

IMPLEMENTASI METODE ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI CUSTOMER CHURN PADA BISNIS USAHA MIKRO KECIL DAN MENENGAH

Stephen¹, Desiyanna Lasut²

Fakultas Sains dan Teknologi, Universitas Buddhi Dharma
Jalan Imam Bonjol No. 41, Tangerang, Indonesia

Email: ¹stephencolz7@gmail.com,

²desiyanna.lasut@ubd.ac.id

Abstrak

Perusahaan UMKM biasanya memiliki data pelanggan yang mereka beli dengan maksud dalam membangun ikatan yang memberikan keuntungan dari masing-masing pelanggan dan perusahaan, tetapi data pelanggan yang sering tidak di organisir dan dianalisis dengan teliti, sehingga pelanggan meninggalkan atau dapat tidak berlangganan dari satu perusahaan atau akan pindah ke perusahaan lainnya. Menganalisa prediksi dalam kehilangan pelanggan (customer churn) memerlukan model algoritma yang bisa melakukan prediksi kehilangan pelanggan secara akurat. Dalam menentukan hasil prediksi tersebut suatu usaha melakukan analisa memakai teknik data mining dengan menggunakan algoritma Naive Bayes sehingga suatu usaha dapat melakukan tindakan preventif. Data latih dan testing diperlukan untuk memasukan ke dalam pemodelan naïve bayes untuk memberi hasil dengan akurasi 84,95% dan AUC 83,20%. Dari akurasi dan AUC ini sebagai tanda hasil Interpretasi dengan nilai yang dilakukan yaitu dalam titik "Baik". Naïve bayes membuktikan bisa memprediksi kemungkinan customer churn berdasarkan data dan klasifikasi sehingga hasilnya bisa mendukung menjadi keputusan bisnis.

Kata Kunci

Algoritma Naïve Bayes, Data mining, Kehilangan pelanggan, RapidMiner, Netbeans

Abstract

A business usually has customer data that they buy with the intention of building a bond that benefits each customer and company, but customer data is often not carefully organized and analyzed, so customers leave or may not subscribe to one company or will move to another company. Analyzing predictions in customer churn requires an algorithm model that can predict customer loss accurately. In determining the prediction results, an attempt is made to analyze using data mining techniques using the Naive Bayes algorithm so that a business can take preventive actions. Training and testing data is needed to be included in the naïve Bayes model to give results with an accuracy of 84.95% and AUC of 83.20%. From this accuracy and AUC as a sign of the results of Interpretation with the value carried out, namely in the "Good" point. Naïve Bayes has proven that it can predict the possibility of customer churn based on data and classification so that the results can support business decisions.

Kata Kunci

Algorithm Naïve Bayes, Data mining, Customer churn, RapidMiner, Netbeans

Latar Belakang

Teknologi terus berkembang pesat dari tahun ke tahun, menyebabkan kebutuhan akan informasi dan komunikasi semakin meningkat. Masyarakat menggunakan berbagai perangkat teknologi yang ada untuk membantu memecahkan masalah informasi dan komunikasi. Pada saat ini, penyebaran informasi tidak terbatas pada media cetak, tetapi jejaring sosial dan media elektronik adalah sarana komunikasi yang paling banyak digunakan dan tersebar luas.

Pelanggan bagi usaha bisnis merupakan aset utama, oleh karena itu berbagai langkah dilakukan oleh UMKM agar pelanggan tidak berhenti membeli. Apalagi di era digital saat ini, sulit mendapatkan pelanggan baru, dan lagi dengan menjamurnya UMKM di pelosok Negara Indonesia. Manajer bisnis tidak hanya dapat mengidentifikasi pelanggan potensial, tetapi juga mempertahankan pelanggan potensial (retensi pelanggan) untuk mencegah pelanggan meninggalkan pembelian suatu usaha dan beralih ke pesaing (churn).

Di suatu usaha bisnis, sejumlah uang dan usaha dihabiskan untuk menarik pelanggan baru, tetapi seiring waktu pelanggan ini dapat berhenti berbelanja karena berbagai alasan. Hal ini tentu tidak diinginkan untuk bisnis ritel, sehingga penting untuk mengetahui terlebih dahulu pelanggan mana yang kemungkinan besar akan berhenti berlangganan pada waktu tertentu. Dalam hal retensi pelanggan, suatu usaha dapat meningkatkan layanan pelanggan, meningkatkan kualitas produk, dan dengan cepat mempelajari pelanggan mana yang kemungkinan besar akan meninggalkan bisnis yang di miliki. Salah satu cara untuk mengetahuinya adalah dengan memprediksi dan mendeteksinya lebih awal dengan membuat churn prediction.

Algoritma Naive Bayes merupakan salah satu algoritma dalam teknik klasifikasi. Naive Bayes adalah klasifikasi menggunakan metode probabilistik dan statistik yang dapat memprediksi peluang masa depan berdasarkan pengalaman masa lalu, oleh karena itu disebut Teorema Bayes. Teorema tersebut dikaitkan dengan kondisi antara sifat-sifat yang dikatakan bebas. Pengklasifikasi Naive Bayes mengasumsikan bahwa ada atau tidak adanya karakteristik tertentu dari suatu kelas tidak ada hubungannya dengan karakteristik kelas lain. Untuk menjelaskan teorema Naive Bayes, perlu diperhatikan bahwa proses klasifikasi memerlukan beberapa petunjuk untuk menentukan kelas mana yang cocok dengan sampel yang dianalisis.

Dalam pembuatan penelitian ini tentunya terdapat suatu tujuan penelitian serta manfaat penelitian yang dapat berguna terutama bagi suatu bisnis UMKM untuk memprediksi customer churn atau kehilangan pelanggan agar menjadi data acuan yang dapat diberikan kepada pemilik usaha untuk digunakan sebagai parameter melakukan promosi atau strategi pemasaran. Berikut ini adalah tujuan dan manfaat yang bisa penulis berikan dari pembuatan penelitian. Ada beberapa tujuan ini dari penelitian yang diharapkan peneliti dalam melakukan penelitian, yaitu :

1. Mendapatkan informasi dalam karakteristik customer churn dalam dasar metode klasifikasi algoritma Naive Bayes.
2. Dapat memahami konsep klasifikasi yang diteliti khususnya untuk memprediksi kelas target dari pelanggan yang melakukan churn dan no-churn suatu bisnis UMKM dengan mempelajari behaviour dari transaksi yang dilakukan oleh pelanggan.
3. Mengetahui hasil prediksi metode klasifikasi algoritma Naive Bayes dengan pengukuran tingkat ketepatan Akurasi yang dapat digunakan sebagai strategi pemasaran.

Adapun beberapa manfaat penelitian yang diharapkan peneliti dalam melakukan penelitian, yaitu:

1. Sebagai bahan penilaian dan informasi perusahaan untuk menentukan jenis strategi bisnis seperti apa yang dibutuhkan untuk mempertahankan pelanggan dan meningkatkan loyalitas dari pelanggan.

2. Bagi pemilik usaha, adanya layanan pada aplikasi yang dapat meningkatkan kenyamanan dan kemudahan dalam melakukan prediksi customer churn.
3. Penelitian ini dapat menjadi referensi penelitian selanjutnya terkait dengan prediksi customer churn maupun tentang penerapan metode Naive Bayes.

Berdasarkan penjelasan di atas, maka tujuan dari penelitian ini adalah untuk membangun model prediksi yang memprediksi churn pelanggan menggunakan Naive Bayes sehingga perusahaan dapat memprediksi kemungkinan terjadinya churn pelanggan pada tahap lebih awal. Dengan mengantisipasi kehilangan pelanggan, pemilik bisnis dapat meningkatkan efektivitas iklan dalam promosi dan mengurangi biaya yang terkait dengan kehilangan pelanggan. Berdasarkan latar belakang yang di jelaskan maka dapat di usulkan "Implementasi Metode Algoritma Naive Bayes Untuk Memprediksi Customer Churn Pada Bisnis UMKM".

Tinjauan Pustaka

A. Teori

i. Data

Data adalah kumpulan fakta yang melukiskan gambaran umum dari suatu situasi. Seseorang yang membuat kebijakan atau keputusan akan sering menggunakan data untuk ditinjau. Berkat data, seseorang dapat menganalisis, menggambarkan, atau menjelaskan suatu situasi. Pada dasarnya penggunaan data (setelah diolah dan dianalisa) menciptakan dasar yang objektif dalam proses pengambilan keputusan atau kebijakan politik untuk memecahkan permasalahan para pengambil keputusan[1]. Keputusan yang baik hanya dapat diperoleh dari mereka yang objektif dan berdasarkan data yang baik.

ii. Data Mining

Data mining merupakan proses mengekstraksi dan mengidentifikasi informasi yang berguna dan pengetahuan yang relevan dari database besar menggunakan teknik statistik, matematika, artificial intelligence, dan pembelajaran mesin[2]. Data mining memakai untuk memandu semua tahapan yang dilakukan untuk mencampurkan metode dan teknik dari macam-macam bidang seperti statistik, database, pembelajaran mesin, dan visualisasi. Data mining biasa diketahui dari suatu Knowledge Discovery in Database (KDD), menuju pada ringkasan informasi tacit yang digunakan dan bermanfaat dari data untuk database.

iii. Prediksi

Istilah prediksi adalah tahap - tahap sistematis yang memperkirakan hal yang sangat mungkin terjadi di masa depan berdasarkan informasi masa lalu dan masa kini yang disimpan untuk tujuan meminimalkan kesalahan dan perbedaan diantara apa yang terjadi dan hasil yang diharapkan. Prediksi yang baik dalam hal ini disebut juga dengan ramalan (*forecast*) pada dasarnya adalah tebakan atau prediksi kejadian yang akan datang[3].

iv. Algoritma Naive Bayes

Naive Bayes adalah pengklasifikasi probabilistik bersahaja yang mencari jumlah dari serangkaian probabilitas dengan menghitung berapa banyaknya frekuensi dan campuran nilai dalam beberapa seluruh data yang diberikan. Secara garis besar, algoritma Naive Bayes dapat digambarkan sebagai persamaan ini[4] :

$$P(R|S) = \frac{P(R)P(S|R)}{P(S)}$$

Keterangan:

R : Data yang belum diketahui kelasnya

S : Hipotesis pada data R yang merupakan class khusus

$P(R|S)$: Nilai probabilitas pada hipotesis R yang berdasarkan kondisi S

$P(R)$: Nilai probabilitas pada hipotesis R

$P(S|R)$: Nilai probabilitas S yang berdasarkan dengan kondisi hipotesis R

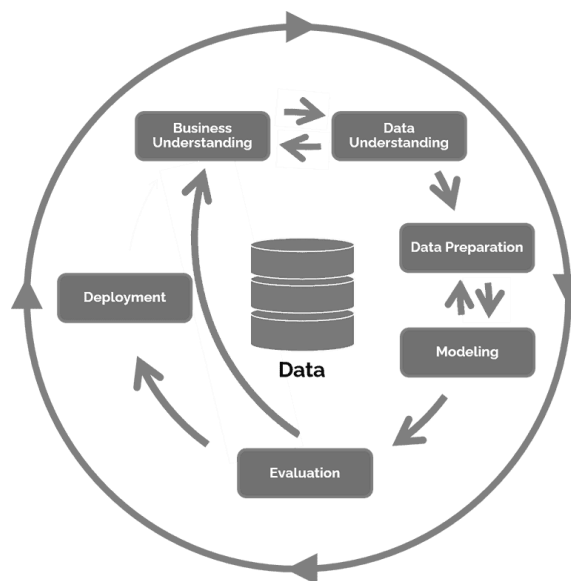
$P(S)$: Nilai probabilitas S

v. Customer Churn

Customer churn merupakan istilah yang mengacu pada pelanggan yang tidak lagi menggunakan layanan atau jasa perusahaan[5]. Perpindahan pelanggan penting untuk diketahui oleh bisnis karena merupakan gambaran holistik tentang keberhasilan bisnis dalam mempertahankan pelanggan. Oleh karena itu, customer churn merupakan salah satu faktor terpenting yang perlu dievaluasi setiap bisnis secara berkelanjutan, terutama untuk bisnis yang sedang berkembang.

B. Analisa dan Perancangan

i. Cross-Industry Standard Process for Data Mining (CRISP-DM)



Gambar 1 CRISP-DM

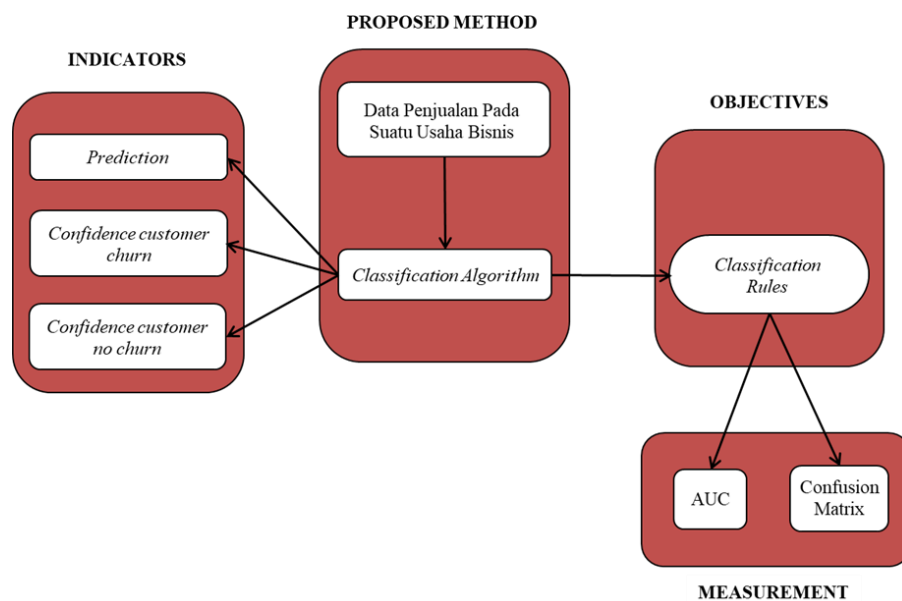
CRISP-DM (Cross Industry Standard Process for Data Mining) adalah sebuah konsorsium perusahaan yang didirikan sang Komisi Eropa dalam tahun 1996 untuk menetapkan dirinya menjadi prosedur standar untuk penambangan data yang bisa digunakan di seluruh industri. CRISP-DM menyediakan proses data mining standar yang dapat diterapkan pada strategi pemecahan masalah umum di unit bisnis atau penelitian[6]. CRISP-DM mempunyai 6 langkah, yaitu :

- Business Understanding, langkah ini membutuhkan pengetahuan tentang objek bisnis, bagaimana data dibuat atau diambil, dan cara terbaik untuk menyelaraskan maksud pemodelan mereka dengan misi usaha bisnis mereka.

- Data Understanding, umumnya dalam fase ini meliputi hasil memeriksa data, pengenalan masalah dalam data.
- Data Preparation, umumnya fase ini digunakan untuk menyelesaikan masalah data dan kemudian membuat variabel anak. Jelas, langkah ini membutuhkan pemikiran dan usaha yang dapat memenuhi kebutuhan untuk mengkonfirmasi bahwa data cocok dengan algoritma yang dipakai.
- Modeling, dasarnya fase ini untuk melakukan model prediktif atau deskriptif. Fase ini menggunakan metode statistik dan data mining dalam menetapkan teknik data mining yang berlaku, perangkat data mining, dan algoritma data mining.
- Evaluation, Langkah evaluasi dilakukan kepada contoh yang diterapkan dalam langkah sebelumnya menggunakan maksud supaya contoh yang dipengaruhi bisa sinkron menggunakan tujuan yang ingin dicapai dalam langkah pertama
- Deployment, tahap implementasi atau perencanaan merupakan tahapan terpenting dari proses CRISP-DM.

Pembahasan

A. Kerangka Pemikiran



Gambar 2 Kerangka Pemikiran

Hasil menurut desain dibutuhkan akan membuat sebuah sistem berita berupa acara pelaksanaan prediksi customer churn. Berikut komponen dari model kerangka pemikiran saya :

1. Indicators

Menggunakan parameter prediction dan confidence untuk mencari hasil association rules.

2. Proposed Method

Data penjualan pada suatu usaha bisnis di olah menggunakan aplikasi rapid miner dengan algoritma Naive Bayes untuk menemukan classification rules.

3. Objectives

Classification rules di gunakan untuk mencari kelas (klasifikasi) yang hasilnya adalah pelanggan mana saja yang churn atau no churn.

4. Measurement

Menghasilkan AUC dan confusion matriks sebagai output dari aplikasi rapid miner.

B. Metode Algoritma Naïve Bayes

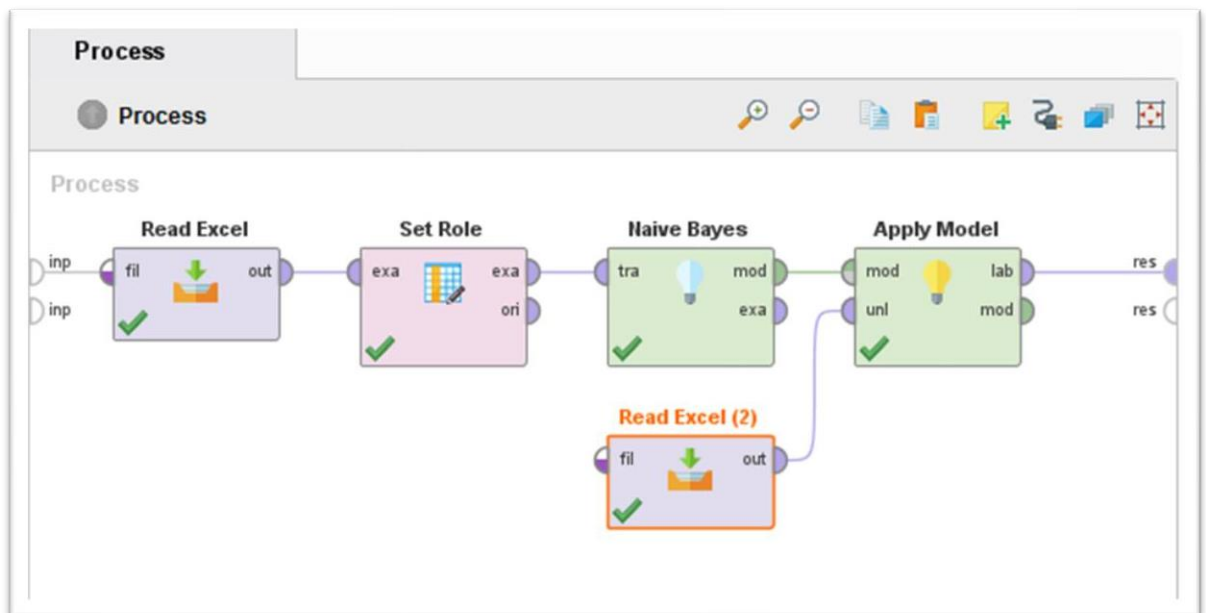
Pada tahap ini algoritma yang dipakai yaitu Naïve Bayes digambarkan ke dalam bentuk model yang akan diolah menggunakan aplikasi RapidMiner. Data set customer churn memiliki 3333 record data, 10 atribut, dan satu atribut churn, di mana data set ini didapatkan penulis di website Kaggle :

Churn	AccountWeeks	ContractRenewal	DataPlan	DataUsage	CustServCalls	DayMins	DayCalls	MonthlyCharge	AverageFee	RoamingMins
Yes	65	1	0	0.29	4	129.1	137	44.9	11.43	12.7
Yes	161	1	0	0	4	332.9	67	84	15.89	5.4
Yes	77	1	0	0	5	62.4	89	26	8.5	5.7
Yes	12	1	0	0	1	249.6	118	64	12.62	11.8
Yes	135	0	1	3.94	0	173.1	85	86.4	10.2	14.6
Yes	119	1	0	0	5	159.1	114	47	11.57	8.8
Yes	87	1	0	0	5	151	83	45	10.99	9.7
Yes	121	1	1	1.57	3	198.4	129	56.7	3.77	5.8
Yes	150	1	0	0	4	178.9	101	45	8.46	13.8
Yes	82	1	0	0.21	0	300.3	109	69.1	9.05	11.7
....
No	28	1	0	0	2	180.8	109	56	14.44	14.1

No	184	0	0	0	2	213.8	105	50	7.98	5
No	74	1	1	3.7	0	234.4	113	100	13.3	13.7

Tabel 1 Sampel Data

Peneliti melakukan evaluasi efektivitas dan kualitas satu model yang dikirim dalam tahap pemodelan sebelum menempatkannya untuk dipakai. Untuk membuat model naïve bayes klasifikasi dari data set “telecom_churn”, kita gunakan Rapid Miner Studio. Berikut cara melakukan nya:



Gambar 3 Proses Rapid Miner

Process operator ini, merupakan testing data yang peneliti yang tentukan dan memasukan data nya ke naïve bayes untuk menentukan prediksi atau hasil prediksi dan confidence Yes dan No.

Pada panel Process, kita gunakan 5 operator yang terdiri atas:

1. Read Excel : digunakan untuk membaca file excel yang telah kita buat dan simpan sebelumnya dengan nama “telecom_churn.xlsx”.
2. Set Role : digunakan untuk menentukan atribut yang akan menjadi label.
3. Naïve Bayes : digunakan untuk pengklasifikasi yang dapat membangun model yang baik bahkan dengan kumpulan data set kami yaitu “Customer Churn” dan membuat prediksi serta confidence dari label data peneliti.
4. Read Excel (2) : digunakan untuk membaca file excel yang telah kita buat dan simpan sebelumnya dengan nama “telecom_churn (Data Testing).xlsx”.
5. Apply Model : digunakan dalam menghasilkan perkiraan dari data yang belum dapat dilihat atau bagian mengatur kembali data dalam mengenakan pada model preprocessing dari example set yang kita tentukan dan di masukan sebelumnya.

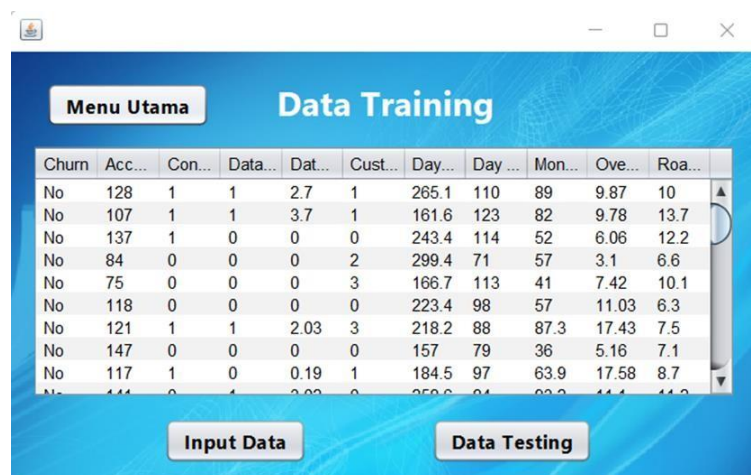
Peneliti membuat ExampleSet atau data testing untuk memprediksi dari data yang belum dapat dilihat atau dalam mengatur kembali data dengan mengenakan pada model preprocessing, seperti berikut hasilnya :

Predicti on(Churn)	Confiden ce (No)	Confiden ce (Yes)	Accou ntWee ks	Contra ctRene wal	DataPI an	Data Usag e
Yes	0.230584731	0.769415269	180	0	1	0.68
	CustServ Calls	DayMins	DayCalls	Monthly Charge	Overag eFee	Roam Mins
	2	135.1	36	65	2.46	1.1

Tabel 2 Hasil Apply Model

C. Tampilan Program

Antarmuka yang digunakan dalam penelitian ini diimplementasikan oleh aplikasi Netbeans dengan Bahasa pemrograman Java. Tampilan dalam program klasifikasi dalam memprediksi customer churn bisa dilihat pada gambar dibawah ini.



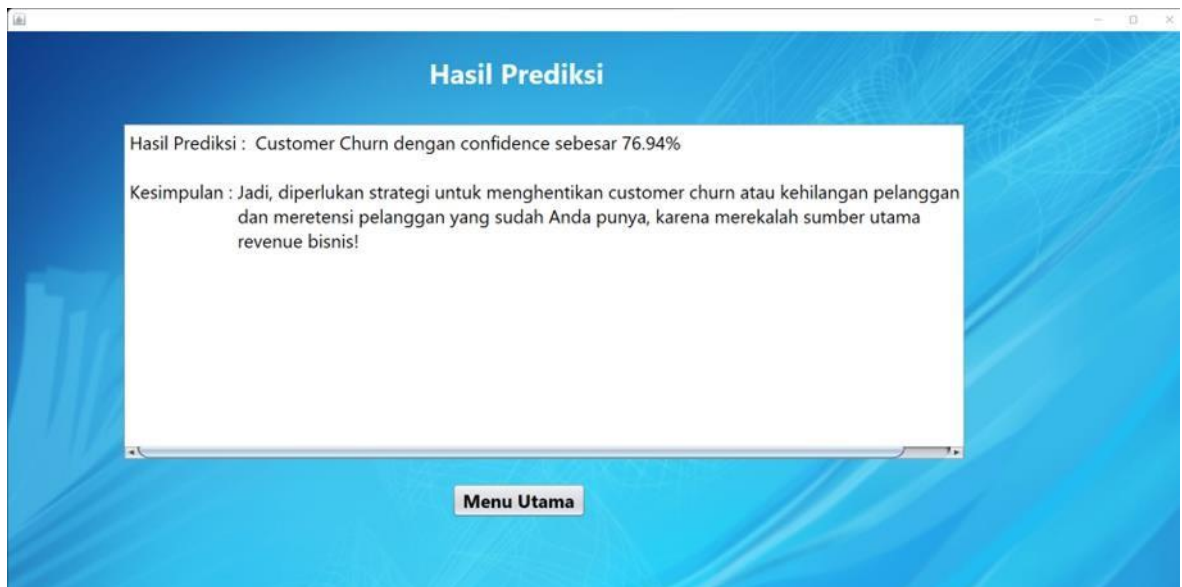
Gambar 4 Menu Data Training

Tampilan pada gambar 4 adalah daftar untuk melihat data set customer churn dan data training yang digunakan untuk melatih algoritma dalam mencari model yang sesuai untuk melakukan prediksi nantinya.



Gambar 5 Tampilan Data Testing

Tampilan pada gambar 5 untuk melakukan input data testing agar dapat mengetahui hasil prediksi dan keakuratan dari data yang di masukan yaitu data testing dan training.



Gambar 6 Tampilan Hasil Prediksi

Tampilan pada gambar 6 adalah menu yang bisa menyerahkan sesuatu informasi terkait prediksi customer atau pelanggan yang dinyatakan churn dan no churn sehingga pemilik usaha dan pihak pengelola usaha bisnis umkm dapat mengetahui seberapa besar tingkat terjadi nya customer churn untuk melakukan strategi pemasaran yang lebih atau lainnya.

D. Evaluasi dan Hasil

Data set yang sudah disebutkan melakukan pengujian dengan metode Naïve Bayes pada software rapid miner. Sehingga mengeluarkan nilai akurasi dari berikut ini.

	true No	true Yes	class precision
pred. No	610	67	90.10%

pred. Yes	48	39	44.83%
class recall	92.71%	36.79%	

Tabel 3 Nilai Akurasi

Nilai akurasi dengan confusion matrix tersebut merupakan berikut ini:

$$akurasi = \frac{(TN + TP)}{TN + FN + TP + P} = \frac{(610 + 39)}{610 + 67 + 48 + 39} = 0.8495 = 84.95\%$$

Beralaskan hasil evaluasi dan validasi dapat mengetahui bahwa akurasi dari algoritma Naive Bayes adalah 84,95%, sehingga aturan yang dihasilkan oleh algoritma Naive Bayes berguna untuk membangun interface untuk memprediksi churn pelanggan potensial dalam usaha bisnis.

Berdasarkan hasil yang telah sedia maka memperoleh nilai AUC dengan tahap klasifikasi berikut sebesar 83,20% yang dapat ditandai hasil Interpretasi dengan nilai AUC berada dengan titik "Baik".

Nilai AUC	Interpretasi
>50-60%	Sangat Lemah
>60-70%	Lemah
>70-80%	Sedang
>80-90%	Baik
90-100%	Sangat Baik

Tabel 4 Interpretasi Nilai Area Under the Curve[7]

Kesimpulan

Beralaskan penelitian yang sudah penulis lakukan bisa diambil kesimpulan :

1. Atribut account weeks, contract renewal, data plan, data usage, customer service calls, day mins, day calls, monthly charge, dan overage fee dapat digunakan untuk melakukan prediksi customer churn.
2. Hasil accuracy prediksi untuk mencegah adanya customer churn atau kehilangan pelanggan dalam suatu usaha bisnis umkm, mengurangi adanya customer churn, serta sebagai hasil pembuat strategi bisnis untuk mempertahankan pelanggan dan meningkatkan loyalitas dari pelanggan. Sebab dari prediksi dalam memakai metode Naive Bayes hal ini cuma mempunyai nilai accuracy yaitu 84,95% dan nilai AUC adalah 83,20% yang ada di titik "Baik" hingga dalam hal klasifikasi ini bisa dipakai dan berjalan sesuai dalam yang dibutuhkan user.
3. Penetapan mendasari dalam memakai nilai akhir pada probabilitas customer churn dan probabilitas customer no churn. Jika nilai probabilitas customer churn lebih besar dari probabilitas customer no churn sehingga pelanggan dapat di prediksi akan churn atau tidak memakai layanan/produk di usaha tersebut. Kebalikannya apabila nilai probabilitas customer no churn lebih besar dari probabilitas churn sehingga pelanggan akan tetap membutuhkan layanan/produk yang tersedia di usaha tersebut.

Ucapan Terima Kasih

Terima Kasih kepada ibu Desiyanna Lasut selaku dosen pembimbing yang selalu membantu dan memberikan masukan dalam proses penyusunan hasil penelitian berlangsung.

Terima Kasih kepada Fakultas Sains dan Teknologi Universitas Buddhi Dharma sebagai tempat penelitian dilakukan.

Referensi :

- [1] Azmi, Z. & Dahria, M., 2013. Decision Tree Berbasis Algoritma Untuk Pengambilan Keputusan. *Jurnal Ilmiah SAINTIKOM*, pp. 157-164.
- [2] Ridwan, M., Suyono, H. & Sarosa, M., 2013. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*, pp. 59-64.
- [3] Manalu, E., Sianturi, F. A. & Manalu, M. R., 2017. Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan Dan Jumlah Pemesanan Pada Cv. Papadan Mama Pastries. *Jurnal Manajemen Dan Informatika Pelita*, pp. 16-21.
- [4] Kurniawan, Y. I., 2018. Perbandingan Algoritma Naive Bayes dan C. 45 Dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, pp. 455-464.
- [5] Prianto, D., Fahmi, I. & Sartono, B., 2020. Strategi Manajemen Pelanggan Internet Rumah Pascabayar Berdasarkan Faktor yang Berpengaruh Terhadap Churn. *Inovbiz: Jurnal Inovasi Bisnis* 8, pp. 128-133.
- [6] Fadillah, A. P., 2015. Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ). *Jurnal Teknik Informatika dan Sistem Informasi*, pp. 260-270.
- [7] Ilham, A., 2017. KOMPARASI ALGORITMA KLASIFIKASI DENGAN PENDEKATAN LEVEL DATA UNTUK MENANGANI DATA KELAS TIDAK SEIMBANG. *Jurnal Ilmiah Ilmu Kompute*, pp. 1-6.