

Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan

Hakam Febtadianrano Putro¹⁾; Retno Tri Vlandari²⁾; Wawan Laksito Yuly Saptomo³⁾

^{1) 3)} Program Studi Informatika, STMIK Sinar Nusantara

²⁾ Program Studi Teknologi Informasi, STMIK Sinar Nusantara

¹⁾hakamputro@gmail.com; ²⁾retnotv@sinus.ac.id; ³⁾wlaksito@sinus.ac.id

ABSTRACT

Business location plays an important role in sales. The business location in cities makes the seller easier to distribute activities for people. Distribution activities are closely related to sales activities. If there is a sales transaction, a classification of potential and non-potential customers will be required. One method that can be used for classification is mining data. One of the most frequently used data mining for classification is the Naive Bayes method. The attributes used in the customer classification process are purchase amount, time interval, and location. The result of classification system are 23 true reactions and 2 false reactions. Based on the results are using the confusion matrix method, it shows that the accuracy value reaches 92%, the precision value reaches 100%, the recall value reaches 91%.

Keywords: Trading Business, Customer Classification, Naive Bayes, Confusion Matrix

I. PENDAHULUAN

Usaha dagang di Sukoharjo semakin berkembang pesat. Semakin banyak pula usaha dagang dalam bidang penjualan ayam. Dengan adanya usaha dagang ini juga meliputi aspek yang bermanfaat bagi masyarakat sekitarnya.

Perkembangan teknologi yang terus berkembang membuat usaha dagang berlomba-lomba mencari terobosan baru

Permasalahan utama dalam usaha dagang ini masih kesulitan dalam mengolah data. Kurangnya pemanfaatan secara optimal dari banyaknya data transaksi untuk melihat pelanggan mana yang berpotensi atau tidak berpotensi.

Data mining merupakan salah satu teknik dalam pengolahan data yang menemukan hubungan dari data yang tidak diketahui oleh pengguna serta menyajikannya kedalam bentuk yang mudah dipahami sehingga dari hubungan data tersebut dapat dijadikan sebagai dasar dalam pengambilan keputusan [1]. Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan yaitu : Deskripsi, Estimasi, Prediksi, Klasifikasi, *Clustering*, dan Asosiasi [2].

Secara singkat, klasifikasi merupakan pengelompokan objek kedalam kelas tertentu berdasarkan kelompoknya yang biasanya disebut dengan kelas (*class*). Salah satu metode klasifikasi yang sering digunakan adalah metode naive bayes. Berdasarkan penelitian yang telah dilakukan sebelumnya menggunakan metode naive bayes.

Metode naive bayes memiliki kelebihan, yaitu cepat dalam perhitungan, algoritma yang sederhana dan berakurasi tinggi [3]. Metode *Naive Bayes* yang hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian [4]. Metode Algoritma *Naive Bayes Classifier* lebih mudah digunakan karena memiliki alur perhitungan yang tidak panjang [5]

Dari permasalahan tersebut maka menerapkan metode naive bayes untuk mengklasifikasikan pelanggan. Adapun Tujuan dari penelitian ini adalah membuat aplikasi yang dapat membantu mengklasifikasikan pelanggan yang berpotensi dan tidak berpotensi, serta digunakan sebagai dasar pertimbangan keputusan salah satunya sebagai strategi pemasaran.

II. TINJAUAN PUSTAKA

2.1 Data Mining

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 mencari dari data yang terdapat pada basis data. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar

sehingga sering disebut *Knowledge Discovery Databases* (KDD) [6].

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu (1) Pembangunan model sebagai *prototype* untuk disimpan sebagai memori dan (2) Penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada suatu objek data lain agar diketahui dikelas mana objek data tersebut dalam model yang sudah disimpnannya [7].

2.2 Personal Home Page Hipyertext Preprocessor (PHP)

PHP merupakan bahasa berbentuk script yang ditempatkan dalam server dan diproses di server. PHP singkatan dari *Hipyertext Preprocessor* yaitu bahasa pemrograman *web server-side* yang bersifat *open source* [8]. Artinya semua sintaks yang diberikan akan sepenuhnya dijalankan pada *server*, sedangkan yang dikirimkan ke *browser* hanya hasilnya saja. PHP tergolong sebagai perangkat lunak *open source* yang diatur dalam aturan *General Public License* (GPL). PHP dirancang khusus untuk membentuk web dinamis, artinya PHP mampu menghasilkan website yang secara terus menerus hasilnya bisa berubah-ubah sesuai dengan pola yang diberikan. Hal ini berkaitan erat dengan basis data sebagai sumber data yang akan ditampilkan [9].

2.3 Penelitian Terkait

Penelitian tentang prediksi besarnya Penggunaan listrik rumah tangga. Tujuan dari penelitian ini mampu untuk memprediksi besarnya penggunaan listrik tiap rumah tangga agar lebih mudah mengatur penggunaan listrik. Variabel yang dipakai meliputi jumlah tanggungan keluarga, luas rumah, pendapatan/bulan, daya listrik, perlengkapan yang dimiliki. Metode yang dipakai menggunakan metode *Naïve Bayes*. Hasil dari penelitian ini persentase 78,3333% untuk keakuratan prediksi, di mana dari 60 data penggunaan listrik rumah tangga yang diuji terdapat 47 data penggunaan listrik rumah tangga yang berhasil diklasifikasikan dengan benar [10].

Penelitian tentang penerapan data mining pada penjualan barang menggunakan metode *Naïve Bayes Classifier* untuk optimasi

strategi pemasaran. Tujuan dari penelitian ini untuk mengetahui minat dan ketertarikan calon pembeli terhadap produk yang tersedia. Variabel yang dipakai meliputi bulan, model barang, harga barang, terjual. Metode yang dipakai menggunakan metode *Naïve Bayes*. Hasil dari penelitian ini didapatkan nilai p value (sig), nilai p value (sig) $0,807 > 0,05$ maka H_0 diterima dan H_a ditolak, nilai t -hitung dan t -tabel, Nilai t -hitung – $0,246 > - t$ -tabel $-2,045$ maka H_0 diterima dan H_a ditolak [11].

Penelitian tentang penentuan penerima beasiswa bidikmisi Universitas Mulawarman. Tujuan dari penelitian ini untuk membantu bagian proses seleksi dengan membuat aplikasi perangkat lunak sistem pendukung keputusan untuk penentuan penerima beasiswa bidikmisi Universitas Mulawarman. Variabel yang dipakai meliputi pekerjaan orang tua, penghasilan orang tua, jumlah tanggungan, daya listrik(*watt*), dan nilai ujian nasional. Metode yang dipakai menggunakan metode *Naïve Bayes*. Hasil dari penelitian ini menghasilkan sebuah aplikasi sistem pendukung keputusan dengan tingkat akurasi sebesar 85.56% [12].

Penelitian tentang penentuan kelayakan pemilihan tempat tinggal. Tujuan dari penelitian ini untuk konsumen dalam mempertimbangkan penentuan pencarian tempat tinggal. Variabel dipakai meliputi keamanan, fasilitas umum, bebas banjir, harga, air bersih, model, sejuk dan nyaman. Metode yang dipakai menggunakan metode *Naïve Bayes*. Hasil dari penelitian ini menghasilkan probabilitas antara layak (sesuai) atau tidak sesuai, yang nantinya dapat digunakan oleh masyarakat umum untuk mempermudah dalam penentuan pencarian tempat tinggal [13].

Penelitian tentang klasifikasi kelayakan keluarga penerima beras rastra. Tujuan dari penelitian ini untuk mengatasi masalah yang belum optimal di desa Bandar Siantar Kecamatan Gunung Malela. Variabel yang dipakai meliputi Nama Kepala Rumah Tangga, PKH, Jml Tanggungan, Kepala Rumah Tangga, Kondisi Rumah, Jml Penghasilan, Status Pemilik Rumah. Metode yang dipakai menggunakan metode *Naïve Bayes*. Hasil dari penelitian ini diharapkan dapat membantu pemerintah khususnya di daerah dalam menentukan kelayakan keluarga penerima beras Rastra [14].

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayes classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [15].

NBC merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki kemampuan dan akurasi tinggi.

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (1)$$

dengan

X : data dengan class yang belum diketahui

C_i : hipotesis data X merupakan suatu class spesifik

P(C_i|X) : probabilitas hipotesis C_i berdasarkan kondisi X (*posteriori probability*)

P(C_i) : probabilitas hipotesis C_i (*prior probability*)

P(X|C_i) : probabilitas X berdasarkan kondisi pada hipotesis C_i

P(X) : probabilitas dari X

Naïve Bayesian Classifier mengasumsikan bahwa keberadaan sebuah atribut (variabel) tidak ada kaitannya dengan beradaan atribut (variabel) yang lain karena asumsi atribut tidak saling terkait (*conditionally independent*), ditulis dengan rumus:

$$P(X|C_i) = \sum_{k=1}^n P(X_k|C_i)$$

Setelah diperoleh hasil dari seluruh data pada setiap class, maka hasil akhirnya dapat menggunakan rumus: (2)

$$P(X|C_i) = \arg \max P(X_i|C_i) * P(C_i)$$

III. METODE PENELITIAN

Dalam penyusunan penelitian ini diperlukan metode-metode penelitian dalam proses pengumpulan data antara lain:

3.1 Sumber Data

a. Data Primer

Data primer diperoleh dari hasil kegiatan pengumpulan data terhadap data transaksi penjualan ayam.

b. Data Sekunder

Data sekunder dalam penelitian ini melakukan pengolahan data dari data laporan penjualan ayam untuk menentukan kriteria status usaha, jumlah pembelian, dan lokasi. Selain itu untuk menunjang data pendukung yang diperoleh dari media pustaka tentang teori-teori yang digunakan dalam penelitian ini. Adapun data sekunder yang diantaranya Data Mining, Klafikasi, Naïve Bayes, MySQL, PHP, UML, Kajian Pustaka.

3.2 Metode Pengumpulan Data

a. Metode Observasi, dalam penelitian ini dilakukan pengamatan terhadap prosedur penjualan ayam.

b. Metode Wawancara, menanyakan langsung kepada pemilik UD.Samodro untuk mendapatkan data yang akan diteliti untuk menyesuaikan proses data yang akan diolah menggunakan metode Naïve Bayes.

c. Studi Pustaka, terdiri dari literatur mengenai penelitian yang sedang dilakukan. Literatur yang dimaksud diantaranya beberapa jurnal ilmiah dan buku yang terkait dengan kasus penelitian ini. Adapun jurnal ilmiah dan buku yang dijadikan sebagai acuan.

3.3 Metode Pengembangan Sistem

a. Perancangan Sistem

Perancangan sistem dalam penelitian ini menggunakan UML diharapkan pengerjaan bahasa visual yang lebih jelas dan terstruktur baik user yang digunakan.

b. Kontruksi Sistem

Kontruksi sistem dalam penelitian ini penunjang pembuatan sebuah aplikasi yang saling berkaitan, yaitu perangkat keras(hardware), perangkat lunak (software).

c. Pengujian Sistem

Proses selanjutnya pengujian aplikasi yang telah dibuat akan diuji untuk mengetahui apakah aplikasi yang dibuat sudah berjalan dengan dan sesuai perancangan yang dilakukan.

IV. HASIL DAN PEMBAHASAN

4.1 Klasifikasi Pelanggan

Data latih dan data uji yang digunakan untuk klasifikasi berjumlah 75 data latih dan

25 data uji, dimana data latih selanjutnya akan proses menggunakan metode naive bayes. Adapun data latih yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Data Latih

Kode Pelanggan	Jumlah Pembelian	Interval Waktu	Lokasi	Target
P1	Sangat Banyak	Harian	Dekat	Berpotensi
P2	Sangat Banyak	Harian	Jauh	Berpotensi
P3	Sangat Sedikit	Mingguan	Jauh	Tidak Berpotensi
P4	Sangat Sedikit	Mingguan	Dekat	Tidak Berpotensi
P5	Banyak	Harian	Jauh	Berpotensi
P6	Banyak	Harian	Dekat	Berpotensi
P7	Sangat Sedikit	Mingguan	Dekat	Tidak Berpotensi
P8	Sangat Sedikit	Mingguan	Dekat	Tidak Berpotensi
P9	Banyak	Harian	Jauh	Berpotensi
P10	Banyak	Harian	Jauh	Berpotensi
P11	Sedikit	Harian	Dekat	Berpotensi
P21	Sedikit	Harian	Dekat	Berpotensi
P31	Sangat Sedikit	Harian	Jauh	Tidak Berpotensi
P41	Sangat Banyak	Harian	Dekat	Berpotensi
P51	Sangat Sedikit	Mingguan	Jauh	Tidak Berpotensi
P61	Sedikit	Harian	Dekat	Berpotensi
P71	Sedikit	Mingguan	Dekat	Berpotensi
P75	Sangat Banyak	Harian	Dekat	Berpotensi

Data yang dipakai sebagai data latih dan uji menggunakan data pada tahun 2019 yang dapat dilihat pada Tabel 2.

Tabel 2. Komposisi Data Latih & Data Uji

Jenis Data	Jumlah	Keterangan
Data Latih	75	Penjualan pada Bulan Juni - Agustus
Data Uji	25	Penjualan pada Bulan Juni - Agustus

Adapun nilai probabilitas setiap kriteria yang didapatkan dari data latih pada Tabel 1. Ditetapkan nilai probabilitas dari setiap kriteria kedalam kategori, yaitu :

- a. Probabilitas jumlah pembelian pada setiap class, seperti pada Tabel 3.

Tabel 3. Probabilitas Jumlah Pembelian

Jumlah Pembelian	Jumlah kejadian "Jumlah Pembelian"		Probabilitas	
	Berpotensi	Tidak Berpotensi	Berpotensi	Tidak Berpotensi
Sangat Banyak	19/55	0/20	0,34	0
Banyak	18/55	0/20	0,33	0
Sedikit	17/55	2/20	0,31	0,1
Sangat Sedikit	1/55	18/20	0,02	0,9

- b. Probabilitas interval waktu pada setiap class, seperti pada Tabel 4.

Tabel 4. Probabilitas Interval Waktu

Interval Waktu	Jumlah kejadian "Interval Waktu"		Probabilitas	
	Berpotensi	Tidak Berpotensi	Berpotensi	Tidak Berpotensi
Harian	53/55	3/20	0,95	0,15
Mingguan	3/55	17/20	0,05	0,85

- c. Probabilitas lokasi pada setiap class, seperti pada Tabel 5.

Tabel 5. Probabilitas Lokasi

Lokasi	Jumlah kejadian "Lokasi"		Probabilitas	
	Berpotensi	Tidak Berpotensi	Berpotensi	Tidak Berpotensi
Dekat	39/55	7/20	0,71	0,35
Jauh	16/55	13/20	0,29	0,65

- d. Probabilitas target pada setiap class, seperti pada Tabel 6.

Tabel 6. Probabilitas Target

Jumlah Kejadian "Target"		Probabilitas	
Berpotensi	Tidak Berpotensi	Berpotensi	Tidak Berpotensi
55/75	20/75	0,73	0,27

4.2 Pengujian Metode Naive Bayes Menggunakan Data Testing

Setelah proses klasifikasi menggunakan metode naive bayes selesai, selanjutnya pengujian data uji yang berjumlah 25 data pelanggan dapat dilihat pada Tabel 7.

Tabel 7. Data Uji

Kode Pelanggan	Jumlah Pembelian	Interval Waktu	Lokasi	Prediksi Target
P76	Banyak	Harian	Dekat	Berpotensi
P77	Sedikit	Mingguan	Dekat	Berpotensi
P78	Banyak	Harian	Jauh	Berpotensi
P79	Banyak	Harian	Jauh	Berpotensi
P80	Banyak	Harian	Dekat	Berpotensi
P81	Sedikit	Harian	Dekat	Berpotensi
P82	Sangat Banyak	Harian	Dekat	Berpotensi
P83	Sedikit	Harian	Jauh	Berpotensi
P84	Sangat Banyak	Harian	Dekat	Berpotensi
P85	Banyak	Harian	Dekat	Berpotensi
P86	Sangat Banyak	Harian	Dekat	Berpotensi
P87	Banyak	Harian	Dekat	Berpotensi
P88	Sangat Sedikit	Mingguan	Jauh	Tidak Berpotensi
P89	Banyak	Harian	Jauh	Berpotensi
P90	Sedikit	Mingguan	Dekat	Berpotensi
P91	Sedikit	Mingguan	Dekat	Berpotensi
P92	Sangat Sedikit	Harian	Dekat	Tidak Berpotensi
P93	Sangat Sedikit	Harian	Dekat	Tidak Berpotensi
P94	Sangat Sedikit	Mingguan	Dekat	Tidak Berpotensi
P95	Sangat Banyak	Harian	Dekat	Berpotensi
P96	Banyak	Harian	Dekat	Berpotensi
P97	Sedikit	Mingguan	Dekat	Berpotensi
P98	Banyak	Harian	Jauh	Berpotensi
P99	Banyak	Harian	Jauh	Berpotensi
P100	Banyak	Harian	Dekat	Berpotensi

Hasil klasifikasi yang disajikan pada Tabel 8, maka selanjutnya dapat dikonversikan menggunakan metode *confusion matrix* seperti bawah ini :

Tabel 8. Pengujian Confusion Matrix

Confusion Matrix	Data Sebenarnya		
	TRUE	FALSE	
Data Prediksi	TRUE	21	0
	FALSE	2	2

Dari tabel *confusion matrix* diatas maka kinerja dari penggunaan metode naive bayes dapat diukur dengan menghitung nilai *precision*, *recall*, *accuracy*. Dalam perhitungan ini sekaligus dikonversikan kedalam bentuk presentase.

a). Maka untuk menghitung nilai *precision* menggunakan rumus (1), hasil perhitungan sebagai berikut :

$$\text{Nilai Precision} = 21 / (21+0) = 100\%$$

b). Maka untuk menghitung nilai *recall* menggunakan rumus (2.1), hasil perhitungan sebagai berikut :

$$\text{Nilai Recall} = 21 / (21+2) = 91\%$$

c). Maka untuk menghitung nilai *accuracy* menggunakan rumus (2.1), hasil perhitungan sebagai berikut :

$$\text{Nilai Accuracy} = 21+2 / (21+2+0+2) = 92\%$$

V. PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian tentang klasifikasi pelanggan menggunakan metode naive bayes dapat ditarik kesimpulan dari beberapa kesimpulan sebagai berikut:

1. Penerapan metode naive bayes untuk mengklasifikasikan pelanggan dapat membantu pemilik memberikan bonus terhadap pelanggan berpotensi dan meningkat kualitas yang lebih baik lagi terhadap pelanggan.
2. Aplikasi metode naive bayes untuk mengklasifikasikan pelanggan dirancang dengan menggunakan aplikasin PHP dan MySQL.
3. Berdasarkan master pelanggan yang dijadikan data latih, telah berhasil mengklasifikasikan 23 data dari 25 data yang diuji. Sehingga berhasil memprediksi pelanggannya dengan nilai *precision* mencapai 100%, nilai *recall* mencapai 91%, nilai *accuracy* mencapai 92%.

5.2 Saran

Saran yang dapat dituliskan untuk penelitian ini adalah sebagai berikut: Sistem yang dibuat masih cukup sederhana karena mengutamakan proses perhitungan dan hasil perhitungan sehingga diharapkan sistem dikembangkan lebih detail.

DAFTAR PUSTAKA

- [1] M. Ridwan, H. Suyono and M. Sarosa, "Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa menggunakan Algoritma Naïve Bayes Classifier," *EECCIS*, vol. VII, 2013.
- [2] W. Muslehatin, M. Ibnu and Mustakim. , "Penerapan Naïve Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau," 2017.
- [3] H. Muhamad, C. Prasojo, N. Sugianto, L. Surtiningsih and I. Cholissodin, "Optimasi Naïve Bayes Classifier dengan Menggunakan Particle Swarm Optimization pada Data Iris," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. IV, pp. 180-184, 2017.
- [4] E. Manulu, F. Sianturi and M. Manulu, "Penerapan Algoritma Naïve Bayes untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan PASTRIES," *Jurnal Mantik Penusa*, vol. I, 2017.
- [5] T. Rosandy, "Perbandingan Metode Naïve Bayes Classifier dengan Metode Decision Tree (C4.5) untuk Menganalisa Kelancaran Pembiayaan (Studi Kasus : KSPPS/BMT Al-Fadhila)," *Jurnal TIM Darmajaya*, vol. II, p. 52, 2016.
- [6] Vlandari, *Data Mining Teori dan Aplikasi Rapidminer*, Yogyakarta: Gava Media, 2017.
- [7] E. Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*, Yogyakarta: ANDI , 2012.
- [8] Anhar. *Panduan Menguasai PHP & MySQL Secara Otodidak*, jakarta: Mediakita, 2010.
- [9] S. Faisal, *Aplikasi Berbasis Web dengan PHP & MySQL*, Yogyakarta: Ram Media, 2011.
- [10] A. Saleh, "Implementasi Metode Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga,"

Citec Journal, 2015.

- [11] O. Nurdiawan and N. Salim, "Penerapan Data Mining Pada Penjualan Barang Menggunakan Metode Naïve Bayes Classifier Untuk Optimasi Strategi Pemasaran," *Jurnal Teknologi Informasi dan Komunikasi*, 2018.
- [12] D. Dahri, F. Agus and D. Khairina, "Metode Naïve Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman," *Jurnal Informatika Mulawarman*, 2016.
- [13] D. Fithri, "Model Data Mining Dalam Penentuan Kelayakan Pemilihan Tempat Tinggal Menggunakan Metode Naïve Bayes," *Jurnal SIMETRIS*, 2016.
- [14] C. Fadlan, S. Ningsih and A. Windarto, "Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra," *JUTIM*, 2018.
- [15] Kusri. and E. Luthfi, *Algoritma Data Mining*, Yogyakarta: ANDI, 2019.