Laman web jurnal: http://ejournal.stikom-db.ac.id/index.php/processor

Processor: Jurnal Ilmiah Sistem Informasi, Teknologi Informasi dan Sistem Komputer

P-ISSN: 1907-6738 | E-ISSN: 2528-0082

Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Naïve Bayes dan ID3

Miryam Clementine Aksama[1], Arum Wahyuniati[2]

[1,2] Fakultas Ekonomi dan Bisnis Program Studi Akuntansi, Universitas Dinamika, Jl. Raya Kedung Baruk No.98, Kedung Baruk, Kec. Rungkut, Kota SBY, Jawa Timur 60298, Indonesia.

ABSTRACT

With the development of the times, the need for life will be higher. The bank provides facilities to facilitate the process. Many banks work hard to meet customer satisfaction to keep customers loyal to the bank. This leads to competition between banks by making various innovations to steal the hearts of customers. Competition becomes an important role for banks because it makes the bank develop and affects its income. The competition also provides consequences for customers who move to other banks. Switching customers to other banks is called Churn because customers decide to leave the bank for other banks. Customer churn should be addressed promptly to avoid a major impact on banks. The more customers churn, the bank must evaluate the services provided. To maintain the customer, the bank manager wants to know what influence makes the customer decide to leave the bank, so it takes churn predictions to know it. In conducting this research, the type of data mining applied is classification because this study has the goal to predict which customers will churn judging from the attributes that have been adjusted. The algorithms applied to the study were Naïve Bayes and ID3. In this study, the algorithm that got the best results was the naïve Bayes algorithm with 10.000 data which was then divided into 2 data with a ratio of 6: 4, 4.000 data which is data testing provides high accuracy results. The result that uses Naïve Bayes is 85.17% and using ID3, the result of the accuracy is 79.17%.

Keywords: Classification, Naïve Bayes, ID3, Customer *Churn*, Bank

ABSTRAK

Dengan berkembangnya zaman, kebutuhan hidup akan semakin tinggi. Bank memberikan fasilitas untuk mempermudah proses tersebut. Banyak bank bekerja keras untuk memenuhi kepuasan pelanggan agar nasabah tetap loyal kepada bank. Hal ini menimbulkan persaingan antar bank dengan melakukan berbagai inovasi untuk mencuri hati nasabah. Persaingan menjadi peran penting untuk bank karena membuat bank menjadi berkembang dan berpengaruh kepada pendapatan milik bank. Persaingan juga memberikan akibat yaitu nasabah yang berpindah ke bank lainnya. Beralihnya nasabah ke bank lainnya ini disebut dengan *Churn* karena nasabah memutuskan untuk meninggalkan bank untuk bank yang lainnya. Customer *Churn* harus ditangani dengan segera untuk menghindari dampak besar untuk bank. Semakin banyak nasabah melakukan *churn*, maka bank harus mengevaluasi layanan yang diberikan. Upaya mempertahankan nasabah tersebut, manajer bank ingin mengetahui pengaruh apa yang membuat nasabah memutuskan untuk meninggalkan bank, sehingga dibutuhkan prediksi *churn* untuk mengetahui hal tersebut. Dalam melakukan penelitian ini, jenis data mining yang diterapkan adalah klasifikasi karena penelitian ini memiliki tujuan untuk memprediksi nasabah mana yang akan melakukan *churn* dilihat dari atribut yang telah disesuaikan. Algoritma yang diterapkan pada penelitian ini adalah Naïve Bayes dan ID3. Pada penelitian ini algoritma yang mendapatkan hasil terbaik adalah algoritma Naïve Bayes dengan 10.000 data yang kemudian dibagi menjadi 2 data dengan ratio 6:4 tersebut, 4.000 data yang merupakan data testing memberikan hasil akurasi yang tinggi. Menggunakan Naïve Bayes, hasil dari akurasinya mendapatkan 85.17% dan menggunakan ID3, hasil dari akurasinya mendapatkan 79.17%.

Kata Kunci: Klasifikasi, Naïve Bayes, ID3, Pelanggan *Churn*, Bank

1. PENDAHULUAN

Dengan berkembangnya zaman, kebutuhan hidup akan semakin tinggi. Terlebih lagi saat rasa ingin memiliki membuat seseorang membutuhkan dan ingin untuk membeli suatu produk tertentu. Bank memberikan fasilitas untuk mempermudah proses tersebut. Banyak bank bekerja keras untuk memenuhi kepuasan pelanggan yang cerdas, sadar harga, dan memiliki informasi produk lain yang banyak didapat dari internet agar tetap loyal kepada bank [1]. Hal ini menimbulkan persaingan antar bank dengan melakukan berbagai inovasi untuk mencuri hati nasabah. Persaingan menjadi peran penting untuk bank karena membuat bank menjadi berkembang dan berpengaruh kepada pendapatan milik bank [2]. Persaingan juga memberikan akibat yaitu nasabah yang berpindah ke bank lainnya.

Beralihnya nasabah ke bank lainnya ini disebut dengan *Churn* karena nasabah memutuskan untuk meninggalkan bank untuk bank yang lainnya. Sehingga dapat disimpulkan bahwa *churn* adalah pemutusan jasa dari suatu bank yang dilakukan oleh nasabah, dikarenakan nasabah merasa bank lain mampu memberikan layanan yang lebih baik [3]. *Customer Churn* harus ditangani dengan segera agar menghindari dampak besar untuk bank [4]. Semakin banyak nasabah melakukan *churn*, maka bank harus mengevaluasi layanan yang diberikan. Sangat jelas bahwa memasukkan nasabah baru akan memakan lebih banyak biaya daripada mempertahankan nasabah yang sudah ada [5]. Upaya mempertahankan nasabah tersebut, manajer bank ingin mengetahui pengaruh apa yang membuat nasabah memutuskan untuk meninggalkan bank, sehingga dibutuhkan prediksi *churn* untuk mengetahui hal tersebut.

Dari penelitian sebelumnya, penelitian tersebut membahas mengenai prediksi *churn* pada kelas pelanggan retail dengan menggunakan analisa komparasi algoritma C4.5 dan Naive Bayes, dan di dalamnya menggunakan metode RFM (*Recency, Frequency, Monetary)* untuk memperoleh kelas pelanggan berdasarkan dari karakter pelanggan yang dimiliki oleh perusahaan retail

ini [6]. Penelitian ini menggunakan data latih meliputi sebanyak 30.765 record untuk *data window* dan sebanyak 17.915 record untuk *forecasting window*. Dapat diperoleh kesimpulan bahwa Naïve Bayes lebih baik dari pada C4.5 dengan hasil akurasi Naïve Bayes yaitu 83.49% dan C4.5 adalah 80.6%.

Pada penelitian selanjutnya yang berkaitan dengan prediksi *customer churn* menggunakan algoritma Naive Bayes yang menggunakan Knowledge Discovery (KDD) dengan pembagian dataset dengan rasio 7:3, 8:2, dan 5:5 menghasilkan akurasi sebesar 83,02%; 82,91%; 82,98% [5]. Dimana hasil akurasi tersebut menunjukkan bahwa 38.551 merupakan hasil prediksi pelanggan dari *non-churn* yang benar dan 9.449 prediksi merupakan hasil prediksi *non-churn* yang salah.

Dari beberapa penelitian yang telah ada dan telah dijelaskan di atas yang menjadi pembeda dengan penelitian terbaru mengenai nasabah bank yang melakukan *churn* bahwa penelitian ini menggunakan data mining dengan menggunakan tahapan SEMMA dengan pembagian dataset menggunakan rasio 6:4 dan menggunakan algoritma Naïve Bayes dan ID3. Peneliti membandingkan metode Naïve Bayes dan ID3 dikarenakan dari beberapa jurnal mengatakan bahwa hasil akurasi milik algoritma ID3 memberikan hasil yang lebih baik dari pada algoritma C4.5 [7].

Dalam penelitian yang pernah ada, peneliti menggunakan naïve bayes dan ID3 sebagai metode klasifikasi untuk menentukan diagnose hipertensi pada manusia, menuliskan bahwa penelitian tersebut memberikan hasil keakuratan bahwa naïve bayes memberikan hasil yang lebih akurat daripada ID3 dan penelitian tersebut masih menggunakan cara manual yang menghabiskan lebih banyak waktu [8]. Pembeda dari jurnal ini dengan jurnal tersebut adalah jurnal ini menggunakan bantuan *tools* RapidMiner sehingga dapat memprediksi lebih banyak data.

Seperti yang kita ketahui di dalam data mining terdapat beberapa metode diantaranya Asosiasi, Klasifikasi, dan Klasterisasi [9]. Dalam melakukan penelitian ini, jenis data mining yang diterapkan adalah klasifikasi karena penelitian ini memiliki tujuan untuk memprediksi nasabah mana yang akan melakukan *churn* dilihat dari atribut yang telah disesuaikan. Algoritma yang diterapkan pada penelitian ini adalah Naïve Bayes dan ID3.

Penerapan data mining pada penelitian ini sangat membantu perbankan untuk menciptakan hubungan dengan para nasabahnya agar tetap loyal dengan bank [9]. Di dalam data mining diperlukan algoritma yang mampu mengklasifikasikan customer tersebut yang akan melakukan *churn* atau tidak, sehingga dalam penelitian ini menggunakan Naïve Bayes dan ID3 untuk memprediksi hal tersebut. Kelebihan dari Naïve bayes adalah mampu melakukan prediksi dengan mempergunakan data yang sudah ada [5]. Sedangkan kelebihan dari ID3 sendiri yaitu memiliki kefleksibelan yang meningkatkan kualitas dari suatu keputusan yang diperoleh [10].

2. TINJAUAN PUSTAKA

2.1. Customer Churn

Churn berhubungan dengan berhentinya nasabah menggunakan jasa/ produk milik perusahaan dan memungkinkan nasabah untuk berpindah ke perusahaan lainnya [4]. Permasalahan tersebut harus segera diatasi karena dapat berdampak besar pada perusahaan [11]. Berpindahnya nasabah dapat dilihat dari beberapa faktor, seperti:

- Faktor Harga
  Harga yang kompetitif dan biaya untuk membeli jasa/ produk milik perusahaan sesuai dengan ekspetasi nasabah.
- Faktor Jasa/ Produk
  Jasa/ Produk milik perusahaan memiliki kualitas yang sesuai dengan ekspetasi nasabah dan sesuai dengan kebutuhan nasabah
- Faktor *Customer*
  Tingkat konsumsi nasabah dan juga pendapatan nasabah menjadi dampak ke loyalitas nasabah kepada pelanggan.

2.2. SEMMA (Sample, Explore, Modify, Model, Assess)

SEMMA merupakan akronim dari Sample, Explore, Modify, Model, dan Assess. Dalam proses yang telah ada SEMMA menawarkan proses yang cukup mudah dipahami serta memungkinkan untuk pengembangan serta pemeliharaan proyek data mining secara terorganisir dan juga memadai [9]. SAS Institute mempertimbangkan dengan 5 tahapan untuk proses data mining :

- Sample
  Didalam tahapan *sample*, ada pengambilan sampel data dengan mengekstraksi dari sebagian dari kumpulan data yang besar dan cukup besar yang berisi informasi yang signifikan, namun cukup kecil untuk dimanipulasi dengan cepat
- Explore
  Didalam tahapan *explore*, akan dilakukan eksplorasi data dengan cara mencari sebuah tren dan anomali yang tidak terduga untuk mendapatkan pemahaman dan ide
- Modify
  Didalam tahapan *modify*, akan dilakukan modifikasi data dengan membuat, memilih, dan mengubah variabel untuk memfokuskan proses pemilihan model
- Model
  Didalam tahapan *model*, berisikan pemodelan data yang menggunakan *software* untuk mendapatkan kombinasi data yang diprediksi secara otomatis sehingga mencapai hasil yang diinginkan
- Assess
  Didalam tahapan *assess*, akan dilakukan penilaian data dengan melakukan evaluasi dari kegunaan dan keandalan hasil dari proses data mining dan memperkirakan seberapa baik kinerjanya.

2.3. Alogritma Naïve Bayes

Algoritma Naïve Bayes merupakan langkah – langkah data mining yang cukup umum digunakan dalam pengolahan data dan juga mengekstrak data yang menggunakan metode serta algoritma tertentu agar menghasilkan informasi yang berguna dalam pengambilan keputusan [11]. Naive Bayes adalah teorema yang dibawa serta dikenalkan oleh Thomas Bayes yang merupakan ilmuan dari Inggris. Naive Bayes mempunyai nilai tambah dalam sisi kecepatan dalam pembelajaran dan juga toleransinya pada nilai yang hilang dari fitur (*missing value*). Algoritma ini pada dasarnya menggunakan perhitungan probabilitas dan asumsi-asumsi. Pada Naïve Bayes, ada sebuah teknik yang bernama *Laplace Correction*, yang merupakan sebuah teknik untuk menghilangkan nilai probabilitas yang memiliki nilai 0 [12]. Hal tersebut ditemukan oleh Pierre Laplace. Teknik tersebut dapat memberi keakuratan untuk dataset yang memiliki data yang banyak.
Persamaan *Teorema Bayes:*

$$P\ (H \mid X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Rumus *Laplace Correction*:

$$\rho_i = \frac{m_i+1}{n+k} \quad (2)$$

2.4. Algoritma ID3

Algoritma ID3 (*Iterative Dichotomicer 3)* merupakan algoritma pembelajaran pohon keputusan atau yang biasa di kenal *decision tree learning* [13]. Pada *decision tree* dimana strukturnya berbentuk pohon dan setiap note yanng ada akan merepresentasikan atribut yang sudah diuji, daun pada note *decision tree* akan merepresentasikan kelompok kelas tertentu. Sedangkan level note teratas dari *decision tree* biasanya merupakan atribut yang memiliki pengaruh besar pada kelas tertentu. Entropy merupakan karakteristik impuryt dan homogenity dari kumpulan data yang diketahui dari ukuran teori informasi. Setelah mengetahui entropy, maka dapat menghitung nilai *information gain* dari tiap atribut. Information gain merupakan sebuah informasi yang diperoleh dari perubahan entropy baik dari observasi, dan bisa juga dengan melakukan partisipasi terhadapat dataset.
Persamaan Entropy:

$$Entropy\ (S) = \sum_{i\ =\ 1}^{c} - Pi\ x\ log_2 Pi \quad (3)$$

Persamaan Information Gain:

$$Gain\ (S, A) = Entropy\ (S) - \left(\sum_{i\ =\ 1}^{n} \frac{|Si|}{|S|} Entropy(Si)\right) \quad (4)$$

2.5. Confusion Matrix

Confusion matrix adalah metode klasifikasi yang digunakan untuk membuktikan kinerja dari klasifikasi tersebut [14]. *Confusion matrix* akan membandingkan hasil informasi dari klasifikasi yang dilakukan oleh sistem dengan hasil informasi dari klasifikasi yang sebenarnya. Tabel dari *confusion matrix* yaitu ada *Accuracy yang* akan memberikan hasil akurasi untuk mendefinisikan tingkat dari dekatnya nilai aktualnya dengan nilai hasil prediksi. Kemudian ada *precision*, merupakan tingkat dari ketepatan informasi yang diinginkan dengan jawaban dari keinginan tersebut akan diberikan oleh sistem. Terakhir ada *recall*, yang merupakan tingkat dari keberhasilan dari sistem untuk mendapatkan sebuah informasi. Pengukuran kinerja dari *confusion matrix* menggunakan 4 istilah yang akan menunjukkan hasil dari proses klasifikasi yaitu:

- TP (*True Positive*)
  Total dari data yang positif dan terklasifikasi oleh sistem dengan benar
- TN (*True Negative*)
- Total dari data yang positif dan terklasifikasi oleh sistem dengan salah
- FP (*False Positive*)
  Total dari data yang negatif tetapi terklasifikasi oleh sistem dengan benar
- FN (*False Negative*)
  Total dari data yang negatif tetapi terklasifikasi oleh sistem dengan salah

Rumus dari tabel *confusion matrix* adalah:
Rumus *Accuracy*:

$$\frac{TP+TN}{TP+TN+FP+FN}\ x\ 100\% \quad (5)$$

Rumus *Precision*:

$$\frac{TP}{TP+FP}\ x\ 100\% \quad (6)$$

Rumus *Recall*:

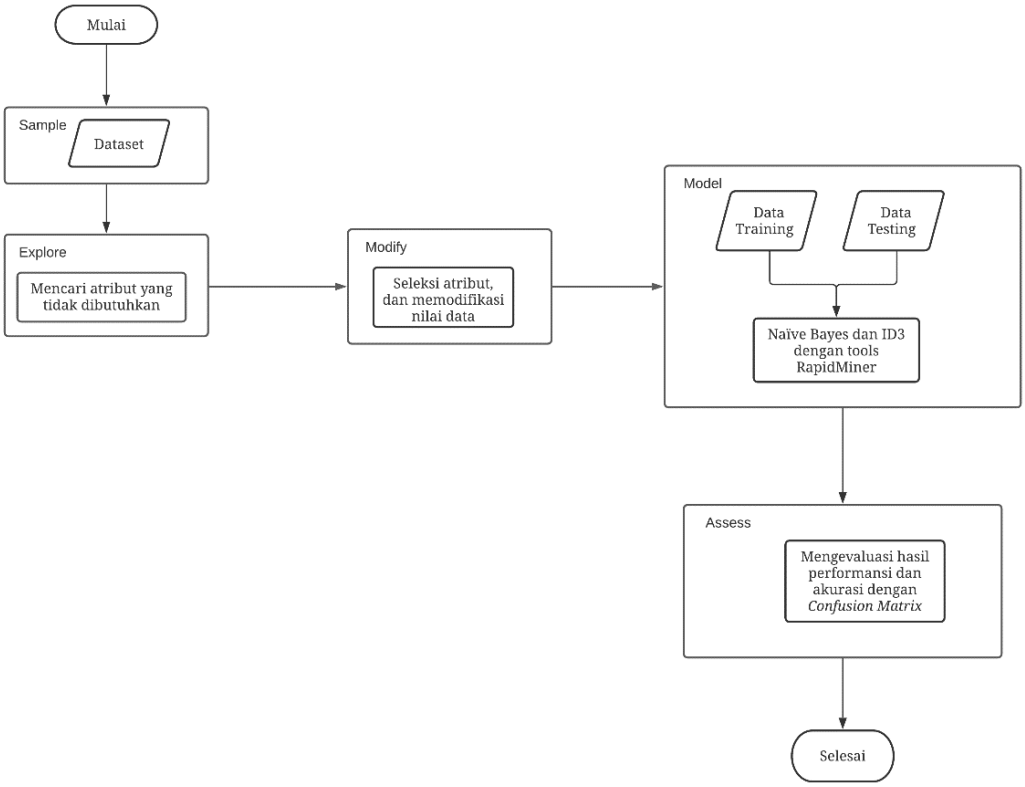
$$\frac{TP}{TP+FN}\ x\ 100\% \quad (7)$$

2.6. RapidMiner

RapidMiner adalah salah satu *software* yang digunakan untuk alat pembelajaran dalam ilmu data mining sebagai alat pengambilan keputusan. RapidMiner adalah suatu sistem yang mendukung sebuah desain dan juga dokumentasi dari keseluruhan proses data mining [15]. RapidMiner memberikan opsi set operator lebih dari satu yang luas, tetapi juga struktur yang mengekspresikan aliran kontrol proses. Software ini juga dikembangkan oleh perusahaan untuk digunakan dalam bidang komersial, penelitian, pendidikan, pelatihan, dan lainnya yang bersangkutan dengan data yang besar. RapidMiner support banyak format data dan menawarkan operator untuk menetapkan tidak hanya domain nilai atribut (jenis atribut), tetapi juga peran mereka dalam proses pembelajaran. Plug-in pilihan fitur secara otomatis menerapkan kriteria yang ditentukan pengguna untuk merancang set fitur terbaik untuk tugas pembelajaran. Selain itu, ia mengevaluasi fitur yang dipilih sehubungan dengan stabilitas. Untuk aplikasi dunia nyata, penting bahwa kinerja yang baik dicapai pada setiap sampel data. Tidak cukup bahwa fitur yang dipilih memungkinkan kinerja yang baik rata-rata dalam validasi silang berjalan, tetapi harus dijamin bahwa fitur memungkinkan kinerja yang cukup baik pada setiap sampel data.

3. METODOLOGI PENELITIAN

Metodologi yang digunakan untuk memproses dataset adalah SEMMA yang merupakan singkatan dari tahapan *Sample*, *Explore*, *Modify*, *Model*, dan *Assess*. Data akan diprediksi menggunakan Naïve Bayes dan ID3 dengan bantuan *tools* RapidMiner. Gambaran umum dari sistem yang akan dibangun dari penelitian ini disusun pada *flowchart* pada gambar 1.



Gambar 1. Gambaran Sistem

Penjelasan dari *flowchart* pada gambar 1 adalah dataset yang digunakan merupakan tahap *sample* dan akan masuk pada tahap *explore* yang akan dicari atribut yang tidak dibutuhkan dalam penelitian ini. Setelah ditemukan, selanjutnya memasuki tahap *modify* dimana data akan diseleksi atribut yang tidak diperlukan, mengelompokkan nilai pada data dan memodifikasi nilai data. Dataset akan dipartisi menjadi 2 data yaitu, data *training* dan data *testing* yang akan melalui perhitungan Naïve Bayes dan ID3 dengan bantuan *tools* RapidMiner yang merupakan tahap *model*. Setelah menemukan performansi dan akurasinya, maka peneliti akan melakukan evaluasi hasil akhir dari proses data mining ini dengan menggunakan *Confusion Matrix*, proses ini masuk dalam tahap *assess*.

4. HASIL DAN PEMBAHASAN

4.1. Sample

Sample yang dipakai untuk melakukan penelitian ini yaitu sebanyak 10.000 data dan berasal dari website Kaggle dan merupakan data nasabah milik bank, dan mempunyai 14 atribut dengan labelnya adalah *churn* dan *non churn* pada atribut *Exited.* Atribut yang ada pada dataset didsekripsikan pada tabel 1. Sedangkan untuk visualisasi data mentahnya ada pada tabel 2.

Tabel 1. Atribut dan Nilai Dataset [16]

Atribut	Type Data	Keterangan
Row Number	Integer	Nomor urut
Customer ID	Integer	Nomor unik kartu kredit milik nasabah

Surname	Polynominal	Nama nasabah bank
Credit Score	Integer	Skor kredit menyatakan kualitas kartu kredit nasabah
Geography	Polynominal	Lokasi tinggal nasabah
Gender	Polynominal	Jenis kelamin nasabah
Age	Integer	Umur nasabah
Tenure	Integer	Jumlah tahun menjadi nasabah
Balance	Real	Saldo yang dimiliki oleh nasabah
NumOfProducts	Integer	Jumlah produk yang telah dibeli nasabah melalui bank
HasCrCard	Integer	Kepemilikan credit card
IsActiveMember	Integer	Keaktifan nasabah menggunakan kartu kredit
Estimate Salary	Real	Gaji yang dimiliki oleh nasabah
Exited	Integer	Keputusan nasabah untuk meninggalkan bank/ tidak

Tabel 2. Visualisasi Data Mentah [16]

RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
1	15634602	Hargrave	619	France	Female	42	2	0	1	1	1	101348.88	1
2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
3	15619304	Onio	502	France	Female	42	8	159660.8	3	1	0	113931.57	1
4	15701354	Boni	699	France	Female	39	1	0	2	0	0	93826.63	0
5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.1	0
…													
…													
10.000	15628319	Walker	792	France	Female	28	4	130142.79	1	1	0	38190.78	0

4.2. Explore

Dalam dataset nasabah bank, beberapa atribut tidak dibutuhkan dalam melakukan prediksi pada penelitian ini. Atribut tersebut yaitu *Row Number*, *Customer ID*, dan *Surname*. Dalam dataset beberapa atribut perlu diubah nilainya untuk mempermudah pembacaannya. Atribut tersebut adalah *HasCrCard, IsActiveMember, dan Exited*. Kemudian untuk atribut *CreditScore, Age, Balance, dan Estimated Salary* akan di kelompokkan untuk memperkecil cakupannya. Nasabah bank akan dikategorikan menjadi 2 kategori(label) yaitu *churn*, dan non-*churn*. Untuk penjelasan yang lebih rinci dapat dilihat pada tabel 3.

Tabel 3. Deskripsi Label

Nilai	**Deskripsi**
Churn	Merupakan label yang diberikan pada nasabah yang akan meninggalkan bank
Non-*Churn*	Merupakan label yang diberikan pada nasabah yang tidak meninggalkan bank

4.3. Modify

Pada tahap ini, dataset yang memiliki 14 atribut berkurang menjadi 11 atribut, dan atribut HasCrCard, IsActiveMember, dan Exited akan dilakukan perubahan nilai untuk mempermudah pembacaannya. Perubahan tersebut tercantum pada tabel 4.

Tabel 4. Perubahan Nilai

Atribut	**Nilai Awal**	**Perubahan Nilai**
HasCrCard	0	No
	1	Yes
IsActiveMember	0	No
	1	Yes
Exited	0	No
	1	Yes

Kemudian pada atribut *CreditScore, Age, Balance, dan Estimated Salary* dilakukan pengelompokkan untuk memperkecil cakupannya. Pengelompokkan tersebut ada pada tabel 5.

Tabel 5. Pengelompokkan Nilai

Atribut			
CreditScore	**Age**	**Balance**	**Estimated Salary**
345 – 366	17 – 30	0 – 50.000	0 – 50.000
367 – 388	31 – 44	50.001 – 100.000	50.001 – 100.000
389 – 410	45 – 58	100.001 – 150.000	100.001 – 150.000
411 – 432	59 – 72	150.001 – 200.000	150.001 – 200.000
…	73 – 86	200.001 – 250.000	
829 – 850	87 – 100	250.001 – 300.000	

4.4. Model

Pada penelitian ini akan dilakukan komparasi 2 metode yaitu Naïve bayes dan juga ID3. Proses klasifikasi ini dibantu dengan tools RapidMiner dan hasil akurasi ke 2 metode akan dikomparasi. Dengan adanya atribut terbaru, dataset akan dibagi menjadi 2 data yaitu, data *training* dan data *testing* dengan ratio 6:4 dan akan menghasilkan akurasi tertinggi dengan bantuan *tools* RapidMiner. Pembagian data ada pada tabel 6.

Tabel 6. Pembagian Dataset

Komposisi Data	**Data**		**Jumlah Seluruh Data**
	Training	**Testing**	
6:4	6.000	4.000	10.000

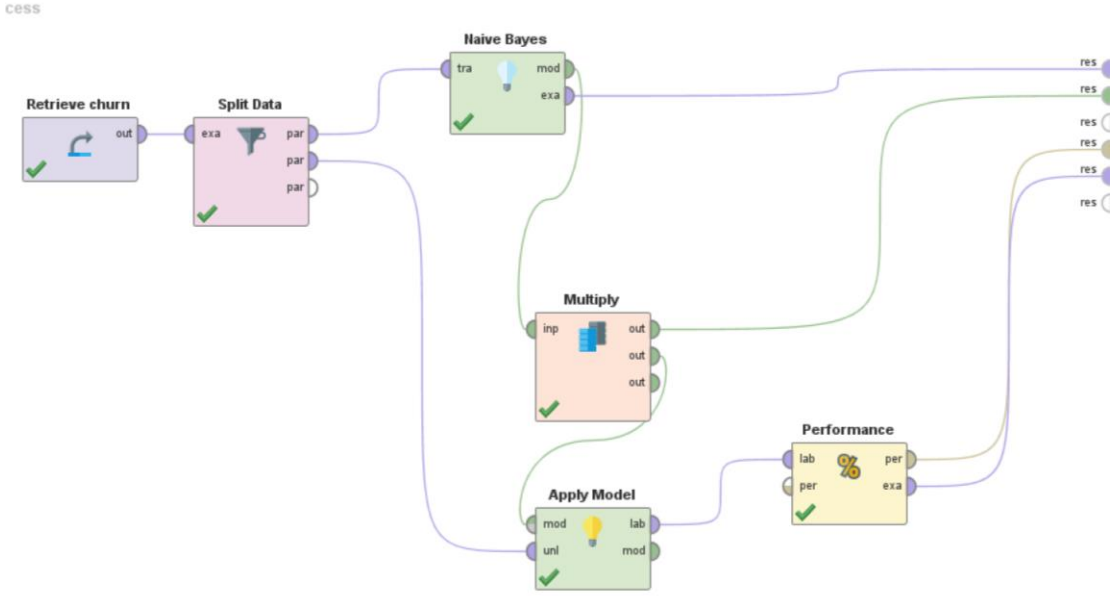
Dalam penelitian ini, *churn* dinyatakan dengan *yes* dan non-*churn* dengan *no* dan dari keseluruh data, data yang memiliki label *churn* dan non-*churn* dapat dilihat pada tabel 7.

Tabel 7. Perbandingan Label

Label Data	**Data**	**Presentase**
Churn	2.037	20,37 %
Non-*Churn*	7.963	79,63 %

4.4.1. Naïve Bayes

Dalam tahap ini, peneliti melakukan proses pengujian menggunakan metode Naïve Bayes dengan bantuan *tools* RapidMiner dan pada operation Naïve Bayes, untuk mendapatkan hasil yang maksimal, tetap mencentang *Laplace Correction*. Proses yang tersusun pada RapidMiner dapat dilihat pada Gambar 2.



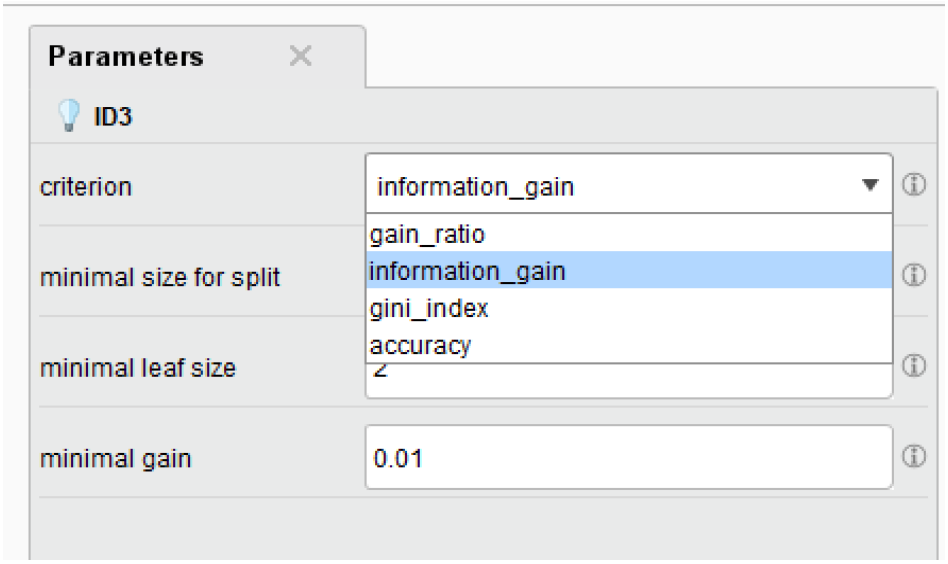
Gambar 2. Model Proses Naïve Bayes

Keterangan:

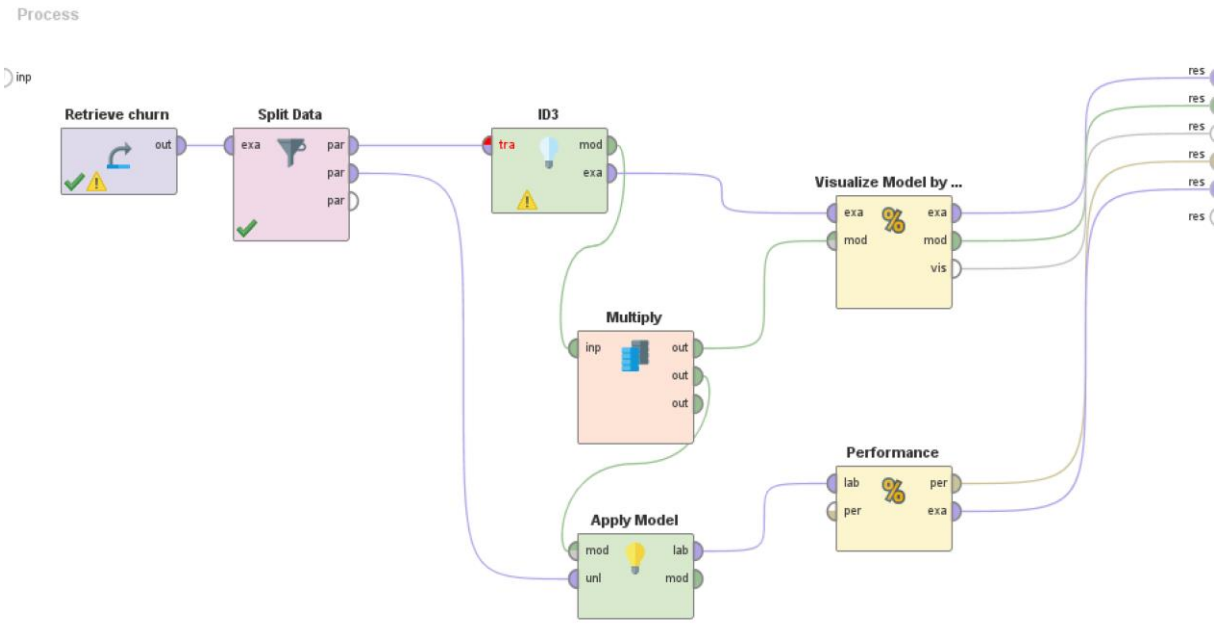
- Retrieve *churn*: Operator ini digunakan untuk mengambil objek yang ada pada data *churn* untuk diproses dengan RapidMiner. Objek ini berupa ExampleSet.
- Split Data: Operator ini digunakan untuk membagi ExampleSet yang telah diinputkan dengan ukuran relatif yang dapat ditentukan pada parameter.
- Naïve Bayes: Operator yang digunakan untuk mengklasifikasikan data dengan metode Naïve Bayes
- Multiply: Operator ini mengambil objek dari input dan mengirimkan salinan independen ke tiap output.
- Apply Model: Operator yang digunakan untuk menghubungkan data testing yang telah di split data ke performance
- Performance: Operator untuk mengukur performa akurasi dari model dan juga evaluasi dari kinerja model tersebut

4.4.2. ID3 (Decision Tree)

Dalam tahap ini, peneliti melakukan proses pengujian menggunakan metode ID3 (Decision Tree) dengan bantuan *tools* RapidMiner. Pada pilihan di operator ID3, akan ada pilihan *criterion* dan untuk perhitungan dari ID3 menggunakan *information gain*. Pemilihan criterion dapat dilihat pada Gambar 3 dan proses pada RapidMiner dapat dilihat pada Gambar 4.



Gambar 3. Pemilihan *Criterion*

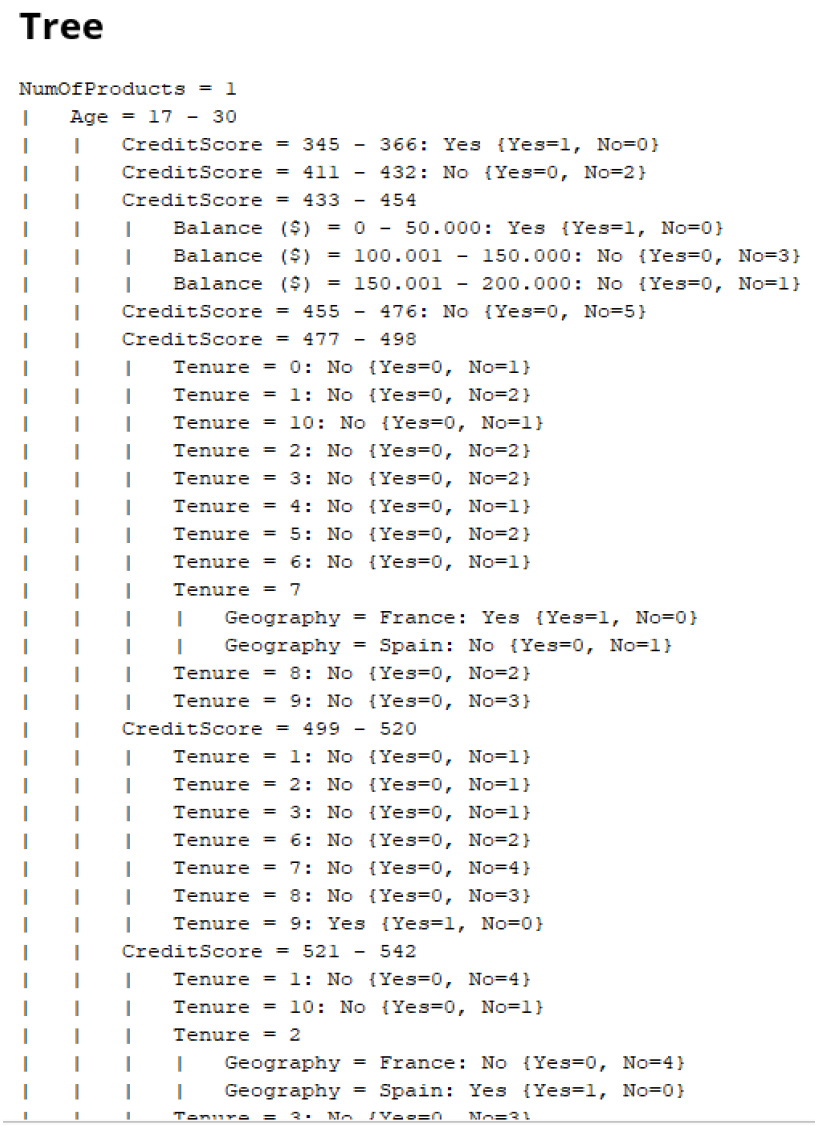


Gambar 4. Model Proses ID3

Keterangan:

- Retrieve *churn*: Operator ini digunakan untuk mengambil objek yang ada pada data *churn* untuk diproses dengan RapidMiner. Objek ini berupa ExampleSet.
- Split Data: Operator ini digunakan untuk membagi ExampleSet yang telah diinputkan dengan ukuran relatif yang dapat ditentukan pada parameter.
- ID3: Operator yang digunakan untuk mengklasifikasikan data dengan metode ID3 dengan pilihan information_gain
- Multiply: Operator ini mengambil objek dari input dan mengirimkan salinan independen ke tiap output.
- Apply Model: Operator yang digunakan untuk menghubungkan data testing yang telah di split data ke performance
- Visualize Model by SOM: Operator yang menyediakan visualisasi model arbitrer dengan bantuan pengurangan dimensi melalui SOM dari kumpulan data dan model.
- Performance: Operator untuk mengukur performa akurasi dari model dan juga evaluasi dari kinerja model tersebut

Hasil ID3 pada Tree pada gambar 5.



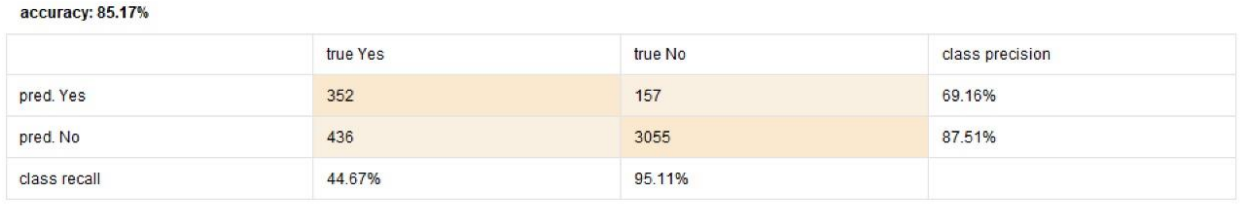
```
Tree

NumOfProducts = 1
|   Age = 17 - 30
|   |   CreditScore = 345 - 366: Yes {Yes=1, No=0}
|   |   CreditScore = 411 - 432: No {Yes=0, No=2}
|   |   CreditScore = 433 - 454
|   |   |   Balance ($) = 0 - 50.000: Yes {Yes=1, No=0}
|   |   |   Balance ($) = 100.001 - 150.000: No {Yes=0, No=3}
|   |   |   Balance ($) = 150.001 - 200.000: No {Yes=0, No=1}
|   |   CreditScore = 455 - 476: No {Yes=0, No=5}
|   |   CreditScore = 477 - 498
|   |   |   Tenure = 0: No {Yes=0, No=1}
|   |   |   Tenure = 1: No {Yes=0, No=2}
|   |   |   Tenure = 10: No {Yes=0, No=1}
|   |   |   Tenure = 2: No {Yes=0, No=2}
|   |   |   Tenure = 3: No {Yes=0, No=2}
|   |   |   Tenure = 4: No {Yes=0, No=1}
|   |   |   Tenure = 5: No {Yes=0, No=2}
|   |   |   Tenure = 6: No {Yes=0, No=1}
|   |   |   Tenure = 7
|   |   |   |   Geography = France: Yes {Yes=1, No=0}
|   |   |   |   Geography = Spain: No {Yes=0, No=1}
|   |   |   Tenure = 8: No {Yes=0, No=2}
|   |   |   Tenure = 9: No {Yes=0, No=3}
|   |   CreditScore = 499 - 520
|   |   |   Tenure = 1: No {Yes=0, No=1}
|   |   |   Tenure = 2: No {Yes=0, No=1}
|   |   |   Tenure = 3: No {Yes=0, No=1}
|   |   |   Tenure = 6: No {Yes=0, No=2}
|   |   |   Tenure = 7: No {Yes=0, No=4}
|   |   |   Tenure = 8: No {Yes=0, No=3}
|   |   |   Tenure = 9: Yes {Yes=1, No=0}
|   |   CreditScore = 521 - 542
|   |   |   Tenure = 1: No {Yes=0, No=4}
|   |   |   Tenure = 10: No {Yes=0, No=1}
|   |   |   Tenure = 2
|   |   |   |   Geography = France: No {Yes=0, No=4}
|   |   |   |   Geography = Spain: Yes {Yes=1, No=0}
|   |   |   Tenure = 3: No {Yes=0, No=3}
```

Gambar 5. Hasil Tree

4.5. Assess

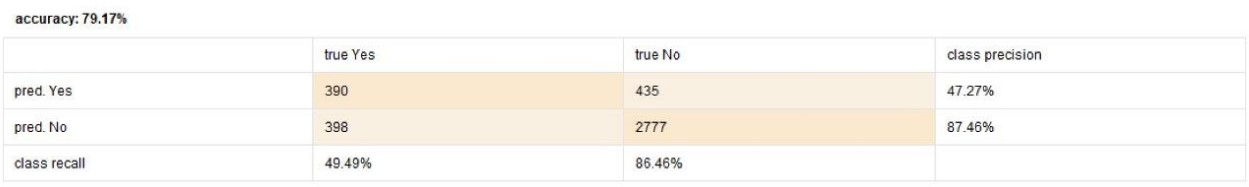
Pada tahapan ini, hasil dari pengujian dengan menggunakan Naïve Bayes dan ID3 maka akan dilakukan evaluasi dengan menggunakan *confusion matrix*. Setelah hasil pengujian tersebut diterapkan ke dalam *confusion matrix*, data yang dihasilkan dapat dilihat pada gambar 6 dan gambar 7.



accuracy: 85.17%

	true Yes	true No	class precision
pred. Yes	352	157	69.16%
pred. No	436	3055	87.51%
class recall	44.67%	95.11%	

Gambar 6. Hasil Akurasi Naïve Bayes



accuracy: 79.17%

	true Yes	true No	class precision
pred. Yes	390	435	47.27%
pred. No	398	2777	87.46%
class recall	49.49%	86.46%	

Gambar 7. Hasil Akurasi ID3

Berdasarkan hasil dari *confusion matrix* dengan menggunakan *tools* RapidMiner, dapat disimpulkan dengan tabel 8 dibawah ini.

Tabel 8. Kesimpulan hasil *confusion matrix*

Keterangan Hasil	Kondisi	Metode	
		Naïve Bayes	**ID3**
Akurasi	-	85.17%	79.17%
Recall	Yes (*Churn*)	44.67%	49.49%
	No (Non-*Churn*)	95.11%	86.46%
Precision	Yes (*Churn*)	69.16%	47.27%
	No (Non-*Churn*)	87.51%	87.46%

Dengan demikian prediksi nasabah yang melakukan *churn* yang diprediksi menggunakan Naïve Bayes mendapatkan akurasi sebesar 85.17% dan yang diprediksi menggunakan ID3 mendapatkan akurasi sebesar 79.17%. Menandakan bahwa Naïve Bayes memberikan akurasi yang lebih baik dibandingkan ID3 pada penelitian ini.

5. KESIMPULAN DAN SARAN

Dilakukannya penelitian ini untuk memprediksi nasabah bank yang *churn* sehingga bank dapat mengetahui alasan nasabah melakukan *churn* dan dapat memberikan pelayanan lebih baik dari sebelumnya. Dengan demikian, kesimpulan dari hasil penelitian yang telah dilakukan, yaitu:

1. Penelitian dilakukan dengan menggunakan metode SEMMA untuk memproses dataset yang masih mentah hingga menghasilkan nilai akurasi.
2. 10.000 data yang kemudian dibagi menjadi 2 data dengan ratio 6:4 tersebut, 4.000 data yang merupakan data testing memberikan hasil akurasi yang lumayan tinggi. Menggunakan Naïve Bayes, hasil dari akurasinya mendapatkan 85.17% dan menggunakan ID3, hasil dari akurasinya mendapatkan 79.17%.
3. Dalam penelitian ini, dilakukan komparasi dengan menggunakan Naïve Bayes dan ID3 untuk mencari metode yang paling baik dalam memprediksi nasabah *churn*.
4. Menggunakan confusion matrix dapat memberikan hasil akurasi, recall, dan juga precision.
5. Secara hasil keseluruhan dari penelitian ini menunjukkan bahwa Naïve Bayes memberikan akurasi yang lebih tinggi dari pada ID3.

Berdasarkan dari kesimpulan diatas, ada beberapa saran untuk prediksi nasabah yang *churn* lebih baik untuk dapat diimplementasikan pada penelitian selanjutnya, yaitu pengujian dengan membandingkan pembagian dataset dengan ratio yang berbeda, pengujian dapat diuji coba dengan menggunakan *k-fold cross validation* untuk mendapatkan pemahaman lebih baik lagi, dan pengujian dapat menggunakan metode C4.5 sebagai pembanding dengan ID3.

DAFTAR PUSTAKA

[1] A. S. Maulana, "Pengaruh Kualitas Pelayanan Dan Harga Terhadap Kepuasaan pelanggan PT. TOI," *J. Ekon. Vol.*, vol. 7, no. 2, pp. 113–125, 2016.

[2] J. Setiyono and S. Sutrimah, "Analisis Teks dan Konteks Pada Iklan Operator Seluler (XL dengan Kartu AS)," *Pedagog. J. Pendidik.*, vol. 5, no. 2, pp. 297–310, 2016, doi: 10.21070/pedagogia.v5i2.263.

[3] T. T. Hanifa, Adiwijaya, and S. Al-faraby, "Analisis Churn Prediction pada Data Pelanggan PT. Telekomunikasi dengan Logistic Regression dan Underbagging," *e-Proceeding Eng.*, vol. 4, no. 2, p. 78, 2017.

[4] H. N. Irmanda, R. Astriratma, and S. Afrizal, "Perbandingan Metode Jaringan Syaraf Tiruan Dan Pohon Keputusan Untuk Prediksi Churn," *JSI J. Sist. Inf.*, vol. 11, no. 2, pp. 1817–1825, 2019, doi: 10.36706/jsi.v11i2.9286.

[5] R. Novendri, R. Andreswari, and O. N. Pratiwi, "Implementasi Data Mining Untuk Memprediksi Customer Churn Menggunakan Algoritma Naive Bayes," *eProceedings …*, vol. 8, no. 2, pp. 2762–2773, 2021, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/download/14678/14455.

[6] N. W. Wardani and N. K. Ariasih, "Analisa Komparasi Algoritma Decision Tree C4.5 dan Naïve Bayes untuk Prediksi Churn Berdasarkan Kelas Pelanggan Retail," *Int. J. Nat. Sci. Eng.*, vol. 3, no. 3, p. 103, 2019, doi: 10.23887/ijnse.v3i3.23113.

[7] Andie, "Penerapan Decision Tree Untuk Menganalisis Kemungkinan Pengunduran Diri Calon Mahasiswa Baru," *Technologia*, vol. 7, no. 1, pp. 8–14, 2016.

[8] Y. Asri, "Analisa Perbandingan Keputusan Metode Klasifikasi Decision Tree dan Naive Bayes Dalam Penentuan Diagnosa Hipertensi," *J. Kaji. Ilmu dan Teknol.*, vol. 4, pp. 41–46, 2015.

[9] V. R. Hananto, *Buku Ajar Kecerdasan Bisnis*. 2017.

[10] N. W. Wardani, G. R. Dantes, and G. Indrawan, "Prediksi Customer Churn Dengan Algoritma Decision Tree C4 . 5," *J. Resist.*, vol. 1, no. 1, pp. 16–24, 2018.

[11] A. Yulianto and F. Firmansyah, "Prediksi Customer Churn Pada Bisnis Retail Menggunakan Algoritma Naïve Bayes," *Remik*, vol. 6, no. 1, pp. 41–47, 2021, doi: 10.33395/remik.v6i1.11196.

[12] M. Rizki, M. Arhami, and Huzeni, "Perbaikan Algoritma Naive Bayes Classifier Menggunakan Teknik Laplacian Correction," *J. Teknol.*, vol. 21, no. 1, p. 7, 2021.

[13] A. Nurzahputra, A. R. Safitri, and M. A. Muslim, "Klasifikasi Pelanggan pada Customer Churn Prediction Menggunakan Decision Tree," *Pros. Semin. Nas. Mat. X 2016*, pp. 717–722, 2016, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/article/download/21528/10288/.

[14] Karsito, "Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naïve Bayes Di Perumahan Azzura

Residencia," *J. Teknol. Pelita Bangsa*, vol. 9, pp. 43–48, 2019.
[15] V. Kotu and B. Deshpande, *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. 2016.
[16] M. Akturk, "Churn for Bank Customers," *Kaggle*, 2020. https://www.kaggle.com/mathchi/churn-for-bank-customers.

NOMENKLATUR

X	Data dengan *class* yang masih tidak diketahui
H	Hipotesis data X merupakan suatu *class* yang spesifik
P(H | X)	Probabilitas hipotesis H berdasar kondisi X (posteriori probability)
P(H)	Probabilitas hipotesis H (prior probability)
P(X | H)	Probabilitas X berdasarkan kondisi pada hipotesis H
P (X)	Probabilitas X
$\boldsymbol{\rho_i}$	Probabilitas dari m_i
m_i	jumlah sampel dalam kelas dari atribut m_i
k	jumlah kelas dari atribut m_i
n	jumlah sampel
C	Jumlah nilai yang ada dalam atribut (jumlah kelas)
Pi	Jumlah sampel yang ada pada kelas i
A	Atribut
S	Sampel
N	Jumlah partisi himpunanyang ada pada atribut A
|Si|	Jumlah sampel yang ada pada partisi
|S|	Jumlah sampel yang ada pada S