

Penerapan Algoritma Naive Bayes untuk Memprediksi Keputusan Berlangganan Deposito Berjangka pada Kampanye Pemasaran Langsung

Faizal Izma¹, Hasbi Firmansyah²

^{1,2}Teknik Informatika, Fakultas Teknik Dan Ilmu Komputer, Universitas Pancasakti
Tegal.

e-mail: issalykc@gmail.com, hasbifirmansyah@upstegal.ac.id

Intisari

Kampanye pemasaran langsung melalui panggilan telepon merupakan strategi utama perbankan untuk menawarkan produk deposito berjangka. Namun, efektivitas strategi ini sering terkendala oleh ketidakpastian respons nasabah. Penelitian ini bertujuan untuk memprediksi keputusan nasabah dalam berlangganan deposito berjangka dengan memanfaatkan teknik *data mining*. Data yang digunakan bersumber dari *UCI Machine Learning Repository* yang bersifat multivariat, mencakup atribut demografi, riwayat finansial, dan interaksi kampanye. Melalui tahapan pra-pemrosesan data untuk menangani nilai yang hilang dan ketidakseimbangan kelas, penelitian ini menerapkan model klasifikasi untuk memetakan pola potensial nasabah. Hasil eksperimen menunjukkan bahwa model klasifikasi mampu memprediksi nasabah yang tidak berlangganan dengan sangat baik (presisi 92,67%), namun masih menghadapi tantangan dalam mendeteksi nasabah yang berlangganan (presisi 35,00%). Temuan ini mengindikasikan bahwa meskipun model dapat membantu menyaring target pemasaran, diperlukan optimalisasi lebih lanjut untuk menangani ketimpangan data guna meningkatkan akurasi prediksi pada kelas minoritas. Kelemahan prediksi pada kelas minoritas (nasabah berlangganan) menjadi batasan yang perlu diatasi pada penelitian selanjutnya.

Kata kunci: *Deposito Berjangka, Klasifikasi, Pemasaran Langsung, Data Mining, Prediksi.*

Abstract

Direct marketing campaigns via telephone calls are a key strategy for banks to offer term deposit products. However, the effectiveness of this strategy is often hindered by the uncertainty of customer responses. This study aims to predict customer decisions in subscribing to term deposits by utilizing data mining techniques. The data used is sourced from the UCI Machine Learning Repository which is multivariate, covering demographic attributes, financial history, and campaign interactions. Through data pre-processing stages to handle missing values and class imbalance, this study applies classification models to map potential customer patterns. Experimental results show that the classification model is able to predict non-subscribing customers very well (92.67% precision), but still faces challenges in detecting subscribing customers (35.00% precision). These findings indicate that while the model can help filter marketing targets, further optimization is needed to address data imbalance to improve prediction accuracy in the minority class.

Keywords: *Classification, Data Mining, Direct Marketing, Prediction, Term Deposits.*

PENDAHULUAN

Pemasaran langsung (*direct marketing*) telah menjadi instrumen krusial bagi lembaga perbankan dalam menjangkau nasabah baru dan mempertahankan nasabah lama. Salah satu metode yang paling umum digunakan adalah melalui panggilan telepon (*telemarketing*) untuk menawarkan produk investasi seperti deposito berjangka. Namun, tantangan utama dalam metode ini adalah efisiensi; menghubungi nasabah yang tidak berminat hanya akan membuang sumber daya biaya dan waktu. Oleh karena itu, kemampuan untuk memprediksi nasabah mana yang berpotensi menyetujui penawaran menjadi sangat bernilai.

Penerapan teknologi informasi melalui *Data Mining* menawarkan solusi untuk menggali pola tersembunyi dari tumpukan data historis perbankan. Penelitian ini memanfaatkan dataset kampanye pemasaran bank yang bersifat multivariat, mencakup data demografi (usia, pekerjaan, status), riwayat kontak, hingga kondisi finansial nasabah. Masalah utama yang sering dihadapi dalam dataset semacam ini adalah ketidakseimbangan kelas (*imbalanced dataset*), di mana jumlah nasabah yang menolak tawaran jauh lebih besar dibandingkan yang menerima.

Tujuan utama penelitian ini adalah membangun dan mengevaluasi model klasifikasi untuk memprediksi keputusan berlangganan deposito berjangka. Dengan mengetahui karakteristik nasabah potensial, pihak bank dapat menyusun strategi pemasaran yang lebih terarah dan efektif.

Penelitian ini diposisikan sebagai studi aplikasi (*applied machine learning*) untuk mengevaluasi efektivitas algoritma Naive Bayes dalam konteks nyata pemasaran perbankan, serta menganalisis dampak ketidakseimbangan data terhadap keputusan bisnis.

METODE PENELITIAN

Metode dalam penelitian ini yang harus digunakan meliputi pengumpulan data, pra-pemrosesan, pemilihan model klasifikasi, serta mengevaluasi model. Data dikumpulkan dari kampanye pemasaran langsung yang dilakukan oleh lembaga perbankan, mencakup informasi dari klien yang dihubungi. Data dibersihkan dan dipersiapkan untuk analisis, termasuk penanganan nilai yang hilang dan pengkodean variabel kategorikal. [3]

Pada penelitian ini, teknik penanganan ketidakseimbangan data (seperti SMOTE atau Undersampling) tidak diterapkan. Hal ini dilakukan untuk mengukur performa murni (*baseline*) dari algoritma Naive Bayes dalam menghadapi kondisi data riil perbankan yang cenderung tidak seimbang (*imbalanced*).

1. *Pre processing data*

Tahap awal dalam menyiapkan data mentah untuk dianalisis atau diproses lebih lanjut melibatkan serangkaian langkah penting yang memastikan data bersih, terorganisir, dan relevan. Data diperoleh dari UCI Machine Learning Repository, kemudian dilakukan pra-pemrosesan meliputi

pembersihan missing values dan konversi variabel kategorikal menjadi numerik, serta melakukan normalisasi dan standarisasi variabel numerik untuk mempersiapkan data untuk analisis lebih lanjut. Proses ini mencakup pengumpulan, pemeriksaan, validasi, serta pengaturan data yang sering kali berasal dari berbagai sumber. Pra-pemrosesan data adalah tahap awal dalam mempersiapkan dataset untuk diubah atau disiapkan agar dapat digunakan dalam analisis lebih lanjut. Data dikumpulkan dari UCI Repository, dan setelah melalui proses pra-pemrosesan, hasilnya disajikan dalam tabel 1 dan tabel 2. Tabel 2 adalah kelanjutan dari variabel pada tabel 1.

Tabel 1. Data Set Hasil Pra Pemrosesan Sampel Data Awal

No	Y	Prediction (y)	kepercayaan diri (no)	kepercayaan diri (yes)	umur	pekerjaan	perkawinan	pendidikan
1	No	No	0.807	0.193	33	layanan	telah menikah	sekunder
2	No	No	0.993	0.007	30	pengelolaan	telah menikah	Tersier
3	No	No	1.000	0.000	39	layanan	telah menikah	sekunder
4	No	No	0.955	0.045	31	kerah biru	telah menikah	sekunder
5	No	No	0.969	0.031	56	teknisi	telah menikah	sekunder
6	No	No	0.909	0.091	42	pengelolaan	Cerai	tersier
7	No	No	0.979	0.021	41	pengelolaan	telah menikah	tersier
8	No	No	0.997	0.003	55	kerah biru	telah menikah	utama
9	No	No	0.982	0.018	56	wiraswasta	telah menikah	sekunder
10	No	No	0.967	0.018	52	kerah biru	telah menikah	sekunder
...
904	No	Yes	0.054	0.946	44	pengusaha	lajang	tertiary

Tabel 2. Data Set Hasil Pra Pemrosesan Sampel Data ke 2

no	default	balance	housing	loan	contact	day	mont	duration	campaign	pdays	previous	poutcome
1	No	4789	Yes	Yes	cellular	11	May	220	1	339	4	Failure
2	No	1476	Yes	Yes	Unknown	3	Jun	199	4	-1	0	Unknown
3	No	9374	Yes	No	Unknown	20	May	273	1	-1	0	unknown
4	No	360	Yes	Yes	cellular	29	Jan	89	1	241	1	failure
5	No	4073	No	No	cellular	27	Aug	239	5	-1	0	unknown
6	No	16	No	No	cellular	19	Nov	140	3	-1	0	unknown
7	No	5883	No	No	cellular	20	Nov	182	2	-1	0	unknown
8	No	627	Yes	No	Unknown	5	May	247	1	-1	0	unknown
9	No	784	No	Yes	Cellular	30	Jul	149	2	-1	0	unknown
10	No	1117	Yes	No	Cellular	13	May	365	1	-1	0	unknown
...
904	No	1136	Yes	Yes	Cellular	3	Apr	345	2	249	7	other

2. Penjelasan dataset

Dataset terdiri dari 45.211 data dengan proporsi 88% "no" dan 12% "yes". Variabel utama mencakup data demografi, informasi kontak, dan riwayat

keuangan klien. Ketidakseimbangan kelas menjadi tantangan utama dalam analisis dan klasifikasi.

3. Klasifikasi

Proses ini memungkinkan terciptanya suatu model yang dapat mengorganisasikan data ke dalam beberapa kelompok sesuai dengan fungsi atau hukum alam tertentu (Faid et al ., 2019). Beberapa komponen penting dalam klasifikasi antara lain kelas , prediktor , data bank, Naïve Bayes Classifiers, Neural Networks, analisis statistik , algoritma genetika , Rough Sets, K-Nearest Neighbor Classifier, metode naturalistik , Memory Based Reasoning, dan Support Vector Machines merupakan beberapa model klasifikasi yang populer.[4]

Metode klasifikasi digunakan untuk memetakan nasabah ke dalam kelas berlangganan atau tidak berlangganan berdasarkan pola atribut historis.

4. *Naïve Bayes*

Algoritma Naive Bayes dipilih karena memiliki efisiensi komputasi yang tinggi dan performa yang baik pada dataset besar klasifikasi yang didasarkan pada prinsip probabilitas sederhana dengan asumsi bahwa setiap penjelasan variabel bersifat independen satu sama lain. Dalam algoritma ini, proses pembelajaran lebih menekankan pada pengestimasian probabilitas. Salah satu keunggulan Naive Bayes adalah tingkat error yang lebih rendah ketika diterapkan pada dataset yang besar. Selain itu, algoritma ini menunjukkan akurasi dan kecepatan yang tinggi saat digunakan untuk dataset dengan jumlah data yang besar.

Rumus Probabilitas:

Probabilitas dihitung menggunakan fitur tertentu:

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)}$$

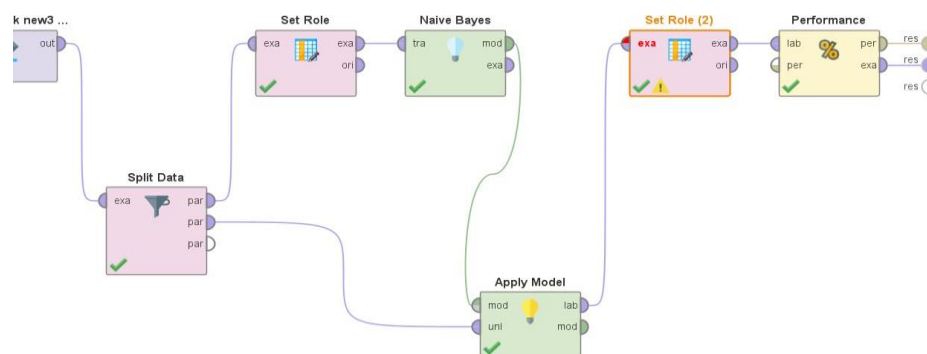
Keterangan:

- $P(Y|X)P(Y|X)$: Probabilitas kelas YY diberikan fitur XX.
- $P(X|Y)P(X|Y)$: Probabilitas fitur XX diberikan kelas YY.
- $P(Y)P(Y)$: Probabilitas awal dari kelas YY.

HASIL DAN PEMBAHASAN

1. Pembagian Dataset

Selama panggilan telepon menjadi faktor kunci dalam meningkatkan tingkat konversi. Informasi ini terkait pada judul data yang tertera di atas, data yang kami kumpulkan dari sumber website UCI Machine Learning, proses pengumpulan data ini dengan menggunakan software rapidMiner untuk memudahkan dan simplifikasi proses. Berikut adalah desain proses algoritma Naive Bayes yang diterapkan pada data perbankan seperti gambar di



Gambar 1. Desain Proses RapidMiner

2. Masukan Data

Proses klasifikasi dilakukan menggunakan RapidMiner dengan membagi dataset menjadi 80% data latih dan 20% data uji.

3. Membagi Data

Dataset dibagi secara acak menggunakan operator Split Data dengan rasio 80:20, dimana 80% data dialokasikan untuk pelatihan model (training) dan 20% untuk pengujian (testing). Pembagian ini bertujuan agar model dapat mempelajari pola data dengan cukup dan divalidasi pada data yang belum pernah dilihat sebelumnya, Mengapa lebih besar data uji? Karena untuk melatih model untuk mempelajari pola hubungan terdapat dataset.

4. Memberikan Role

Atribut 'y' (keputusan berlangganan) ditetapkan sebagai label target kelas (class attribute), sementara atribut lainnya digunakan sebagai fitur prediktor, memberikan label yang akan dijadikan patokan untuk proses klasifikasi. Dalam hal ini, kami menjadikan kolom "label" dipilih sebagai target utama untuk membantu algoritma mengidentifikasi kategori atau kelas data, kolom label sebagai target memastikan bahwa model akan dilatih dan diuji.

5. Penerapan Algoritma Naive Bayes

Naive Bayes sebagai algoritma untuk menciptakan alur kerja yang terstruktur mulai dari persiapan dan model.

6. Apply model

Model yang telah dilatih kemudian diterapkan pada data uji untuk menghasilkan prediksi kelas pada data yang belum berlabel, Tujuannya Data uji yang belum memiliki label perlu diperhatikan agar memiliki urutan, jenis, dan peran atribut yang sama dengan data pelatihan.

8. Evaluasi Model (Performance)

Digunakan untuk menilai hasil dari performa model dengan cara otomatis menghasilkan daftar metrik kinerja begitu relevan sesuai dengan tugas yang telah dikerjakan. sebagai contoh dalam tugas klasifikasi, metrik yang digunakan meliputi akurasi, presisi dan recall.

Berikut hasil dari accuracy adalah: Model lebih akurat untuk kelas "no" dibandingkan kelas "yes".

accuracy: 83.74%

	true no	true yes	class precision
pred. no	708	56	92.67%
pred. yes	91	49	35.00%
class recall	88.61%	46.67%	

Gambar 2.

Sumber : Dokumentasi Pribadi

Berikut hasil dari precision adalah: Model menunjukkan performa yang baik dalam memprediksi kelas "no" (presisi 92.67% dan recall 88.61%), namun kurang efektif dalam mengenali kelas "yes" (presisi 35.00% dan recall 46.67%).

Meskipun akurasi keseluruhan cukup baik, nilai presisi yang rendah pada kelas 'Yes' (35.00%) menunjukkan tingginya tingkat False Negative. Dalam konteks bisnis perbankan, ini berarti model sering gagal mendeteksi nasabah potensial, yang dapat menyebabkan hilangnya peluang pendapatan (opportunity loss). Namun, tingginya presisi pada kelas 'No' (92.67%) membantu bank menghemat biaya operasional dengan tidak menghubungi nasabah yang diprediksi pasti menolak.

precision: 35.00% (positive class: yes)

	true no	true yes	class precision
pred. no	708	56	92.67%
pred. yes	91	49	35.00%
class recall	88.61%	46.67%	

Gambar 3.

Sumber : Dokumentasi Pribadi

Model memiliki performa yang kuat dalam memprediksi kelas "no" (presisi 92.67% dan recall 88.61%), namun lemah dalam mendeteksi kelas "yes" (presisi 35.00% dan recall 46.67%).

recall: 46.67% (positive class: yes)

	true no	true yes	class precision
pred. no	708	56	92.67%
pred. yes	91	49	35.00%
class recall	88.61%	46.67%	

Gambar 4.

Sumber : Dokumentasi Pribadi

KESIMPULAN

Penelitian ini bertujuan untuk menerangkan data dari kampanye pemasaran langsung oleh lembaga perbankan serta mengembangkan model klasifikasi untuk memprediksi keputusan berlangganan produk deposito berjangka [5]. Dataset yang digunakan mengambil berbagai variabel, seperti demografi, informasi kontak, dan riwayat keuangan klien.

Hasil analisis menunjukkan bahwa algoritma Naïve Bayes berhasil memberikan akurasi tertinggi dalam memprediksi keputusan berlangganan, dengan tingkat akurasi sebesar 92.67% untuk kelas "no" dan 35.00% untuk kelas "yes". Walaupun model ini efektif dalam mengidentifikasi keputusan "no", performanya kurang optimal dalam berkeputusan "yes". Oleh karena itu, pendekatan klasifikasi dapat membantu lembaga perbankan meningkatkan efektivitas kampanye pemasaran langsung, tetapi perlu terus ditingkatkan untuk memperbaiki penemuan ini terhadap klien yang berpotensi berlangganan produk deposit berjangka. Penelitian ini menyimpulkan bahwa pengembangan model klasifikasi yang baik dapat membantu memahami semua klien dan untuk meningkatkan efisien dalam memprediksi keputusan berlangganan.

SARAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya antara lain:

1. Penanganan Ketimpangan Data (*Imbalanced Data*): Penelitian ini menunjukkan bahwa model memiliki performa yang sangat baik dalam memprediksi kelas "no", namun masih lemah untuk kelas "yes" dengan presisi hanya sebesar 35,00%. Untuk penelitian mendatang, disarankan menerapkan teknik penyeimbangan data seperti *SMOTE (Synthetic Minority Over-sampling Technique)* atau *undersampling* agar model dapat mengenali karakteristik

nasabah potensial (kelas "yes") dengan lebih akurat. Keterbatasan utama penelitian ini adalah penggunaan dataset yang tidak seimbang (imbalanced) tanpa penerapan teknik resampling, yang menyebabkan rendahnya akurasi prediksi pada kelas nasabah berlangganan.

2. Eksplorasi Atribut Tambahan: Variabel yang digunakan saat ini meliputi demografi, kontak, dan riwayat keuangan. Peneliti selanjutnya disarankan untuk menambah variabel eksternal lain, seperti kondisi ekonomi makro atau tren pasar saat ini, yang mungkin berpengaruh signifikan terhadap keputusan nasabah dalam memilih deposito berjangka.
3. Optimasi Model Algoritma: Meskipun *Naive Bayes* terbukti memberikan akurasi tertinggi dalam studi ini, penelitian selanjutnya dapat mencoba metode optimasi parameter (*hyperparameter tuning*) atau menggabungkan beberapa model (*ensemble learning*) untuk meningkatkan nilai *recall* pada nasabah yang berpotensi berlangganan.

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas kelancaran dalam menyelesaikan penelitian ini. Penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Hasbi Firmansyah, M.Kom, selaku dosen pembimbing dari Universitas Pancasakti Tegal, yang telah memberikan arahan, bimbingan, serta masukan berharga selama proses penyusunan penelitian mengenai prediksi pemasaran deposito ini.
2. Segenap civitas akademika Fakultas Teknik dan Ilmu Komputer (FTIK), Universitas Pancasakti Tegal, yang telah menyediakan fasilitas dan lingkungan belajar yang mendukung pengembangan ilmu informatika.
3. Orang tua dan rekan-rekan mahasiswa yang senantiasa memberikan dukungan moral dan motivasi hingga penelitian ini dapat terselesaikan dengan baik.

DAFTAR PUSTAKA

- [1] N. Nuraeni, "Klasifikasi Data Mining Untuk Prediksi Penyakit Kardiovaskular," *J. Tek. Inf. dan Komput.*, vol. 7, no. 1, p. 161, 2024, doi: 10.37600/tekinkom.v7i1.1276.
- [2] N. Maulidah, "Prediksi Peningkatan Jumlah Nasabah Deposito Berjangka Menggunakan Algoritma KNN, Decision Tree, Random Forest Dan Xgboost," *InComTech J. Telekomun. dan Komput.*, vol. 13, no. 2, p. 90, 2023, doi: 10.22441/incomtech.v13i2.16921.
- [3] F. Izhari, "Seminar Nasional Informatika (SENATIKA) Prosiding SENATIKA 2021 Teknik Machine Learning untuk Bank Marketing Dataset," *Senatika*, pp. 393–397, 2021, [Online]. UCI Machine Learning

- Repository. [Online]. Diakses: 2025.
- [4] R. Wahyudi, K. Impana Manik, M. Alfin, J. Bush Henrydunan, A. Arnita, and F. Ramadhani, “Analisis Faktor Keputusan Nasabah Berlangganan Term Deposit: Perbandingan Random Forest Dan Xgboost,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 4, pp. 5600–5607, 2025, doi: 10.36040/jati.v9i4.13881.
- [5] Rasikh Muhammad Riyyasy Azfa, Nouval Aghniya Wahyu, and Tantyok Henri, “Penerapan Algoritma Machine Learning Untuk Memprediksi Term Deposit Nasabah Perbankan,” *LEDGER J. Inform. Inf. Technol.*, vol. 2, no. 3, pp. 145–56, 2023, [Online]. Available: <https://journal.ittelkom-pwt.ac.id/index.php/ledger/article/download/1416/425/>