

Analisis Prediksi Nasabah yang Berpotensi Membuka Deposito pada Bank Umum di Bekasi Menggunakan Algoritma C4.5 dan Naive Bayes

Dwi Handayani dan Nuryuliani

Magister Manajemen Sistem Informasi, Universitas Gunadarma

Jl. Margonda Raya No. 100, Depok 16424

E-mail : handayani.dwi57@yahoo.com, nryulia@staff.gunadarma.ac.id

Abstrak

Salah satu jasa perbankan yang ditawarkan oleh bank umum adalah deposito. Deposito masih merupakan primadona bagi masyarakat yang ingin berinvestasi. Marketing harus mengambil kesempatan ini dengan melakukan strategi pemasaran langsung yang lebih efektif dan efisien, salah satu cara yang dapat digunakan yaitu memprediksi nasabah yang berpotensi membuka simpanan deposito menggunakan teknik data mining khususnya metode klasifikasi algoritma Naive Bayes dan C4.5. Klasifikasi tersebut menggunakan tools Weka versi 3.8.4.

Data yang digunakan dalam penelitian ini berasal dari data nasabah marketing suatu Bank Umum di Bekasi pada tahun 2018. Data yang diuji menggunakan tujuh atribut sebagai predictor yang terdiri dari Pekerjaan, Status, Pendidikan, KPR, *Loan*, Kontak, Pemasaran_Sebelumnya, dan satu atribut sebagai penentu predictor yaitu Pembukaan_Deposito yang berjumlah 9374 *record* data dengan komposisi data training 80% berjumlah 7499 *record* dan data testing 20% berjumlah 11875 *record*.

Hasil teknik klasifikasi dengan algoritma C4.5 pada penelitian ini menghasilkan persentase keakuratan prediksi sangat baik pada nasabah yang berpotensi membuka deposito dengan nilai *accuracy* berjumlah 91,9467%, *precision* berjumlah 92,40%, *recall* berjumlah 91,90%, dan *f-measure* berjumlah 92% dibandingkan dengan algoritma Naive Bayes dengan nilai *accuracy* berjumlah 89,8133%, *precision* berjumlah 90,2%, *recall* berjumlah 89,8%, dan *f-measure* berjumlah 89,8% . Melihat nilai *accuracy* hampir mendekati 100%, maka teknik klasifikasi dengan algoritma C4.5 merupakan kategori klasifikasi sangat baik digunakan untuk memprediksi nasabah yang berpotensi membuka deposito.

Kata Kunci: Metode Klasifikasi, C4.5, Naive Bayes, Weka, Deposito

Pendahuluan

Salah satu dalam jasa perbankan yang ditawarkan oleh salah satu bank umum adalah deposito. Menurut Statistik perbankan Indonesia OJK, deposito masih merupakan primadona bagi masyarakat yang ingin berinvestasi namun dengan sedikit resiko. Hal ini bisa dilihat dari keunggulan nominal deposito dibandingkan dengan nilai giro dan tabungan yaitu deposito pada bulan Juni 2019 senilai Rp. 2.566.756 Milyar, giro senilai Rp. 1.406.536 Milyar, dan tabungan senilai Rp.1.826.202 Milyar [1]. Hal itu dikarenakan deposito memiliki tingkat suku bunga per tahun yang diterima lebih tinggi dibandingkan dengan suku bunga giro atau simpanan tabungan biasa. Untuk itu, suatu bank harus mengambil kesempatan tersebut dengan melakukan

pemasaran langsung yang lebih efektif dalam pekerjaan marketing dan efisien waktu, salah satu cara yang dapat digunakan yaitu memprediksi nasabah yang berpotensi membuka simpanan deposito pada bank tersebut. Prediksi tersebut dapat digunakan dengan menggunakan data-data nasabah yang sudah ada lalu diproses sehingga menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik Data Mining [2].

Dalam proses data mining terdapat tahap yang harus dilakukan untuk pengolahan suatu data, yaitu tahap data training dan data testing. Terdapat beberapa algoritma klasifikasi data mining yang dapat digunakan untuk strategi pemasaran dan promosi, seperti yang ditulis oleh Sergio Moro

dan Raul M.S. Laureano diantaranya Naive Bayes (NB), Decision Trees (DT) dan Support Vector Machines (SVM) [3].

Pada penelitian ini menggunakan data nasabah marketing pada suatu Bank Umum di daerah Bekasi untuk diuji menggunakan metode klasifikasi algoritma C4.5 dan Naive Bayes.dengan tools Weka Versi 3.8.4. Output dalam tools Weka menghasilkan Confusion Matrix. Evaluasi dengan menggunakan Confusion Matrix menghasilkan nilai Precision, *Recall*, F-Measure dan *Accuracy*.

Algoritma C4.5 merupakan pengembangan dari algoritma konvensional induksi pohon keputusan yaitu ID3. Algoritma yang merupakan pengembangan dari ID3 ini dapat mengklasifikasikan data dengan metode pohon keputusan yang memiliki beberapa kelebihan. Adapun kelebihanannya dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah di interpretasikan, dan tercepat diantara algoritma-algoritma yang menggunakan memori utama di komputer [4]. Naive Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema atau aturan bayes dengan asumsi independensi yang kuat pada fitur, artinya bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama [5]. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian [6].

Hasil teknik klasifikasi dengan algoritma C4.5 pada penelitian ini menghasilkan persentase keakuratan prediksi sangat baik pada nasabah yang berpotensi membuka deposito dengan memiliki nilai (data testing) *accuracy* prediksi sebesar 91,9467%, precision sebesar 92,40%, *recall* sebesar 91,90%, dan f-measure sebesar 92%. Pada penelitian ini metode klasifikasi dengan algoritma C4.5 dipilih untuk memperoleh keakuratan dalam memprediksi nasabah yang berpotensi membuka deposito dibandingkan metode klasifikasi dengan algoritma Naive Bayes.

Penelitian Terdahulu

Penelitian terdahulu menjadi salah satu acuan dalam melakukan penelitian sehingga dapat memperkaya teori yang digunakan dalam mengkaji penelitian yang akan dilakukan.. Berikut penelitian terdahulu yang pertama oleh Wachid Darmawan menggunakan algoritma Decision Tree, KNN, dan Naive Bayes. Penelitian ini menyimpulkan bahwa algoritma Decision Tree mendapatkan *accuracy* yang paling baik yaitu sebesar 88.48% dibandingkan algoritma KNN, dan Naive Bayes [7].

Penelitian kedua oleh Reny Wijayanti dan Sulastri menggunakan algoritma Naive Bayes dengan

membandingkan nilai *accuracy* pada jumlah variabel yang berbeda. Penelitian ini menyimpulkan bahwa dengan menggunakan 3 variabel mempunyai nilai *accuracy* data paling tinggi yaitu 80,59% dibandingkan menggunakan 23 variabel, 11 variabel, dan 6 variabel [8].

Penelitian ketiga oleh Yuni Eka Achyani menggunakan algoritma SVM. Penelitian ini menyimpulkan bahwa pengujian model dengan menggunakan algoritma SVM dapat digunakan secara tepat dan akurat untuk prediksi pemasaran langsung dengan nilai *accuracy* adalah 88.71% dan nilai AUC adalah 0.896 [9].

Penelitian keempat oleh Larissa Navia Rani menggunakan algoritma C4.5. Penelitian ini menyimpulkan bahwa algoritma C4.5 dianggap sebagai algoritma yang sangat membantu dalam melakukan klasifikasi dalam bentuk struktur pohon keputusan (decision tree) maupun dalam bentuk aturan atau rule If – Then [10].

Penelitian kelima oleh Harry Dhika dan Fitriana Destiawati menggunakan algoritma C4.5. Penelitian ini menyimpulkan bahwa Penerapan algoritma C4.5 memiliki tingkat *accuracy* yang termasuk dalam klasifikasi sangat baik yakni 97,71% [11].

Penelitian keenam oleh Ahmad Asifuddin Aqham dan Kristoko Dwi Hartomo menggunakan algoritma Naive Bayes dan Genetika dengan feature selection. Penelitian ini menyimpulkan bahwa algoritma Naive Bayes mendapatkan *accuracy* sebesar 86,71% sedangkan algoritma Naive Bayes menggunakan algoritma Genetika dengan feature selection mendapatkan *accuracy* sebesar 90,27% [12].



Gambar 1: Kerangka Kerja Penelitian

Metode Penelitian

Pada metode penelitian ini diuraikan dalam kerangka kerja penelitian untuk memberikan gambaran yang jelas, teratur, dan sistematis, dapat dilihat pada Gambar 1.

Kerangka kerja penelitian ini menggunakan tahapan Cross Industry Standard Process for Data Mining (CRISP-DM) [13]. Terdapat tiga tahapan penelitian, pertama merupakan tahapan pengumpulan data. Tahapan kedua adalah tahapan persiapan data. Proses pada tahapan ini merupakan bagian dari Metodologi CRISP-DM. Proses

pada tahapan ini terdiri dari pemahaman bisnis (Business Understanding), pemahaman data (Data Understanding), dan proses persiapan data (Data Preparation). Tahapan ketiga merupakan tahapan Klasifikasi dan Evaluasi. Proses pada tahapan ini juga merupakan bagian dari metodologi CRISP-DM. Proses pada tahapan ini terdiri dari proses pemodelan (Model Building), proses validasi dan evaluasi (Validation and Evaluation) proses penyebaran (Deployment).

1. Tahap Pengumpulan Data. Pengumpulan data yang diambil dari suatu Bank Umum di Bekasi dari data nasabah marketing pada Tahun 2018 berjumlah 39.260 data.

2. Tahapan Persiapan Data :

(a) Pemahaman Bisnis (Business Understanding). Pemahaman Bisnis berisi tentang seorang marketing di suatu bank umum daerah bekasi menawarkan produk deposito kepada seluruh nasabah tanpa memprediksi nasabah yang berpotensi membuka deposito, sehingga dirasakan kurang tepat sasaran dalam menentukan langkah dan strategi bagi para nasabah marketing tersebut.

(b) Pemahaman Data (Data Understanding). Data mentah yang digunakan dalam penelitian ini berasal dari marketing suatu Bank Umum di Bekasi yaitu data nasabah marketing pada tahun 2018. Data nasabah marketing terdiri dari 11 atribut. Atribut tersebut hasil penggabungan dari 4 tabel yaitu tabel data nasabah (nasabah yang hanya mempunyai produk buku tabungan), tabel data nasabah kredit, tabel data nasabah marketing, dan tabel data nasabah deposito. Jenis atribut tersebut yaitu No_Cif, No_Rek, Nama, Pekerjaan, Status, Pendidikn, KPR, Loan, Kontak, Pemasaran_Sebelumnya, dan Pembukaan_deposito.

(c) Proses Persiapan Data (Data Preparation). Proses persiapan data (Data Preparation) yang berisi pre-processing data, transformasi data, dan pembagian data. Data pre-processing yang dilakukan melalui langkah pembersihan data (cleaning) untuk menghapus atribut-atribut yang tidak relevan (tidak digunakan dalam pengujian klasifikasi) seperti atribut Nama, No_CIF, dan No_Rekening. Proses cleaning data juga menghapus beberapa *record* yang kosong atau bernilai null dari kolom pekerjaan, pendidikan, kontak, dan kolom pemasaran_sebelumnya. Selanjutnya Transformasi data yang dilakukan yaitu

menginisialisaasi data ke dalam bentuk yang dapat dan mudah digunakan pada data mining tools WEKA untuk mengidentifikasi pola dengan teknik Klasifikasi data mining. Selanjutnya melakukan pembagian data training 80% dan data testing 20% dari data nasabah marketing yang sudah dilakukan proses *cleaning*.

3. Tahapan Klasifikasi dan Evaluasi.

(a) Proses Pemodelan (Model Building).



Gambar 2: Proses Permodelan

Pada Gambar 2 Proses Pemodelan menjelaskan cara pembentukan dataset nasabah marketing yang kemudian diolah menjadi data yang siap digunakan untuk proses klasifikasi data mining menggunakan tools Weka (melakukan pembentukan data training). Setelah data tersebut siap untuk diuji, data training dan data testing pada data training di uji menggunakan metode klasifikasi algoritma C4.5 dan Naive Bayes pada tools weka dan menghasilkan Confusion Matrix. Dari Confusion Matrix bisa mendapatkan nilai precision, *recall*, *accuracy*, dan f-measure.

(b) Proses Validasi dan Evaluasi (Validation and Evaluation). Dalam tahapan proses ini akan dilakukan validasi serta pengukuran keakuratan hasil yang dicapai. Sebagai alat pengukuran penelitian ini akan menggunakan tools Weka. Tahap fase evaluasi dilakukan untuk mendapatkan kualitas dan efektifitas sebelum disebarkan. Apakah model yang digunakan sudah tepat dan memenuhi tujuan awal dalam memecahkan permasalahan pada penelitian ini, serta mengambil keputusan berkaitan dengan penggunaan hasil dari data mining.

(c) Proses Penyebaran (Deployment). Fase penyebaran yaitu hasil dari penelitian ini dapat dijadikan rekomendasi oleh pihak perusahaan dalam menentukan strategi bisnis dalam memasarkan produk deposito kepada nasabah sehingga lebih efektif dan efisien dalam memprediksi serta

mendapatkan nasabah yang berpotensi membuka deposito.

Langkah – langkah Algoritma C4.5

Adapun langkah – langkah dalam penyelesaian Algoritma C4.5 adalah:

1. Pilih Atribut Sebagai Akar didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada.
2. Bentuk Cabang Untuk Masing – Masing Nilai
3. Bagi Kasus Dalam Cabang
4. Ulangi Proses untuk masing – masing cabang sampai semua kasus pada cabang memiliki kelas yang sama [14].

Adapun langkah – langkah dalam penyelesaian Algoritma C4.5 adalah 1. Pilih Atribut Sebagai Akar didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. 2. Bentuk Cabang Untuk Masing – Masing Nilai 3. Bagi Kasus Dalam Cabang 4. Ulangi Proses untuk masing – masing cabang sampai semua kasus pada cabang memiliki kelas yang sama [14].

1. Hitung Entropy dari setiap atribut dengan menggunakan rumus (1) , sebagai berikut :

$$Entropy[S] = -p + \log_2 p + -p \cdot \log_2 p \quad (1)$$

Dengan keterangan sebagai berikut :

S = ruang data sampel yang digunakan untuk training.

P+ = jumlah yang bersolusi positif yang mendukung pada data sampel untuk kriteria tertentu.

P- = jumlah yang bersolusi negatif yang tidak mendukung pada data sampel untuk kriteria tertentu.

2. Information Gain dari setiap atribut dengan menggunakan rumus (2), sebagai berikut :

$$Gain[S, A] = Entropy[S] - \sum_{v \in \text{value}[Sv] / |S|} * Entropy[Sv] \quad (2)$$

Dengan keterangan sebagai berikut:

S = ruang data sampel yang digunakan untuk training.

A = atribut.

V = nilai yang mungkin untuk atribut A.

|Sv| = jumlah sampel untuk nilai V.

|S| = jumlah seluruh sampel data.

Entropy [Sv] = entropy untuk sampel-sampel yang memiliki nilai V.

Nilai[A] = himpunan nilai-nilai yang mungkin untuk atribut A.

3. Algoritma C4.5 menghitung gain ratio untuk masing-masing atribut, dan atribut yang memiliki nilai yang tertinggi akan dipilih sebagai simpul. Penggunaan gain ratio ini memperbaiki kelemahan dari ID3 yang menggunakan informasi gain. Gain Ratio dari setiap atribut menggunakan rumus (3), sebagai berikut:

$$Gainratio[S, A] = Gain[S, A] / SplitInfo[S, A] \quad (3)$$

Tujuan dari perhitungan Information Gain dan Split Info adalah untuk mendapatkan nilai Gain Ratio. Atribut pada Decision tree dapat ditentukan berdasarkan nilai Gain Ratio.

Langkah – langkah Algoritma Naive Bayes

Adapun tahapan algoritma Naive Bayes adalah:

1. Menghitung jumlah kelas/label.
2. Menghitung Jumlah Kasus Per Kelas
3. Kalikan Semua Variable Kelas
4. Bandingkan Hasil Per Kelas

Langkah kerja dalam perhitungan Algoritma Naive Bayes [15]:

1. Pembacaan data training dengan melihat data yang digunakan berbentuk numerik atau tidak.
2. Hitung jumlah dan probabilitas, namun apabila data numerik maka:

Cari nilai mean dan standar deviasi dari masing – masing parameter yang merupakan data numerik. Adapun persamaan yang digunakan untuk menghitung nilai rata – rata (mean) menggunakan rumus

(4), sebagai berikut:

$$\mu = (x_1 + x_2 + x_3 + \dots + x_n) / n \quad (4)$$

Keterangan:

μ = rata – rata hitung (mean)

xi = nilai sampel ke-i

n = jumlah sampel

3. Cari nilai probabilistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
4. Mendapatkan nilai dalam tabel mean, standar deviasi dan probabilitas

Confusion Matrix

Hasil penelitian dari metode klasifikasi menggunakan tools Weka menghasilkan confusion matrix. Dalam proses evaluasi hasil penelitian akan digunakan confusion matrix. Evaluasi dengan menggunakan confusion matrix menghasilkan nilai *accuracy*, *precision*, dan *recall*. Menurut Ian H. Witten, Dapat dijelaskan juga bahwa confusion matrix merupakan sebuah metode untuk evaluasi yang menggunakan tabel matrix seperti pada tabel 1 [16]:

Tabel 1: *Confusion Matrix*

Correct Classification	Classified as	
	(+)	(-)
(+)	True positive (TP)	False Negative (FN)
(-)	False Positive (FP)	True Negative (TN)

Menurut Kurniawan, *Recall* adalah perbandingan jumlah dokumen relevan yang terambil sesuai dengan query yang diberikan dengan jumlah kumpulan dokumen yang relevan dengan query. *Precision* adalah perbandingan jumlah dokumen yang relevan terhadap query dengan jumlah dokumen yang terambil dari hasil pencarian [17]. Menurut David M W Powers, *Accuracy* didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual [18]. Berdasarkan tabel 1 confusion matrix didapatkan rumus *precision*, *recall*, dan *accuracy* sebagai berikut :

1. Rumus penilaian *precision* yang digunakan dalam penelitian ini, lihat persamaan (5):

$$Precision = TP / [TP + FP] \quad (5)$$

2. Rumus penilaian *recall* yang digunakan dalam penelitian ini, lihat persamaan (6):

$$Recall = TP / [TP + FN] \quad (6)$$

3. Rumus penilaian *accuracy* yang digunakan dalam penelitian ini, lihat persamaan (7):

$$Accuracy = (TP + TN) / (TP + FN + FP + TN) \quad (7)$$

Menurut Powers David M W, F-Measure merupakan salah satu perhitungan evaluasi dalam informasi temu kembali yang mengkombinasikan *recall* dan *precision* [19]. Rumus penilaian f-measure yang digunakan dalam penelitian ini, lihat persamaan (8) :

$$F - Measure = \frac{2(precision * recall)}{(precision + recall)} \quad (8)$$

Hasil Penelitian dan Pembahasan

Data yang digunakan dalam penelitian ini diambil dari data mentah nasabah marketing Bank Umum di Bekasi pada Tahun 2018 berjumlah 39.260 *record* data dengan 11 atribut pada dataset nasabah marketing. Data tersebut dilakukan proses pre-processing data dan transformasi data agar mudah digunakan pada data mining tools Weka untuk mengidentifikasi pola dengan teknik klasifikasi data mining. Proses tersebut menghasilkan dataset nasabah marketing yang berisi 8 atribut dan 9374 *record* data. Data yang diuji menggunakan tujuh atribut sebagai predictor yang terdiri dari Pekerjaan, Status, Pendidikan, KPR, Loan, Kontak, Pemasaran_Sebelumnya, dan satu atribut sebagai penentu predictor yaitu Pembukaan_Deposito. Atribut-atribut tersebut dapat digambarkan pada tabel 2 sebagai berikut:

Tabel 2: Atribut Data Nasabah Marketing

Atribut	Tipe Data	Kategori Data
Pekerjaan	Teks	BUMN, Guru, Logistik, Ojek_Online, Pengajar, Pensiunan, PNS, PRT, Swasta, Usaha_Online, Wiraswasta
Status	Teks	Nikah, Belum_Nikah, Cerai
Pendidikan	Teks	SMA, D3, S1/S2
KPR	Teks	Ada, Tidak_Ada
Loan	Teks	Ada, Tidak_Ada
Kontak	Teks	Telephone, No_HP
pemasaran_sebelumnya	Teks	Sukses, Gagal, Mendekati_iya, Mendekati_gagal
Pembukaan_Deposito	Teks	Y, N

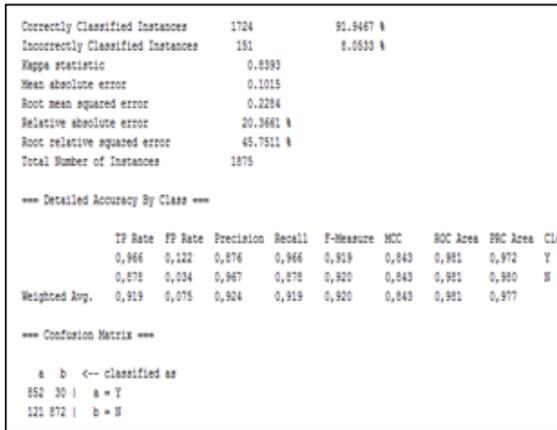
Data yang diuji pada penelitian ini berjumlah 9374 *record* pada dataset nasabah marketing yang akan dibagi menjadi data training 80% berjumlah 7499 *record* dan data testing 20% berjumlah 11875 *record*. Dapat dilihat pada Tabel 3.

Tabel 3: Jumlah *Record* Data yang Digunakan

Pembukaan_Deposito	Training	Testing
Y	3959	993
N	3540	882

Klasifikasi Data Testing pada Data Training Menggunakan Algoritma C4.5

Penelitian ini menggunakan metode klasifikasi pada Algoritma C4.5 dengan tools Weka untuk menguji prediksi nasabah yang berpotensi membuka deposito dengan menghasilkan nilai *accuracy*, *precision*, *recall*, dan *f-measure*. Berikut hasil uji coba data testing berjumlah 11875 *record* menggunakan algoritma C4.5 dengan tools Weka pada Klasifikasi atribut Pembukaan_Deposito, dapat dilihat pada gambar 3 berikut:



Gambar 3: Tampilan Hasil Klasifikasi Atribut Pembukaan_Deposito pada Data Testing C4.5

Seperti yang dapat dilihat pada gambar 3, klasifikasi algoritma C4.5 pada data testing dengan confusion matrix yang dapat dilihat pada tabel 4, sebagai berikut:

Tabel 4: Confusion Matrix Atribut Pembukaan_Deposito pada Data Testing C4.5

	Y	T
Y	852	30
T	121	872

Berikut ini adalah perhitungan *Precision*, *Recall*, *F-measure*, dan *Accuracy* :

1. Menghitung *precision* menggunakan persamaan (5), sebagai berikut:

$$Precision (y) = 852 / (852+121) = 0,876$$

real (0,876)

$$Precision (n) = 872 / (872+30) = 0,967$$

real (0,967)

$$Average precision = 0,9215$$

real (0,924)

2. Menghitung *recall* menggunakan persamaan (6), sebagai berikut :

$$Recall (y) = 852 / (852+30) = 0,966$$

real (0,966)

$$Recall (n) = 872 / (872+121) = 0,878$$

real (0,878)

$$Average Recall = 0,922$$

real (0,919)

3. Menghitung *f-measure* menggunakan persamaan (8), sebagai berikut:

$$F-measure (y) = (2*(0,876*0,966)) / (0,876+0,966) = 0,919$$

real (0,919)

$$F-measure (n) = (2*(0,967*0,878)) / (0,967+0,878) = 0,920$$

real (0,920)

$$Average F-measure = 0,920$$

real (0,920)

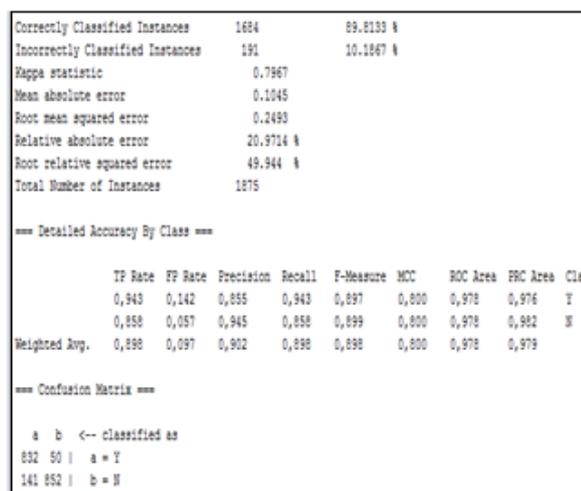
4. Menghitung *Accuracy* menggunakan persamaan (7), sebagai berikut:

$$Accuracy = (852+872) / (852+872+121+30) = 0,919467$$

real (0,919467)

Klasifikasi Data Testing pada Data Training Menggunakan Algoritma Naive Bayes

Penelitian ini menggunakan metode klasifikasi pada Algoritma Naive Bayes dengan tools Weka untuk menguji prediksi nasabah yang berpotensi membuka deposito dengan menghasilkan nilai *accuracy*, *precision*, *recall*, dan *f-measure*. Berikut hasil uji coba data testing berjumlah 11875 *record* menggunakan algoritma Naive Bayes dengan tools Weka pada Klasifikasi atribut Pembukaan_deposito, dapat dilihat pada gambar 4 berikut:



Gambar 4: Tampilan Hasil Klasifikasi Atribut Pembukaan_Deposito pada Data Testing Naive Bayes

Seperti yang dapat dilihat pada gambar 4, klasifikasi algoritma Naive Bayes pada data testing dengan confusion matrix yang dapat dilihat pada tabel 5, sebagai berikut:

Tabel 5: Confusion Matrix Atribut Pembukaan_Deposito pada Data Testing Naive Bayes

Y	T
832	50
141	852

Berikut ini adalah perhitungan *Precision*, *Recall*, *F-measure*, dan *Accuracy*:

1. Menghitung *precision* menggunakan persamaan (5), sebagai berikut:

$$Precision (y) = 832 / (832+141) = 0,855 \text{ real } (0,855)$$

$$Precision (n) = 852 / (852+50) = 0,945 \text{ real } (0,945)$$

$$Average\ precision = 0,900 \text{ real } (0,902)$$

2. Menghitung *recall* menggunakan persamaan (6), sebagai berikut:

$$Recall (y) = 832 / (832+50) = 0,943 \text{ real } (0,943)$$

$$Recall (n) = 852 / (852+141) = 0,858 \text{ real } (0,858)$$

$$Average\ Recall = 0,901 \text{ real } (0,898)$$

3. Menghitung *f-measure* menggunakan persamaan (8), sebagai berikut:

$$F\text{-measure } (y) = (2 * (0,855 * 0,943)) / (0,855 + 0,943) = 0,897 \text{ real } (0,897)$$

$$F\text{-measure } (n) = (2 * (0,945 * 0,858)) / (0,945 + 0,858) = 0,899 \text{ real } (0,899)$$

$$Average\ F\text{-measure} = 0,898 \text{ real } (0,898)$$

4. Menghitung *Accuracy* menggunakan persamaan (7), sebagai berikut:

$$Accuracy = (832+852) / (832+852+141+50) = 0,898133 \text{ real } (0,898133)$$

Analisis Hasil Komparasi

Berdasarkan hasil pengujian terhadap dataset nasabah marketing yang mana berisi info para nasabah yang membuka deposito atau tidak membuka deposito. Dengan menggunakan jumlah persentase data training 80% dan data testing 20%,

maka diperoleh hasil prediksi terbaik dari komparasi perhitungan nilai *precision*, *recall*, *f-measure*, dan *accuracy* untuk metode Klasifikasi C4.5 dan Naive Bayes yang digunakan dan dapat dilihat pada tabel 6 sebagai berikut:

Tabel 6: Hasil Komparasi Nilai *Precision*, *Recall*, *F-measure*, dan *Accuracy* pada Data Training dan Testing

Algoritma	<i>F-measure</i>		<i>Accuracy</i>	
	data training	data testing	data training	data testing
C4.5	92,90%	92%	92,87%	91,95%
Naive Bayes	91,10%	89,80%	91,08%	89,81%

Algoritma	<i>Precision</i>		<i>Recall</i>	
	data training	data testing	data training	data testing
C4.5	93,20%	92,40%	92,90%	91,90%
Naive Bayes	91,40%	90,20%	91,10%	89,80%

Pada tabel 6 menjelaskan bahwa perbandingan *precision*, *recall*, *f-measure*, dan *accuracy* pada setiap metode Klasifikasi C4.5 dan Naive Bayes dari data training dan data testing. Terlihat bahwa pada metode klasifikasi C4.5 memiliki nilai *precision*, *recall*, *f-measure*, dan *accuracy* lebih tinggi dibandingkan Naive Bayes. Selisih dari nilai *precision* dari kedua algoritma pada data training yaitu 1,8% dan pada data testing yaitu 2,2%. Selisih dari nilai *recall* dari kedua algoritma pada data training yaitu 1,8% dan pada data testing yaitu 2,1%. Selisih dari nilai *F-measure* dari kedua algoritma pada data training yaitu 1,8% dan pada data testing yaitu 2,2%. Selisih dari nilai *accuracy* dari kedua algoritma pada data training yaitu 1,79% dan pada data testing yaitu 2,13%.

Dari penjelasan hasil prediksi yang diperoleh maka dapat dibuat penggolongan pengukuran tingkat *accuracy* yang dapat dibagi menjadi beberapa kelompok seperti pada tabel 7 berikut:

Tabel 7: Nilai Kualitas *Accuracy*

Rentang <i>Accuracy</i>	Kualitas <i>Accuracy</i>
0,90 – 1,00	Excellent
0,80 – 0,90	Good
0,70 – 0,80	Fair
0,60 – 0,70	Poor
0,50 – 0,60	Failure

Memperhatikan nilai *accuracy* yang diperoleh pada teknik klasifikasi yang dilakukan, maka sesuai dengan pengelompokan nilai *precision*, *recall*, *f-measure*, dan *accuracy* yang dihasilkan pada penerapan teknik klasifikasi adalah dengan menggunakan metode algoritma C4.5 dalam analisa metode klasifikasi untuk prediksi nasabah yang berpotensi membuka deposito masuk kategori klasifikasi sangat baik.

Penutup

Berdasarkan hasil penelitian yang dilakukan pada data nasabah marketing Bank Umum di Bekasi dengan data training 80% (7499 record data) dan data testing 20% (1875 record data) terhadap tujuh atribut sebagai predictor dan satu atribut sebagai penentu predictor, bahwa teknik klasifikasi dengan algoritma C4.5 merupakan Klasifikasi yang cocok diterapkan untuk memprediksi nasabah yang berpotensi membuka deposito. Hal tersebut terlihat pada algoritma C4.5 mempunyai nilai paling tinggi dengan nilai (data testing) *accuracy* prediksi sebesar 91,9467%, *precision* sebesar 92,40%, *recall* sebesar 91,90%, dan *f-measure* sebesar 92%, dengan begitu tingkat akurasi yang hampir mendekati 1,00 atau sangat valid dibandingkan dengan metode klasifikasi pada algoritma Naive Bayes dengan nilai (data testing) *accuracy* prediksi sebesar 89,8133%, *precision* sebesar 90,20%, *recall* sebesar 89,80%, dan *f-measure* sebesar 89,80%.

Saran yang dapat menjadi masukan bagi penelitian serupa dimasa yang akan datang, diantaranya yaitu: 1. Penelitian ini menggunakan dataset nasabah marketing suatu Bank Umum di Bekasi, untuk penelitian selanjutnya dapat menggunakan dataset yang lainnya, dari dataset private, dataset public, atau dataset bank lainnya untuk menghasilkan model klasifikasi yang lebih baik lagi. 2. Penelitian ini membandingkan antara algoritma Naive Bayes dengan C4.5, untuk penelitian selanjutnya dapat dikembangkan dengan membandingkan algoritma klasifikasi lainnya atau clustering lainnya yang dapat dikombinasikan dengan algoritma optimasi lainnya seperti Particle Swarm Optimization atau feature selection lainnya seperti backward elimination, forward selection, dan stepwise selection, sehingga dapat diketahui algoritma yang lebih baik performanya.

Daftar Pustaka

- [1] OJK., "Statistik Perbankan Indonesia juni 2019", diakses daring pada <https://www.ojk.go.id>, [Accessed 14 Februari 2020].
- [2] Daniel T. Larose, "Discoverin Knowledge In Data", United States of America: John Wiley & Sons, Inc, 2005.
- [3] Sergio Moro and Raul M.S. Laureano, "Using Data Mining for Bank Direct Marketing: An Application of The CRISP-DM Methodology". Lisboa: Instituto Universitário de Lisboa, 2011.
- [4] J.R. Quinlan, "C4. 5: programs for machine learning", Elsevier, 2014.
- [5] IJNS, "Indonesian Journal on Networking and Security", Volume 3 No 2, <http://ijns.org>, SSN: 2302-5700, 2014.
- [6] S. A. Pattekari & A.Parveen, "Prediction System for Heart Disease Using Naive Bayes", International Journal of Advanced Computer and Mathematical Sciences, ISSN 2230-9624, Vol. 3, No 3, Hal 290-294, 2012.
- [7] Wachid Darmawan, "Komparasi Metode Data Mining dalam Memprediksi Nasabah Bank yang Akan Memilih Tabungan Deposito Menggunakan Algoritma Klasifikasi", IC-Tech, Vol XIII(1), 49-55, 2018.
- [8] Reny Wijayanti dan Sulastrri, "Analisa Klasifikasi Kartu Kredit Menggunakan Algoritma Naïve Bayes", Prosiding SINTAK, ISBN: 978-602 8557-20-7 ,443-447, 2018.
- [9] Yuni Eka Achyani, "Prediksi Pemasaran Langsung Menggunakan Metode Support Vector Machine", Jurnal Teknik Komputer, Vol III(2), 1-6, 2017.
- [10] Larissa Navia Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit". Jurnal Invotek Polbeng-Seri Informatika, Vol 1(2), 126-132, 2016.
- [11] Harry Dhika dan Fitriana Destiawati, "Penerapan Algoritma C4.5 Untuk Penilaian Karyawan Pada Restoran Cepat Saji". Seminar Nasional Informatika dan Aplikasinya (SNIA), ISBN: 978-602-50525-0-7, A-55 – A-59, Cimahi, 2017.
- [12] Ahmad Asifuddin Aqham dan Kristoko Dwi Hartomo, "Data Mining untuk Nasabah Bank Telemarketing Menggunakan kombinasi Algoritma Naive Bayes Dan Algoritma Genetik", Jurnal Nasional Informatika dan Teknologi Jaringan, Vol 4(1), 47-56, 2019.
- [13] P. Chapman, dkk, "CRISP-DM v.1.0 Step-by-step Data Mining Guide", SPSS, Inc., USA, 2000.
- [14] D. H. Kamagi & S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa", ULTIMATICS. 6(1), 15-18, 2014.
- [15] A. Saleh, "Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga", 2015.

- [16] Eibe Frank and Ian H. Witten, “ Data Mining: Practical Machine Learning Tools and Techniques, Second Edition”, San Frasco: Morgan Kaufman Publisers, 2005.
- [17] Didik Kurniawan, “Evaluasi sistem temu kembali informasi model ruang vector dengan pendekatan user judgement”, Jurnal Sains MIPA, Vol 16(3), 155–162, Lampung, 2010.
- [18] David M W Powers, “The Problem with Kappa”, Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 345–355, Avignon France, 2012.
- [19] David M W Powers, “Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation”, Journal of Machine Learning Technologies, Vol 2(1), 37–63, South Australia, 2011.

Halaman ini sengaja dikosongkan.