

Pada tahap ini, model prediksi dibuat menggunakan metode Regresi Linear untuk mengidentifikasi pola atau wawasan berharga dari data. Pembuatan model dilakukan dengan menggunakan Google Colaboratory.

3.5. Interpretasi/Evaluasi

Hasil analisis data disajikan dalam format yang mudah dipahami. Pola atau informasi yang ditemukan dievaluasi untuk memverifikasi konsistensinya dengan fakta atau hipotesis sebelumnya. Presentasi hasil disederhanakan untuk mempermudah pemahaman dan evaluasi pola informasi.

3.6. Deployment

Untuk meningkatkan pengalaman pengguna, peneliti telah mengembangkan antarmuka berbasis situs web yang dapat memprediksi ketersediaan voucer. Dalam pengembangan sistem ini, Visual Studio Code digunakan untuk merancang dan mengatur struktur serta logika program. Model hasil analisis data diintegrasikan ke dalam platform web, dengan Visual Studio Code membantu dalam manajemen dan modifikasi kode, memastikan kualitas dan kontinuitas situs web prediksi persediaan.

4. HASIL DAN PEMBAHASAN

4.1. Seleksi Data

Penelitian ini menggunakan data *voucher* yang dikumpulkan selama tahun 2023 dari Raffa Cell Sukabumi. Data ini telah diproses dan akan digunakan untuk memprediksi penjualan *voucher* di tahun 2024. Atribut yang dipertimbangkan dalam prediksi termasuk nama *provider*, jumlah byte, masa aktif, harga awal, harga jual, stok awal, dan stok akhir.

Tabel 1. Deskripsi Atribut

No	Atribut	Deskripsi
1	Nama *Provider*	Nama layanan yang menyediakan layanan
2	Jumlah Byte	Jumlah data yang ditawarkan dalam paket layanan
3	Masa Aktif	Durasi atau periode waktu selama layanan digunakan setelah diaktifan
4	Harga Awal	Harga yang dikenakan oleh penyedia layanan
5	Harga Jual	Harga yang dikenakan pada konsumen
6	Stok Awal	Jumlah unit pada layanan yang tersedia pada awal
7	Stok Akhir	Jumlah unit pada layanan yang tersisa

Langkah pertama dalam proses *Knowledge Discovery in Databases* (KDD) adalah seleksi data, di mana data yang relevan dipilih untuk analisis lebih lanjut. Pada tahap ini, pandas digunakan untuk memuat data prediksi *voucher*, dan beberapa baris pertama ditampilkan untuk memberikan gambaran tentang atribut dan isinya. Langkah ini penting untuk memahami data yang tersedia dan formatnya.

```
dataset =
pd.read_excel('/content/updated_vo
ucherkuota (23).xlsx')
dataset
```

	Nama_Provider	Jumlah_Byte_(gb)	Masa_Aktif_(hari)	Harga_Awal	Harga_Jual	Stok_Awal	Stok_Akhir	Bulan	Tahun
0	2	1	1	5000	7000	100	50	1	2023
1	2	2	2	5500	8000	300	150	1	2023
2	2	2	3	6000	10000	150	50	1	2023
3	2	3	5	10000	13000	150	100	1	2023
4	2	4	5	12000	14000	200	150	1	2023
...	...	...	...	...	...	...	...	...	...

Gambar 2. Atribut Dataset

Fungsi `dataset` digunakan untuk membaca data dari file Excel yang berisi fitur-fitur terkait hasil prediksi, seperti stok awal, stok akhir, dan harga jual. Tahap ini memastikan bahwa data yang diperlukan untuk analisis lebih lanjut dapat diakses dan siap untuk diproses.

4.2. Pra-processing

Langkah kedua dalam proses KDD adalah pra-pemrosesan awal data, yang melibatkan pembersihan data untuk meningkatkan kualitasnya. Proses ini dimulai dengan menampilkan informasi dasar tentang data, seperti tipe data dan jumlah nilai yang hilang, menggunakan `data.info()`. Jumlah baris duplikat

dihitung dengan `data.duplicated().sum()` dan jumlah data yang hilang dengan `data.isna().sum()`. Informasi ini membantu mengidentifikasi masalah potensial dalam data, seperti duplikasi atau nilai yang hilang, yang perlu diperbaiki.

a. Mencari Informasi Dataset
Langkah ini bertujuan untuk menentukan tindakan yang diperlukan untuk setiap kolom yang akan digunakan. Beberapa kolom mungkin mengandung data yang perlu diperhatikan.

```
Nama_Provider        object
Jumlah_Byte_(gb)      int64
Masa_Aktif_(hari)     int64
Harga_Awal            int64
Harga_Jual            int64
Stok_Awal             int64
Stok_Akhir            int64
Bulan                object
Tahun                 int64
dtype: object
```

Gambar 3. Informasi Dataset

b. Mengidentifikasi Nilai Duplikat

```
dataset_duplikat =
dataset[dataset.duplicated()]
print(len(dataset_duplikat))
dataset.isna().sum()
```

Fungsi `data.duplicated().sum()` digunakan untuk menghitung jumlah baris duplikat dalam DataFrame. Hasil yang diperoleh adalah 0, menunjukkan bahwa tidak ada baris duplikat dalam data tersebut.

c. Mengidentifikasi Nilai yang Hilang
Jumlah nilai yang hilang dihitung untuk memastikan semua data yang diperlukan tersedia.

```
Nama_Provider        0
Jumlah_Byte_(gb)     0
Masa_Aktif_(hari)    0
Harga_Awal           0
Harga_Jual           0
Stok_Awal            0
Stok_Akhir           0
Bulan                0
Tahun                0
dtype: int64
```

Gambar 4. Identifikasi Nilai Hilang

d. Menghitung Matriks Korelasi
Setelah itu, fitur numerik relevan seperti nama *provider*, jumlah byte, stok awal, stok akhir, harga awal, dan harga akhir dipilih untuk analisis lebih lanjut. Matriks korelasi dihitung dan ditampilkan menggunakan seaborn dan matplotlib. Matriks ini membantu mengidentifikasi hubungan antara variabel yang berbeda, serta pola atau tren dalam data. Dengan menganalisis matriks korelasi, hubungan antar variabel dapat dipahami lebih baik, dan pola dalam data dapat diungkap.

```
columns =
dataset.select_dtypes(include=[np.n
umber])
kolom_numerik =
columns.columns.tolist()
datanumerik =
dataset[kolom_numerik]
correlation_matrix =
datanumerik.corr()
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix,
annot=True, cmap='viridis',
linewidths=0.5, fmt=".2f")
plt.title('korelasi antar kolom
dataset')
plt.show()
```

4.3. Transformation

Fase ketiga dalam proses KDD dikenal sebagai transformasi data, yang melibatkan pengubahan data yang telah diproses sebelumnya menjadi format yang siap untuk data mining. Pada tahap ini, fitur-fitur seperti nama *provider*, jumlah GB, masa aktif, harga awal, harga jual, stok awal, stok akhir, dan bulan dipilih untuk dimasukkan ke dalam model prediksi. Fitur-fitur ini digunakan sebagai variabel (X) dalam prediksi, sementara hasil prediksi ditempatkan sebagai target (y) dalam data.

Tabel 2. Detail Penggunaan Atribut

Atribut	Detail Penggunaan
Tahun	×
Nama *Provider*	√
Jumlah GB	√
Masa Aktif	√
Harga Awal	√
Harga Jual	√
Stok Awal	√
Stok Akhir	√
Bulan	√

Atribut-atribut yang digunakan dalam penelitian dicantumkan pada tabel di atas. Atribut yang dipilih untuk digunakan ditandai dengan tanda centang (√), sedangkan atribut yang dihilangkan pada tahap penyiapan data ditandai dengan tanda silang (×).

a. Menentukan Label yang Akan Diprediksi

```
X =
dataset.drop(columns=['Stok_Awal'])
y = dataset['Stok_Awal']
```

Pada tahap ini, data yang relevan untuk analisis dipilih. Fitur X digunakan untuk melakukan prediksi, sementara Y adalah nilai yang ingin diprediksi.

b. Menormalisasi Data

```
normalisasi = MinMaxScaler()
normalisasi.fit(X)
data_digunakan_normalisasi =
normalisasi.transform(X)
with open('normalisasi.pkl', 'wb')
as f:
pickle.dump(normalisasi, f)
X =
pd.DataFrame(data_digunakan_normal
isasi,columns=X.columns)
print(X)
```

Normalisasi data menggunakan StandardScaler bertujuan untuk memastikan semua fitur berada dalam skala yang sama. Langkah ini membantu meningkatkan performa model *machine learning* dengan mengurangi variasi skala antar fitur.

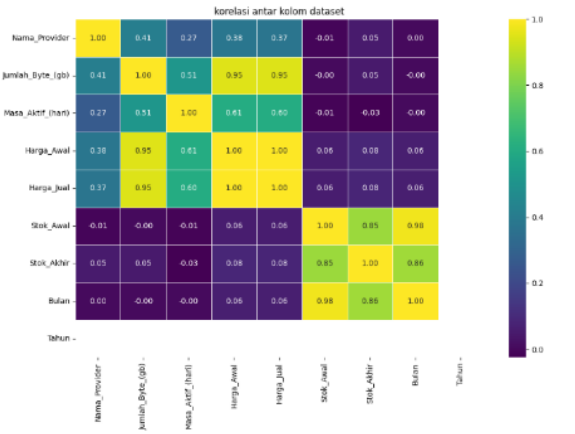
```
      Nama_Provider  Jumlah_Byte_(gb)  Masa_Aktif_(hari)  Harga_Awal  \
0          0.666667          0.000000           0.000000    0.000000
1          0.666667          0.011236           0.034483    0.004525
2          0.666667          0.011236           0.068966    0.009050
3          0.666667          0.022472           0.137931    0.045249
4          0.666667          0.033708           0.137931    0.063348
...             ...               ...                ...         ...
1519       0.000000          0.235955           1.000000    0.411765
1520       0.000000          0.269663           1.000000    0.429864
1521       0.000000          0.325843           1.000000    0.457014
1522       0.000000          0.382022           1.000000    0.475113
1523       0.000000          0.438202           1.000000    0.502262

      Harga_Jual  Stok_Akhir  Bulan  Tahun
0       0.000000    0.041096    0.0    0.0
1       0.008811    0.219178    0.0    0.0
2       0.026432    0.027397    0.0    0.0
3       0.052863    0.082192    0.0    0.0
4       0.061674    0.315068    0.0    0.0
...          ...         ...    ...    ...
1519    0.392070    0.931507    1.0    0.0
1520    0.427313    0.876712    1.0    0.0
1521    0.444934    0.767123    1.0    0.0
1522    0.462555    0.726027    1.0    0.0
1523    0.488987    0.712329    1.0    0.0

[1524 rows x 8 columns]
```

Gambar 5. Normalisasi Data

c. Memeriksa Korelasi



Gambar 6. Visualisasi Matriks Korelasi

Visualisasi heatmap dari nilai matriks korelasi, dengan penjelasan sebagai berikut:

- +1: Korelasi positif sempurna.
- -1: Korelasi negatif sempurna.
- 0: Tidak ada korelasi linier.
- 0.0 hingga 0.3 (atau -0.0 hingga -0.3): Korelasi lemah.
- 0.3 hingga 0.7 (atau -0.3 hingga -0.7): Korelasi sedang.
- 0.7 hingga 1.0 (atau -0.7 hingga -1.0): Korelasi kuat.

Visualisasi heatmap korelasi tersebut menunjukkan hubungan antar variabel dalam dataset. Hasil menunjukkan ada korelasi sangat kuat antara "Jumlah_Byte_(gb)" dengan "Harga_Awal" dan "Harga_Jual", masing-masing dengan nilai korelasi 0,95, yang menunjukkan bahwa peningkatan jumlah byte cenderung diikuti dengan kenaikan harga awal dan harga jual. Selain itu, ada juga korelasi kuat antara "Stok_Awal" dan "Stok_Akhir" (0,85), serta antara "Bulan" dengan "Stok_Awal" (0,98) dan "Stok_Akhir" (0,86), menunjukkan bahwa jumlah stok awal dan akhir cenderung konsisten di berbagai bulan. Sementara itu, "Nama_Provider" dan "Tahun" memiliki korelasi yang rendah atau hampir tidak ada dengan variabel lainnya, yang berarti kedua variabel ini tidak memiliki hubungan linear yang kuat dengan variabel lain dalam dataset.

4.4. Data Mining

```
X_latih, X_test, y_latih, y_test =
train_test_split(X, y,
test_size=0.2, random_state=13)
```

Tahap keempat dalam penelitian ini yaitu data mining di mana algoritma regresi linear digunakan untuk menemukan pola dalam data. Proses ini dimulai dengan membagi data menjadi set pelatihan dan set pengujian menggunakan "*train_test_split*" dari scikit-learn, dengan 80% data untuk pelatihan dan 20% untuk pengujian. Ini penting untuk menguji kinerja model pada data baru dan mencegah *overfitting*. Setelah itu, model regresi linear dilatih pada set pelatihan untuk memodelkan hubungan antara variabel independen dan dependen. Model yang sudah dilatih kemudian disimpan menggunakan Pickle, memungkinkan prediksi di masa depan tanpa perlu melatih ulang model, sehingga efisien dan siap digunakan dalam aplikasi nyata.

4.5. Interpretasi/Evaluasi

Tahap kelima dalam proses KDD adalah evaluasi, di mana pola yang telah ditemukan diperiksa untuk memastikan keakuratan dan relevansinya. Berbagai metrik digunakan untuk menilai kinerja model, termasuk *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Mean Squared Error* (MSE), dan R^2. RMSE memberikan estimasi kesalahan prediksi dalam satuan yang sama dengan keluaran, menunjukkan sejauh mana prediksi model menyimpang dari nilai aktual. MAE mengukur total kesalahan prediksi tanpa mempertimbangkan arah

kesalahan, mengkuantifikasi perbedaan rata-rata antara nilai yang sebenarnya dan yang diprediksi. MSE memberikan bobot lebih pada kesalahan besar karena sifat pengkuadratannya. R^2 mengukur seberapa besar variabilitas dalam data yang dapat dijelaskan oleh model.

a. Uji Performa Model

```
rmse_lr =
np.sqrt(mean_squared_error(y_test,
y_pred))
mae_lr =
mean_absolute_error(y_test,
y_pred)
mse_lr =
mean_squared_error(y_test, y_pred)
r2_lgb = r2_score(y_test,
y_pred_lgb)print(f'Linear
Regression - MSE: {mse_lr}, MAE:
{mae_lr}, R²: {r2_lgb}, RMSE:
{rmse_lr}')
```


Gambar 7. Evaluasi Performa Model

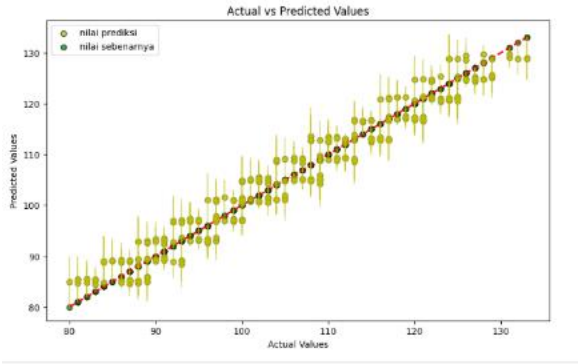
Evaluasi performa model menunjukkan:

- RMSE sebesar 2.91003 menunjukkan kesalahan prediksi yang relatif kecil.
- MAE sebesar 2.56135 menunjukkan bahwa rata-rata prediksi model berbeda dari nilai aktual sebesar 2.56135.
- MSE sebesar 8.4682 menunjukkan kesalahan kuadrat rata-rata yang kecil.
- R^2 sebesar 0.98359 menunjukkan bahwa model hampir sepenuhnya menjelaskan variasi dalam data

b. Interpretasi Model

Tahap berikutnya adalah presentasi pengetahuan, di mana hasil temuan disajikan secara mudah dipahami. Plot sebar dibuat untuk membandingkan nilai aktual dengan prediksi, memberikan gambaran yang jelas tentang seberapa baik model memprediksi hasil.

```
# Calculate errors
errors = y_test - y_pred
plt.figure(figsize=(10, 6))
plt.scatter(y_test, y_pred,
color='y', edgecolors=(0, 0, 0),
alpha=0.7, label='nilai prediksi')
plt.plot([y_test.min(),
y_test.max()], [y_test.min(),
y_test.max()], 'r--', lw=2)
plt.scatter(y_test, y_test,
color='g', edgecolors=(0, 0, 0),
alpha=0.7, label='nilai
sebenarnya')
# Add error bars
plt.errorbar(y_test, y_pred,
yerr=np.abs(errors), fmt='o',
color='y', alpha=0.5)
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
plt.title('Actual vs Predicted
Values')
plt.legend()
plt.show()
```

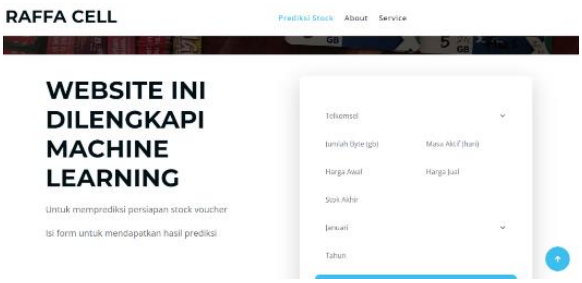


Gambar 8. Interpretasi Model

Gambar yang menunjukkan "Nilai Aktual vs. Nilai Prediksi" menunjukkan semua titik prediksi (lingkaran biru) berada sangat dekat atau tepat pada garis identitas merah, mencerminkan performa model yang sangat baik dengan kesalahan prediksi minimal. Nilai RMSE, MAE, dan MSE yang sangat rendah serta nilai R^2 yang mendekati 1 menunjukkan efektivitas model dalam memprediksi dengan presisi tinggi.

4.6. Halaman Home

Halaman ini merupakan halaman utama website prediksi ini.



Gambar 9. Halaman *Home*

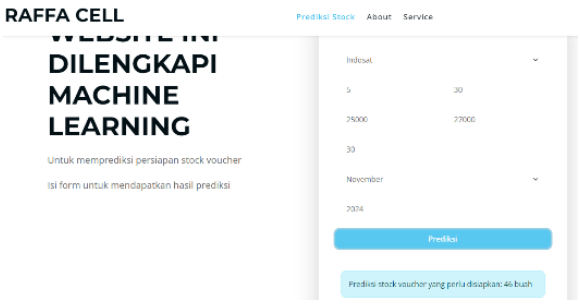
4.7. Halaman About

Halaman ini berisi deskripsi singkat tentang Raffa Cell.

Gambar 10. Halaman *About*

4.8. Halaman Hasil Prediksi Voucher

Halaman ini berisi hasil prediksi stok *voucher* yang harus disediakan berdasarkan data yang dimasukkan.



Gambar 11. Halaman Hasil Prediksi

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan situs web yang memanfaatkan atribut standar seperti nama penyedia, jumlah GB, stok awal dan akhir, harga awal, serta harga jual, untuk meramalkan kebutuhan voucher di masa depan. Algoritma regresi linear yang diintegrasikan ke dalam situs web tersebut menunjukkan kinerja optimal, dengan hasil prediksi yang akurat ditunjukkan oleh nilai *Root Mean Square Error* (RMSE) sebesar 29%, *Mean Absolute Error* (MAE) sebesar 25%, *Mean Squared Error* (MSE) sebesar 84%, dan tingkat akurasi 90%. Hal ini menunjukkan bahwa model yang dikembangkan merupakan alat yang andal untuk memprediksi inventaris *voucher*, memberikan wawasan penting bagi manajemen inventaris dan pengambilan keputusan.

Untuk meningkatkan prediksi di masa depan dan menyempurnakan model, disarankan agar penelitian berikutnya memasukkan faktor tambahan seperti data penjualan sebelumnya, promosi, dan lokasi penjualan, serta membandingkan regresi linear dengan algoritma lain seperti *Random Forest*, *K-Nearest Neighbors* (KNN), atau *Support Vector Machine* (SVM). Menambahkan fitur seperti dasbor untuk melacak *voucher* yang terjual juga akan memberikan wawasan yang lebih komprehensif.

DAFTAR PUSTAKA

[1] A. F. Qadafi and A. D. Wahyudi, "Sistem Informasi Inventory Gudang Dalam Ketersediaan Stok Barang Menggunakan Metode Buffer Stok," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 1, no. 2, pp. 174–182, 2020, doi: 10.33365/jatika.v1i2.557.

[2] M. Kafil, "Penerapan Metode K-Nearest Neighbors," *J. Mhs. Tek. Inform.*, vol. 3, no. 2, pp. 59–66, 2019.

[3] H. D. Yulianto and D. F. Maulana, "Perancangan Sistem Informasi Akuntansi Persediaan Barang Dagang Menggunakan SAK EMKM Berbasis Web," *is Best Account. Inf. Syst. Inf. Technol. Bus. Enterp. this is link OJS us*, vol. 5, no. 2, pp. 121–135, 2020, doi: 10.34010/aisthebest.v5i2.3244.

[4] K. Amalia and S. Hidayat, "Analisis Kemandirian Belajar Menggunakan Model Discovery Learning dalam Pembelajaran Jarak Jauh," *PEDADIDAKTIKA J. Ilm. Pendidik. Guru Sekol. Dasar*, vol. 8, no. 3, pp. 621–631, 2021, doi: 10.17509/pedadidaktika.v8i3.39231.

[5] W. A. P. Wanto Anjar, "Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan MWanto, A. (2019). Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan Metode Backpropagation. Jurnal & Penelitian Teknik Infor," *J. Penelit. Tek. Inform.*, vol. 2, no. 2, pp. 37–44, 2017, [Online]. Available: https://zenodo.org/record/1009223#.Wd7norlTbhQ

[6] D. Derisma, "Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining," *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 84–88, 2020, doi: 10.30871/jaic.v4i1.2152.

[7] S. Supardi *et al.*, "Peran Data Mining dalam Memprediksi Tingkat Penjualan Sepatu Adidas Menggunakan Metode Algoritma Regresi Linear Sederhana," *J. Ekon. Manjemen Sist. Inf.*, vol. 4, no. 5, pp. 883–890, 2023.

[8] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 9, no. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.

[9] A. N. Maharadja, I. Maulana, and B. A. Dermawan, "Penerapan Metode Regresi Linear Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi," *J. Appl. Informatics Comput.*, vol. 5, no. 1, pp. 95–102, 2021, doi: 10.30871/jaic.v5i1.3184.

[10] A. Triono, A. S. Budi, and R. Abdillah, "Implementasi Peretasan Sandi Vigenere Chipper Menggunakan Bahasa Pemograman Python," *J. JOCOTIS - J. Sci. Inform. Robot.*, vol. 1, no. 1, pp. 1–9, 2023, [Online]. Available: https://jurnal.ittc.web.id/index.php/jumri

[11] A. Sitio, A. Sindar, M. Marbun, D. Tiara, and A. Aswin, "Pengenalan Data Scientist Pada Peserta PKBM AL HABIB Melalui Belajar Dasar Coding Python," *J. Pengabdi. Pada Masy.*, vol. 7, no. 1, pp. 194–200, 2022, doi: 10.30653/002.202271.44.

[12] Agustini and W. J. Kurniawan, "Sistem E-Learning Do'a dan Iqro' dalam Peningkatan Proses Pembelajaran pada TK Amal Ikhlas," *J. Mhs. Apl. Teknol. Komput. dan Inf.*, vol. 1, no. 3, pp. 154–159, 2019, [Online]. Available: http://www.ejournal.pelitaindonesia.ac.id/JMApTeKsi/index.php/JOM/article/view/526