BAB II
LANDASAN TEORI

2.1 Penelitian Terkait

Data pendukung dalam sebuah penelitian sangatlah penting untuk di jadikan sebagai acuan penyusunan penelitian, oleh karena itu dalam penelitian ini juga mengambil teori – teori dari peneitian sebelumnya.

Adapun penelitian sebelumnya yang terkait dengan penelitian ini antara lain yaitu :

Hera Wasiati dan Dwi Wijayanti dalam penelitiannya yang berjudul "Sistem Pendukung Keputusan Penentuan Kelayakan Calon Karyawan Tenaga Kerja Indonesia Menggunakan Metode *Naive Bayes* (Studi Kasus : PT. Karyatama Mitra Sejati Yogyakarta)" pada tahun 2014 menyatakan dalam penyeleksian tenaga kerja indonesia dengan metode naive bayes yang diharapkan mampu membantu staff dalam menentukan siapa yang layak diterima atau tidak. Dalam penyeleksiannnya ada beberapa kreteria yang digunakan yaitu : pendidikan, tinggi badan, nilai tes, usia, dan berat badan. Data yang digunakan dalam penelitian ini sebanyak 542 dengan 362 sebagai data training dan 180 sebagai data tes, akurasi polanya 73,89% dan *error*nya 26,11% jadi data yang tepat sebanyak 133 dan yang tidak tepat 47 (Wasiati & Wijayanti, 2014).

Alfa Saleh dalam penelitiannya yang berjudul "Implementasi Metode Klasifikasi *Naive Bayes* Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga" pada tahun 2015 menyatakan bahwa pentingnya peranan listrik membuat permintaan listrik meningkat pesat sehingga tidak *linier* dengan persediaan listrik, oleh karena itu setiap rumah tangga harus paham memprediksi kebutuhan listriknya. Pada penelitian ini metode naive bayes digunakan untuk memprediksi penggunaan listrik tiap rumah tangga, dari 60 data yang di uji memperoleh hasil sebesar 78,3333% untuk keakuratan

prediksi, dimana dari 60 data terdapat 47 data pengguna listrik rumah tangga yang berhasil diklasifikasikan dengan benar (Saleh, 2015).

Mujib Ridwan, hadi Suyono dan M. Sarosa dalam penelitiannya yang berjudul "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes *Classifier*" pada tahun 2013 menyatakan penelitian ini di fokuskan untuk mengevaluasi kinerja akademik mahasiswa pada tahun ke-2 dan diklasifikasikan dalam kategori mahasiswa yang daat lulus tepat waktu atau tidak. Input dari sistem ini adalah data induk mahasiswa dan data akademik mahasiswa, sampel mahasiswa angkatan 2005-2009 yang sudah dinyatkan lulus akan dijadikan sebagai data *training*dan *testing,* sedangkan dara angkatan 2010-2011 akan dijadikan target. Data tersebut akan diproses menggunakan metode *Naive Bayes Classifier.* Hasil pengujian menunjukkan bahwa faktor yang sangat berpengaruh dalam penentuan ini adalah IPK, IP semester 1, IP semester 4 dan jenis kelamin. Pengujian pada data mahasiswa angkatan 2005-2009 mempunyai akurasi 83%, 50% dan 70% berturut-turut (Ridwan, Suyono, & Saroso, 2013).

Aris Nugroho dan Subanar dalam penelitiannya yang berjudul "Klasifikasi *Naive Bayes* untuk Prediksi Kelahiran pada Data Ibu Hamil" pada tahun 2013 menyatakan dd\engan model pendekatan Bayesian berupa Klasifikasi *Naive Bayes* dengan HMAP (Hipotesis Maksimum A Posteriori) dipakai memprediksi kelahiran yang akan dialami ibu hamil dengan karakteristik Usia ibu, Tinggi Badan, Jumlah Hb, Tekanan Darah, Riwayat Kehamilan lalu dan Penyakit bawaan. Semua data didiskritkan berdasar batasan yang dipakai Departemen Kesehatan dan hasil prediksi berupa probabilitas terjadinya resiko, bisa dipakai sebagai rujukan tempat melahirkan ataupun penilaian kinerja dari penyelenggara jasa persalinan. Dengan fungsi klasifNB dalam bahasa R, fase Training untuk estimasi maksimum likelihood dan sesuai dengan karakteristik ibu hamil, aplikasi

menjadi dinamis melakukan prediksi sesuai wilayah dipilih (Nugroho & Subanar, 2013).

Sukmawati Anggraeni Putri dalam penelitiannya yang berjudul "Kombinasi Integrasi Metode *Sampling* dengan *NaiveBayes* Untuk Ketidakseimbangan Kelas Pada Prediksi Cacat Perangkat Lunak" pada tahun 2015 menyatakan bahwa penerapan teknik sampling terutama pada teknik SMOTE dan *Resample*terbukti dapat meningkatkan kinerja penklasifikasi *Naive Bayes* (Putri, 2015).

Intan Cahya Gumilang, Sudjalwo dan Aris Rakhmadi dalam penelitiannya yang berjudul "Prediksi Persediaan Obat Dengan Metode *Naive Bayes* (Studi Kasus: Apotek Saputra)" menyatakan pada apotek saputro dalam pengolahan data obatnya masih menggunakan sistem manual dan belum dapat memprediksi stok obat, penggunaan metode *interpolasi linier* yang nantinya digunakan untuk menghitung prediksi stok obat yang terjual, sedangkan metode *naive bayes* digunakan untuk menghitung peluang. Untuk stok obat Alleron yang terjual pada bulan April 2013 adalah 21. Hasil dari perhitungan prediksi dengan metode interpolasi linieruntuk obat Alleron pada bulan April 2013 adalah 26 dengan error 0,23, dan hasil perhitungan peluang dengan menggunakan metode naïve bayesuntuk obat Alleron di bulan April 2013 adalah 0,37 (Gumilang, Sudjalwo, & Rakhmadi, 2014).

2.2 Data Mining

Data mining dapat diartikan sebagai proses penemuan pengetahuan yang bermanfaat dan menarik di dalam kumpulan data yang besar, tujuan utama data mining yaitu prediksi (*prediction*) dan uraian (*description*). Data mining juga mempunyai beberapa tugas utama yaitu *classification* (klasifikasi), *regression* (regresi), *clustring* (pengelompokan),*summarization* (ringkasan), *dependencymodeling* (pemodelan ketergantungan), *changeanddeviationdetection* (pendeteksian perubahan dan deviasi) (Nurani, Susanto, & Proboyekti, 2007).

Data mining juga dapat diartikan sebagai rangkaian kegiatan untuk menemukan pola yang menarik dari data dalam jumlah besar, data – data tersebut dapat disimpan dalam *database,data warehouse* atau penyimpanan informasi lainnya. Ilmu – ilmu yang berkaitan dengan data mining diantaranya adalah *databasesystem*, *datawarehouse*, *statistik*, *meachinelearning*, *informationretrieval* dan koputasi tingkat tinggi. Data mining juga didukung oleh ilmu lain seperti spatial data analysis, signal processing, neural network dan pengenalan pola (Meilani & Susanti, 2014).

Menurut Han dan Kamber alasan utama mengapa data mining diperlukan adalah karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan *knowledge* yang berguna (Junanto, 2013).

2.3 Teknik Data Mining

Dalam data mining terdapat beberapa teknik yang dipakai yaitu (Hermawati, 2013):

2.3.1 *Classification*

Klasifikasi adalah menentukan *record*data baru kesalah satu dari beberapa kategori (klas) yang telah didefinisikan sebelumnya. Biasanya hal ini disebut juga dengan *supervised learning*.

Mujib Ridwan, hadi Suyono dan M. Sarosa dalam penelitiannya yang berjudul "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes *Classifier*" pada tahun 2013 menyatakan penelitian ini di fokuskan untuk mengevaluasi kinerja akademik mahasiswa pada tahun ke-2 dan diklasifikasikan dalam kategori mahasiswa yang daat lulus tepat waktu atau tidak. Input dari sistem ini adalah data induk mahasiswa dan data akademik mahasiswa, sampel mahasiswa angkatan 2005-2009 yang sudah dinyatkan lulus akan dijadikan sebagai data *training*dan *testing,* sedangkan dara

angkatan 2010-2011 akan dijadikan target. Data tersebut akan diproses menggunakan metode *Naive Bayes Classifier.* Hasil pengujian menunjukkan bahwa faktor yang sangat berpengaruh dalam penentuan ini adalah IPK, IP semester 1, IP semester 4 dan jenis kelamin. Pengujian pada data mahasiswa angkatan 2005-2009 mempunyai akurasi 83%, 50% dan 70% berturut-turut (Ridwan, Suyono, & Saroso, 2013).

2.3.2 *Clustering*

Klasterisasi adalah membuat beberapa sub-set atau kelompokdari data-set yang telah tersedia sedemikian rupa sehingga elemen – elemen dari suatu kelompok tertentu memiliki set properti yang di *share* bersama, di tingkat similiaritas yang tinggi dalam satu kelompok dan tingkat similiaritas antar kelompok yang rendah. Disebut *unsupervised learning.*

Dari penelitian sebelumnya pada teknik data mining clustering yang berjudul "Analisa Perbandinganmetode *Hierarchical Clustering,* K-Means Dan Gabungan Keduanya Dalam *Cluster* Data (Studi Kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS)" oleh Tahta Alfina mengatakan bahwa dalam studi kasusnya tersebut metode yang cocok digunakan dalam *cluster* data adalah kombinasi algoritma *Hierarchical Clustering* danK-Means menghasilkan pengelompokan data yang lebih baik jika dibandingkan dengan k-means dalam semua pengujian (Alfina, Santosa, & Barakbah, 2012).

2.3.3 *AssociationRule Discovery*

Mendeteksi kumbulan atribut – atribut yang muncul bersamaan dalam frekuensi yang sering dan membentuk sejumlah kaida dari kumpulan – kumpulan tersebut.

Pada penelitian "Data Mining Untuk Analisa Tingkat Kejahatan Jalanan Dengan Algoritma *Association Rul*e Metode Apriori" oleh Fadlina bahwa hasil dari perancangan data mining dengan algoritma apriori ini diperoleh informasi yang dibutuhkan oleh pihak kepolisian berupa prosesntase yang digunakan oleh bagian reskrim untuk mengetahui kejahatan jalanan apa yang sering terjadi sehingga persoalan tingkat kejahatan jalanan dapat diminimalisasi (Fadlina, 2014)

2.3.4 *SequentialPattern Discovery*

Mencari sejumlah *event*yang secara umum terjadi bersama – sama.

Pada penelitian "Penentuan Pola Sekuensial Pada Data Transaksi Perpustakaan IPB Menggunakan Algoritma *Graph Search Techniques*" oleh Imam S Sitanggang bahwa berdasarkan pola sekuensial yang diperoleh dalam penelitiannya maka dapat disimpulkan bahwa minimum support tertinggi hingga masih terbentuk *large sequence* berada pada nilai 30% dan time constraint 6 bulan dengan transaksi pinjaman terbanyak dilakukan oleh mahasiswa yang berasal dari departemen pemuliaan tanaman dan teknologi benih sebanyak 209 transaksi (Sitanggang, Ardiansyah, & Agung).

2.3.5 *Regression*

Regresi meperediksi nilai dari suatu *variable* kontinyu yang diberikan berdasarkan nilai dari *variable*yang lain, dengan mengansumsikan sebuah model ketergantungan *linier* atau *nonlinier*.

Pada penelitian yang berjudul "Penerapan Metode Generalized Ridge Regression Dalam Mengatasi Masalah Multikolinearitas" oleh Ni Ketut Utami mengatakan bahwa pada

data yang mengalami masalah multikolinearitas, metode kuadrat terkecil (*Ordinary Least Square*) tidak dapat melakukan pendugaan koefisien regresi dengan tepat. Metode *generalized Ridge Regression* merupakan salah satu metode alternatif yang dapat megatasi masalah multikolinearitas dengan sangat baik, dibuktikan dari nilai VIF dari masing – masing peubah bebas yang lebih kecil dari 5 (Utami, Sukarsa, & Kencana, 2013).

2.4 Tahap – Tahap Data Mining

Ada beberapa tahapan dalam data mining, yaitu (Meilani & Susanti, 2014):

1. Pembersihan Data (*Data Cleaning*)

   Pembersihan data merupakan langkah menghilangkan *noise*dan data yang tidak konsisten atau data yang tidak relevan.pada umumnya data yang diperole, baik dari database auatu perusahaan maupn 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. Permbersihan data juga akan mempengaruhi performasi dari teknik data mining karena data yang ditangani akan berkurang jumlah 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 teks. Integrasi data dlakukan pada atribut – atribut yang mengidentifikasikan entitas – entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi

data perlu dilakukan secara cermat karena kesalaha 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.*

4. Tranformasi Data (*Data Transformation*)

   Data di ubah 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 motode 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*)

   Mengidentifikasi pola –pola menarik ke dalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data

mining berupa pola – pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik 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 ini data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Adakalanya hal ini harus melibatkan orang – orang yang tidak paham data mining. Karenanya hasil presentasi data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisai juga bisa membantu mengkomunikasikan hasil data mining.

2.5 Klasifikasi

Klasifikasimerupakan proses untuk menemukan fungsi dan model yang dapat membedakan atau menjelaskan konsep atau kelas data dengan tujuan memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam proses pengklasifikasian biasa terdapat dua proses yang harus dilakukan, yaiut (Nugroho & Subanar, 2013) :

1. Proses Training

Pada proses ini akan digunakan data training set atau data sampel yang telah diketahui label – label atau atribut dari data sampel tersebut untuk membangun model.

2. Proses Testing

   Pada proses testing ini dilakukan untuk mengetahui keakuratan model yang telah dibuat pda proses training maka dibangun data yang disebut dengan data testing untuk mengklasifikasi label – labelnya.

Klasifikasi merupakan penempatan objek – objek kesalah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Klasifikasi sekarang ini telah banyak digunakan dalam berbagai aplikasi, sebagai contoh pendeteksian pesan email, spam berdasarkan header dan isi atau mengklasifikasikan galaksi berdasarkan bentuk – bentuknya. Pada proses klasifikasi data yang diinputkan adalah data record atau data sampel. Pada setiap record dikenal sebagai instance atau atau contoh yang ditentukan oleh sebuah tuple (x,y). Dimana x adalah himpunan atribut dan y adalah atribut tertentu yang menyatakan sebagai label class (Nugroho & Subanar, 2013).

2.6 Naive Bayes

Naive Bayes merupakan sebuah pengklasifikasian probalistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang diberikan. Algoritma menggunakan teorema bayes dan mengansumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. *Naive Bayes* juga didefinisikan sebagai pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuan inggis Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Saleh, 2015).

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama

adalah produk dari probabilitas individu. Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan *estimasi* parameter yang diperlukan dalam proses pengklasifikasian. *Naive Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang komples dari pada yang diharapkan (Saleh, 2015).

Persamaan dari teorema *Bayes*dapat dilihat di bawah ini :

$$P(H|X) = \frac{P(X|H).P(H)}{P(H)} \quad \text{.....(2.1)}$$

Dimana :

X : data dengan *class*yang belum diketahui

H : hipotesis data menggunakan suatu *class* spesifik

P(H|X) : probabilitas hipotesis H berdasar kondisi X (*parteriori* probabilitas)

P(H) : probabilitas hipotesis H (prior probabilitas)

P(X|H) : probabilitas X bedasarkan kondisi pada hipotesis H

P(X) : probabilitas H

Untuk menjelaskan metode *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang di analisis tersebut. Karena itu, metode *Naive Bayes*di atas disesuaikan sebagai berikut (Saleh, 2015) :

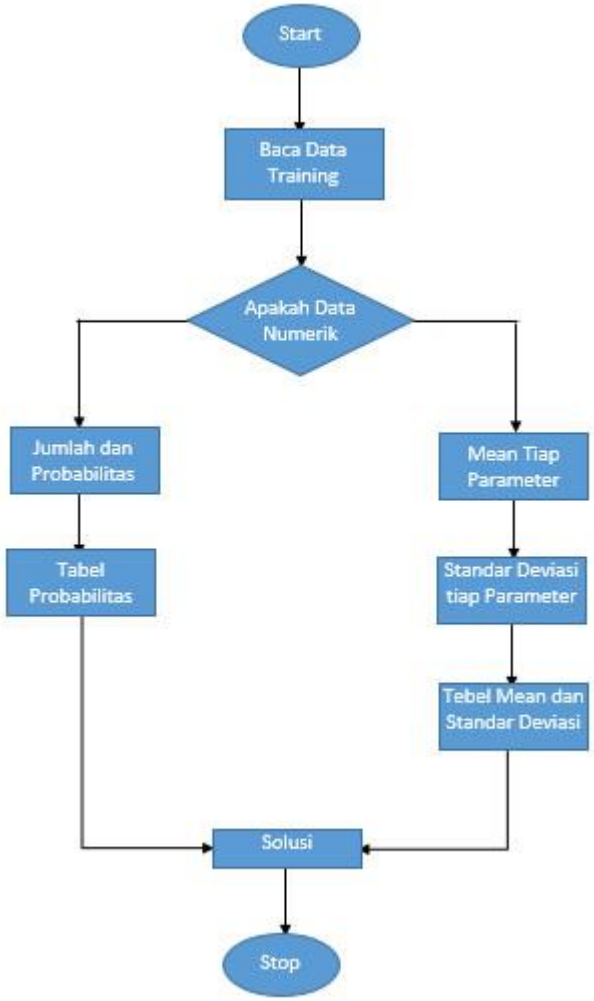
$$P(C|F1 \ldots . Fn) = \frac{P(C)P(F1 \ldots Fn|C)}{P(F1 \ldots Fn)} \quad \text{.....(2.2)}$$

Di mana Variabel C mempresentasikan kelas, sementara variabel F1...Fn mempresentasikan karakteristik petunjuk yang dibutuhkan untuk menentukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik – karakteristik sampel pada kelas C (disebut *likelihood*), dibagi dengan peluang kemunculan karakteristik – karakteristik secara global (disebut juga *evidence*). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut (Saleh, 2015):

$$posterior = \frac{prior\ x\ likelihood}{evidence} \qquad \text{.....(2.3)}$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *Posterior* tersebut nantinya akan dibandingkan dengan nilai – nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan.

Alur metode *Naive Bayes* dapat digambarkan sebagai berikut :



Gambar 2.1 Alur Metode Naive Bayes

Adapun penjelasan dari Gambar 1 adalah sebagai berikut (Saleh, 2015):

1. Baca data *trining*
2. Hitung jumlah dan probabilitas, namun apabila data numerik maka :
   a. Cari nilai *mean* dan standar deviasi dari masing –masing parameter yang merupakan data numerik.

      Adapun persamaan yang digunakan untuk menghitung nilai rata – rata (*mean*) dapat dilihat sebagai berikut :

      $$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{.....(2.4)}$$

      Atau

$$\mu = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} \quad \text{.....(2.5)}$$

Di mana :

μ : rata – rata hitung (*mean*)

x_i : nilai sampel ke-i

n : jumlah sampel

Dan persamaan untuk menghitung nilai simpangan baku (standar deviasi) dapat dilihat sebagai berikut :

$$\sigma = \sqrt{\frac{\sum_{x_i}^{n}(x_i - \mu)^2}{n-1}} \quad \text{.....(2.6)}$$

Dimana :

σ : standar deviasi

x_i : nilai x ke- i

μ : rata – rata hitung

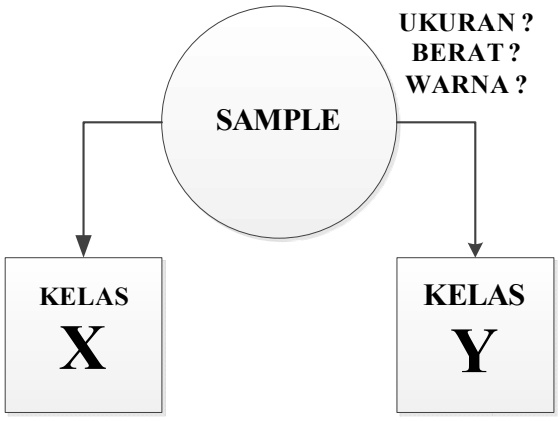
n : jumlah sampel

b. Cari nilai probabilistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.

3. Mendapatkan nilaidalam tabel *mean,* standar deviasi dan probabilitas.
4. Solusi yang dihasilkan

2.6.1 Contoh kasus naive bayes

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu obyek (Agus Mulyanto 2009). Oleh karena itu, kelas yang ada tentulah lebih dari satu. Penentuan kelas dari suatu dokumen dilakukan dengan cara membandingkan nilai probabilitas suatu sampel berada di kelas yang satu dengan nilai probabilitas suatu sampel berada di kelas yang lain.



Gambar 2.2 Ilustrasi contoh proses klasifikasi

Dengan persamaan teorema *Naïve Bayes* yang telah diturunkan di subbab A, kita mendapatkan nilai P(C|F1...F), yaitu nilai peluang suatu sampel dengan karakteristik Fn1...F berada dalam kelas C, atau dikenal dengan istilah *Posterior*.

Umumnya kelas yang ada tidak hanya satu, melainkan lebih dari satu. Sebagai contoh, ahli statistik ingin mengklasifikasikan sampel kucing ke dalam jenis kelaminnya. Oleh karena itu, terdapat dua kelas yaitu jantan dan betina.

Suatu sampel kucing akan diklasifikasikan ke dalam satu kelas saja, entah itu jantan atau betina, dengan melihat petunjuk-petunjuk yang ada (misalnya berat badan, panjang ekor, dll). Penentuan kelas yang cocok bagi suatu sampel dilakukan dengan cara membandingkan nilai *Posterior* untuk masing-masing kelas, dan mengambil kelas dengan nilai *Posterior* yang tinggi. Secara matematis klasifikasi dirumuskan sebagai berikut:

$$C_{NB} = argmax_{c \in C}\ P(c) \prod_{i=1}^{n} P(f_i|c)$$

dengan c yaitu variabel kelas yang tergabung dalam suatu himpunan kelas C. Dapat dilihat bahwa rumusan di atas tidak memuat

nilai *Evidence* (Z). Hal ini disebabkan karena *evidence*memiliki nilai yang positif dan tetap untuk semua kelas sehingga tidak mempengaruhi perbandingan nilai *Posterior*. Karena itu, faktor Z ini dapat dihilangkan. Perlu menjadi perhatian pula bahwa metoda *Naïve Bayes classifier* ini dapat digunakan bila sebelumnya telah tersedia data yang dijadikan acuan untuk melakukan klasifikasi.

Sebagai contoh, terdapat dua kelompok merek sepatu (X dan Y), dimana terdapat 3 petunjuk yang digunakan misalnya warna sepatu (merah, hitam), bahan sepatu (kulit, sintetis) dan model sepatu (Tali, Velkro).

Tabel1.1 Contoh Data Klasifikasi Metoda Naïve Bayes Classifier

Warna	**Bahan**	**Model**	**Jenis**
Merah	Kulit	Tali	X
Hitam	Kulit	Tali	X
Merah	Sintetis	Velkro	Y
Hitam	Kulit	Velkro	Y
Hitam	Sintetis	Tali	Y
Hitam	Sintetis	Velkro	X

Bila terdapat sampel sepatu Hitam, Sintetis, Tali (tidak ada pada data di atas), klasifikasi dapat dilakukan dengan menggunakan *Naïve Bayes classifier.*

Pertama-tama harus dicari terlebih dahulu *Posterior* X dan Y untuk sampel tersebut.

P(X) = 3/6 = 0.5 P(Y) = 0.5

P(Hitam|X) = 2/3 = 0.66 P(Hitam|Y)=1/3=0.33

P(Sintetis|X)=1/3=0.33 P(Sintetis|Y)=2/3=0.66

P(Tali|X)=2/3=0.66 P(Tali|Y)=1/3=0.33

Posterior X = P(X) P(Hitam|X) P(Sintetis|X) P(Tali|X)

= 0.5 x 0.66 x 0.33 x 0.66 = 0.072

Posterior Y = P(Y) P(Hitam|Y) P(Sintetis|Y) P(Tali|Y)
= 0.5 x 0.33 x 0.66 x 0.33 = 0.034

Karena *Posterior X* > *Posterior Y,*
maka sampel sepatu tersebut bermerek X.

2.7 Donor Darah

Darah adalah cairan yang terdapat pada semua makhluk hidup kecuali tumbuhan yang berfungsi mengirimkan zat – zat dan oksigen yang dibutuhkan oleh jaringan tubuh, mengangkut bahan – bahan kimia hasil metabolisme dan juga sebagai pertahanan tubuh terhadap virus atau bakteri. Donor adalah memberikan jaringan hidup agar dapat digunakan pada tubuh lain untuk tujuan bertahan hidup. Sedangkan donor darah atau *transfusi* darah adalah segala macam tindakan atau kegiatan kesehatan untuk menghasilkan penggunaan darah dengan cara khusus yang kemudian darah tersebut disumbangkan atau diserahkan kepada pasien yang membutuhkannya melalui pelayanan kesehatan dengan tujuan pengobatan dan pemulihan kesehatan terhadap pasien (U, 2010).

Dalam penentuan calon pendonor darah terdapat beberapa anjuaran yang disarankan oleh pihak penyedia layanan kesehatan donor darah, berikut adalah beberapa anjuran dari layanan kesehatan PMI Kabupaten Demak kepada calon pendonor darah :

1. Perhatikan waktu istirahat, tidur minimal 4 jam sehari sebelum donor dilakukan
2. Mengkonsumsi cairan lebih banyak sebelum dan pada hari donor darah
3. Makan dan minum minimal 2 jam sebelum donor darah dilakukan
4. Tambah ataupun makanan yang banyak mengandung zat besi seperti kacang – kacangan,daging merah dan sayuran hijau.

Sedangkan syarat dan ketentuan untuk calon pendonor darah adalah sebagai berikut (S, Soebandi, R, & S, 2012):

1. Usia pendonor antara 17 – 60 Tahun
2. Berat badan minimal 45 kg
3. Tekanan darah 110/70 sampai 160/100 mmHg
4. Tekanan darah *sistole*100 – 160 mmHg, *diastole* 60 – 100 mmHg
5. Kadar hemoglobin (hb) minimal 12,5 gr/dl dan maksimal 17,5 gr/dl
6. Nadi 50 – 100 kali/menit
7. Kesehatan baik, tidak mempunyai riwayat penyakit dalam dan menular
8. Tidak mengkonsumsi narkoba dan alkohol
9. Tidak demam
10. Kulit lengan sehat
11. Tidak menerima transfusi darah dalam jangka waktu 6 bulan terakhir
12. Tidak mendapat imunisasi dalam 4 minggu terakhir
13. Interval donor darah minimal tiga bulan
14. Wanita tidak sedang haid, tidak hamil dan tidak menyusui.

Apabila semua syarat – syarat tersebut dapat dipenuhi oleh calon pendonor maka proses donor darah akan dilakukan oleh petugas, tetapi apabila calon pendonor darah tidak bisa memenuhi syarat di atas maka proses donor darah tidak dapa dilanjutkan karena akan membehayakan penerima darah tersebut.

2.8 Pengolahan Atribut Data

Pada tahap ini akan dilakukan pengolahan atribut data yang telah diperoleh dari PMI Kabupaten Demak untuk digunakan sebagai data latih penentuan calon pendonor darah.

2.8.1 Pengelompokan Data

Sebelumnya pada tahap awal dari penelitian ini telah dilakukan pengelompokan variabel berdasarkan klasifikasi calon pendonor darah variabel tersebut adalah data *diskrit* dan data *kontinu.* Dari dataset yang telah diperoleh terdapat 2 data *diskrit* dan 4 data *kontinu,* sebagai berikut :

Data *Diskrit*:

- Jenis Kelamin
- Status Pendonor Darah

Data *Kontinu :*

- Usia (th)
- Berat Badan (kg)
- Kadar Hemoglobin (gr/dl)
- Tekanan Darah (mmHg)

2.8.2 Pengelompokan Nilai Atribut

Dataset yang telah dikelompokkan berdasarkan variabelnya menjadi data *diskrit* dan data *kontinu,* selanjutnya data tersebut akan dihitung nilai *mean* dan nilai *standart deviasi* dari data *kontinu*. Sebelum menentukan nilai *mean* dan *standartdeviasi* terlebih dahulu akan ditentukan ambang batas dari setiap atribut untuk inputan sistem. Berikut tabel ambang batas oleh PMI Kabupten Demak dalam klasifikasi calon pendonor darah, yaitu :

Tabel2.2 Ambang Batas Usia

Muda	Dewasa	Tua
18 s/d 30 Tahun	31 s/d 60 Tahun	> 60 Tahun

Tabel2.3 Ambang Batas Berat Badan

Kurus	Normal	Gemuk
< 45 Kg	45 s/d 65 Kg	> 65 Kg

Tabel2.4 Ambang Batas Tekanan Darah

Rendah	Normal	Tinggi
< 110/70 mmHg	110/70 - 160/100 mmHg	> 160/100 mmHg

Tabel2.5 Ambang Batas Kadar Hemoglobin

Rendah	Normal	Tinggi
< 12,5 gr/dl	12,5 s/d 17,5 gr/dl	> 17,5 gr/dl

2.9 Java

Java adalah bahasa pemrograman dan platform komputasi pertama kali yang dirilis oleh *Sun Microsystem* pada tahun 1995. Java merupakan teknologi yang mendasari kekuatan program untuk *utilitas,* permainan dan aplikasi bisnis. Java berjalan pada lebih dari 850 juta komputer pribadi di seluruh dunia, dan ada miliaran perangkat di seluruh dunia, termasuk ponsel dan perangkat televisi.

Salah satu karakteristik java adapah *portabilitas,* yang artinya bahwa program komputer yang ditulis dalam java harus dijalankan secara sama, pada setiap *hardware / platform* sistem operasi. Hal ini dicapai dengan menyusun kode bahasa java ke sebuah java *bytecode.* Pengguna aplikasi biasanya menggunakan java Runtime Environment (JRE) diinstal pada mesin mereka sendiri untuk menjalankan aplikasi java atau dalam browser web untuk applet java.

Untuk pembuatan dan pengembangan aplikasi berbasis java diperlukan Java Development Kit (JDK), dimana saat ini pemilik lisensi dari JDK adalah Oracle Corporation ang telah resmi mengakuisisi Sun Microsystem pada awal tahun 2010.

2.10 *MySQL*

MySQL (*My Structured Query Language*) merupakan perangkat lunak sistem basis data, perangkat lunak ini juga sering disebut DBMS (*Database Management System*). Namun berbeda dengan basis data konvensional seperti dbf, dat, dan mdb, *MySQL* sendiri memiliki beberapa

kelebihan diantaranya yaitu *multi user,* bersifat *multithread* serta mendukung sistem jaringan (Prihartanto, 2102).

MySQL adalah program database yang mampu mengirim dan menerima data dengan sangat cepat dan *multiuser*.*MySQL* memiliki dua bentuk lisensi, yaitu *free software* dan *sharewere* . *MySQL* yang *free software* bebas digunakan untuk keperluan pribadi atau usaha tanpa harus membaar atau membeli lisensi, yang berada di bawah lisensi GNU/GPL.(Muhammad Luqman,2012)

2.11 Kerangka Pemikiran

Kerangka pemikiran merupakan garis besar dari langkah – langkah penelitian yang dilakukan. Langkah – langkah tersebut disusun sedemikian rupa sebagai acuan untuk tahap – tahap yang dilakukan dalam proses penelitian.

Tabel2.6 Kerangka Pemikiran

Permasalahan

<table>
<tr><td colspan="3">Dalam menentukan calon pendonor darah pada PMI Kabupaten Demak masih menggunakan cara manual sehingga cara kerjanya dirasa kurang efektif</td></tr>
<tr><td colspan="3">Tujuan</td></tr>
<tr><td colspan="3">Memepermudah dan mempercepat dalam menentukan calon pendonor darah khususnya pada PMI Kabupaten Demak dengan akurasi tinggi</td></tr>
<tr><td colspan="3">Eksperimen</td></tr>
<tr><td>Inputan</td><td>Metode</td><td>Implementasi</td></tr>
<tr><td>Data pendonor darah sebelumnya pada PMI Kabupaten Demak</td><td>Naive Bayes Classifier</td><td>Java Netbeans</td></tr>
<tr><td colspan="3">Hasil</td></tr>
<tr><td colspan="3">Menghasilkan sebuah sistem yang akan mengklasifikasi dan memprediksi calon pendonor darah layak atau tidak</td></tr>
<tr><td colspan="3">Manfaat</td></tr>
<tr><td colspan="3">Membantu mengurangi kesalahan dalam menentukan calon pendonor darah dan mengurangi resiko tertularnya penyakit menular melalui donor darah dan juga mempermudah petugas dalam menentukan calon pendonor darah</td></tr>
</table>