

BAB 2

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Pada observasi ini penulis menggunakan beberapa jurnal yang digunakan untuk refrensi, berikut merupakan jurnal yang memiliki hubungan dengan penjualan dengan menggunakan berbagai macam metode dan jurnal yang menggunakan metode naïve bayes untuk membantu penyelesaian masalah observasi.

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

Persaingan yang terjadi dalam dunia bisnis memaksa para pelakunya untuk selalu memikirkan strategi – strategi dan terobosan yang dapat menjamin kelangsungan dari bisnis yang dijalankannya. 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.

2. Penerapan Metode Naïve Bayes Untuk Klasifikasi Pelanggan Telekomunikasi Selular (STUDI KASUS PAKDD 2006 KOMPETISI DATA MINING)

Perkembangan teknologi telekomunikasi hingga saat ini berkembang semakin pesat. Perusahaan-perusahaan yang bergerak dalam bidang jasa telekomunikasi juga mulai mengembangkan jaringan telekomunikasi generasi ketiga (3G) untuk para pelanggan dengan fitur-fitur yang lebih modern. Seiring dengan kemajuan teknologi, kebutuhan yang akan digunakan dalam penyimpanan data juga semakin besar. Akan tetapi, data yang besar tidak akan berguna jika informasi yang terkandung di dalamnya tidak diketahui. Oleh karena itu digunakan teknik data mining untuk mengakses informasi tersebut. Dalam pembahasan ini, data mining digunakan untuk menentukan pelanggan 2G dan 3G serta memprediksi pelanggan 2G yang berpotensi beralih ke layanan 3G dengan studi kasus PAKDD 2006 Kompetisi Data Mining. Metode yang digunakan dalam memecahkan permasalahan tugas akhir ini dengan menggunakan Naive Bayes. Hal ini dikarenakan Naive Bayes merupakan salah satu metode yang mudah digunakan dalam mode klasifikasi. Dengan banyaknya atribut yang ada, maka digunakan metode feature selection untuk menentukan atribut yang akan digunakan dalam mode klasifikasi. Semakin besar peluang pelanggan 2G yang beralih ke jaringan 3G, maka pihak perusahaan pun akan semakin intensif dalam memenuhi target pasar yang potensial. Diharapkan semakin bertambahnya pelanggan 3G akan semakin meningkatkan profit bagi perusahaan.

Tabel 0.1 Ringkasan Penelitian

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
1.	Dicky Nofriansyah ¹ , Kamil Erwansyah ² , Mukhlis Ramadhan ³ , Mei 2016	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 Naïve Bayes	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.
2.	Yudo Indra Dewa, Moch. Arif Bijaksana, Arie Ardiyanti Suryani, 2009	Seiring dengan kemajuan teknologi, kebutuhan yang akan digunakan dalam penyimpanan data juga semakin besar. Akan tetapi, data yang besar tidak akan berguna jika informasi yang terkandung di dalamnya	Metode Naïve Bayes	Data mining digunakan untuk menentukan pelanggan 2G dan 3G serta memprediksi pelanggan 2G yang berpotensi beralih ke layanan 3G dengan studi kasus PAKDD 2006 Kompetisi Data Mining. Metode yang digunakan

		tidak diketahui.		dalam memecahkan masalah tugas akhir ini yaitu Naive Bayes.
--	--	------------------	--	---

2.1.1 Simpulan dari Dua Jurnal yang Terkait

Bahwa data mining bermanfaat untuk mengelola data perusahaan yang tak terhitung jumlahnya dan sangat efisien dalam penyelesaian karena tidak menghabiskan waktu yang cukup lama, sehingga dapat memajukan pendapatan perusahaan dan meningkatkan strategi perusahaan dalam bidang bisnis.

2.1.2 Perbedaan Penelitian yang Terkait

Berikut perbedaan penelitian penulis yang diteliti dengan penelitian sebelumnya adalah.

1. Data yang digunakan oleh penulis yaitu data penjualan produk Yamaha di PT Arista Mitra Lestari Cabang Pamanukan periode Januari 2015 – Juli 2016.
2. Data yang dianalisis dan diproses oleh penulis akan dijadikan klasifikasi data penjualan dengan menggunakan parameter produk, warna, dan alamat customer.
3. Metode yang digunakan penulis yaitu data mining klasifikasi naïve bayes yang digunakan untuk memprediksi selera konsumen dalam membeli produk Yamaha di PT Arista Mitra Lestari Cabang Pamanukan, untuk melihat keakurasian data yang diteliti menggunakan perangkat lunak WEKA, perhitungan memakai bahasa PHP dan aplikasi interface menggunakan bahasa pemrograman PHP.

2.2 Landasan Teori

2.2.1 Penjualan

Penjualan merupakan sebuah kegiatan yang dilakukan untuk dikembangkan beberapa *planning* strategis yang ditujukan dalam usaha pemasaran kebutuhan dan keinginan konsumen untuk memperoleh penjualan yang mendapatkan keuntungan[3]. Penjualan

diartikan sebagai sumber hidup perusahaan, karena perusahaan mendapatkan keuntungan serta usaha menarik *customer* yang mengusahakan untuk melihat pesona pembeli sehingga dapat mengetahui hasil jual produk yang dihasilkan[3]. Dari pengertian tersebut dapat disimpulkan, bahwa arti penjualan yaitu pemindahan barang dan jasa yang dibutuhkan oleh beberapa orang yang bekerja dibidang penjualan seperti melaksanakan dagangan, agen, pelayanan wakil, dan wakil pemasaran.

2.2.2 Jenis Penjualan

Jenis penjualan dibedakan menjadi beberapa jenis[3], sebagai berikut.

1. Penjualan secara langsung yaitu penjualan yang dilakukan menggunakan cara memperoleh barang dari penyetok dan langsung dikirim ke konsumen.
2. Penjualan *stock warehouse* yaitu penjualan barang dari penyediaan yang telah ada pada *warehouse*.
3. Penjualan *combination* adalah kegiatan berjualan yang dilakukan dengan barang diambil sebagian penyetok dan sebagian dari *stock* yang tersedia di *warehouse*.

Dari beberapa jenis penjualan yang diuraikan sebelumnya, penulis dapat menyimpulkan bahwa PT Arista Mitra Lestari Cabang Pamanukan, termasuk jenis penjualan stock.

2.2.3 Definisi Produk Yamaha

Produk Yamaha merupakan salah satu produk kendaraan roda yang komponennya didatangkan langsung dari Jepang dan produk Yamaha juga bersaing dengan produk lainnya seperti : Honda, Suzuki, Kawasaki, dan lain-lain. Adapun teknologi Yamaha saat ini, yaitu teknologi *Blue Core* yang diterapkan pada kendaraan Yamaha bertipe matic, seperti : Mio Z, New Fino 125, Aerox 125LC, All New Soul GT, NMAX, dan Mio M3 125. Produk Yamaha yang lainnya yaitu kendaraan bertipe *Naked Bike*, seperti : New Vixion Advance, Xabre, All New Byson FI, dan MT-25. Produk Yamaha bertipe *Special Edition*. Produk Yamaha bertipe Moped, seperti : Vega

Force, MX King 150, Jupiter Z1. Produk Yamaha bertipe Sport, seperti : R25, dan R15[4].

2.2.4 Teori Perilaku (Selera) Konsumen

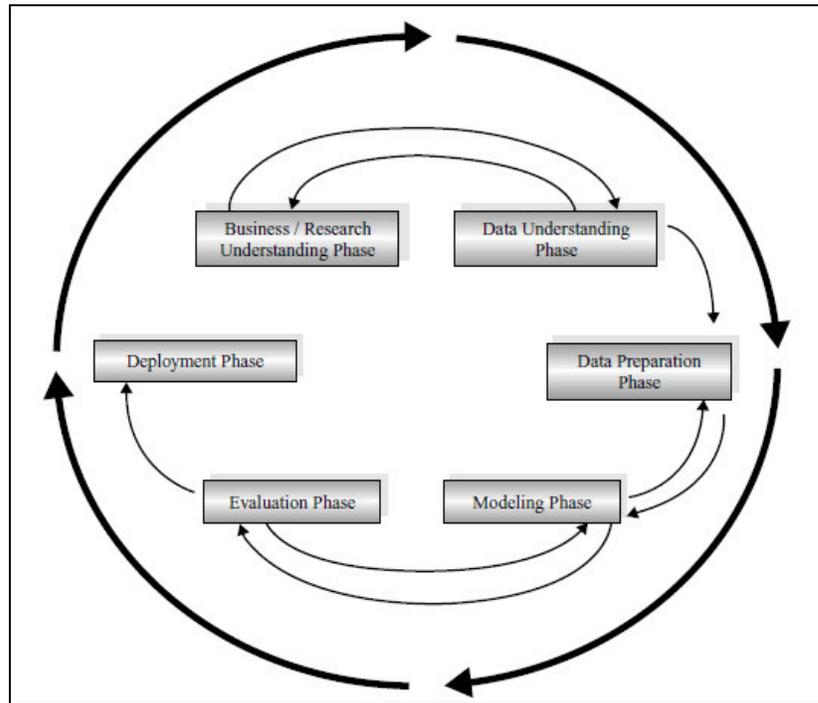
Perilaku permintaan konsumen terhadap barang dan jasa akan dipengaruhi oleh beberapa faktor, diantaranya: pendapatan, selera konsumen, dan harga barang, disaat kondisi yang lain tidak berubah (*ceteris paribus*). Perilaku konsumen ini didasarkan pada Teori Perilaku Konsumen yang menjelaskan bagaimana seseorang dengan pendapatan yang diperolehnya, dapat membeli berbagai barang dan jasa sehingga tercapai kepuasan tertentu sesuai dengan apa yang diharapkannya[17].

2.2.5 Definisi Data Mining

Data Mining yaitu sebuah *process* pengekstrasian *information* dari sekumpulan data yang memiliki jumlah besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data[5]. Data Mining adalah salah satu langkah menganalisis dari data perspektif yang beda dan menyimpulkan menjadi informasi penting yang dapat digunakan untuk peningkatan keuntungan, meminimalkan biaya pengeluaran, atau bisa juga semuanya[5]. Dari satu makna ini dapat disimpulkan bahwa Data Mining merupakan mode ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi-informasi yang nantinya dapat digunakan.

2.2.6 CRISP-DM

CRISP-DM (*Cross-Industry Standart Process for Data Mining*) adalah standar progress data mining sebagai penyelesaian masalah secara meluas dari *business* atau penelitian. Enam fase CRISP-DM[7].



Gambar 0.1 Mode CRISP-DM

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)
 - a. Menentukan penyeluruhan dalam suatu usaha dengan proyek yang dituju dan kebutuhan yang lengkap.
 - b. Penerjemahan masalah data mining berdasarkan perumusan yang dituju dengan pembatasan yang ada.
 - c. Persiapan yang diawali dengan taktis.
2. Fase Pemahaman Data (*Data Understanding Phase*)
 - a. Pengumpulan dataset.
 - b. Penggunaan dengan diselidikinya analis data dalam pengenalan dan mencari dengan diawali informasi.
 - c. Evaluasi dataset dengan akurat.
 - d. Keinginan sejumlah dengan akurat.
 - e. Keinginan sejumlah *group* dalam pencarian data dengan dikandungnya dari suatu masalah.

3. Fase Pengolahan Data (*Data Understanding Phase*)
 - a. Persiapan dari dataset awal, kumpulan dari dataset dengan menggunakan dalam sebagian besar informasi selanjutnya.
 - b. Memilih masalah dari atribut yang diinginkan dengan analisis disesuaikan analisis yang dituju.
 - c. Lakukan dengan mengubah beberapa *variable* bila diinginkan.
 - d. Persiapan dengan awlan dataset agar persiapan model yang dituju.
4. Fase Pemodelan (*Modelling Phase*)
 - a. Pemilihan dan pengaplikasian permodelan yang akurat.
 - b. *Calibration* dengan pemodelan untuk mengoptimalkan keberhasilan.
 - c. Butuh pemerhatian dengan sebgaiain teknik dalam penggunaan dengan masalah data mining yang ada.
 - d. Keperluan memproses yang diolah pada suatu dataset dengan diolahnya dataset yang di dalamnya bentuk yang disesuaikan kebutuhan teknik data mining.
5. Fase Evaluasi (*Evaluation Phase*)
 - a. Pengevaluasian sebagian besar pemodelan dalam pendapatan dengan penyebaran yang efektif dan kualitas dengan banyaknya suatu pemodelan dalam fase-fase untuk penggunaan evaluasi yang akurat.
 - b. Penetapan yang diawali dengan pemenuhan model yang dituju.
 - c. Penentuan masalah kepentingan bisnis dengan tidaknya penanganan masalah yang benar.
 - d. Pengambilan hal yang terkait dengan digunakan data mining dengan penghasilan yang benar.
6. Fase penyebaran (*Deployment Phase*)
 - a. Pemakaian dengan dihasilkannya pemodelan yang membentuk model tidak ditindakan telah menyelesaikan proyek.
 - b. Contoh penyebaran sederhana : Penyusunan laporan.

- c. Contoh penyebaran kompleks : Penerapan *process* data mining secara parallel pada perusahaan lain.

2.2.7 Metode Pelatihan Data Mining

Metode pelatihan yaitu cara berlangsungnya pembelajaran dan pelatihan dalam menganalisis dataset untuk mendapatkan pola data tertentu[7]. Metode pelatihan data mining memiliki 3 kelompok, seperti : *supervised learning*, *unsupervised learning*, dan *association learning*. 3 kelompok tersebut memiliki definisi, sebagai berikut[7].

1. *Supervised Learning*

Kumpulan record dari inputan yang digunakan dan telah diketahui output , dengan kata lain variable yang menjadi target telah ditentukan dalam *dataset* yang sedang dianalisis. Sebagian besar algoritma dalam kelompok tersebut terdiri dari : klasifikasi, estimasi, dan prediksi. Algoritma yang digunakan akan melakukan *process* pembelajaran yang berdasarkan *value* dari *variable* sasaran yang telah terasosiasi dengan *value* pada variabel *predictor*.

2. *Unsupervised Learning*

Pada metode tersebut data yang dianalisa diterapkan tanpa adanya guru serta pelatihan pada data lampau, dengan kata lain diartikan sebagai pencarian pola pada setiap atribut yang digunakan. Tidak termasuk penetapan atribut atau kelas pada sasaran. Contoh algoritma yang menerapkan metode *unsupervised learning* adalah Clustering.

3. *Association Learning*

Berbeda dengan dua ^{kelompok} yang terdapat di atas, pada mode ini mempunyai tujuan untuk mencari atribut yang muncul pada transaksi yang sama. Algoritma asosiasi biasanya berfungsi untuk mencari dan menganalisa transaksi belanja dengan konsep mencari produk yang dibeli secara bersamaan dalam satu transaksi yang sama. Algoritma yang digunakan dalam kelompok asosiasi adalah Apriori.

2.2.8 Jenis-jenis Variabel Data Mining

Variable dalam nilainya dapat dikelompokkan menjadi 4 jenis, sebagai berikut[6].

1. Nominal

Variable nominal mempunyai nilai yang terkandung di dalamnya berupa *symbol* dengan fungsi sebagai label atau pemberian nama. Tipe ini tidak dapat diurutkan maupun diukur jaraknya akan tetapi hanya dapat diuji persamaannya. Dengan mempunyai ciri khusus antara nilai nominal satu dengan nominal yang lainnya dan tidak mempunyai hubungan.

2. Ordinal

Variable tipe ini hampir sama seperti tipe nominal yaitu berupa *symbol*, tetapi antara satu *symbol* dengan *symbol* lainnya dapat diurutkan dan tidak dapat dijumlahkan

3. Interval

Varibel interval mempunyai sifat yang dapat diurutkan nilainya antar data satu dengan data yang lain dan dapat ditukar menggunakan ketetapan dan unit yang sama.

4. Rasio

Variable tipe rasio merupakan *variable* dengan nilai nol yang bersifat absolute. Nilai *variable* rasio ini digunakan sebagai bilangan *rill*. Dalam mode perhitungan aritmatika termasuk penjumlahan, pengurangan, pembagian, perkalian dan lain-lain. *Variable* ini dapat melakukan terhadap nilai data rasio.

2.2.9 Metode dalam Data Mining

Data mining mempunyai cara menganalisis *dataset* yang akan diteliti dan dapat dibagikan menjadi lima metode dalam menarik simpulan tentang pola data yang tersembunyi[6]. Berikut metode-metode yang terdapat didalam data mining.

1. Estimasi (*Estimation*)

Merupakan teknik untuk mendapatkan pola tersembunyi pada sebuah *dataset* dengan cara melihat sasaran *variable* kategori. Teknik ini sama seperti teknik klasifikasi, tetapi teknik estimasi cenderung difungsikan untuk tipe data *numeric* dan memiliki label. Metode estimasi digunakan pada pembelajaran *supervised learning*. Contoh algoritma yang dapat digunakan pada teknik ini adalah : *Linear Regression*, *Support Vector Machine*, dan *Neural Network*.

2. Klasifikasi (*Classification*)

Teknik klasifikasi berfungsi sebagai mengelompokkan pola data yang sama berdasarkan atribut yang dimiliki. Pada teknik tersebut data yang diolah cenderung menggunakan tipe data nominal, tetapi tidak menutup kemungkinan untuk pengolahan data dengan tipe data *numeric*. Teknik klasifikasi mempunyai sifat yaitu *supervised learning* yang memiliki arti penemuan pola yang baru memerlukan guru atau label sasaran. Contoh algoritma yang menggunakan teknik klasifikasi, diantaranya : *Naïve Bayes*, *C4.5*, *ID3*, *K-Nearest Neighbord*, *Linear Discriminant Analysis*, dan *CART*.

3. Prediksi (*Prediction*)

Teknik prediksi mempunyai kesamaan dengan teknik estimasi dan klasifikasi dalam melakukan analisis kumpulan data, tetapi tipe data yang digunakan adalah *numeric* baik dalam *variable* atau label. Ciri yang dimiliki teknik prediksi yaitu salah satu *variable* yang dimanfaatkan tipe data *time series*. Teknik ini juga termasuk pada kategori *supervised learning*. Algoritma yang menerapkan teknik prediksi, yaitu : *Linear Regretion*, *Support Vector Machine*, dan *Neural Network*.

4. Klastering (*Classtering*)

Teknik klastering sering disebut sebagai langkah *segmentation*, metode klastering berfungsi untuk mengidentifikasi kelompok *rill* dari kasus yang didasarkan pada satu kelompok atribut, dengan cara mengelompokkan data-data yang mempunyai kesamaan pada setiap atribut-atribut. Teknik tersebut mempunyai ciri khusus yaitu *dataset* yang digunakan tidak mempunyai label sasaran. Metode pembelajaran teknik klastering termasuk *unsupervised learning* yang dimana tidak

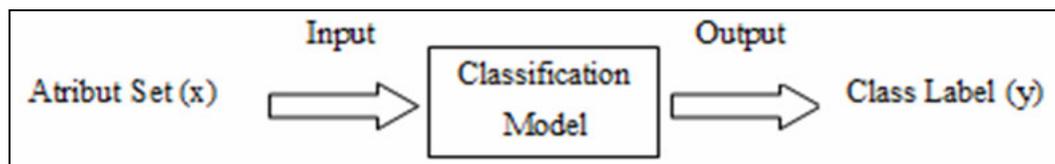
membutuhkan guru dalam menemukan sebuah pola. Algoritma yang menggunakan teknik ini, diantaranya : *K-Means*, *K-Medoids*, dan *Self-Organization Map* (SOM).

5. Asosiasi (*Assosiation*)

Teknik asosiasi biasa dikenal sebagai *association rule* berfungsi untuk menemukan relasi diantara item-item data serta menemukan atribut yang muncul bersamaan. Teknik ini menggunakan pembelajaran yang berbeda dari empat teknik diatas, karena menggunakan *association learning* yang bertujuan untuk menemukan pola item yang muncul pada transaksi yang sama. Algoritma yang menggunakan teknik tersebut, diantaranya : *Apriori*, dan *FP-Growth*.

2.2.10 Definisi Klasifikasi

Classification merupakan sebuah tugas yang terpenting dalam mode data mining yang mengorganisasikan serta mengelompokkan data ke dalam *class* yang berbeda merupakan salah satu tujuan utama dari klasifikasi. Definisi pengklasifikasian merupakan salah satu fungsi yang mempunyai sifat *prediction* dan mengelompokkan data sub tertentu ke dalam sebuah *class*. Mode pengklasifikasian diproses dari beberapa data dengan class yang telah ditentukan dan dikenal dengan ciri-ciri yang lampau. Tampilan pengklasifikasian biasanya dapat diukur dengan ketepatan[8].



Gambar 0.2 Blok Diagram Model Klasifikasi

2.2.11 Definisi Metode Naive Bayes

Algorithm Naive Bayes, sebuah metode *algorithm* yang ada pada metode data mining *classification*. Naive Bayes ialah teknik pengklasifikasian menggunakan teknik peluang dan statistik yang dikemukakan oleh ilmuwan Inggris yaitu Thomas Bayes. Naive Bayes digunakan untuk memprediksi peluang dimasa yang akan datang

berdasarkan pengalaman pada waktu lampau, sehingga disebut dengan Teorema Bayes. Teorema ini digabungkan dengan Naive yang mengasumsikan kondisi antar elemen saling bebas. *Classification* Naive Bayes diupayakan bahwa ada atau tidak ada ciri tertentu dari sebuah *class* tidak ada hubungannya dengan ciri dari kelas lainnya[8].

Persamaan dari Teorema Bayes yaitu :

$$P(H|X) = P(X|H) \cdot P(H) \div P(X) \quad (2.1)$$

Keterangan :

X :Data dengan class yang belum diketahui

H :Hipotesis data X merupakan suatu class spesifik

$P(H|X)$:Probabilitas hipotesis H berdasarkan kondisi X (posteriori probability)

$P(H)$:Probabilitas hipotesis H(prior probability)

$P(X|H)$: Probabilitas X berdasarkan kondisi hipotesis H

$P(X)$: Probabilitas X

Adapun alur dari metode Naive Bayes, sebagai berikut.

1. Membaca data training
2. Hitung Jumlah dan probabilitas, namun apabila data numerik maka.
 - a. Mencari *value* rata-rata dan standar deviasi dari tiap-tiap parameter yang merupakan data berangka.

Adapun kesamaan yang menggunakan dalam nilai rata-rata hitung (*mean*) dapat dijabarkan sebagai berikut.

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (2.2)$$

Atau (2.3)

$$\mu = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

dimana :

μ : rata-rata hitung (*mean*)

X_i : nilai sample ke-i

n : jumlah sample

Dan persamaan dalam hitung *value* simpangan baku (standar deviasi), sebagai berikut.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (2.4)$$

dimana :

σ : standar Deviasi

X_i : nilai x ke-i

μ : rata-rata hitung (*mean*)

n : jumla sample

- b. Mencari *value* probabilitas dengan langkah hitung jumlah data yang sesuai dari bilangan yang sama dibagi dengan jumlah data pada kategori tersebut.
3. Memperoleh *value* dalam tabel rata-rata, standar deviasi dan peluang.

2.2.12 Kelebihan Algoritma Naive Bayes

Classification Naive Bayes terdapat beberapa keuntungan/kelebihan, diantaranya sebagai berikut[2].

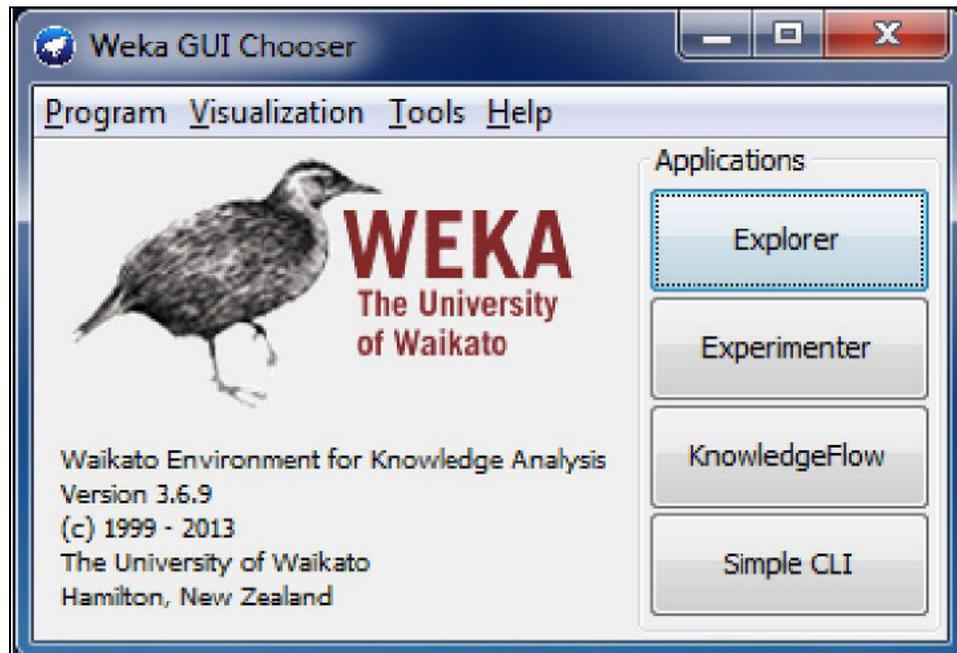
1. Mengolah data kuantitatif dan data diskrit
2. Kuat untuk titik noise yang disolasi, contohnya yang dirata-ratakan ketika diestimasi syarat probabilitas data
3. Membutuhkan jumlah data *training* kecil untuk diestimasi parameter (rata-rata dan variansi dari variabel) yang membutuhkan untuk pengklasifikasi
4. Mengatasi *value* yang hilang dengan melupakan instansi selama perhitungan peluang estimasi
5. Cepat dan ekonomis ruang
6. Kuat terhadap atribut yang tidak relevan

2.2.13 Perangkat Lunak yang Terkait pada Penelitian

1. Waikato Environment For Knowledge Analysis (WEKA)

Weka ialah *system* data mining yang dikembangkan oleh Universitas Waikato di Selandia Baru yang menjalankan *algorithm* data mining[9]. Weka ialah sebuah kumpulan dari *machine learning agorithm* dalam pekerjaan data mining. Algoritmanya mampu diaplikasikan secara serta-merta ke set data atau dipanggil dari *coding* Java itu sendiri. WEKA memuat berbagai alat untuk data pra-pengolahan (*pre-processing*), *classification*, regresi, *clustering*, aturan asosiasi, dan visualisasi. WEKA juga pantas untuk dikembangkan *scema* pelajaran perangkat baru[10]. WEKA mengadakan implementasi dari pembelajaran *algorithm* yang beroleh dengan mudah untuk menerapkan pada *dataset*. Penerapan mencakup berbagai peralatan untuk ubah *dataset*, *pre-process dataset*, menyerahkan *scema* pembelajaran, dan dianalisis *classification* yang dihasilkan dan kemampuannya tanpa diharuskan menulis kode program[11]. Salah satu pengguna WEKA ialah sebagai menerapkan cara pemrosesan untuk *dataset* dan menganalisis hasil/keluaran untuk data yang dipelajari secara berkelanjutan menerus. Aktor lainnya yaitu dipakai sebagai langkah pembelajaran bagi meramal pada masalah baru. Penerapan lainnya dilakukan pada beberapa

pembelajaran yang beda dan membandingkan cara kerja dari mereka dan memilih salah satu untuk dipergunakan dalam meramal. Pada *interface* utama dapat anda pilih metode pembelajaran yang diinginkan pada menu. Banyak metode yang mempunyai parameter yang senada, yang dapat diproses melewati lembar properti atau editor entitas. Sebuah modul penilaian umum dipakai untuk mengukur kinerja semua pengklasifikasian[12].



Gambar 0.3 Tampilan WEKA

Sebagaimana yang ditampilkan pada Gambar 2.3 WEKA GUI Chooser mempunyai empat tombol utama, yaitu :

a. *Explore*

Explore diartikan sebuah pilihan aspek untuk data yang dijelajahi dengan WEKA. *Explore* mempunyai 6 macam tab di halaman kerja dengan mempunyai tugas sebagai berikut[12].

- 1) *Preprocess* ialah aspek untuk memilih data set dan mengubahnya dengan bermacam cara.

- 2) *Classify* adalah pembelajaran *training* skema yang melakukan klasifikasi atau regresi dan evaluasinya.
- 3) *Cluster* yaitu pembelajaran *cluster* atau pengelompokan untuk *dataset*.
- 4) *Associate* ialah pembelajaran aturan berasosiasi untuk data dan ujiannya.
- 5) *Select attributes* merupakan bidang dalam memilih aspek yang paling relevan dalam *dataset*.
- 6) *Visualize* merupakan bagian tampilan plot dari dua dimensi yang beda dari data itu dan interaksinya.

b. *Experimenter*

Experimenter adalah sebuah pilihan bagian untuk melakukan pengujian dan melakukan mode statistik antara skema pembelajaran. *Experimenter* kemungkinan *user* untuk dibuat eksperimen dalam skala besar, mulai dari eksperimen yang diproses sampai eksperimen selesai sampai dilakukan analisis kinerja secara statistik terhadap apa yang telah diperoleh.

c. *Knowledge Flow*

Knowledge Flow mengartikan, sebuah pilihan bagian yang mendukung fungsi dasar yang cocok seperti menjelajahi namun dengan antarmuka *drag* dan *drop*. Salah satu kegunaannya adalah *knowledge flow* saling mendorong adanya pembelajaran tambahan. Pada *interface* awal *knowledge flow* ini user dapat melihat posisi sebuah kinerja dari mode yang melakukannya, *user* yang disambungkan kedalam sebuah graf berarah yang memproses dan menganalisis data. Pada bagian ini ialah langkah secara akurat bagaimana data berjalan dalam sistem dimana tidak disediakan dalam *explore*.

d. *Simple CLI*

Simple CLI ialah pilihan bagian yang memberikan tampilan garis perintah biasa yang memungkinkan adanya perintah langsung eksekusi dari WEKA untuk sistem operasi yang tidak memberikan tampilan garis perintah sendiri. Dibalik *interface* interaktif *explore*, *experimenter*, *knowledge flow* pada WEKA

mempunyai fungsi dasar yang dapat diakses secara langsung pada tampilan garis perintah. Garis perintah terdapat pada *simple CLI*, pada *interface* utama WEKA *tools simple CLI* terletak disebelah kanan bawah.

WEKA sebagai *machine learning* yang memiliki tugas dalam *user* salah satu langkah, WEKA mempunyai beberapa cara utama dalam pemersalahan *data mining*, yaitu regresi, klasifikasi, *clustering*, *association rule mining*, dan memilih atribut. Pengenalan data ialah bagian yang tidak terpisahkan dari sebuah pekerjaan, dan banyak fasilitas visualisasi data dan alat data *preprocessing* yang tersedia. Semua *algorithm* di WEKA mengambil masukan dalam bentuk tabel penghubung tunggal dengan format *Attribute Relation File Format (ARFF)*, yang dapat dibaca dari sebuah *file* atau menghasilkan dari permintaan basis data[12].

2. PHP (*Hypertext Preprocessor*)

PHP ialah bahasa server-side-scripting yang bersatu dengan HTML untuk menciptakan halaman web yang dinamis. Karena PHP merupakan server-side-scripting maka sintaks dan arahan PHP akan diesksekusi diserver lalu hasilnya akan dikirimkan ke browser dengan format HTML[13].

Untuk menciptakan halaman web, aktualnya PHP tidaklah bahasa pemrograman yang wajib digunakan. Kita bisa saja membuat website hanya menggunakan HTML saja. Web yang menghasilkan dengan HTML (dan CSS) ini disebut dengan website statis, dimana konten dan halaman web bersifat tetap.

3. MySQL

MySQL merupakan database yang mendukung berbagai macam pemrograman *script* untuk aplikasi *web* maupun aplikasi berbasis *desktop*. Sejarah MySQL pertama kali dikembangkan oleh perusahaan bernama MySQL AB yang merupakan perusahaan asal Swedia pada tahun 1994-1995. Pada mulanya sebelum menjadi MySQL AB perusahaan tersebut bernama TCX yang merupakan perusahaan yang bergerak dibidang konsultan *database* dan

pengembang perangkat lunak. Perkembangan saat ini MySQL dibawai atau diakuisisi oleh perusahaan Oracle. Kepopuleran yang dimiliki MySQL yaitu menggunakan bahasa SQL sebagai *language basic* untuk mengakses databasenya selain itu juga didukung dengan kinerja *query* yang cepat dan dapat digunakan oleh perusahaan skala menengah kecil serta bersifat *open source*[14].