



## Penerapan *K-Nearest Neighbor Regression* untuk Memprediksi Penjualan Tas Branded di Toko Shan Bogor

Sejahtera Abadi Hia<sup>1</sup>

Rahmawati<sup>2</sup>

<sup>1,2</sup> Universitas Pamulang

Sejahterahia24@gmail.com.<sup>1</sup> dosen02394@unpam.ac.id<sup>2</sup>

### Kata kunci:

K-Nearest Neighbor  
Regresion, Knowledge  
Discovery in Database,  
Prediksi, RMSE.

### Abstrak

Toko Shan Bogor merupakan pusat perbelanjaan yang menjual berbagai jenis merek tas seperti merk Gucci, Zara, Longcamp, dan Louis. Sistem pengelolaan data di Toko Shan bogor masih menggunakan pencacatan secara konvensional dimana data hasil penjualan masih diolah kedalam bentuk MS.excel. dilihat. Untuk mempermudah pihak pengelola Toko maka dibutuhkan prediksi penjualan dimasa yang akan datang dengan menggunakan teknik klasifikasi data mining algoritma *K-Nearest Neighbor Regresion*. Hasil dari penelitian prediksi penjualan tas terlaris adalah nilai  $K = 2$  menghasilkan nilai RMSE 0.43494,  $k = 3$  menghasilkan nilai RMSE 0.46214 untuk tas Gucci,  $k = 13$  menghasilkan nilai RMSE 0.45539 untuk tas Zara,  $k = 6$  menghasilkan nilai RMSE 0.48950 untuk tas Louis. Seperti yang tertera dalam pedoman RMSE dapat disimpulkan bahwa seluruh model yang diuji memiliki tingkan kesalahan yang sedang yaitu 0.30-0,559.

### Pendahuluan

Perkembangan dunia industri saat ini dapat dikatakan sangat pesat. Berbagai produk baru bermunculan dan menjamur di masyarakat. Salah satu industri yang tidak ada habisnya untuk melakukan inovasi adalah industri tas. Produksi tas Indonesia menduduki peringkat ke-4 di dunia. Indonesia memproduksi total 1,4 miliar pasang tas pada tahun 2018 lalu. Direktur jenderal industri kecil, menengah, dan aneka (ikma) dan kementerian perindustrian mengatakan bahwa angka tersebut memberi kontribusi sebesar 4,6% dari total produksi tas di dunia.

Toko Shan Bogor merupakan pusat perbelanjaan yang menjual berbagai jenis merk tas yang branding seperti merk Gucci, Zara, Longcamp, dan Louis, Yang memiliki bermacam-macam model yang trendy dan modis. Selama ini Shan senantiasa konsisten dalam meningkatkan dan mengembangkan usahanya agar dapat memenuhi kebutuhan konsumen baik dari segi kualitas tas, harga maupun pemasarannya. Dikarenakan Persaingan yang terus berkembang harus berusaha menempatkan produk tas dibenak konsumen. Atribut tas dapat membedakan dan mengetahui kelebihan tas dengan tas yang lainnya. Kepuasan konsumen terhadap suatu tas merupakan salah satu hal yang paling penting untuk meraih pasar.

Sistem pengelolaan data penjualan di Toko Shan Bogor masih menggunakan pencatatan secara konvensional dimana data hasil penjualan direkapitulasi kedalam Ms.Excel, sehingga data yang dimasukkan kedalam Ms.Excel semakin lama semakin menumpuk dan tidak teratur, yang menyebabkan data sulit dipahami dan kurang informasi yang bisa dikembangkan untuk data penjualan tas setiap tahunnya. Untuk itu perlu dilakukan prediksi hasil penjualan ditahun yang akan datang. Dengan ini perlu dilakukan tindakan untuk mengolah data hasil penjualan dengan menggunakan perhitungan teknik data mining. Salah satu algoritma yang terdapat pada teknik data mining yaitu algoritma K-Nearest Neighbor.

Dalam penelitian sebelumnya yang dilakukan oleh Yuwantara et al., (2019) yang menggunakan algoritma regresi K-Nearest Neighbor ( KNN) tentang memprediksi harga beras premium diperoleh RMSE ( Root Mean Square Error ) dengan nilai 0.125 dan parameter k=2 yang sudah dinormalisasi dan sudah membandingkan dengan 3 metode -yang dinormalisasi lainnya seperti SVR RBF dengan RMSE 579,391, Decision Tree Regressor dengan RMSE 1850,04 dan yang paling besar MLP dengan RMSE 3954,792.

## **Metode**

Peneliti melakukan penelitian di Toko Shan Bogor di Jl. megasentul. alamanda, cijujung, kec.sukaraja, Kota bogor, Jawa barat 16710. Penelitian ini dilakukan kurang lebih 1 bulan lamanya yaitu terhitung mulai bulan Maret 2021.

Metode yang penulis gunakan yaitu dengan melakukan pengumpulan data. Adapun cara penulis melakukan pengumpulan data adalah:

1. Metode Pengumpulan Data
  - a. Observasi  
Dilakukan dengan cara mengamati secara langsung atas penelitian yang dilakukan.
  - b. Wawancara  
Peneliti melakukan wawancara secara langsung pada pihak yang dijadikan sebagai objek penelitian.
  - c. Studi pustaka  
Peneliti mengumpulkan data dengan tinjauan pustaka dan pengumpulan buku-buku, bahan-bahan tertulis serta referensi yang relevan.
2. Metode Analisis  
Metode K-Nearest Neighbor melakukan klasifikasi perhitungan suatu data dengan data lain. Data yang akan digunakan adalah data penjualan Tas branded di Toko Shan Bogor.
3. Metode Perancangan  
Metode perancangan dilakukan dalam sebuah proses sesuai dengan tahapan data mining meliputi pengumpulan data, analisis, dan sintesis konsep. Tahapan-tahapan pada metode rancangan ini mulai dari pencarian ide kemudian mencari permasalahan dan tujuan dari rancangan itu sesuai dengan Knowledge Discovery in Database (KDD) data mining.
4. Metode Pengujian  
Metode pengujian dilakukan dengan menggunakan Cros Validation yang dibagi secara acak dengan perbandingan yang sama. Kemudian error akan dihitung untuk memperoleh hasil error secara keseluruhan.

## Hasil dan Pembahasan

### Evaluasi

Pada tahap evaluasi ini dilakukan pengujian model menggunakan cross validation untuk setiap nilai k dari k-nearest neighbor dalam range 1-15.

Implementasi pada program sebagai berikut :

```
k_range = range(1, 15)
rmse_err = []
for k in k_range:
    knn = KNeighborsRegressor(n_neighbors=k, metric='euclidean')
    scores = cross_val_score(knn, X, y, scoring='neg_mean_absolute_error', cv=cv)
    RMSE = sqrt(mean(abs(scores)))
    rmse_err.append(RMSE)
    print('RMSE value for k= ', k , 'is:', RMSE)
plt.plot(k_range, rmse_err, '-ok')
plt.xlabel('value of K for KNN')
plt.ylabel('RMSE')
plt.grid(True)
plt.show()
```

Gambar 1. Script Evaluasi.

Penjelasan script:

Fungsi `range(1, 15)` berfungsi menciptakan ukuran angka dari 1-14 yang dideklarasikan kedalam variabel `k_range` dimana selanjutnya akan digunakan sebagai looping.

Variabel `rmse_err = []` akan digunakan untuk menyimpan nilai-nilai yang dihasilkan dari `rmse_err.append(RMSE)`.

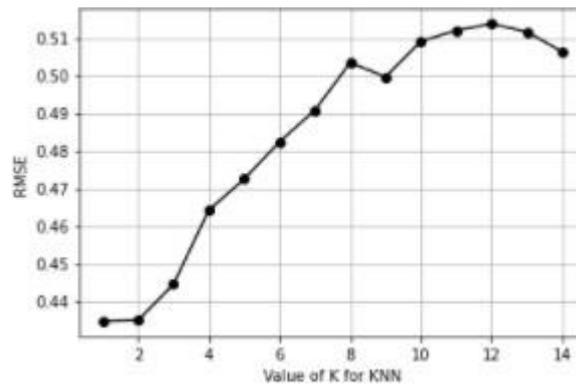
`For k in k_range` berarti melakukan perulangan variabel `k` sebanyak ukuran angka yang ada pada variabel `k_range` serta menjalankan perintah di dalamnya.

`KNeighborsRegressor(n_neighbors=k, metric='euclidean')` berarti Fungsi K-NN Regresi dengan parameter `n_neighbors` sebagai nilai ketetangaan adalah variabel `k` yang terus bertambah selama proses looping dan `metric` yang digunakan adalah `euclidean` kemudian fungsi ini dideklarasikan kedalam variabel `knn`.

`Cross_val_Score(knn, X, y, scoring='neg_mean_absolute_error', cv=cv)` berfungsi untuk mengevaluasi dengan parameter yang diisi dengan model, data training, data testing, model scoring, serta crossvalidasi yang menggunakan variabel `cv` dimana

berisi fungsi K-Fold Cross Validation. Fungsi tersebut disimpan dalam variabel `score`

Sqrt(mean(absolute(scores))) berarti mengkuadratkan nilai rata-rata dari hasil score yang telah di absolutkan. Fungsi tersebut disimpan dalam variabel RMSE.



rmse\_err.append(RMSE) berarti menambahkan nilai array dengan nilai dari variabel RMSE yang kemudian masuk dalam list array pada rmse\_err.

plt.plot(k\_range, rmse\_err) berfungsi untuk menampilkan grafik dengan sumbu x adalah variabel k\_range dan sumbu y adalah variabel rmse\_err.

Berikut Hasil Implementasi :

Gambar 2. Grafik Nilai MRSE

Grafik tersebut menunjukkan perbedaan nilai RMSE berdasarkan nilai k pada k-nearst neighbor. Nilai k tersebut sangat mempengaruhi hasil dari keakuratan model prediksi.

```
RMSE value for k= 1 is: 0.43469649062425036
RMSE value for k= 2 is: 0.4349453864308589
RMSE value for k= 3 is: 0.4444624815964758
RMSE value for k= 4 is: 0.4641941301612834
RMSE value for k= 5 is: 0.4724441376538217
RMSE value for k= 6 is: 0.48237328147813363
RMSE value for k= 7 is: 0.49073040106621907
RMSE value for k= 8 is: 0.5033975475795358
RMSE value for k= 9 is: 0.4996150514297979
RMSE value for k= 10 is: 0.5090735154728089
RMSE value for k= 11 is: 0.511958528426676
RMSE value for k= 12 is: 0.5137187230449426
RMSE value for k= 13 is: 0.5115710719977838
RMSE value for k= 14 is: 0.5062686813704254
```

Gambar 3. Nilai RMSE per-k

Nilai tersebut menunjukkan detail nilai pada setiap nilai k dalam range 1-15. Proses pengujian dilakukan setelah model dihasilkan dari proses training sebelumnya. Evaluasi ini dilakukan untuk mengetahui akurasi dari implementasi metode k-NN.

### Performance Measure

Setelah didapati nilai error yang terkecil maka model yang dibentuk akan menggunakan nilai k dengan hasil error terkecil tersebut.

Implementasi pada program sebagai berikut :

```
#akurasi RMSE
model = KNeighborsRegressor(n_neighbors=3, metric='euclidean')
scores = cross_val_score(model, X, y, scoring='neg_mean_absolute_error', cv=cv)
sqrt(mean(abs(scores)))

0.4621474298463427
```

Gambar 4. Performance Measure

Proses tersebut dilakukan untuk semua data training yang ada, setelah itu maka didapat nilai RMSE terbaik untuk setiap produk.

Berikut merupakan hasil nilai akurasi terbaik untuk setiap produk Tas di Toko Shan Bogor.

Tabel 1. Hasil Nilai Akurasi Terbaik di Toko Shan Bogor.

Produk	Nilai K Terbaik	RMSE Terbaik	Prediksi Penjualan
Gucci	2	0.43494	Maret
Zara	3	0.46214	Januari
Longcamp	13	0.45539	Juli
Louis	14	0.48781	November

#### Model Terbaik

Setelah didapati model dengan error terkecil maka model terbaik telah didapatkan, selanjutnya adalah menguji model tersebut dengan data testing yang telah disiapkan.

Implementasi dan hasil pada program sebagai berikut :

```
model = KNeighborsRegressor(n_neighbors=2, metric='euclidean')
model.fit(X, y)

KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='euclidean',
                    metric_params=None, n_jobs=None, n_neighbors=2, p=2,
                    weights='uniform')
```

Gambar 5 Model Terbaik

#### Penjelasan script:

model.fit(X, y) berfungsi untuk menyesuaikan transformator dengan data variabel X dan y.

#### Prediksi

Berikut ini merupakan script untuk melakukan prediksi berdasarkan model yang telah dibuat sebelumnya.

```
y_pred = model.predict(x_tes)
y_pred
dft['prediksi'] = y_pred
dft
```

Gambar 6. Script Prediksi.

Penjelasan script:

Model.predict(x\_tes) berfungsi untuk mengembalikan hasil prediksi untuk data dalam variabel x\_tes. Fungsi tersebut disimpan dalam variabel y\_pred.

dft['prediksi'] = y\_pred berfungsi untuk menambahkan kolom dengan nama prediksi yang berisi data dalam variabel y\_pred.

Berikut merupakan hasil implemetasi:

	bln-1	bln-2	bln-3	bln-4	bln-5	bln-6	bln-7	bln-8	bln-9	bln-10	bln-11	bln-12	prediksi
0	42	49	61	60	55	60	62	62	57	50	87	89	0.519481
1	49	61	60	55	60	62	62	57	50	87	89	65	0.616883
2	61	60	55	60	62	62	57	50	87	89	65	87	0.941558
3	60	55	60	62	62	57	50	87	89	65	87	98	0.753247
4	55	60	62	62	57	50	87	89	65	87	98	93	0.344156
5	60	62	62	57	50	87	89	65	87	98	93	30	0.428571
6	62	62	57	50	87	89	65	87	98	93	30	21	0.532468
7	62	57	50	87	89	65	87	98	93	30	21	37	0.116883
8	57	50	87	89	65	87	98	93	30	21	37	23	0.058442
9	50	87	89	65	87	98	93	30	21	37	23	28	0.344156
10	87	89	65	87	98	93	30	21	37	23	28	33	0.142857
11	89	65	87	98	93	30	21	37	23	28	33	31	0.246753

Gambar 7. Prediksi Penjualan Zara.

Nilai prediksi merupakan ukuran skala dari normalisasi dengan nilai tertinginya adalah nilai penjualan yang paling laris.

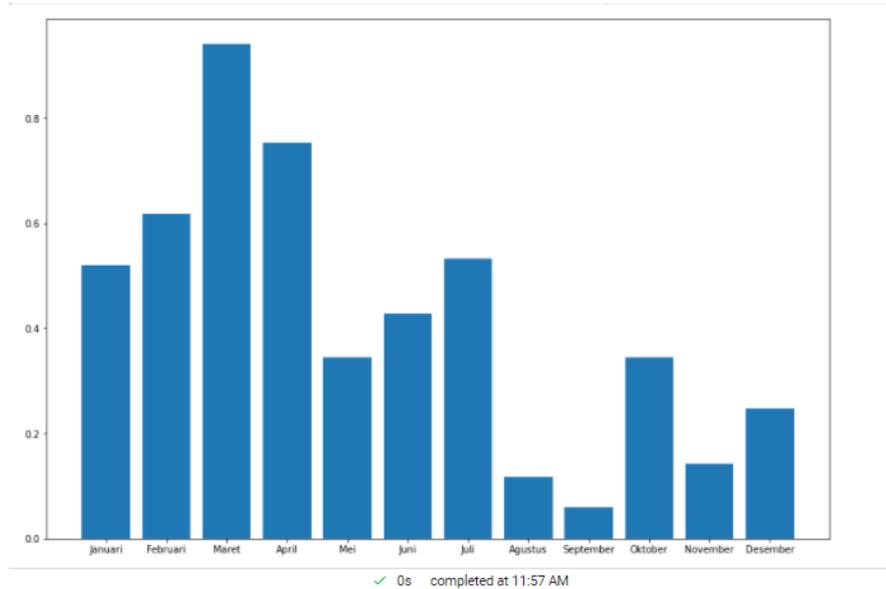
Visualisasi

Untuk memudahkan dalam memahami data tersebut penulis melakukan visualisasi terhadap data prediksi.

```
fig = plt.figure()
ax = fig.add_axes([0,0,1.9,1.9])
Prediksi = np.array(dft['prediksi'])
Bulan = ['Januari', 'Februari', 'Maret', 'April', 'Mei', 'Juni', 'Juli', 'Agustus', 'September', 'Oktober', 'November', 'Desember']
ax.bar(Bulan, Prediksi)
plt.show()
```

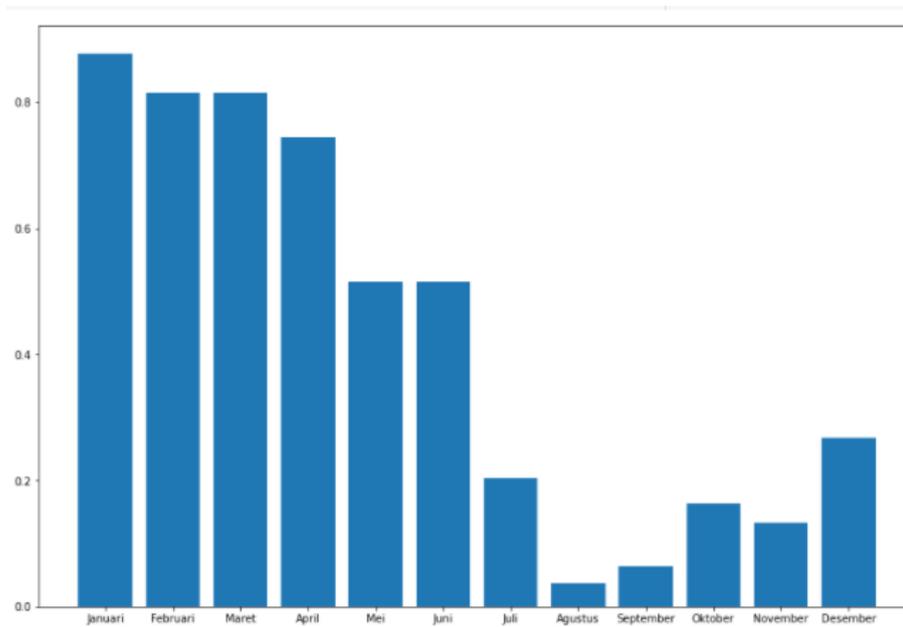
Gambar 8. Script Visualisasi.

Berikut merupakan hasil visualisasi prediksi setiap penjualan produk.



Gambar 9. Visualisasi Hasil Prediksi Penjualan Gucci.

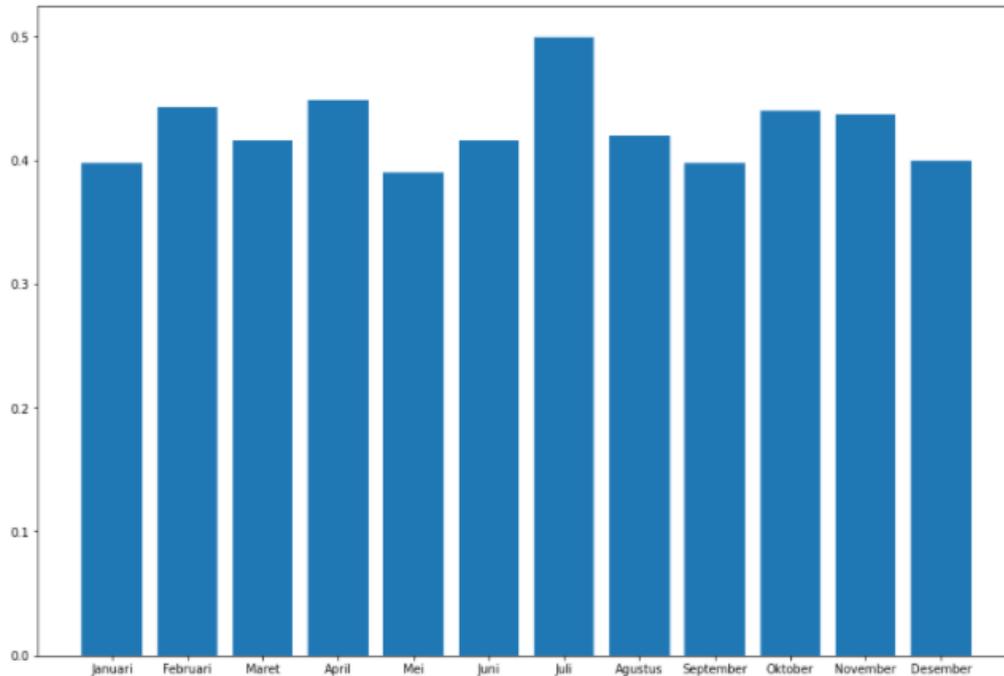
Pada gambar diatas merupakan hasil prediksi penjualan produk Gucci dimana prediksi



penjualan terbanyak terdapat pada bulan Maret.

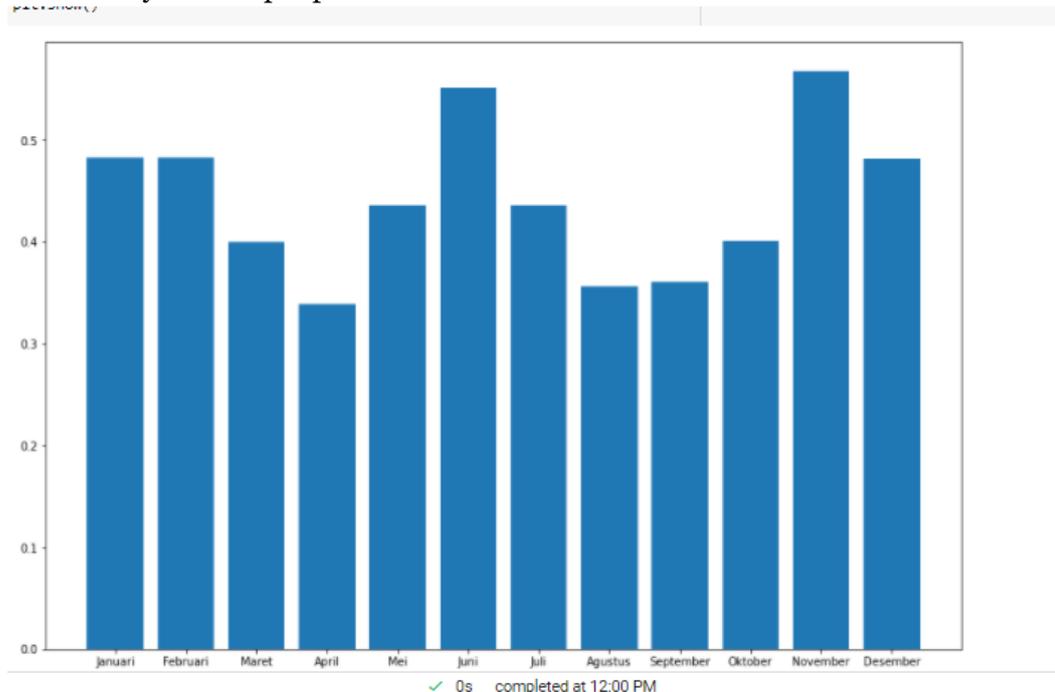
Gambar 10. Visualisasi Hasil Prediksi Penjualan Zara.

Pada gambar diatas merupakan hasil prediksi penjualan produk Zara dimana prediksi penjualan terbanyak terdapat pada bulan Januari.



Gambar 11. Hasil Prediksi Penjualan Longcamp.

Pada gambar diatas merupakan hasil prediksi penjualan produk Longcamp dimana prediksi penjualan terbanyak terdapat pada bulan Juli.



Gambar 12. Hasil Prediksi Penjualan Louis.

Pada gambar diatas merupakan hasil prediksi penjualan produk Gucci dimana prediksi penjualan terbanyak terdapat pada bulan November.

### **Kesimpulan**

Dari hasil penelitian dan pengujian model prediksi penjualan menggunakan metode K-Nearest Neighbor Regression, pada akhir penelitian ini peneliti dapat memberikan kesimpulan sebagai berikut :

1. Metode yang diusulkan yaitu K-Nearest Neighbor Regression untuk menyelesaikan kasus prediksi penjualan produk di Toko Shan Bogor.
2. Diperoleh hasil prediksi penjualan produk terlaris untuk Gucci pada bulan Maret, Zara pada bulan Januari, Longcamp pada bulan Juli, dan Louis pada November.
3. Diperoleh kesimpulan nilai k yang paling optimal dari rentang 1 sampai 15 adalah nilai  $k = 2$  menghasilkan nilai RMSE 0.43494 untuk produk sepatu Gucci,  $k = 3$  menghasilkan nilai RMSE 0.46214 untuk produk Zara,  $k = 13$  menghasilkan nilai RMSE 0.45539 untuk produk Longcamp,  $k = 14$  menghasilkan nilai RMSE 0.48789 untuk produk Louis, Dengan demikian berdasarkan pedoman interpretasi RMSE dapat disimpulkan bahwa tingkat kesalahan seluruh model yang penulis uji memiliki kesalahan sedang dikarenakan nilai RMSE yang didapat berada dalam range 0,30 – 0,599.

### **Daftar Pustaka**

- Argina, Andi Maulida. 2020. Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabete. Vol. 1, No.2
- Bode, A. (2017). K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika. Ilkom Jurnal Ilmiah, 8.
- Baharuddin, M.M., Hasanuddin T., & Azis H.2019. Analisa Perfoma Metode K- Nearest Neighbor Untuk Identifikasi Jenis Kaca. Vol. 11, No.3
- <https://repository.bsi.ac.id/index.php/unduh/item/20880/file-9-daftar-simbol.pdf> (2017, Oktober).
- Hendini a. 2016. Pemodelan Uml Sistem Informasi Monitoring Penjualan Dan Stok Barang (Studi Kasus : Distro Zhezha Pontianak). Vol. 4, No.2
- Ismael. (2017). Rancang Bangun Sistem Informasi Penyaluran Semen Padang Untuk Daerah Bengkulu Selatan Di CV.Mutia Bersaudara. Jurnal Edik Informatika, 10.
- Khoiri. (2020). Pengertian Dan Cara Menghitung Room Mean Square Error (RMSE). 5.
- Lestari, I. D. (2016). Klasifikasi Online Google. Jurnal Iqra', 12.
- Mukhlisin Y., Imrona M., & Murdiansyah D T. 2020. Prediksi Harga Beras Premium dengan Metode Algoritma K-Nearest Neighbor. Vol. 7, No.1

- Maulana, A., & Fajrin, A. A. (2018). Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Penjualan Spare Part Motor. *Jurnal Ilmiah*.
- Nasution D.A., Khotimah H H., & Chamidah N. 2019. Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-NN. Vol. 4, No.1
- Pratama T C. 2018. Penerapan Metode K-Nearest Neighbor Dalam Menentukan Kelayakan Calon Nasabah Yang Layak Untuk Kredit Mobil (Studi Kasus PT. Astra International, Tbk-Toyota) *Jurnal Riset Komputer* Vol. 5 No. 4
- Suyanto. (2017). Data Mining Untuk Klasifikasi Dan Klaterisasi Data . *Informatika*.
- Sholado, R. M. (2021). LKP : Deteksi Kanker Kulit Menggunakan Deep Learning  
[Universitas Dinamika]. <https://repository.dinamika.ac.id/eprint/5434/>
- Santoso & Nurmalina, Radna. 2017. Perencanaan dan Pengembangan Aplikasi Absensi Mahasiswa Menggunakan Smart Card Guna Pengembangan Kampus Cerdas (Studi Kasus Politeknik Negeri Tanah Laut) *Jurnal Integrasi*, Vol. 9, No. 1.
- STIMK Royal, Ksieran. 2016. Penerapan Data Mining Untuk Prediksi Penjualan Walpaper Menggunakan Algoritma C4.5. Vol. 2, No. 2
- Sianturi C J., Ardini E., & Sembiring N S. 2020. Sales Forecasting Information System Using Theleast Square Method In Windi Mebel. Vol. 1, No. 2
- Suyanto . 2017. Data Mining Untuk Klasifikasi dan Klaterisasi Data. *Bandung Informatika*.
- Tanuwijaya, J., & Seng, H. (2019). LQ45 Stock Index Prediction Using K-Nearest Neighbor Regression. *International Journal Of Recet Technology And Engineering*, 5.