

ALGORITMA C4.5 UNTUK MEMPREDIKSI PENGAMBILAN KEPUTUSAN MEMILIH DEPOSITO BERJANGKA

Hendri Mahmud Nawawi¹; Sri Rahayu²; Muhammad Ja'far Shidiq³; Jajang Jaya Purnama⁴

^{1,2,3,4} Program Studi Ilmu Komputer
STMIK Nusa Mandiri Jakarta
www.nusamandiri.ac.id

mahmudhend94@gmail.com, srahayu110527@gmail.com, ash.shidiq.mj@gmail.com,
jajangja2412@bsi.ac.id



Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi-NonKomersial 4.0 Internasional.

Abstract— Deposits are one form of investment offered by the Bank or other financial institutions with the nature of regulating and binding according to the rules set by the manager and the investor or commonly called investors. The advantage of being an investor is getting a fee or profit calculated based on the agreed time period at the beginning of the agreement. Whereas for investor fund managers can be used to advance and develop their business and business. Finding and determining potential customers is the first step to running a financial business in the form of this deposit, before the transaction decision is taken which is a favorable decision for both parties, investors or managers, one of the decision making techniques can be done using Data Mining using the C4.5 Algorithm which is a structured decision-making technique based on input variables so that it can produce the most potential typical information for customers to participate in time deposits.

Keywords: Deposits, Data Mining, C4.5 Algorithms

Intisari— Deposito merupakan salah satu bentuk investasi yang ditawarkan oleh pihak Bank atau lembaga keuangan lainnya dengan sifat mengatur dan mengikat sesuai aturan yang telah ditetapkan oleh pihak pengelola dan pihak pemodal atau biasa disebut investor. Keuntungan menjadi investor adalah mendapatkan sejumlah *fee* atau keuntungan yang dihitung berdasarkan tenggang waktu yang telah disepakati diawal perjanjian. Sedangkan bagi pengelola dana investor bisa digunakan untuk memajukan dan mengembangkan bisnis dan usahanya. Mencari dan menentukan pelanggan potensial adalah langkah awal untuk menjalankan bisnis keuangan dalam bentuk deposito ini, sebelum terjadinya transaksi keputusan yang diambil adalah keputusan yang menguntungkan kedua pihak baik

investor atau pengelola salah satu teknik pengambilan keputusan ini dapat dilakukan dengan menggunakan proses Data Mining dengan menggunakan Algoritma C4.5 yang merupakan teknik pengambilan keputusan secara terstruktur berdasarkan variable inputan sehingga dapat menghasilkan informasi typical pelanggan paling potensial untuk mengikuti deposito berjangka.

Kata Kunci : Deposito, Data Mining, Algoritma C4.5

PENDAHULUAN

Perbankan merupakan lembaga yang berperan penting untuk kestabilan ekonomi suatu Negara. Perbankan mencakup layanan penyimpanan asset keuangan Negara dan menghimpunnya kemudian disalurkan ke berbagai sektor ekonomi. Selain untuk menghimpun dan menyalurkan dana saat ini layanan perbankan memang menjadi lahan bisnis yang cukup diminati oleh orang-orang yang memiliki asset atau uang banyak, selain dengan tujuan keamanan menyimpan uang di Bank menjadikan pilihan supaya uang yang disimpan bertambah dengan adanya komisi bagi hasil atau biasa disebut bunga. Salah satu produk yang sering ditawarkan bank adalah deposito. Deposito atau sering disebut juga deposito berjangka merupakan salah satu produk bank jenis jasa tabungan yang ditawarkan kepada masyarakat. Dana deposito merupakan tabungan berjangka yang sifatnya mengikat tetapi dengan keunggulan penwaran bunga yang cukup menggiurkan.

Menurut Undang-Undang No. 10/1998, Pasal 1 ayat 7 (1998:7) yang memberikan pengertian deposito adalah simpanan yang penarikannya hanya dapat dilakukan pada waktu tertentu berdasarkan perjanjian nasabah penyimpan

dengan bank. Berdasarkan definisi deposito tersebut dapat disimpulkan bahwa pengertian deposito adalah simpanan jenis ketiga yang dikeluarkan bank yang penarikannya hanya dapat dilakukan dalam waktu tertentu berdasarkan perjanjian antara nasabah penyimpan dengan bank yang bersangkutan.

Dengan sifatnya yang mengikat antara bank selaku penyalur dana dan nasabah selaku penyimpan dana deposito menjadi sebuah layanan transaksi yang harus betul-betul matang difikirkan terutama bagi nasabah, hal ini bertujuan untuk menghindari dana kena cas tau kena denda apabila disuatu hari dana yang di depositokan diambil sebelum masa perjanjian berakhir.

Pada penulisan makalah ini penulis membuat sebuah analisa untuk membuat klasifikasi terhadap calon nasabah potensial untuk membuat keputusan bahwa ia akan melakukan deposito atau tidak. Data yang diambil berdasarkan dataset yang dilakukan oleh marketing di salah satu bank diportugis dengan melakukan konfirmasi terhadap beberapa nasabah melalui telpon bahwa ia memilih deposito atau tidak untuk dana yang ia miliki. Untuk mengetahui hubungan antara atribut dalam menentukan sebuah keputusan Sistem Penunjang Keputusan (SPK) atau Decision Support System (DSS), merupakan sebuah sistem yang digunakan sebagai alat bantu bagi para pengambil keputusan untuk memperluas kemampuan manusia dengan tidak menggantikan perannya. SPK dikembangkan dengan tujuan untuk mendukung pengambilan keputusan dengan memilih alternatif keputusan yang merupakan hasil pengolahan informasi-informasi yang diperoleh atau tersedia dengan menggunakan model-model pengambil keputusan serta untuk menyelesaikan masalah-masalah bersifat terstruktur, semi terstruktur dan tidak terstruktur (Yunus, Dahlan, & Santoso, 2014).

Berry dan linoff dalam bukunya data mining technique for marketing, Sales, and Customers Support mendefenisikan data mining sebagai suatu proses eksplorasi dan analisis secara otomatis maupun semiotomatis terhadap data dalam jumlah besar dengan tujuan menemukan pola atau aturan yang berarti. Dalam situasi ini yang melekat dapat berupa korelasi dikontrol untuk, atau dihapus sama sekali, selama kontruksi desain eksperimental (Wijaya, Hasibuan, & Ramadhani, 2018).

Untuk menentukan sebuah keputusan Penelitian ini menggunakan Algoritma Klasifikasi *decision tree* C4.5 dengan melakukan klasifikasi terhadap kategori yang paling mempengaruhi untuk membuat keputusan ya atau tidak untuk melakukan deposito.

Konsep dasar dari algoritma RS-C4.5 mirip dengan algoritma C4.5 yaitu mencapai klasifikasi untuk membangun pohon keputusan. Sebelum membuat sebuah pohon keputusan, proses reduksi atribut dilakukan terlebih dahulu pada set atribut asli untuk mendapatkan set minimal atribut, lalu atribut perolehan dengan rasio maksimal dipilih untuk memberikan informasi klasifikasi dari tabel keputusan setelah pengurangan atribut untuk menghasilkan pohon keputusan secara rekursif, mengatasi kekurangan C4.5 algoritma ketika informasi diperoleh secara langsung diadopsi ketika atribut dipilih (Bao & Guan, 2016).

Rumusan Masalah

Pada rumusan masalah yang akan dibahas adalah membuat prediksi terhadap investor potensial yang memungkinkan untuk melakukan deposito berjangka berdasarkan hasil kampanye yang dilakukan oleh marketing disalah satu bank di portugis yang melakukan kampanye melalui telephone untuk mencari pelanggan potensial untuk mengikuti program deposito berjangka, hasil dari riset dengan menggunakan 17 atribut dimana 9 dari 17 atribut tersebut merupakan data kuantitatif dan 8 atribut kualitatif hasil dari kampanye dengan menggunakan 17 atribut inputan tersebut menghasilkan sebuah keputusan Ya dan Tidak dari pelanggan yang ditawarkan untuk mengikuti deposito berjangka.

Tujuan

Tujuan dari penelitian ini untuk mengetahui tipe pelanggan seperti apa yang potensial untuk memutuskan memilih deposito berjangka.

Metode Penelitian

Metode penelitian yang digunakan pada permasalahan ini adalah dengan membuat klasifikasi menggunakan algoritma *decision tree* C4.5. Pohon keputusan adalah penggolong yang melakukan partisi rekursif. Contohnya sebuah pohon keputusan khas terdiri dari node internal, tepi dan node daun. Setiap simpul daun adalah simpul terminal dari pohon dengan label kelas. Misalnya memberikan ilustrasi dari pohon keputusan dasar, di mana lingkaran berarti simpul dan persegi keputusan berarti simpul daun. Dalam contoh ini, kami memiliki tiga atribut pemisahan, yaitu usia, jenis kelamin, dan kriteria 3, bersama dengan dua label kelas, yaitu, YA dan TIDAK. Setiap jalur dari simpul akar ke daun node membentuk aturan klasifikasi (Amin, Indwiarti, & Sibaroni, 2015).

BAHAN DAN METODE

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. Data mining merupakan teknologi yang digunakan dalam disiplin ilmu yang berbeda-beda untuk mencari hubungan yang signifikan antara variabel dalam set data yang besar. Metode data mining dapat digunakan untuk melakukan klasifikasi data dengan mengubah data menjadi pohon keputusan (decision tree) dan aturan-aturannya (rule), yang dapat digunakan untuk memprediksi atau mengklarifikasi suatu kejadian (Yunus et al., 2014).

Dataset ini diambil dari UCI Repository data Bank di Protugis yang diakses secara *Online*. attribute yang digunakan tidak semuanya, mengacu pada penelitian ini attribut yang digunakan merupakan attribut pokok (Wiharto, Kusnanto, & Herianto, 2016). Dari 17 attribut yang didapatkan 9 diantaranya merupakan data kuantitatif kemudian data tersebut diolah menjadi data kualitatif sehingga attribut yang dipakai menjadi 9 attribut kualitatif yang merupakan inputan paling berpengaruh. Teknik pemangkasan ini bertujuan untuk membuat struktur pohon yang dihasilkan oleh algoritma C4.5 lebih sederhana dan distribusi kelasnya dapat dipertahankan (Amin et al., 2015). Evaluasi dan validasi hasil dalam penelitian ini adalah hasil klasifikasi pada data *training* berdasarkan waktu proses yang dibutuhkan, tingkat akurasi, serta jumlah data yang diklasifikasikan ke dalam kelas YA/TIDAK. Untuk mengetahui evaluasi dari kinerja model klasifikasi didasarkan pada banyaknya (*count*) *dataset record* yang diprediksi secara benar dan tidak benar pada model klasifikasi tersebut. *Count* ini ditabulasikan dalam sebuah hasil yang dikenal sebagai *confusion matrix* (Santoso, 2011).

Algoritma C4.5 yang merupakan algoritma pengambilan keputusan yang melakukan partisi rekursif atas ruang instance, sebuah pohon keputusan tipikal terdiri dari simpul internal, tepi dan simpul daun. Setiap simpul internal disebut simpul keputusan yang mewakili tes pada atribut atau subset atribut, dan masing-masing edge diberi label dengan nilai spesifik atau rentang nilai atribut input (Dai & Ji, 2014).

Secara umum alur proses algoritma C4.5 untuk membangun pohon keputusan dalam data mining adalah:

- 1) Pilih atribut sebagai simpul akar.
- 2) Buat cabang untuk tiap-tiap nilai.
- 3) Bagi kasus dalam cabang.

- 4) Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama (Yunus et al., 2014).

Pemilihan atribut sebagai simpul, baik simpul akar (root) atau simpul internal didasarkan pada nilai Gain tertinggi dari atribut-atribut yang ada.

Penghitungan nilai Gain digunakan rumus seperti dalam Persamaan 1.

$$\text{Gain}(S, A) = \frac{\text{Gain}(S, A)}{S_i(S, A)} \dots\dots\dots (1)$$

$$S_i = -\sum_i^n - p_i \log 2p_i \dots\dots\dots (2)$$

- S : Himpunan kasus
- A : Atribut
- n : Jumlah partisi atribut A
- |S_i |: Jumlah kasus pada partisi ke-i
- |S|: Jumlah kasus dalam S

untuk menghitung nilai Entropy dapat dilihat pada Persamaan 2.

$$\text{Entropy}(S) = -\sum_{i=1}^c s_i/s \log 2 s_i/s \dots\dots\dots (3)$$

(Widayu, Darma, Silalahi, & Mesran, 2017).

Algoritma C4.5 memiliki aturan input: sampel pelatihan, label pelatihan, atribut

1. Membangun simpul akar untuk pohon
2. Jika semua sampel positif, berhenti setelah pohon dengan simpul akar, beri label dengan (+)
3. Jika semua sampel negatif, berhenti setelah pohon dengan simpul akar, beri label dengan (-)
4. jika atribut hilang, berhenti setelah pohon dengan simpul root, beri label dengan nilai yang sering muncul dalam label pelatihan
5. untuk yang lain
 - Atribut ← yang mengklasifikasikan sampel dengan hasil terbaik (berdasarkan rasio keuntungan)
 - Atribut decision untuk simpul akar ← A untuk setiap nilai V_i yang dimungkinkan untuk A
 - a) cabang tambahan di bawah root yang terkait dengan A = V_i
 - b) Menentukan sampel S_{vi} sebagai subset dari sampel yang memiliki nilai V_i untuk atribut A
 - c) Jika sampel S_{vi} tidak ada
 - di bawah cabang, tambahkan simpul daun dan label = yang sering
 - untuk yang lain, tambahkan cabang baru di bawah cabang saat ini (Amin et al., 2015).

Performance

Kinerja sistem informasi dapat dievaluasi dengan menghitung akurasi berdasarkan input data pelatihan dan data uji. dalam klasifikasi, pengukuran yang biasa dilakukan adalah presisi, daya ingat (*recall*) dan akurasi. Keputusan untuk menentukan suatu kepastian merupakan klasifikasi biner yang hasilnya bisa didapatkan Ya atau Tidak.

Tabel 1 Performance Sistem

	identifikasi tidak memenuhi syarat	identifikasi memenuhi syarat
tidak memenuhi syarat	a	b
memenuhi syarat	c	d

Sumber : (Amin et al., 2015)

Presisi

Presisi adalah bagian dari data yang diambil sesuai dengan informasi yang diperlukan. Rumus presisi adalah:

$$Precision = \frac{(d)}{(b+d)} \times 100\% \dots\dots\dots (4)$$

Dalam klasifikasi biner, presisi dapat dianggap setara dengan nilai prediksi positif.

Recall

Recall adalah pengumpulan data yang berhasil diambil dari bagian data yang relevan dengan queri. rumus recall adalah:

$$Recall = \frac{(d)}{(c+d)} \times 100\% \dots\dots\dots (5)$$

Accuracy

Akurasi adalah persentase dari total data uji yang diidentifikasi dengan benar. Rumus *accuracy* adalah:

$$Accuracy = \frac{(a+d)}{Totalsample} \times 100\% \dots\dots\dots (6)$$

HASIL DAN PEMBAHASAN

Objek permasalahan yang penulis angkat pada masalah ini adalah dataset yang diperoleh dari hasil kampanye pemasaran langsung dari lembaga perbankan di Portugis. Kampanye pemasaran didasarkan pada panggilan telepon. Seringkali, lebih dari satu kontak ke klien yang sama diperlukan, untuk mengakses jika produk (deposito berjangka bank) akan ('ya') atau tidak ('tidak') berlangganan.

Pada pembahasan ini penulis tertarik untuk membuat kalsifikasi terhadap pelanggan yang memilih ya dan tidak untuk mengikuti deposito jangka panjang serta melakukan analisa terhadap faktor-faktor yang dapat mempengaruhi keputusan pelanggan dalam memilih atau menyetujui mengikuti deposito berjangka panjang.

Variabel input dari dataset yang diambil sebanyak 9 attribut yaitu, attribute kategori usia, penghasilan, pendidikan, status pernikahan, pinjaman di Bank, pinjaman rumah, type contac yang dihubungi, hasil kampanye sebelumnya dan keputusan Ya atau Tidak yang menjadi *variable class* untuk memiliki deposito berjangka. untuk lebih jelasnya dijelaskan pada tabel dibawah ini.

Tabel 2 Data Training

No	Kategori usia	Penghasilan	Pendidikan	Status Pernikahan	Pinjaman Bank	Pinjaman Rumah	Typ contact	Kampanye sebelumnya	Keputusan
1	Produktif	Middle	Primary	Menikah	Y	Y	Celluler	Sukses	Y
2	Produktif	Middle	Secondary	Singel	Y	Y	Celluler	Gagal	T
3	Produktif	Low	Tertiary	Bercerai	Y	T	unknow n	Tidak diket ahui	T
4	Non Produktif	Low	Secondary	Bercerai	T	T	unknow n	Tidak diket ahui	T
...									
248	Non produktif	Middle	Secondary	Menikah	T	T	Celluler	Gagal	T
249	Produktif	Hight	Primary	Singel	T	T	Celluler	Sukses	Y
250	Produktif	Hight	Primary	Menikah	T	Y	tele phone	sukses	Y

Sumber: (Nawawi, Rahayu, Shidiq, & Purnama, 2018)

Setelah mendapatkan *data training* maka langkah selanjutnya adalah menghitung nilai Entropy dan nilai Gain dari masing-masing variable inputan.

Menghitung Nilai Entropy dari data training diketahui jumlah kasus sebanyak 250 data, yang menyatakan Ya untuk deposito sebanyak 30 data dan yang menyatakan Tidak sebanyak 220. Maka didapat entropy nilai akarnya totalnya:

$$Entropy (S) = - \sum_{i=1}^n - pi * \log 2pi$$

$$=(-30/250 * \log2(30/250)) + (-220/250 * \log2(220/250))$$

$$= -0,2372$$

Menghitung nilai Entropy dari data kategori usia produktif yang diketahui jumlah kasus 221, yang menyatakan Ya ada 23 dan yang Tidak ada 198, sehingga didapat nilai produktifnya :

$$\begin{aligned} \text{Entropy (S)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-23/221 * \log_2(23/221)) + (-198/221 * \log_2 \\ & \quad (198/221)) \\ &= 0,4818 \end{aligned}$$

Menghitung nilai Entropy dari data kategori usia Non Produktif yang diketahui jumlah kasus ada 29, yang menyatakan Ya ada 7 dan yang Tidak ada 22, sehingga nilai yang didapat untuk kategori Non Produktif adalah:

$$\begin{aligned} \text{Entropy (S)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-7/29 * \log_2 (7/29)) + (-22/29 * \log_2 (22/29)) \\ &= 0,7973 \end{aligned}$$

Menghitung nilai gain untuk atribut kategori usia, dengan nilai gain yang didapat sebagai berikut:

$$\begin{aligned} \text{Gain (S,A)} &= \text{Entropy(S)} - \sum_{i=0}^n s_i/s * \text{entropy (S}_i) \\ \text{Gain (S,A)} &= -0,2372 - ((221/250(0,4818) + \\ & \quad 29/250(0,7973)) \\ &= -0.75558 \end{aligned}$$

Menghitung nilai Entropy dari kategori data pekerjaan Rendah yang diketahui jumlah datanya ada 87, data yang memilih Ya ada 10 dan yang tidak ada 77, maka nilai entropy nya didapat sebagai berikut :

$$\begin{aligned} \text{Entropy (S)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-10/87 * \log_2 (10/87)) + (-77/87 * \log_2 \\ & \quad (77/87)) \\ &= 0,5146 \end{aligned}$$

Menghitung nilai Entropy dari data kategori pekerjaan Medium, Jumlah datanya ada 115, yang menyatakan Ya ada 13 dan yang menyatakan Tidak ada 102, maka nilai Entropy nya didapat sebagai berikut :

$$\begin{aligned} \text{Entropy (S)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-13/115 * \log_2 (13/115)) + (-102/115 * \log_2 \\ & \quad (102/115)) \\ &= 0,5090 \end{aligned}$$

Menghitung nilai Entropy dari data kategori pekerjaan Tinggi, Jumlah datanya ada 48, yang menyatakan Ya ada 7 dan yang menyatakan Tidak ada 41, maka nilai Entropy nya didapat sebagai berikut :

$$\begin{aligned} \text{Entropy (S)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-7/48 * \log_2 (7/48)) + (-41/48 * \log_2 (41/48)) \\ &= 0,5993 \end{aligned}$$

Menghitung nilai gain untuk atribut kategori Pekerjaan, dengan nilai gain yang didapat sebagai berikut:

$$\begin{aligned} \text{Gain (S,A)} &= \text{Entropy(S)} - \sum_{i=0}^n s_i/s * \text{entropy (S}_i) \\ \text{Gain (S,A)} &= -0,2372 - ((87/250(0,05146) + \\ & \quad 115/250 (0,5090) + ((48/250) (0,5993)) \\ &= -0,06708 \end{aligned}$$

Menghitung nilai Entropy dari data kategori Status Pernikahan Single, Jumlah datanya ada 78, yang menyatakan Ya ada 11 dan yang menyatakan Tidak ada 67, maka nilai Entropy nya didapat sebagai berikut :

$$\begin{aligned} \text{Entropy (S)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-11/78 * \log_2(11/78)) + (-67/78 * \log_2 (67/78)) \\ &= 0,5869 \end{aligned}$$

Menghitung nilai Entropy dari data kategori Status Pernikahan Menikah, Jumlah datanya ada 146, yang menyatakan Ya ada 14 dan yang menyatakan Tidak ada 132, maka nilai Entropy nya didapat sebagai berikut :

$$\begin{aligned} \text{Entropy (S)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-14/146 * \log_2(14/146)) + (-132/146 * \log_2 \\ & \quad (132/146)) \\ &= 0.4558 \end{aligned}$$

Menghitung nilai Entropy dari data kategori Status Pernikahan Bercerai, Jumlah datanya ada 26, yang menyatakan Ya ada 5 dan yang menyatakan Tidak ada 21, maka nilai Entropy nya didapat sebagai berikut :

$$\begin{aligned} \text{Entropy (S)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-5/26 * \log_2 (5/26)) + (-21/26 * \log_2 (21/26)) \\ &= 0,7063 \end{aligned}$$

Menghitung nilai gain untuk atribut kategori Pernikahan, dengan nilai gain yang didapat sebagai berikut:

$$\begin{aligned} \text{Gain (S,A)} &= \text{Entropy(S)} - \sum_{i=0}^n s_i/s * \text{entropy (S}_i) \\ \text{Gain (S,A)} &= -0,2372 - ((78/250 (0,5869) + 146/250 \\ & \quad (0,4558) + ((26/250 (0,7063)) \\ &= -0,29622 \end{aligned}$$

Menghitung nilai Entropy dari data kategori Pendidikan Primary, Jumlah datanya ada 33, yang menyatakan Ya ada 4 dan yang menyatakan Tidak ada 29, maka nilai Entropy nya didapat sebagai berikut :

$$\begin{aligned} \text{Entropy (S)} &= \sum_{i=1}^n - p_i * \log_2 p_i \\ &= (-4/33 * \log_2(4/33)) + (-29/33 * \log_2 (29/33)) \\ &= 0,5328 \end{aligned}$$

Menghitung nilai Entropy dari data kategori Pendidikan Secondary, Jumlah datanya ada 130, yang menyatakan Ya ada 13 dan yang menyatakan

Tidak ada 117, maka nilai Entropy nya didapat sebagai berikut :

$$\text{Entropy (S)} - \sum_{i=1}^n - p_i * \log 2p_i$$

$$= (-13/130 * \log_2(13/130)) + (-117/130 * \log_2(117/130))$$

$$= 0,4690$$

Menghitung nilai Entropy dari data kategori Pendidikan Tertiary, Jumlah datanya ada 76, yang menyatakan Ya ada 12 dan yang menyatakan Tidak ada 64, maka nilai Entropy nya didapat sebagai berikut :

$$\text{Entropy (S)} - \sum_{i=1}^n - p_i * \log 2p_i$$

$$= (-12/76 * \log_2(12/76)) + (-64/76 * \log_2(64/76))$$

$$= 0,6292$$

Menghitung nilai Entropy dari data kategori Pendidikan tidak diketahui, Jumlah datanya ada 11, yang menyatakan Ya ada 1 dan yang menyatakan Tidak ada 10, maka nilai Entropy nya didapat sebagai berikut :

$$\text{Entropy (S)} - \sum_{i=1}^n - p_i * \log 2p_i$$

$$= (-1/11 * \log_2(1/11)) + (-10/11 * \log_2(10/11))$$

$$= 0,4395$$

Menghitung nilai gain untuk attribut kategori Pendidikan, dengan nilai gain yang didapat sebagai berikut:

$$\text{Gain (S,A)} = \text{Entropy(S)} - \sum_{i=0}^n s_i/s * \text{entropy (Si)}$$

$$\text{Gain (S,A)} = -0,2372 - ((33/250 (0,5328) + 130/250 (0,4690) + ((76/250(0,6292)) + ((11/250 (0,4395))$$

$$= -0,146972$$

Menghitung nilai Entropy dari data kategori Memiliki Cicilan Rumah, Jumlah datanya ada 139, yang menyatakan Ya ada 11 dan yang menyatakan Tidak ada 128, maka nilai Entropy nya didapat sebagai berikut :

$$\text{Entropy (S)} - \sum_{i=1}^n - p_i * \log 2p_i$$

$$= (-11/139 * \log_2(11/139)) + (-128/139 * \log_2(128/139))$$

$$= 0,3991$$

Menghitung nilai Entropy dari data kategori Tidak Memiliki Cicilan Rumah, Jumlah datanya ada 111, yang menyatakan Ya ada 19 dan yang menyatakan Tidak ada 92, maka nilai Entropy nya didapat sebagai berikut :

$$\text{Entropy (S)} - \sum_{i=1}^n - p_i * \log 2p_i$$

$$= (-19/111 * \log_2(19/111)) + (-92/111 * \log_2(92/111))$$

$$= 0,6604$$

Menghitung nilai gain untuk attribut kategori Memiliki Cicilan Rumah, dengan nilai gain yang didapat sebagai berikut:

$$\text{Gain (S,A)} = \text{Entropy(S)} - \sum_{i=0}^n s_i/s * \text{entropy (Si)}$$

$$\text{Gain (S,A)} = -0,2372 - ((139/250 (0,3991) + 111/250 (0,6604)$$

$$= -0,16591$$

Selanjutnya untuk menentukan nilai Entropy dan Gain dari kategori Cicilan Bank, Contac yang dihubungi dan hasil kampanye sebelumnya dapat dilakukan dengan cara dan rumus yang sama seperti menentukan nilai Entropy diatas. Maka hasil dari keseluruhannya bisa dilihat dari tabel dibawah ini secara rinci.

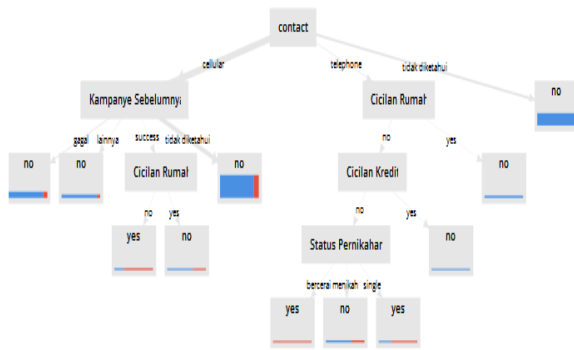
Tabel 3 Nilai Entropy dan Gain

	Data	Ya	Tidak	Pi (yes)	Log 2 Pi	Pj (No)	Log 2 Pj	Entropy	Gain
Akar Total	250	330	20	1,32	0,4005	0,0800	3,6439	0,2372	
Usia									
Produktif	221	23	198	0,10407	3,2643	0,8959	0,1585	0,4818	
Non Produktif	297	7	22	0,24138	2,0506	0,7586	0,3985	0,7973	
									0,7558
Kategori Pekerjaan									
Rendah	87	10	77	0,11494	3,1210	0,8851	0,1762	0,5146	
Medium	115	13	102	0,11304	3,1451	0,8870	0,1731	0,5090	
Tinggi	487	7	41	0,14583	2,7776	0,8542	0,2274	0,5993	
									0,0678
Status Pernikahan									
Single	78	11	67	0,14103	2,8260	0,8590	0,2193	0,5869	
Menikah	146	14	132	0,09589	3,3825	0,9041	0,1454	0,4558	
Bercerai	265	5	21	0,19231	2,3785	0,8077	0,3081	0,7063	
									0,296222
Pendidikan									
Primary	33	4	29	0,12121	3,0444	0,8788	0,1864	0,5328	
Secondary	130	13	117	0,10000	3,3219	0,9000	0,1520	0,4690	
Tertiary	76	12	64	0,15789	2,6630	0,8421	0,2479	0,6292	
Tidak Diketahui	11	1	10	0,09091	3,4594	0,9091	0,1375	0,4395	
									0,146972
Cicilan Rumah									
Ya	139	11	128	0,07914	3,6595	0,9209	0,1189	0,3991	
Tidak	11	19	92	0,17117	2,5465	0,8288	0,2709	0,6604	
									0,16591
Cicilan Bank									
Ya	43	1	42	0,02326	5,4263	0,9767	0,0339	0,1594	

Tidak	20			0,140	2,83	0,85	0,21	0,584
	7	29	178	10	55	99	78	5
								0,2193
								49
Contac								
Cellular	16			0,136	2,86	0,86	0,21	0,576
	8	23	145	90	88	31	24	1
Telephone	19	6	13	0,315	1,66	0,68	0,54	0,899
				79	30	42	75	7
Tidak Diketahui	63	1	62	0,015	5,97	0,98	0,02	0,117
				87	73	41	31	6
								0,5263
								1
Kampanye Sebelumnya								
Sukses	9	5	4	0,555	0,84	0,44	1,16	0,991
				56	80	44	99	1
Gagal	32	12	20	0,375	1,41	0,62	0,67	0,954
				00	50	50	81	4
Lainnya	13	1	12	0,076	3,70	0,92	0,11	0,391
				92	04	31	55	2
Tidak Diketahui	19	6	20	0,102	3,29	0,89	0,15	0,475
				04	28	80	53	4
								0,2423
								7

Sumber: (Nawawi, Rahayu, Shidiq, & Purnama, 2018)

Dari tabel 3 yang telah dijabarkan diatas dapat dilihat dari nilai gain yang paling tinggi adalah contac yang dihubungi, contac yang dihubungi memiliki nilai gain sebesar -0,52631 → (0,52631), maka dapat disimpulkan dan digambarkan bahwa contac yang dihubungi sebagai akar dari pohon keputusan dengan algoritma *decision tree* C4.5 sebagai berikut.



Sumber: (Nawawi, Rahayu, Shidiq, & Purnama, 2018)

Gambar 1 Pohon keputusan dari hasil pengolahan dengan Algoritma C4.5

Berdasarkan gambar diatas pada pohon keputusan dapat diambil daftar aturan dari pohon keputusan yaitu:

1. Jika Nasabah dihubungi melalui contac celluler dan kampanye sebelumnya sukses dan tidak memiliki cicilan rumah maka keputusannya adalah Ya akan melakukan deposito
2. Jika Nasabah dihubungi melalui contac celluler dan kampanye sebelumnya gagal, tidak diketahui atau lainnya maka keputusannya adalah Tidak.

3. Jika nasabah dihubungi melalui nomor telephone, tidak memiliki cicilan rumah, tidak memiliki cicilan bank dan statusnya single atau bercerai maka keputusannya adalah Ya
4. Jika nasabah yang dihubungi melalui nomor telephone, tidak memiliki cicilan rumah, tidak memiliki cicilan bank dan statusnya menikah maka keputusannya adalah Tidak.
5. Jika nasabah yang dihubungi melalui telephone dan memiliki cicilan rumah maka keputusannya adalah Tidak.
6. Jika nasabah yang dihubungi melalui telephone dan tidak memiliki cicilan rumah tetapi memiliki cicilan kredit, maka keputusannya adalah Tidak
7. Jika contacnya tidak diketahui maka keputusannya adalah Tidak.

Berdasarkan aturan diatas maka dapat diketahui calon nasabah yang memungkinkan melakukan deposito berjangka adalah nasabah yang dihubungi melalui contac selluler akan mengatakan Ya apabila kampanye sebelumnya sukses dan tidak memiliki cicilan rumah dan nasabah yang dihubungi melalui telephone akan melakukan deposito apabila tidak memiliki cicilan rumah atau cicilan kredit serta status pernikahannya single atau bercerai.

Hasil dari algoritma C4.5 berdasarkan hasil keseluruhan digambarkan dalam algoritma *decision tree* dengan menggunakan *tolls* rapid miner adalah sebagai berikut:

Tree

```

contact = cellular
| Kampanye Sebelumnya = gagal: no {no=25, yes=3}
| Kampanye Sebelumnya = lainnya: no {no=11, yes=1}
| Kampanye Sebelumnya = sukses
| | Cicilan Rumah = no: yes {no=1, yes=3}
| | Cicilan Rumah = yes: no {no=2, yes=1}
| Kampanye Sebelumnya = tidak diketahui: no {no=106, yes=15}
contact = telephone
| Cicilan Rumah = no
| | Cicilan Kredit = no
| | | Status Pernikahan = bercerai: yes {no=0, yes=2}
| | | Status Pernikahan = menikah: no {no=4, yes=2}
| | | Status Pernikahan = single: yes {no=1, yes=2}
| | Cicilan Kredit = yes: no {no=2, yes=0}
| Cicilan Rumah = yes: no {no=6, yes=0}
contact = tidak diketahui: no {no=62, yes=1}
    
```

Sumber : Hasil Penelitian (2019)

Gambar 2 Tampilan Hasil Pengujian Algoritma C.45 dengan Rapid Miner

Performance

Tabel 4 Presisi, Recall dan Accuracy

	true no	true yes	class precision
pred no	211	30	87,55%
pred yes	9	0	0,00%
class recall	95,91%	0,00%	

Sumber: (Nawawi, Rahayu, Shidiq, & Purnama, 2018)

Performance dari pengujian data training adalah
Presisi :

$$\frac{0}{30 + 0} \times 100\% = 0,00\%$$

Recall

$$\frac{0}{9 + 0} \times 100\% = 0,00\%$$

Accuracy

$$\frac{211 + 0}{250} \times 100\% = 84,4\%$$

KESIMPULAN

Berdasarkan hasil penelitian diatas untuk menentukan keputusan pemilihan deposito berjangka yang dilakukan di sebuah bank dari dataset yang diperoleh menunjukkan bahwa pelanggan potensial adalah pelanggan yang dihubungi melalui celluler dan telephon jika status single atau bercerai dan tidak memiliki kredit rumah ataupun kredit bank maka pelanggan tersebut besar kemungkinan untuk mengatakan Ya untuk melakukan Deposito berjangka. Accuracy nilai yang didapatkan dari hasil penelitian terhadap jumlah sample dengan 9 attribut input paling berpengaruh terhadap pengambilan keputusan adalah sebesar 84,4% nilai keakuratannya.

REFERENSI

- Amin, R. K., Indwiarti, & Sibaroni, Y. (2015). Implementation of decision tree using C4.5 algorithm in decision making of loan application by debtor (Case study: Bank pasar of Yogyakarta Special Region). *2015 3rd International Conference on Information and Communication Technology, ICoICT 2015*, 0, 75–80. <https://doi.org/10.1109/ICoICT.2015.7231400>
- Bao, X., & Guan, X. (2016). A Method of Predicting Crude Oil Output Based on RS-C4.5 Algorithm. *Proceedings - 2016 3rd International Conference on Information Science and Control Engineering, ICISCE 2016*, (3), 63–66. <https://doi.org/10.1109/ICISCE.2016.24>
- Dai, W., & Ji, W. (2014). A mapreduce implementation of C4.5 decision tree algorithm. *International Journal of Database Theory and Application*, 7(1), 49–60.
- Nawawi, H. M., Rahayu, S., Shidiq, M. J., & Purnama, J. J. (2018). *Laporan Akhir Penelitian: Algoritma C4.5 Untuk Memprediksi Pengambilan Keputusan Memilih Deposito Berjangka*. Jakarta.
- Santoso, B. T. (2011). ANALISA DAN PENERAPAN METODE C4.5 UNTUK PREDIKSI LOYALITAS PELANGGAN. *Jurnal Ilmiah Fakultas Teknik*, 10(1), 1–47.
- Widayu, H., Darma, S., Silalahi, N., & Mesran. (2017). Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5. *Issn 2548-8368, Vol 1, No(June)*, 7.
- Wiharto, W., Kusnanto, H., & Herianto, H. (2016). Interpretation of clinical data based on C4.5 algorithm for the diagnosis of coronary heart disease. *Healthcare Informatics Research*, 22(3), 186–195. <https://doi.org/10.4258/hir.2016.22.3.186>
- Wijaya, A. C., Hasibuan, N. A., & Ramadhani, P. (2018). Implementasi Algoritma C5.0 Dalam Klasifikasi Pendapatan Masyarakat (Studi Kasus : Kelurahan Mesjid Kecamatan Medan Kota). *Informasi Dan Teknologi Ilmiah (INTI)*, 13, 192–198.
- Yunus, M., Dahlan, H. S., & Santoso, P. B. (2014). SPK Pemilihan Calon Pendorong Darah Potensial dengan Algoritma C4.5 dan Fuzzy Tahani. *Jurnal EECCIS, Vol. 8 No.(1)*, 47–54.