

Perbandingan Metode *Random Forest* Dan *Naïve Bayes* Dalam Prediksi Keberhasilan Klien Telemarketing

*Ricky Leonardo¹, Janis Pratama², Chrisnatalis³

Address: Universitas Prima Indonesia, Teknik Informatika, Jl. Sekip Sei Kambing Medan 20111, Indonesia ¹²³

Email: rickyleonardooo@gmail.com¹, janispratama6@gmail.com², christ.natalis@gmail.com³

* Corresponding author

Abstrak

Telemarketing adalah salah satu metode promosi menggunakan media media telepon. Salah satu contoh dari penggunaan telemarketing ada pada bank seperti promosi deposito. Perlu dilakukan prediksi terhadap keputusan klien terhadap telemarketing untuk membantu kinerja telemarketing. Pada penelitian ini, data yang digunakan pada penelitian ini adalah data sekunder yang berasal dari website Machine Learning UCI. Metode yang digunakan dalam penelitian ini adalah *Naïve Bayes*, dan *Random Forest*. Sebelum melakukan pelatihan dan pengujian, dilakukan praproses data terhadap data-data tersebut. Metode praproses yang digunakan adalah normalisasi data, pengubahan data kategori. Setelah melakukan pengujian, dapat disimpulkan bahwa Algoritma *Random Forest* lebih tepat digunakan untuk kasus prediksi keputusan klien. Hal ini terlihat dimana akurasi yang didapatkan adalah 90%, dimana lebih tinggi 5% dibandingkan algoritma *Naïve Bayes*. Nilai dari AUC dari algoritma *Random Forest* adalah 0.97 dimana lebih tinggi 1,3 dibandingkan algoritma *Naïve Bayes*.

Keywords – Prediction, Telemarketing, *Naïve Bayes*, *Random Forest*, Comparison

1. Latar Belakang

Telemarketing adalah salah satu metode promosi dalam memasarkan atau mensosialisasikan produk menggunakan media telepon. Telemarketing dianggap menjadi salah satu metode yang cukup efektif dibandingkan metode lain seperti memasang iklan. Biaya yang diperlukan untuk melakukan telemarketing relatif rendah dibandingkan metode promosi lainnya. Menurut para pakar, penawaran menggunakan metode telemarketing mempunyai kesempatan untuk diterima oleh klien, karena sifat dari telemarketing adalah berbicara secara personal terhadap klien [1].

Salah satu contoh penggunaan telemarketing adalah bank dalam hal promosi deposito. Bank akan melakukan penawaran terhadap klien melalui telepon dengan menjelaskan produk produk deposito yang mereka tawarkan. Dalam proses telemarketing, biasanya data-data klien akan disimpan untuk dianalisa pada masa depan. Salah satu yang dapat dianalisa adalah prediksi

keputusan klien untuk melakukan deposito pada bank tersebut [2].

Beberapa algoritma yang dapat digunakan untuk memprediksi adalah *Naïve Bayes* dan *Random Forest*. *Naïve Bayes* adalah sebuah metode yang menggunakan asumsi keidependenan atribut (tidak ada kaitan antar atribut dalam melakukan klasifikasi [3]). Algoritma *Naïve Bayes* pernah digunakan dalam beberapa kasus prediksi seperti kelulusan [3], harga emas [4], usia kelahiran [5], dan lain lain. *Random Forest* adalah sebuah metode yang menggunakan pohon keputusan sebagai *base classifier* yang dibangun dan dikombinasikan untuk proses pembelajaran [6]. Algoritma *Random Forest* pernah digunakan dalam beberapa kasus prediksi seperti curah hujan [6], harga ponsel [7], lama studi mahasiswa [8], dan lain-lain.

Kedua algoritma tersebut memiliki perbedaan dalam hal performansi, terutama pada akurasi. Misalnya pada kasus klasifikasi kualitas kredit, algoritma *random forest*

memiliki performansi yang cukup baik dimana akurasi yang didapatkan lebih baik dibandingkan naïve bayes [9]. Sedangkan pada kasus klasifikasi jenis text, naïve bayes yang memiliki performansi yang cukup baik. Dari hal tersebut, perlu dilakukan perbandingan kedua algoritma tersebut untuk menentukan algoritma yang lebih tepat untuk kasus prediksi keberhasilan telemarketing [10].

Berdasarkan latar belakang diatas, maka penulis ingin melaksanakan penelitian berjudul "Perbandingan Metode *Random Forest* dan *Naïve Bayes* dalam Prediksi Keberhasilan Klien Telemarketing". Penelitian ini bertujuan untuk membandingkan metode *Random Forest* dengan *Naïve Bayes* dalam melakukan prediksi keputusan klien terhadap telemarketing untuk deposito bank.

2. Metode

Dalam penelitian ini, adapun prosedur penelitian ini adalah sebagai berikut:

1. Pengumpulan Data

Data yang digunakan pada penelitian ini didapatkan melalui website Machine learning UCI dengan alamat website <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>. Repositori dataset ini sering digunakan untuk menguji metode *machine learning*. Dataset ini memiliki 17 atribut, yaitu:

- a. age : umur dari klien
- b. job : jenis pekerjaan dari klien. Nilai pada atribut ini adalah admin, dan blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, technician, unemployed, dan unknown.
- c. marital : status pernikahan. Nilai pada atribut ini adalah divorced, married, single, unknown.
- d. education : status pendidikan. Nilai pada atribut ini adalah basic.4y, basic.6y, basic.9y, high.school, illiterate, professional.course, university.degree, dan unknown.
- e. default : status kredit pada bank. Nilai pada atribut ini adalah no, yes, dan unknown.
- f. balance : saldo rekening klien.
- g. housing : status hutang rumah. Nilai pada atribut ini adalah no, yes, dan unknown.
- h. loan : status hutang pribadi. Nilai pada atribut ini adalah no, yes, dan unknown.
- i. contact : metode telemarketing. Nilai pada atribut ini adalah cellular dan telephone.
- j. month : bulan kontak terakhir.
- k. day : hari terakhir kontak terakhir.
- l. duration : jumlah waktu melakukan telemarketing terhadap klien.

- m. campaign : jumlah klien yang telah dihubungi setelah klien ini.
- n. pdays : jumlah hari terakhir setelah klien dihubungi telemarketing sebelumnya.
- o. previous : jumlah klien yang telah dihubungi sebelum klien ini.
- p. poutcome : keputusan dari telemarketing sebelumnya.
- q. y : keputusan klien untuk melakukan deposito. Nilai pada atribut ini adalah yes, dan no.

2. Praproses Data

Praproses data adalah tahap yang penting dalam proses pengolahan data. Tahap yang dilakukan adalah:

- a. Normalisasi, adalah proses penskalaan data sehingga data terletak pada rentang tertentu.
- b. Pengubahan data kategori, adalah proses mengubah data data kategori menjadi beberapa kolom.
- c. Pembagian data, adalah proses membagi data menjadi 2 kelompok yaitu kelompok latih, dan kelompok uji.

3. Pelatihan Data

Pelatihan data adalah proses pelatihan data data dari kelompok latih menggunakan metode *Random Forest* dan *Naïve Bayes* untuk menghasilkan sebuah model yang berfungsi untuk memprediksi keberhasilan telemarketing.

4. Pengujian Data

Pengujian Data adalah proses pengujian model yang telah didapatkan dari hasil pelatihan terhadap data kelompok uji. Pada proses ini akan dibandingkan hasil pengujian model dari metode *Random Forest* dan *Naïve Bayes*.

2.1. Naïve Bayes

Naïve Bayes adalah suatu metode klasifikasi dengan menggunakan metode probabilitas dan statistik berdasarkan pengalaman masa sebelumnya. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian. Pada sebuah dataset, setiap baris/dokumen I diasumsikan sebagai vector dari nilai-nilai atribut $\langle x_1, x_2, \dots, x_n \rangle$ dimana tiap nilai-nilai menjadi peninjauan atribut X_i ($i \in [1, n]$).

Setiap baris mempunyai label kelas $c_i \in \{c_1, c_2, \dots, c_k\}$ sebagai nilai variabel kelas C , sehingga untuk melakukan klasifikasi dapat dihitung nilai probabilitas $p(C=c_i | X=x_j)$, dikarenakan pada *Naïve Bayes* diasumsikan setiap atribut

saling bebas, maka persamaan yang didapat adalah sebagai berikut (3):

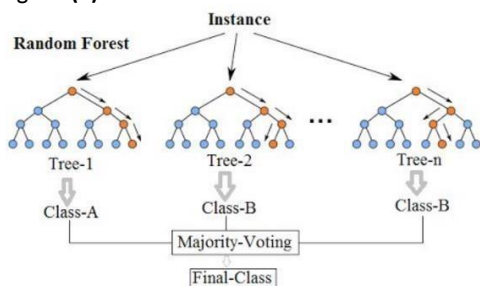
$$P = (X_i = x_i | Y_i = y_i) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - u_{ij})^2}{2\sigma_{ij}^2}} \quad (1)$$

2.2. Random Forest

Random Forest adalah sebuah metode pembelajaran melalui pembangunan pohon (tree) keputusan dengan melalui pelatihan. Ada tiga aspek penting dalam metode random forest, yaitu:

1. Melakukan bootstrap sampling untuk membangun pohon prediksi;
2. Masing-masing pohon keputusan memprediksi dengan prediktor acak;

Lalu random forest melakukan prediksi dengan mengombinasikan hasil dari setiap pohon keputusan dengan cara majority vote untuk klasifikasi atau rata-rata untuk regresi (4).



Gambar 1. Random Forest

3. Hasil dan Pembahasan

3.1. Hasil

3.1.1. Praposes Data

Data yang digunakan berjumlah 45.211 data. Atribut yang akan digunakan sebagai variabel prediktor adalah atribut age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous dan poutcome. Sedangkan untuk atribut y akan digunakan sebagai variabel target.

age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome		
0	1.606965	management	married	tertiary	no	0.256419	yes	no	unknown	5	may	0.011016	-0.569351	-0.411453	unknown	n	
1	0.288529	technician	single	secondary	no	-0.437895	yes	no	unknown	5	may	-0.416127	-0.569351	-0.411453	unknown	n	
2	-0.747384	entrepreneur	married	secondary	no	-0.446762	yes	yes	unknown	5	may	-0.707361	-0.569351	-0.411453	unknown	n	
3	0.571051	blue-collar	married	unknown	no	0.047205	yes	no	unknown	5	may	-0.645231	-0.569351	-0.411453	unknown	n	
4	-0.747384	unknown	single	unknown	no	-0.447091	no	no	unknown	5	may	-0.233620	-0.569351	-0.411453	unknown	n	
...	
45206	0.947747	technician	married	tertiary	no	-0.176460	no	no	cellular	17	nov	2.791329	0.076230	-0.411453	unknown	ye	
45207	2.831227	retired	divorced	primary	no	0.120447	no	no	cellular	17	nov	0.768224	-0.246560	-0.411453	unknown	ye	
45208	2.925401	retired	married	secondary	no	1.429593	no	no	cellular	17	nov	3.373797	0.721811	1.436189	1.050473	success	ye
45209	1.512791	blue-collar	married	secondary	no	-0.228024	no	no	telephone	17	nov	0.970146	0.399020	-0.411453	unknown	n	
45210	-0.370689	entrepreneur	married	secondary	no	0.528364	no	no	cellular	17	nov	0.399328	-0.246560	1.476138	4.523577	other	n

Gambar 2. Data Penelitian

	age	balance	duration	campaign	pdays	previous
0	1.606965	0.256419	0.011016	-0.569351	-0.411453	-0.251940
1	0.288529	-0.437895	-0.416127	-0.569351	-0.411453	-0.251940
2	-0.747384	-0.446762	-0.707361	-0.569351	-0.411453	-0.251940
3	0.571051	0.047205	-0.645231	-0.569351	-0.411453	-0.251940
4	-0.747384	-0.447091	-0.233620	-0.569351	-0.411453	-0.251940
...
45206	0.947747	-0.176460	2.791329	0.076230	-0.411453	-0.251940
45207	2.831227	0.120447	0.768224	-0.246560	-0.411453	-0.251940
45208	2.925401	1.429593	3.373797	0.721811	1.436189	1.050473
45209	1.512791	-0.228024	0.970146	0.399020	-0.411453	-0.251940
45210	-0.370689	0.528364	0.399328	-0.246560	1.476138	4.523577

Gambar 3. Hasil Normalisasi

Untuk atribut dengan jenis kategori akan dilakukan perubahan menjadi nilai dummy. Atribut yang diubah menjadi nilai dummy adalah job, marital, education, default, loan, contact, month, y.

day	job_blue-collar	job_entrepreneur	job_housmaid	job_management	job_retired	job_self-employed	job_services	job_student	job_technician	job_unemployed	job_unknown	marital_married
0	0	0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	1
...
45206	17	0	0	0	0	0	0	0	0	0	1	0
45207	17	0	0	0	0	1	0	0	0	0	0	0
45208	17	0	0	0	0	1	0	0	0	0	0	0
45209	17	1	0	0	0	0	0	0	0	0	0	0
45210	17	0	0	0	0	0	0	0	0	0	0	0

Gambar 4. Hasil Perubahan Variabel Dummy

Data akan dikelompokkan menjadi 2 kelompok yaitu kelompok pelatihan dan pengujian Rasio pembagiandata adalah 4:1. Melalui pengelompokkan data ini jumlah data pelatihan menjadi 36.168 data, dan jumlah data pengujian adalah 9.043 data.

3.1.2. Pengujian

Hasil pengujian akan divisualisasikan menggunakan metode confusion matrix. Berikut adalah hasil dari pengujian model yang dihasilkan oleh metode tersebut terhadap data pengujian.

Tabel 1. Confusion Naïve Bayes

	Prediksi +	Prediksi -
Aktual +	7225	779
Aktual -	537	502

Tabel 2. Confusion Matrix Random Forest

	Prediksi +	Prediksi -
Aktual +	7768	236
Aktual -	624	415

Berdasarkan confusion matrix pada tabel 1. dan 2., beberapa jenis nilai evaluasi seperti akurasi, presisi dan recall dari algoritma Naïve Bayes dan Random Forest dapat diambil yaitu sebagai berikut:

1. Naïve Bayes

Tabel 3. Nilai Akurasi, Presisi, dan Recall Naïve Bayes

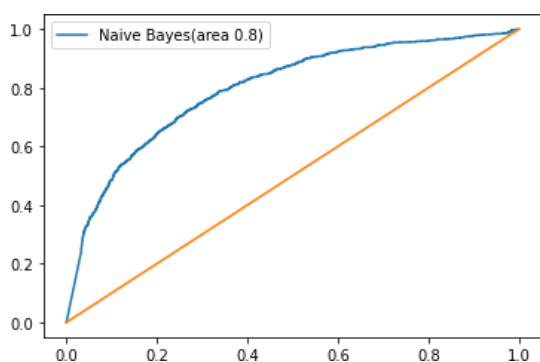
Akurasi	Presisi	Recall
85%	93%	90%

2. Random Forest

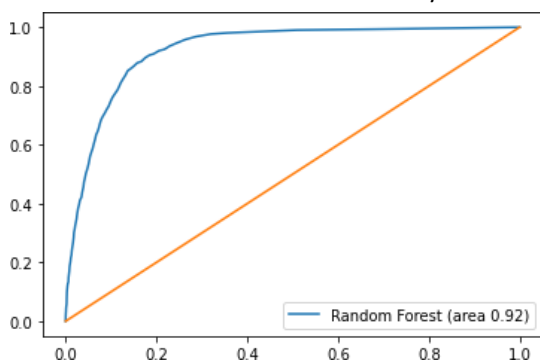
Tabel 4. Nilai Akurasi, Presisi, dan Recall Random Forest

Akurasi	Presisi	Recall
90%	92%	97%

Nilai AUC dari Algoritma *Naïve Bayes* dan *Random Forest* dalam melakukan prediksi adalah 0.8 dan 0.92 dalam Berikut adalah kurva ROC (*Receiver Operating Characteristic*) setiap algoritma dalam melakukan prediksi terhadap keputusan klien.



Gambar 5. Kurva ROC Naïve Bayes

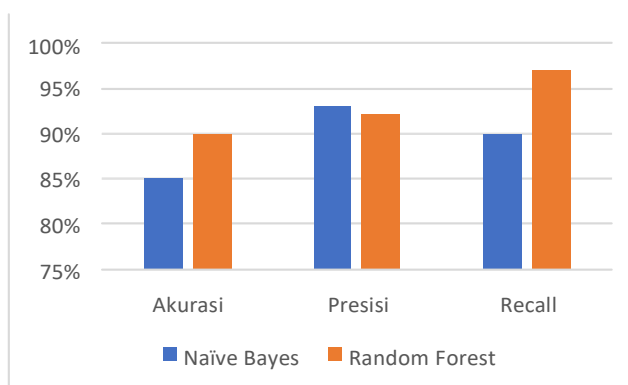


Gambar 6. Kurva ROC Random Forest

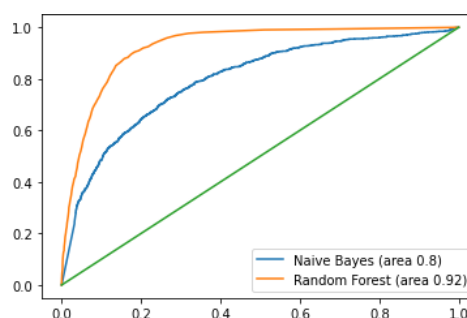
tinggi 0.13 dibandingkan algoritma *Naïve Bayes*. Nilai AUC dari *Naïve Bayes* juga terlalu rendah yaitu 0.79 dimana nilai tersebut termasuk kedalam kinerja klasifikasi algoritma yang sedang dalam kasus ini.

Tabel 5. Hasil Perbandingan Algoritma

No.	Nama Algoritma	Akurasi	Presisi	Recall	AUC
1.	Naïve Bayes	85%	93%	92%	0.8
2.	Random Forest	90%	92%	97%	0.92



Gambar 7. Grafik Perbandingan Nilai Akurasi, Presisi, dan Recall



Gambar 8 Perbandingan Kurva ROC

3.2. Pembahasan

Berdasarkan hasil pengujian, Algoritma *Naïve Bayes* dan *Random Forest* cocok dalam kasus ini, dimana akurasi dari kedua algoritma diatas 80%. Dengan membandingkan hasil dari performasi dari Algoritma *Naïve Bayes* dan *Random Forest* dalam melakukan prediksi, terlihat bahwa algoritma *Random Forest* lebih tepat dalam kasus ini.

Hal ini terlihat dimana algoritma *Random Forest* mencapai akurasi sebesar 90%, lebih tinggi 5% dibandingkan dengan algoritma *Naïve Bayes*. *Random Forest* juga mendapatkan nilai AUC sebesar 0.92, lebih

4. Kesimpulan

Setelah melakukan penelitian ini, maka dapat diambil kesimpulan sebagaiberikut:

1. Metode *Naïve Bayes*, dan *Random Forest* dapat digunakan untuk melakukan prediksi.
2. Metode *Random Forest* lebih tepat dalam kasus ini dikarenakan memiliki performasi yang paling baik dibandingkan metode *Naïve Bayes*.

5. Daftar Pustaka

- [1] J. A. Monareh, A. Fauzi, and I. P. Nuralam, "PENGARUH TELEMARKETING TERHADAP KEPUTUSAN PEMBELIAN (Survei Online pada Pelanggan Produk Multiguna Astra Credit Companies Priority di PT Astra Sedaya Finance)," *J. Adm. Bisnis*, vol. 58, no. 2, 2018.
- [2] E. P. Saputra, "PREDIKSI KEBERHASILAN TELEMARKETING BANK UNTUK Mencari ALGORITMA DENGAN PERFORMA TERBAIK," *J. Ilmu Pengetah. Dan Teknol. Komput.*, vol. 2, no. 2, pp. 66–72, 2017.
- [3] S. Syarli and A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilm. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2016.
- [4] M. Guntur, J. Santony, and Y. Yuhandri, "Prediksi Harga Emas dengan Menggunakan Metode Naïve Bayes dalam Investasi untuk Meminimalisasi Resiko," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 1, pp. 354–360, 2018, doi: 10.29207/resti.v2i1.276.
- [5] N. R. Indraswari and Y. I. Kurniawan, "Aplikasi Prediksi Usia Kelahiran Dengan Metode Naive Bayes," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 129–138, 2018, doi: 10.24176/simet.v9i1.1827.
- [6] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.
- [7] V. W. Siburian and I. E. Mulyana, "Prediksi Harga Ponsel Menggunakan Metode Random Forest," *Pros. Annu. Res. Semin.*, vol. 4, no. 1, pp. 144–147, 2018.
- [8] I. M. Budi Adnyana, "Prediksi Lama Studi Mahasiswa Dengan Metode Random Forest (Studi Kasus : Stikom Bali)," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 8, no. 3, pp. 201–208, 2016, doi: 10.22303/csrid.8.3.2016.201-208.
- [9] S. N. Edusaintek, B. Bawono, R. Wasono, and U. M. Semarang, "Perbandingan Metode Random Forest Dan Naïve Bayes Untuk Klasifikasi Debitur Berdasarkan Kualitas Kredit," *Fmipa Unimus*, pp. 343–348, 2019.
- [10] T. Pranckevičius and V. Marcinkevičius, "Comparison of Naive Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification," *Balt. J. Mod. Comput.*, vol. 5, no. 2, 2017, doi: 10.22364/bjmc.2017.5.2.05.
- [11] A. P. Wibawa *et al.*, "Naïve Bayes Classifier for Journal Quartile Classification," *Int. J. Recent Contrib. from Eng. Sci. IT*, vol. 7, no. 2, p. 91, 2019, doi: 10.3991/ijes.v7i2.10659.