

SISTEM PENDUKUNG PENGAMBILAN KEPUTUSAN MENGGUNAKAN METODE NAIVE BAYES (STUDI KASUS KREDIT SEPEDA MOTOR)

¹ **Jatmika, S.Si, M.Kom** ² **Lilis Anggraeni**

^{1,2}Jurusan Komputer, Teknik Informatika, Fakultas Sains dan Komputer, Universitas Kristen Immanuel
Jalan Solo Km. 11 PO Box 4 YKAP Yogyakarta, ph: (0274) 496256-296247 fax: (0274) 496258
lilisanggraeni9@gmail.com

Abstract

In a company or enterprise decision-making as an important element that affects the development of the company. Therefore, it is necessary to design a decision support system that aims to help a manager to take a decision.

Decision Support System is designed to be a system that can help the user to help make a decision. In this case the user is the manager. In order to give consideration to give approval for credit application or not. To design such a system is required to support the method of system performance. The algorithm used is Naive Bayes. Existing data on the company will be used for processing using Naive Bayes algorithm. Naive Bayes algorithm will calculate the posterior probability for each case values of the target attribute in every case (sample data).

Furthermore, Naive Bayes will classify the sample data to a class that has a posterior probability value highest. . The output of the system is on a potential creditors will qualify creditor to credit approval or not . Testing is done using old data or data from the company as training data that will be tested with new data (data potential creditors) .

Key Word: *Naive Bayes, Probabilitas Posterior, Kreditor.*

1. Pendahuluan

Fasilitas jalan raya di Indonesia yang semakin rumit memaksa para pengguna jasa angkutan umum dan juga pejalan kaki beralih menggunakan kendaraan pribadi, salah satu kendaraan yang banyak diminati adalah kendaraan roda dua yaitu sepeda motor. Banyak masyarakat yang ingin memiliki sepeda motor, tetapi tidak memiliki dana yang cukup untuk membeli motor secara tunai. Tentu membutuhkan suatu cara untuk memiliki motor tetapi dengan sistem pembayaran secara kredit.

Para produsen motor memanfaatkan peluang ini untuk bersaing memberikan kredit yang mudah sehingga diminati banyak konsumen. Secara umum pemberian kredit pada sebuah perusahaan produsen motor didasarkan pada kegiatan analisa kredit. Untuk mendapatkan kredit motor konsumen harus mengikuti prosedur yang berlaku di perusahaan yang bersangkutan. Analisa yang dilakukan dengan cara interview, pengisian proposal dan kelengkapan data serta dilakukan survey lapangan. Untuk kemudian produsen motor melakukan kegiatan administrasi kredit. Sejauh ini untuk dapat menentukan konsumen yang berhak mendapatkan kredit dengan kegiatan

administrasi yang meliputi kegiatan analisis secara kuantitatif dan kualitatif memerlukan waktu yang cukup lama, karena data yang disajikan tidak berbentuk angka dan membutuhkan pertimbangan yang cukup lama.

Untuk membantu mengatasi masalah tersebut, maka penulis membuat sistem pendukung pengambilan keputusan dalam kredit sepeda motor, agar user yang dalam hal ini adalah manajer, tidak kesulitan untuk menentukan memberikan kredit atau tidak. serta mempersingkat pengerjaan waktu dalam analisa kredit, agar konsumen tidak cenderung mencari produsen lain yang dapat memproses analisa kredit secara lebih cepat.

2. Decision Support system

DSS (Decision Support System) atau Sistem Pendukung Pengambilan Keputusan adalah seperangkat sistem yang mampu memecahkan masalah secara efisien dan efektif, yang bertujuan untuk membantu pengambil keputusan memilih berbagai alternative keputusan yang merupakan hasil pengolahan informasi – informasi yang diperoleh / tersedia dengan menggunakan model – model pengambilan keputusan. proses pengambilan keputusan terdiri atas 3 fase yaitu:

a. Fase Intelligence

Tahap ini merupakan proses penelusuran dan pendeteksian dari lingkup problematika serta proses pengenalan masalah. Data masukan diperoleh, diproses, dan diuji dalam rangka mengidentifikasi masalah.

b. Fase Design

Tahap ini merupakan proses menemukan, mengembangkan dan menganalisa alternative tindakan yang bisa dilakukan. Tahap ini meliputi proses untuk mengerti masalah, menurunkan solusi dan menguji kelayakan solusi.

c. Fase Choice

Pada tahap ini dilakukan proses pemilihan diantara berbagai alternatif tindakan yang mungkin dijalankan. Hasil pemilihan tersebut kemudian diimplementasikan dalam proses pengambilan keputusan. Rumusan masalah dalam penelitian ini adalah bagaimana membuat aplikasi yang dapat digunakan user untuk mendukung pengambilan keputusan dalam pemberian pinjaman kredit motor.

Karakteristik sistem pendukung pengambil keputusan adalah sebagai berikut:

Sistem pendukung pengambil keputusan dirancang untuk membantu pengambilan keputusan dalam memecahkan masalah yang sifatnya semi terstruktur ataupun tidak terstruktur.

Sistem pendukung keputusan dirancang sedemikian rupa sehingga dapat digunakan dengan mudah oleh orang – orang yang tidak memiliki dasar kemampuan pengoperasian komputer yang tinggi.

Sistem pendukung keputusan dirancang dengan menekankan pada aspek fleksibilitas serta kemampuan adaptasi yang tinggi. Sehingga mudah disesuaikan dengan berbagai perubahan lingkungan yang terjadi dalam kebutuhan pemakai.

Dengan berbagai karakteristik seperti dijabarkan diatas, sistem pendukung keputusan tidak hanya memiliki keuntungan tetapi juga memiliki kerugian. Keuntungan yang dimaksud diantaranya meliputi:

- a. Sistem pendukung pengambilan keputusan memperluas kemampuan pengambil keputusan dalam memproses data atau informasi bagi pemakainya.
- b. Sistem pendukung pengambilan keputusan membantu pengambil keputusan dalam hal penghematan waktu yang dibutuhkan untuk memecahkan masalah yang sangat kompleks dan tidak terstruktur.
- c. Sistem pendukung pengambilan keputusan dapat menghasilkan solusi dengan lebih cepat serta mendapatkan hasil yang lebih baik.

Adapun kerugian dari sistem pendukung pengambil keputusan meliputi beberapa hal, antara lain:

- a. Pada pemodelan sistem bisnis akan menghasilkan model yang tidak dapat menangkap semua pengaruh pada entity.

- b. Sistem pendukung pengambilan keputusan membutuhkan keterampilan matematika yang tinggi untuk mengembangkan model yang lebih kompleks secara pribadi.

3. Klasifikasi

Klasifikasi dan prediksi adalah dua bentuk analisis data yang bisa digunakan untuk mengekstrak model dari data yang berisi kelas – kelas atau untuk memprediksi trend data yang akan datang. Klasifikasi memprediksi data dalam bentuk kategori, sedangkan prediksi memodelkan fungsi – fungsi dari nilai kontinyu. Misalkan model klasifikasi bisa dibuat untuk mengelompokkan aplikasi peminjaman pada bank apakah beresiko atau aman, sedangkan model prediksi bisa dibuat untuk memprediksi pengeluaran untuk membeli peralatan komputer dari pelanggan potensial berdasarkan pendapatan dan lokasi tinggal.

Prediksi bisa dipandang sebagai pembentukan dan penggunaan model untuk menguji kelas dari sampel yang tidak berlabel, atau menguji nilai atau rentang nilai dari suatu atribut. Dalam pandangan ini, klasifikasi dan regresi adalah dua jenis masalah prediksi, dimana klasifikasi digunakan untuk memprediksi nilai – nilai yang kontinyu. Untuk selanjutnya penggunaan istilah *prediction* untuk memprediksi kelas yang berlabel disebut *classification*, dan penggunaan istilah prediksi untuk memprediksi nilai – nilai yang kontinyu sebagai *prediction*.

4. Model klasifikasi.

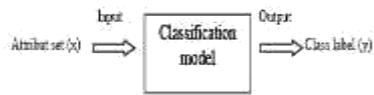
Data input untuk klasifikasi adalah koleksi dari *record*. setiap *record* dikenal sebagai *instance* atau contoh, yang ditentukan oleh sebuah tuple (x,y) , dimana x adalah himpunan atribut dan y adalah atribut tertentu, yang dinyatakan sebagai label kelas (juga dikenal sebagai kategori atau atribut target). Klasifikasi adalah tugas pembelajaran sebuah fungsi target f yang memetakan setiap himpunan atribut x ke salah satu label kelas y yang telah didefinisikan sebelumnya. Fungsi target juga dikenal secara informal sebagai model klasifikasi.

➤ Pemodelan Deskriptif.

Model klasifikasi dapat bertindak sebagai alat penjelas untuk membedakan objek– objek dari kelas – kelas yang berbeda. Sebagai contoh untuk para ahli Biologi, model deskriptif yang meringkas data.

➤ Pemodelan Prediktif.

Model klasifikasi juga dapat digunakan untuk memprediksi label kelas dari record yang tidak diketahui. Seperti pada Gambar 2 tampak sebuah model klasifikasi dapat dipandang sebagai kotak hitam yang secara otomatis memberikan sebuah label ketika dipresentasikan dengan himpunan atribut dari *record* yang tidak diketahui.



Gambar 1

Gambar 1 adalah Klasifikasi sebagai pemetaan sebuah himpunan atribut input x ke dalam label kelas. Beberapa teknik klasifikasi yang digunakan adalah *decision tree classifier*, *rule-based classifier*, *neural-network*, *support vector machine*, dan *naïve Bayes classifier*. Setiap teknik menggunakan algoritma pembelajaran untuk mengidentifikasi model yang memberikan hubungan yang paling sesuai antara himpunan atribut dan label kelas dari data input.

Pendekatan umum yang digunakan dalam masalah klasifikasi adalah, pertama, *training set*, berisi record yang mempunyai label kelas yang diketahui haruslah tersedia. *Training set* digunakan untuk digunakan untuk membangun model klasifikasi, yang kemudian diaplikasikan ke *test set*, yang berisi *record – record* dengan label kelas yang tidak diketahui.

5. Pengertian Teorema Bayes

Teori keputusan Bayes atau sering disebut Teorema Bayes adalah pendekatan statistic yang fundamental dalam pengenalan pola (*pattern recognition*). Pendekatan Teorema Bayes ini didasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas dan nilai yang muncul dalam keputusan-keputusan tersebut.

Jika X adalah bukti atau kumpulan data pelatihan dan adalah hipotesis. Jika class variable memiliki hubungan tidak deterministic dengan atribut, maka dapat diperlukan X dan sebagai atribut acak dan menangkap hubungan peluang menggunakan . Peluang bersyarat ini juga dikenal dengan probabilitas posterior untuk , dan () adalah probabilitas prior. Untuk mengestimasi peluang posterior secara akurat untuk setiap kombinasi label kelas yang mungkin dan nilai atribut adalah masalah sulit karena membutuhkan training set sangat besar, meski untuk jumlah *moderate* atribut. Penggunaan teorema Bayes untuk melakukan klasifikasi sangat bermanfaat karena menyediakan pernyataan istilah peluang posterior dari peluang prior (), peluang kelas bersyarat dan bukti () seperti pada Rumus 2.1 berikut [3] :

$$(1) \quad \frac{P(Y|X)}{P(Y)} = \frac{P(X_1|Y)P(X_2|Y) \dots P(X_n|Y)}{P(X_1)P(X_2) \dots P(X_n)} \quad (1)$$

Dalam hal ini :

X = himpunan data training

Y = hipotesis.

(|) = probabilitas posterior, yaitu probabilitas bersyarat dari hipotesis Y berdasarkan kondisi X .

() = probabilitas prior dari hipotesis Y , yaitu probabilitas bahwa hipotesis Y bernilai benar sebelum data X muncul.

() = probabilitas dari data X .

(|) = probabilitas bersyarat dari X berdasarkan kondisi pada hipotesis Y , dan biasa disebut dengan *likelihood*. *Likelihood* ini mudah untuk dihitung ketika memberikan nilai 1 saat X dan Y konsisten, dan memberikan nilai 0 saat X dan Y tidak konsisten.

6. Klasifikasi Naïve Bayesian

Klasifikasi Naïve Bayesian merupakan salah satu metode pengklasifikasi yang berdasarkan pada penerapan Teorema Bayes dengan asumsi antar atribut penjas saling bebas (*independent*). Algoritma ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman dimasa sebelumnya.

Klasifikasi Naïve Bayes diasumsikan dimana nilai atribut dari sebuah kelas dianggap terpisah dan independen dengan nilai atribut lainnya:

$$(1) \quad \frac{P(Y|X)}{P(Y)} = \frac{P(X_1|Y)P(X_2|Y) \dots P(X_n|Y)}{P(X_1)P(X_2) \dots P(X_n)} \quad (2)$$

Keterangan :

X = himpunan data training

Y = hipotesis.

(|) = probabilitas posterior, yaitu probabilitas bersyarat dari hipotesis Y berdasarkan kondisi X . () = probabilitas prior dari hipotesis Y , yaitu probabilitas bahwa hipotesis $Y|X_N$ bernilai benar sebelum data X muncul.

() = probabilitas dari data X .

(| 1) , (| 2) , (|) = probabilitas dari X_1, X_2, X_n untuk hipotesis Y , biasa disebut dengan *likelihood*.

Karena $P(X)$ *irrelevant* maka untuk mencari peluang hanya menggunakan rumus berikut ini :

$$P((Y|X) = P(X_1|Y)P(X_2|Y) \dots P(X_n|Y)P(Y) \quad (3)$$

Jika ada $P(X_n|Y)$ yang memiliki nilai 0, maka $P(Y|X) = 0$. Maka klasifikasi Naïve Bayesian tidak bisa memprediksi record yang salah satu atributnya memiliki probabilitas bersyarat (*likelihood*) = 0. Untuk mengatasi hal itu maka dilakukan penambahan nilai 1 ke setiap *evidence* dalam perhitungan sehingga probabilitas tidak akan bernilai 0. Langkah ini sering disebut Laplace Estimator dengan rumus sebagai berikut :

$$P(X_i|Y_j) = \frac{n_{ij} + 1}{n + m} \quad (4)$$

Dimana :

= total jumlah instances dari kelas

= jumlah contoh training dari yang menerima nilai

= parameter yang dikenal sebagai ukuran sampel ekuivalen

$$Akurasi = \frac{\sum \text{Data yang sesuai data testing}}{\sum \text{Jumlah data testing}} \times 100 \%$$

7. Analisis Hasil Pengujian

Setelah sistem berhasil diimplementasikan dan dapat langsung digunakan oleh *user*, tahap akhir dari penelitian ini adalah pengujian sistem yang telah dibangun. Sistem pendukung pengambilan keputusan yang dibangun menggunakan algoritma *Naive Bayesian*.

Tahap pengujian sistem sebagai berikut: Pengujian dilakukan sebanyak 10 kali. Dengan menggunakan total data pelatihan 100 data dan data uji total 90 data. Di bawah ini adalah beberapa contoh pengujian yang dilakukan penulis:

Pengujian I dengan total data pelatihan 50 data dan data uji 10 data dari data asli. Dan hasilnya adalah :

Tabel 1

harga motor	uang muka	cicilan	banyak cicilan	pekerjaan	penghasilan	pengeluaran	rekomendasi
14.555.000	2.950.000.00	1.373.000.00	12	Guru	5.000.000.00	3.000.000.00	ya
11.300.000	2.750.000.00	1.649.000.00	34	wiraswasta	1.500.000.00	900.000.00	tidak
11.300.000	3.000.000.00	1.192.000.00	12	guru	4.000.000.00	2.500.000.00	ya
12.770.000	2.500.000.00	1.204.000.00	11	Dagang	3.200.000.00	1.000.000.00	tidak
15.890.000	3.400.000.00	1.691.000.00	34	pegawai	900.000.00	900.000.00	tidak
15.335.000	3.100.000.00	1.413.000.00	12	PNS	6.300.000.00	4.000.000.00	ya
12.775.000	4.500.000.00	1.191.000.00	12	Wiraswasta	2.100.000.00	1.000.000.00	tidak
14.120.000	2.850.000.00	1.332.000.00	14	Guru	4.450.000.00	2.900.000.00	ya
23.810.000	4.790.000.00	2.155.000.00	14	Perawat	2.300.000.00	1.200.000.00	tidak
13.500.000	2.700.000.00	1.295.000.00	13	PNS	1.250.000.00	1.000.000.00	tidak

Pengujian II dengan total data pelatihan 50 data uji 10 data dari data acak. Dan hasilnya ada di tabel 2:

Tabel 2

harga motor	uang muka	cicilan	banyak cicilan	pekerjaan	penghasilan	pengeluaran	rekomendasi
11.300.000	3.500.000.00	604.000.00	24	Guru	3.000.000.00	2.800.000.00	ya
17.700.000	3.550.000.00	1.636.000.00	14	Pegawai Sw.	5.000.000.00	2.000.000.00	ya
11.300.000	2.100.000.00	640.000.00	24	Guru	2.200.000.00	900.000.00	tidak
11.300.000	2.100.000.00	1.000.000.00	14	Guru	2.000.000.00	2.000.000.00	ya
16.500.000	2.300.000.00	2.100.000.00	12	Karyawan T.	1.000.000.00	900.000.00	tidak
17.500.000	2.000.000.00	1.500.000.00	12	Guru	1.500.000.00	3.000.000.00	ya
11.300.000	2.800.000.00	1.050.000.00	24	Pegawai	3.000.000.00	1.500.000.00	ya
11.300.000	2.000.000.00	2.000.000.00	12	Pegawai Sw.	2.000.000.00	2.000.000.00	tidak
1.000.000.00	1.000.000.00	1.200.000.00	12	PNS	1.000.000.00	900.000.00	tidak
11.300.000	4.500.000.00	2.100.000.00	12	PNS	10.000.000.00	7.000.000.00	ya

Pengujian yang lain dari pengujian III hingga pengujian x terlampir. Untuk membuktikan keakuratan sistem maka penulis membuat tingkat akurasi dari setiap pengujian dan mengambil rata – rata untuk menilai persentase keakuratan sistem. Tingkat akurasi dilakukan dengan perhitungan sebagai berikut:

Dari pengujian dan perhitungan akurasi yang sudah dilakukan penulis, maka hasil pengujian adalah sebagai berikut:

Tabel 3

Peng ujian	Total Data Training	Total Data Testing	Data Tidak Sesuai	Data Sesuai	Akurasi
I	50	10	1	9	90%
II	50	10	4	6	60%
III	100	10	0	10	100%
IV	100	10	6	4	40%
V	25	5	2	3	60%
VI	25	10	6	4	40%
VII	75	10	3	7	70%
VIII	75	10	1	9	90%
IX	10	5	4	1	20%
X	10	10	7	3	30%
Rata – Rata = 60%					

Pada pengujian I sampai pengujian X memiliki nilai akurasi yang berbeda - beda. Hal ini dapat disebabkan karena variasi data untuk tiap atribut data pengujian berbeda-beda. Asumsi lain yang dimiliki adalah dengan adanya data-data anomali pada data *testing* dapat mempengaruhi nilai akurasi sistem. Contoh pada pengujian ke IV dari data *testing* sebanyak 10 data,terdapat 6 *record* yang memiliki pola data seperti pada tabel 4.

Tabel 4

34	Heri	2000000	2000000	2100000	1000000	11300000	Penghasilan = Pengeluaran
35	Asnawi	1000000	900000	2300000	2100000	16500000	Cicilan > Penghasilan
36	Dewi	1500000	3000000	2000000	1500000	17500000	Penghasilan > Pengeluaran
37	Henry	3000000	1500000	2800000	1050000	11300000	Jum.Cicilan > Harga Motor
38	Mia Ratih	2000000	2000000	2000000	2000000	11300000	Semua Atribut Sama
39	Gundu	1000000	900000	1000000	1200000	1000000	Harga motor tidak Normal

8. Kesimpulan Dan Saran

Berdasarkan hasil pengujian dapat diketahui bahwa algoritma Naïve Bayes memiliki peluang untuk dipergunakan sebagai sistem pendukung pengambilan keputusan. Namun algoritma Naïve Bayes akan lebih baik nilai ke akuratan jika data pelatihan dan data uji tidak terdapat data yang bernilai abnormal. Sehingga untuk memaksimalkan fungsi dari sistem tersebut yang menggunakan algoritma naïve bayes data yang dipakai untuk data

pelatihan harus memenuhi standar. Dalam hal ini nilai standar data adalah sesuai dengan persyaratan kredit sepeda motor.

Daftar Pustaka:

- [1] Indhalina, Asteria, (2010), Tugas Akhir :Klasifikasi Data Penerimaan Mahasiswa Baru Dan Prestasi Akademik Mahasiswa Jurusan Teknik Informatika Universitas Sanata Dharma Dengan Menggunakan Algoritma *Naïve Bayes*, Universitas Sanata Dharma, Yogyakarta.
- [2] Kurniawan,Arief.2010.*Sistem Pendukung Pengambilan Keputusan Kelayakan Kredit Pemilikan Motor Dengan Metode Scoring System*.
- [3] Kusriani dan Luthfi, EmhaTaufiq. (2009). *Algoritma Data Mining*, Penerbit Andi. Yogyakarta.
- [4] Santoso, Budi.(2007). *Data Mining : Teknik Pemanfaatan Data Untuk Keperluan Bisnis*.
- [5] Han, Jiawei & Micheline Kamber. (2006). *Data Mining : Concept and Techniques*, Second Edition, Morgan Kaufman Publishers. San Fransisco, USA.
- [6] Tan, P., Michael, S., dan Vipin, K. 2005. *Introduction to Data Mining*. Boston: Pearson Education,Inc.