

KLASIFIKASI TELEMARKETING MENGGUNAKAN NAÏVE BAYES CLASSIFICATION DAN WRAPPER SEQUENTIAL FEATURE SELECTION

TELEMARKETING CLASSIFICATION USING NAÏVE BAYES CLASSIFICATION AND WRAPPER SEQUENTIAL FEATURE SELECTION

Muhammad Fikri Eina¹, Yulison Herry Chrisnanto², Melina³

^{1,2,3}Program Studi Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani
mfikrieina20@if.unjani.ac.id

ABSTRACT

Telemarketing is an important marketing strategy in the banking industry, but its effectiveness requires a deeper understanding of the factors that influence customer decisions. Data analysis is key in understanding customer behavior towards bank products offered, with the main challenge being to process large datasets to identify the most relevant features. This research develops a classification model using Naïve Bayes Classification combined with the Wrapper Sequential Feature Selection (WSFS) feature selection method on the Bank Marketing dataset. The goal is to compare Naïve Bayes classification accuracy with and without WSFS feature selection. Identification of the most influential feature subsets in predicting customer decisions is applied with WSFS, followed by Naïve Bayes classification. The results showed that the application of WSFS significantly improved the accuracy of the Naïve Bayes classification model in predicting customer decisions to choose deposit products, with an accuracy of 77.88%. Additional evaluation using precision, recall, and F1-score metrics also showed consistent improvement after using WSFS. This research contributes to optimizing telemarketing strategies in the banking industry by providing a more efficient method of analyzing and predicting customer decisions. The implication is that it can improve the marketing of bank products.

Keywords: Bank, Classification, Feature Selection, Naïve Bayes, Telemarketing, WSFS.

ABSTRAK

*Telemarketing merupakan strategi pemasaran penting dalam industri perbankan, namun efektivitasnya memerlukan pemahaman mendalam tentang faktor-faktor yang mempengaruhi keputusan nasabah. Analisis data menjadi kunci dalam memahami perilaku nasabah terhadap produk bank yang ditawarkan dengan tantangan utama adalah mengolah *dataset* besar untuk mengidentifikasi fitur-fitur yang paling relevan. Penelitian ini mengembangkan model klasifikasi menggunakan *Naïve Bayes Classification* yang dikombinasikan dengan metode seleksi fitur *Wrapper Sequential Feature Selection* (WSFS) pada *dataset* Bank Marketing. Tujuannya adalah membandingkan akurasi klasifikasi *Naïve Bayes* dengan dan tanpa seleksi fitur WSFS. Identifikasi *subset* fitur paling berpengaruh dalam memprediksi keputusan nasabah diterapkan dengan WSFS, dilanjutkan dengan klasifikasi *Naïve Bayes*. Hasil penelitian menunjukkan bahwa penerapan WSFS signifikan meningkatkan akurasi model klasifikasi *Naïve Bayes* dalam memprediksi keputusan nasabah memilih produk deposito, dengan akurasi sebesar 77.88%. Evaluasi tambahan menggunakan metrik *precision*, *recall*, dan *F1-score* juga menunjukkan peningkatan yang konsisten setelah menggunakan WSFS. Penelitian ini berkontribusi dalam mengoptimalkan strategi *telemarketing* di industri perbankan dengan menyediakan metode yang lebih efisien dalam menganalisis dan memprediksi keputusan nasabah. Implikasinya dapat meningkatkan pemasaran produk bank.*

Kata Kunci: Bank, Klasifikasi, Naïve Bayes, Seleksi Fitur, Telemarketing, WSFS

PENDAHULUAN

Industri perbankan merupakan sektor yang sangat penting dalam perekonomian suatu negara. Sebagai institusi keuangan, bank menawarkan berbagai layanan dan produk, termasuk deposito, pinjaman, dan pemasaran (Cahya & Jannah, 2022). Dalam upaya mempromosikan produknya, bank memanfaatkan strategi *telemarketing*

sebagai alat telekomunikasi. Saat ini, cara *telemarketing* digunakan oleh perusahaan-perusahaan operator besar dalam mempromosikan produknya karena *telemarketing* cenderung mudah diterima oleh calon nasabah (Saputra, 2017).

Telemarketing adalah strategi pemasaran jarak jauh yang menggunakan teknologi informasi dalam pemasaran yang

terkoordinasi dan terorganisir, serta merupakan program komunikasi persuasif ke pasar sasaran yang menghubungkan pasar sasaran tersebut dengan produk yang ditawarkan (Damayanti & Sastika, 2021). Mempromosikan produk deposito merupakan salah satu contoh dari kegiatan *telemarketing* bank. Deposito adalah jenis simpanan di bank yang memungkinkan nasabah untuk menyetorkan dan menarik dana dalam jangka waktu tertentu (Muchlisin Riadi, 2021). Proses *telemarketing* bank melibatkan kontak langsung dengan calon nasabah melalui telepon, dimana staf menjelaskan berbagai produk deposito yang tersedia. Dalam proses ini, data-data calon nasabah dicatat dan dianalisis untuk keperluan masa depan. Salah satu aspek analisis tersebut adalah memprediksi keputusan calon nasabah untuk membuka deposito di bank tersebut (Leonardo, Pratama, & Chrisnatalis, 2020).

Dalam upaya meningkatkan efektivitas strategi pemasaran, bank memanfaatkan berbagai metode pengolahan data, seperti teknik *data mining*. Beberapa algoritma yang umum digunakan adalah *Naïve Bayes*, *Decision Tree*, *Support Vector Machine (SVM)* dan lain-lain. Dibandingkan dengan *Decision Tree* dan *SVM*, *Naïve Bayes* adalah salah satu algoritma klasifikasi dengan akurasi terbaik. Tingkat akurasi dari masing-masing model bisa berbeda-beda tergantung pada model pembelajaran (Syafina, Toker, Santoso, Haryono, & Kurniawan, 2023). Temuan dari salah satu penelitian menyatakan bahwa metode *Naïve Bayes* menunjukkan performa yang lebih rendah dibandingkan dengan metode lain yang diuji yaitu *Decision Tree* dan *Logistic Function* (Fikriah & Hayati, 2022). Temuan ini menegaskan pentingnya pemilihan metode yang tepat dalam analisis data untuk mendapatkan hasil yang akurat.

Analisis data dalam pemasaran perbankan, digunakan untuk memahami perilaku pelanggan dan membuat

keputusan pemasaran yang lebih baik. Namun, tantangan utama dalam analisis data adalah mengekstraksi fitur yang relevan dari *dataset* yang besar (Saeys, Inza, & Larrañaga, 2007). Seleksi fitur menjadi elemen krusial, khususnya ketika berhadapan dengan *dataset* yang memiliki jumlah variabel dan fitur yang besar. Proses ini dapat membantu dalam mengeliminasi variabel yang kurang relevan dan pada akhirnya dapat meningkatkan presisi dan efisiensi. Oleh karena itu, metode seleksi fitur seperti *Wrapper Sequential Feature Selection (WSFS)* menjadi penting. *WSFS* adalah teknik yang digunakan untuk memilih *subset* fitur yang paling relevan dari *dataset*. Teknik ini kemudian digabungkan dengan algoritma klasifikasi seperti *Naïve Bayes* untuk meningkatkan akurasi prediksi (Syafina et al., 2023).

Terdapat beberapa penelitian terdahulu yang telah mengkaji efektivitas kombinasi *Feature Selection* dan *Naïve Bayes* dalam berbagai kasus. Salah satunya adalah penelitian (Yusa, Coastera, & Yandika, 2022), yang menggunakan metode *WSFS* dalam memberikan rekomendasi fitur yang dapat digunakan untuk meningkatkan performa algoritma klasifikasi, dan diuji dengan menggunakan metode validasi *10-Fold Cross validation*. Diperoleh hasil bahwa kombinasi 6 fitur dari total 11 fitur pada *dataset* medis *Cardiovascular Disease (CVD)* dapat meningkatkan performa algoritma klasifikasi *Naïve Bayes* dengan nilai akurasi sebesar 81,575%. Nilai *Recall* sebesar 54,889%. Nilai *Precision* sebesar 87,787%. Jika dibandingkan dengan hanya menerapkan algoritma klasifikasi *Naïve Bayes*, nilai performa yang dihasilkan hanya 75,241% untuk performa akurasi, *recall* hanya 43,556% dan *Precision* yaitu 67,662%. Selanjutnya, penelitian (Rahmansyah, Dewi, Andini, Ningrum, & Suryana, 2018), yang melakukan analisis komparatif terhadap pengaruh *Feature Selection* pada algoritma *Naïve Bayes* dan *Support Vector Machine*. Penelitian ini

melibatkan penerapan beberapa algoritma *Machine Learning*. Hasil penelitiannya menunjukkan bahwa pengaruh terbaik dari *Feature Selection* ada pada algoritma *Naive Bayes*, dan menghasilkan peningkatan nilai *presisi* dan *recall* menjadi 0.5 dan 0.52 dari nilai sebelumnya yaitu 0.455 dan 0.434. Sebaliknya, untuk *Support Vector Machine*, hasil terbaik diperoleh ketika semua fitur digunakan tanpa ada yang dikurangi.

Berdasarkan uraian di atas, telah ada penelitian terdahulu yang mengkaji efektivitas berbagai algoritma klasifikasi dan metode seleksi fitur. Namun, masih terdapat kesenjangan dalam literatur mengenai kombinasi spesifik WSFS dengan *Naive Bayes* pada *dataset* Bank Marketing. Oleh karena itu, penelitian ini menggunakan pendekatan WSFS untuk seleksi fitur dan *Naive Bayes* untuk klasifikasi pada *dataset* Bank Marketing, yang merupakan kebaruan. Urgensinya adalah peningkatan akurasi dalam memprediksi keputusan nasabah. Tujuan penelitian ini adalah untuk membandingkan nilai akurasi klasifikasi *Naive Bayes* dengan dan tanpa seleksi fitur WSFS pada *dataset* Bank Marketing dapat meningkatkan akurasi prediksi terhadap keputusan nasabah untuk membuka deposito. Hasil penelitian diharapkan dapat memberikan kontribusi praktis bagi industri perbankan dalam mengoptimalkan strategi pemasaran, serta kontribusi teoritis dalam pengembangan data mining dan *Machine Learning*.

TINJAUAN PUSTAKA

Telemarketing

Telemarketing adalah strategi pemasaran yang memanfaatkan telepon untuk menjual produk secara langsung kepada konsumen dan pelanggan bisnis. Ini melibatkan penggunaan telepon dan pusat panggilan untuk menarik prospek, menjual produk, dan memberikan layanan kepada pelanggan (Liana & Suryawardani, 2018).

Data Mining

Data mining merupakan serangkaian metode dan prosedur yang dikembangkan dari beragam sumber data, mulai dari gudang data, basis data relasional, hingga *file* datar tanpa format (Kumar et al., 2021). Analisis prediktif dalam data *mining* memanfaatkan teknik-teknik statistik untuk membuat prediksi atau antisipasi terhadap ukuran statistik kepastian berdasarkan fakta yang ada (Sghir, Adadi, & Lahmer, 2023). *Data mining* adalah proses menemukan pola-pola yang tersembunyi yang terdiri dari pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari sekumpulan data. Pola-pola ini dapat ditemukan di database, data *warehouse*, atau media penyimpanan informasi lainnya (Kusrini & Emha Taufiq, 2009).

Naive Bayes Classification

Naive Bayes adalah kumpulan algoritma klasifikasi berdasarkan *Teorema Bayes* (Vaibhav Jayaswal, 2020). Bukan satu algoritma, tetapi keluarga algoritma di mana semuanya berbagi prinsip umum, yaitu setiap pasangan fitur yang diklasifikasikan tidak tergantung satu sama lain (Yang, 2018). Klasifikasi dapat digunakan sebagai alat umum dan dapat diterapkan ke berbagai domain klasifikasi (Yang, 2018). Rumus dasar dari *Naive Bayes* diberikan pada persamaan (1) (Vaibhav Jayaswal, 2020).

$$P(A|B) = \frac{P(A) P(B|A)}{P(B)} \quad (1)$$

Dimana

$P(A|B)$: kemungkinan kejadian A jika kejadian B benar

$P(B|A)$: kemungkinan kejadian B jika kejadian A benar

$P(A)$ dan $P(B)$: kemungkinan kejadian A dan B.

Wrapper Sequential Feature Selection

Metode *wrapper* untuk pemilihan fitur berbeda dengan metode filter, metode *wrapper* menggunakan kinerja algoritma

pengklasifikasi yang dipilih sebagai metrik untuk membantu pemilihan *subset* fitur terbaik (Remeseiro & Bolon-Canedo, 2019). Metode *wrapper* memberikan hasil yang lebih baik dibandingkan dengan metode filter, tetapi ketika jumlah fitur menjadi sangat besar, cenderung lebih mahal secara komputasi (Piao, Piao, Park, & Ryu, 2012).

Metode *wrapper* mengidentifikasi set fitur dengan performa terbaik untuk algoritma pengklasifikasi yang dipilih (Wah, Ibrahim, Hamid, Abdul-Rahman, & Fong, 2018). Hal ini menjadi keuntungan utama dari metode *wrapper*, dan telah terbukti menghasilkan kinerja prediksi yang lebih tinggi daripada yang dapat diperoleh dengan metode filter. Secara praktis, strategi pencarian dan algoritma pengklasifikasi apa pun dapat digabungkan untuk menghasilkan metode *wrapper*. Metode *wrapper* secara implisit mempertimbangkan ketergantungan fitur, termasuk interaksi dan redundansi, selama pemilihan subset terbaik. Namun, karena tingginya jumlah komputasi yang diperlukan untuk menghasilkan subset fitur dan mengevaluasinya, metode *wrapper* memiliki beban komputasi yang berat (relatif terhadap metode filter dan *embedded*) (Chandrashekar & Sahin, 2014). Berikut ini adalah beberapa metode yang umum digunakan dalam pemilihan fitur menggunakan pendekatan *wrapper*:

1) *Sequential Forward Selection*

Sequential Forward Selection (SFS) dimulai dengan himpunan kosong. Metode ini berkinerja baik ketika hanya melibatkan jumlah fitur yang kecil. Namun, kelemahan utama dari *sequential forward selection* adalah ketidakmampuannya untuk menghapus fitur yang menjadi tidak signifikan setelah penambahan fitur lainnya.

2) *Sequential Backward Elimination*

Sequential Backward Selection (SBS), juga disebut sebagai *Sequential Backward Elimination* (SBE), bekerja berlawanan dengan seleksi maju berurutan. Eliminasi mundur berurutan pada dasarnya

bekerja paling baik dengan banyak fitur dalam *dataset*.

3) *Sequential Floating Forward Selection and Backward Selection (SFFS and SFBS)*

Metode *SFFS* dan *SFBS* diperkenalkan oleh Pudil, Novovičová, dan Kittler, yang memungkinkan kontrol *backtracking* tanpa perlu mengatur parameter apa pun. Keunikan dari metode ini adalah fleksibilitasnya, di mana jumlah langkah maju dan mundur ditentukan secara dinamis, bukan sebelumnya. *SFFS* dan *SFBS* mungkin merupakan metode pemilihan fitur yang paling efektif. Metode-metode dikenal sebagai metode mengambang, memungkinkan penambahan dan penghapusan *subset* fitur sampai diperoleh jumlah *subset* fitur yang tepat (Wah, Ibrahim, Hamid, Abdul-Rahman, & Fong, 2018).

Hipotesis Penelitian

Berdasarkan tinjauan pustaka di atas, rumusan hipotesis dalam penelitian ini, yaitu

H0: Tidak ada perbedaan signifikan dalam akurasi klasifikasi Naïve Bayes dengan dan tanpa seleksi fitur WSFS pada *dataset* Bank Marketing.

H1: Terdapat peningkatan signifikan dalam akurasi klasifikasi pada model Naïve Bayes dengan dan tanpa seleksi fitur WSFS pada *dataset* Bank Marketing.

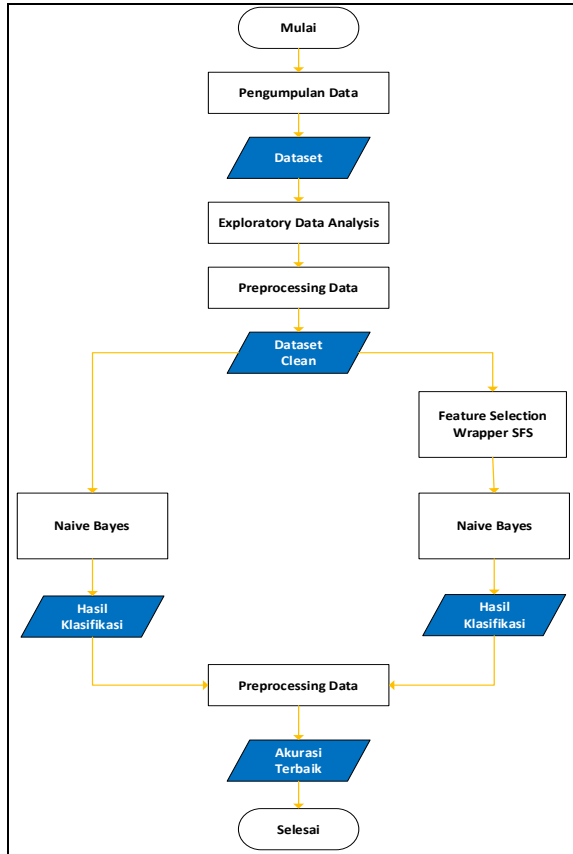
Beberapa penelitian terkait yaitu penelitian (Yusa et al., 2022), menunjukkan bahwa kombinasi WSFS dengan Naïve Bayes dapat meningkatkan akurasi klasifikasi. Pada *dataset* medis Cardiovascular Disease (CVD). Hasil penelitiannya memperoleh peningkatan akurasi dari 75,241% menjadi 81,575% setelah penerapan WSFS. Kemudian penelitian (Rahmansyah et al., 2018) menunjukkan adanya peningkatan presisi dan *recall* pada algoritma *Naive Bayes* setelah penerapan seleksi fitur.

Oleh karena itu, penelitian ini bertujuan membandingkan nilai akurasi

klasifikasi *Naïve Bayes* dengan dan tanpa seleksi fitur WSFS pada *dataset* Bank *Marketing* dengan menguji apakah kombinasi tersebut dapat menghasilkan peningkatan akurasi yang signifikan pada *dataset* Bank *Marketing*

METODE

Terdapat beberapa tahapan dalam penelitian ini, yang dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

Gambar 1 menjelaskan secara rinci langkah-langkah yang dilakukan dalam penelitian ini, dimulai dari pengumpulan *dataset* hingga evaluasi hasil, untuk menguji pengaruh penggunaan metode seleksi fitur *Wrapper SFS* terhadap akurasi klasifikasi *Naïve Bayes* dalam memprediksi keputusan calon nasabah membuka deposito pada bank.

Pengumpulan Data

Pengumpulan data merupakan langkah awal dalam penelitian ini. Data yang digunakan berasal dari *dataset* Bank

Marketing yang tersedia publik di *Kaggle*(Bargava, 2016), sebuah sumber terpercaya yang menyediakan *dataset* untuk penelitian ilmiah.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) digunakan untuk mengurangi dimensi data atau meningkatkan pemahaman analisis data melalui visualisasi data. Langkah ini mencakup deskripsi *dataset*, jenis data pada setiap fitur, nilai indeks, informasi *dataset*, dan korelasi antar data untuk mengevaluasi pengaruh fitur data lainnya.

Preprocessing Data

Preprocessing data dilakukan untuk memastikan tidak ada nilai yang hilang (Melina et al., 2022). Proses ini meliputi :

- Penanganan nilai yang hilang: semua fitur dalam *dataset* tidak memiliki nilai yang hilang.
- Konversi jenis data: Mengubah tipe data jika diperlukan.
- Pemeriksaan data duplikat: Memastikan tidak ada data duplikat dalam *dataset*.

Split Data

Dataset dibagi menjadi data latih dan data uji. Pembagian dilakukan dengan proporsi tertentu, misalnya 80% untuk data latih dan 20% untuk data uji, untuk pelatihan dan pengujian model.

Naive Bayes Tanpa Proses Feature Selection

Untuk eksperimen pertama, metode *Naive Bayes* diimplementasikan pada *dataset* bersih yang telah melalui proses *preprocessing* sebelumnya, tanpa melibatkan proses seleksi fitur. Selanjutnya, pengujian dilakukan dengan membagi *dataset* menjadi dua bagian: data latih dan data uji. Hasil dari pengujian ini adalah nilai akurasi atau skor lainnya yang ditentukan berdasarkan evaluasi model pada data uji.

Naive Bayes dengan Sequential Feature Selection

Untuk eksperimen kedua, dilakukan seleksi fitur menggunakan metode *wrapper* pada *dataset* yang telah dibersihkan. Dalam proses ini, menggunakan metode *Sequential Feature Selector* (SFS), yang dimulai dengan mengidentifikasi fitur-fitur terbaik. Selanjutnya, metode ini menggabungkan fitur-fitur yang telah diidentifikasi sebelumnya dengan fitur-fitur lain yang memiliki performa lebih baik daripada yang lain. Proses ini berulang hingga berbagai kombinasi fitur terbentuk. Hasil dari setiap kombinasi fitur yang diperoleh melalui SFS kemudian digunakan dalam tahap klasifikasi dengan metode *Naive Bayes*. Akhirnya, model dievaluasi menggunakan *K-Fold Cross Validation* dengan nilai k yang telah ditentukan, yaitu $k = 10$.

Evaluasi

Setelah menghitung akurasi, langkah selanjutnya adalah mengevaluasi kinerja algoritma yang digunakan. Evaluasi ini melibatkan penggunaan *Confusion Matrix* dan menampilkan nilai-nilai pada *Classification Report*, seperti *Precision*, *Recall*, dan *F1-Score*. Setelah semua tahapan eksperimen selesai, selanjutnya adalah menentukan hasil pengujian mana yang memiliki akurasi terbaik: apakah eksperimen *Naive Bayes* tanpa menggunakan *Feature Selection Wrapper* atau dengan menggunakan *Wrapper Sequential Feature Selector* (WSFS). Dalam penelitian ini, dilakukan analisis untuk menentukan fitur-fitur mana saja yang memiliki korelasi tinggi dan berpengaruh terhadap label yang ada pada *dataset* serta menghasilkan akurasi tertinggi dari klasifikasi *Naive Bayes* yang telah divalidasi.

HASIL DAN PEMBAHASAN

Berikut ini adalah hasil penelitian berdasarkan tahapan dalam metode penelitian ini, termasuk *dataset* yang

diperoleh, hasil *pre-processing dataset*, hasil seleksi fitur menggunakan *Wrapper Sequential Feature Selection*, serta perbandingan antara dua pendekatan dalam klasifikasi menggunakan algoritma *Naive Bayes* pada *dataset* Bank Marketing.

Dataset

Dalam penelitian ini menggunakan *dataset* dari *Kaggle* terkait dengan kampanye pemasaran bank. Berikut adalah atribut/fitur serta deskripsi yang digunakan pada *dataset* tersebut dapat dilihat pada Tabel 1.

Tabel 1. Dataset

No.	Atribut/Fitur	Deskripsi
1	age	Usia
2	job	Jenis pekerjaan
3	marital	Status pernikahan
4	education	Tingkat pendidikan: 'primary', 'secondary', 'tertiary', 'unknown'
5	default	Memiliki kredit default?
6	balance	Rata-rata saldo tahunan
7	housing	Memiliki pinjaman rumah?
8	loan	Memiliki pinjaman pribadi?
9	contact	Jenis komunikasi kontak
10	day_of_week	Hari kontak terakhir dalam minggu
11	month	Bulan kontak terakhir
12	duration	Durasi kontak terakhir, dalam detik.
13	campaign	Jumlah kontak yang dilakukan selama kampanye.
14	pdays	Jumlah hari yang berlalu sejak klien terakhir dihubungi dari kampanye sebelumnya.
15	previous	Jumlah kontak yang dilakukan sebelum kampanye ini.
16	poutcome	Hasil dari kampanye pemasaran sebelumnya: 'failure', 'nonexistent', 'success'
17	deposit	Label apakah melakukan deposito?

Preprocessing

Data awal diproses untuk menangani nilai yang hilang, mengkodekan variabel kategorikal, dan normalisasi fitur numerik. Transformasi dilakukan untuk mengubah data kategorikal menjadi format numerik agar dapat digunakan oleh *library machine learning* yang digunakan.

Split Data

Pemisahan data dilakukan untuk memastikan bahwa model dapat dilatih dengan data yang cukup serta diuji dengan data yang representatif. *Dataset* dibagi menjadi data latih (80%) dan data uji (20%) untuk melatih dan menguji model. Setelah pemisahan, terdapat 8,929 data untuk latihan dan 2,233 data untuk pengujian, seperti yang ditunjukkan pada Tabel 2.

Tabel 2. Hasil Split Data

Data latih	Data uji
8929	2233

Naive Bayes tanpa SFS

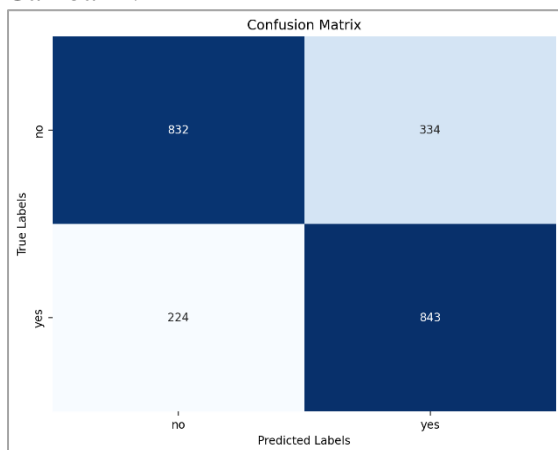
Pada percobaan pertama, *Naive Bayes* diterapkan tanpa *Sequential Feature Selection* (SFS). Hasil klasifikasi ditunjukkan pada Tabel 3.

Tabel 3. Hasil Klasifikasi Naive Bayes (Tanpa SFS)

Metrik	Kelas 0	Kelas 1	Rata-rata
Precision	0.79	0.72	0.75
Recall	0.71	0.79	0.75
F1-Score	0.75	0.75	0.75

Berdasarkan Tabel 3 hasil dari percobaan ini menunjukkan bahwa akurasi klasifikasi menggunakan *Naive Bayes* sebesar 75.01%, dengan rata-rata nilai *precision*, *recall*, dan *f1-score* berturut-turut adalah 75 %, 75%, dan 75%.

Adapun *Confusion Matrix* untuk *Naive Bayes* tanpa SFS ditampilkan pada Gambar 2.



Gambar 1. Confusion Matrix Naive Bayes

Naive Bayes dengan SFS

Pada percobaan kedua, dilakukan seleksi fitur menggunakan SFS sebelum menerapkan *Naive Bayes*. Proses SFS

bertujuan untuk mencari kombinasi fitur terbaik dengan skor tertinggi berdasarkan akurasi, menggunakan validasi silang 10-fold.

Tabel 4. Sequential Feature Selection (SFS)

Iterasi	Fitur index	Skor
1.	11	0.691416
2.	0, 11	0.709429
3.	0, 11, 12	0.721981
4.	0, 6, 11, 12	0.731664
5.	0, 6, 11, 12, 14	0.740355
6.	0, 6, 7, 11, 12, 14	0.747889
7.	0, 2, 6, 7, 11, 12, 14	0.752190
8.	0, 2, 3, 6, 7, 11, 12, 14	0.756133
9.	0, 1, 2, 3, 6, 7, 11, 12, 14	0.756582
10.	0, 1, 2, 3, 6, 7, 9, 11, 12, 14	0.754880
11.	0, 1, 2, 3, 6, 7, 9, 10, 11, 12, 14	0.751655
12.	0, 1, 2, 3, 6, 7, 9, 10, 11, 12, 14, 15	0.750314
13.	0, 1, 2, 3, 5, 6, 7, 9, 10, 11, 12, 14, 15	0.744847
14.	0, 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15	0.747000
15.	0, 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15	0.738579
16.	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15	0.733130

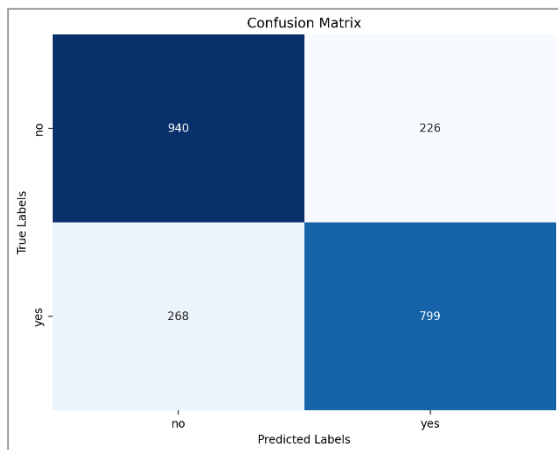
Pada Tabel 4, SFS dilakukan untuk memilih kombinasi fitur terbaik berdasarkan nilai skor pada setiap iterasi. Kombinasi fitur yang dipilih pada iterasi ke-9 adalah *age*, *job*, *marital*, *education*, *housing*, *loan*, *duration*, *campaign*, dan *previous* dengan skor 0.757, yang kemudian digunakan untuk klasifikasi dengan *Naive Bayes*, yang hasilnya ditunjukkan pada Tabel 5.

Tabel 5. Hasil Klasifikasi Naive Bayes dengan SFS

Metrik	Kelas 0	Kelas 1	Rata-rata
Precision	0.78	0.78	0.78
Recall	0.81	0.75	0.78
F1-Score	0.79	0.76	0.78

Tabel 5 menunjukkan hasil klasifikasi *Naive Bayes*. Setelah menerapkan SFS, terjadi peningkatan signifikan dalam akurasi menjadi 77.88%, dengan rata-rata nilai *precision* 78%, *recall* 78%, dan *f1-score* 78%.

Adapun hasil pengukuran kinerja menggunakan *Confusion Matrix* ditampilkan pada Gambar 3.



Gambar 2. Confusion Matrix SFS+NB

Evaluasi

Penerapan SFS menghasilkan peningkatan signifikan dalam kinerja model *Naive Bayes*, dengan akurasi meningkat menjadi 77.88%. Seleksi fitur ini membantu mengidentifikasi kombinasi fitur yang paling relevan untuk memprediksi keputusan calon nasabah dalam membuka deposito, meningkatkan *precision*, *recall*, dan *F1-score* secara keseluruhan

SIMPULAN

Penelitian ini menunjukkan bahwa penggunaan SFS secara signifikan meningkatkan akurasi model klasifikasi *Naive Bayes* dalam memprediksi keputusan calon nasabah untuk membuka deposito di bank. SFS memungkinkan model untuk fokus pada informasi yang paling penting dalam proses pengambilan keputusan dengan mengidentifikasi kombinasi fitur yang paling relevan dari *dataset* *Bank Marketing*. Hasil eksperimen menunjukkan bahwa *Naive Bayes* setelah penerapan SFS mencapai akurasi sebesar 77.88%, lebih tinggi dibandingkan dengan penggunaan *Naive Bayes* tanpa SFS yang hanya mencapai 75.01%. Selain itu, *precision*, *recall*, dan *f1-score* juga mengalami peningkatan yang signifikan setelah menggunakan SFS, menggambarkan kemampuan model dalam memprediksi dengan lebih akurat. Implikasi penelitian ini adalah bahwa teknik seleksi fitur seperti SFS menjadi

strategi efektif dalam meningkatkan efisiensi dan efektivitas strategi pemasaran perbankan, membantu bank untuk lebih tepat sasaran dalam menjangkau dan menarik calon nasabah potensial.

DAFTAR PUSTAKA

- Anusuya, R., & Chinnadorai, K. M. (2015). A study on consumer awareness and satisfaction towards handloom products with special reference to Coimbatore district. *International Journal in Management & Social Science*, 3(9), 296–300.
- Bargava. (2016). Bank Marketing. Retrieved June 2, 2024, from Kaggle website: <https://www.kaggle.com/datasets/rouseguy/bankbalanced>
- Cahya, A. D., & Jannah, N. (2022). Analisis Strategi Pemasaran Meningkatkan Jumlah Nasabah Pada Produk Deposito iB Ibadah di PT Bank Sumut Syariah KCPSy HM Yamin. *JIKEM: Jurnal Ilmu Komputer, Ekonomi Dan Manajemen*, 2(1), 1311–1324.
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16–28.
- Damayanti, M. C., & Sastika, W. (2021). Analisis Telemarketing (outbound Call) Produk Indihome Di Witel Bandung Tahun 2021. *EProceedings of Applied Science*, 7(6).
- Fikriah, F. K., & Hayati, N. (2022). Feature Selection dengan Komparasi Algoritma untuk Prediksi Telemarketing Bank. *Jurnal Ilmu Komputer*. Retrieved from <https://archive.ics.uci.edu/ml/dataset-s/Bank+Marketing>
- Kumar, A., Kumar, A., Bashir, A. K., Rashid, M., Kumar, V. D. A., & Kharel, R. (2021). Distance based pattern driven mining for outlier detection in high dimensional big dataset. *ACM Transactions on*

- Management Information System (TMIS)*, 13(1), 1–17.
- Kusrini, & Emha Taufiq. (2009). *Algoritma Data Mining*.
- Leonardo, R., Pratama, J., & Chrisnatalis, C. (2020). Perbandingan Metode Random Forest Dan Naïve Bayes Dalam Prediksi Keberhasilan Klien Telemarketing. *Jurnal Teknologi Dan Ilmu Komputer Prima (Jutikomp)*, 3(2), 455–459.
- Liana, S., & Suryawardani, B. (2018). Pengaruh Telemarketing Terhadap Keputusan Pembelian. *EProceedings of Applied Science*, 4(3).
- Melina, Sukono, Napitupulu, H., Sambas, A., Murniati, A., & Kusumaningtyas, V. A. (2022). Artificial Neural Network-Based Machine Learning Approach to Stock Market Prediction Model on the Indonesia Stock Exchange During the COVID-19. *Engineering Letters*, 30(3).
- Muchlisin Riadi. (2021, December 7). Deposito - Pengertian, Karakteristik, Jenis, Manfaat dan Perhitungan Bunga. Retrieved May 7, 2024, from KAJIANPUSTAKA website: <https://www.kajianpustaka.com/2021/12/deposito.html>
- Piao, Y., Piao, M., Park, K., & Ryu, K. H. (2012). An ensemble correlation-based gene selection algorithm for cancer classification with gene expression data. *Bioinformatics*, 28(24), 3306–3315. <https://doi.org/10.1093/bioinformatics/bts602>
- Rahmansyah, A., Dewi, O., Andini, P., Ningrum, T. H. P., & Suryana, M. E. (2018). Membandingkan Pengaruh Feature Selection Terhadap Algoritma Naïve Bayes dan Support Vector Machine. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*.
- Remeseiro, B., & Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in Biology and Medicine*, 112, 103375.
- Saeyes, Y., Inza, I., & Larrañaga, P. (2007, October 1). A review of feature selection techniques in bioinformatics. *Bioinformatics*, Vol. 23, pp. 2507–2517. Oxford University Press. <https://doi.org/10.1093/bioinformatics/btm344>
- Saputra, E. P. (2017). Prediksi Keberhasilan Telemarketing Bank Untuk Mencari Algoritma Dengan Performa Terbaik. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 2(2), 66–72.
- Sghir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). *Education and Information Technologies*, 28(7), 8299–8333.
- Syafina, N. T., Toker, E. Y., Santoso, A. B., Haryono, C. A. D., & Kurniawan, M. (2023). Implementasi Metode Wrapper Sequential Feature Selection (WSFS) pada Dataset Stroke Menggunakan Metode Naïve Bayes Multinomial. *Prosiding Seminar Nasional Teknik Elektro, Sistem Informasi, Dan Teknik Informatika (SNESTIK)*, 1(1), 340–346.
- Vaibhav Jayaswal. (2020). Understanding Naïve Bayes algorithm. Retrieved November 30, 2023, from Towards Data Science website: <https://towardsdatascience.com/understanding-na%C3%AFve-bayes-algorithm-f9816f6f74c0>
- Wah, Y. B., Ibrahim, N., Hamid, H. A., Abdul-Rahman, S., & Fong, S. (2018). Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy. *Pertanika Journal of Science & Technology*, 26(1).
- Wah, Y. B., Ibrahim, N., Hamid, H. A., Abdul-Rahman, S., & Fong, S.

- (2018). Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy. *Pertanika Journal of Science & Technology*, 26(1).
- Yang, F.-J. (2018). An implementation of naive bayes classifier. *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, 301–306. IEEE.
- Yusa, M., Coastera, F. F., & Yandika, M. R. (2022). Reduksi Dimensi Data menggunakan Metode Wrapper Sequential Feature Selection untuk Peningkatan Performa Algoritma Naïve Bayes terhadap Dataset Medis. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 8(2), 364–370.