

**PENERAPAN ALGORITMA K-NEAREST NEIGHBOR (K-NN) UNTUK
PREDIKSI LANGGANAN NASABAH PADA DEPOSITO BERJANGKA
MENGUNAKAN DATA BANK TERM DEPOSIT**



Diajukan Untuk Memenuhi Tugas Data Warehouse Dan Business Intelligence

Kelompok

19220737 - Rafif Setyo Nugroho

19220918 - Matius Dimas Prasetya

19221464 - Fahmi Akmal Aziz Pane

19220827 - Anugrah Akbar Riyadi

19220021 - Muhammad Reza Pahlevy

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS BINA SARANA INFORMATIKA JATIWARINGIN
TAHUN 2025**

DAFTAR ISI

COVER	1
DAFTAR ISI.....	2
BAB I.....	3
LATAR BELAKANG	3
1.1 Latar Belakang Masalah.....	3
1.2 Jurnal Referensi.....	3
BAB II.....	9
PEMODELAN.....	9
2.1 Metodologi Penelitian	9
2.2 Dataset dan Sumber Data	9
2.2.1 Atribut Demografis.....	9
2.2.2 Atribut Keuangan	9
2.2.3 Atribut Kampanye Pemasaran	10
2.2.4 Atribut Historis Kampanye Sebelumnya	10
2.2.5 Atribut Target (Label yang Diprediksi).....	10
2.3 Preprocessing Data.....	10
2.3.1 Preprocessing dengan Excel	11
2.3.2 Import dan Processing dengan RapidMiner	11
2.4 Implementasi Model K-NN dengan RapidMiner.....	12
2.4.1 Tampilan RapidMiner.....	12
2.4.2 Desain Process.....	12
BAB III	14
HASIL DAN PEMBAHASAN.....	14
3.1 Hasil Implementasi RapidMiner	14
3.2 Contoh Sebagian Data Prediksi.....	15
BAB IV	16
PENUTUP.....	16
4.1 Kesimpulan.....	16
4.2 Saran.....	16

BAB I

LATAR BELAKANG

1.1 Latar Belakang Masalah

Industri perbankan menghadapi tantangan kompetitif yang semakin ketat dalam menawarkan produk keuangan kepada nasabah. Term deposit merupakan salah satu produk investasi yang penting bagi bank karena memberikan sumber dana yang stabil dan menguntungkan. Namun, strategi pemasaran tradisional yang bersifat massal seringkali tidak efisien dan menghabiskan biaya yang besar tanpa memberikan hasil yang optimal.

Permasalahan utama yang dihadapi adalah bagaimana mengidentifikasi nasabah yang memiliki potensi tinggi untuk berlangganan term deposit. Pendekatan pemasaran acak kepada seluruh nasabah tidak hanya membuang-buang sumber daya, tetapi juga dapat mengganggu nasabah yang tidak berminat. Oleh karena itu, diperlukan suatu model prediktif yang dapat membantu bank dalam mengidentifikasi target nasabah yang tepat.

Machine learning, khususnya algoritma klasifikasi, menawarkan solusi untuk permasalahan ini. Dengan memanfaatkan data historis nasabah dan karakteristik demografis serta perilaku keuangan mereka, model klasifikasi dapat memprediksi probabilitas nasabah untuk berlangganan term deposit. Salah satu algoritma yang populer dan efektif untuk masalah klasifikasi adalah K-Nearest Neighbors (K-NN).

K-NN dipilih karena kesederhanaannya dalam implementasi, tidak memerlukan asumsi distribusi data, dan memberikan performa yang baik pada dataset dengan ukuran sedang. Algoritma ini bekerja berdasarkan prinsip bahwa data dengan karakteristik serupa cenderung memiliki hasil klasifikasi yang sama.

1.2 Jurnal Referensi

1. Prediksi Penjualan Produk Unilever Menggunakan Metode KNN

a) Link :

<https://www.jurnal.stkipppgritulungagung.ac.id/index.php/jipi/article/view/1910/936>

b) Permasalahan :

Toko Rizky Barokah di Nganjuk mengalami kesulitan dalam merencanakan stok barang, terutama untuk produk Unilever, karena permintaan konsumen yang bervariasi. Tidak adanya sistem prediksi menyebabkan risiko kelebihan atau kekurangan stok.

c) Solusi :

Menggunakan metode klasifikasi dalam data mining, khususnya algoritma K-Nearest Neighbor (K-NN), untuk memprediksi kategori penjualan produk Unilever (laris, sedang, atau kurang). Dengan prediksi ini, toko dapat lebih baik dalam menyusun strategi penyediaan stok.

d) Metode :

1. Algoritma K-Nearest Neighbor (K-NN): Klasifikasi berdasarkan kedekatan data baru terhadap data latih menggunakan rumus *Euclidean Distance*.
2. Data yang digunakan: penjualan produk dari Unilever, Indofood, Nestlé, dan P&G selama 3 tahun (2017–2019).
3. Nilai K yang diuji: 10, 20, 30, dan 40.
4. Evaluasi menggunakan Confusion Matrix dan perhitungan akurasi klasifikasi.

e) Hasil

1. Akurasi tertinggi: 86,66% (pada K = 30).
2. Akurasi terendah: 40% (pada K = 10).
3. Nilai K yang optimal berada pada kisaran 20–30.
4. Prediksi mingguan penjualan menunjukkan hasil yang konsisten, seperti kategori *sedang* untuk minggu tertentu.

2. Rancang Bangun Aplikasi Data Mining Prediksi Kelulusan Ujian Nasional Menggunakan Algoritma KNN Dengan Metode Euclidean Distance Pada SMPN 2 Pagedangan

a) Link :

<https://jurnal.umt.ac.id/index.php/jika/article/view/2288/1664>

b) Permasalahan :

Guru di SMPN 2 Pagedangan mengalami kesulitan dalam memprediksi kelulusan siswa pada Ujian Nasional (UN) berdasarkan nilai tryout. Belum tersedia sistem prediksi yang dapat mengolah data secara efisien dan akurat, sehingga pengambilan keputusan sulit dilakukan tepat waktu.

c) Solusi :

Dibuat sebuah aplikasi prediksi kelulusan Ujian Nasional menggunakan algoritma K-Nearest Neighbor (K-NN) dengan metode Euclidean Distance. Aplikasi ini mengklasifikasikan apakah siswa kemungkinan lulus atau tidak lulus berdasarkan nilai-nilai tryout.

d) Metode :

1. Algoritma: K-Nearest Neighbor (K-NN)

2. Pengukuran Jarak: Euclidean Distance
 3. Data: 744 data siswa dari tahun ajaran 2016–2019
 4. Atribut: Nilai tryout pertama dan kedua (MTK, B. Indonesia, B. Inggris, IPA)
 5. Model Pengembangan Sistem: SDLC model Waterfall
 6. Evaluasi: Akurasi, presisi, dan recall
- e) Hasil
1. Akurasi rata-rata: 88,42%
 2. Akurasi tertinggi ($K = 7$): 96,26%
 3. Presisi: 96,17%
 4. Recall: 97,32%
 5. Aplikasi mampu memproses data dalam jumlah besar dengan cepat dan akurat
 6. Sistem memberikan bantuan signifikan bagi sekolah dalam strategi peningkatan kelulusan
3. Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT.Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor
- a) Link :
- <https://ejournal.sisfokomtek.org/index.php/jikom/article/view/83/72>
- b) Permasalahan :
- PT. Nippon Indosari Corpindo Tbk (produsen Sari Roti) mengalami kesulitan menentukan produk roti terlaris karena sistem pemantauan penjualan masih dilakukan secara manual. Hal ini membuat data tidak akurat dan kurang efisien, sehingga sulit merencanakan produksi dengan optimal.
- c) Solusi :
- Mengembangkan aplikasi prediksi penjualan roti terlaris menggunakan metode K-Nearest Neighbor (K-NN) berbasis desktop. Sistem ini bertujuan untuk membantu perusahaan dalam menentukan jenis roti yang paling diminati masyarakat berdasarkan data penjualan 3 bulan terakhir.
- d) Metode :
1. Algoritma: K-Nearest Neighbor (K-NN)
 2. Pengukuran Jarak: Euclidean Distance
 3. Data: Kuantitas produk dan kuantitas terjual (dalam 3 bulan)
 4. Langkah-langkah:
 - a) Menentukan nilai **k** (dalam penelitian ini, $k = 5$)
 - b) Melakukan perhitungan jarak antara data uji dan data latih

c) Menentukan kategori produk berdasarkan mayoritas dari tetangga terdekat

5. Output prediksi: Apakah penjualan produk akan “Naik” atau “Menurun”

e) Hasil :

1. Aplikasi berhasil mengelompokkan dan memprediksi produk roti terlaris dengan akurasi logis berdasarkan klasifikasi bobot.
2. Contoh hasil: Produk dengan nilai prediksi mayoritas dari 5 tetangga terdekat menghasilkan kategori “Naik”.
3. Sistem membantu perusahaan menentukan prioritas produksi produk roti yang paling laku.

4. Sistem Informasi Prediksi Penjualan Produk Thrift di Toko Manado Menggunakan Algoritma KNN

a) Link :

<https://media.neliti.com/media/publications/579949-sistem-informasi-prediksi-penjualan-prod-edd31f5d.pdf>

b) Permasalahan :

Toko thrift di Manado mengalami kesulitan dalam menentukan produk thrift mana yang paling diminati dan layak untuk dijadikan stok. Tanpa sistem prediksi yang tepat, toko berisiko mengalami kerugian karena stok barang tidak laku terjual akibat salah perencanaan.

c) Solusi :

Dibuat sebuah sistem informasi berbasis klasifikasi menggunakan algoritma K-Nearest Neighbor (K-NN) untuk memprediksi penjualan produk thrift berdasarkan data penjualan per bulan. Sistem ini membantu toko dalam menganalisis produk-produk yang berpotensi tinggi untuk laku di masa depan.

d) Metode :

1. Algoritma: K-Nearest Neighbor (K-NN)
2. Metode jarak: Euclidean Distance
3. Nilai K yang digunakan: $K = 5$
4. Tahapan pengembangan: Model waterfall (analisis kebutuhan, desain, coding, pengujian, instalasi & pemeliharaan)

e) Hasil :

1. Sistem mampu mengklasifikasikan jenis produk thrift berdasarkan tingkat penjualan.

2. Contoh hasil: Produk thrift di Toko Benny diklasifikasikan ke dalam kategori tinggi dengan bobot nilai sebesar 102.77, menandakan minat beli konsumen yang kuat.
3. Sistem ini mendukung pengambilan keputusan untuk penyediaan stok secara lebih akurat dan efisien.

5. Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma KNN

a) Link :

<https://e-journal.stmiklombok.ac.id/index.php/misi/article/view/125/84>

b) Permasalahan :

Informasi prakiraan cuaca seringkali membingungkan dan tidak akurat karena bergantung pada interpretasi subjektif dari prakirawan cuaca dan berasal dari berbagai sumber data yang kompleks. Padahal, prakiraan yang tepat sangat dibutuhkan oleh banyak sektor seperti pertanian, perikanan, pariwisata, dan konstruksi.

c) Solusi :

Mengembangkan aplikasi prediksi cuaca harian berbasis data mining menggunakan algoritma K-Nearest Neighbor (KNN) yang dapat mengklasifikasikan kondisi cuaca secara otomatis berdasarkan data historis dari BMKG Gorontalo.

d) Metode :

1. Algoritma: K-Nearest Neighbor (KNN)
2. Pengukuran jarak: Euclidean Distance
3. Data: Curah hujan harian dengan atribut:
 - a) Temperatur rata-rata
 - b) Penyinaran matahari
 - c) Peristiwa cuaca khusus
 - d) Tekanan udara
 - e) Kelembaban rata-rata
 - f) Angin rata-rata
4. Pra-pemrosesan: Data harian diakumulasi menjadi bulanan
5. Validasi model: Menggunakan K-Fold Cross Validation dan evaluasi Root Mean Square Error (RMSE)
6. Pengujian: Menggunakan 335 data, dengan split 70% untuk training dan 30% untuk testing

e) Hasil :

1. Sistem mampu memprediksi cuaca harian ke dalam kategori: tidak hujan, hujan ringan, hujan lebat, dan sangat lebat.
2. Nilai RMSE yang diperoleh dari pengujian: 9.899 ± 0.000 (semakin rendah RMSE, semakin baik akurasi prediksi)
3. Aplikasi mampu mempermudah masyarakat mengakses informasi potensi hujan secara cepat dan mudah dipahami

BAB II

PEMODELAN

2.1 Metodologi Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode Cross-Industry Standard Process for Data Mining (CRISP-DM) yang terdiri dari tahapan business understanding, data understanding, data preparation, modeling, evaluation, dan deployment.

2.2 Dataset dan Sumber Data

Dataset yang digunakan bersumber dari file Assignment-2_Data.csv yang berisi informasi historis kampanye pemasaran bank. Dari total 45.212 records, dipilih 499 sampel secara random untuk efisiensi komputasi dan proof of concept. Dataset terdiri dari 17 atribut kondisi dan 1 atribut target (keputusan berlangganan term deposit).

Id	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
1001	99	managem	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
1002	44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	unknown	no
1003	33	entreprene	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	0	unknown	no
1004	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	0	unknown	no
1005	33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	0	unknown	no
1006	35	managem	married	tertiary	no	231	yes	no	unknown	5	may	139	1	-1	0	unknown	no
1007	28	managem	single	tertiary	no	447	yes	yes	unknown	5	may	217	1	-1	0	unknown	no

Atribut-atribut dalam dataset meliputi:

2.2.1 Atribut Demografis

1. Age: Usia nasabah (dalam tahun). Digunakan untuk memahami preferensi usia terhadap produk keuangan.
2. Job: Jenis pekerjaan nasabah, seperti *management*, *technician*, *blue-collar*, dll. Dapat mencerminkan penghasilan dan stabilitas kerja.
3. Marital: Status pernikahan nasabah (*single*, *married*, *divorced*). Mempengaruhi prioritas keuangan.
4. Education: Tingkat pendidikan terakhir nasabah (*primary*, *secondary*, *tertiary*, *unknown*). Pendidikan dapat menunjukkan tingkat pemahaman terhadap risiko keuangan.

2.2.2 Atribut Keuangan

1. Balance: Saldo rata-rata tahunan di rekening nasabah. Angka ini merefleksikan kekuatan finansial dan kemungkinan untuk mengambil produk keuangan.
2. Housing: Status pinjaman rumah. Nilai *yes* berarti nasabah sedang memiliki cicilan rumah.

3. Loan: Status pinjaman pribadi. *Yes* berarti nasabah memiliki pinjaman pribadi aktif.

2.2.3 Atribut Kampanye Pemasaran

1. Contact: Jenis saluran komunikasi terakhir yang digunakan, seperti *cellular*, *telephone*, atau *unknown*.
2. Duration: Durasi panggilan terakhir (dalam detik). Semakin lama, sering kali semakin besar kemungkinan nasabah tertarik.
3. Campaign: Jumlah total kontak (panggilan) yang dilakukan ke nasabah selama kampanye saat ini. Angka tinggi bisa berarti nasabah belum tertarik meskipun telah dihubungi berkali-kali.

2.2.4 Atribut Historis Kampanye Sebelumnya

1. Pdays: Jumlah hari sejak nasabah terakhir kali dihubungi dalam kampanye sebelumnya.
 - a. Nilai -1 berarti nasabah belum pernah dihubungi sebelumnya.
 - b. Angka positif menunjukkan jarak waktu (dalam hari) sejak kontak terakhir dari kampanye sebelumnya.
2. Previous: Jumlah total kontak yang dilakukan ke nasabah dalam kampanye sebelumnya.
 - a. Nilai 0 berarti tidak pernah dihubungi sebelumnya.
 - b. Nilai lebih besar menunjukkan berapa kali nasabah pernah dihubungi sebelumnya.
3. Poutcome: Hasil dari kampanye pemasaran sebelumnya.
 - a. *success* = nasabah membeli produk pada kampanye sebelumnya.
 - b. *failure* = tidak membeli.
 - c. *other/unknown* = hasil tidak diketahui atau tidak relevan.

2.2.5 Atribut Target (Label yang Diprediksi)

1. y: Apakah nasabah akhirnya berlangganan produk *term deposit*?
 - a. *yes* = berlangganan (positif class)
 - b. *no* = tidak berlangganan

2.3 Preprocessing Data

Tahapan preprocessing dilakukan dengan kombinasi Microsoft Excel dan RapidMiner

2.3.1 Preprocessing dengan Excel

1. Sebelum:

Id	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
1001,999		management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
1002,44		technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	unknown	no
1003,33		entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	0	unknown	no
1004,47		blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	0	unknown	no
1005,33		unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	0	unknown	no
1006,35		management	married	tertiary	no	231	yes	no	unknown	5	may	139	1	-1	0	unknown	no
1007,28		management	single	tertiary	no	447	yes	yes	unknown	5	may	217	1	-1	0	unknown	no
1008,,		entrepreneur	divorced	tertiary	yes,,	yes	no	unknown	5	may	380	1	-1	0	unknown	no	
1009,58		retired	married	primary	no,,	yes	no	unknown	5	may	50	1	-1	0	unknown	no	
1010,43		technician	single	secondary	no,,	yes	no	unknown	5	may	55	1	-1	0	unknown	no	
1011,41		admin.,	divorced	secondary	no	270	yes	no	unknown	5	may	222	1	-1	0	unknown	no
1012,29		admin.,	single	secondary	no	390	yes	no	unknown	5	may	137	1	-1	0	unknown	no
1013,53		technician	married	secondary	no	6	yes	no	unknown	5	may	517	1	-1	0	unknown	no
1014,58		technician	married	unknown	no	71	yes	no	unknown	5	may	71	1	-1	0	unknown	no
1015,57		services	married	secondary	no	162	yes	no	unknown	5	may	174	1	-1	0	unknown	no
1016,51		retired	married	primary	no	229	yes	no	unknown	5	may	353	1	-1	0	unknown	no
1017,45		admin.,	single	unknown	no	13	yes	no	unknown	5	may	98	1	-1	0	unknown	no

2. Sesudah:

Id	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
1001	99	managem	married	tertiary	no	2143	yes	no	unknown		5 may	261	1	-1	0	unknown	no
1002	44	technician	single	secondary	no	29	yes	no	unknown		5 may	151	1	-1	0	unknown	no
1003	33	entreprene	married	secondary	no	2	yes	yes	unknown		5 may	76	1	-1	0	unknown	no
1004	47	blue-collar	married	unknown	no	1506	yes	no	unknown		5 may	92	1	-1	0	unknown	no
1005	33	unknown	single	unknown	no	1	no	no	unknown		5 may	198	1	-1	0	unknown	no

- Data Cleaning: Menghapus baris yang kosong atau memiliki atribut yang tidak lengkap
- Data Formatting: Merapikan format data agar konsisten dan siap untuk import
- Initial Validation: Memastikan struktur data sesuai dengan kebutuhan RapidMiner

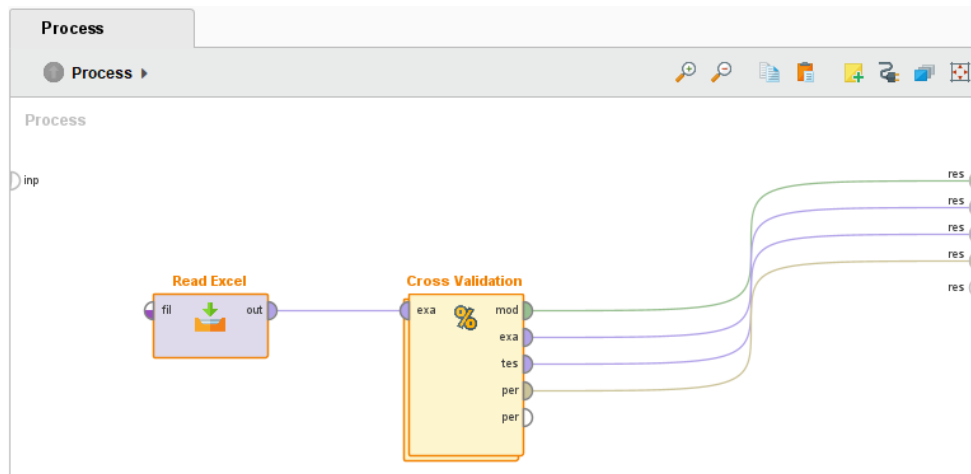
2.3.2 Import dan Processing dengan RapidMiner

- Read Excel Operator: Menggunakan operator "Read Excel" untuk mengimport dataset yang telah diproses
- Data Type Conversion: Memastikan setiap atribut memiliki tipe data yang sesuai (numerical/nominal)
- Data Validation: Validasi final sebelum masuk ke tahap modeling

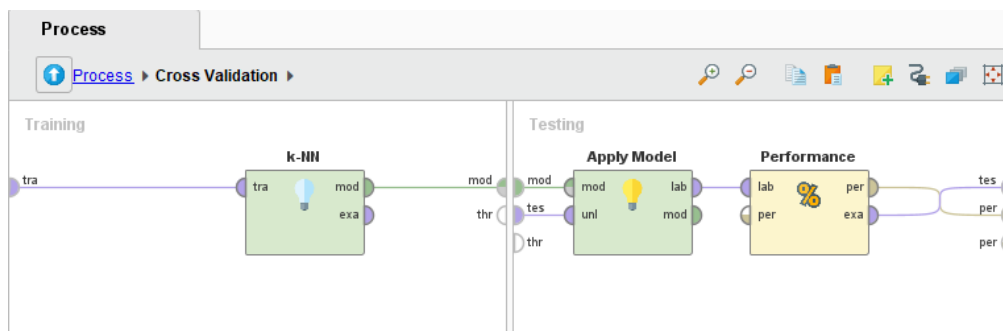
2.4 Implementasi Model K-NN dengan RapidMiner

2.4.1 Tampilan RapidMiner

1. Tampilan Proses



2. Tampilan Cross Validation



2.4.2 Desain Process

1. Read Excel: Operator utama untuk membaca dataset
2. Cross Validation: Operator untuk implementasi K-Fold Cross Validation yang berisi:
 - a) Training Subprocess:
 1. k-NN Operator: Untuk membangun model K-NN dengan parameter:
 2. $k = 5$ (jumlah tetangga terdekat)
 3. Distance measure = Euclidean distance
 - b) Testing Subprocess:
 1. Apply Model: Menerapkan model K-NN yang telah dilatih pada data testing
 2. Performance (Binomial Classification): Evaluasi performa dengan metrik:
 - a. Accuracy, Precision, Recall, F1-Score
 - b. ROC Curve dan AUC
 - c. Confusion Matrix

c) Konfigurasi Cross Validation:

1. Number of Folds: 10
2. Sampling Type: Automatic

d) K-NN dalam RapidMiner:

1. k: 5 (optimal berdasarkan grid search)
2. Measure Types: Mixed measures untuk menangani atribut numerical dan nominal
3. Mixed Measure: Mixed Euclidean distance

BAB III

HASIL DAN PEMBAHASAN

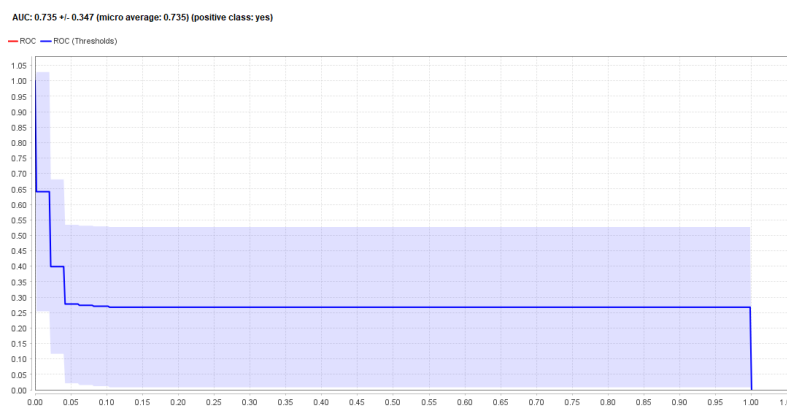
3.1 Hasil Implementasi RapidMiner

Setelah dilakukan proses klasifikasi terhadap dataset menggunakan RapidMiner, diperoleh hasil sebagai berikut:

1. Akurasi: $98.00\% \pm 1.33\%$

Menunjukkan bahwa model mampu memprediksi sebagian besar data dengan benar.

2. AUC (Area Under Curve): 0.735 ± 0.347



Nilai ini menggambarkan kemampuan model untuk membedakan antara kelas positif dan negatif. Nilai AUC > 0.7 menunjukkan performa klasifikasi yang cukup baik, meskipun belum optimal.

3. Precision (kelas positif): 66.67%

Dari seluruh prediksi "yes", hanya 66.67% yang benar-benar benar. Ini menunjukkan bahwa model memiliki akurasi sedang saat memutuskan kelas positif.

4. Recall (kelas positif): 18.18%

Dari seluruh data yang seharusnya "yes", hanya 18.18% yang berhasil dideteksi. Ini memperlihatkan kelemahan model dalam mengenali kelas minoritas.

5. F-Measure (kelas positif): 28.57%

Kombinasi harmonis dari precision dan recall. Nilai ini rendah karena recall-nya sangat rendah.

6. Confusion Matrix

	True No	True Yes	Class Precision
Pred. No	487	9	98.19%
Pred. Yes	1	2	66.67%
Class Recall	99.80%	18.18%	

Analisis Hasil

1. Akurasi Tinggi \neq Model Baik

Akurasi sebesar 98% terlihat mengesankan, tetapi perlu dicermati bahwa dataset bersifat tidak seimbang. Kelas “no” jauh lebih banyak dibandingkan kelas “yes”, sehingga model cenderung mengklasifikasikan data sebagai “no” untuk mendapatkan skor akurasi tinggi.

2. Recall Rendah pada Kelas Positif

Hanya 18.18% dari pelanggan yang benar-benar berlangganan term deposit berhasil dikenali. Ini sangat rendah dan mengindikasikan model tidak sensitif terhadap kelas minoritas.

3. Precision Kelas Positif Tinggi

Meskipun recall rendah, precision pada kelas positif 66.67% menunjukkan bahwa jika model memprediksi “yes”, kemungkinan besar prediksi itu benar. Namun, karena terlalu banyak kasus “yes” yang tidak dikenali, precision ini kurang berarti tanpa recall yang seimbang.

4. Nilai AUC 0.735

Mengindikasikan model memiliki kemampuan sedang dalam membedakan antara pelanggan yang akan dan tidak akan berlangganan produk.

5. Trade-off ROC Curve

Berdasarkan AUC dan grafik ROC (tidak ditampilkan di sini), terdapat trade-off antara True Positive Rate (TPR) dan False Positive Rate (FPR). Perlu analisis ambang batas klasifikasi untuk meningkatkan sensitivitas terhadap kelas “yes”.

3.2 Contoh Sebagian Data Prediksi

ROW NO	y	pred	confiden	confidenc	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome
1	no	no	0	1	39	management	single	tertiary	no	255	yes	no	unknown	5	may	296	1	-1	0	unknown
2	no	no	0	0.999999	50	management	married	secondary	no	49	yes	no	unknown	5	may	180	2	-1	0	unknown
3	no	no	0	1	44	technician	married	secondary	no	0	yes	no	unknown	5	may	225	2	-1	0	unknown
4	no	no	0	0.999999	55	blue-collar	married	primary	no	23	yes	no	unknown	5	may	291	1	-1	0	unknown
5	no	no	0	0.999999	43	technician	married	secondary	no	1937	yes	no	unknown	5	may	181	1	-1	0	unknown

BAB IV

PENUTUP

4.1 Kesimpulan

Penelitian ini telah berhasil menerapkan algoritma *K-Nearest Neighbor* (K-NN) untuk menyelesaikan permasalahan klasifikasi dalam prediksi apakah seorang nasabah akan melakukan langganan pada produk deposito berjangka. Dataset yang digunakan adalah dataset yang terdiri dari 499 data nasabah dengan 17 atribut prediktor dan 1 atribut target (y) yang menyatakan apakah nasabah melakukan langganan ("yes") atau tidak ("no").

Hasil yang diperoleh dari pemodelan menggunakan *RapidMiner* dengan pendekatan *10-fold cross validation* menunjukkan bahwa model memiliki akurasi sebesar $98.00\% \pm 1.33\%$, yang secara umum menunjukkan kinerja sangat baik dalam mengklasifikasikan data. Namun demikian, nilai recall sebesar 18.18% dan F1-score sebesar 28.57% untuk kelas "yes" (positif) mengindikasikan bahwa model masih kurang optimal dalam mendeteksi nasabah yang benar-benar akan melakukan langganan.

Confusion matrix yang dihasilkan memperlihatkan bahwa sebagian besar prediksi berada pada kelas mayoritas ("no"), sedangkan hanya sebagian kecil dari kelas "yes" yang berhasil terklasifikasi dengan benar. Hal ini menandakan adanya ketimpangan kelas (class imbalance) yang signifikan, di mana proporsi antara kelas "no" dan "yes" tidak seimbang. Kurva ROC dan nilai AUC sebesar 0.735 ± 0.347 menunjukkan bahwa model memiliki kapabilitas sedang dalam membedakan kedua kelas tersebut.

Dengan demikian, dapat disimpulkan bahwa meskipun algoritma K-NN memberikan akurasi tinggi, namun sensitivitas terhadap kelas minoritas masih rendah. Ini menjadi pertimbangan penting dalam penerapan model prediktif di sektor keuangan, khususnya dalam strategi pemasaran produk perbankan.

4.2 Saran

Untuk meningkatkan performa model dalam mendeteksi nasabah yang berpotensi melakukan langganan terhadap produk deposito berjangka, terdapat beberapa langkah strategis yang dapat dilakukan. Pertama, perlu diterapkan metode penyeimbangan data, seperti *SMOTE* (*Synthetic Minority Over-sampling Technique*) atau *undersampling* terhadap kelas mayoritas untuk mengurangi ketimpangan kelas. Dengan pendekatan ini, model akan lebih peka terhadap pola-pola dari kelas minoritas.

Kedua, disarankan untuk melakukan tuning parameter K pada algoritma K-NN. Pemilihan nilai K yang optimal sangat memengaruhi akurasi dan generalisasi model. Pemanfaatan *grid search* atau *elbow method* bisa menjadi pendekatan yang baik untuk eksplorasi parameter ini.

Ketiga, perlu dilakukan eksperimen dengan algoritma pembandingan seperti *Random Forest*, *Support Vector Machine (SVM)*, dan *Logistic Regression* untuk melihat sejauh mana performa K-NN dapat dikalahkan atau dilengkapi oleh model lain. Komparasi ini penting