ANALISIS TINGKAT PENANGANAN SAMPAH DI JAWA BARAT MENGGUNAKAN *REGRESI LINIER*

Alfian Nur Alam [1], Ade Irma Purnamasari [2], Irfan Ali [3]
[1] Teknik Informatika, STMIK IKMI Cirebon
[2] Teknik Informatika, STMIK IKMI Cirebon
[3] Rekayasa Perangkat Lunak, STMIK IKMI Cirebon
Jalan Perjuangan No. 10 B Majasem, Kec.Kesambi, Kota Cirebon, Jawa Barat 45135, Indonesia
alfiannuralam68@gmail.com

ABSTRAK

Sampah merupakan masalah besar bagi masyarakat dan lingkungan di Jawa Barat. Volume sampah yang meningkat setiap tahunnya menyebabkan penumpukan di tempat penampungan sementara (TPS). Hal ini disebabkan oleh berbagai faktor, seperti kepadatan penduduk, karakteristik lingkungan, kondisi sosial ekonomi, norma budaya, dan sikap masyarakat Dalam penelitian ini menerapkan proses *Knowledge Discovery InDatabases (KDD)*, data mining digunakan untuk memprediksi volume dan mengidentifikasi faktor yang paling berpengaruh pada timbunan sampah yang tidak tertangani di Jawa Barat. Hasil penelitian menunjukkan bahwa jumlah produksi sampah dan jumlah sampah terkelola dapat mempengaruhi tingkat timbunan sampah dan, jumlah timbunan sampah di Jawa Barat pada tahun 2023 sebesar 38615 ton per *hari*. Jumlah ini menurun sebesar 300169 ton per hari dari jumlah timbunan 5 tahun sebelumnya. Hasil penelitian ini menunjukkan bahwa diperlukan upaya yang lebih komprehensif dari berbagai pihak untuk mengatasi masalah sampah di Jawa Barat. Pemerintah daerah perlu menyusun tindakan yang tepat, baik dari segi prasarana dan sarana, maupun dari segi sumber daya manusia. Selain itu, diperlukan juga kesadaran dan partisipasi masyarakat dalam pengelolaan sampah.

Kata kunci: *Prediksi, Linear Regression, Pengelolaan Sampah, Sampah*

1. PENDAHULUAN

Penanganan sampah merupakan isu serius yang berdampak pada lingkungan dan kehidupan rumah tangga. Solusi untuk mengatasi hal ini harus merangkul semua pihak dan dilakukan secara menyeluruh. Berbagai faktor seperti kepadatan penduduk, karakteristik lingkungan, aspek sosial ekonomi, budaya, sikap, dan perilaku masyarakat menjadi penghambat utama dalam upaya pengelolaan sampah. Meskipun begitu, sampah tetap menjadi masalah kompleks dan berkelanjutan yang memerlukan keterlibatan berbagai pemangku kepentingan dengan mempertimbangkan karakteristik sampah serta aspek sosial-budaya masyarakat lokal[1]. Volume sampah yang terus meningkat dan menumpuk di Tempat Penampungan Sementara (TPS) perumahan dan pemukiman adalah faktor utama. Pengelolaan yang kurang optimal, fasilitas yang tidak memadai, transportasi sampah, produksi, dan akumulasi sampah turut berperan. Dalam ranah Informatika, data mining menjadi solusi penting yang berdampak signifikan pada berbagai aspek.

Metode ini memproses data untuk mendukung pengambilan keputusan, terutama dalam menangani akumulasi sampah di Provinsi Jawa Barat. Melalui Algoritma Linear Regression, penelitian ini bertujuan memprediksi akumulasi sampah berdasarkan wilayah kabupaten dan kota serta volume harian dan tahunan di Provinsi Jawa Barat. Penelitian ini bertujuan untuk memprediksi faktor dan timbunan sampah berdasarkan tahun, wilayah dan volume pada sampah. Penelitian ini juga memiliki potensi untuk memberikan kontribusi pada bidang Informatika dan lingkungan hidup, dalam penanganan kasus akumulasi sampah, penelitian ini dapat menjadi pendukung keputusan dalam penanganan kasus sampah untuk pemerintah setempat. Hasil penelitian ini dapat digunakan sebagai referensi untuk penelitian lebih lanjut yang berfokus pada penggunaan teknologi informasi untuk memecahkan masalah lingkungan. Hasil ini juga dapat berfungsi sebagai dasar untuk desain metodologi dan algoritma yang lebih baik serta pemilihan algoritma yang tepat.

2. TINJAUAN PUSTAKA

Upaya Memasyarakatkan Bank Sampah untuk Meningkatkan Pendapatan dan Memberdayakan Perempuan di Margasari. Mendiskusikan inisiatif bank sampah dalam meningkatkan kesadaran masyarakat, terutama perempuan, dalam mengelola sampah dan menjaga lingkungan. Selain itu, program ini bertujuan untuk mendorong pemerintah dan masyarakat dalam mengatasi permasalahan sampah, sambil juga menjadi solusi ekonomi bagi para ibu rumah tangga[2].

Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi. Dalam penelitian ini, metode regresi linear digunakan untuk memprediksi hasil panen tanaman padi berdasarkan data petani Lamongan. Langkah-langkah yang digunakan meliputi pra-pemrosesan data, penerapan regresi linear, dan validasi hasil. Selain itu, pembahasan mencakup penggunaan uji korelasi parsial dan koefisien determinasi untuk mengevaluasi bagaimana variabel independen mempengaruhi variabel dependen dalam model regresi linear[3].

Segmentasi dan Pembentukan Model Regresi Nasabah Berbasis Analisis Recency, Frequency dan Monetary. Studi ini mengeksplorasi cara perusahaan sekuritas dapat memanfaatkan analisis data untuk membagi nasabah ke dalam kelompok-kelompok yang berbeda dan memprediksi bagaimana mereka akan bertransaksi. Dengan memahami kelompok-kelompok dan kebiasaan transaksi nasabah, perusahaan dapat meningkatkan layanan dan nilai bisnisnya[4].

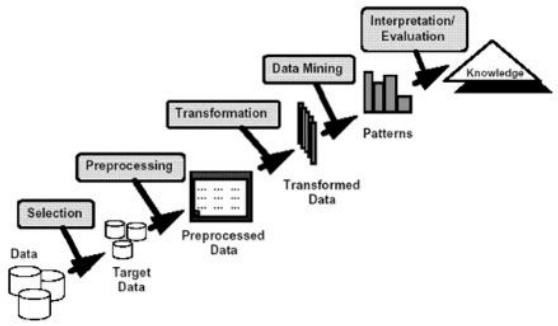
Analisis kinerja karyawan dilakukan untuk menentukan target dan menilai kinerja karyawan. Penilaian kinerja karyawan digunakan untuk menentukan kenaikan golongan atau promosi pegawai. Metode klasifikasi decision tree dan regresi logistik digunakan untuk memprediksi kinerja karyawan[5].

2.1. Algoritma *Regresi Linear*

Algoritma regresi linear adalah satu algoritma machine learning yang digunakan untuk memprediksi nilai dari sebuah variabel target berdasarkan nilai dari beberapa variabel input (atau fitur). Algoritma ini mengasumsikan bahwa hubungan antara variabel target dan variabel input adalah linier. Regresi linier digunakan untukmencari informasi tentang variabel bebas yang memiliki korelasi dengan variabel terikat[6][7][8].

2.2. Data *Mining*

Knowledge Discovery in Database (KDD) adalah metode yang digunakan untuk menemukan informasi yang berharga dari database yang ada. Hasil pengetahuan yang diperoleh dapat digunakan untuk membangun basis pengetahuan *(knowledge base)* yang digunkan dalam keperluan mengambil keputusan. Proses KDD melibatkan beberapa tahapan, seperti pembersihan data, transformasi data, data mining, dan interpretasi dan evaluasi[9][10], proses KDD seperti pada gambar berikut:



Gambar 1. Tahapan proses KDD (Sumber: https://dosbing.id/wp-content/uploads/2019/11/Annotation-2019-11-26-192902.png)

2.2.1. *Selection*

Pada tahap *selection* dalam data mining, variabel yang akan digunakan untuk analisis perlu dipilih secara hati-hati agar tidak ada data yang sama dan terjadi perulangan.

2.2.2. *Prepocessing*

Pada prepocessing terdapat dua tahap, yaitu data *cleaning*, langkah untuk membersihkan data dari data yang tidak relevan. Data yang tidak relevan tersebut dapat berupa data yang hilang, data yang berisi noise, dan data yang tidak konsisten, dan tahap data *Integration* adalah proses menggabungkan data dari berbagai sumber yang berbeda menjadi satu kumpulan data.

2.2.3. *Transformation*

Tahap ini bertujuan untuk mengubah format data agar sesuai dengan kebutuhan analisis. Data *transformation* dapat dilakukan dengan berbagai cara, seperti mengubah tipe data, mengubah ukuran data, atau mengubah struktur data.

2.2.4. *Data Mining*

Data mining adalah proses mengekstrak pola atau pattern dari data menggunakan berbagai teknik seperti statistik, machine learning, dan artificial intelligence.

2.2.5. *Interpretation/evaluation*

Tahapan ini bertujuan untuk memahami dan mengevaluasi pola atau pattern yang telah ditemukan. Tahapan ini penting untuk memastikan bahwa pola atau pattern yang ditemukan adalah pola atau pattern yang valid dan dapat digunakan untuk tujuan penelitian

2.3. *Root Mean Squared Error (RMSE)*

Root Mean Squared Error (RMSE) merupakan metode evaluasi yang digunakan untuk mengukur akurasi model regresi linear. RMSE dihitung dengan melakukan pengkuadratan terhadap selisih antara nilai prediksi dengan nilai observasi, kemudian hasilnya dibagi dengan jumlah data dan diambil akar kuadratnya. RMSE tidak memiliki satuan tertentu dan sering digunakan sebagai standar untuk mengukur tingkat kesalahan model dalam melakukan prediksi terhadap data[11]. Secara matematis, rumus RMSE dapat dijabarkan sebagai berikut:

$$\text{RMSE} = \frac{\sqrt{\sum yt - yt^2}}{n}$$

RMSE = Root Mean Square Error
n = Jumlah Sampel
yt = Nilai Aktual Indeks
yt^2 = Nilai Prediksi Ind

2.4. *R2-Score*

R2 score, atau koefisien determinasi, merupakan metrik yang digunakan untuk mengukur besar kecilnya sumbangan atau kontribusi perubahan variabel bebas dalam model regresi linier. Persamaan regresi linier dapat ditulis sebagai berikut:

$$R^2 = \frac{(y-y)^2}{(Y-y)^2 + (X-X)^2}$$

Y = Variabel dependen
n = Variabel independen
yt = Nilai rata-rata dari variabel dependen
yt^2 = Nilai rata-rata dari variabel dependen

3. METODE PENELITIAN

3.1. Sumber Data

Data penelitian ini diperoleh dari situs web Open Data Jabar yang berisi informasi mengenai faktor-faktor kasus permasalahan sampah berdasarkan kota/kabupaten di provinsi jawa barat. Data ini mencakup atribut-atribut seperti id, kode_provinsi, nama_kabupaten/kota, jumlah_timbunan, jumlah_terkelola, satuan, dan tahun. Dataset yang digunakan terdapat beberapa atribut yang berbeda dari 3 dataset yang akan digunakan, berkut detail atribut pada masing-masing dataset yang digunakan.

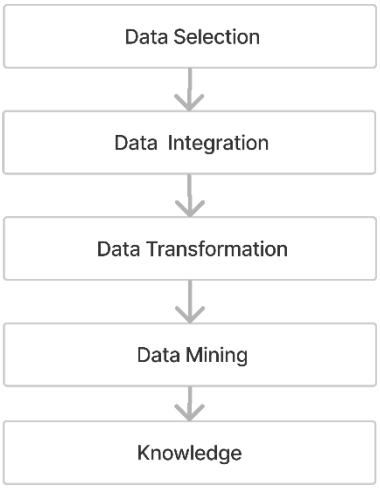
Tabel 3. Dataset Penelitian

No	Nama Dataset
1	Sampah Yang Ditangani
2	Timbunan Sampah
3	Produksi Sampah

1. Sampah Yang Ditangani
   dataset ini meliputi atribut id, kode provinsi, nama provinsi, kode kabupaten kota, nama kabupaten kota, jumlah sampah, satuan dan tahun, data ini berdasarkan wilayah perumahan dan permukiman di provinsi jawa barat.
2. Timbunan Sampah
   dataset ini meliputi atribut id, kode provinsi, nama provinsi, kode kabupaten kota, nama kabupaten kota, jumlah sampah, satuan dan tahun, data ini berdasarkan wilayah perumahan dan permukiman di provinsi jawa barat.
3. Produksi Sampah
   dataset ini meliputi atribut id, kode provinsi, nama provinsi, kode kabupaten kota, nama kabupaten kota, jumlah sampah, satuan dan tahun, data ini berdasarkan wilayah perumahan dan permukiman di provinsi jawa barat.

3.2. Metode Penelitian

1. Data *Selection*
   Melakukan pengambilan data yang relevan dari kumpulan data yang lebih besar, dan akan digunakan sebagai sampel dalam langkah-langkah berikutnya.
2. Data *Integration*
   proses menggabungkan data dari berbagai sumber menjadi satu set data yang koheren dan terpadu.



Gambar 2. Diagram metode penelitian

3. Data *Transformation*
   Melakukan transformasi data untuk menyederhanakan dan meningkatkannya agar cocok dengan metode data mining yang akan digunakan.
4. Data *Mining*
   Melakukan pembangunan model menggunakan Algoritma Linear Regression pada tools *google colab* untuk prediksi timbunan sampah.
5. Data *Mining*
   informasi yang berguna dan berharga yang dapat diperoleh dari data. Knowledge ini dapat berupa pola, trend, atau hubungan antar data yang dapat digunakan untuk analisis permasalahan.

4. HASIL DAN PEMBAHASAN

Hasil penelitian menunjukkan bahwa linear regression dapat digunakan untuk memprediksi volume akumulasi sampah berdasarkan korelasi dari faktor tertentu. Pada penelitian ini menggunakan dataset jumlah akumulasi sampah berdasarkan faktor tertentu di wilayah Provinsi Jawabarat sebanyak 136 data dari tahun 2018 – 2022.

4.1. Data *Selction*

Pada tahap ini di lakukan data selection tahapan ini bertujuan untuk memilih data beberapa data yang akan digunakan dalam proses prediksi yaitu dataset, Sampah Yang ditangani, Timbunan Sampah, dan Produksi Sampah Berdasarkan Kabupaten/Kota di Jawa Barat. tahapan seleksi dilakukan menggunakan Microsoft Excel, pada proses penelitian menggunakan data tahun 2018-2022. Hasil dari seleksi tiga dataset diketahui pada masing-masing dataset terdapat 4 atribut yang dapat dijadikan atribut dalam penelitian diantaranya.

a. Atribut dataset sampah yang ditangani
   yaitu, Kode, Nama Kabupaten/Kota, Jumlah Sampah dan Tahun.

b. Atribut dataset timbunan sampah
yaitu, Kode, Nama Kabupaten/Kota, Jumlah Sampah dan Tahun.
c. Atribut dataset produksi sampah
yaitu, Kode, Nama Kabupaten/Kota, Jumlah Sampah dan Tahun.

4.2. Data *Integration*

Pada tahap ini di lakukan data *integration* tahapan ini bertujuan untuk menggabungkan beberapa data yang akan digunakan. Proses penggabungan data melalui tools Micrososft Excel, hasil penggabungan data dapat dilihat pada tabel sebagai berikut:

Tabel 4. Data *Integration*

Nama_Kabupaten_Kota	Kode	Jml_produksi	Jml_terkelola	Jml_timbunan	Tahun
KABUPATEN BOGOR	3201	1511,15	878169,85	3874,75	2018
KABUPATEN SUKABUMI	3202	419,01	209863,15	1074,39	2018
KABUPATEN CIANJUR	3203	981,41	364065,95	251,44	2018
KABUPATEN BANDUNG	3204	1895,94	1177662,73	4861,38	2018
KABUPATEN GARUT	3205	464,74	242110,63	1191,63	2018
KABUPATEN TASIKMALAYA	3206	464,52	298488,65	1191,08	2018
KABUPATEN CIAMIS	3207	260,91	96562,73	669	2018
KABUPATEN KUNINGAN	3208	251,7	124431,17	645,39	2018
KABUPATEN CIREBON	3209	465,75	327501,54	1194,23	2018
KABUPATEN MAJALENGKA	3210	254,63	137589,95	652,9	2018
…	…	…	…	…	…
KOTA CIMAHI	3277	193,36	173,8	193,36	2022
KOTA TASIKMALAYA	3278	295,87	194,29	295,87	2022
KOTA BANJAR	3279	60,29	44	60,29	2022

Hasil dari penggabungan tiga dataset diketahui pada masing-masing dataset Dari hasil penggabungan dataset terdapat 6 atribut yaitu, Nama Kabupaten/Kota, Kode, Jumlah Produksi, Jumlah Terkelola, Jumlah Timbunan, dan Tahun.

4.3. Data *Transformation*

Pada tahap transformasi data, data yang telah dikumpulkan diubah ke dalam format yang dapat dibaca oleh algoritma data mining. Hal ini dilakukan untuk memudahkan proses analisis data. Data yang telah diubah kemudian disesuaikan dengan masalah dan tujuan penelitian, yaitu prediksi akumulasi sampah di wilayah permukiman dan perumahan di Provinsi Jawa Barat untuk tahun 2023. Data ini terdiri dari 5 atribut yaitu, kode, jumlah produksi, jumlah terkelola, jumlah timbunan, dan tahun. Transformasi data dapat dilihat pada gambar 3 berikut.

	kode	tahun	jml_produksi	jml_terkelola	jml_timbunan
0	3201	2018	1511.15	878169.85	3874.75
1	3202	2018	419.01	209863.15	1074.39
2	3203	2018	981.41	364065.95	251.44
3	3204	2018	1895.94	1177662.73	4861.38
4	3205	2018	464.74	242110.63	1191.63
...	...	...	...	...	...
130	3275	2022	1500.77	1061.00	1500.77
131	3276	2022	1418.87	1194.20	1418.87
132	3277	2022	193.36	173.80	193.36
133	3278	2022	295.87	194.29	295.87
134	3279	2022	60.29	44.00	60.29

135 rows × 5 columns

Gambar 3. Data *Transformation*

4.4. Data *Mining*

Proses data mining yang diterapkan dalam penelitian ini adalah menggunakan regresi linear untuk memprediksi timbunan sampah pada permukiman dan perumahan di Provinsi Jawa Barat. Evaluasi hasil prediksi dilakukan dengan mengukur *Root Means Squared Error (RMSE)* dan Nilai *R-squared (R2).* Pada tahap ini, fokus utama adalah pada identifikasi korelasi pada masing-masing variabel untuk mengetahui faktor yang mempengaruhi timbunan sampah dan jumlah tmbunan sampah. Proses data mining menggunakan algoritma *Linear Regression* pada *Google Colab* untuk memodelkan prediksi timbunan sampah tahun 2023 berdasarkan Kabupaten/Kota di Jawa Barat. Langkah pertama import *library* yang di digunakan. Bisa dilihat pada gambar 4 sebagai berikut:

```
#import Library
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import linear_model
```

Gambar 4. *Import Library*

a. *Import pandas as pd*
Digunakan untuk mengimport modul pandas ke dalam skrip Python, sehingga dapat menggunakan fungsi dan kelasnya.
b. *Import numpy as np*
Untuk mengimpor *library NumPy* ke dalam program Python. Library NumPy adalah library yang memiliki fungsi untuk melakukan komputasi numerik.

c. *Import seaborn as sns*
Digunakan untuk mengimpor library Seaborn ke dalam program *Python. Library* Seaborn dibangun di atas library Matplotlib dan Pandas dan berfokus pada membuat visualisasi data statistik.

d. *Import matplotlib pyplot as plt*
Digunakan untuk mengimpor *library Matplotlib* ke dalam program *Python. Library Matplotlib* memiliki fungsi visualisasi data seperti *plot, histogram,* dan s*catter plot.*

e. *From sklearn import linear model*
Digunakan untuk mengimpor modul linear_model ke dalam program Python dari library scikit-learn.

Selanjutnya membuat data *frame* baru yaitu akumulasi1 yang berisi data jml_produksi dan jml_timbunan yang dapat dilihat pada gambar 5 sebagai berikut:

```
akumulasi1=akumulasi[['jml_produksi', 'jml_timbunan']]
akumulasi1
```

	jml_produksi	jml_timbunan
0	1511.15	3874.75
1	419.01	1074.39
2	981.41	251.44
3	1895.94	4861.38
4	464.74	1191.63
...	...	...
130	1500.77	1500.77
131	1418.87	1418.87

Gambar 5. Data *Frame*

Selanjutnya mengambil variabel *input* yaitu kolom ke 1 dan 2 pada *dataframe* akumulasi1 yaitu kolom jml_produksi dan jml_timbunan, perintah yang di *input* ke dalam program yang dapat dilihat pada gambar 6 sebagai berikut:

```
[19] x=akumulasi1.iloc[:,:-1].values
     y=akumulasi1.iloc[:,1].values
```

Gambar 6. Input variabel x dan y

Langkah berikutnya melakukan proses training pada data frame akumulasi1, pada proses ini dilakukan split pada data yang di bagi menjadi data testing, untuk data testing menggunakan 10% dan selebihnya digunakan untuk data training. Kemudian untuk mengetahui data yang digunakan untuk training dan testing penulis memeriksa jumlah data *training* dan data *testing* yang dapat dilihat pada gambar 7 sebagai berikut:

```
print(x_train.shape)
print(y_train.shape)
print(x_test.shape)
print(y_test.shape)
```

```
(121, 1)
(121,)
(14, 1)
(14,)
```

Gambar 7. Memeriksa data training dan testing

Langkah berikutnya membuat model prediksi menggunakan *Linear Regression*, dalam hal penulis membuat objek *linear regression* dengan nama reg. Dapat dilihat pada gambar 8 sebagai berikut:

```
from sklearn.linear_model import LinearRegression
reg=LinearRegression()
reg.fit(x_train, y_train)
```

```
▾ LinearRegression
LinearRegression()
```

Gambar 8. Model *Linear Regression*

Selanjutnya melakukan proses prediksi menggunakan model *linear regression* dengan perintah yang di *input* kedalam program. Bisa dilihat pada gambar 9 sebagai berikut:

```
#prediksi menggunakan linear regression
y_pred = reg.predict(x_test)
print(y_pred)
```

```
[2501.3619992  1848.44520372 2953.19373783 2474.12687908 3018.39088905
 2126.19410973 2125.48172228 3565.01125749 4758.36983028 1724.11989435
 1910.58730876 3043.05045454 2259.71195712 4311.23710884]
```

Gambar 9. Hasil prediksi

Dari hasil pemodelan menggunakan regresi linear pada gambar 9 terdapat hasil prediksi namun bentuk pada data masih dalam format acak, untuk memudahkan agar data dapat dibaca dengan baik maka proses selanjutnya merubah bentuk data hasil prediksi kedalam data frame. Bisa dilihat pada gambar 10 sebagai berikut:

```
il prediksi kedalam data frame
= pd.DataFrame({'Nilai Aktual':y_test, 'N
```

Nilai Prediksi
2501.361999
1848.445204
2953.193738
2474.126879
3018.390889
2126.194110

Gambar 10. Data hasil prediksi

Langkah selanjutnya menghitung data prediksi pada data *frame* dan dijumlahkan untuk mengetahui data keseluruhan dari data prediksi untuk tahun 2023 dengan perintah yang di input kedalam program. Bisa dilihat pada gambar 11. sebagai berikut:

```
akumulasidf =pd.DataFrame({
    "Data Prediksi" :[2575,1921,3029,2548,3094,2199,2198,3642,
                      4839,1796,1983,3119,2333,4391,2471,2271,
                      5564,2843,3509,3039,1905,1738,2854,1886,
                      1564,1808,2406]
})
```

Gambar 11. Menjumlahkan data hasil prediksi

Setelah proses penjumlahan selanjutnya melihat hasil dari penjumlahan pada data hasil prediksi tahun 2023 yaitu 38615. Bisa dilihat pada gambar 12 sebagai berikut:

```
#hasil dari penjumlahan pada data hasil prediksi tahun 2023
jumlah = akumulasidf["Data Prediksi"].sum()
print('Prediksi Jumlah Timbunan Sampah Tahun 2023 Yaitu', jumlah)
```

```
Prediksi Jumlah Timbunan Sampah Tahun 2023 Yaitu 38615
```

Gambar 12. Hasil prediksi tahun 2023

Selanjutnya, cari nilai *coefficient* dan nilai *intercept.* Kemiringan garis regresi yang positif menunjukkan hubungan positif antara variabel independen dan variabel dependen, sedangkan nilai intercept menunjukkan titik potong garis regresi dengan sumbu y. Titik potong garis regresi adalah titik di mana variabel independen bernilai nol. Hasil bisa dilihat pada gambar 13 sebagai berikut:

```
#Mencari Niali coefficient dan intercept
print('coefficient:',reg.coef_)
print('intercept:' ,reg.intercept_)
```

```
coefficient: [1.36997586]
intercept: 1488.127852579766
```

Gambar 13. Nilai Coefficient dan Intercept

Mencari nilai akurasi yang dihasilkan dari proses pemodelan untuk mengetahui tingkat akurasi prediksi pada model yang dibuat. Penulis mencari akurasi RMSE dan nilai R-squared (R2). Hasilnya adalah 1,586,55 untuk akurasi RMSE dan 0,38 untuk nilai R2-Score. Bisa dilihat pada gambar 14 sebagai berikut:

```
# Root Means Squared Error (RMSE) dan Nilai R-squared (R2).
import math
y_test_ =reg.predict(x_test)

from sklearn.metrics import mean_absolute_error, mean_squared_error,r2_score
print("Roots Mean Squared Error (RMSE): %.2f" % math.sqrt
(mean_squared_error(y_test,y_pred)))
print("R2-Score: %.2f" % r2_score(y_test,y_pred))
```

```
Roots Mean Squared Error (RMSE): 1702.37
R2-Score: 0.25
```

Gambar 14. Akurasi *Root Means Squared Error (RMSE)* dan nilai *R-squared (R2)*

Setelah diketahui akurasi pada model yang dibuat, langkah selanjutnya mencari nilai korelasi antar variabel, hasil korelasi di *visualisasi* menggunakan *heatmap*. Bisa dilihat pada gambar 15 sebagai berikut:

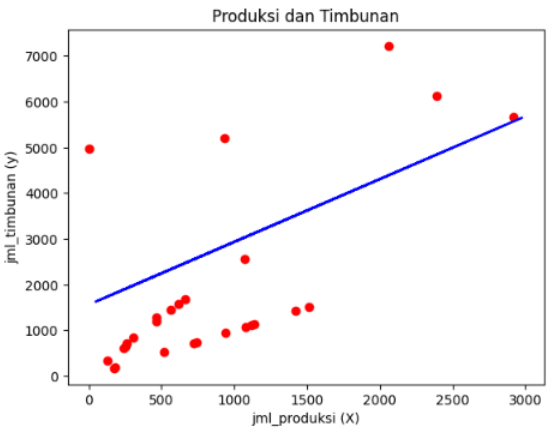
```
# Mencari Nilai korelasi
sns.heatmap(akumulasi.corr(),
            cmap = "RdBu_r",
            square=True,
            annot=True,
            cbar=True)
plt.title("Korelasi Antar Variabel")
plt.show()
```

Gambar 15. Mencari korelasi antar variabel

Hasil evaluasi pada model prediksi di ukur dengan nilai *RMSE (Root Squared Mean Error)* untuk hasil *RMSE* mendapat nilai 1,586,55 sedangkan untuk nilai *R2-Score* mendapat nilai 0,38. bisa dikatakan sangat baik jika hasil *RMSE* dan *R2-Score* mendekati 0 atau prediksi yang mendekati nilai aktual. Selain itu untuk nilai Koefisien sebesar 1.36997586 menunjukkan bahwa variabel bebas memberikan dampak positif pada variabel terikat. Ini berarti jika nilai variabel bebas naik, nilai variabel terikat juga akan naik. Sementara nilai intercept sebesar 1488.12 mengindikasikan nilai variabel terikat ketika variabel bebas bernilai nol. Untuk hasil prediksi timbunan sampah menunjukan hasil 38615 pada tahun 2023.

4.5. *Knowledge*

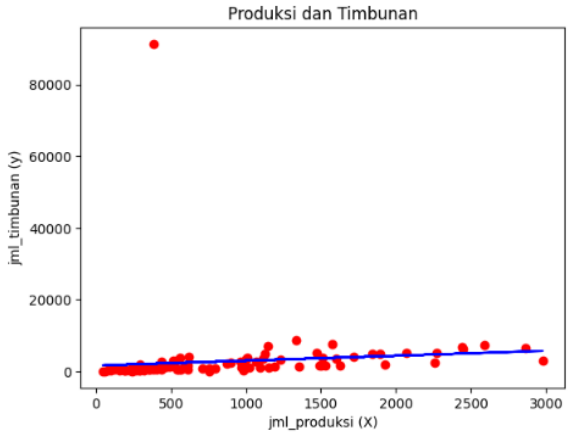
Pada langkah yang telah dilakukan, hasil prediksi timbunan sampah di tahun 2023 sebesar 38615 telah diperoleh menggunakan algoritma regresi linear proses KDD dalam pembuatan model, tingkat kualitas data pelatihan dan pengujian juga divisualisasikan. Dapat dilihat pada gambar 16 dan 17 sebagai berikut:



Gambar 16. Visualisasi akurasi data testing

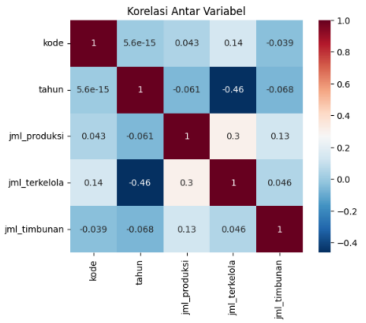
Pada gambar 16 diatas, pola garis lurus menunjukkan hubungan positif nilai y (jumlah timbunan) cenderung meningkat seiring dengan nilai x (jumlah produksi). Secara keseluruhan, grafik scatterplot ini menunjukkan bahwa data train masih dapat digunakan untuk meramalkan jumlah timbunan

berdasarkan jumlah produksi yang telah dicatat sebelumnya.



Gambar 17. Visualisasi akurasi data training

Pada gambar 17 di atas, pola garis lurus menunjukkan hubungan positif antara nilai y, yang merupakan jumlah timbunan, dan nilai x, yang merupakan jumlah produksi. Secara keseluruhan, grafik scatterplot ini menunjukkan bahwa data test dapat digunakan namun masih bisa untuk diperbaiki untuk meramalkan jumlah timbunan dengan menggunakan jumlah produksi yang telah dicatat sebelumnya. Berdasarkan hasil identifikasi *faktor* terdapat variabel yang paling berpengaruh dalam timbunan berdasarkan hasil korelasi, variabel yang paling berpengaruh yaitu, *jml_produksi dan jml_terkelola*, hasil bisa dilhat pada gambar *plot heatmap* 18 sebagai berikut:



Gambar 18. *Plot Heatmap* korelasi

5. KESIMPULAN

Berdasarkan hasi prediksi timbunan sampah di wilayah Provinsi Jawa Barat menggunakan algoritma *Regresi Linear*, diperoleh hasil prediksi untuk tahun 2023. Prediksi ini didasarkan pada data akumulasi sampah berdasarkan permukiman dan perumahan, analisis data menggunakan *Knowledge Discovery in Database* dan algoritma *Linear Regression* memproyeksikan jumlah timbunan sampah tahun 2023 sebesar memprediksi jumlah timbunan sampah pada tahun 2023 dengan jumlah timbunan 38615 ton per hari, setelah dilakukan analisis dari hasil prediksi dimana pada data 5 tahun terakhir 2018-2022 jumlah timbunan sampah sebanyak 338784 ton per hari dan mengalami penurunan sebesar 300169 ton per hari di tahun 2023, variabel utama yang memengaruhi akumulasi sampah adalah jumlah produksi sampah. Evaluasi model melalui RMSE menghasilkan nilai 1,586.55, menandakan tingkat akurasi yang relatif baik dengan nilai 1,586.55, memperlihatkan potensi peningkatan akurasi yang dapat dilakukan. Meski demikian, model perlu diperbaiki untuk menjelaskan variasi data lebih baik karena nilai R2-Score-nya masih rendah, mencapai 0.38. Kenaikan nilai R2-Score akan mengindikasikan kemampuan model dalam menjelaskan variasi data secara lebih optimal.

DAFTAR PUSTAKA

[1] P. Pasande and E. Tari, “Daur Ulang Sampah di Desa Paisbuloli Sulawesi Tenggara,” *Din. J. Pengabdi. Kpd. Masy.*, vol. 5, no. 1, pp. 147–153, 2020, doi: 10.31849/dinamisia.v5i1.4380.

[2] Kusuma Wardany, Reni Permata Sari, and Erni Mariana, “Sosialisasi Pendirian ‘Bank Sampah’ Bagi Peningkatan Pendapatan Dan Pemberdayaan Perempuan Di Margasari,” *Din. J. Pengabdi. Kpd. Masy.*, vol. 4, no. 2, pp. 364–372, 2020, doi: 10.31849/dinamisia.v4i2.4348.

[3] H. W. Herwanto, T. Widiyaningtyas, and P. Indriana, “Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi,” *Natl. J. Electr. Eng. Inf. Technol.*, vol. 8, no. 4, p. 364, 2019, [Online]. Available: https://journal.ugm.ac.id/v3/JNTETI/article/view/2563

[4] R. Cristover, H. Toba, and B. R. Suteja, “Segmentation and Formation of Customer Regression Model Based on Recency, Frequency and Monetary Analysis,” *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 2, pp. 474–484, 2022, doi: 10.28932/jutisi.v8i2.5075.

[5] E. D. Anggara, A. Widjaja, and B. R. Suteja, “Prediksi Kinerja Pegawai sebagai Rekomendasi Kenaikan Golongan dengan Metode Decision Tree dan Regresi Logistik,” *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 1, pp. 218–234, 2022, doi: 10.28932/jutisi.v8i1.4479.

[6] B. Pradito and D. S. Purnia, “Komparasi Algoritma Linear Regression dan Neural Network Untuk Memprediksi Nilai Kurs Mata Uang,” *EVOLUSI J. Sains dan Manaj.*, vol. 10, no. 2, pp. 64–71, 2022, doi: 10.31294/evolusi.v10i2.13284.

[7] H. K. Prakosa and N. Rokhman, “Anomaly Detection in Hospital Claims Using K-Means and Linear Regression,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 15, no. 4, pp. 391–402, 2021, [Online]. Available: https://journal.ugm.ac.id/ijccs/article/view/68160

[8] F. F. A. Fetrus Jari Harianto, “Linear Regression Algorithm Analysis to Predict the Effect of Inflation on the Indonesian Economy Fetrus,”

Indones. J. Comput. Sci., vol. 12, no. 1, pp. 1673–1681, 2023, [Online]. Available: http://3.8.6.95/ijcs/index.php/ijcs/article/view/3224

[9] Gustiendina, Mh. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, [Online]. Available: https://teknosi.fti.unand.ac.id/index.php/teknosi/article/view/773

[10] I. Priyadi, J. Santony, and J. Na, "Data Mining Predictive Modeling for Prediction of Gold Prices Based on Dollar Exchange Rates , Bi Rates and World Crude Oil Prices," vol. 2, no. 2, pp. 93–100, 2019, [Online]. Available: https://ejournal.uin-suska.ac.id/index.php/IJAIDM/article/view/6864

[11] R. Kurniawan, A. Halim, and H. Melisa, "Prediksi Hasil Panen Pertanian Salak di Daerah Tapanuli Selatan Menggunakan Algoritma SVM (Support Vector Machine)," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 2, pp. 903–912, 2023, doi: 10.30865/klik.v4i2.1246.