

## Penerapan Data Mining dengan Algoritma *Naive Bayes Clasifier* untuk Mengetahui Minat Beli Pelanggan terhadap Kartu *Internet XL* (Studi Kasus di CV. Sumber Utama Telekomunikasi)

Dicky Nofriansyah<sup>#1</sup>, Kamil Erwansyah<sup>#2</sup>, Mukhlis Ramadhan<sup>#3</sup>

<sup>#1</sup> Prodi Sistem Informasi, STMIK Triguna Dharma Medan

<sup>#2,3</sup> Prodi Sistem Komputer, STMIK Triguna Dharma Medan

Email. : <sup>#1</sup>[dickynofriansyah@gmail.com](mailto:dickynofriansyah@gmail.com)<sup>1</sup>

### Abstrak

Persaingan yang terjadi dalam dunia bisnis memaksa para pelakunya untuk selalu memikirkan strategi-strategi dan terobosan yang dapat menjamin kelangsungan dari bisnis yang dijalankannya. Hal ini akan memunculkan persaingan antar sesama provider kartu terhadap kartu internet. Para provider kartu internet berlomba-lomba menarik minat pelanggan dengan berbagai macam strategi pemasaran agar tidak kalah saing dan tetap eksis. Dan perusahaan ingin selalu meluncurkan kartu internet terbaru tanpa memikirkan kartu internet tersebut akan laku atau tidak dipasaran. Konsep data mining akan memudahkan cara menyelesaikan masalah yang terjadi di CV. Sumber Utama Telekomunikasi. Maka, metode klasifikasi mampu menemukan model yang membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Oleh sebab itu, algoritma naive bayes dapat memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya. Hasil dari penelitian ini agar dapat memprediksi atau memperkirakan laku atau tidak kartu internet yang baru, sehingga perusahaan dapat mengambil keputusan dan meningkatkan strategi pemasaran.

**Kata Kunci** : data mining, kartu internet, klasifikasi, algoritma naive bayes

### Abstract

*Competition is happening in the business world to force the perpetrators to always think of the strategy - and a breakthrough strategy that can guarantee the continuity of the business being operated. This leads to competition among card providers on the internet card. The provider of internet card race-the race to attract customers with a variety of marketing strategies in order not to lose competitiveness and still exist. And the company wants to always launch the latest internet card without thinking of internet cards will be sold in the market or not. The concept of data mining will ease the way to solve the problem that occurred in the CV. Main source of Telecommunications. Thus, the classification method is able to find a model that distinguishes the concept or class of data, in order to be able to predict the class of an unknown object label. Therefore, Naive Bayes algorithm can predict future opportunities based on the experience of earlier. Results from this study in order to predict or estimate the behavior or the new Internet card, so that the company can make decisions and improve marketing strategy.*

*Keywords: data mining, internet cards, classification, naive Bayes algorithm*

## A. PENDAHULUAN

Persaingan disuatu perusahaan membuat para pelakunya harus selalu memikirkan strategi-strategi terobosan yang dapat menjamin kelangsungan bisnis mereka. Pada suatu perusahaan memiliki ketersediaan data yang melimpah. Ini melahirkan kebutuhan akan adanya teknologi yang dapat memanfaatkannya untuk membangkitkan pengetahuan-pengetahuan baru, yang dapat membantu dalam pengaturan strategi dalam menjalankan bisnis. Prediksi minat konsumen akan sangat penting bagi perusahaan, dimana dengan adanya prediksi minat konsumen perusahaan dapat mengambil suatu keputusan atau strategi yang benar dan tepat bagi konsumennya.

*Database* dalam jumlah yang besar merupakan salah satu aset berharga yang dimiliki sebuah perusahaan. Sebagai salah satu perusahaan yang bergerak di bidang bisnis telekomunikasi, pihak marketing CV. Sumber Utama Telekomunikasi haruslah memikirkan strategi dalam pemasaran untuk mampu menghasilkan suatu informasi yang siap digunakan untuk membantu pihak marketing dalam mengambil keputusan strategis pemasaran. Untuk memenuhi kebutuhan-kebutuhan di atas, banyak cara yang dapat ditempuh. Salah satunya adalah dengan melakukan pemanfaatan *database* perusahaan menggunakan teknik *Data Mining*.

Saat ini perusahaan banyak sekali mengeluarkan kartu internet yang berbagai macam jenis tanpa memperkirakan kartu tersebut laku atau tidaknya dipasarkan dimasyarakat. Hal ini tentu akan membuat perusahaan merugi jika kartu internet tersebut tidak laku terjual dimasyarakat.

Metode *Klasifikasi* adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan

konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses *klasifikasi* membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Model itu sendiri bisa berupa aturan "jika-maka", berupa pohon keputusan, atau formula matematis.

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan *Inggris Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. *Teorema* tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas.

### 1. Tujuan Penelitian

Secara umum, tujuan penelitian ini adalah untuk memberikan gambaran mengenai minat beli pelanggan terhadap kartu *internet* XL pada CV. Sumber Utama Telekomunikasi.

- Untuk memanfaatkan *data mining* menghasilkan informasi dalam mengetahui minat beli kartu *internet* XL dengan menggunakan *metode klasifikasi*.
- Untuk menetapkan kriteria dalam menentukan kartu *internet* XL yang baru dengan menggunakan algoritma *naive bayes*.

### 2. Manfaat Penelitian

Adapun manfaat yang dapat diambil dari hasil penelitian ini adalah:

- Sebagai acuan dalam mengembangkan pengetahuan tentang bagaimana

- mendapatkan dan menarik minat pelanggan terhadap kartu *internet* XL.
- b. Bagi pihak CV. Sumber Utama Telekomunikasi dapat dijadikan sebagai masukan-masukan untuk menarik minat pelanggan yang lebih baik lagi.
  - c. Sebagai bahan acuan bagi peneliti selanjutnya dalam mengembangkan penelitian yang lebih lanjut.
  - d. Dapat menghasilkan informasi dalam mengetahui minat beli kartu dengan menggunakan metode *Klasifikasi*.

### 3. Metodologi Penelitian

Metodologi penelitian yang digunakan untuk mendukung pembuatan Penelitian "Penerapan Data Mining dengan Algoritma *Naive Bayes Clasifier* Untuk Mengetahui Minat Beli Pelanggan Terhadap Kartu *Internet* XL Menggunakan Metode Klasifikasi (Studi Kasus CV. Sumber Utama Telekomunikasi)" ini adalah:

#### a. Pengumpulan Data

Pada tahap ini yang dikumpulkan berupa data-data transaksi penjualan kartu *internet*XL sebagai sampel pada tahun 2014 pada CV. Sumber Utama Telekomunikasi.

#### 1) Studi Literatur

Studi literatur yang digunakan dalam penulisan Penelitian ini adalah mengumpulkan bahan referensi mengenai metode *klasifikasi* dari berbagai buku, jurnal, dan beberapa referensi lainnya.

#### 2) Observasi

Metode pengumpulan data yang dilakukan dengan mengadakan pengamatan atau peninjauan langsung terhadap sumber permasalahan serta berkomunikasi langsung dengan pihak yang

bersangkutan serta bertanggungjawab dalam penentuan sertifikasi pada sekolah tersebut.

#### 3) Wawancara

Metode pengumpulan data yang dilakukan melalui tanya jawab yang diajukan secara langsung kepada narasumber untuk mendapatkan data atau informasi yang berkaitan dengan objek penelitian.

#### 1. Analisa Permasalahan

Pada tahap ini akan dilakukan analisis dengan metode *Klasifikasi* pada data transaksi penjualan kartu *internet* XL.

#### 2. *Preprocessing* Data

Mempersiapkan data yang akan ditambah.

#### 3. Pengujian

Pada tahap ini dilakukan proses pengujian dan percobaan terhadap aplikasi sesuai dengan kebutuhan yang ditentukan sebelumnya agar berjalan seperti yang diharapkan.

### B. TINJAUAN PUSTAKA

*Data Mining* dan Analisis Kebutuhan Sistem memiliki suatu keterkaitan antara satu dengan yang lainnya sesuai dengan bidangnya masing-masing. Untuk itu pemanfaatan teknologi dan sumber daya yang ada merupakan salah satu faktor yang perlu diperhatikan bagi seseorang yang akan melakukan pengolahan data.

Dalam perkembangannya *data mining* memiliki banyak definisi yang cukup beragam sehingga *data mining* dapat menambah ilmu pengetahuan. Berikut ini adalah beberapa definisi *data mining* pada umumnya:

Menurut Turban, dkk (dalam Kusriani, 2009:3)

*Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

Menurut Turban, dkk (dalam Kusri, 2009:4)

*Data Mining* merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari database yang besar.

Menurut Pramudiono (dalam Kusri, 2009:3) *Data Mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual”.

Jadi dapat disimpulkan bahwa, *data mining* adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data.

Berdasarkan definisi-definisi yang telah disampaikan, hal penting yang terkait dengan *data mining* adalah:

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan *data mining* adalah mendapatkan hubungan atau pola

yang mungkin memberikan indikasi yang bermanfaat.

Istilah *data mining* dan *Knowledge Discovery in Database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses *Knowledge Discovery in Database* (KDD) adalah *data mining*.

Proses *Knowledge Discovery in Database* (KDD) secara garis besar dapat dijelaskan sebagai berikut:

1. Pembersihan data (*data cleaning*)  
Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari *database* suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa *data mining* yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.
2. Integrasi data (*data integration*)  
Integrasi *data* merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari satu *database* tetapi juga berasal dari beberapa *database* atau *file text*. Integrasi data

dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

3. Seleksi Data (*Data Selection*)

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus *market basket analysis*, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

4. Transformasi data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan *clustering* hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi - bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses *mining*

Merupakan suatu proses utama saat

metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi Pola (*Pattern Evaluation*)

Untuk mengidentifikasi pola-pola *knowledge based* yang ditemukan. Dalam tahap ini hasil dalam teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila hasil yang dicapai tidak sesuai hipotesa, ada beberapa alternative yang dapat diambil seperti menjadikan untuk memperbaiki proses data *mining*, mencoba metode *data mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7. Presentasi pengetahuan (*knowledge presentation*),

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami *data mining*. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses *data mining*. Data Mining dibagi menjadi beberapa metode berdasarkan tugas yang dapat dilakukan, yaitu:

1. Asosiasi (*Association*)

Digunakan untuk mengenali kelakuan dari kejadian - kejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian. Salah

satu contohnya *Market Basket Analysis*, yaitu salah satu metode asosiasi yang menganalisa kemungkinan pelanggan untuk membeli beberapa item secara bersamaan. Penting tidaknya suatu aturan asosiasi dapat diketahui dengan dua parameter yaitu:

a. *Support*

Suatu ukuran yang menunjukkan seberapa besar tingkat dominasi *item / itemset* dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu *item / itemset* layak untuk dicari *confidence* tersebut (contoh, dari keseluruhan transaksi yang ada, seberapa besar tingkat dominasi yang menunjukkan *item A* dan *B* dibeli bersamaan).

b. *Confidence*

Suatu ukuran yang menunjukkan hubungan antar dua *item* secara *conditional* (contoh, seberapa sering *item B* dibeli jika orang membeli *item A*) (Kusrini dan Lutfi 2009:150)

2. Pengklusteran (*clustering*)

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan

(*homogeny*), yang mana kemiripan dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

Contoh pengklusteran dalam bisnis dan penelitian adalah:

- a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari satu suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- b. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap *ial* dalam baik dan mencurigakan.
- c. Melakukan pengklusteran terhadap ekspresi dari *gen*, untuk mendapatkan kemiripan perilaku dari *gen* dalam jumlah besar. (Kusrini dan Lutfi 2009:12)

3. Prediksi (*prediction*)

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.

Contoh prediksi bisnis dan penelitian adalah:

- a. Prediksi harga beras dalam tiga bulan yang akan datang.
- b. Prediksi persentasi kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi. (Kusrini dan Lutfi 2009:11).

4. Estimasi (*estimation*)

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik dari pada kearah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari

variabel target sebagai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

Contoh lain yaitu estimasi nilai indeks prestasi kumulatif mahasiswa program pasca sarjana dengan melihat nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana. (Kusrini dan Lutfi 2009:10)

#### 5. Klasifikasi (*classification*)

Klasifikasi adalah fungsi pembelajaran yang memetakan (mengklasifikasi) sebuah unsur (*item*) data ke dalam salah satu dari beberapa kelas yang sudah didefinisikan. Contoh lain klasifikasi dalam bisnis dan penelitian adalah:

- a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau tidak.
- b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- c. Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa. (Kusrini dan Lutfi 2009:11)

Algoritma *Naive Bayes Classifier* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik

yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. *Teorema* tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. *Klasifikasi Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya.

Persamaan dari *teorema Bayes* adalah : 
$$P(H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Keterangan:

**X**: Sampel data yang memiliki kelas (label) yang tidak diketahui

**H**: Hipotesa bahwa X adalah data kelas (label)

**P(H)**: Peluang dari hipotesa H

**P(X)**: Peluang dari data sampel yang diamati

**P(X|H)**: Peluang dari data sampel X bila diasumsikan bahwa hipotesa benar

Sehingga *Naive Bayesian Classifier* dapat didefinisikan juga sebagai metode klasifikasi yang berdasarkan teori probabilitas dan teorema bayesian dengan asumsi bahwa setiap variabel ataupun parameter penentu keputusan bersifat bebas (*independence*) sehingga keberadaan setiap variabel tidak ada kaitannya dengan keberadaan atribut yang lain.

Adapun alur dari metode *Naive Bayes* adalah sebagai berikut:

1. Menghitung Nilai Peluang Kasus Baru Dari Setiap Hipotesa dengan Klas (Label) yang ada " $P(X|Ci)$ "
2. Menghitung Nilai Akumulasi Peluang Dari Setiap Klas " $P(X|Ci)$ "
3. Menghitung Nilai  $P(X|Ci) \times P(Ci)$
4. Menentukan Klas dari Kasus baru tersebut.

### C. ANALISA

Dalam menentukan peminatan kartu internet baru yang akan diluncurkan adalah dengan mengumpulkan data – data kartu internet yang sudah dipasarkan sebelumnya. Untuk mengatasi permasalahan dalam menentukan peminatan kartu baru yang akan diluncurkan, diperlukan suatu alat analisis bagi perusahaan untuk mengetahui minat masyarakat terhadap kartu internet baru yang diluncurkan sehingga dapat diketahui kartu tersebut diminati atau tidak oleh masyarakat.

Maka, metode klasifikasi mampu menemukan model yang membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Oleh sebab itu, algoritma *naive bayes* dapat memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya. Sehingga dapat mengetahui minat masyarakat pada kartu internet yang akan diluncurkan. Berikut ini adalah daftar jenis kartu internet XL pada CV. Sumber Utama Telekomunikasi.

Tabel 3.1 Data Kasus

No.	Kode Kartu	Kuota	Masa Aktif	Harga	Klasifikasi
1	BBS5	Dibawah 7GB	1 Minggu	Rendah	Tidak
2	BGD25	Dibawah 7GB	1 Bulan	Rendah	Tidak
3	BGD49	Diatas 7GB	1 Bulan	Rendah	Tidak
4	IH10	Dibawah 7GB	1 Minggu	Rendah	Tidak
5	IH25	Dibawah 7GB	1 Bulan	Rendah	Minat
6	IH75	Dibawah 7GB	1 Bulan	Sedang	Tidak
7	IH99	Dibawah 7GB	1 Bulan	Sedang	Minat
8	IHP199	Dibawah 7GB	1 Bulan	Tinggi	Tidak
9	IHP449	Diatas 7GB	1 Bulan	Tinggi	Tidak
10	IU149	Diatas 7GB	1 Bulan	Tinggi	Minat
11	IU49	Dibawah 7GB	1 Bulan	Rendah	Minat
12	IUN12	Dibawah 7GB	1 Minggu	Rendah	Minat
13	SN30	Dibawah 7GB	1 Bulan	Rendah	Minat
14	SN55	Diatas 7GB	1 Bulan	Rendah	Minat
15	XO30	Dibawah 7GB	6 Bulan	Rendah	Minat
16	XO60	Dibawah 7GB	6 Bulan	Sedang	Minat
17	XO40	Dibawah 7GB	6 Bulan	Rendah	Tidak
18	IH49	Dibawah 7GB	1 Bulan	Rendah	Minat
19	IUN49	Dibawah 7GB	1 Bulan	Rendah	Minat
20	IHPRO75	Dibawah 7GB	1bulan	Sedang	????

Sumber : Data CV. Sumber Utama Telekomunikasi

#### Keterangan:

- Kriteria 1 = Menjelaskan tentang Kriteria “Kuota”
- Kriteria 2 = Menjelaskan tentang Kriteria “Masa Aktif”
- Kriteria 3 = Menjelaskan tentang Kriteria “Harga”

Tahap awal cara kerja dari proses perhitungan *Naive Bayes* adalah dengan cara melakukan pengambilan data training dari penjualan kartu *internet*. Adapun variabel penentu yang digunakan dalam mengklasifikasikan data kartu *internet* yaitu:



1. Kode kartu

Merupakan variabel jenis kartu internet yang dikelompokkan dalam beberapa kode kartu internet.

2. Kuota

Merupakan variabel yang dikelompokkan dalam dua kategori yaitu kuota diatas 7GB dan kuota dibawah 7GB. Kuota diatas 7GB merupakan batas kuota dari kuota 7GB, 8GB ,11GB dan 12GB. Kuota dibawah 7GB merupakan batas kuota dari kuota 6GB, 5GB , 4GB, 2GB dan 500MB.

Kriteria	Kategori	Keterangan
Kuota	Diatas 7GB	7GB – 12 GB
	Dibawah 7GB	200MB – 6,5GB

3. Masa Aktif

Merupakan variabel lamanya masa pakai kartu internet yang dikelompokkan dalam 3 kategori yaitu 1 minggu, 1 bulan, dan 6 bulan. Kriteria masa aktif dapat dilihat dari tabel 3.3.

Tabel 3.3 Kriteria Masa Aktif

Kriteria	Kategori	Keterangan
Masa Aktif	1 Minggu	7 hari di mulai dari mengaktifkan Kartu
	1 Bulan	30 hari di mulai dari mengaktifkan Kartu
	6 Bulan	90 hari di mulai dari mengaktifkan Kartu

4. Harga

Merupakan variabel jenis harga kartu internet yang dikelompokkan dalam tiga kategori yaitu tinggi, rendah dan sedang.

Adapun kategori tersebut mempunyai nilai batas ambang yaitu:

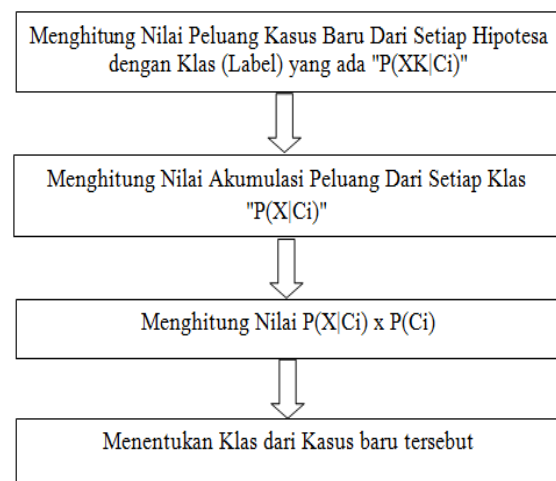
- Kategori “tinggi” mulai dari harga Rp 80.000,00 s/d Rp 199.000,00
- Kategori “sedang” mulai dari harga Rp 50.000,00 s/d Rp 79.000,00
- Kategori “rendah” mulai dari harga Rp 5.000,00 s/d 40.000,00.

Kriteria harga dilihat dari tabel 3.4.

Tabel 3.4 Kriteria Harga

Kriteria	Kategori	Keterangan
Harga	Tinggi	Rp 80.000 – Rp 199.000
	Sedang	Rp 50.000 – Rp 79.000
	Rendah	Rp 5.000 – Rp 49.000

Penyelesaian masalah menggunakan Algoritma *Naive Bayes* pada gambar 3.1 untuk memudahkan pemahaman dan alur proses yang akan dibahas. Penyelesaian masalah menggunakan Algoritma *Naive Bayes* pada gambar 3.1 untuk memudahkan pemahaman dan alur proses yang akan dibahas.



Gambar 3.1 Diagram Penyelesaian Masalah *Naive Bayes*

Sehingga Algoritma dari sistem yang dibangun adalah sebagai berikut:

1. Menghitung nilai peluang kasus baru dari setiap hipotesa dengan klas (label) yang ada "P(XK|Ci)"
2. Menghitung nilai akumulasi peluang dari setiap klas "P(X|Ci)"
3. Menghitung Nilai dari setiap klas
4. Menentukan kelas dari kasus tersebut

Berdasarkan tabel 3.1 dapat dihitung klasifikasi data kartu internet apabila diberikan input berupa kuota, masa aktif, dan harga menggunakan algoritma *Naive Bayes*.

Apabila diberikan input data baru, maka klasifikasi data kartu internet dapat ditentukan melalui langkah berikut:

a. Menghitung nilai P(XK|Ci)  
 P(Kriteria 1= "Di bawah 7 GB" | Klasifikasi = "Minat")  
 $P(\text{Kriteria 1} = 9/15 = 0,6$

P(Kriteria 1= "Di bawah 7 GB" | Klasifikasi = "Tidak")  
 $P(\text{Kriteria 1} = 6/15 = 0,4$

P(Kriteria 2= "1 Bulan" | Klasifikasi = "Minat")  
 $P(\text{Kriteria 2} = 8/13 = 0,62$

P(Kriteria 2= "1 Bulan" | Klasifikasi = "Tidak")  
 $P(\text{Kriteria 2} = 5/13 = 0,38$

P(Kriteria 3= "Sedang" | Klasifikasi = "Minat")  
 $P(\text{Kriteria 3} = 2/3 = 0,66$

P(Kriteria 3= "Sedang" | Klasifikasi = "Tidak")  
 $P(\text{Kriteria 3} = 1/3 = 0.33$

b. Hitung nilai P(X|Ci) untuk setiap Kelas (label)  
 P(Kriteria 1= "Di bawah 7 GB"

| Klasifikasi = "Minat")  
 $P(\text{Kriteria 1} = 9/15 = 0,6$

- P(Kriteria 1= "Di bawah 7 GB" | Klasifikasi = "Tidak")  
 $P(\text{Kriteria 1} = 6/15 = 0,4$

- P(Kriteria 2= "1 Bulan" | Klasifikasi = "Minat")  
 $P(\text{Kriteria 2} = 8/13 = 0,62$

- P(Kriteria 2= "1 Bulan" | Klasifikasi = "Tidak")  
 $P(\text{Kriteria 2} = 5/13 = 0,38$

- P(Kriteria 3= "Sedang" | Klasifikasi = "Minat")  
 $P(\text{Kriteria 3} = 2/3 = 0,66$

- P(Kriteria 3= "Sedang" | Klasifikasi = "Tidak")  
 $P(\text{Kriteria 3} = 1/3 = 0.33$

c. Hitung nilai P(X|Ci) untuk setiap Kelas (label)

$P(X|Klasifikasi = "Minat") = 0,6 \times 0,62 \times 0,66 = 0,245$   
 $P(X|Klasifikasi = "Tidak") = 0,4 \times 0,38 \times 0,33 = 0,050$

d. Hitung nilai P(X|Ci) \* P(Ci)  
 (P(X|Klasifikasi = "Minat") x P(Klasifikasi = "Minat")) =  $0,245 \times 11/19 = 0.141842$

(P(X|Klasifikasi = "Tidak") x P(Klasifikasi = "Tidak")) =  $0,050 \times 8/19 = 0,021053$

5. Menentukan Kelas Dari Kasus Baru  
 Berdasarkan perhitungan akhir dengan mengalikan nilai peluang dari kasus yang di angkat, kita melihat bahwa nilai P(X|Keterangan="Minat") lebih tinggi dari P(X|Keterangan="Tidak") = 0.141842 banding 0.021053. Sehingga dapat

disimpulkan bahwa Kartu Internet tersebut masuk dalam klasifikasi “Minat”

#### D. SIMPULAN DAN SARAN

Berdasarkan perancangan, analisis, implementasi dan pengujian pada Penerapan Data Mining dengan Algoritma *Naive Bayes Classifier* Untuk Mengetahui Minat Beli Pelanggan Terhadap Kartu Internet XL Menggunakan Metode Klasifikasi, maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Sistem klasifikasi data kartu internet ini digunakan untuk menampilkan informasi klasifikasi minat atau tidak minat pada kartu internet yang baru di luncurkan dengan menggunakan algoritma *Naive Bayes*.
2. Algoritma *Naive Bayes* sangat cocok diterapkan dalam memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga memudahkan perusahaan untuk memprediksi peminatan masyarakat terhadap kartu internet yang baru diluncurkan.
3. Dengan mengetahui minat atau tidak minat kartu internet yang baru diluncurkan, akan meminimalisir kerugian pada perusahaan. Perusahaan juga akan lebih selektif dalam meluncurkan produk baru.

Untuk lebih mengembangkan dan meningkatkan Implementasi dari Algoritma *Naive Bayes*, ada beberapa saran yang dapat dijadikan pertimbangan, yaitu :

1. Produk – produk kartu internet yang akan diluncurkan hendaknya di *display* sesuai dengan tingkat *Support* dan *Confidence* dari produk *furniture* tersebut.
2. Volume data yang digunakan untuk uji coba disarankan ditambah. Jumlah data yang besar dan luas akan benar – benar dapat digunakan sebagai salah

satu gambaran pengambilan keputusan serta penggunaan nilai *Support* dan *Confidence* yang bervariasi, sehingga dapat menghasilkan lebih banyak aturan asosiasi antar data yang mengandung informasi penting yang lebih bermanfaat.

3. Penulis mengharapkan agar penelitian ini dilakukan juga dengan menggunakan metode atau algoritma selain apriori untuk tujuan pengembangan penelitian di masa yang akan datang.

#### E. DAFTAR PUSTAKA

- Adi, N. 2004. *Analisis dan Perancangan Sistem Informasi Dengan Metodologi Berorientasi Objek, Informatika*. Yogyakarta: Andi Offset.
- Dicky Nofriansyah. 2015. *Algoritma Data Mining dan Pengujiannya*. Yogyakarta: Deepublish
- Edi Suhendi. 2011. *Buku Pintar Microsoft Office Excel 2007*. Bandung: Super Computer Publishing.
- Fajar A. H. 2013. *Data Mining*. Yogyakarta: Andi Offset.
- Kusrini & Emha, T. L. 2009. *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Tim Penyusun. 2014. *Panduan Sisten Informasi*. Medan: STMIK Triguna Dharma.
- Tim Penyusun. 2014. *Pembekalan Analisis & Perancangan Sistem Informasi*. Medan: STMIK Triguna Dahrma.

Wicaksana, & dkk. 2009. *Belajar Data Mining Dengan Rapid Miner*. Yogyakarta: Andi Offset.

[http://www.insidebigdata.com/2014/10/10/rapidminer-moves-predictive\\_analytics-data-mining-machine-learning-cloud](http://www.insidebigdata.com/2014/10/10/rapidminer-moves-predictive_analytics-data-mining-machine-learning-cloud),  
Diakses, 20 Februari 2015.