

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



**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS BUDDHI DHARMA
TANGERANG**

2022

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

SKRIPSI

Diajukan sebagai salah satu syarat untuk kelengkapan gelar kesarjanaan pada

Program Studi Teknik Informatika

Jenjang Pendidikan Strata 1



Disusun oleh:

STEPHEN

20181000031

TEKNIK INFORMATIKA

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS BUDDHI DHARMA
TANGERANG**

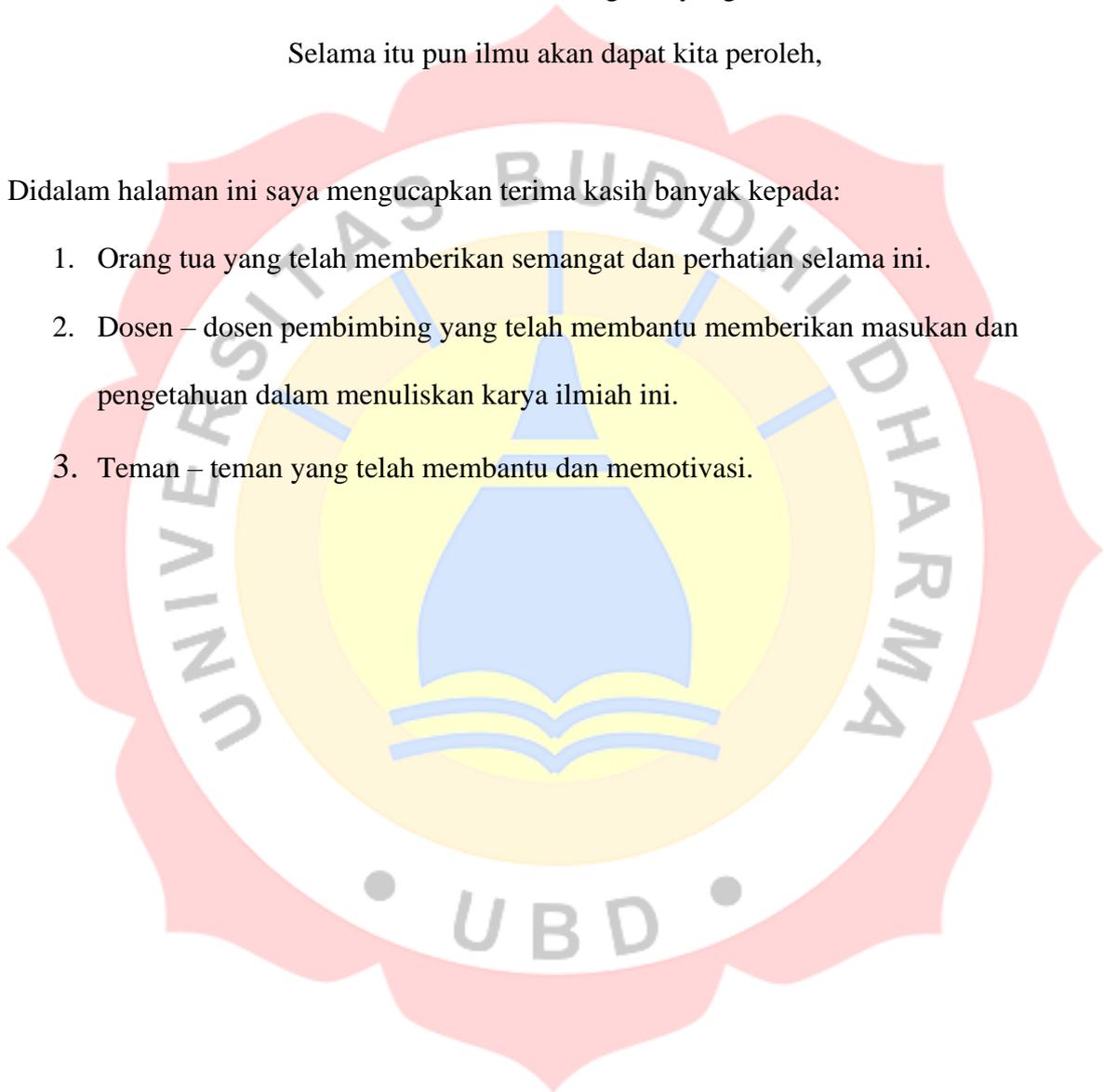
2022

LEMBAR PERSEMBAHAN

Ilmu dapat pelajari dimana pun dan kapan pun,
Tanpa terhalang oleh tempat dan waktu,
Selama ada niat dan keinginan yang kuat,
Selama itu pun ilmu akan dapat kita peroleh,

Didalam halaman ini saya mengucapkan terima kasih banyak kepada:

1. Orang tua yang telah memberikan semangat dan perhatian selama ini.
2. Dosen – dosen pembimbing yang telah membantu memberikan masukan dan pengetahuan dalam menuliskan karya ilmiah ini.
3. Teman – teman yang telah membantu dan memotivasi.



UNIVERSITAS BUDDHI DHARMA

LEMBAR PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini,
NIM : 20181000031
Nama : Stephen
Jenjang Studi : Strata 1
Program Studi : Teknik Informatika
Peminatan : Database

Dengan ini saya menyatakan bahwa:

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapat gelar akademik (Diploma/Sarjana) atau kelengkapan studi, baik di Universitas Buddhi Dharma maupun di Perguruan Tinggi lainnya.
2. Skripsi ini saya buat sendiri tanpa bantuan dari pihak lain, kecuali arahan dosen pembimbing.
3. Dalam Skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dan dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan daftar pustaka.
4. Dalam Skripsi ini tidak terdapat pemalsuan (kebohongan), seperti buku, artikel, jurnal, data sekunder, pengolahan data, dan pemalsuan tanda tangan dosen atau Ketua Program Studi Universitas Buddhi Dharma yang dibuktikan dengan keasliannya.
5. Lembar pernyataan ini saya buat dengan sesungguhnya, tanpa paksaan dan apabila dikemudian hari atau pada waktu lainnya terdapat penyimpangan dan ketidak benaran dalam pernyataan ini, saya bersedia menerima sanksi akademik berupa pencabutan gelar akademik yang telah saya peroleh karena Skripsi ini serta sanksi lainnya sesuai dengan peraturan dan norma yang berlaku.

Tangerang, 28 Juni 2022

Penulis,



Stephen

UNIVERSITAS BUDDHI DHARMA

LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH

NIM : 20181000031
Nama : Stephen
Jenjang Studi : Strata 1
Program Studi : Teknik Informatika
Peminatan : DATABASE

Dengan ini menyetujui untuk memberikan ijin kepada pihak Universitas Buddhi Dharma, Hak Bebas Royalti Non – Eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah kami yang berjudul: “Implementasi Metode Algoritma *Naive Bayes* Untuk Memprediksi *Customer Churn* Pada Bisnis UMKM”, beserta alat yang diperlukan (apabila ada).

Dengan Hak Bebas Royalti Non – Eksklusif ini pihak Universitas Buddhi Dharma berhak menyimpan, mengalih-media atau format-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya, dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis atau pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Universitas Buddhi Dharma, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 28 Juni 2022

Penulis,



Stephen

20181000031

UNIVERSITAS BUDDHI DHARMA

LEMBAR PENGESAHAN PEMBIMBING

**IMPLEMENTASI METODE ALGORITMA *NAIVE BAYES* UNTUK
MEMPREDIKSI *CUSTOMER CHURN* PADA BISNIS UMKM**

Dibuat Oleh:

NIM : 20181000031

Nama : Stephen

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian
Komprehensif

Program Studi Teknik Informatika

Peminatan Basis Data

Tahun Akademik 2021/2022

Disahkan oleh,

Tangerang, 28 Juni 2022

Pembimbing,



Desiyanna Lasut, S.Kom., M.Kom

NIDN. 0402128601

UNIVERSITAS BUDDHI DHARMA

LEMBAR PENGESAHAN SKRIPSI

**IMPLEMENTASI METODE ALGORITMA NAIVE BAYES UNTUK
MEMPREDIKSI CUSTOMER CHURN PADA BISNIS UMKM**

Dibuat Oleh:

NIM : 20181000031

Nama : Stephen

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian
Komprehensif

Program Studi Teknik Informatika

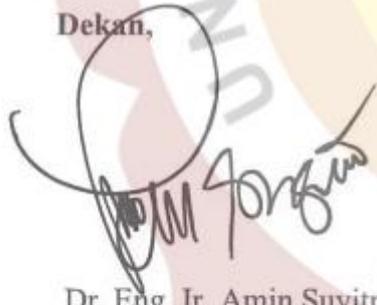
Peminatan Basis Data

Tahun Akademik 2021/2022

Disahkan oleh,

Tangerang, 28 Juni 2022

Dekan,



Dr. Eng, Ir. Amin Suyitno, M.Eng.

NIDK. 8826333420

Ketua Program Studi,



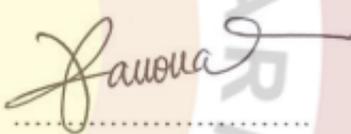
Hartana Wijaya, S.Kom., M.Kom.

NIDN. 0412058102

LEMBAR PENGESAHAN TIM PENGUJI

Nama : Stephen
NIM : 20181000031
Fakultas : Sains dan Teknologi
Judul Skripsi : IMPLEMENTASI METODE ALGORITMA *NAIVE BAYES*
UNTUK MEMPREDIKSI *CUSTOMER CHURN* PADA
BISNIS UMKM

Dinyatakan LULUS setelah mempertahankan di depan Tim Penguji pada hari Senin, 01 Agustus 2022

	Nama penguji :	Tanda Tangan :
Ketua Sidang	: Rino, M.Kom NIDN. 0420058502	
Penguji I	: Ramona Dyah Safitri, S.Si, M.Si NIDN. 0420039301	
Penguji II	: Desiyanna Lasut, S.Kom, M.Kom NIDN. 0402128601	

Mengetahui,
Dekan Fakultas Sains dan Teknologi


Dr. Eng. Ir. Amin Suyitno, M.Eng.

NIDK. 8826333420

KATA PENGANTAR

Dengan mengucapkan Puji Syukur kepada Tuhan Yang Maha Esa, yang telah memberikan Rahmat dan karunia-Nya kepada penulis sehingga dapat menyusun dan menyelesaikan Skripsi ini dengan judul **Implementasi Metode Algoritma Naive Bayes Untuk Memprediksi Customer Churn Pada Bisnis UMKM**. Tujuan utama dari pembuatan Skripsi ini adalah sebagai salah satu syarat kelengkapan dalam menyelesaikan program pendidikan Strata 1 Program Studi Teknik Informatika di Universitas Buddhi Dharma. Dalam penyusunan Skripsi ini penulis banyak menerima bantuan dan dorongan baik moril maupun materiil dari berbagai pihak, maka pada kesempatan ini penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

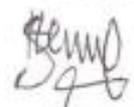
1. Bapak Dr. Suryadi Winata, SE, MM, M.Si., Ak., CA, CMA,CBV, ACPA, CPA (Aust.), sebagai Plt. Rektor Universitas Buddhi Dharma
2. Bapak Dr. Eng, Ir. Amin Suyitno, M.Eng., Dekan Fakultas Sains dan Teknologi.
3. Bapak Hartana Wijaya, S.Kom., M.Kom., sebagai Ketua Program Studi Teknik Informatika
4. Ibu Desiyanna Lasut, S.Kom., M.Kom, sebagai pembimbing yang telah membantu dan memberikan dukungan serta harapan untuk menyelesaikan penulisan Skripsi ini.
5. Orang tua dan keluarga yang selalu memberikan dukungan baik moril dan materiil.
6. Teman-teman yang selalu membantu dan memberikan semangat

Serta semua pihak yang terlalu banyak untuk disebutkan satu-persatu sehingga terwujudnya penulisan ini. Penulis menyadari bahwa penulisan Skripsi ini masih belum sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan di masa yang akan datang.

Akhir kata semoga Skripsi ini dapat berguna bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Tangerang 06 Juli 2022

Penulis,


Stephen

IMPLEMENTASI METODE ALGORITMA *NAIVE BAYES* UNTUK MEMPREDIKSI *CUSTOMER CHURN* PADA BISNIS UMKM.

(84 halaman + xii / 9 tabel / 26 gambar / 21 pustaka / lampiran)

ABSTRAK

Penerapan teknologi saat ini mulai merambah dan semakin meningkat di berbagai bidang. Keseluruhan aktivitas di sektor korporasi semakin terkait dengan pengaruh dan dampak teknologi. Macam - macam aplikasi atau *software* komputer tersedia secara melebar yang memungkinkan pengusaha menggunakannya untuk mengelola operasi bisnis mereka. 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 *naive 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". *Naive bayes* membuktikan bisa memprediksi kemungkinan *customer churn* berdasarkan data dan klasifikasi sehingga hasilnya bisa mendukung menjadi keputusan bisnis.

Kata kunci : Algoritma *Naive Bayes*, Data mining, Kehilangan pelanggan, RapidMiner, Netbeans

IMPLEMENTATION OF NAIVE BAYES ALGORITHM METHOD FOR PREDICTING CUSTOMER CHURN IN MSME BUSINESS.

(84 pages + xii / 9 tables / 26 pictures / 21 libraries / appendices)

ABSTRACT

Currently the application of technology is growing and is starting to penetrate various fields. All activities carried out by a business sector are increasingly inseparable from the influence of technology. Various computer applications are widely available that enable entrepreneurs to apply them in managing their business activities. In a UMKM business, in general, they have customer data that is a member of shopping with the aim of establishing a mutually beneficial relationship between a customer and a business, but customer data is often not managed and analyzed properly, causing customers to stop shopping in a business or move to another business. other competitors. To be able to predict customer loss, we need an algorithm model that can predict customer churn accurately. To determine the prediction results, an attempt is made to analyze data mining techniques with the Naive Bayes algorithm so that a business can take preventive actions. The training and testing data were entered into the nave Bayes modeling with an accuracy of 84.95% and AUC of 83.20%. From the accuracy and AUC this indicates the result of the interpretation of the value is at the "Good" point. Naïve Bayes is proven to be able to predict the possibility of customer churn based on data and classification so that the results can be used to support business decisions.

Keywords: *Algorithm Naïve Bayes, Data mining, Customer churn, RapidMiner, Netbeans*

DAFTAR ISI

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

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

LEMBAR PERSEMBAHAN

LEMBAR PERNYATAAN KEASLIAN SKRIPSI

LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH

LEMBAR PENGESAHAN PEMBIMBING

IMPLEMENTASI METODE ALGORITMA *NAIVE BAYES* UNTUK MEMPREDIKSI
CUSTOMER CHURN PADA BISNIS UMKM

LEMBAR PENGESAHAN SKRIPSI

LEMBAR PENGESAHAN TIM PENGUJI

KATA PENGANTAR..... i

ABSTRAK ii

ABSTRACT iii

DAFTAR ISI iv

DAFTAR GAMBAR..... ix

DAFTAR TABEL xi

DAFTAR LAMPIRAN xii

BAB I PENDAHULUAN..... 1

1.1 Latar Belakang Masalah 1

1.2	Identifikasi Masalah.....	3
1.3	Rumusan Masalah.....	4
1.4	Ruang Lingkup	4
1.5	Tujuan dan Manfaat	5
1.5.1	Tujuan.....	5
1.5.2	Manfaat.....	5
1.6	Metodologi Penelitian dan Pengumpulan Data	6
1.6.1	Teknik Pengumpulan Data	6
1.6.2	Metode Penelitian.....	7
1.7	Sistematika Penulisan	8
BAB II	LANDASAN TEORI.....	10
2.1	Teori Umum.....	10
2.1.1	Data.....	10
2.1.2	Informasi.....	12
2.1.3	Prediksi	13
2.1.4	Usaha Mikro, Kecil, dan Menengah.....	14
2.1.5	Pelanggan	16
2.1.6	<i>Customer Churn</i>	17
2.2	Teori Khusus.....	18
2.2.1	<i>Data Mining</i>	18
2.2.2	Bahasa Pemrograman <i>Java</i>	19
2.2.3	Algoritma <i>Naive Bayes</i>	21

2.3	Teori Analisa dan Perancangan	22
2.3.1	<i>Cross-Industry Standard Process for Data Mining (CRISP-DM)</i>	22
2.3.2	Aplikasi <i>Rapid Miner</i>	26
2.3.3	Aplikasi <i>Netbeans IDE</i>	26
2.4	Tinjauan Studi	27
2.4.1	Penelitian	27
2.4.1	Rangkuman Model Penelitian	31
2.5	Kerangka Pemikiran	33
BAB III	ANALISA MASALAH & PERANCANGAN APLIKASI	35
3.1	Analisa Kebutuhan	35
3.1.1	Data Set Customer Churn	35
3.1.2	Analisa Perangkat Lunak Untuk Membangun Sistem	37
3.1.3	Analisa Kebutuhan Pemakai	37
3.1.4	Analisa Kebutuhan Aplikasi	38
3.2	Latar Belakang Penggunaan Metode Algoritma Naïve Bayes	39
3.2.1	<i>Bussiness Understanding</i>	40
3.2.2	<i>Data Understanding</i>	40
3.2.3	<i>Data Preparation</i>	43
3.2.4	<i>Modeling</i>	45
3.2.5	<i>Evaluation</i>	48
3.2.6	<i>Deployment</i>	51
3.3	Perancangan Antarmuka	51

3.3.1	Perancangan Halaman Login.....	51
3.3.2	Perancangan Menu Utama.....	52
3.3.3	Perancangan Menu Data Training	53
3.3.4	Perancangan Input Data Training	54
3.3.5	Perancangan Data Testing	55
3.3.6	Perancangan Menu Hasil Prediksi.....	55
3.4	Jadwal Penelitian	56
BAB IV PENGUJIAN DAN IMPLEMENTASI		58
4.1	Hasil Penelitian	58
4.1.1	Hitungan Manual	58
4.1.2	<i>Flowchart Dengan Naïve Bayes</i>	67
4.2	Tampilan Program	71
4.2.1	Tampilan Halaman Login.....	71
4.2.2	Tampilan Menu Utama.....	72
4.2.3	Tampilan Menu Data Training	73
4.2.4	Tampilan Menu Input Data Training.....	74
4.2.5	Tampilan Data <i>Testing</i>	74
4.2.6	Tampilan Menu Hasil Prediksi.....	75
4.3	Pengujian Prediksi	75
4.4	Pengujian Sistem.....	77
4.5	Evaluasi dan Hasil	81
BAB V SIMPULAN DAN SARAN		83

5.1 Kesimpulan	83
5.2 Saran	84
DAFTAR PUSTAKA.....	85
DAFTAR RIWAYAT HIDUP	88



DAFTAR GAMBAR

Gambar 2. 1 CRISP-DM	25
Gambar 2. 2 Kerangka Pemikiran	33
Gambar 3. 1 Analisa Kebutuhan Aplikasi	38
Gambar 3. 2 <i>Data Preparation Customer Churn</i>	43
Gambar 3. 3 Import Data Set ke Rapid Miner	44
Gambar 3. 4 Proses Set Role Data Transform di Rapid Miner	44
Gambar 3. 5 Hasil Data Transform	45
Gambar 3. 6 Proses <i>Modeling</i> di <i>Rapid Miner</i>	45
Gambar 3. 7 Proses Prediksi di <i>Rapid Miner</i>	47
Gambar 3. 8 <i>Simple Distribution</i>	49
Gambar 3. 9 Tampilan Menu Login	52
Gambar 3. 10 Tampilan Utama	53
Gambar 3. 11 Tampilan Menu Data Training	54
Gambar 3. 12 Tampilan Input Data Training	54
Gambar 3. 13 Tampilan Menu Data Testing	55
Gambar 3. 14 Tampilan Hasil Prediksi	56
Gambar 4. 1 Tampilan Halaman Login	71
Gambar 4. 2 Tampilan Menu Utama	72
Gambar 4. 3 Tampilan Menu Data Training	73
Gambar 4. 4 Menu Input Data Training	74
Gambar 4. 5 Tampilan Data Testing	74
Gambar 4. 6 Tampilan Menu Hasil Prediksi	75
Gambar 4. 7 Pengujian Data Testing	76

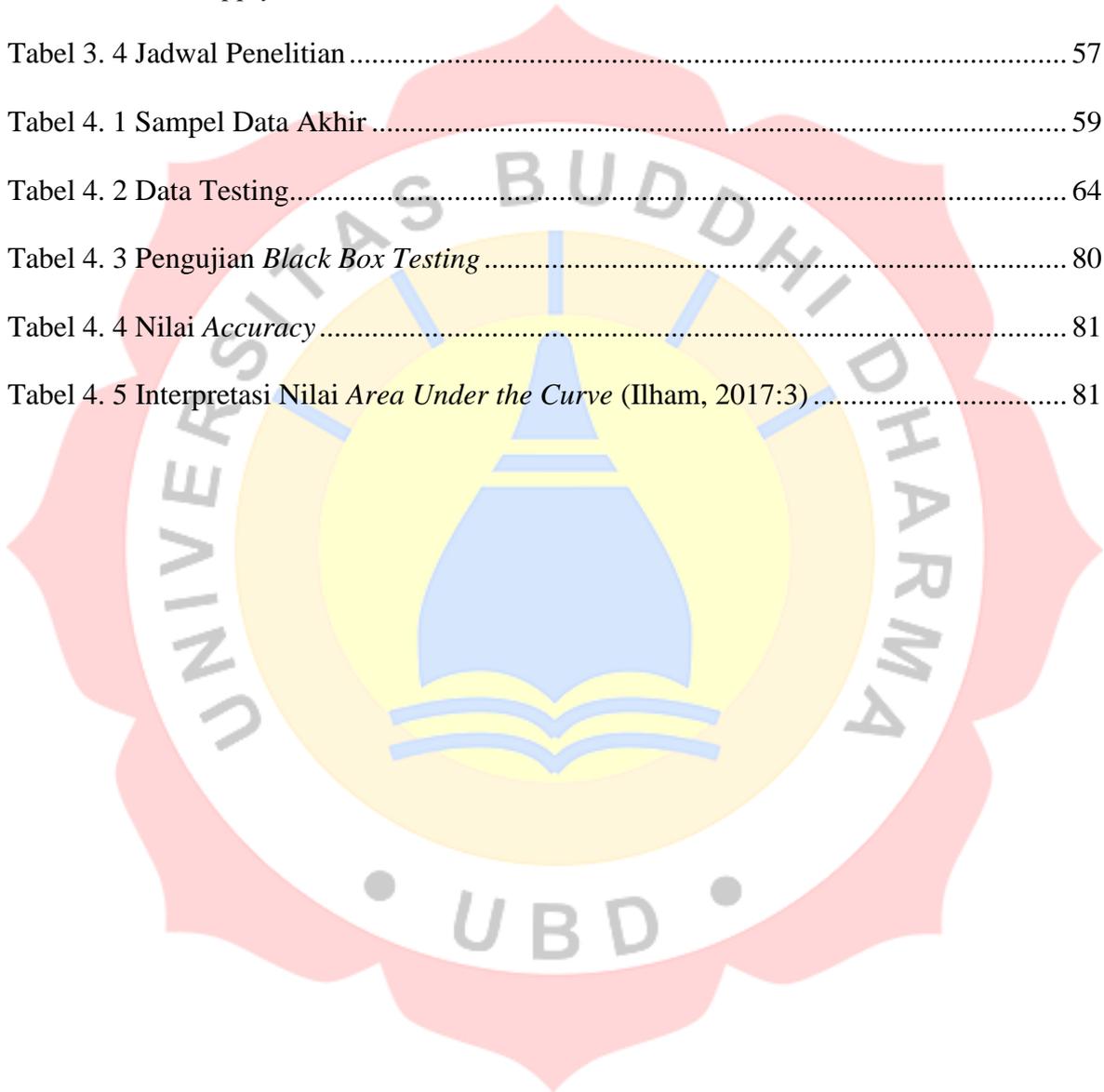
Gambar 4. 8 Pengujian Hasil Prediksi..... 77

Gambar 4. 9 Nilai AUC dengan metode *Naive Bayes*..... 82



DAFTAR TABEL

Tabel 3. 1 Data Set Customer Churn.....	37
Tabel 3. 2 Distribution Table.....	49
Tabel 3. 3 Hasil <i>Apply Model</i>	50
Tabel 3. 4 Jadwal Penelitian	57
Tabel 4. 1 Sampel Data Akhir	59
Tabel 4. 2 Data Testing.....	64
Tabel 4. 3 Pengujian <i>Black Box Testing</i>	80
Tabel 4. 4 Nilai <i>Accuracy</i>	81
Tabel 4. 5 Interpretasi Nilai <i>Area Under the Curve</i> (Ilham, 2017:3)	81



DAFTAR LAMPIRAN

Lampiran 1 Data

Lampiran 2 Requirement Elicitation

Lampiran 3 Kartu Bimbingan TA/Skripsi



BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

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.

Usaha Mikro, Kecil dan Menengah (UMKM) merupakan salah satu metode bisnis yang populer di masyarakat (Maulidasari dan Setiyana, 2020:63). Banyaknya penjual UMKM menjadikan bisnis ini sebagai salah satu penggerak roda perekonomian negara. Usaha mikro, kecil dan menengah merupakan aktivitas komersial atau bisnis yg dijalankan perorangan, rumah tangga & bisnis kecil. Klasifikasi tersebut didasarkan pada jumlah penjualan tahunan, jumlah kekayaan atau aset, dan jumlah karyawan yang dipekerjakan.

Pelanggan bagi usaha bisnis merupakan aset utama (Warnaningtyas, 2020:55), 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 usah dan beralih ke pesaing (*churn*).

Customer churn adalah bisnis yang kehilangan pelanggan. Tingkat churn dihitung berdasarkan jumlah pelanggan yang meninggalkan bisnis yang di miliki dalam waktu

tertentu (Arifin, 2014:279). *Customer churn* penting untuk diketahui bisnis karena merupakan gambaran keberhasilan perusahaan dalam mempertahankan pelanggan. Di persaingan pasar yang kompetitif saat ini, sebagian besar pelanggan mencari produk yang memenuhi kebutuhan mereka dan layanan yang lebih baik dengan harga yang lebih rendah. Ketika perusahaan dapat mengurangi atau mencegah churn pelanggan, mereka dapat meningkatkan *customer lifetime value* (CLV). CLV adalah jumlah total yang dapat di harapkan dari rata-rata pelanggan untuk dibelanjakan pada bisnis yang dimiliki selama masa hidup mereka.

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*.

Dengan menggunakan prediksi *churn* atau pelanggan dapat menentukan dan mengidentifikasi prediksi pelanggan sejak awal sebelum pindah ke usaha pesaing lain, ini juga dapat membantu CRM (*Customer Relationship Management*) yaitu untuk menjaga pelanggan, dan mengurangi potensi kerugian yang ditemui (Dewi dan Samuel, 2015:2). Teknik data mining, atau klasifikasi, diperlukan untuk membuat prediksi dari sejumlah besar data atau data berbentuk. Salah satu metode yang digunakan dalam klasifikasi adalah *Naive Bayes*.

Algoritma *Naive Bayes* merupakan salah satu algoritma dalam teknik klasifikasi (Indrayuni, 2019:30). *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.

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”**.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dijelaskan, masalah yang dapat diidentifikasi, yaitu :

1. Belum adanya pendataan yang akurat terhadap *customer churn* pada suatu bisnis UMKM.
2. Belum adanya hasil prediksi dalam ketepatan akurasi untuk memprediksi *customer churn* dalam beberapa atribut klasifikasi data suatu usaha bisnis.

3. Belum adanya software yang efisien untuk membantu memprediksi *customer churn* untuk pemilik usaha bisnis UMKM.

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah tersebut, maka terdapat beberapa rumusan masalah yaitu:

1. Bagaimana cara membuat hasil prediksi pada *customer churn* dari pelanggan suatu bisnis yang di miliki ?
2. Bagaimana tingkat akurasi untuk memprediksi *customer churn* dalam beberapa atribut klasifikasi data suatu usaha bisnis ?
3. Bagaimana cara mengimplementasikan Data Mining untuk membuat *software* yang dapat memprediksi *customer churn* dengan menggunakan metode *Naïve Bayes* ?

1.4 Ruang Lingkup

Ruang lingkup yang akan diteliti yaitu :

1. Data yang digunakan dalam Skripsi ini berupa data sekunder yang diperoleh dari website dan platform Kaggle. Dalam data ini adalah data berhenti berlangganan pelanggan dalam bisnis.
2. Penelitian ini dilakukan untuk memprediksi *customer churn* dalam suatu bisnis UMKM dari data pelanggan yang pernah melakukan transaksi.
3. Metode Statistik yang digunakan adalah data mining klasifikasi dengan algoritma *Naïve Bayes*.
4. Software dan perangkat lunak untuk membantu penelitian yang digunakan adalah Netbeans, Rapid Miner, dan Microsoft Excel.

5. Atribut yang akan dipakai berkaitan dengan industri telekomunikasi

1.5 Tujuan dan Manfaat

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.

1.5.1 Tujuan

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 *Naïve 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 *Naïve Bayes* dengan pengukuran tingkat ketepatan Akurasi yang dapat digunakan sebagai strategi pemasaran.

1.5.2 Manfaat

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*.

1.6 Metodologi Penelitian dan Pengumpulan Data

1.6.1 Teknik Pengumpulan Data

A. Metode Dokumentasi

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah penggunaan studi dokumenter yang dilakukan dengan pengumpulan data menggunakan data berupa buku, catatan (dokumen) sebagai metode dokumenter, sumber informasi berupa dokumen atau catatan. Dalam pendekatan ini, pengumpulan data cukup mentransfer dokumen teks yang relevan ke dalam lembar yang telah disiapkan untuk mereka sesuai kebutuhan.

Dokumentasi adalah teknik pengumpulan data melalui studi profil data pribadi responden. Salah satu metode pengumpulan data adalah kualitatif dengan melihat dan menganalisis subjek atau dokumen yang ditulis oleh orang lain tentang suatu hal tersebut. Teknik ini digunakan untuk mendapatkan data yang berkenaan dengan atribut-atribut penyebab terjadinya *customer churn* pada suatu data di usaha bisnis yang di gunakan dalam penelitian ini.

B. Studi Pustaka

Untuk mendukung penelitian ini, penulis juga menggunakan penelitian studi pustaka melalui berbagai sumber media seperti internet, buku, dan sumber lainnya.

1.6.2 Metode Penelitian

Metode penelitian ini akan memakai model standarisasi data mining yang bisa disebut *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk mendapatkan sebuah visualisasi pola-pola yang terdapat dalam basis data. Langkah-langkah yang ada didalamnya yaitu:

A. *Business Understanding*

Pengetahuan tentang objek bisnis dilakukan pada langkah ini, dengan cara membuat atau mendapatkan data, dan untuk menyatukan menjadi sesuai tujuan pemodelan pada suatu usaha bisnis untuk membuat model yang paling baik. Tahapan yang dapat digunakan antara lain: menentukan dengan jelas tujuan dan persyaratan penulis secara menyeluruh, mengubah tujuan ini untuk dapat di tentukan batasan untuk merumuskan masalah data mining penulis, dan kembangkan rencana awal yang dapat mencapai suatu tujuan ini.

B. *Data Understanding*

Dalam langkah ini antara lain melakukan dengan memberikan dasar analitis untuk studi dengan meringkas dan menganalisa kemungkinan masalah dalam suatu data yang di dapatkan. Sekali lagi, pada tahap ini harus bisa melakukan dengan lebih teliti dan tidak tergesa-gesa, seperti halnya visualisasi data, yang mungkin susah dipahami dalam hubungannya dengan peringkasan data.

C. *Data Preparation*

Langkah ini memerlukan pertimbangan yang efektif dan bisa mengupayakan dengan baik dalam membuat kepastian bahwa data tersebut cocok dalam algoritma yang dipakai.

D. Modelling

Dalam langkah ini peneliti menggunakan metode statistik dan data mining dalam menentukan teknik data mining, alat data mining, dan algoritma data mining, atau *Naive Bayes* yang akan diterapkan.

E. Evaluation

Pada langkah ini mencakup pemeriksaan informasi yang dihasilkan dari proses sebelumnya terhadap fakta maupun hipotesa yang telah ada dan melakukan pengujian akurasi terhadap hasil penelitian.

F. Deployment

Selama fase ini, wawasan dan informasi yang diperoleh, dipresentasikan, diatur dan disajikan dalam format khusus untuk konsumsi pengguna.

1.7 Sistematika Penulisan

Untuk lebih jelas dalam memahami penelitian ini, peneliti mengelompokkan materi yang tertulis dalam penelitian ini menjadi beberapa sub bab dengan sistematika penulisan sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini berisi pendahuluan mengenai latar belakang masalah, identifikasi masalah, rumusan masalah, ruang lingkup, metodologi penelitian dan pengumpulan data serta sistematika penelitian.

BAB II LANDASAN TEORI

Pada bab ini berisi landasan teori – teori yang digunakan untuk menganalisis masalah yang muncul serta perencanaan penelitian yang dilakukan oleh peneliti. Selain itu berisi teori pengertian dan definisi yang berasal dari kutipan buku, jurnal, dan website yang berkaitan dengan penelitian tersebut.

BAB III ANALISA KEBUTUHAN DAN PERANCANGAN APLIKASI

Pada bab ini berisi tentang analisis masalah dalam *customer churn* di suatu usaha UMKM dengan menggunakan algoritma *Naïve Bayes* pada klasifikasi pelanggan khususnya yang pernah melakukan transaksi di suatu usaha serta konstruksi algoritma yang akan digunakan.

BAB IV PENGUJIAN DAN IMPLEMENTASI

Pada bab ini berisi pembahasan serta implementasi algoritma *Naïve Bayes*, flowchart aplikasi, serta tampilan sederhana aplikasi, serta evaluasi penelitian.

BAB V KESIMPULAN DAN SARAN

Pada bab ini berisi pembahasan kesimpulan dari penelitian serta saran yang diberikan peneliti terhadap penelitian yang telah dilakukan dan yang akan dilanjutkan jika ingin diangkat kembali sebagai perbaikan penelitian berikut

BAB II

LANDASAN TEORI

2.1 Teori Umum

Dalam bab ini penulis menjelaskan pembahasan tentang materi penelitian secara teoritis yang menyangkut tentang teori - teori para ahli tentang penelitian data mining khususnya dalam metode klasifikasi ini.

2.1.1 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 (Azmi dan Dahria, 2013:158). Keputusan yang baik hanya dapat diperoleh dari mereka yang objektif dan berdasarkan data yang baik.

Data masih bersifat mentah atau masih utuh. Jadi, jika ingin mendapatkan data yang baik dan akurat, sangat penting untuk mengandalkan data yang dapat diandalkan karena kebenaran, akurasi, ketepatan waktu, dan jangkauannya yang luas. Karena suatu data yang digunakan masih mentah, penulis tidak mendapatkan informasi yang lengkap. Tidak heran jika penulis perlu mengolah data yang penulis terima hingga benar-benar dapat menghasilkan informasi yang dapat dipahami dari data tersebut.

Dalam penelitian, seseorang selalu membutuhkan data. Data dapat diubah menjadi sumber informasi yang tepercaya. Pernyataan seperti asumsi yang mendasari relatif tidak menjadi perhatian besar untuk penelitian kualitatif, karena alat terbaik untuk memahami data penelitian adalah yang paling mungkin untuk mengarah pada pemahaman yang maksimal tentang seluruh fenomena yang diteliti dalam konteks (Nugrahani, 2014:57). Jadi penulis membutuhkan data untuk menjelaskan hal-hal yang berbeda. Di bawah ini adalah beberapa fungsi data, yaitu :

1. Data yang digunakan sebagai dokumen penilaian untuk menjawab suatu masalah tertentu
2. Data dipakai buat memecahkan masalah dan memilih kebijakan & keputusan
3. Data yang diperoleh juga bisa dijadikan acuan pada setiap aplikasi suatu aktivitas atau aktivitas
4. Data yang mendasari penelitian atau perencanaan dalam suatu kegiatan

Dalam penggunaan statistik, sangat penting untuk memiliki pemahaman yang jelas tentang definisi data dan jenis data (Sholikhah, 2016:344). Data dapat dikelompokkan ke dalam banyak kategori yang berbeda, salah satunya didasarkan pada bagaimana diperolehnya. Ada dua cara dalam mendapatkan data tersebut, yaitu :

- a. Data primer, merupakan data yang diteliti, didapatkan dan dikumpulkan langsung mengenai orang-orang oleh suatu organisasi atau individu.
- b. Data sekunder, merupakan data yang memiliki asal dari sumber lain yang sudah ada. Ini berarti bahwa penggunaan data sekunder menghilangkan kebutuhan untuk mengumpulkan data langsung dari subjek yang pernah

di teliti sebelumnya. Seringkali jenis data ini dapat diperoleh dari survei yang dilakukan sebelumnya. Bisa dalam bentuk grafik, tabel, atau diagram.

2.1.2 Informasi

Binarso dkk (2012:73) mengungkapkan: “Informasi adalah data yang diproses dalam bentuk yang berguna bagi pengguna”. Orang membutuhkan informasi untuk menambah wawasan, memperbarui pengetahuan, dan mendokumentasikan suatu opini. Bahkan tidak jarang informasi dijadikan sebagai dasar pengambilan keputusan.

Suatu informasi harus memenuhi persyaratan yang sesuai agar dapat dijadikan langkah dalam pengambilan keputusan yang berkualitas. Berikut persyaratan untuk memenuhi suatu informasi yang berkualitas yaitu :

a. Informasi harus akurat

Sebuah informasi dapat dikategorikan baik dan layak apabila terbebas dari kesalahan – kesalahan serta mencerminkan kondisi yang sesuai dengan kenyataan agar tidak menyebabkan suatu kesalahan pemahaman terhadap hasil keputusan yang diperoleh.

b. Informasi harus lengkap

Sebuah informasi harus disajikan secara lengkap dan utuh sebagai pengetahuan serta agar tidak mempengaruhi seseorang dalam menentukan keputusan.

c. Informasi wajib relevan

Dalam memberikan informasi harus memiliki manfaat yang sesuai bagi pembaca maupun pendengarnya, karena setiap orang memiliki tingkat relevansi yang berbeda.

d. Informasi harus tepat pada waktu

Informasi yang sangat baik dan terorganisir perlu menggunakan data yang tepat waktu. Ketepatan waktu berkaitan dapat memiliki ketersediaan data ketika harus digunakan dalam tujuan tertentu. Jadi informasi yang berkualitas memiliki asal dari sebuah kumpulan data yang bisa diolah dan menghasilkan dalam waktu singkat dan akurat.

Tujuan penerapan teknologi informasi untuk bisnis adalah untuk mencapai rantai nilai teknologi informasi yang berguna di semua aspek bisnis dan menuju peningkatan efisiensi dan efektivitas karyawan dan bisnis, dan produktivitas untuk mencapai tingkat pengembalian yang ditargetkan dengan risiko minimal, dan biaya (Indrayani, 2012:54). Dengan demikian, informasi memiliki peran penting dalam mengubah keadaan suatu masyarakat.

2.1.3 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 (Manalu dkk, 2017:18).

Prediksi tidak selalu memberikan jawaban pasti dan benar tentang apa yang akan terjadi, tetapi mereka mencoba menemukan jawaban yang paling sedekat mungkin dengan apa yang akan terjadi. Prakiraan merupakan apa yang akan terjadi dalam situasi yang memungkinkan dan menginformasikan perencanaan dan mengambil keputusan. Peramalan (*forecasting*) merupakan

ketentuan dalam menciptakan suatu informasi fakta bahwa situasi sosial masa akan datang berdasarkan informasi yang pernah ada berkaitan masalah kebijakan (Kafil, 2019:61). Prediksi datang dalam 3 bentuk utama, yaitu :

1. Proyeksi, merupakan perkiraan berdasarkan ekstrapolasi tren masa lalu dan sekarang untuk masa akan datang. Proyeksi menimbulkan beberapa pertanyaan afirmatif dalam argumen yang diambil dari metode dan kasus paralel.
2. Prediksi, merupakan ramalan yang berdasarkan dalam asumsi teoretis yang mempunyai keunggulan. Pendapat ini bisa berupa hukum teoritis (contohnya hukum yang mengurangi nilai uang), pernyataan teoritis (contohnya menyatakan bahwa disintegrasi masyarakat sipil adalah akibat dari kesenjangan antara keinginan dan kesanggupan), atau analogi (contohnya analogi antara perkembangan lembaga negara dan perkembangan bioorganisme).
3. Perkiraan (*conjecture*), merupakan ramalan berdasarkan suatu kondisi atau penilaian ahli dalam keadaan masa depan masyarakat.

Tujuan dari peramalan kebijakan merupakan untuk mendapatkan informasi tentang mengubah dimasa yang akan datang dapat timbul dari sesuatu implementasi pada suatu keputusan dan dampaknya. Oleh karena itu, sebelum membuat rekomendasi, perlu dibuat kebijakan untuk mendapatkan rekomendasi yang benar-benar akurat untuk penerapan di masa mendatang.

2.1.4 Usaha Mikro, Kecil, dan Menengah

Usaha Mikro, Kecil dan Menengah (UMKM) merupakan salah satu praktik bisnis yang paling populer di masyarakat. UMKM adalah kegiatan bisnis atau usaha yang dijalankan oleh perorangan, rumah tangga dan usaha

kecil. Perkembangan sektor usaha kecil dan menengah (UMKM) memiliki implikasi khusus bagi upaya pencapaian pertumbuhan ekonomi dan pengentasan kemiskinan di suatu negara (Wibowo dkk, 2015:60).

Klasifikasi UMKM biasanya dilakukan dengan pembatasan pendapatan tahunan, jumlah kekayaan atau aset, dan jumlah karyawan. Hal ini penting untuk digunakan untuk izin usaha di masa depan dan juga menentukan jumlah pajak yang akan dibayarkan kepada pemilik UMKM. Berikut masing-masing kriterianya :

- a Usaha mikro, adalah suatu usaha manufaktur yang dimiliki oleh orang perseorangan dan/atau satu badan yang memenuhi syarat usaha mikro yang dipersyaratkan oleh undang-undang.
- b Usaha kecil, adalah usaha ekonomi produktif yang berdiri sendiri, yang dilakukan oleh orang perseorangan atau badan usaha yang bukan merupakan anak perusahaan atau afiliasi dari usaha yang dimiliki, atau dikuasai secara langsung atau tidak langsung menjadi bagian dari usaha menengah atau usaha besar yang memenuhi kriteria usaha kecil dalam perusahaan dalam arti hukum
- c Usaha menengah, adalah usaha ekonomi dan produktif yang berdiri sendiri, yang dilakukan oleh orang perseorangan atau badan usaha yang bukan merupakan anak perusahaan atau cabang usaha yang memiliki, menguasai, atau merupakan bagian dari usaha kecil, baik langsung maupun tidak langsung. atau badan usaha besar dengan kekayaan bersih atau pendapatan tahunan, sebagaimana diwajibkan oleh undang-undang.

UMKM akan mampu bertahan dan bersaing jika dapat melakukan manajemen yang baik (Wibowo dkk, 2015:60). Meskipun tidak semua

UMKM memiliki pengelolaan keuangan, perizinan, dan aspek hukum lainnya yang profesional, namun keberadaan UMKM memiliki peran penting dalam perekonomian negara. Berikut peran UMKM yaitu: memajukan pemerataan ekonomi, memperluas kesempatan kerja, mendukung perekonomian dalam situasi kritis, meningkatkan mata uang negara, dan merespon kebutuhan masyarakat.

2.1.5 Pelanggan

Pelanggan adalah setiap individu, rumah tangga, atau bisnis yang membeli produk dari vendor atau pemasok tertentu, terlepas dari ide, barang, atau layanannya. Peneliti tampaknya setuju bahwa kepuasan pelanggan adalah faktor nomor satu yang dapat membangun loyalitas pelanggan (Laurent, 2016:95).

Dari beberapa definisi pelanggan diatas, terdapat beberapa jenis – jenis pelanggan yang dapat dipelajari, antara lain (Dewa dan Setyohadi, 2017:35) :

1. Pelanggan internal, merupakan individu atau sekelompok orang dalam suatu perusahaan. Secara umum, jenis pelanggan ini mempunyai dampak yang sangat besar terhadap kelancaran operasi perusahaan.
2. Pelanggan antara, merupakan orang perseorangan atau beberapa kelompok orang yang bertindak dalam penyalur barang yang barangnya dibeli tidak untuk dipakai sendiri melainkan untuk dijual kembali.
3. Pelanggan eksternal biasanya, adalah individu atau kelompok di luar perusahaan yang membeli produk perusahaan. Pelanggan eksternal sering mencari beberapa hal penting saat membeli produk dari perusahaan yang dimiliki, antara lain: harga, kesesuaian, kualitas produk dan layanan, metode pembayaran dan proses pengiriman.

2.1.6 *Customer Churn*

Customer churn merupakan istilah yang mengacu pada pelanggan yang tidak lagi menggunakan layanan atau jasa perusahaan (Prianto dkk, 2020:129). *Churn rate* dihitung berdasarkan jumlah pelanggan yang meninggalkan perusahaan selama periode waktu tertentu. 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.

Ketika perusahaan dapat mengurangi atau mencegah *churn* pelanggan, mereka dapat meningkatkan *customer lifetime value* (CLV). CLV adalah jumlah total yang dapat di harapkan dari rata-rata pelanggan untuk dibelanjakan pada bisnis selama perkembangan. Untuk mengurangi tingkat *churn* secara keseluruhan, strategi manajemen pelanggan perlu diterapkan sepenuhnya dari akuisisi pelanggan hingga pelanggan tidak lagi menggunakan layanan (Prianto dkk, 2020:131). Oleh karena itu, menemukan strategi untuk mengurangi dan mencegah *churn* pelanggan dapat menjadi cara yang bagus untuk meningkatkan keuntungan bisnis. Cara menghentikan *customer churn*, yaitu dengan cari tahu penyebab *churn*, meningkatkan *customer engagement*, memberikan *reward* pelanggan, dan memberikan keuntungan jangka panjang.

2.2 Teori Khusus

2.2.1 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 (Ridwan dkk, 2013:60). Data mining pula bisa dipercaya menjadi proses ekskavasi pengetahuan berdasarkan sejumlah akbar data. 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.

Sebagai urutan proses, data mining bisa dipisahkan menjadi sekitar langkah prosedur. Langkah - langkah ini bersifat saling melakukan aksi, pengguna secara melintas berpartisipasi atau melalui *knowledge base*. Data mining didalam teknik *Knowledge Discovery in Database* (KDD) dilakukan beserta langkah-langkah sebagai berikut (Ridwan dkk, 2013:60) :

1. Pembersihan data (*data cleaning*), merupakan teknik di mana data yang belum lengkap, rawan kesalahan, dan tidak konsisten dihapus dari kumpulan data. Ketahui lagi tentang tata laksana daur hidup data untuk mengetahui tentang perbuatan memproses data.
2. Integrasi data (*data integration*), merupakan teknik integrasi data di mana data repetitif disatukan.

3. Seleksi data (*data selection*), merupakan teknik memilih atau menyeleksi data yang relevan untuk analisis yang dihasilkan mulai pengumpulan data yang tersedia.
4. Transformasi data (*data transformation*), proses perubahan data telah dipilih sebagai prosedur eksplorasi melalui sarana dan entri data.
5. Proses *mining*, proses paling penting di mana teknik yang berbeda akan diterapkan untuk mengekstraksi banyak data potensial yang berbeda untuk data yang bermanfaat.
6. Evaluasi pola (*pattern evaluation*), proses di mana model - model yang atratif telah ditemukan dengan identifikasi terukur yang diukur untuk pengetahuan presentasi.
7. *Knowledge presentation*, merupakan langkah terakhir dalam proses. Teknik presentasi yang digunakan disini dimaksudkan untuk membantu pengguna memahami dan menginterpretasikan hasil data mining.

2.2.2 Bahasa Pemrograman *Java*

Java bisa disebut sebagai bahasa pemrograman populer yang dipakai dalam meluaskan *back-end* perangkat lunak, aplikasi android, & situs web. Bahasa pemrograman ini bersifat lintas platform, yaitu dapat digunakan di macam - macam platform, seperti desktop, Android dan bahkan untuk sistem operasi *Linux*. *Java* dapat berjalan diberbagai platform atau media sistem informasi dan berbagai *hardware* (Ali, 2019:8). Salah satu aplikasi yang menggunakan bahasa pemrograman *java* adalah *NetBeans IDE*.

Keunggulan utama *Java* adalah dapat berjalan di banyak platform komputer/sistem operasi dengan prinsip menulis sekali, berjalan di mana saja

(Komputer, 2013:3). Selain itu ada beberapa kelebihan lainnya dengan memerlukan bahasa pemrograman *Java* :

1. Membangun Aplikasi yang Fleksibel

Java memiliki compiler atau semacam mesin virtual yang dapat menerjemahkan sintaks *Java* menjadi *bytecode* untuk setiap platform. Jadi mengapa *Java* bisa berjalan di platform yang berbeda.

2. Mempunyai Library yang Cukup

Java memiliki *library* yang lengkap dan dalam bahasa pemrograman dapat di artikan *library* merupakan sekumpulan kegunaan dan aplikasi yang dapat langsung digunakan dalam pembuatan aplikasi.

3. Berorientas pada Objek

Pemrograman berorientasi objek adalah program yang berhubungan dengan desain aplikasi berbasis objek. Objek di sini dapat dipahami sebagai bidang data dengan perilaku dan properti.

4. Mirip dengan Bahasa C++

Bahasa pemrograman C++ sebenarnya tergolong salah satu bahasa pemrograman yang sangat baik, tetapi tidak sefleksibel *Java*. Sementara, *Java* dapat dibuat dengan lebih fleksibel dan sederhana.

5. Membuat *Coding* Lebih Sederhana

Dibanding dengan bahasa pemrograman yang lain, *Java* memiliki desain *coding* yang lebih sederhana dan ringkas.

6. Mendukung *Native Method*

Manfaat nyata *Java* untuk pengembangan aplikasi adalah mendukung *native method*. Ini adalah fungsi dari program *Java* yang dapat ditulis dalam bahasa selain *Java*, seperti C atau C++.

2.2.3 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. Algoritma ini menggunakan teorema *Bayes* dan membuat perhitungan kira-kira yaitu diberikan nilai variabel kelas, semua atribut adalah independen. Teorema keputusan *Bayes* merupakan metode statistik dasar untuk pengenalan pola (Ridwan dkk, 2013:61). *Naive Bayes* mendasari kepada asumsi sederhana bahwa nilai atribut yang berdiri sendiri bersyarat ketika nilai *output* di sediakan. Dengan makna berbeda, probabilitas pengamatan umum merupakan produk dalam probabilitas pribadi. Kegunaan dalam memakai *Naive Bayes* adalah metode ini sekedar memerlukan sedikit data latih untuk menetapkan jangkauan yang diperkirakan parameter yang membutuhkan untuk memproses klasifikasi. *Naive Bayes* umumnya berbuat amat sangat bagus di sebagian besar situasi kehidupan yang lebih menyeluruh dari pada yang diharapkan. Langkah-langkah dari tahapan algoritma *Naive Bayes* adalah, sebagai berikut:

1. Menghitung jumlah kelas atau label.
2. Menghitung jumlah kasus setiap kelas.
3. Kalikan semua variabel kelas.
4. Bandingkan hasil setiap kelas.

Secara garis besar, algoritma *Naive Bayes* dapat digambarkan sebagai persamaan ini (Kurniawan, 2018:457):

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

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

2.3 Teori Analisa dan Perancangan

2.3.1 *Cross-Industry Standard Process for Data Mining (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 (Fadillah, 2015:261). Model proses CRISP-DM memberikan wawasan tentang siklus hidup proyek penambangan data. CRISP-DM mempunyai 6 langkah, yaitu :

1. *Business Understanding*

Ini adalah fase pertama CRISP-DM dan termasuk bagian yang cukup penting. 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. Kegiatan yang dilakukan meliputi: mendefinisikan secara jelas keseluruhan tujuan dan persyaratan, menerjemahkan tujuan tersebut dan mendefinisikan batasan untuk perbuatan merumuskan kasus data

mining, dan kemudian menyiapkan rencana permulaan dalam pencapaian tujuan maksud tersebut.

2. *Data Understanding*

Umumnya dalam fase ini meliputi hasil memeriksa data, pengenalan masalah dalam data. Langkah ini membutuhkan dasar analitis dalam studi dengan meringkas dan mengidentifikasi kemampuan yang mempunyai kesanggupan kasus dalam data. Langkah ini juga wajib melakukan dengan teliti dan tanpa tergesa-gesa, karena dalam visualisasi data terkadang sulit dipahami jika dikaitkan dengan ringkasan data. Apabila terdapat perkara dalam termin ini yg belum teratasi, maka akan mengganggu termin pemodelan. Ringkasan atau *summary* data bisa bermanfaat dalam memastikan bahwa data yang didistribusikan sesuai dengan yang diharapkan atau untuk mengungkapkan bias yang tidak diharapkan yang harus ditangani dalam tahap berikutnya, yaitu *Data Preperation*.

3. *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. Ini tidak mengandung maksud bahwa ketika data awal bekerja sebelum masalah dalam data dipecahkan, data tersebut dapat digunakan sampai tahap akhir. Langkah ini sering diperiksa ketika masalah terdeteksi selama pengembangan model. Jadi iterasi ini dilakukan sampai menemukan sesuatu yang cocok dengan data tersebut. Langkah

pengambilan sampel dapat dilakukan di sini dan data biasanya dibagi menjadi dua, data latih dan data uji. Kegiatan yang dilakukan meliputi: pemilihan masalah dan parameter yang hendak dianalisa (*Select data*), mentransformasikan parameter sudah pasti (*Transformation*) dan membersihkan data sehingga data sudah disediakan dalam tahap pemodelan (*Cleaning*).

4. *Modeling*

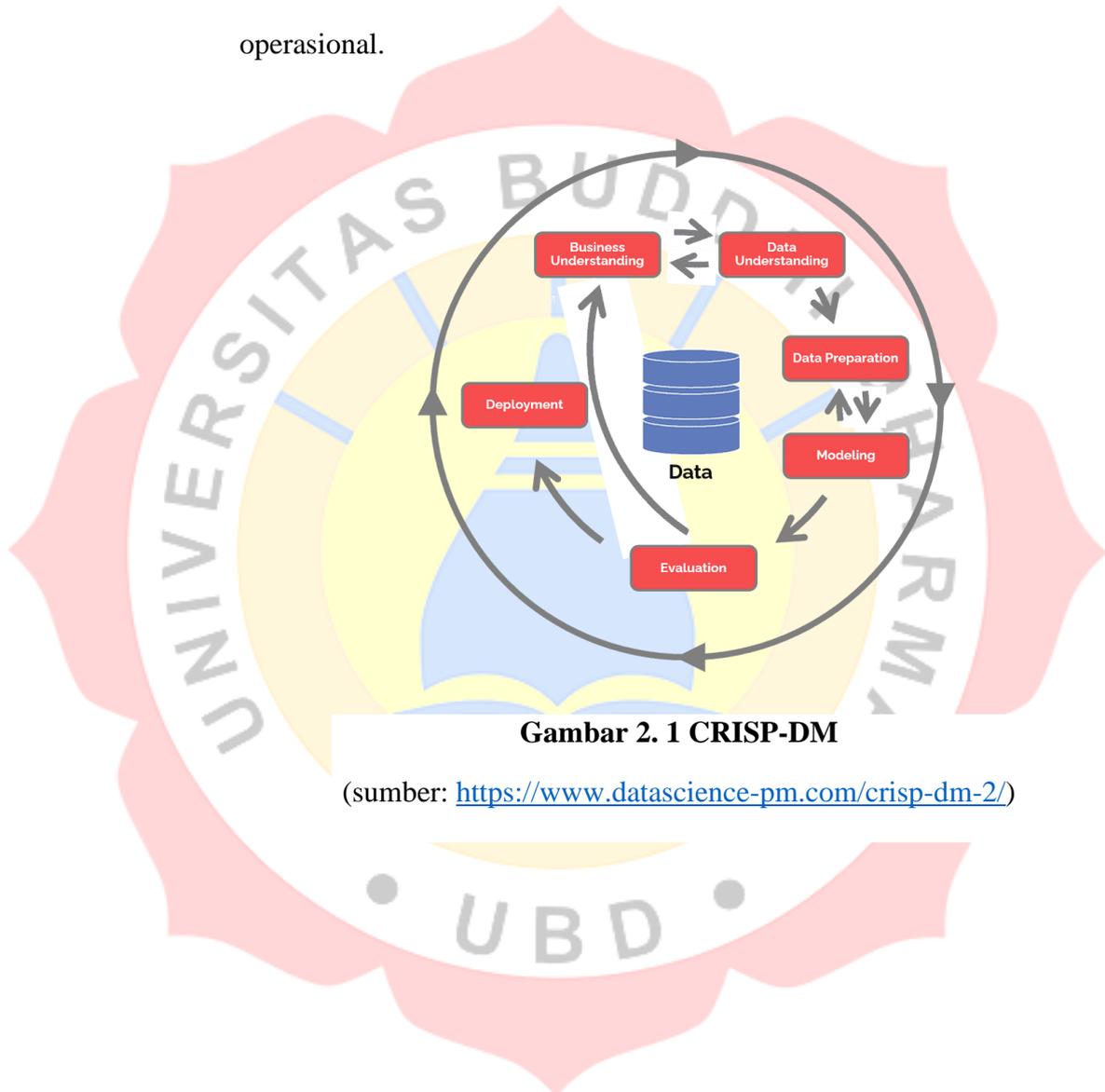
Pada dasarnya, fase ini untuk melakukan model prediktif atau deskriptif. Fase ini menggunakan metode statistik dan data mining dalam menetapkan teknik data mining yang berlaku, perangkat data mining, dan algoritma data mining. Langkah setelahnya merupakan memakai perangkat dalam mengimplementasikan teknik dan algoritme data mining ke data. Jika ada kebutuhan untuk mengadaptasi data untuk proses data mining khusus, maka bisa berulang lagi ke langkah *data preparation*. Beberapa model yang umum digunakan adalah klasifikasi, scoring, ranking, clustering, menemukan hubungan, dan karakterisasi.

5. *Evaluation*

Langkah selanjutnya interpretasikan hasil data mining yang diperoleh selama tahap pemodelan dalam langkah sebelumnya. 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.

6. *Deployment*

Tahap implementasi atau perencanaan merupakan tahapan terpenting dari proses CRISP-DM. Perencanaan penerapan dimulai dalam proses pembelajaran bisnis dan harus memasukkan tidak hanya bagaimana menghasilkan poin model, tetapi juga bagaimana mentransfer poin keputusan dan mengintegrasikan keputusan ke dalam sistem operasional.



Gambar 2. 1 CRISP-DM

(sumber: <https://www.datascience-pm.com/crisp-dm-2/>)

2.3.2 Aplikasi *Rapid Miner*

Rapid Miner adalah software dalam mengerjakan data mining. Memakai perangkat lunak ini maka bisa melakukan penambangan teks, termasuk analisis teks, memperoleh pola dan menggabungkannya dengan metode statistik, basis data, dan *Artificial Intelligence* (AI). *Rapid Miner* dirancang dalam tim analitik, yang menggabungkan semua siklus hidup pengetahuan data, mulai awal persediaan data sampai *machine learning* sampai perkiraan hingga implementasi model (Prasetyo et al, 2021:10).

Selain itu, *Rapid Miner* juga digunakan untuk pembelajaran mendalam, pembelajaran mesin, dan analitik prediktif. Alat ini dapat diterapkan untuk bisnis, perdagangan, pelatihan, pendidikan, penelitian, dll. Proses data mining yang dikerjakan bagi *Rapid Miner* merupakan ETL (*Extract, Transform, Load*), data *pre-processing*, *visualization*, *modeling* dan *evaluation*. Proses tersebut dijelaskan dalam XML, dibikin menggunakan GUI (*Graphical User Interface*) dan memakai bahasa pemrograman *Java*.

2.3.3 Aplikasi *Netbeans IDE*

Netbeans merupakan sebuah aplikasi atau lingkungan pengembangan, dan alat bagi pemrogram untuk menulis, mengkompilasi, memecahkan masalah, dan menyebarkan program (Afrizal, 2014:4). *Netbeans* adalah aplikasi *Java* (IDE) oleh *Sun Microsystems* yang melakukan proses di *Swing*. *Swing* adalah ilmu pengetahuan terapan *Java* dalam mengembangkan program desktop yang bisa melakukan di macam-macam program seperti *Windows*, *Linux*, *Mac OS X*, dan *Solaris*. IDE adalah domain pemrograman yang tertanam untuk program *software* yang mempersiapkan antarmuka pengguna grafis (GUI), yaitu kode atau editor teks, kompiler, dan debugger.

NetBeans mempunyai pengembangan tujuan umum dengan fokus khusus pada *Java*. Ini juga memiliki editor dan fitur berkualitas tinggi untuk bahasa populer lainnya seperti *JavaScript*, *PHP*, dan *C++*. *NetBeans* menuang dalam 2 hal, sebagai platform dalam pembuatan mengembangkan desktop *Java* dan lingkungan pengembangan terintegrasi yang dibangun menggunakan platform *NetBeans*. Fitur *NetBeans* dapat dipakai pada programmer dalam menulis, mengkompilasi, memecahkan masalah, dan mendistribusikan aplikasi *NetBeans* yang dibuat untuk bahasa pemrograman *Java*, tetapi lain dari pada *NetBeans* bahasa pemrograman lain.

2.4 Tinjauan Studi

2.4.1 Penelitian

Jurnal 1

No.	Data Jurnal/Makalah	Keterangan
1	Judul	Prediksi <i>Customer Churn</i> Dengan Algoritma <i>Decison Tree C4.5</i> Berdasarkan Segmentasi Pelanggan Pada Perusahaan Retail
2	Jurnal	Jurnal Rekayasa Sistem Komputer (RESISTOR)
3	Volume dan halaman	Vol. 1, No. 1
4	Tanggal & Tahun	April, 2018
5	Penulis	Ni Wayan Wardani, Gede Rasben Dantes, Gede Indrawan
6	Penerbit	Universitas Pendidikan Ganesha
7	Tujuan penelitian	Untuk melakukan prediksi churn pelanggan untuk tiap kelas customer dan menyaksikan tinggi rendahnya akurasi untuk mengenal penerapan model RFM untuk menciptakan kelas customer menempuh tahap pembagian dalam segmen customer.
8	Lokasi dan Sumber Penelitian	UD. Mawar Sari
9	Perancangan Sistem	Rancangan sistem : a RFM (<i>Recency, Frequency, Monetary</i>) b <i>Confusion Matrix</i>

10	Hasil Penelitian	<p>Hasil dari penelitian dengan Algoritma <i>Decision Tree C4.5</i> adalah :</p> <ul style="list-style-type: none"> a Hasil kinerja pada kelas pelanggan <i>dormant</i> yaitu recall 97.51%, precision 75.18%, dan accuracy 76.18%. b Hasil kinerja pada kelas pelanggan <i>everyday</i> yaitu recall 100%, precision 99.04%, dan accuracy 99.04%. c Hasil kinerja pada kelas pelanggan <i>golden</i> yaitu <i>recall</i> 100%, precision 98.84%, dan accuracy 98.84%. d Hasil kinerja pada kelas pelanggan <i>superstar</i> yaitu recall 96.15%, precision 99.43%, dan accuracy 95.63%.
11	Kekuatan Penelitian	<p>Kekuatan dari penelitian ini adalah :</p> <ul style="list-style-type: none"> a Dalam mempraktikkan dasar teknik CRM (Customer Relationship Management), bisnis bisa menerapkan segmentasi dan mengidentifikasi customer, memprediksi potensi churn berdasarkan kelas pelanggan dan mengimplementasikan program loyalitas pelanggan..
12	Kelemahan Penelitian	<p>Kelemahan dari penelitian ini adalah :</p> <ul style="list-style-type: none"> a Teknik RFM yang digunakan fokus hanya pada pelanggan terbaik, memiliki variabel terbatas, fokus hanya pada pelanggan yang sudah ada, dan tidak berlaku untuk pelanggan baru atau potensial. b Perhitungan manual tanpa menggunakan <i>software rapid miner</i> sehingga data yang tidak efektif
13	Kesimpulan	<p>Berdasarkan evaluasi matriks konfusi, kita dapat menyimpulkan bahwa kelas pelanggan yang tidak aktif adalah kelas pelanggan yang cenderung churn.</p>

Jurnal 2

No.	Data Jurnal/Makalah	Keterangan
1	Judul	Prediksi <i>Customer Churn</i> Perusahaan Telekomunikasi Menggunakan <i>Naive Bayes</i> Dan <i>K-Nearest Neighbor</i>
2	Jurnal	Jurnal Informasi Interaktif

3	Volume dan halaman	Vol. 4 No. 3
4	Tanggal & Tahun	September 2019
5	Penulis	Kaharudin, Musthofa Galih Pradana, Kusri
6	Penerbit	Program Studi Teknik Informatika Fakultas Teknik Universitas Janabadra
7	Tujuan penelitian	Untuk membandingkan dua metode yakni metode <i>Naïve Bayes</i> dan metode <i>K-Nearest Neighbor</i> agar mencapai hasil optimal saat membuat perkiraan yang membantu organisasi menetapkan kebijakan dan keputusan.
8	Lokasi dan Sumber Penelitian	Perusahaan Telekomunikasi
9	Perancangan Sistem	Rancangan sistem : a <i>Data Flow Diagram</i> (DFD)
10	Hasil Penelitian	Hasil dari penelitian adalah : a Metode <i>Naïve Bayes</i> menghasilkan akurasi sebesar 76%. b Untuk melakukan prediksi menggunakan algoritma <i>K-Nearest Neighbor</i> untuk menghasilkan akurasi terbaik, $K = 5$, memiliki tingkat akurasi 78%.
11	Kekuatan Penelitian	Kekuatan dari penelitian ini adalah : a Diagram aliran data atau DFD membantu memahami kemampuan dan keterbatasan sistem, memvisualisasikan konten, dan menjelaskan komponen sistem secara lebih rinci. b Dengan melakukan perbandingan 2 metode maka bisa melihat mana yang lebih baik dan akurat.
12	Kelemahan Penelitian	Kelemahan dari penelitian ini adalah : a Perancangan sistem DFD membutuhkan waktu lama untuk dibuat b Tidak dapat memberi informasi yang tidak diketahui dari data <i>training</i> seperti pembelajaran tanpa pengawasan
13	Kesimpulan	Dari hasil analisis perbandingan kedua metode tersebut dapat disimpulkan bahwa metode yang lebih akurat adalah metode <i>K-Nearest Neighbor</i> dengan nilai $K=5$.

Jurnal 3

No.	Data Jurnal/Makalah	Keterangan
1	Judul	Prediksi <i>Customer Churn</i> Pada Bisnis Retail Menggunakan Algoritma <i>Naïve Bayes</i>

2	Jurnal	Riset dan E-Jurnal Manajemen Informatika Komputer
3	Volume dan halaman	Volume 6, Nomor 1
4	Tanggal & Tahun	Oktober 2021
5	Penulis	Firmansyah, Agus Yulianto
6	Penerbit	Universitas Bina Sarana Informatika
7	Tujuan penelitian	Untuk menciptakan contoh prediktif memakai Naive Bayes buat memprediksi customer churn sebagai akibatnya perusahaan ritel bisa memprediksi kerugian timbal lebih awal.
8	Lokasi dan Sumber Penelitian	Data keanggotaan toko retail di 10 toko yang diambil secara acak (random sampling).
9	Perancangan Sistem	Rancangan sistem : a CRISP-DM (<i>Cross-Industry Standard Process for Data Mining</i>)
10	Hasil Penelitian	Hasil dari penelitian adalah : a Dari pemodelan <i>naïve bayes</i> untuk memprediksi <i>customer churn</i> memperoleh hasil akurasi mencapai 80% dan presisi sebesar 100%.
11	Kekuatan Penelitian	Kekuatan dari penelitian ini adalah : a Dengan menggunakan metode CRISP-DM dapat dikatakan untuk versi lebih cepat, ringkas, dan tepat dalam mengerjakan beberapa penelitian maupun menyelesaikan suatu kasus. b Mudah diimplementasi dan memberikan hasil yang presisi yang tepat.
12	Kelemahan Penelitian	Kelemahan dari penelitian ini adalah : a Jika ada penambahan data training maka akan mempengaruhi tingkat akurasi dari metode <i>naïve bayes</i> . b Dalam proses perencanaan awal yang termasuk dalam CRISP-DM, ada sedikit penekanan atau kurangnya analisis pada proses bisnis.
13	Kesimpulan	Awal model <i>Naive Bayes</i> dalam memprediksi <i>customer churn</i> , hasilnya adalah akurasi 80% dengan presisi 100% sehingga model bisa menerapkan pada prakiraan dalam asal data yang diklasifikasi. Perbuatan meningkatkan banyaknya data latihan wajibnya hendak mensugesti nilai ketepatan contoh Naive Bayes, sebagai akibatnya perlu diimplementasikan contoh adonan memakai prosedur pemecahan lain pada mempertinggi akurasi menjadi Bayesian network atau C5.0. Model <i>Naive Bayes</i> sama saja sesuai dalam data klasifikasi, dan

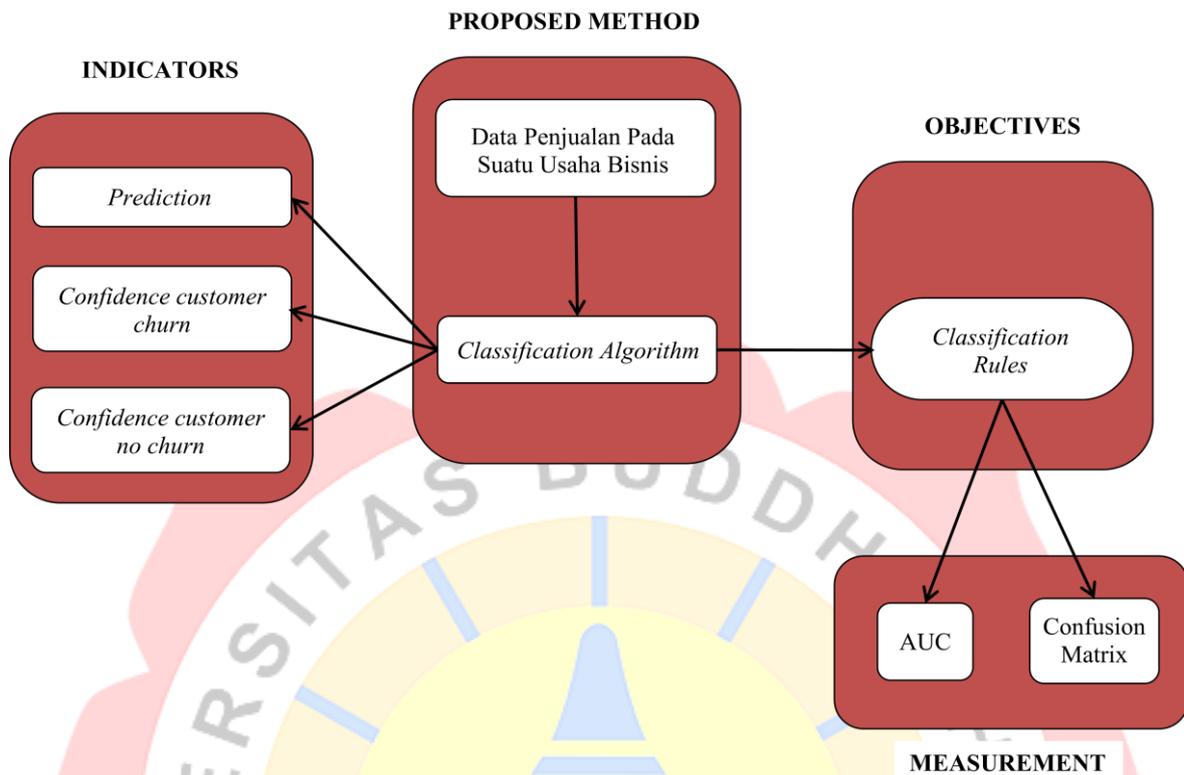
		dalam data numerik algoritma <i>Bayesian network</i> dapat digunakan.
--	--	---

2.4.1 Rangkuman Model Penelitian

Peneliti	Nama Jurnal	Tahun	Institusi	Judul dan metode yang digunakan	Kesimpulan
Ni Wayan Wardani, Gede Rasben Dantes, Gede Indrawan	Jurnal Rekayasa Sistem Komputer (RESISTOR) Vol. 1 No. 1 p-ISSN: 2598-7542 e-ISSN: 2598-9650	2018	Universitas Pendidikan Ganesha	Prediksi <i>Customer Churn</i> Dengan Algoritma <i>Decision Tree</i> C4.5 Berdasarkan Segmentasi Pelanggan Pada Perusahaan Retail	Berdasarkan dalam penilaian menggunakan <i>confusion matrix</i> bisa disimpulkan bahwa kelas pelanggan dormant merupakan kelas pelanggan yang berpotensi <i>churn</i> .
Kaharudin, Musthofa Galih Pradana, Kusri	Jurnal Informasi Interaktif Vol. 4 No. 3 ISSN: 2527-5240	2019	Program Studi Teknik Informatika Fakultas Teknik Universitas Janabadra	Prediksi <i>Customer Churn</i> Perusahaan Telekomunikasi Menggunakan <i>Naive Bayes</i> Dan <i>K-Nearest Neighbor</i>	Berdasarkan pada hasil analisa perbandingan kedua metode dapat disimpulkan bahwa metode yang menghasilkan akurasi yang lebih akurat merupakan metode <i>K-Nearest Neighbor</i> dengan nilai K=5.
Firmansyah, Agus Yulianto	Riset dan E-Jurnal Manajemen Informatika Komputer Vol 6 No 1 e-ISSN: 2541-1330 p-ISSN: 2541-1332	2021	Universitas Bina Sarana Informatika	Prediksi <i>Customer Churn</i> Pada Bisnis Retail Menggunakan Algoritma <i>Naive Bayes</i>	Dari model <i>Naive Bayes</i> untuk memprediksi <i>customer churn</i> , hasilnya adalah akurasi 80% dengan presisi 100% sehingga model dapat diterapkan pada prediksi dengan sumber data yang diklasifikasi.

					<p>Peningkatan jumlah data latih tentunya akan mempengaruhi nilai akurasi model <i>Naive Bayes</i>, sehingga perlu diimplementasikan model gabungan menggunakan algoritma lain untuk meningkatkan akurasi seperti <i>Bayesian network</i> atau C5.0. Model <i>Naive Bayes</i> juga hanya cocok untuk data klasifikasi, dan untuk data numerik algoritma <i>Bayesian network</i> dapat digunakan.</p>
--	--	--	--	--	--

2.5 Kerangka Pemikiran



Gambar 2. 2 Kerangka Pemikiran

Rancangan pengertian di atas bisa menerangkan menjadi berikut:

Kerangka bentuk data mining pada untuk menurut yang dibutuhkan data buat sistem *customer churn* suatu usaha. Data mining ini dilaksanakan menggunakan sistem melakukan analisis data berhubungan dengan operasi dalam data penjualan data customer churn memakai teknik data mining *naive bayes*. 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*

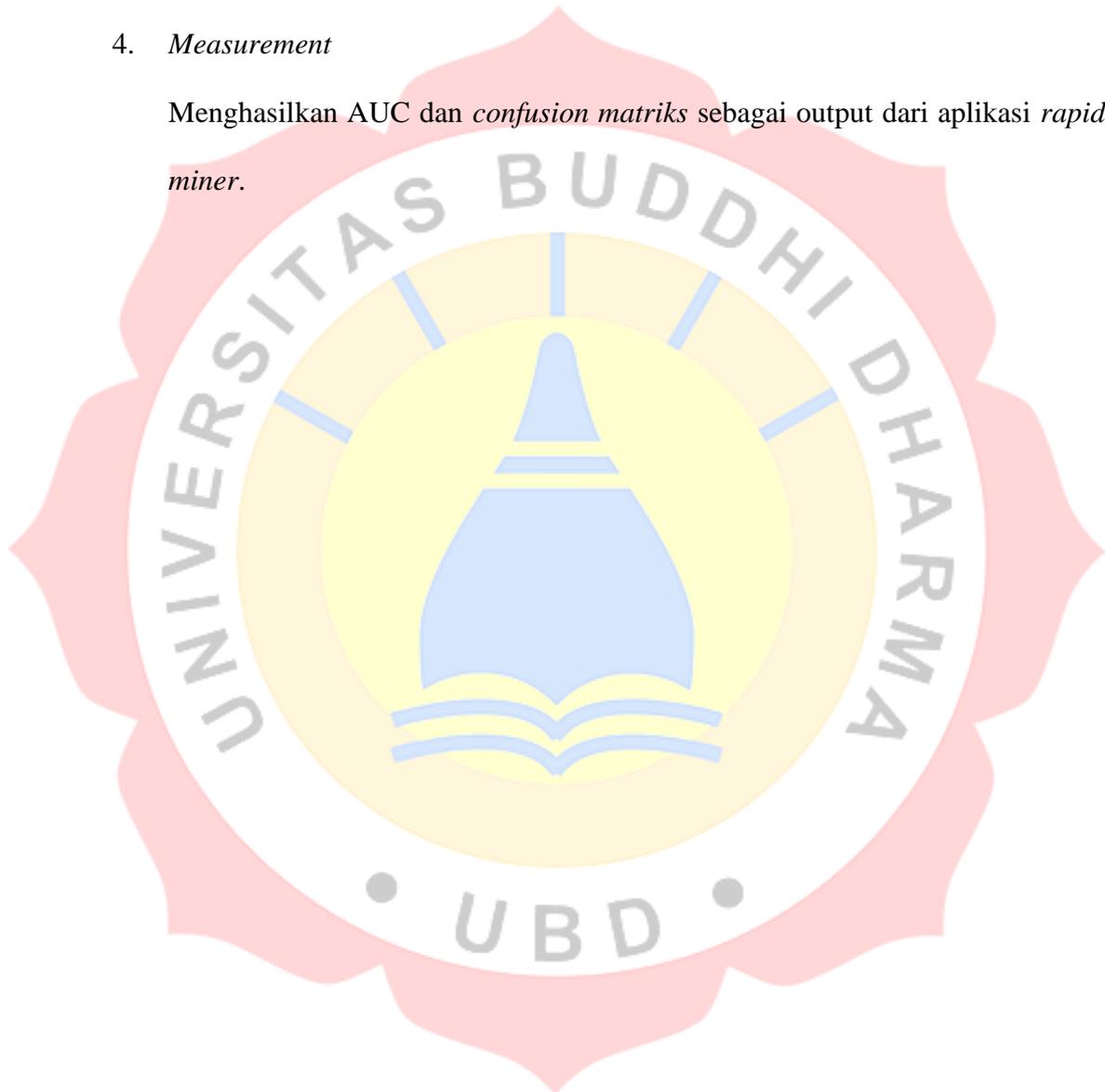
Data penjualan pada suatu usaha bisnis di olah menggunakan aplikasi *rapid miner* dengan algoritma *Naive Bayes* untuk menemukan *classification rules*.

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*.



BAB III

ANALISA MASALAH & PERANCANGAN APLIKASI

3.1 Analisa Kebutuhan

Berikut adalah analisa kebutuhan yang di lakukan oleh penulis dalam penelitian ini :

3.1.1 Data Set Customer Churn

Data set *customer churn* memiliki 3333 record data, 10 atribut, dan satu atribut churn, di mana data set ini didapatkan penulis di *website Kaggle* yang bisa mengakses dari link di bawah ini :

<https://www.kaggle.com/datasets/barun2104/telecom-churn>

11 atribut yang dipakai untuk penelitian Skripsi ini yaitu sebagai berikut :

1. *Churn*
2. *Account Weeks*
3. *Contract Renewal*
4. *Data Plan*
5. *Data Usage*
6. *Cust Serv Calls*
7. *Day Mins*
8. *Day Calls*
9. *Monthly Charge*
10. *Overage Fee*
11. *Roam Mins*

Keterangan data :

- *Churn* yang di artikan kehilangan pelanggan. Dibagi menjadi 2 yaitu : angka 1 jika pelanggan membatalkan atau berhenti layanan, dan angka 0 jika pelanggan yang masih membutuhkan atau memakai layanan.
- *Account Weeks* merupakan jumlah minggu pelanggan memiliki akun aktif.
- *Contract Renewal* adalah perpanjangan kontrak yang dibagi menjadi 2, yaitu : angka 1 jika pelanggan baru saja memperbarui kontrak, dan angka 0 jika tidak memperbarui kontrak.
- *Data Plan* yang dapat di artikan paket data di bagi menjadi 2 bagian, yaitu : angka 1 jika pelanggan memiliki paket data, dan angka 0 jika tidak memiliki paket data.
- *Data Usage* merupakan penggunaan data dengan gigabyte di setiap bulanan nya.
- *Customer Service Calls* adalah jumlah panggilan ke layanan pelanggan.
- *Day Mins* adalah rata-rata dalam sehari per bulan untuk penggunaan data nya.
- *Day Calls* merupakan jumlah rata-rata panggilan dalam sehari.
- *Monthly Charge* yaitu rata-rata tagihan bulanan.
- *Overage Fee* yaitu biaya kelebihan penggunaan terbesar dalam 12 bulan terakhir.
- *Roam Mins* merupakan jumlah rata-rata menit roaming.

Churn	AccountWeeks	ContractRenewal	DataPlan	RoamMins
0	128	1	1	10

0	107	1	1	13.7
0	137	1	0	12.2
0	84	0	0	6.6
0	75	0	0	10.1
0	118	0	0	6.3
0	121	1	1	7.5
0	147	0	0	7.1
0	117	1	0	8.7
0	141	0	1	11.2
1	65	1	0	12.7
0	74	1	0	9.1
0	168	1	0	11.2
0	95	1	0	12.3
0	62	1	0	13.1
1	161	1	0	5.4
....	13.8
0	93	1	0	0	8.1

Tabel 3. 1 Data Set Customer Churn

3.1.2 Analisa Perangkat Lunak Untuk Membangun Sistem

Perangkat lunak yang hendak digunakan untuk bangkit berdiri sistem dari penelitian ini yaitu :

1. *Processor i5-1135G7*
2. *Sistem operasi Windows 11 Home*
3. *Memory 8 Gb DDR 4*
4. *Microsoft Office 2019*
5. *Rapid Miner*
6. *Netbeans*

3.1.3 Analisa Kebutuhan Pemakai

Berdasarkan pengamatan pada hal-hal yang dibutuhkan oleh pemilik usaha, maka diperlukannya pembuatan beberapa fitur yang dapat

memudahkan pemilik usaha untuk membaca dan menerima informasi prediksi pelanggan dengan mudah, berikut kebutuhan yang di butuhkan untuk aplikasi, yaitu :

1. Menampilkan hasil prediksi
2. Menampilkan informasi yang jelas
3. Aplikasi tidak berat
4. Tampilan menarik dan mudah digunakan

3.1.4 Analisa Kebutuhan Aplikasi

Bersumber pada tabel analisis yang dibutuhkan pemakai atau user, di buatlah rancangan aplikasi untuk melengkapi kebutuhan user, berikut merupakan tabel kebutuhan aplikasi :

Analisa Kebutuhan		Keterangan
Saya ingin sistem ini dapat :		
1	Menampilkan sistem user yang <i>friendly</i> (mudah dipahami user)	V
2	Tampilan aplikasi simple dan menarik	V
3	Dapat digunakan secara online	X
4	Menampilkan hasil prediksi	V
5	Memiliki keamanan data yang baik	V
6	Dapat menyarankan pelanggan agar tetap menggunakan layanan atau membeli produk nya	V
7	Mengetahui alasan mengapa <i>customer churn</i>	X
8	Menampilkan informasi yang jelas	V
9	Tampilan aplikasi mudah digunakan dan dipahami user	V

Gambar 3. 1 Analisa Kebutuhan Aplikasi

3.2 Latar Belakang Penggunaan Metode Algoritma Naïve Bayes

Pada tahap ini algoritma yang dipakai yaitu *Naïve Bayes* digambarkan ke dalam bentuk model yang akan diolah menggunakan aplikasi *RapidMiner*. Berikut ini merupakan pseudocode model dalam melakukan pengolahan :

```
function naivebayesclassification()

    load hasilmean;

    load hasilStd;

    load DataPerkelas;

    for k = 1:2

        perhitungan1=sqrt(2*pi);

        perhitungan2 (:,:,k)=1./ (perhitungan1 stdD(:,:, k));

        perhitungan3 (:,:,k)=-((dataCustomerChurn (:,:,) hasilMean (:,:,

        k)).^2);

        perhitungan4 (:,:,k)= 2 stdD (:,:,k).^2;

        perhitungan5 (:,:,k) perhitungan3 (:,:,k)/perhitungan4 (:,:,k);

        perhitungan6 (:,:,k)= exp (1). ^perhitungan5 (:,:, k);

        perhitungan7 (:,:, k)=perhitungan2 (:,:, k). *perhitungan6 (:,:, k);

        hasil1 (:,:,k) = perhitungan7 (1, 1, k) perhitungan7 (1,2, k)*

        perhitungan7 (1,3, k)* perhitungan7 (1,4,k) *perhitungan7 (1,5, k)

        perkalianchurn (:,:,k)* perkalianno churn (:,:,k);

    end

    [argmax (:, :), keles]-max (hasil1);

    hasilee = keles;

end
```

3.2.1 *Bussiness Understanding*

Fase proses bisnis berpusat dari perbuatan memahami tujuan yang dibutuhkan menurut evaluasi bisnis. Setelahnya, pemahaman ini diterjemahkan ke dalam rencana penambangan data pertama dalam menyampaikan tujuan. Tujuan yang dimaksudkan ini didasarkan dari kemampuan analitis dari algoritma *Naïve Bayes* (Klasifikasi). Tujuannya yaitu bagian melakukan klasifikasi data dalam kelas sudah pasti, setelah itu pola sudah disebutkan bisa menggunakan dalam membuat perkiraan pelanggan yang dapat terindikasi *churn*, sehingga pemilik usaha dapat mengambil tindakan promosi atau strategi marketing jika pelanggan terindikasi *churn* atau *no churn* tersebut.

Dataset ini akan digunakan untuk mendapatkan informasi yang diperlukan dalam melakukan penelitian ini. Informasi yang diperoleh diinginkan dapat memberikan informasi bagi proses memahami bisnis. Tujuan dari data mining untuk penelitian ini adalah untuk menemukan akurasi paling baik dalam memperkirakan sejauh mana *churn* pelanggan.

Kita akan menggunakan data dari *data usage*, *account weeks*, *contract renewal*, *data plan* dan lain nya yang ada di data set yang ada indikasi nya dari penyakit jantung dengan total 3333 entries row data secara keseluruhan. Sehingga kita dapat mengklasifikasi berdasarkan *data usage*, *account weeks*, *contract renewal*, *data plan* dan lainnya agar accuracy nya tinggi dalam memprediksi *customer churn* tersebut.

3.2.2 *Data Understanding*

Dari data set yang peneliti miliki terdapat beberapa attribut sebagai berikut :

- *Account Weeks* (Mingguan Akun) : Dalam bahasa Inggris istilah *Account Weeks* berarti *account* = akun (nasabah/klien), sementara *weeks* =

mingguan. Karena keberadaan klien itu sangat penting bagi perusahaan telekomunikasi atau bisnis umkm lainnya, maka jumlah minggu pelanggan telah memiliki akun aktif dapat menjadi salah satu atribut dalam *churn* di perusahaan telekomunikasi.

- *Contract Renewal* (Perpanjang Kontrak): Kontrak kerja ini biasanya memuat aturan kerja, termasuk hak dan kewajiban antara dua pihak, dalam hal ini pemberi kerja (perusahaan/pabrik) dan pelanggan itu sendiri. Perpanjangan kontrak adalah salah satu hal utama yang wajib dipahami dan juga dipelajari bagi masing-masing orang. Memang surat ini merupakan bukti kesepakatan antara pelanggan dan perusahaan yang mengatur hak dan kewajiban setiap orang. Maka atribut perpanjangan kontrak ini menjadi salah satu dalam penyebab *customer churn*.
- *Data Plan* (Paket Data): Istilah data di dunia seluler melekat pada jumlah yang dibebankan di setiap paket data di smartphone. Di perusahaan telekomunikasi menawarkan paket berbeda untuk mengimbangi jumlah data yang dapat dinikmati pengguna ponsel cerdas. Maka dari sini peneliti dapat melihat pelanggan yang memiliki paket data atau tidak.
- *Data Usage* (Penggunaan Data): Penggunaan data sebagai sebagai bisnis di industri telekomunikasi dapat melihat gigabyte penggunaan data bulanan yang di gunakan oleh pelanggan/klien. Jadi penggunaan data disini juga menjadi faktor yang mempengaruhi *customer churn*.
- *Customer Service Calls* (Panggilan Layanan): Peran *customer service* sangat penting dalam menyelesaikan keluhan pelanggan demi tercapainya kepuasan pelanggan dan membangun citra perusahaan yang baik. Maka jika

suatu perusahaan tidak dapat melihat jumlah keluhan pelanggan maka pelanggan akan menghilang atau tidak berlangganan kembali.

- *Day Mins*: Merupakan rata-rata menit penggunaan pelanggan sehari - hari per bulan. Besaran dari suatu trafik telekomunikasi diukur dengan satuan waktu.
- *Day Calls* (Panggilan Sehari-hari): Atribut ini untuk melihat rata-rata jumlah panggilan pelanggan dalam per hari. Rata-rata penggunaan berkelanjutan tergantung pada jumlah panggilan masuk dan durasi panggilan.
- *Monthly Charge* (Biaya Bulanan): Merupakan rata-rata tagihan bulanan pelanggan atau klien. Tagihan biaya adalah kewajiban Pelanggan untuk membayar penggunaan layanan atau sarana untuk melancarkan pelaksanaan fungsi sudah tentu, sudah masuk pengeluaran administrasi, denda, bunga, dan jika ada biaya lainnya.
- *Overage Fee* (Biaya Kelebihan): Yaitu biaya kelebihan penggunaan terbesar dalam 12 bulan terakhir. Biaya kelebihan adalah perbuatan mengorbankan atau biaya yang dilakukan bagi klien dengan tujuan mendapatkan guna lewat dari semestinya dalam kegiatan yang dilakukan.
- *Roam Mins*: Dapat melihat jumlah rata-rata menit roaming pada pelanggan. *Roaming* itu sendiri, merupakan proses pemeliharaan koneksi ke Internet atau jaringan seluler menggunakan layanan telekomunikasi yang disediakan oleh penyedia telekomunikasi lain.

3.2.3 Data Preparation

Data Preparation adalah proses atau langkah-langkah yang dilakukan peneliti dalam mengubah data mentah berubah keadaan jadi data mempunyai kualitas tinggi (input yang sesuai untuk alat data mining) di rapid miner. Berikut sumber data peneliti :

<https://www.kaggle.com/datasets/barun2104/telecom-churn>

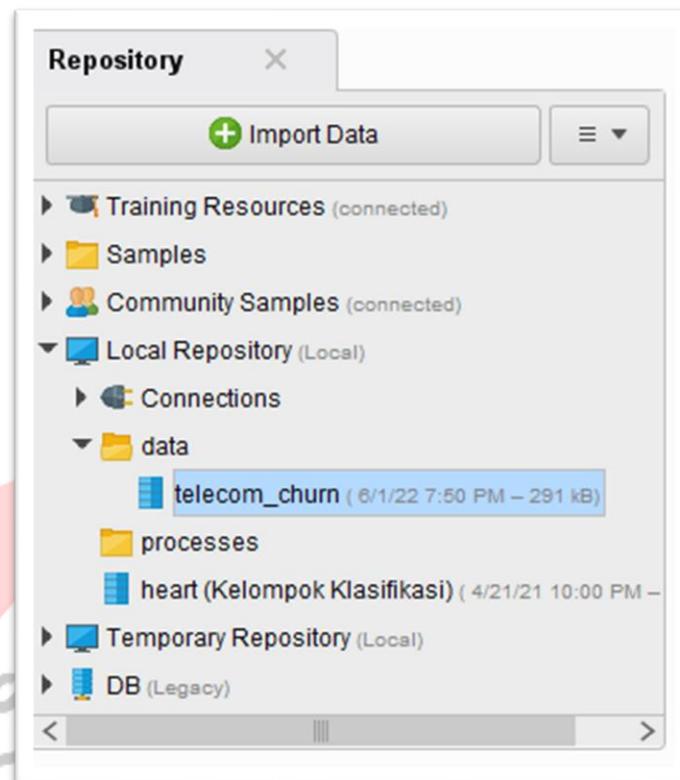
Churn	AccountWeeks	ContractRenewal	DataPlan	DataUsage	CustServCalls	DayMins	DayCalls	MonthlyCharge	OverageFee	RoamMins
0	128	1	1	2.700	1	265.100	110	89	9.870	10
0	107	1	1	3.700	1	161.600	123	82	9.780	13.700
0	137	1	0	0	0	243.400	114	52	6.060	12.200
0	84	0	0	0	2	299.400	71	57	3.100	6.600
0	75	0	0	0	3	166.700	113	41	7.420	10.100
0	118	0	0	0	0	223.400	98	57	11.030	6.300
0	121	1	1	2.030	3	218.200	88	87.300	17.430	7.500
0	147	0	0	0	0	157	79	36	5.160	7.100
0	117	1	0	0.190	1	184.500	97	63.900	17.580	8.700
0	141	0	1	3.020	0	258.600	84	93.200	11.100	11.200
1	65	1	0	0.290	4	129.100	137	44.900	11.430	12.700
0	74	1	0	0.340	0	187.700	127	49.400	8.170	9.100
0	168	1	0	0	1	128.800	96	31	5.250	11.200
0	95	1	0	0.440	3	156.600	88	52.400	12.380	12.300
0	62	1	0	0	4	120.700	70	47	15.360	13.100

ExampleSet (3,333 examples, 0 special attributes, 11 regular attributes)

Gambar 3. 2 Data Preparation Customer Churn

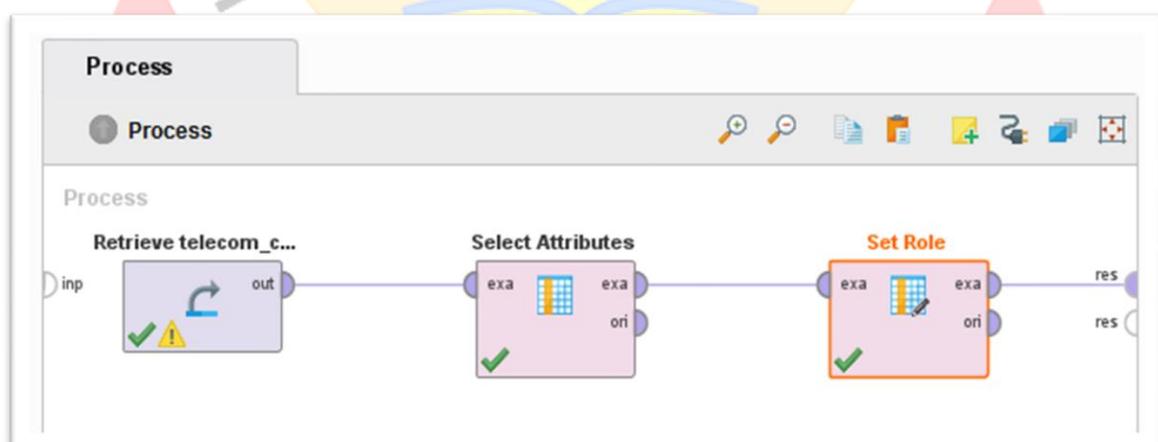
Setelah mendapatkan data set nya, peneliti melakukan data transform yang merupakan suatu usaha yang dilakukan terutama dalam menjadikan lain dari semula skala pengukuran dari data murni ke wujud beda sehingga data tersebut bisa melaksanakan dugaan – dugaan yang diterima sebagai dasar yang menjadi dasar analisa varians. Berikut cara nya :

1. Pertama kami masukan atau import data set ke aplikasi rapid miner seperti ini :



Gambar 3. 3 Import Data Set ke Rapid Miner

- Setelah itu data set di masukan ke operator “*Set Role*” lalu kita jadikan atribut “*output*” sebagai label target dari data set yang akan kita gunakan untuk melakukan klasifikasi, seperti ini proses nya :



Gambar 3. 4 Proses Set Role Data Transform di Rapid Miner

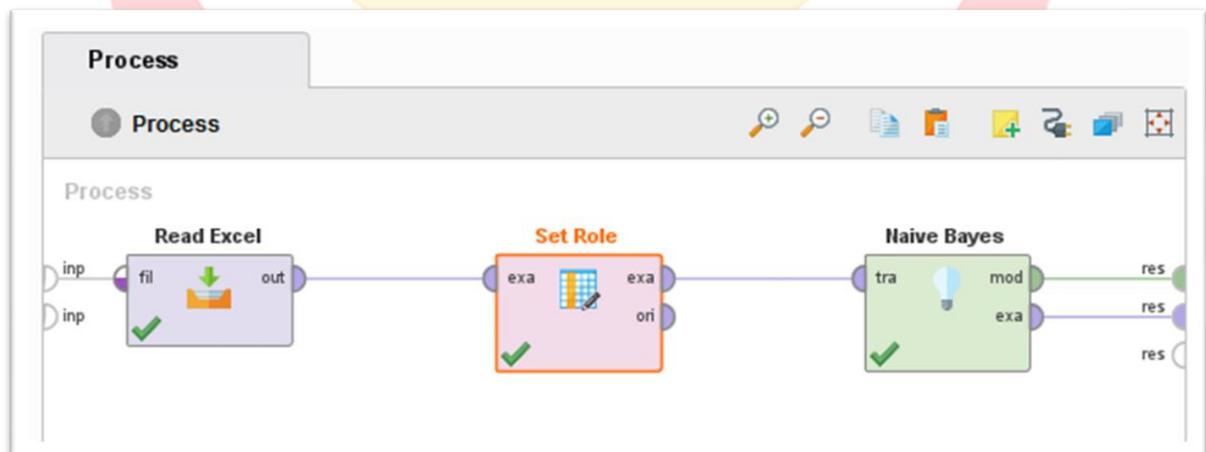
- Lalu di klik tombol run atau start pada proses nya dan output atau hasil data nya akan seperti ini :

Row No.	Churn	AccountWeeks	ContractRenewal	CustServCalls	DataPlan	DataUsage	DayCalls	DayMins	MonthlyCharge	OverageFee	RoamMins
1	0	128	1	1	1	2.700	110	265.100	89	9.870	10
2	0	107	1	1	1	3.700	123	161.600	82	9.780	13.700
3	0	137	1	0	0	0	114	243.400	52	6.060	12.200
4	0	84	0	2	0	0	71	299.400	57	3.100	6.600
5	0	75	0	3	0	0	113	166.700	41	7.420	10.100
6	0	118	0	0	0	0	98	223.400	57	11.030	6.300
7	0	121	1	3	1	2.030	88	218.200	87.300	17.430	7.500
8	0	147	0	0	0	0	79	157	36	5.160	7.100
9	0	117	1	1	0	0.190	97	184.500	63.900	17.580	8.700
10	0	141	0	0	1	3.020	84	258.600	93.200	11.100	11.200
11	1	65	1	4	0	0.290	137	129.100	44.900	11.430	12.700
12	0	74	1	0	0	0.340	127	187.700	49.400	8.170	9.100
13	0	168	1	1	0	0	96	128.800	31	5.250	11.200
14	0	95	1	3	0	0.440	88	156.600	52.400	12.380	12.300
15	0	62	1	4	0	0	70	120.700	47	15.360	13.100

Gambar 3. 5 Hasil Data Transform

3.2.4 Modeling

Pada tahap ini, 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 melakukannya:



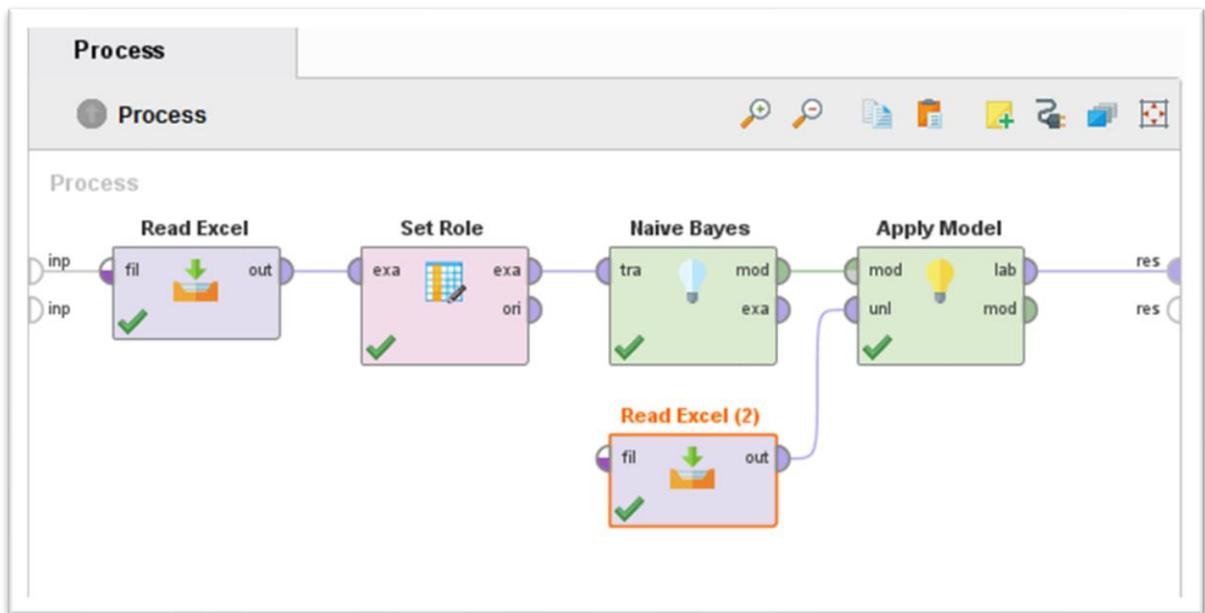
Gambar 3. 6 Proses Modeling di Rapid Miner

Keterangan:

Process operator ini, merupakan data set peneliti yang akan mencari distribution model dari label atribut *churn* yaitu *class yes* dan *class no*. Pada panel *process*, kita gunakan 3 operator yang terdiri atas:

1. *Read Excel* : digunakan untuk membaca file excel yang telah kita buat dan simpan sebelumnya dengan nama "*telecom_churn.xlsx*". *Sheet* yang peneliti gunakan adalah sheet data jadi dan peneliti memilih semua kolom dan baris.
2. *Set Role* : digunakan untuk menentukan atribut yang akan menjadi label. Kita pilih *Churn* sebagai label, tetapi sebelumnya data peneliti kelas 1 dan 0 di atribut *Churn* di ubah menjadi *Yes* dan *No*.
3. *Naïve Bayes* : digunakan untuk pengklasifikasi yang dapat membangun model yang baik bahkan dengan kumpulan data set kami yaitu "*Customer Churn*".

Setelah semua operator sudah di setting, kemudian peneliti klik tombol *start* atau *run* untuk menjalankan prosesnya di *rapid miner*. Hasilnya akan peneliti bahas pada tahap *evaluation*. Selanjutnya peneliti akan memasukan data testing di *rapid miner* untuk melakukan prediksi pada data tersebut antara *churn* atau *no churn*, berikut cara melakukan proses nya :



Gambar 3. 7 Proses Prediksi di *Rapid Miner*

Keterangan:

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*”. Sheet yang kita gunakan adalah sheet data jadi dan kita pilih semua kolom dan baris.
2. *Set Role* : digunakan untuk menentukan atribut yang akan menjadi label. Kita pilih Churn sebagai label, tetapi sebelumnya data kami 1 dan 0 kami ubah menjadi *Yes* dan *No*.
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*”. Sheet yang kita gunakan adalah sheet data testing untuk melihat prediksi dan confidence nya.
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.

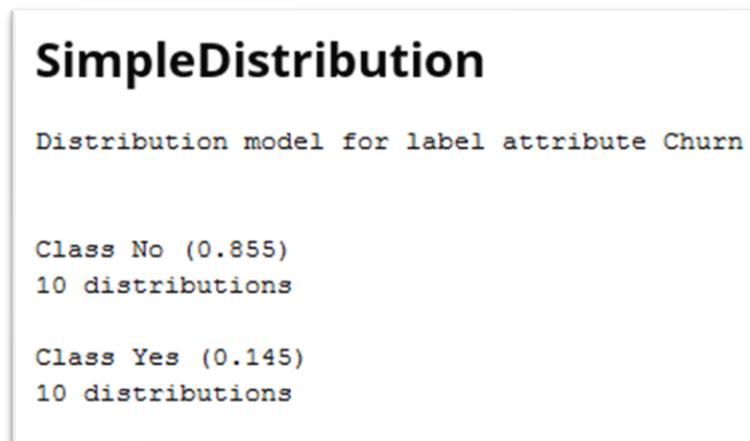
Kemudian kita klik tombol *start* atau *run* untuk menjalankan prosesnya.

Hasilnya akan peneliti bahas pada tahap *evaluation*.

3.2.5 *Evaluation*

Fase ini adalah perubahan sambungan untuk keperluan data mining. Penilaian dirinci dengan tujuan supaya menghasilkan fase pemodelan sesuai dengan tujuan yang menyampaikan untuk fase pemahaman bisnis, yakni mengeluarkan data customer yang teratur. Fase ini juga mengevaluasi seberapa baik perolehan pemodelan data mining mencukupi tujuan data mining yang ditetapkan dalam fase pemahaman bisnis. Berikut hasil pemodelan data mining nya :

- *Simple Distribution*



Gambar 3. 8 Simple Distribution

– Distribution Table

Attribute	Parameter	No	Yes
AccountWeeks	mean	100.7936842	102.6645963
AccountWeeks	standard deviation	39.88235018	39.46782041
ContractRenewal	mean	0.934736842	0.716356108
ContractRenewal	standard deviation	0.247032975	0.451233411
DataPlan	mean	0.295438596	0.16563147
DataPlan	standard deviation	0.456319728	0.372134926
DataUsage	mean	0.862150877	0.546956522
DataUsage	standard deviation	1.286095431	1.155351634
CustServCalls	mean	1.449824561	2.229813665
CustServCalls	standard deviation	1.163883327	1.853274932
DayMins	mean	175.1757544	206.9140787
DayMins	standard deviation	50.18165518	68.99779154
DayCalls	mean	100.2831579	101.3354037
DayCalls	standard deviation	19.80115695	21.58230721
MonthlyCharge	mean	55.81624561	59.19006211
MonthlyCharge	standard deviation	16.43901114	16.06547885
OverageFee	mean	9.954617544	10.62308489
OverageFee	standard deviation	2.514629336	2.586432805
RoamMins	mean	10.15887719	10.7
RoamMins	standard deviation	2.784489231	2.793189702

Tabel 3. 2 Distribution Table

– *Apply Model*

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 hasil nya :

Prediction (Churn)	Confidence (No)	Confidence (Yes)	Account Weeks	Contract Renewal	DataPlan	DataUsage
Yes	0.230584731	0.769415269	180	0	1	0.68
	CustServCalls	DayMins	DayCalls	Monthly Charge	Overage Fee	Roam Mins
	2	135.1	36	65	2.46	1.1

Tabel 3. 3 Hasil Apply Model

Disini peneliti memasukan data testing dengan atribut, sebagai berikut:

- *Account Weeks* : 180 minggu pelanggan memiliki akun aktif
- *Contract Renewal* : 0 yang dapat di artikan tidak memperbarui kontrak
- *Data Plan* : 1 yang dapat di artikan pelanggan memiliki paket data
- *Data Usage* : 0.68 gigabyte penggunaan data bulanan
- *Customer Service Calls* : 2 panggilan ke layanan pelanggan
- *Day Mins* : 135.1 menit dalam satu hari rata-rata per bulan
- *Day Calls* : 36 jumlah panggilan dalam sehari
- *Monthly Charge* : 65 rata-rata tagihan bulanan
- *Overage Fee* : 2.46 biaya kelebihan penggunaan terbesar dalam 12 bulan terakhir
- *Roam Mins* : 1.1 jumlah rata-rata menit *roaming*

Maka hasil dari atribut di atas ini di prediksi akan *customer churn* dengan *confidence churn*: 0.769415269 lebih tinggi dari *confidence no churn*: 0.230584731.

3.2.6 Deployment

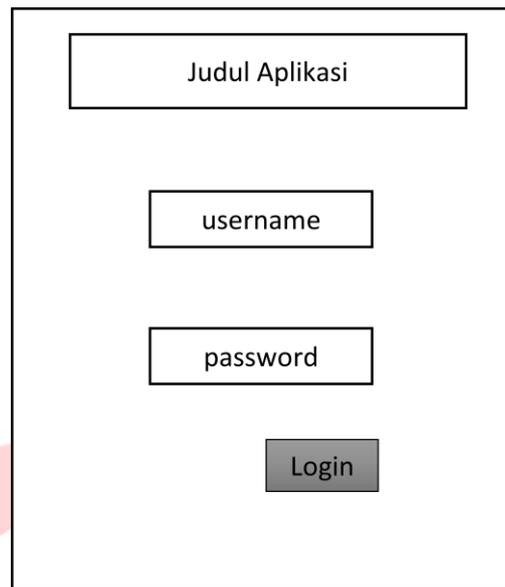
Penggunaan dataset berdasarkan indikasi *customer churn* ini dilakukan untuk mengklasifikasi yang memprediksi kemungkinan terindikasi *churn* (kehilangan pelanggan) atau sebaliknya yaitu *no churn* (pelanggan masih menggunakan layanan atau produknya) berdasarkan nilai atribut yang mempengaruhi dalam *customer churn* dari data yang sudah dikumpul sebelumnya. Dengan adanya pemodelan data mining *naïve bayes* ini kita dapat meramalnya kesempatan di masa akan depan bersumber pada yang pernah dialami di masa lalu, dan di harapkan mampu membuat strategi pemasaran bagi suatu pemilik usaha secara otomatis.

3.3 Perancangan Antarmuka

Proses merancang antarmuka menyelubungi perbuatan merancang layar, menu, dan cara merancang tampilan dalam tampilan user.

3.3.1 Perancangan Halaman Login

Rancangan ini dimana desain program dimulai dengan tampilan awal pada halaman login untuk mengidentifikasi pengguna. Pengguna wajib menginput username dan password pada field yang telah disediakan, untuk penelitian ini penulis mencantumkan *username* nya yaitu “*user*” dan *passwordnya* yaitu “*user*” sebagai bentuk default dari proses login tersebut.



Judul Aplikasi

username

password

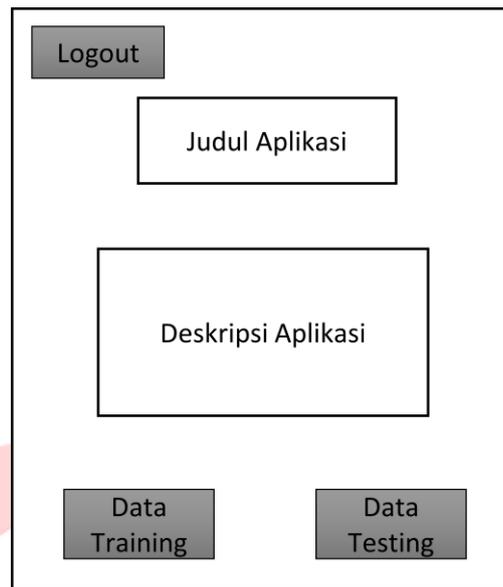
Login

Gambar 3. 9 Tampilan Menu Login

Pada rancangan ini meliputi teks field username, password, button login dan ada beberapa pemberitahuan yang akan ditampilkan seperti label yang akan muncul ketika salah mengetikkan username dan password seperti tulisan error message jika gagal login.

3.3.2 Perancangan Menu Utama

Pada rancangan menu utama akan menampilkan tampilan awal program untuk memudahkan pengguna untuk memakai aplikasi ini. Berikut rancangan menu utama :

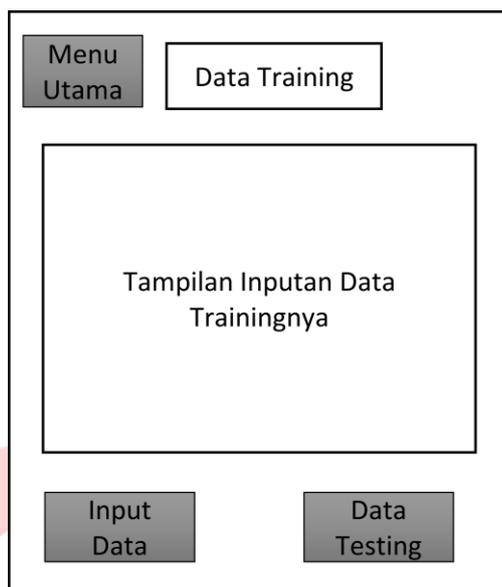


Gambar 3. 10 Tampilan Utama

Pada rancangan ini meliputi judul aplikasi, deskripsi atau keterangan aplikasi, tombol logout, tombol perpindahan menuju data training, dan data testing.

3.3.3 Perancangan Menu Data Training

Perancangan data training ini adalah bagian dataset yang dipakai dalam melatih algoritma, atau melakukan fungsi pada sebuah algoritma *naïve bayes* sesuai tujuannya yaitu untuk melakukan prediksi *customer churn*. Pembuat menyerahkan sesuatu kepada petunjuk menempuh algoritma supaya mesin yang pembuat latih dapat berusaha mendapatkan hubungan timbal baliknya otomatis. Berikut rancangan menu data training :



Gambar 3. 11 Tampilan Menu Data Training

Pada rancangan ini meliputi tabel tampilan data training, button input data training, button ke menu data testing dan tombol pindah menuju halaman utama.

3.3.4 Perancangan Input Data Training

Pada rancangan ini berfungsi untuk memasukan data set ke tampilan menu data training untuk membuat prediksi. Berikut rancangan tampilan input data training nya :

Gambar 3. 12 Tampilan Input Data Training

Pada rancangan ini meliputi field atribut-atribut, button input, tombol cancel yang menuju pada menu halaman data training.

3.3.5 Perancangan Data Testing

Perancangan data testing ini adalah bagian dataset yang di tes untuk memprediksi customer churn untuk mengamati keakuratannya, atau dengan maksud lain mengamati performanya. Berikut rancangan tampilan input data testing nya:

The image shows a software interface for data testing. It contains a 'Menu Utama' button in the top left corner and a 'Data Testing' button in the top right corner. Below these buttons, there are four rows of input fields, each starting with the label 'Atribut :'. At the bottom of the interface is a 'Prediksi' button.

Gambar 3. 13 Tampilan Menu Data Testing

Pada rancangan ini meliputi field atribut-atribut, button input, tombol menu utama dan tombol menu prediksi yang menuju pada menu hasil prediksi.

3.3.6 Perancangan Menu Hasil Prediksi

Perancangan tampilan hasil prediksi ini dapat memberikan 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 terjadinya customer churn untuk melakukan

strategi pemasaran yang lebih atau lainnya. Berikut rancangan tampilan menu hasil prediksi nya:



Gambar 3. 14 Tampilan Hasil Prediksi

Halaman ini memuat judul perolehan prediksi, pendapatan dari perkiraan yang berupa informasi ketepatan, tombol yang wajib di pilih dan tombol balik ke tempat dalam ke tampilan sebelumnya yaitu halaman menu utama.

3.4 Jadwal Penelitian

Jadwal kegiatan dalam merancang dan pembuatan penelitian ini di rencanakan hendak sejak bulan Maret hingga pada bulan Juli 2022. Berikut ini merupakan tabel hasil menyusun kegiatan pembuatan rancangan yang hendak penulis laksanakan.

No	Kegiatan	2022																			
		Maret				April				Mei				Juni				Juli			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengumpulan Data				■	■															
2	Analisis Data						■	■	■	■											
3	Design Aplikasi									■	■	■	■	■	■						
4	Coding Program													■	■	■	■	■			

