



## K- NEAREST NEIGHBOR REGRESSION UNTUK PREDIKSI PENJUALAN DONUT

**Ratna Salkiawati<sup>1</sup>, Nurfiyah<sup>2</sup>, Hendarman Lubis<sup>3</sup>,  
Ilham Nur El Ghozy<sup>4</sup>**

Program Studi Informatika<sup>1</sup>, Program Studi Informatika<sup>2</sup>,  
Program Studi Informatika<sup>3</sup>, Program Studi Informatika<sup>4</sup>  
Fakultas Ilmu Komputer<sup>1</sup>, Fakultas Ilmu Komputer<sup>2</sup>,  
Fakultas Ilmu Komputer<sup>34</sup>  
Universitas Bhayangkara Jakarta Raya<sup>1</sup>  
Universitas Bhayangkara Jakarta Raya<sup>2</sup>  
Universitas Bhayangkara Jakarta Raya<sup>3</sup>

[ratna\\_tind@dsn.ubharajaya.ac.id](mailto:ratna_tind@dsn.ubharajaya.ac.id)<sup>1</sup>, [nurfiyah@dsn.ubharajaya.ac.id](mailto:nurfiyah@dsn.ubharajaya.ac.id)<sup>2</sup>,  
[hendarman.lubis@dsn.ubharajaya.ac.id](mailto:hendarman.lubis@dsn.ubharajaya.ac.id)<sup>3</sup>,  
[ilham.nur17@mhs.ubharajaya.ac.id](mailto:ilham.nur17@mhs.ubharajaya.ac.id)<sup>4</sup>

**Received:** May 23, 2025. **Revised:** June 18, 2025. **Accepted:** June 19, 2025. **Issue Period:** Vol.9 No.1 (2025), Pp. 180-190

**Abstrak:** Toko Donut “X” merupakan sebuah toko yang memproduksi variasi roti dan kue dengan menggunakan bahan pilihan terbaik tanpa menggunakan pengawet. Permasalahan yang terjadi toko tersebut adalah kesulitan untuk memprediksi permintaan konsumen dan juga belum bisa memprediksi penjualan di masa yang akan datang. Agar mempermudah pihak pengusaha dalam merencanakan penjualan serta penjadwalan produksi. Maka dilakukan prediksi penjualan produk menggunakan teknik K-Nearest Neighbor Regression. Metode yang diterapkan dalam penelitian ini adalah Knowledge Discovery in Database (KDD). Metode yang diterapkan melewati 5 tahap yaitu; Data Selection, Preprocessing, Transformation, Data Mining, Interpretation/Evaluation. Hasil pada penelitian ini menunjukkan prediksi penjualan terlaris pada produk Blackforest di bulan Juni dengan  $k = 12$  RMSE = 0.41876, Brownies di bulan Januari dengan  $k = 4$  RMSE = 0.46398, Coklat di bulan Februari dengan  $k = 9$  RMSE = 0.4827, Donat di bulan Juli dengan  $k = 14$  RMSE = 0.47980, Tart di bulan Juli dengan  $k = 12$  RMSE = 0.47381, Triple Disert Box di bulan November dengan  $k = 13$  RMSE = 0.49336. Sesuai dengan pedoman RMSE dapat disimpulkan bahwa seluruh model yang diuji memiliki tingkat kesalahan sedang dalam range 0,30 – 0,599.

**Kata kunci:** *K-Nearest Neighbor; Regression; Prediksi; RMSE; KDD*

**Abstract:** “X” Donut Shop is a shop that produces a variety of breads and cakes using the best selected ingredients without using preservatives. The problem that occurs is that the store still has difficulty predicting consumer demand and also cannot predict sales in the future. In order to facilitate the entrepreneur in planning sales and production scheduling. Then the prediction of product sales is made using the K-Nearest Neighbor Regression technique. The method applied in this research is





*Knowledge Discovery in Database (KDD). The method applied passes through 5 stages, namely; Data Selection, Preprocessing, Transformation, Data Mining, Interpretation/Evaluation. The results of this study indicate the bestselling predictions for Blackforest products in June with  $k = 12$  RMSE = 0.41876, Brownies in January with  $k = 4$  RMSE = 0.46398, Chocolate in February with  $k = 9$  RMSE = 0.4827, Donuts in July with  $k = 14$  RMSE = 0.47980, Tart in July with  $k = 12$  RMSE = 0.47381, Triple Desert Box in November with  $k = 13$  RMSE = 0.49336. In accordance with the RMSE guidelines, it can be concluded that all tested models have a moderate error rate in the range of 0.30 – 0.599.*

**Keywords:** K-Nearest Neighbor; Regression; Prediction; RMSE; KDD

## I. PENDAHULUAN

Tidak dipungkiri bahwa teknologi dan informasi dapat membawa perubahan dalam segala aspek, termasuk kepada perkembangan dunia bisnis. perkembangan teknologi yang pesat saat ini banyak menarik perhatian masyarakat luas. Demikian pula pada perkembangan komputer yang dinamis diiringi perkembangan zaman dapat menghasilkan informasi yang dibutuhkan, termasuk pada penjualan produk dimana kegiatan usahanya yang selalu membutuhkan perhitungan yang signifikan terhadap informasi untuk menjalankan sebuah penjualan yang dikelola setiap harinya.

Penjualan merupakan suatu unsur penting dalam suatu perusahaan di bidang pemasaran, berharap mendapat keuntungan yang lebih supaya bisa melanjutkan usaha tersebut. perusahaan yang didirikan untuk tujuan menghasilkan produk untuk kebutuhan para konsumen. perusahaan yang baik tentu memiliki tujuan jangka panjang supaya pada masa yang akan datang dapat mengalami perubahan yang baik.

Toko Donat “X” merupakan sebuah toko yang memproduksi variasi roti dan kue dengan menggunakan bahan pilihan terbaik dan tanpa menggunakan pengawet. Permasalahan yang terjadi adalah perusahaan masih mengalami kesulitan untuk memprediksi permintaan konsumen dan juga belum bisa memprediksi penjualan dimasa yang akan datang. Pembuatan produk yang tidak menggunakan pengawet membuat produk tidak bisa bertahan lama, sehingga ketika penjualan tidak habis maka produk harus dibuang, hal ini menyebabkan terjadinya pemborosan bahan baku. Permintaan yang tidak pasti juga mengakibatkan cara bekerja yang tidak efisien. sehingga dibutuhkan untuk memprediksi penjualan produk yang terdapat pada Toko Donat “X” dengan menerapkan metode Data Mining.

Data mining merupakan proses menemukan pola menarik dan pengetahuan dari sejumlah besar data. Sumber data bisa berupa database, warehouse, web, repositori dan informasi lainnya. terdapat beberapa metode data mining yang dapat digunakan salah satunya adalah K-Nearest Neighbor (KNN). KNN merupakan metode klasifikasi yang menentukan kategori berdasarkan mayoritas kategori pada k-Nearest Neighbor. KNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing. Data mining sering disebut juga Knowledge Discovery in Database, dengan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set berukuran besar, keluaran dari data mining ini dapat dipakai untuk memperbaiki pengambilan keputusan dimasa depan [1], [2], [3].

Prinsip kerja algoritma K-Nearest Neighbor (KNN) adalah mencari jarak terdekat dengan k tetangga (neighbor) terdekat dalam data training dengan data yang akan diuji. Teknik mengelompokkan data baru k dengan cara menghitung jarak data baru ke beberapa data/tetangga (neighbor) terdekat. Algoritma K-Nearest Neighbor (KNN) merupakan instead-based learning, dimana data training disimpan sehingga klasifikasi untuk record baru yang belum diklasifikasi dapat ditemukan dengan membandingkan kemiripan yang paling banyak dalam data training [4].

Sebelumnya terdapat penelitian yang dilakukan oleh [5] berupa perbandingan algoritma K-Nearest Neighbor dengan Decision Tree dalam memprediksi penjualan makanan hewan peliharaan di Petshop Dore Vet Clinic didapatkan hasil 6 data diklasifikasikan terlaris sesuai dengan prediksi yang dilakukan dengan metode K-





Nearest Neighbor (KNN) dari 30 data yang digunakan, selanjutnya [6] berupa prediksi harga beras premium dengan metode algoritma K-Nearest Neighbor (KNN) didapatkan hasil dapat melakukan prediksi terhadap harga beras pada tahun 2014-2019 menggunakan metode K-Nearest Neighbor (KNN) dengan nilai RMSE 0,125 dan parameter  $k = 2$  yang sudah dinormalisasi, selanjutnya [7] berupa komparasi metode data mining K-Nearest Neighbor (KNN) dengan Naive Bayes untuk klasifikasi kualitas air bersih (studi kasus PDAM tirta kencana kabupaten jombang) didapatkan hasil rata-rata nilai akurasi metode KNN sebesar 82.42%, selanjutnya [8] berupa penerapan metode K-Nearest Neighbor (KNN) dalam memprediksi masa studi mahasiswa (Studi Kasus : Mahasiswa STIKOM Dinamika Bangsa) didapatkan hasil untuk dua semester yaitu nilai  $k$  yang terbaik untuk digunakan memprediksi mahasiswa adalah nilai  $k=10$  dengan tingkat keberhasilan 69%, selanjutnya [9] berupa analisis performa metode K-Nearest Neighbor (KNN) untuk identifikasi jenis kaca didapatkan hasil nilai performa paling baik pada  $k=3$ , dimana tingkat akurasi mencapai 64%, presisi 63%, recall 71% dan F-Measure sebesar 67%. selanjutnya [10] berupa implementasi metode K-Nearest Neighbor (KNN) untuk prediksi penjualan furniture pada CV. Octo Agung Jepara dengan hasil menunjukkan metode yang diusulkan berhasil diimplementasikan untuk menyelesaikan kasus prediksi penjualan dengan tingkat error atau MSE sebesar 6% dan akurasi 94%.

Berdasarkan penelitian sebelumnya maka akan dilakukan penelitian dengan menerapkan Data Mining menggunakan algoritma K-Nearest Neighbor (KNN) untuk memprediksi penjualan produk pada tahun 2018 sampai 2020, dengan metode pengujian yang digunakan adalah metode k-fold cross validation. sebuah metode untuk membagi data training, testing, serta validasi, sedangkan evaluasi menggunakan RMSE (Root Mean Square error) untuk menguji akurasi model yang dibuat. sehingga perusahaan dapat melakukan estimasi berapa bahan baku yang dibutuhkan tanpa harus memakan waktu, tenaga, dan biaya tambahan lainnya dan juga untuk menambah pemasukan secara optimal.

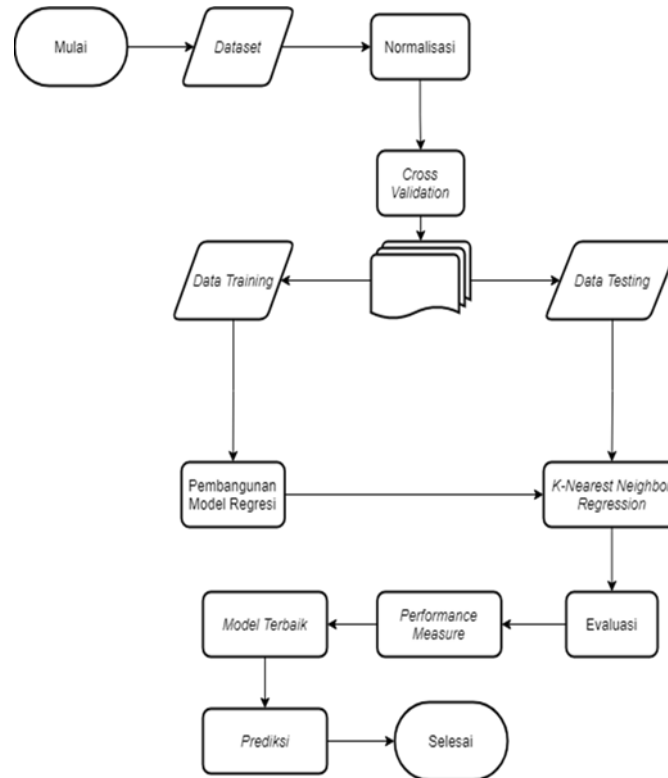
## II. METODE DAN MATERI

Tahapan penelitian ini seperti pada gambar dibawah ini:



DOI: 10.52362/jisicom.v9i1.1964

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



Gambar 1. Tahapan penelitian

Penelitian ini dimulai dari:

1. membangun dataset dari data penjualan mulai dari bulan ke-1 tahun ke-1 sampai dengan bulan ke-12 tahun ke -3
2. lalu dilakukan normalisasi data dengan cara membuat data yang sudah ada menjadi nilai yang lebih kecil
3. hasil dari normalisasi dilakukan cross Validation, data yang digunakan untuk training adalah data penjualan tahun ke-1 sampai tahun ke-3 sedangkan data yang digunakan untuk testing adalah data penjualan tahun ke 3
4. dalam tahap ini model k-nearest neighbor regression dibangun sebagai Sampel pengujian model yang akan digunakan pada evaluasi menggunakan cross validation[11]
5. pada tahap evaluasi ini dilakukan pengujian model menggunakan k-fold cross validation untuk setiap nilai k dari K-Nearest Neighbor dalam range 1-15
6. setelah didapati nilai error yang terkecil maka model yang dibentuk akan menggunakan nilai k dengan hasil error terkecil tersebut
7. setelah didapati model dengan error terkecil maka model terbaik telah didapatkan, selanjutnya adalah menguji model tersebut dengan data testing yang telah disiapkan
8. nilai prediksi merupakan ukuran skala dari normalisasi dengan nilai tertingginya adalah nilai penjualan yang paling laris.

### III. PEMBAHASAN DAN HASIL



DOI: 10.52362/jisicom.v9i1.1964

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



3.1. Data Penelitian

Pada bab ini akan dibahas mengenai hasil penelitian dengan menggunakan dataset penjualan produk Toko Donut ‘X’ dari tahun 1-3, dimana data tersebut akan digunakan training. Data training itu sendiri dikelompokan menjadi 2 bagian yaitu data input dan data target. Data input merupakan data penjualan dari bulan ke-1 sampai bulan ke-12, sedangkan data target menggunakan data bulan ke-13. Misal target penjualan bulan ke-1 tahun ke-2 maka data input yang digunakan adalah data penjualan dari bulan 1-12 tahun ke-1, dan seterusnya sampai data yang ada yaitu penjualan bulan ke-12 tahun ke-3. Data testing hanya berupa data input dimulai dari bulan ke-1 sampai bulan ke-12 tahun ke-3 yang akan menghasilkan prediksi penjualan tahun berikutnya.

Tabel 1. Data Penjualan Produk Donut

NO	PRODUK	JUMLAH		
		Tahun 1	Tahun 2	Tahun 3
1	BLACKFOREST	724	551	600
2	BROWNIES	1016	985	759
3	COKLAT	716	766	638
4	DONAT	15976	16047	15143
5	TART	1376	1271	1177
6	TRIPLE DISERT BOX	836	1030	978

Pada tabel 1. diatas, dapat dilihat pada data tersebut terlihat adanya peningkatan penjualan produk tertentu namun cenderung mengalami penurunan pada sebagian produk lainnya.

3.2. Pembentukan Dataset

Berikut ini merupakan contoh data training dan contoh data testing dari beberapa produk Toko Donut ‘X’:

Tabel 2. Data Training Donut

Data input												target
1367	1331	1291	1292	1329	1342	1315	1316	1366	1355	1293	1379	1369
1331	1291	1292	1329	1342	1315	1316	1366	1355	1293	1379	1369	1252
1291	1292	1329	1342	1315	1316	1366	1355	1293	1379	1369	1252	1361
1292	1329	1342	1315	1316	1366	1355	1293	1379	1369	1252	1361	1323
1329	1342	1315	1316	1366	1355	1293	1379	1369	1252	1361	1323	1262
1342	1315	1316	1366	1355	1293	1379	1369	1252	1361	1323	1262	1340
1315	1316	1366	1355	1293	1379	1369	1252	1361	1323	1262	1340	1419
1316	1366	1355	1293	1379	1369	1252	1361	1323	1262	1340	1419	1326
1366	1355	1293	1379	1369	1252	1361	1323	1262	1340	1419	1326	1317
1355	1293	1379	1369	1252	1361	1323	1262	1340	1419	1326	1317	1308
1293	1379	1369	1252	1361	1323	1262	1340	1419	1326	1317	1308	1345
1379	1369	1252	1361	1323	1262	1340	1419	1326	1317	1308	1345	1425
1369	1252	1361	1323	1262	1340	1419	1326	1317	1308	1345	1425	1368
1252	1361	1323	1262	1340	1419	1326	1317	1308	1345	1425	1368	1378





1361	1323	1262	1340	1419	1326	1317	1308	1345	1425	1368	1378	1143
1323	1262	1340	1419	1326	1317	1308	1345	1425	1368	1378	1143	1150
1262	1340	1419	1326	1317	1308	1345	1425	1368	1378	1143	1150	1152
1340	1419	1326	1317	1308	1345	1425	1368	1378	1143	1150	1152	1276
1419	1326	1317	1308	1345	1425	1368	1378	1143	1150	1152	1276	1303
1326	1317	1308	1345	1425	1368	1378	1143	1150	1152	1276	1303	1322
1317	1308	1345	1425	1368	1378	1143	1150	1152	1276	1303	1322	1299
1308	1345	1425	1368	1378	1143	1150	1152	1276	1303	1322	1299	1243
1345	1425	1368	1378	1143	1150	1152	1276	1303	1322	1299	1243	1254
1425	1368	1378	1143	1150	1152	1276	1303	1322	1299	1243	1254	1255

Tabel 3. Data Testing Donut

Data Input											
1252	1361	1323	1262	1340	1419	1326	1317	1308	1345	1425	1368
1361	1323	1262	1340	1419	1326	1317	1308	1345	1425	1368	1378
1323	1262	1340	1419	1326	1317	1308	1345	1425	1368	1378	1143
1262	1340	1419	1326	1317	1308	1345	1425	1368	1378	1143	1150
1340	1419	1326	1317	1308	1345	1425	1368	1378	1143	1150	1152
1419	1326	1317	1308	1345	1425	1368	1378	1143	1150	1152	1276
1326	1317	1308	1345	1425	1368	1378	1143	1150	1152	1276	1303
1317	1308	1345	1425	1368	1378	1143	1150	1152	1276	1303	1322
1308	1345	1425	1368	1378	1143	1150	1152	1276	1303	1322	1299
1345	1425	1368	1378	1143	1150	1152	1276	1303	1322	1299	1243
1425	1368	1378	1143	1150	1152	1276	1303	1322	1299	1243	1254
1368	1378	1143	1150	1152	1276	1303	1322	1299	1243	1254	1255

### 3.3. Normalisasi dataset

Proses normalisasi data, dilakukan dengan cara membuat data yang sudah ada menjadi nilai yang lebih kecil. Data hasil penelitian yang sudah diolah dinormalisasi dengan menjadikan data menjadi jarak [0, 1], artinya nilai minimal dari data menjadi 0 dan nilai maksimal dari data menjadi 1 dan data di antara minimal dan maksimal menyesuaikan antara jarak yang digunakan. Berikut hasil implementasi normalisasi:



DOI: 10.52362/jisicom.v9i1.1964

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).





0.705882	0.763158	0.780488	0.288462	0.769231	0.777778	0.351852	0.870370	0.611111	0.333333	0.685185	0.833333	0.240741
0.852941	0.842105	0.097561	0.769231	0.769231	0.351852	0.870370	0.611111	0.333333	0.685185	0.833333	0.240741	0.481481
0.941176	0.105263	0.707317	0.769231	0.326923	0.870370	0.611111	0.333333	0.685185	0.833333	0.240741	0.481481	0.481481
0.117647	0.763158	0.707317	0.326923	0.865385	0.611111	0.333333	0.685185	0.833333	0.240741	0.481481	0.481481	0.333333
0.852941	0.763158	0.146341	0.865385	0.596154	0.333333	0.685185	0.833333	0.240741	0.481481	0.481481	0.333333	0.481481
0.852941	0.157895	0.829268	0.596154	0.307692	0.685185	0.833333	0.240741	0.481481	0.481481	0.333333	0.481481	0.444444
0.176471	0.894737	0.487805	0.307692	0.673077	0.833333	0.240741	0.481481	0.481481	0.333333	0.481481	0.444444	0.518519
1.000000	0.526316	0.121951	0.673077	0.826923	0.240741	0.481481	0.481481	0.333333	0.481481	0.444444	0.518519	0.240741
0.588235	0.131579	0.585366	0.826923	0.211538	0.481481	0.481481	0.333333	0.481481	0.444444	0.518519	0.240741	0.277778
0.147059	0.631579	0.780488	0.211538	0.461538	0.481481	0.333333	0.481481	0.444444	0.518519	0.240741	0.277778	0.296296
0.705882	0.842105	0.000000	0.461538	0.461538	0.333333	0.481481	0.444444	0.518519	0.240741	0.277778	0.296296	0.481481
0.941176	0.000000	0.317073	0.461538	0.307692	0.481481	0.444444	0.518519	0.240741	0.277778	0.296296	0.481481	0.370370
0.000000	0.342105	0.317073	0.307692	0.461538	0.444444	0.518519	0.240741	0.277778	0.296296	0.481481	0.370370	0.944444
0.382353	0.342105	0.121951	0.461538	0.423077	0.518519	0.240741	0.277778	0.296296	0.481481	0.370370	0.944444	1.000000
0.382353	0.131579	0.317073	0.423077	0.500000	0.240741	0.277778	0.296296	0.481481	0.370370	0.944444	1.000000	0.037037
0.147059	0.342105	0.268293	0.500000	0.211538	0.277778	0.296296	0.481481	0.370370	0.944444	1.000000	0.037037	0.074074
0.382353	0.289474	0.365854	0.211538	0.250000	0.296296	0.481481	0.370370	0.944444	1.000000	0.037037	0.074074	0.000000
0.323529	0.394737	0.000000	0.250000	0.269231	0.481481	0.370370	0.944444	1.000000	0.037037	0.074074	0.000000	0.648148
0.441176	0.000000	0.048780	0.269231	0.461538	0.370370	0.944444	1.000000	0.037037	0.074074	0.000000	0.648148	0.537037
0.000000	0.052632	0.073171	0.461538	0.346154	0.944444	1.000000	0.037037	0.074074	0.000000	0.648148	0.537037	0.407407
0.058824	0.078947	0.317073	0.346154	0.942308	1.000000	0.037037	0.074074	0.000000	0.648148	0.537037	0.407407	0.481481
0.088235	0.342105	0.170732	0.942308	1.000000	0.037037	0.074074	0.000000	0.648148	0.537037	0.407407	0.481481	0.351852
0.382353	0.184211	0.926829	1.000000	0.000000	0.074074	0.000000	0.648148	0.537037	0.407407	0.481481	0.351852	0.537037
0.205882	1.000000	1.000000	0.000000	0.038462	0.000000	0.648148	0.537037	0.407407	0.481481	0.351852	0.537037	0.537037

Gambar 2. Hasil Normalisasi Data Training

Setelah data dinormalisasi maka dilakukan pembagian data kembali. Pembentukan variabel untuk data input adalah huruf X dengan data dimulai dari index ke-0 sampai 11, sedangkan data target menggunakan huruf y dengan data yang merupakan index ke 12.

### 3.4. Implementasi K-Fold Cross Validation

Dalam tahap ini digunakan K-Fold Cross Validation untuk membagi serta melakukan validasi data.

TRAIN: [ 0 1 2 4 5 6 7 8 9 10 11 12 14 15 16 17 19 20 21 22 23] TEST: [ 3 13 18]
TRAIN: [ 0 1 2 3 4 5 6 7 8 9 10 11 12 13 15 16 18 19 21 22 23] TEST: [14 17 20]
TRAIN: [ 0 1 3 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20 21 22 23] TEST: [ 2 4 10]
TRAIN: [ 0 1 2 3 4 5 8 9 10 11 12 13 14 15 16 17 18 20 21 22 23] TEST: [ 6 7 19]
TRAIN: [ 0 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 22 23] TEST: [ 1 21]
TRAIN: [ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 17 18 19 20 21 22 23] TEST: [ 0 16]
TRAIN: [ 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 16 17 18 19 20 21 22] TEST: [15 23]
TRAIN: [ 0 1 2 3 4 5 6 7 8 10 11 12 13 14 15 16 17 18 19 20 21 23] TEST: [ 9 22]
TRAIN: [ 0 1 2 3 4 5 6 7 9 10 11 13 14 15 16 17 18 19 20 21 22 23] TEST: [ 8 12]
TRAIN: [ 0 1 2 3 4 6 7 8 9 10 12 13 14 15 16 17 18 19 20 21 22 23] TEST: [ 5 11]

Gambar 3. Hasil Implementasi K-Fold Cross Validation





```

-----
Index TRAIN: [ 0 1 2 4 5 6 7 8 9 10 11 12 14 15 16 17 19 20 21 22 23] Index TEST: [ 3 13 18]
X_Train:
[[0.70588235 0.76315789 0.7804878 0.28846154 0.76923077 0.77777778
0.35185185 0.87037037 0.61111111 0.33333333 0.68518519 0.83333333]
[0.85294118 0.84210526 0.09756098 0.76923077 0.76923077 0.35185185
0.87037037 0.61111111 0.33333333 0.68518519 0.83333333 0.24074074]
[0.94117647 0.10526316 0.70731707 0.76923077 0.32692308 0.87037037
0.61111111 0.33333333 0.68518519 0.83333333 0.24074074 0.48148148]
[0.85294118 0.76315789 0.14634146 0.86538462 0.59615385 0.33333333
0.68518519 0.83333333 0.24074074 0.48148148 0.48148148 0.33333333]
[0.85294118 0.15789474 0.82926829 0.59615385 0.30769231 0.68518519
0.83333333 0.24074074 0.48148148 0.48148148 0.33333333 0.48148148]
[0.17647059 0.89473684 0.48780488 0.30769231 0.67307692 0.83333333
0.24074074 0.48148148 0.48148148 0.33333333 0.48148148 0.44444444]
[1. 0.52631579 0.12195122 0.67307692 0.82692308 0.24074074
0.48148148 0.48148148 0.33333333 0.48148148 0.44444444 0.51851852]
[0.58823529 0.13157895 0.58536585 0.82692308 0.21153846 0.48148148
0.48148148 0.33333333 0.48148148 0.44444444 0.51851852 0.24074074]
[0.14705882 0.63157895 0.7804878 0.21153846 0.46153846 0.48148148
0.33333333 0.48148148 0.44444444 0.51851852 0.24074074 0.27777778]
[0.70588235 0.84210526 0. 0.46153846 0.46153846 0.33333333
0.48148148 0.44444444 0.51851852 0.24074074 0.27777778 0.2962963 ]
[0.94117647 0. 0.31707317 0.46153846 0.30769231 0.48148148
0.44444444 0.51851852 0.24074074 0.27777778 0.2962963 0.48148148]
[0. 0.34210526 0.31707317 0.30769231 0.46153846 0.44444444
0.51851852 0.24074074 0.27777778 0.2962963 0.48148148 0.37037037]
[0.38235294 0.13157895 0.31707317 0.42307692 0.5 0.24074074
0.27777778 0.2962963 0.48148148 0.37037037 0.94444444 1. ]
[0.14705882 0.34210526 0.26829268 0.5 0.21153846 0.27777778
0.2962963 0.48148148 0.37037037 0.94444444 1. 0.03703704]
[0.38235294 0.28947368 0.36585366 0.21153846 0.25 0.2962963
0.48148148 0.37037037 0.94444444 1. 0.03703704 0.07407407]
[0.32352941 0.39473684 0. 0.25 0.26923077 0.48148148
0.37037037 0.94444444 1. 0.03703704 0.07407407 0. ]
[0. 0.05263158 0.07317073 0.46153846 0.34615385 0.94444444
1. 0.03703704 0.07407407 0. 0.64814815 0.53703704]
[0.05882353 0.07894737 0.31707317 0.34615385 0.94230769 1.
0.03703704 0.07407407 0. 0.64814815 0.53703704 0.40740741]
[0.08823529 0.34210526 0.17073171 0.94230769 1. 0.03703704
0.07407407 0. 0.64814815 0.53703704 0.40740741 0.48148148]
[0.38235294 0.18421053 0.92682927 1. 0. 0.07407407
0. 0.64814815 0.53703704 0.40740741 0.48148148 0.35185185]
[0.20588235 1. 1. 0. 0.03846154 0.
0.64814815 0.53703704 0.40740741 0.48148148 0.35185185 0.53703704]]
X_Test:
[[0.11764706 0.76315789 0.70731707 0.32692308 0.86538462 0.61111111
0.33333333 0.68518519 0.83333333 0.24074074 0.48148148 0.48148148]
[0.38235294 0.34210526 0.12195122 0.46153846 0.42307692 0.51851852
0.24074074 0.27777778 0.2962963 0.48148148 0.37037037 0.94444444]
[0.44117647 0. 0.04878049 0.26923077 0.46153846 0.37037037
0.94444444 1. 0.03703704 0.07407407 0. 0.64814815]]
Y_Train:
[0.24074074 0.48148148 0.48148148 0.48148148 0.44444444 0.51851852
0.24074074 0.27777778 0.2962963 0.48148148 0.37037037 0.94444444
0.03703704 0.07407407 0. 0.64814815 0.40740741 0.48148148
0.35185185 0.53703704 0.53703704]
Y_Test: [0.33333333 1. 0.53703704]

```

Gambar 4. Hasil Split K-Fold Cross Validation

### 3.5. Model K- Nearest Neighbor Regression

Dalam tahap ini model *k-nearest neighbor regression* dibangun sebagai sampel pengujian model yang akan digunakan pada evaluasi menggunakan *cross validation*. Implementasi pada program sebagai berikut:

```

## membangun model
model = KNeighborsRegressor(metric='euclidean')
model

KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='euclidean',
metric_params=None, n_jobs=None, n_neighbors=5, p=2,
weights='uniform')

```

Gambar 5. Implementasi K-NN dan Cross Validation

### 3.6. Evaluasi





Pada tahap evaluasi ini dilakukan pengujian model menggunakan *k-fold cross validation* untuk setiap nilai k dari *k-nearest neighbor* dalam range 1-15. Fungsi range (1, 15) berfungsi menciptakan ukuran angka dari 1-14 yang dideklarasikan ke dalam variabel k\_range dimana selanjutnya akan digunakan sebagai looping.

```

RMSE value for k= 1 is: 0.5630056026736993
RMSE value for k= 2 is: 0.5333043973631884
RMSE value for k= 3 is: 0.4816595112459242
RMSE value for k= 4 is: 0.46638549730750145
RMSE value for k= 5 is: 0.4505141096158509
RMSE value for k= 6 is: 0.43955645433005514
RMSE value for k= 7 is: 0.44444444444444444
RMSE value for k= 8 is: 0.4459609543889534
RMSE value for k= 9 is: 0.4384628343023749
RMSE value for k= 10 is: 0.43336894440569756
RMSE value for k= 11 is: 0.4214706229036839
RMSE value for k= 12 is: 0.4187601727736277
RMSE value for k= 13 is: 0.4209325783514135
RMSE value for k= 14 is: 0.4283803218827481
    
```

Gambar 5. 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.

### 3.7. Model Terbaik dan Prediksi

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

	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	51	51	43	51	49	53	38	40	41	51	45	76	0.453704
1	51	43	51	49	53	38	40	41	51	45	76	79	0.390432
2	43	51	49	53	38	40	41	51	45	76	79	27	0.427469
3	51	49	53	38	40	41	51	45	76	79	27	29	0.445988
4	49	53	38	40	41	51	45	76	79	27	29	25	0.490741
5	53	38	40	41	51	45	76	79	27	29	25	60	0.510802
6	38	40	41	51	45	76	79	27	29	25	60	54	0.483025
7	40	41	51	45	76	79	27	29	25	60	54	47	0.424383
8	41	51	45	76	79	27	29	25	60	54	47	51	0.458333
9	51	45	76	79	27	29	25	60	54	47	51	44	0.429012
10	45	76	79	27	29	25	60	54	47	51	44	54	0.467593
11	76	79	27	29	25	60	54	47	51	44	54	54	0.481481

Gambar 6. Hasil Prediksi Penjualan

Nilai prediksi merupakan ukuran skala dari normalisasi dengan nilai tertingginya adalah nilai penjualan yang paling laris. Berikut merupakan tabel hasil prediksi bulan penjualan terlaris untuk setiap produk Toko Donut “X”.

Tabel 4. Hasil Prediksi setiap produk





Produk	Prediksi
Blackforest	Bulan ke-6
Brownis	Bulan ke-1
Coklat	Bulan ke-2
Donat	Bulan ke-7
Tart	Bulan ke-6
Triple disert box	Bulan ke-11

#### IV. KESIMPULAN

Hasil penelitian dan pengujian model prediksi penjualan menggunakan metode K-Nearest Neighbor Regression, dapat ditarik kesimpulan bahwa nilai k yang paling optimal dari rentang 1 sampai 15 adalah nilai k = 12 menghasilkan nilai RMSE 0.41876 untuk produk Blackforest, k = 4 menghasilkan nilai RMSE 0.46398 untuk produk brownies, k = 9 menghasilkan nilai RMSE 0.46827 untuk produk coklat, k = 14 menghasilkan nilai RMSE 0.47980 untuk produk Donat, k = 12 menghasilkan nilai RMSE 0.47381 untuk produk Tart, dan k = 13 menghasilkan nilai RMSE 0.41876 untuk produk Triple Disert Box. Dengan demikian berdasarkan pedoman interpretasi RMSE dapat disimpulkan bahwa tingkat kesalahan seluruh model pengujian dalam penelitian prediksi penjualan produk dengan metode K-Nearest Neighbor Regression yang penulis uji memiliki kesalahan sedang dikarenakan nilai RMSE yang didapat berada dalam range 0,30 – 0,599. Kesalahan sedang berarti hasil prediksi sedikit mendekati nilai yang sebenarnya. Diperoleh hasil prediksi penjualan produk terlaris untuk Blackforest pada bulan ke-6, Brownies pada bulan ke-1, Coklat pada bulan ke-2, Donat pada bulan ke-7, Tart pada bulan ke-6, dan Triple Disert Box pada bulan ke-11. Penelitian lebih lanjut dapat menerapkan dan melakukan optimasi terhadap metode k-Nearest Neighbor dengan menambah jumlah data yang lebih banyak sehingga dapat menghasilkan training yang bervariasi serta mendapat hasil error yang lebih kecil.

#### REFERENSI

- [1] M. Reza Noviansyah, T. Rismawan, D. Marisa Midyanti, J. Sistem Komputer, and F. H. MIPA Universitas Tanjungpura Jl Hadari Nawawi, "PENERAPAN DATA MINING MENGGUNAKAN METODE K-NEAREST NEIGHBOR UNTUK KLASIFIKASI INDEKS CUACA KEBAKARAN BERDASARKAN DATA AWS (AUTOMATIC WEATHER STATION) (STUDI KASUS: KABUPATEN KUBU RAYA)," 2018.
- [2] I. A. Nikmatun and I. Waspada, "IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," *Jurnal SIMETRIS*, vol. 10, no. 2, 2019.
- [3] P. N. Harahap and S. Sulindawaty, "Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus PT.Arma Anugerah Abadi Cabang Sei Rampah)," *MATICS*, vol. 11, no. 2, p. 46, Jan. 2020, doi: 10.18860/mat.v11i2.7821.



DOI: 10.52362/jisicom.v9i1.1964

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



- [4] V. Kotu and B. Deshpande, *Predictive analytics and data mining : concepts and practice with RapidMiner*.
- [5] D. M. Meliala and P. Hasugian, "Perbandingan Algoritma K-Nearest Neighbor Dengan Decision Tree Dalam Memprediksi Penjualan Makanan Hewan Peliharaan Di Petshop Dore Vet Clinic," *Jurnal Teknologi Informasi*, vol. 3, 2020.
- [6] Y. Mukhlisin, M. Imrona, and D. T. Murdiansyah, "Prediksi Harga Beras Premium dengan Metode Algoritma K-Nearest Neighbor," 2020.
- [7] M. A. Rahman, N. Hidayat, and A. A. Supianto, "Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih (Studi Kasus PDAM Tirta Kencana Kabupaten Jombang)," 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] Jasmir, D. Zaenal Abidin, S. Nurmaini, and R. Firsandaya Malik, "Penerapan Metode K-Nearest Neighbor dalam Memprediksi Masa Studi Mahasiswa," 2017.
- [9] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "ANALISIS PERFORMA METODE K-NEAREST NEIGHBOR UNTUK IDENTIFIKASI JENIS KACA," *ILKOM Jurnal Ilmiah*, vol. 11, no. 3, pp. 269–274, Dec. 2019, doi: 10.33096/ilkom.v11i3.489.269-274.
- [10] R. Hutami and E. Z. Astuti, "Implementasi MetodemK-Nearest Neighbor Untuk Prediksi Penjualan Furniture Pasa CV.Octo Agung Jepara," 2016.
- [11] D. Berrar, "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1–3, Elsevier, 2018, pp. 542–545. doi: 10.1016/B978-0-12-809633-8.20349-X.

