

## Komparasi Metode Naïve Bayes dan C45 pada Prediksi Pelanggan Deposito Berjangka

**Meilany Nonsi Tentua<sup>1</sup>, Turiyan<sup>2</sup>, Karandi Nurbagja<sup>3</sup>**

<sup>1,2,3</sup>Program Studi Informatika, Universitas PGRI Yogyakarta

e-mail: <sup>1</sup>[meilany@upy.ac.id](mailto:meilany@upy.ac.id), <sup>2</sup>[turiyan295@gmail.com](mailto:turiyan295@gmail.com),

<sup>3</sup>[karandinurbagja@gmail.com](mailto:karandinurbagja@gmail.com)

### **Intisari**

Penentuan pelanggan produk pada dunia perbankan dalam memasarkan produk seperti produk deposito berjangka merupakan permasalahan tersendiri. Penelitian ini bertujuan untuk mengkomparasi metode dalam pengembangan model prediksi langganan deposito berjangka dengan pemasaran secara langsung dari lembaga perbankan portugis. Pemasaran ini didasarkan pada panggilan telepon. Dengan memasukan data informasi kontak yang dilakukan kepada klien untuk mendapatkan jawaban “ya” atau “tidak”. Model prediksi yang dibandingkan dibangun dengan metode Naïve Bayes dan metode C45. Hasil penelitian memperlihatkan model dengan nilai tertinggi adalah model yang dibangun dengan metode Naivebayes yaitu 63.77% . Model yang dibangun dengan metode Naïve Bayes juga dapat menghubungkan antar atribut sehingga lebih akurat dan menjadikan data sederhana dan efisien.

**Kata kunci**— Naïve Bayes, C45, Deposito Berjangka

### ***Abstract***

*Determining product customers in the banking world in marketing products such as time deposit products is a problem. This study aims to compare the methods in*

*developing a predictive model for time deposit subscriptions with direct marketing from Portuguese banking institutions. This marketing is based on phone calls, by entering contact information data made to the client to get a "yes" or "no" answer. The comparison prediction model was built using the Naïve Bayes method and the C45 method. The results showed that the model with the highest value was built using the Naive Bayes method, which was 63.77%. The model built using the Naïve Bayes method can also link between attributes to be more accurate and make the data simple and efficient.*

**Keywords**— Naïve Bayes, C45, deposit products

## PENDAHULUAN

Pada saat ini untuk di perbankan banyak yang bersaing supaya masyarakat bisa mempercayai Lembaga keuangan. Pada saat ini, peran perbankan sangat penting untuk memajukan perekonomian di suatu negara. Bank merupakan tempat untuk memudahkan masyarakat dalam hal menyimpan uang atau untuk menyiapkan investasi di masa depan [1]. Jasa bank pasti berhubungan dengan semua keuangan. Oleh karena itu, dunia perbankan banyak menyediakan tempat untuk penyimpanan keuangan bagi klien yang berlangganan. Tempat penyimpanan tersebut yaitu deposito, deposito merupakan simpanan klien yang dilakukan perjanjian penyimpanan dengan bank bersangkutan dalam kurun waktu tertentu untuk melakukan penarikan [2]. Sebuah perbankan di portugis mengadakan bank marketing dalam memberikan deposit yang berjangka pada pelanggannya. Di dalam dataset ini berisikan data tentang Bank marketing yang berada di portugis yang akan mengkampanyekan deposit berjangka dengan cara menghubungi pelanggannya melalui media telepon dengan cara memberikan informasi kontak yang dilakukan pelanggan untuk mendapatkan jawaban “yes” atau “no” untuk deposit berjangka milik mereka [3]. Karena pada zaman modern ini kita bisa memanfaatkan teknologi-teknologi yang akan bisa dimanfaatkan sebagai alat komunikasi jarak jauh yang akan semakin canggih, seperti telephone dan lain – lainnya. Pemasaran dan melalui telephone ini bisa dibilang sangat berpengaruh

bagi pelaku pemasaran karena bisa mengenalkan produk mereka ke masyarakat secara luas dengan singkat [4]. Teknik pemasaran ini disebut dengan Telemarketing, yaitu salah satu cara untuk mempromosikan produk yang paling efektif, karena strategi pemasaran ini melakukan penawaran produk pada klien, salah satunya yaitu deposito berjangka yang ditawarkan oleh bank tersebut [4]. Dan menggunakan strategi pemasaran seperti ini bisa meningkatkan peminjaman modal dengan menawarkan deposito berjangka bagi masyarakat dan dalam meningkatkan keberhasilan tersebut dilakukan dengan memprediksi keputusan masyarakat untuk mengurangi biaya pada saat melakukan panggilan telephone [4].

Pada model Naivebayes dapat menghubungkan atribut dan antar fitur yang independen sehingga menjadi sederhana dan efisien [5]. Model C45 merupakan algoritma yang sangat detail dalam perhitungannya dengan menggunakan pohon keputusan untuk melihat hasil akhirnya [6]. Metode Naivebayes merupakan metode yang mungkin sangat sederhana untuk mengklasifikasikan data yang efisien, relevan dan sederhana.[7]. Namun pada saat ini belum ditemukan algoritma yang akurat dalam penentuan prediksi-prediksi deposit berjangka. Berdasarkan uraian diatas penelitian ini akan mengkomparasi dua metode utk memprediksi pelanggan deposito berjangka, yaitu metode Naive Bayes dan metode C45.

## METODE PENELITIAN

Metode yang akan dilakukan untuk dataset “Bank Marketing” adalah memprediksi dan mengkomparasi kedua metode tersebut :

### A. Pengumpulan data

Data yang digunakan untuk penelitian ini bersumber dari : <https://www.kaggle.com/sonujha090/bank-marketing>. Data tersebut merupakan hasil data kumpulan kampanye pemasaran deposit bank di portugis. Data dibentuk oleh tim pemasaran deposit yang berjangka melalui media telephone yang

## Komparasi Metode Naïve Bayes dan C45 pada Prediksi Pelanggan Deposito Berjangka

(Meilany Nonsi Tentua, Turiyan, Karandi Nurbagja)

dilakukan pelanggan mendapatkan informasi apakah pelanggan memilih “ya” atau “tidak”. Dataset terdiri dari 4522 data dengan 17 atribut (16 atribut dan 1 atribut target). Untuk data (train) digunakan sebanyak 4000 data, sedangkan untuk data (test) digunakan sebanyak 521 data.

Atribut itu yang disajikan dataset Bank Marketing adalah:

1. Usia : age (numerik)
2. Tipe pekerjaan job : (Kategorikal)
3. Status perkawinan (marital) : married, single, dan divorced
4. Education : primary, secondary, unknown, dan tertiary
5. memiliki kredit?(default) : yes dan no
6. memiliki pinjaman perumahan? (housing) : yes dan no
7. memiliki pinjaman pribadi? (loan) : yes dan no
8. jenis komunikasi (contact) : unknown, telephone, dan cellular
9. kontak terakhir Bulan (month) : jan, feb, mar,....., nov, dec
10. durasi kontak terakhir, dalam detik (duration) : numerik
11. jumlah kontak yang dilakukan selama kampanye ini dan untuk klien ini (numerik, termasuk kontak terakhir) : campaign
12. jumlah hari yang berlalu setelah klien terakhir dihubungi dari kampanye sebelumnya (numerik; 999 berarti klien sebelumnya tidak dihubungi) : pdays
13. jumlah kontak yang dilakukan sebelum kampanye ini dan untuk klien ini (numerik) : previous
14. hasil dari kampanye pemasaran sebelumnya (kategoris) : poutcome
15. balance: saldo tahunan rata-rata
16. y - apakah klien telah berlangganan deposito berjangka? (biner: ‘yes’, ‘no’)

### B. Processing data

Processing data merupakan suatu cara agar membuat data mentah menjadi data yang bisa di gunakan atau berkualitas. Pemilahan data yang tidak sesuai dengan cara menghilangkan Sebagian data yang salah termasuk kedalam processing

data. Pada saat pemilahan data dilakukan dengan menghapus atribut data yang bernilai tidak sesuai karena nilai data yang tidak sesuai akan merusak data yang akan dihitung. Pada perhitungan dataset ini akan dilakukan penyederhanaan data dengan melakukan range data di setiap atributnya karena nilai disetiap data terlalu tinggi jika tidak melakukan range maka akan mempengaruhi atau mengubah data di setiap atributnya.

## HASIL DAN PEMBAHASAN

Model prediksi yang dibangun menggunakan metode Naivebayes dikerjakan dengan cara mencari probabilitas “ya” dan “tidak”. Selanjutnya mencari rata-rata(mean) dan deviasi standar dari data training. Formula untuk prediksi pada metode Naïve Bayes menggunakan persamaan (1).

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

Dengan  $\mu$  adalah rata-rata,  $\sigma$  adalah nilai deviasi standart dan  $\pi$  adalah nilai ketetapan yaitu 3,14.

Hasil klasifikasi dari metode Naivebayes diuji dengan perbandingan Akurasi, precision, dan recall dengan menggunakan tabel confusion matriks. Tabel 1 menyajikan hasil perhitungan confusion matrik.

Tabel 1. Tabel perbandingan akurasi di metode Naivebayes

akurasi	precision	recall
2883	3597.4471	2883
0.6376908	0.7957193	0.637691
63.77%	80%	64%

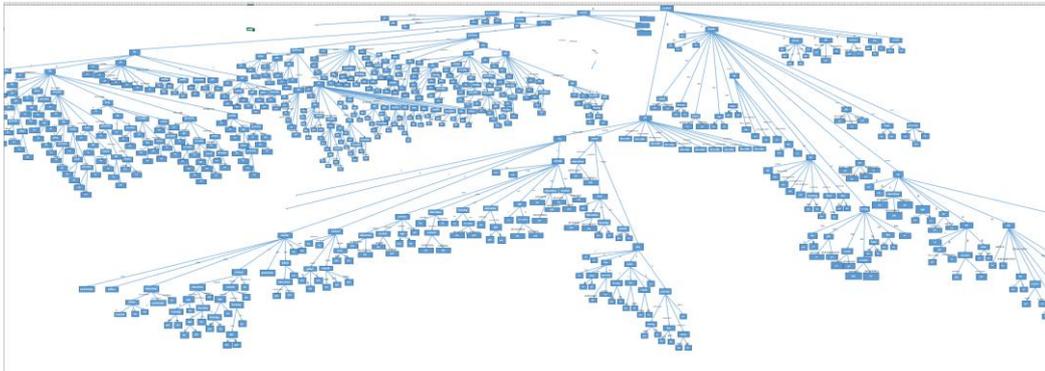
Model prediksi yang dibangun dengan metode C45 [8] akan membentuk pohon penelusuran dengan perhitungan yang dimulai dengan mencari nilai entropi pada persamaan (2). Setelah itu, nilai gain pada persamaan (3) akan dicari dengan menggunakan nilai entropi sebelumnya.

$$Entropi(S) = E(S) = \sum_{j=1}^k -p_j \log_2 p_j \quad (2)$$

Dengan  $S$  adalah seluruh data,  $k$  adalah jumlah partisi pada  $S$  dan  $p_j$  adalah nilai probabilitas masing-masing target ( $y/n$ ).

$$Gain(A) = Entropi(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi(S_i) \quad (3)$$

Dengan  $S$  adalah seluruh data,  $A$  adalah atribut,  $|S_i|$  adalah jumlah attribute  $I$ ,  $|S|$  adalah jumlah seluruh data dan  $Entropi(S_i)$  adalah nilai entropi pada atribut ke- $i$ .



Gambar 1. Pohon penelusuran yang terbentuk pada metode C45

Pohon penelusuran yang terbentuk untuk memprediksi pelanggan deposito berjangka tersaji pada Gambar1. Hasil klasifikasi dari metode 45 diuji dengan perbandingan Akurasi, precision, dan recall dengan menggunakan tabel confusion matriks. Tabel 2 menyajikan hasil perhitungan confusion matrik.

Tabel 2. Tabel perbandingan akurasi di metode C45

akurasi	precision	recall
521	60.0400354	4000
0.1152400	0.0132803	0.884760
12%	1%	88%

Hasil penelitian memperlihatkan bahwa tingkat akurasi di metode Naivebayes ini lebih tinggi dari metode C45. Karena dalam dalam metode

Naivebayes lebih akurat kebenarannya dibandingkan dengan metode C45. Tingkat precision di metode Naivebayes lebih tinggi dari pada metode C45. Dikarenakan tingkat informasi yang didapatkan dari perhitungan lebih banyak. Perbandingan tingkan akurasi disajikan pada tabel 3.

Tabel 3. Tabel Perbandingan Metode Naivebayes, dan C45

Evaluasi	c45	Naivebayes
Akurasi	12%	64%
Precision	1%	80%
Recall	88%	64%

Tabel IV. Gambar grafik C45

Pada perhitungan diatas sudah bisa membandingkan antara metode Naivebayes dan metode C45, karena pada kedua algoritma itu bisa mendapatkan hasil akurasi algoritma dari Naivebayes sebanyak 63.77%, sedangkan nilai akurasi dari algoritma C45 sebesar 12%.

## KESIMPULAN

Hasil perbandingan antara Naïve Bayes dan C45 pada dataset "Bank Marketing" mendapatkan tingkat akurasi 63.77%, 80% untuk precision, dan 64% untuk recall. Metode C45 mempunyai tingkat akurasi 12%, 1% untuk precision, dan 88% untuk recall. Pada dataset "Bank marketing" ini bisa menunjukkan bahwa metode Naivebayes mempunyai tingkat akurasi yang lebih tinggi sehingga hasil prediksi menjadi lebih akurat.

## DAFTAR PUSTAKA

- [1] Nurfaizah dan R. D. A. Parmitasari, "Pengaruh tingkat suku bunga deposito terhadap jumlah dana deposito pada bank mandiri cabang utama makassar," *J. Akunt.*, hal. 1–13, 2013.

## Komparasi Metode Naïve Bayes dan C45 pada Prediksi Pelanggan Deposito Berjangka

(Meilany Nonsi Tentua, Turiyan, Karandi Nurbagja)

- [2] Amanitanovi, “Bab 4 – Deposito,” *bp1.blogger.com*, hal. 26–36, 2013.
- [3] B. Kencana, “Klasifikasi Logistic dan KNN pada data direct Campaign Bank (Phone Calls),” *RPosts - Analisis Campaign Bank*, 2020. .
- [4] R. Sulaehani, “Prediksi Keputusan Klien Telemarketing Untuk Deposito Pada Bank Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination,” *Ilk. J. Ilm.*, vol. 8, no. 3, hal. 182–189, 2016, doi: 10.33096/ilkom.v8i3.83.182-189.
- [5] F. Husaini, “Algoritma Klasifikasi Naive Bayes Untuk Menilai Kelayakan Kredit (Studi Kasus : Bank Mandiri Kredit Mikro),” *Progr. Stud. Tek. Inform.*, vol. 1, no. 3, hal. 2–12, 2016.
- [6] W. Darmawan, “Komparasi Metode Data Mining Dalam Memprediksi Nasabah Bank yang akan Memilih Tabungan Deposito Menggunakan Algoritma Klasifikasi,” *IC-Tech*, vol. XIII, no. 1, hal. 49–55, 2018.
- [7] A. & T. Sugiana, “Komparasi Algoritma Naive Bayes , K-Nearest Neighbor dan C45 Pada Penyakit Melanoma,” *Semin. Nas. Din. Inform. 2020 Univ. PGRI Yogyakarta*, hal. 163–165, 2020, doi: 10.5507/bp.2012.081.P.
- [8] M. N. Tentua dan A. Sihabuddin, “Improved C4 . 5 Algorithm Using The L ’ hospital Rule And Prunning On The Recommendation System,” *Int. J. Sci. Technol. Res.*, vol. 8, no. 11, 2019.