JURNAL RHOMADONA BAMBANG PAMUNGKAS .pdf

by divapeniagustinn@gmail.com 1

Submission date: 23-Oct-2024 08:47PM (UTC+1100)
Submission ID: 2441794249
File name: JURNAL_RHOMADONA_BAMBANG_PAMUNGKAS_.pdf (505.38K)
Word count: 4213
Character count: 26211

JURNAL INFORMATIKA, Vol.11 No.1 April 2024, Halaman 1-8
ISSN: 2355-6579 I E-ISSN: 2528-2247

Penerapan metode K-Nearest Neighbor untuk peminatan konsumen terhadap jenis tabungan.

Rhomadona Bambang Pamungkas[1], Rachmat Adi Purnama[2]

[1,2] Universitas Bina Sarana Informatika Jakarta
Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450, Indonesia

E-mail : [1]bambangrbp@gmail.com, [2]rachmat.rap@bsi.ac.id

Informasi Artikel | Diterima: 00-00-2024 | Direvisi: 00-00-2024 | Disetujui: 00-00-2024

Abstrak

Penelitian ini bertujuan menganalisis preferensi nasabah terhadap berbagai jenis tabungan di Bank Woori Saudara KC Bogor dengan metode *K-Nearest Neighbor (KNN)* di platform Google Colab. Latar belakangnya adalah perkembangan teknologi informasi yang mengubah sektor perbankan, menuntut inovasi layanan yang lebih efisien. Penelitian ini bertujuan memahami preferensi dan kebutuhan nasabah untuk mempertahankan serta meningkatkan pangsa pasar. Permasalahan yang dijawab mencakup penerapan *KNN* dalam menganalisis preferensi konsumen dan bagaimana hasilnya dapat membantu bank merancang strategi pemasaran dan meningkatkan pelayanan. Data nasabah dari tahun 2006 hingga 2023 digunakan, dengan fokus pada sampel 5.000 data dari tahun 2018 hingga 2022. Metode *KNN* diterapkan untuk klasifikasi preferensi nasabah terhadap berbagai jenis tabungan. Variabel independen meliputi produk tabungan, usia nasabah, kebangsaan, pendapatan, saldo awal, saldo minimum, dan lainnya. Hasil evaluasi model *KNN* menunjukkan kinerja sangat baik dengan akurasi 95,66%, *precision* tertinggi 1,00, *recall* tertinggi 0,98, dan *F1-score* tertinggi 0,98. Hasil penelitian diharapkan membantu Bank Woori Saudara KC Bogor merancang strategi pemasaran yang lebih efektif dan meningkatkan layanan. Penelitian ini juga memperdalam pemahaman penulis tentang metode KNN dalam analisis preferensi konsumen di sektor perbankan serta memberikan contoh penerapan algoritma machine learning dalam kasus nyata.

Kata Kunci: Rekomendasi Tabungan, *K-Nearest Neighbor*, dan Nasabah

Abstract

This research aims to analyze customer preferences for various types of savings at Bank Woori Saudara KC Bogor using the K-Nearest Neighbor (KNN) method on the Google Colab platform. The background is the development of information technology which is changing the banking sector, demanding more efficient service innovation. This research aims to understand customer preferences and needs to maintain and increase market share. The problems answered include the application of KNN in analyzing consumer preferences and how the results can help banks design marketing strategies and improve services. Customer data from 2006 to 2023 is used, focusing on a sample of 5,000 data from 2018 to 2022. The KNN method is applied to classify customer preferences for various types of savings. Independent variables include savings products, customer age, nationality, income, initial balance, minimum balance, and others. The KNN model evaluation results show very good performance with an accuracy of 95.66%, the highest precision of 1.00, the highest recall of 0.98, and the highest F1-score of 0.98. The research results are expected to help Bank Woori Saudara KC Bogor design more effective marketing strategies and improve services. This research also deepens the author's understanding the KNN method in analyzing consumer preferences in the banking sector and provides examples of the application of machine learning algorithms in real cases.

Keywords: *Savings Recommendations, K-Nearest Neighbor, and Customer*

1. Pendahuluan

Perkembangan teknologi informasi mendorong bank untuk memahami preferensi nasabah dalam produk tabungan. Penelitian ini menggunakan metode K-Nearest Neighbor (KNN) di Google Colab untuk menganalisis data nasabah Bank Woori Saudara KC Bogor, bertujuan mengidentifikasi pola preferensi nasabah da

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

memberikan rekomendasi untuk strategi pemasaran yang lebih efektif.

Fokus penelitian adalah analisis data 5.000 sampel nasabah dari 2018 hingga 2022. Hasil analisis diharapkan dapat membantu Bank Woori Saudara dalam meningkatkan layanan dan efektivitas strategi pemasaran berdasarkan preferensi nasabah baru.

Data mining adalah proses untuk menggali informasi tersembunyi dari data, membantu bank dalam mengidentifikasi pola perilaku nasabah yang berguna untuk pengambilan keputusan. Teknologi ini memiliki beberapa tugas utama, seperti deskripsi, prediksi, klasifikasi, dan clustering, yang memungkinkan bank mengelola data nasabah secara efektif (Tarigan, Hardinata, Qurniawan, Safii, dan Winanjaya, 2022).

Bank berfungsi sebagai lembaga yang mengumpulkan dan menyalurkan dana, baik dalam bentuk simpanan maupun pinjaman. Bank konvensional menerapkan bunga, sedangkan bank syariah mengikuti prinsip-prinsip syariah seperti mudharabah dan wadiah, yang menghindari riba (Berlian, Andri, dan Apriana, 2023).

Tabungan adalah salah satu produk bank, yang disesuaikan dengan kebutuhan nasabah. Bank Woori Saudara menawarkan produk seperti Tabungan Woori Saudara, Tabungan Premium, dan Tabungan SimPel, yang masing-masing memberikan berbagai keuntungan dan fitur menarik (Putri & Aravik, 2021).

K-Nearest Neighbor (KNN) adalah algoritma klasifikasi yang digunakan untuk menganalisis preferensi nasabah berdasarkan jarak terdekat dengan data sebelumnya, membantu bank mengelompokkan nasabah berdasarkan produk tabungan yang sesuai (A'yuniyah dan Reza, 2023).

Microsoft Excel dan Google Colab mendukung pengolahan data dengan fitur penyimpanan dan visualisasi yang kuat, memudahkan analisis dan pengambilan keputusan berbasis data (Sartika, Lubis dan Lisdayanti, 2022).

Google Colab dilengkapi dengan berbagai Pustaka Python seperti Pandas, Matplotlib, dan Plotly yang dapat digunakan untuk memproses data dan membuat visualisasi data (Gelar Guntara, 2023).

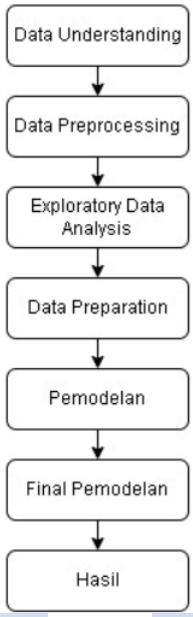
Pandas dan Numpy adalah dua library Python yang sangat populer digunakan dalam analisis data. Pandas memfasilitasi manipulasi data tabular, terutama dalam format seperti CSV, dengan menggunakan struktur data DataFrame yang memungkinkan pengguna mengolah data secara efisien. Sementara itu, Numpy mendukung manipulasi objek numerik dalam bentuk array multidimensi, yang sangat berguna untuk operasi komputasi matematis, seperti matriks dalam aljabar linier. Keduanya merupakan alat penting dalam ekosistem Python untuk analisis data yang efisien dan terstruktur (Irsan, Santos.B, dan Husain, 2024).

Python menyediakan alat untuk memecahkan sistem pertidaksamaan linier, membuat grafik pertidaksamaan dan melakukan operasi pada ekspresi pertidaksamaan (Palgunadi, Kadju, Alfaris, Mauludin, dan Rosyani, 2023).

Penelitian terkait menunjukkan penggunaan algoritma K-Nearest Neighbor (KNN) dalam berbagai konteks aplikasi (Rahmadhani, Vica, Arum dan Widya, 2022) memprediksi harga pangan di Indonesia dengan KNN, menghasilkan MAE sebesar 55.55 dan RMSE 81.64, menunjukkan prediksi yang baik. Penelitian oleh (Atmaja, Hakim, Haryadi dan Suwaryo 2021) menggunakan KNN untuk mengklasifikasikan risiko penyebaran Covid-19 di Jawa Barat, mencapai akurasi 85%. Di bidang pendidikan, Kartarina et al. (2021) membandingkan KNN dan Naive Bayes untuk memprediksi kelulusan mahasiswa, dengan KNN lebih unggul dengan akurasi 96,18%. Penelitian A'yuniyah dan Reza (2023) menggunakan KNN untuk klasifikasi jurusan siswa di SMA, dengan akurasi 93,52%. Secara keseluruhan, KNN efektif dalam prediksi dan klasifikasi di berbagai bidang.

2. Metode Penelitian

Metode penelitian ini dapat dilihat pada gambar berikut.



Gambar 1. Tahapan Penelitian

Berikut adalah penjelasan secara detail pada tahapan penelitian ini:

2.1. Data Understanding

Data understanding adalah tahap pengumpulan koleksi data awal, dan melakukan proses pengenalan terhadap data tersebut dengan tujuan untuk lebih mengenal nature dari data yang akan dipakai (Farissa, Mayasari, Umaidah, dan Yuyun, 2021).

2.2. *Data Preprocessing*

Data preprocessing adalah untuk membersihkan, menormalkan, dan mempersiapkan data sehingga dapat diolah lebih efektif oleh algoritma analisis atau model *machine learning*. Dataset ini terlebih dahulu dilakukan pembersihan, perbaikan, dan transformasi tanpa merubah validitas isi yang terkandung dalam dataset (Putra, Tahiyat, Ihsan, Rahmaddeni, Efrizoni, dan Lusiana, 2024).

2.3. *Exploratory Data Analysis*

"*Exploratory Data Analysis* adalah tahapan memiliki karakteristik yang fleksibel dan diperlukan untuk melakukan identifikasi dan investigasi suatu fenomena yang muncul pada saat melaukan penelitian empiris"(Khairu, 2020)

2.4. Data Preparation

"*Data preparation* adalah tahapan untuk memperbaiki masalah yang terdapat pada data sebelum data masuk ke tahap modeling sehingga menghasilkan modeling yang bagus"(Suhanda, 2020)

2.5. Pemodelan

Pemodelan adalah proses dalam data science di mana model matematis atau algoritma dibuat, dilatih, dan dievaluasi untuk memprediksi atau mengklasifikasikan data

2.6. Final Pemodelan

Final pemodelan adalah tahap akhir dalam pembuatan model prediktif di mana model terbaik dipilih, dilatih ulang dengan seluruh dataset, dievaluasi akhir, dan dioptimasi untuk performa optimal

2.7. Hasil

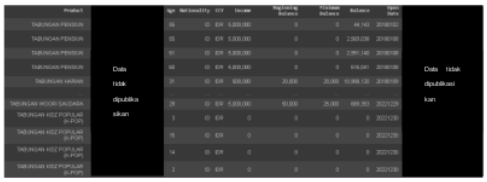
Hasil adalah tahapan penentuan dari model terlatih yang dievaluasi dengan metrik seperti akurasi dan error, visualisasi performa seperti confusion matrix, serta kesimpulan dan model yang siap digunakan untuk prediksi.

3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini mencakup 5.061 sampel dari periode 2018 hingga 2022. Hasil analisis menunjukkan bahwa variabel usia dan pendapatan sangat memengaruhi preferensi nasabah terhadap produk tabungan. Produk "Tabungan Woori Saudara" merupakan yang paling diminati oleh nasabah dengan jumlah pemilih terbanyak.

3.1. Data Understanding

Data penelitian ini berasal dari Bank Woori Saudara KC Bogor, diambil melalui sistem internal bank dengan teknik kuesioner. Data diperoleh dalam format CSV pada 1 November 2023, mencakup 5061 data dari periode 2018-2022, dengan 11 atribut meliputi jenis produk, nama, umur, kewarganegaraan, pendapatan, saldo awal, saldo minimum, saldo akhir, tanggal pembukaan akun, dan nama perusahaan.Berikut sampel datanya:



Gambar 2. Sampel Data

Tahap ini melibatkan pengenalan terhadap dataset yang berisi atribut-atribut seperti usia, pendapatan, kebangsaan, dan saldo awal. Dari eksplorasi awal, ditemukan bahwa data memiliki distribusi normal pada atribut usia, sementara pendapatan menunjukkan distribusi yang tidak merata.

Tabel 1. Jumlah Jenis Produk

NO	Produk	Jumlah
1	Tabungan Woori Saudara.	1798
2	Tabungan Pensiun	1482
3	Tabungan Kidz Popular (K-Pop)	901
4	Tabungan Cerdas	401
5	Tabungan Harian	204
6	Simpanan Pelajar	128
7	Tabungan Kami	97
8	Tabungankku	28
9	Tabungan Premium	22

Dataset menunjukkan bahwa "Tabungan Woori Saudara" paling populer dengan 1798 kemunculan, diikuti oleh "Pensiun Tabungan" (1482 kemunculan). Produk lain memiliki frekuensi lebih rendah, dengan "Tabungan Premium" paling jarang muncul (22 kali). Informasi ini membantu memahami distribusi dan popularitas produk bank.

Tabel 2. Jumlah Kebangsaan Nasabah

NO	Kode	Negara	Jumlah
1	ID	Indonesia	5050
2	KR	Korea	11

Dataset menunjukkan bahwa mayoritas nasabah berasal dari kewarganegaraan "ID" dengan 5050 kemunculan, sedangkan kewarganegaraan "KR" hanya muncul 11 kali. Ini menandakan dominasi kewarganegaraan "ID" dalam dataset.

3.2. Data Preprocessing

Pada tahap ini penting dalam analisis data yang mencakup pembersihan, transformasi, penskalaan, pengurangan dimensi, dan pembagian data. Langkah-langkah ini memastikan data berkualitas dan terstruktur, sehingga model analisis atau machine learning dapat menghasilkan hasil yang lebih akurat dan andal.

Pada tahap pembersihan data, atribut "Company Name" memiliki 1 nilai yang hilang, sementara atribut lainnya tidak ada yang hilang. Nilai hilang tersebut perlu diisi dengan nilai modus. Tidak ditemukan duplikasi data dalam dataframe, sehingga langkah pembersihan dan persiapan data dapat dilanjutkan.

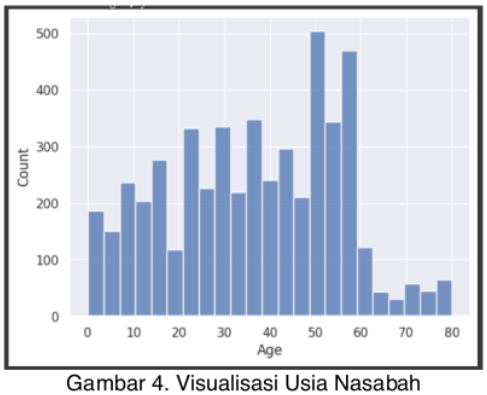
```
[33] # Daftar kolom numerik yang perlu diubah
     numerical_features = ['Age','Income', 'Beginning Balance', 'Minimum Balance', 'Balance']
     # Mengonversi kolom numerik menjadi string dan menghapus koma
     for feature in numerical_features:
         df[feature] = df[feature].astype(str).str.replace(',', '').astype(float)
```

Gambar 3. *Source Code Preprocessing*

Kode melakukan preprocessing fitur numerik dengan mengubah atribut menjadi tipe string, menghapus tanda koma, lalu mengonversinya kembali ke tipe float. Langkah ini memastikan format data numerik sesuai untuk analisis atau pemodelan lebih lanjut.

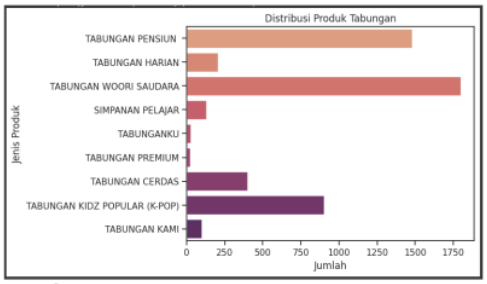
3.3. *Exploratory Data Analysis*

Tahapan ini adalah proses awal dalam analisis data untuk memahami struktur, pola, dan karakteristik data. EDA menggunakan statistik deskriptif, visualisasi, dan analisis korelasi untuk mengidentifikasi pola, tren, hubungan antar variabel, dan anomali. Teknik umum meliputi histogram, boxplot, scatter plot, heatmap, dan analisis korelasi. EDA memastikan data siap untuk analisis lanjutan dan pembuatan model.



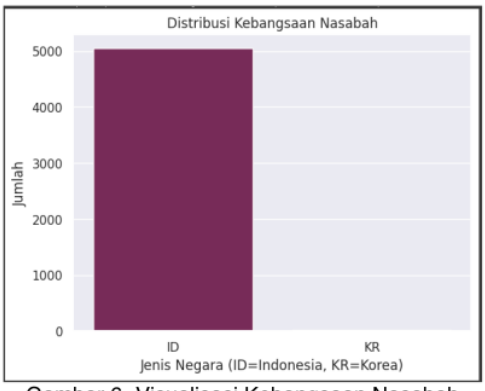
Gambar 4. Visualisasi Usia Nasabah

Visualisasi distribusi usia nasabah menggunakan histogram menunjukkan bahwa nasabah tersebar merata dari usia muda hingga pertengahan, dengan peningkatan signifikan pada kelompok usia 30-50 tahun dan puncaknya di usia sekitar 50 tahun dengan lebih dari 500 nasabah. Setelah usia 60 tahun, jumlah nasabah menurun drastis. Mayoritas nasabah berada di usia produktif (20-60 tahun). Distribusi ini membantu bank dalam menyesuaikan produk atau layanan sesuai dengan demografi nasabah.



Gambar 5. Visualisasi Produk Tabungan

Visualisasi distribusi produk tabungan nasabah menunjukkan bahwa Tabungan Woori Saudara mendominasi dengan sekitar 1798 nasabah. Tabungan Pensiun dan Tabungan Kidz Popular juga signifikan, masing-masing dengan sekitar 1482 dan 901 nasabah. Produk lainnya, seperti Tabungan Cerdas, Harian, Pelajar, Kami, Tabunganku, dan Premium, memiliki jumlah nasabah yang lebih kecil, berkisar antara 22 hingga 401. Diagram ini menyoroti bahwa Tabungan Woori Saudara, Pensiun, dan Kidz Popular adalah produk tabungan utama.



Gambar 6. Visualisasi Kebangsaan Nasabah

Visualisasi distribusi kebangsaan nasabah menunjukkan dominasi nasabah Indonesia (ID) dengan sekitar 5050 orang, sedangkan nasabah Korea (KR) sangat sedikit dan hampir tidak terlihat. Diagram ini dengan jelas memperlihatkan mayoritas nasabah berasal dari Indonesia.

3.4. *Data Preparation*

Data Preparation adalah tahap penting dalam analisis data yang mencakup pengumpulan, pembersihan, transformasi, dan pengurangan dimensi data. Proses ini juga melibatkan penskalaan, pembagian data menjadi set pelatihan dan pengujian, serta validasi data. Langkah-langkah ini memastikan data siap untuk analisis atau pemodelan dengan hasil yang akurat dan andal.

```
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
categorical_features = ['Product', 'Customer Name', 'Nationality', 'CCY', 'Company Name']
for feature in categorical_features:
    df[feature] = le.fit_transform(df[feature])
```

Gambar 7. *Source Code Lebel Encoder*

Kode *LabelEncoder* digunakan untuk mengubah nilai kategori menjadi numerik agar data kategori dapat diproses oleh algoritma machine learning. Kode menerapkan *label encoding* pada kolom kategorikal seperti *'Product', 'Customer Name', 'Nationality', 'CCY',* dan *'Company Name'*. Nilai *string* diubah menjadi integer dan disimpan kembali dalam *dataframe* untuk analisis atau pemodelan lebih lanjut.

Tabel 3. Hasil *Lebel Encoder*

Atribut	Lebel Encoding
Product	0: Simpanan Pelajar, 1: Tabungan Cerdas, 2: Tabungan Harian, 3: Tabungan Kami, 4: Tabungan *Kidz Popular* (K-Pop), 5: Tabungan Pensiun, 6: Tabungan Premium, 7: Tabungan Woori Saudara, 8: Tabungankku
Customer Name	Setiap nama nasabah diberikan label numerik yang unik dari 1 sampai 5061
Nationality	0: ID, 1: KR
CCY	0: IDR, 1: USD
Company Name	Setiap nama Perusahaan diberikan label numerik yang unik dari 1 sampai 5061

Kolom dalam tabel telah diubah menjadi nilai numerik: Kolom Product sekarang memiliki nilai dari 0 hingga 8, masing-masing mewakili berbagai jenis tabungan, mulai dari Simpanan Pelajar (0) hingga Tabunganku (8). Kolom Customer Name berisi label numerik unik dari 1 hingga 5061 untuk setiap nama nasabah. Kolom Nationality diubah menjadi 0 untuk ID dan 1 untuk KR, sedangkan kolom CCY menggunakan 0 untuk IDR dan 1 untuk USD. Terakhir, kolom Company Name juga berisi label numerik unik dari 1 hingga 5061 untuk setiap nama perusahaan.

```
[23] x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
```

Gambar 8. *Source Code Train Test Split*

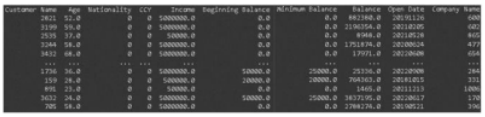
Dataframe dipisah menjadi dua bagian: x, yang berisi semua atribut kecuali "product", dan y, yang hanya berisi atribut "product". Pemisahan ini mempersiapkan x sebagai fitur input dan y sebagai target untuk prediksi dalam tugas machine learning. Dimensi x adalah (5061, 10) dan y (5061, 1), menunjukkan pemisahan berhasil. Selanjutnya, fungsi train_test_split digunakan untuk membagi dataset menjadi x_train, x_test, y_train, dan y_test dengan rasio 80:20, yaitu 4048 sampel untuk pelatihan dan 1007 untuk pengujian, menggunakan random_state=42 agar hasil dapat direproduksi.

```
print(f"x_train : {x_train.shape}")
print(f"y_train : {y_train.shape}")
print(f"x_test : {x_test.shape}")
print(f"y_test : {y_test.shape}")
```

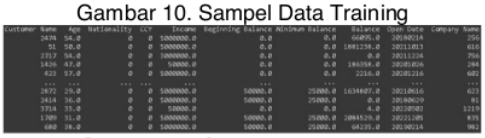
```
x_train : (4048, 10)
y_train : (4048,)
x_test : (1013, 10)
y_test : (1013,)
```

Gambar 9. *Source Code Testing* Dan *Training*

Dataset telah berhasil dibagi menjadi dua subset utama untuk machine learning: training set dan testing set. x_train terdiri dari 4048 sampel dengan 10 fitur, sementara y_train berisi label yang sesuai untuk setiap sampel. x_test memiliki 1013 sampel dengan jumlah fitur yang sama, dan y_test berisi label untuk setiap sampel dalam x_test. Subset ini siap digunakan untuk pelatihan dan evaluasi model machine learning. Berikut sampel data taining dan sampel data testing :



Gambar 10. Sampel Data Training



Gambar 11. Sampel Data Testing

3.5. *Pemodelan*

Pemodelan adalah proses dalam data science untuk membuat, melatih, dan mengevaluasi model prediksi atau klasifikasi data. Langkah-langkahnya meliputi pemilihan model, pelatihan dengan dataset, evaluasi kinerja, validasi, dan optimasi untuk meningkatkan akurasi. Tujuan utamanya adalah menghasilkan model yang dapat digunakan untuk pengambilan keputusan atau memberikan wawasan berdasarkan data.

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(x_train, y_train)#Melatih Model
```

Gambar 12. Melatih Pemodelan

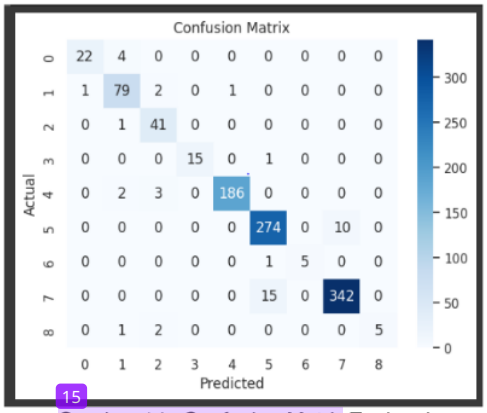
Peneliti akan melatih model K-Nearest Neighbors (KNN) menggunakan data latih yang telah disiapkan, dengan menginisialisasi model menggunakan parameter 3 tetangga terdekat untuk membuat prediksi.

	precision	recall	f1-score	support
0	0.96	0.85	0.90	26
1	0.91	0.95	0.93	83
2	0.85	0.98	0.91	42
3	1.00	0.94	0.97	16
4	0.99	0.97	0.98	191
5	0.94	0.96	0.95	284
6	1.00	0.83	0.91	6
7	0.97	0.96	0.96	357
8	1.00	0.62	0.77	8
accuracy			0.96	1013
macro avg	0.96	0.90	0.92	1013
weighted avg	0.96	0.96	0.96	1013

Akurasi KNN: 95.66%

Gambar 13. Hasil Evaluasi

Evaluasi model K-Nearest Neighbor (KNN) menunjukkan hasil yang sangat baik dalam klasifikasi data. Precision tertinggi mencapai 1.00 untuk kelas 3, 4, dan 6, menunjukkan ketepatan tinggi dalam memprediksi sampel di kelas tersebut. Recall tertinggi adalah 0.98 untuk kelas 2, mengindikasikan model mengenali hampir semua sampel di kelas tersebut. F1-score tertinggi, 0.98, diperoleh di kelas 4, menunjukkan keseimbangan precision dan recall. Model mencapai akurasi keseluruhan 95.66%. Macro average untuk precision, recall, dan F1-score masing-masing 0.96, 0.90, dan 0.92, sementara weighted average untuk ketiganya adalah 0.96. *Confusion matrix* adalah matriks yang menampilkan jumlah data terprediksi, baik yang benar maupun yang salah (Luthfi Bangun Permadi & Gumilang Sekolah Tinggi Teknologi Bandung, n.d.).



Gambar 14. *Confusion Matrix* Evaluasi

Evaluasi model K-Nearest Neighbor (KNN) menggunakan confusion matrix menunjukkan performa yang cukup baik dalam memprediksi kelas pada data uji. Mayoritas sampel diprediksi dengan benar, terlihat dari nilai besar pada diagonal utama matriks. Kinerja per kelas menunjukkan, misalnya, kelas 0 diprediksi dengan benar sebanyak 22 kali, kelas 1 sebanyak 79 kali,

dan seterusnya. Namun, terdapat beberapa kesalahan prediksi, seperti kelas 4 yang memiliki 3 sampel diprediksi sebagai kelas 3 dan kelas 5 yang memiliki 9 sampel diprediksi sebagai kelas 7. Secara keseluruhan, matriks ini menunjukkan kinerja model yang baik, meskipun masih terdapat kesalahan pada kelas dengan sampel lebih sedikit.

3.6. *Final Pemodelan*

Final pemodelan adalah tahap akhir dalam pembuatan model prediktif di mana model terbaik dipilih, dilatih ulang dengan seluruh dataset, dievaluasi akhir, dan dioptimasi untuk performa optimal. Model ini kemudian diimplementasikan dalam lingkungan produksi dan dipantau serta dipelihara secara berkala untuk memastikan kinerjanya tetap andal dan akurat dalam aplikasi nyata. "*K-fold cross validation* adalah metode yang digunakan untuk mengestimasi tingkat kesalahan dalam pengujian pemrosesan citra digital. Teknik ini bekerja dengan membagi data latih dan data uji ke dalam k kelompok yang saling terpisah, kemudian melakukan pengujian ulang sebanyak k kali untuk setiap kelompok tersebut" (Rozikin et al., 2020).

```
Cross-Validation Scores: [0.93089832 0.95948617 0.97332016 0.95948617 0.93379447]
Mean Cross-Validation Score: 95.14%
```

Gambar 15. Source Code Cross-validation

Cross-validation digunakan untuk mengevaluasi kinerja model K-Nearest Neighbor (KNN) dengan 3 tetangga terdekat menggunakan 5 fold. Hasil validasi silang menghasilkan *Cross-Validation Scores*: [0.93089832 0.95948617 0.97332016 0.95948617 0.93379447], dengan *Mean Cross-Validation Score*: 95.14%. Untuk menghitung nilai rata-rata skor validasi silang, dilakukan dengan cara:

a. Menjumlahkan seluruh skor validasi silang yang diperoleh pada setiap *fold*: 0.93089832 + 0.95948617 + 0.97332016 + 0.95948617 + 0.93379447 = 4.75698529
b. Menghitung rata-rata dengan membagi jumlah skor tersebut dengan jumlah *fold* (5): 4.75698529 / 5 = 0.9513970578
c. Mengalikan hasil rata-rata dengan 100 untuk mendapatkan persentase: 0.9513970578 * 100 = 95.14%

Proses ini menunjukkan bahwa model KNN memiliki performa yang sangat baik dalam melakukan klasifikasi pada dataset yang digunakan, dengan akurasi rata-rata yang konsisten di setiap fold. *GridSearchCV* adalah fungsi library pada python yang mengulang melalui parameter yang ditentukan terlebih dahulu dan mengimplementasikannya kedalam model untuk menemukan parameter terbaik (Jefika et al., 2020).

```
from sklearn.model_selection import GridSearchCV

param_grid = {'n_neighbors': np.arange(1, 20)}
knn_gscv = GridSearchCV(knn, param_grid, cv=5)
knn_gscv.fit(x_train, y_train)

print(f'Best k: {knn_gscv.best_params_["n_neighbors"]}')
```

```
Best k: 1
```

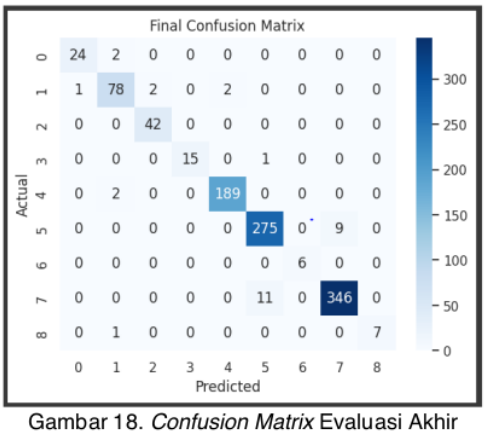
Gambar 16. Source Code GridSearchCV

Hyperparameter tuning pada algoritma K-Nearest Neighbors (KNN) bertujuan untuk menemukan parameter optimal seperti jumlah tetangga terdekat (n_neighbors) dan metrik jarak, guna meningkatkan akurasi model. Dalam kode yang menggunakan GridSearchCV dari scikit-learn, dilakukan pencarian nilai terbaik untuk n_neighbors dari 1 hingga 19. Hasil terbaik menunjukkan bahwa n_neighbors = 1 memberikan performa terbaik pada data. Tuning parameter dengan teknik seperti grid search sangat penting untuk meningkatkan kinerja model dalam memprediksi data baru.

	precision	recall	f1-score	support
0	0.96	0.92	0.94	26
1	0.94	0.94	0.94	83
2	0.95	1.00	0.98	42
3	1.00	0.94	0.97	16
4	0.99	0.99	0.99	191
5	0.96	0.97	0.96	284
6	1.00	1.00	1.00	6
7	0.97	0.97	0.97	357
8	1.00	0.88	0.93	8
accuracy			0.97	1013
macro avg	0.98	0.96	0.96	1013
weighted avg	0.97	0.97	0.97	1013

Gambar 17. Hasil Evaluasi Akhir

Hasil evaluasi akhir model K-Nearest Neighbor (KNN) menunjukkan performa yang sangat baik dengan akurasi keseluruhan mencapai 97.00%. Precision tertinggi adalah 1.00 untuk kelas 3, 6, dan 8, dan recall tertinggi adalah 1.00 untuk kelas 2, 4, 5, dan 6, menandakan ketepatan dan kemampuan model dalam mengenali sampel dengan sangat baik. F1-score tertinggi adalah 1.00 untuk kelas 6, menunjukkan keseimbangan optimal antara precision dan recall. Support tertinggi ditemukan pada kelas dengan 284 sampel. Macro average untuk precision, recall, dan F1-score masing-masing adalah 0.98, 0.96, dan 0.97, sedangkan weighted average ketiganya adalah 0.97, menandakan konsistensi model yang sangat baik di seluruh kelas.



Gambar 18. *Confusion Matrix* Evaluasi Akhir

Confusion matrix akhir untuk model K-Nearest Neighbor (KNN) menunjukkan kinerja yang baik dengan akurasi tinggi, di mana mayoritas sampel diprediksi dengan benar, seperti terlihat dari nilai besar pada diagonal utama. Kinerja per kelas bervariasi: kelas 0 (24 sampel benar), kelas 1 (78 sampel benar), kelas 2 (42 sampel benar), kelas 3 (15 sampel benar), kelas 4 (189 sampel benar), kelas 5 (275 sampel benar), kelas 6 (6 sampel benar), kelas 7 (346 sampel benar), dan kelas 8 (7 sampel benar). Beberapa kesalahan prediksi terjadi terutama pada kelas dengan jumlah sampel lebih sedikit, seperti 2 sampel kelas 1 yang diprediksi sebagai kelas 2, dan 9 sampel kelas 5 yang diprediksi sebagai kelas 7. Total 300 sampel diuji, dan matriks ini membantu menghitung metrik performa seperti akurasi, presisi, recall, dan F1-score, menunjukkan model KNN sebagai solusi klasifikasi yang efektif meskipun masih ada beberapa kesalahan.
4o mini

3.7. *Hasil*

Hasil preprocessing data untuk nasabah baru melibatkan pembersihan, transformasi kategori ke numerik, normalisasi atau standarisasi, serta penanganan nilai hilang. Setelah data siap, model yang telah dilatih digunakan untuk memprediksi nasabah baru, menghasilkan data yang siap dianalisis dan prediksi yang dioptimasi.

```
testing = {
    'Customer Name': ["Joko", "Bambang", "Yusup"],
    'Age': [56, 40, 20],
    'Nationality': [1, 1, 1],
    'CCY': [1, 1, 1],
    'Income': [1000000, 3000000, 5000000],
    'Beginning Balance': [50000, 75000, 75000],
    'Minimum Balance': [25000, 60000, 1000000],
    'Balance': [0, 30000, 200000],
    'Open Date': [20230101, 20230503, 20240606],
    'Company Name': ["PT JAYA", "TNI", "Perawat"],
}

testing = pd.DataFrame(testing)
testing_df = preprocess_data(testing)
testing_df
```

Gambar 19. *Source Code Data Testing*

Preprocessing data untuk nasabah baru melibatkan input atribut seperti 'Customer Name', 'Age', 'Nationality', 'CCY', 'Income', 'Beginning Balance', 'Minimum Balance', 'Balance', 'Open Date', dan 'Company Name'. Data testing disimpan dalam bentuk kamus, dikonversi menjadi dataframe, dan kemudian diproses lebih lanjut untuk analisis atau pemodelan.

```
pred_coba = knn_final.predict(testing_df)
print("Hasil Prediksi dari Nasabah Baru")
print(pred_coba)

Hasil Prediksi dari Nasabah Baru
[8 3 7]
```

Gambar 20. *Source Code* Prediksi

Model memprediksi jenis tabungan untuk nasabah baru berdasarkan fitur relevan dari data mereka. Hasil prediksi menunjukkan bahwa nasabah Joko direkomendasikan produk "Tabunganku," nasabah Bambang untuk produk "Tabungan Kami," dan nasabah Yusup untuk produk "Tabungan Woori Saudara." Model memberikan rekomendasi tabungan yang personal dan relevan.

4. Kesimpulan

Penelitian ini menunjukkan bahwa metode K-Nearest Neighbor (K-NN) di Google Colab efektif untuk menganalisis preferensi tabungan nasabah di Bank Woori Saudara KC Bogor dengan akurasi mencapai 96.94%. Model KNN berhasil mengidentifikasi pola preferensi nasabah dari data 2018-2022 dan memberikan rekomendasi tabungan yang personal untuk nasabah baru seperti Joko, Bambang, dan Yusup. Hasil ini memungkinkan bank untuk merancang strategi pemasaran yang lebih efektif dan meningkatkan layanan kepada nasabah.

Referensi

Berlian, D., Apriana, A., & Al-Amar Subang, S. (2023). PERBANDINGAN PEMBERIAN KREDIT ANTARA BANK KONVENSIONAL DAN PEMBIAYAAN BANK SYARIAH KEPADA USAHA KECIL DAN MENENGAH. In *Jurnal Perbankan Syariah Indonesia)* (Vol. 2, Issue 2).

Dewi Marini Umi Atmaja, A. R. H. , D. H. , N. S. (n.d.). tik-52-9-hal-penerapan-algoritma-k-nearest-neighbor-untuk-prediksi-pengelompokkan-tingkat-risiko-penyebaran-covid-19. *2021*.

Farissa, R. A., Mayasari, R., & Umaidah, Y. (2021). Perbandingan Algoritma K-Means

dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 5, Issue 2). http://jurnal.polibatam.ac.id/index.php/JAIC

Gelar Guntara, R. (2023). Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab. *Jurnal Ilmiah Multidisiplin*, *2*(6).

Homepage, J., A'yuniyah, Q., & Reza, M. (n.d.). *IJIRSE: Indonesian Journal of Informatic Research and Software Engineering Application Of The K-Nearest Neighbor Algorithm For Student Department Classification At 15 Pekanbaru State High School Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru*.

Jefika, M., Kosasi, H., Prayogi, G., & Dharma, A. (2020). Prediksi Gelombang Corona Dengan Metode Neural Network. *JIKOMSI Jurnal Ilmu Komputer Dan Sistem Informasi*, *3*(2), 102–107.

Khairu, N., Divisi, N., Data, A., Nugraha, Y., Febria, C., Divisi, F., Divisi, A. E., Intan, J., Divisi, K., Suherman, A. L., Penelitian, D., & Pengabdian, D. (2020). Evaluasi Berbasis Data: Kebijakan Pembatasan Mobilitas Publik dalam Mitigasi Persebaran COVID-19 di Jakarta. In *Jurnal Sistem Cerdas*.

Luthfi Bangun Permadi, M., & Gumilang Sekolah Tinggi Teknologi Bandung, R. (n.d.). *PENERAPAN ALGORITMA CNN (CONVOLUTIONAL NEURAL NETWORK) UNTUK DETEKSI DAN KLASIFIKASI TARGET MILITER BERDASARKAN CITRA SATELIT*.

Muhamad Irsan, Forkas Tiroy Santos.B, & Ahmad Husain. (2024). *Implementasi Aplikasi Pandas (Phyton) Dalam Mengelola Data Excel Sebagai Media Persiapan Pelaporan Nilai Raport Siswa*. https://jurnalpengabdianmasyarakatbangsa.com/index.php/jpmba/index

Palgunadi, Reinardus Di Caprio Kadju, Muhammad Rifqi Alfaris, Dion Mauludin, & Perani Rosyani. (2023). *Pemodelan Dan Penyelesaian Pertidaksamaan Dalam Kalkulus Menggunakan Python*.

Putra, F., Tahiyat, H. F., Ihsan, R. M., Rahmaddeni, R., & Efrizoni, L. (2024). Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(1), 273–281. https://doi.org/10.57152/malcom.v4i1.1085

Putri, N. S., & Aravik, H. (n.d.). *ANALISIS PRODUK TABUNGAN WADI'AH PADA PT. BANK PEMBIAYAAN RAKYAT SYARIAH AL FALAH BANYUASIN*.

Rahmadhani, V., Arum, W., Bhayangkara, U., & Raya, J. (2022). *LITERATURE REVIEW INTERNET OF THINK (IOT): SENSOR, KONEKTIFITAS DAN QR CODE*. *3*(2). https://doi.org/10.38035/jmpis.v3i2

Rozikin, C., Susilo Yuda Irawan, A., & Singaperbangsa Karawang, U. (2020). *Klasifikasi Jenis Buah Mangga dengan Metode Backpropagation*.

Sartika, A. R., Lubis, E., Lisdayanti, S., & Yudha, R. K. (2022). Pelatihan Aplikasi Microsoft Word, Microsoft Excel dan Power Point Pada siswa-siswi di SMPN 4 Kutacane. *Empowerment: Jurnal Pengabdian Masyarakat*, *1*(5), 712–721. https://doi.org/10.55983/empjcs.v1i5.249

Suhanda, Y., Kurniati, I., & Norma, S. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika Dan Komputer*, *6*(2), 12–20. https://doi.org/10.37012/jtik.v6i2.299

Tarigan, P. M. S., Hardinata, J. T., Qurniawan, H., Safii, M., & Winanjaya, R. (2022). Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang. *Jurnal Janitra Informatika Dan Sistem Informasi*, *2*(1), 9–19. https://doi.org/10.25008/janitra.v2i1.142

JURNAL RHOMADONA BAMBANG PAMUNGKAS .pdf

ORIGINALITY REPORT

15%	14%	9%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Nahdlatul Ulama Surabaya Student Paper	4%
2	Submitted to Universitas Bina Sarana Informatika Student Paper	1%
3	journal.irpi.or.id Internet Source	1%
4	Regita Santo Vara, Dwi Yuni Utami, Sari Dewi. "Perancangan dan Pengembangan Sistem Informasi Perpustakaan Buku Berbasis Web pada DITJEN POTHAN KEMHAN Jakarta", Jurnal Infortech, 2023 Publication	1%
5	jurnal.polibatam.ac.id Internet Source	1%
6	jutif.if.unsoed.ac.id Internet Source	1%
7	journal.thamrin.ac.id Internet Source	1%

8	journal-nusantara.com Internet Source	1 %
9	sostech.greenvest.co.id Internet Source	1 %
10	ejournal.sisfokomtek.org Internet Source	1 %
11	Nurussakinah Nurussakinah, Muhammad Faisal. "Klasifikasi Penyakit Diabetes Menggunakan Algoritma Decision Tree", Jurnal Informatika, 2023 Publication	<1 %
12	ejournal.bsi.ac.id Internet Source	<1 %
13	ojs.unud.ac.id Internet Source	<1 %
14	123dok.com Internet Source	<1 %
15	Arif Bijaksana Putra Negara, Hafiz Muhardi, Fahmi Sajid. "Perbandingan Algoritma Klasifikasi terhadap Emosi Tweet Berbahasa Indonesia", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2021 Publication	<1 %
16	medium.com Internet Source	<1 %

17	Nayny Frastika, Yunita Yunita. "Comparison of Social Media BOT Functions Using the K-Nearest Neighbor Method Against User Satisfaction", Jurnal Informatika, 2024 Publication	<1 %
18	caritulisan.com Internet Source	<1 %
19	es.scribd.com Internet Source	<1 %
20	ijazbaznas.com Internet Source	<1 %
21	jurnal.dharmawangsa.ac.id Internet Source	<1 %
22	repository.usd.ac.id Internet Source	<1 %
23	eprintslib.ummgl.ac.id Internet Source	<1 %
24	mpm.spbstu.ru Internet Source	<1 %
25	zahrannisakp.blogspot.com Internet Source	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		