APLIKASI DATA MINING MENGGUNAKAN NAIVE BAYES CLASSIFIER UNTUK PERSETUJUAN PENGAJUAN KREDIT

Agung Kurniadi Hernawan[1)], Selvia Lorena Br G, M.T.[2)]
[1), 2)]Jurusan Sistem Komputer Universitas Komputer Indonesia Bandung
[1]agung13_kun@yahoo.com, [2] selvia@yahoo.com

ABSTRAK

Data mining adalah sebuah teknik yang memanfaatkan data dalam jumlah yang besar untuk memperoleh informasi berharga yang sebelumnya tidak di ketahui dan dapat dimanfaatkan untuk pengambilan keputusan penting. Pada penelitian ini penulis berusaha menambang data (data mining) nasabah sebuah perusahaan bank negeri yang ada di Bandung untuk mengetahui apakah pengajuan kredit yang diajukan nasabah layak atau tidak layak. Data yang ada dianalisis menggunakan Naïve Bayes Classifier. Naïve Bayes merupakan salah satu metode pada probabilistic reasoning. Naïve Bayes Classifier bertujuan untuk melakukan klasifikasi data pada kelas tertentu, kemudian pola tersebut dapat digunakan untuk memperkirakan kriteria nasabah kredit, sehingga perusahaan bisa mengambil keputusan menerima atau menolak pengajuan kredit nasabah tersebut.

Kata Kunci: Pengajuan Kredit, Algoritma Naïve Bayes, Data Mining, klasifikasi

1. PENDAHULUAN

Bank adalah lembaga keuangan yang berfungsi sebagai perantara keuangan antara pihak yang kelebihan dana *(surplus unit)* dan pihak yang kekurangan dana *(deficit unit)*. Melalui bank kelebihan dana tersebut dapat disalurkan kepada pihak-pihak yang memerlukan dan memberikan manfaat bagi kedua belah pihak. Bank menerima simpanan uang dari masyarakat (dana pihak ketiga) dan meyalurkannya kembali dalam bentuk kredit. Oleh karena itu, peranan bank tidak lepas dari kegiatannya dalam pengaturan lalu lintas pembayaran dari waktu ke waktu, bahkan disetiap saat dikala bank itu beroperasi. Dari aktifitas bank tersebut tersalurlah berbagai produk bank sesuai dengan kebijakan kebijakan yang ditetapkan oleh bank yang bersangkutan.

Permasalahan yang sering timbul dalam perusahaan bank yang memberikan jasa pemberian kredit adalah banyaknya nasabah yang menunggak dalam membayar cicilan kreditnya, sehingga diperlukan sebuah sistem yang dapat mengklasifikasikan nasabah mana yang masuk kedalam kriteria yang pengajuan kredit layak dan kriteria yang pengajuan kreditnya tidak layak, sehingga perusahaan dapat mengatasi permasalahan tersebut sejak dini.

Sebuah bank pastilah mempunyai data yang begitu besar. Banyak yang belum

menyadari bahwa dari pengolahan data – data tersebut dapat memberikan informasi yang bermanfaat berupa klasifikasi data nasabah yang akan bergabung pada perusahaan itu sendiri. Penggunaan teknik data mining diharapkan mampu memberikan informasi yang berguna tentang teknik klasifikasi data nasabah yang akan bergabung dalam kelompok kriteria yang kreditnya layak dan kriteria yang kreditnya tidak layak. Dengan menggunakan teknik data mining ini juga diharapkan dapat memberi gambaran mengenai target pemasaran kreditnya di masa yang akan datang yang pada akhirnya dapat menekan biaya operasional *marketing* seminimal mungkin.

Klasifikasi data adalah suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan objek di dalam sebuah basis data, dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan. Tujuan dari klasifikasi adalah untuk menemukan model dari *training set* yang membedakan atribut ke dalam kategori atau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan atribut yang kelasnya belum diketahui sebelumnya.

Naïve Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema *Bayes.* Ciri utama dari *Naïve Bayes Classifier* ini adalah asusmsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi atau kejadian. Sebelum menjelaskan *Naïve Bayes Classifier* ini, akan dijelaskan terlebih dahulu Teorema *Bayes* yang menjadi dasar dari metoda tersebut.

Pada teorema *Bayes*, bila terdapat dua kejadian yang terpisah (misalkan A dan B), maka teorema *Bayes* dirumuskan sebagai berikut:

$$P(A|B) = \frac{P(A)}{P(B)} P(A|B) \text{.....(1)}$$

Teorema Bayes sering juga dikembangkan mengingat berlakunya hukum probabilitas total, menjadi seperti berikut:

$$P(A|B) = \frac{P(A)P(B|A)}{\sum_{i=1}^{n} P(A_i|B)} \text{.....(2)}$$

Untuk menjelaskan teorema *Naïve Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang akan dianalisis tersebut. Karena itu, teorema *Bayes* di atas disesuaikan sebagai berikut:

$$P(C|F_1, \dots, F_n) = \frac{P(C)P(F_1,\dots,F_n|C)}{P(F_1,..,F_n)} \text{.....(3)}$$

Dimana variable C merepresentasikan kelas, sementara variable F_1,……., F_n merepresentasikan karakteristik-karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel dengan karakteristik tertentu dalam kelas C (*posterior*) adalah peluang munculnya kelas C (sebelum masuk sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga evidence) karena itu, rumus (3) dapat ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{prior\ x\ likelihood}{evidence} \text{.....(4)}$$

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

Penjabaran lebih lanjut rumus *Bayes* tersebut dilakukan dengan menjabarkan $P(F_1,\dots,F_n|C)$ menggunakan aturan perkalian, menjadi sebagai berikut:

$$P(F_1,..,F_n|C = P(F_1|C)P(F_2,..,F_n|C,F_1)$$

$$= P(F_1|C)P(F_2|C,F_1)P(F_3,\dots,F_n|C,F_1,F_2)$$

$$P(F_1,\dots,F_n|C) = P(F_1|C)P(F_2|C,F_1)\dots P(F_n|C,F_1,F_2,..,F_{n-1})$$ .....(5)

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya factor-faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu-persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan.

Di sinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing-masing petunjuk (F_1,F_2…..F_n) saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(F_i - F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i)$$

Untuk i ≠ j, sehingga

$$P(F_i|C,F_j) = P(F_i|C)$$.....(6)

Dari persamaan di atas dapat disimpulkan bahwa asumsi independensi naïf tersebut membuat syarat peluang menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran (F_1,F_2…..F_n |C) dapat disederhanakan menjadi seperti berikut:

$$P(F_1 \dots . F_n|C) = P(F_1|C)P(F_2|C) \dots \dots P(F_n|C)$$

$$P(F_1 \dots . F_n|C) = \prod_{i=1}^{n} P(F_i|C)$$.....(7)

Dengan kesamaan di atas, persamaan teorema *Bayes* dapat dituliskan sebagai berikut:

$$P(F_1 \dots F_n|C) = \frac{1}{P(F_1,F_2 \dots .,F_n)} P(C) \prod_{i=1}^{n} P(F_i|C)$$

$$P(C|F_1 \dots . F_n) = \frac{P(C)}{Z} \prod_{i=1}^{n} P(F_i|C)$$.....(8)

Persamaan di atas merupakan model dari teorema *Naïve Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Adapun Z mempresentasikan *evidence* yang nilainya konstan untuk semua kelas pada satu sampel.

Table 1.1 Contoh Data untuk Klasifikasi Metode *Naive 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

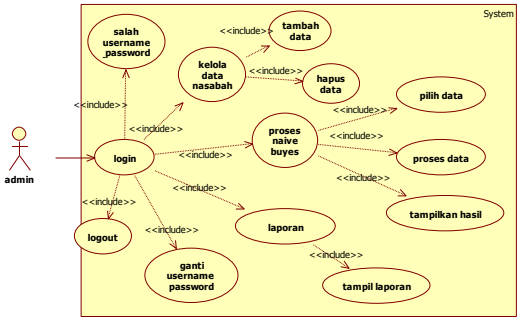
Bila terdapat sampel sepatu Hitam, Sintetis, Tali (tidak ada pada data di atas), klasifikasi dapat dilakukan dengan menggunakan *Naïve Bayes Classifier*. Pertama-tama harus dicari terlebih dahulu *Posterior* X dan Y untuk sampel tersebut.

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

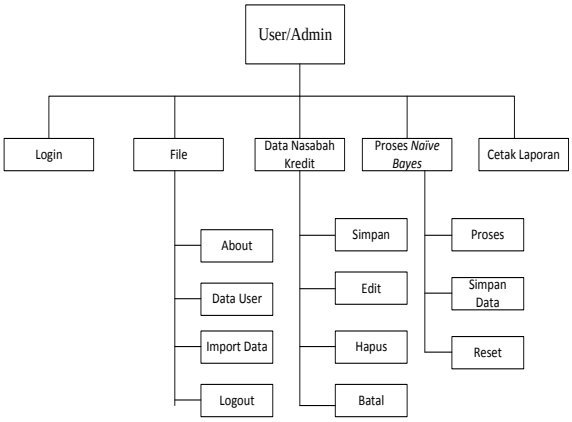
Posterior X = P(X) P(Hitam|X) P(Sintetis|X) P(Tali|X)
=0.5 x 0.66 x 0.33 x 0.66 = 0.072
Posterior Y = P(Y) P(Hitam|Y) P(Sintetis|Y) P(Tali|Y)
= 0.5 x 0.33 x 0.66 x 0.33 = 0.034
Karena *Posterior X > Posterior Y*, maka sampel sepatu tersebut bermerek X.

2. PERANCANGAN SISTEM

Pada tahap ini ditujukan untuk memperoleh gambaran mendasar mengenai desain perancangan sistem aplikasi data mining menggunakan *naïve bayes classifier* untuk persetujuan pengajuan kredit yang akan dibuat. Gambar di bawah merupakan *use case diagram* pada aplikasi persetujuan pengajuan kredit menggunakan *Naïve Bayes Classifier* .

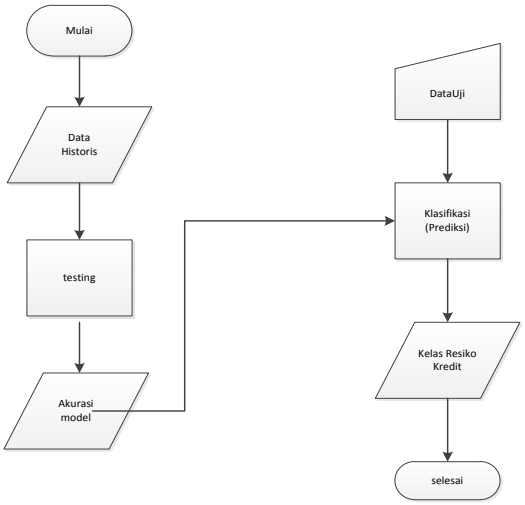


Gambar 3.1 Use Case Diagram



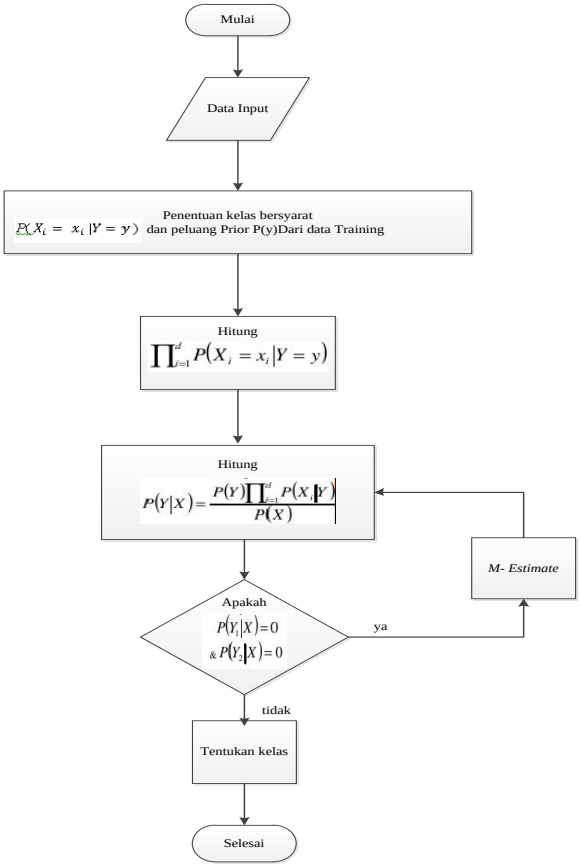
Gambar 3.2 Perancangan Struktur Menu

Flowchart sistem yang akan di buat adalah sebagai berikut:



Gambar 3.3 Flowchart Sistem Secara Keseluruhan

Flowchart tahap klasifikasi dengan *naïve bayes classifier* sebagai berikut:



Gambar 3.4 Flowchart Tahap Klasifikasi Dengan Naïve Bayes Classifier

3. IMPLEMENTASI DAN PENGUJIAN

Implementasi Aplikasi

a. Implementasi Antar Muka Pengguna (*User Interface*) Aplikasi

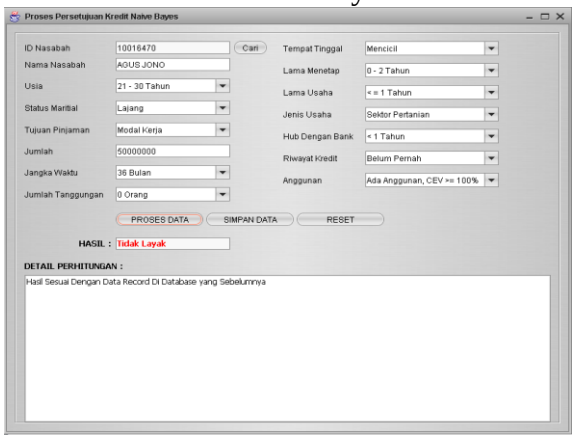
Berikut adalah tampilan antar muka pengguna (*user interface*) menu login aplikasi.



Gambar 4.1 Tampilan Antar Muka Pengguna (*User Interface*) Menu Login Aplikasi

b. Implementasi Persetujuan pengajuan kredit

Berikut contoh tampilan untuk menentukan kriterian nasabah apakah di tolak atau di terima kreditnya.



Gambar 4.2 Tampilan Prediksi Nasabah Kredit pada Aplikasi

4. KESIMPULAN DAN SARAN

Setelah melalui tahap implementasi dan pengujian sistem dilakukan maka dapat diambil beberapa kesimbulan sebagai berikut:

1. Metode klasifikasi dengan Naïve Bayes Classifier dapat digunakan untul melakukan prediksi kelas resiko pempberian kredit dengan cepat.
2. jika data yang dimiliki atau data trining nol, *Naïve Bayes Classifier* tidak dapat mengklasifikasikan record, sehingga data yang di prediksi akan bernilai nol atau tidak layak.
3. Dari hasil pengujian sistem yang diperoleh, melihat hasil tingkat keberhasilan dari testing data 11 data dengan tingkat keberhasilan 63,6% dan dari testing data sebanyak 50Data dengan tingkat keberhasilan 76%. Aplikasi data mining menggunakan naïve bayes classifier ini sudah dapat membantu dalam proses menentukan kriteria nasabah kredit.

Adapun saran untuk mengoptimalkan aplikasi di masa mendatang adalah sebagai berikut:

1) Jika data historis yang di dalam data base berjumlah banyak, maka dapat dibuat kemungkinan pertimbangan baru yaitu menambah kelas output
2. Data harus lebih bervariasi untuk menghindari adanya data traning nol, yang akan menyebabkan data yang di uji menjadi bernilai nol.

5. DAFTAR PUSTAKA

[1] Darmawan, A., 2012. *Pembuatan Aplikasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighborhood*. Fakultas Teknik dan Ilmu Komputer, Universitas Komputer Indonesia

[2] Erwin. 2009. *Analisis Market Basket dengan Algoritma Apriori dan FP-Growth.* Jurnal Generik Vol.4 No2, Juli 2009

[3] Gunadi, G., & Sensuse, D. I. 2012. *Penerapan metode data mining market basket Analysis terhadap data penjualan produk buku Dengan menggunakan algoritma apriori dan Frequent pattern growth (fp-growth) : Studi kasus percetakan pt. Gramedia.* Jurnal Telematika M.Kom Vol.4 No.1, Maret 2012

[4] Han, J., & Kamber, M. 2006. *Data Mining: Concept and Techniques.* Morgan Kaufman: San Fransisco

[5] Hermawati, F. A., 2013. *Data Mining.* Yogyakarta: Andi

[6] Huda, N. M., 2010. *Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa*]. Fakultas Mipa, Universitas Diponegoro.

[7] Liao,Shu-Hsi., Chu Hui-Pei dan Hsia, Pei-Yuan. 2012. *Data mining techniques and applications – A decade review from 2000 to 2011*. Department of Management Sciences, Tamkang University.

[8] Pramudiono, I., (2007). *Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data.* Diakses tanggal 26 April 2014, dari http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.

http://id.wikipedia.org/wiki/Kota_Bandung.