

**KLASIFIKASI KARAKTERISTIK KONSUMEN  
SEPEDA MOTOR MERK T DI JAWA BARAT MENGGUNAKAN  
METODE NAÏVE BAYES CLASSIFIER PADA DATA MINING**

**Jaka Aulia Pratama**

Departemen Statistika, Universitas Padjadjaran  
[jakajek@gmail.com](mailto:jakajek@gmail.com)

**Zulhanif**

Departemen Statistika, Universitas Padjadjaran

**Yadi Suprijadi**

Departemen Statistika, Universitas Padjadjaran

**ABSTRACT.** PT. JKL has a role as a main dealer of T's brand are handling three types of motorcycle products in West Java. These are type of Sport, CUB, and Scooter (Automatic Transmissions). The company records the buyer of T's brand motorcycle in the Customer Database (CDB). CDB collected from 2011 to 2013 yielded information of consumer characteristics which is necessary in market planning. Consumer characteristics are classified into two groups: Repeated Order and New Customer. Classification methods used in the study of Data Mining is the Naïve Bayes Classifier. Model classification is done by calculating the conditional probability to choose the greatest value of probability. The accuracy of the classification is 83% and the error classification is 17%.

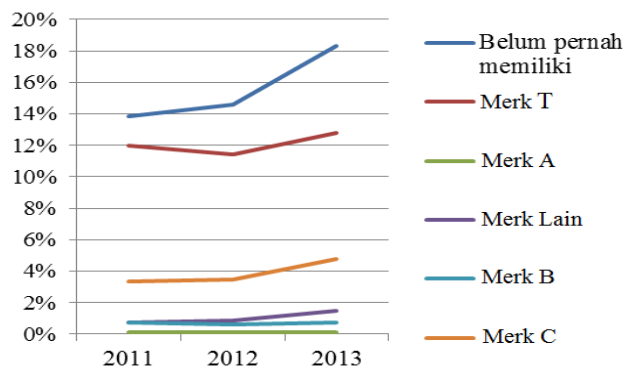
**Keywords:** Database, Data Mining, Classification, Naïve Bayes Classifier, Conditional Probability

**ABSTRAK.** PT. JKL berperan sebagai main dealer sepeda motor merk T menangani tiga jenis produk sepeda motor yaitu jenis Sport, CUB (Bebek), dan Skuter (*Matic*) di Jawa Barat. Perusahaan ini mencatat konsumen yang membeli produk sepeda motor merk T dalam *Customer Database* (CDB). CDB yang dihimpun dari tahun 2011 sampai dengan tahun 2013 menghasilkan informasi karakteristik konsumen yang diperlukan dalam perencanaan pemasaran. Karakteristik konsumen diklasifikasikan menjadi dua yaitu *Repeat Order* dan *New Customer*. Metode klasifikasi yang digunakan dalam penelitian Data Mining ini adalah *Naïve Bayes Classifier*. Model pengklasifikasian dilakukan dengan cara menghitung peluang bersyarat dengan memilih nilai peluang yang paling besar. Ketepatan dalam klasifikasi adalah sebesar 83% dan kesalahan klasifikasi adalah sebesar 17%.

**Kata Kunci:** Database, Data Mining, Klasifikasi, *Naïve Bayes Classifier*, Peluang Bersyarat

## 1. PENDAHULUAN

PT. JKL adalah perusahaan yang bergerak dalam bidang distribusi sepeda motor dan suku cadang sepeda motor merk T yang menangani tiga jenis produk sepeda motor yaitu jenis *Sport*, CUB (Bebek), dan AT (*Automatic*) di 18 area *dealer-dealer* cabang di Jawa Barat. PT. JKL mencatat data konsumen yang membeli produk sepeda motor merk T dalam *database* yang selanjutnya dinamakan *Customer Database* (CDB). Salah satu pemanfaatan *Customer Database* yaitu untuk menghasilkan informasi atau pengetahuan mengenai karakteristik konsumen sepeda motor merk T yang selanjutnya sangat diperlukan dalam rencana pemasaran. Karakteristik tersebut akan bermanfaat jika berlandaskan kepada hasil pengklasifikasian terhadap karakteristik konsumen *Repeat Order* dan *New Customer*. Untuk mendapatkan informasi dari *Customer Database* yang menunjukkan apakah konsumen tersebut *Repeat Order* atau *New Customer*, dapat dilihat dari merk sepeda motor yang dimiliki sebelumnya. Dalam *Customer Database* karakteristik kepemilikan sepeda motor sebelumnya dibagi menjadi merk T, merk A, merk motor lain, merk B, merk C dan belum pernah memiliki. Konsumen dengan merk sepeda motor yang dimiliki sebelumnya adalah merk T disebut sebagai konsumen dengan karakteristik *Repeat Order*, sedangkan konsumen dengan karakteristik kepemilikan sepeda motor bukan merk T dan belum pernah memiliki sepeda motor merupakan konsumen dengan karakteristik *New Customer*. Persentase kepemilikan sepeda motor sebelumnya dari konsumen PT. JKL tahun 2011 – 2013 ditunjukkan pada Gambar 1.



**Gambar 1.** Persentase Berdasarkan Kategori Kepemilikan Sepeda Motor Sebelumnya Konsumen PT. JKL tahun 2011 – 2013

Gambar 1 menunjukkan bahwa tahun 2011 – 2013 konsumen dengan kepemilikan sepeda motor sebelumnya merk T dan konsumen yang belum pernah memiliki sepeda motor menjadi pangsa pasar utama PT. JKL, sedangkan persentase dari konsumen dengan kepemilikan sepeda motor sebelumnya merk C, merk motor lain, merk B dan merk A termasuk rendah yaitu dibawah 6% dari total penjualan selama tiga tahun tersebut. Oleh karena itu diperlukan strategi pemasaran melalui klasifikasi karakteristik konsumennya yang akan diambil dari *Customer Database* tahun 2011 – 2013.

Salah satu bahan yang diperlukan dalam membuat strategi pemasaran adalah klasifikasi karakteristik konsumen *Repeat Order* dan *New Customer* sepeda motor merk T. Oleh karena itu yang menjadi permasalahan dalam penelitian ini adalah, bagaimana membuat klasifikasi tersebut dengan melibatkan nilai peluang pada setiap klasifikasinya apabila data yang akan diolah memiliki volume yang besar dan tersimpan dalam *database*.

Tujuan dari penelitian ini adalah, mengaplikasikan teknik *Data Mining* dalam pengklasifikasian dengan metode *Naïve Bayesian Classifier* pada *Customer Database* PT. JKL. Hasil penelitian berupa klasifikasi karakteristik konsumen *Repeat Order* dan *New Customer* sepeda motor merk T di Jawa Barat diharapkan dapat membantu PT. JKL dalam menyusun strategi pemasaran yang lebih efektif kepada calon konsumen sepeda motor merk T di Jawa Barat.

## 2. METODE PENELITIAN

### 2.1 *Data Mining*

*Data Mining* adalah proses seleksi, eksplorasi, dan pemodelan data dalam jumlah besar untuk menemukan pola atau relasi yang ada dengan tujuan memperoleh hasil yang jelas dan berguna bagi pemilik *database* (Giudici, 2003). Pada *Data Mining*, terdapat dua pendekatan yang digunakan yaitu pendekatan statistika klasik dan metode kecerdasan buatan. Pendekatan statistika klasik yang dimaksud adalah data dieksplorasi dan dianalisis melalui metode-metode statistika.

Pada umumnya, *Data Mining* digunakan untuk tujuan prediksi dan deteksi. Prediksi tidak dipergunakan dalam tugas *Data Mining* pada penelitian ini karena tujuan dari penelitian ini bukan untuk melakukan peramalan yang umumnya menggunakan regresi demikian pula dengan asosiasi yang pada umumnya diaplikasikan dalam hal penentuan antar hubungan. Penelitian ini juga tidak mempunyai tujuan untuk mendeteksi kejadian yang tidak biasa (deteksi). Permasalahan dalam penelitian ini adalah menelaah karakteristik konsumen yang dapat digolongkan kepada kategori persoalan klasifikasi.

## 2.2 *Naïve Bayes Classifier*

*Naïve Bayes Classifier* adalah metode klasifikasi dalam *Data Mining* berdasarkan probabilitas dan teorema *Bayes*, dengan asumsi bahwa setiap variabel bersifat bebas (*independent*). Asumsi tersebut akan menghilangkan kebutuhan banyaknya jumlah data latih (*Training Data*) dari perkalian Kartesius seluruh variabel yang dibutuhkan untuk mengklasifikasikan suatu data (Berson dkk., 2001). Pada artikelnya Domingos dan Pazzani (1997) dijelaskan performa *Naïve Bayes Classifier* dalam fungsi *zero-one loss* yaitu fungsi yang mendefinisikan *error* hanya sebagai pengklasifikasian yang salah. Tidak seperti fungsi *error* yang lain misalnya *squared error*, fungsi *zero-one loss* tidak memberi nilai suatu kesalahan perhitungan peluang selama peluang maksimum ditugaskan ke dalam kelas yang benar. Ini berarti bahwa *Naïve Bayes Classifier* dapat mengubah peluang *posterior* dari tiap kelas, tetapi kelas dengan nilai peluang *posterior* maksimum jarang diubah. Berikut adalah algoritma *Naïve Bayes Classifier* dengan asumsi semua atribut independen:

$$\begin{aligned}
 P(X_1, X_2, \dots, X_n | Y) &= P(X_1 | X_2, \dots, X_n, C) P(X_2, \dots, X_n | Y) \\
 &= P(X_1 | C) P(X_2, \dots, X_n | Y) \\
 &= P(X_1 | C) P(X_2 | C) \dots P(X_n | Y)
 \end{aligned} \tag{1}$$

## 2.3 Penurunan Algoritma *Naïve Bayes*

Menurut Mitchell (2010), algoritma *Naïve Bayes* merupakan algoritma klasifikasi yang didasarkan pada aturan *Bayes*. Asumsi yang dipergunakan adalah

bahwa atribut  $X_1, \dots, X_n$  independen satu dengan yang lainnya yang diberikan oleh  $Y$ . apabila diberikan  $X = (X_1, X_2, X_3)$ , maka langkah untuk mengestimasi  $P(X | Y)$  adalah sebagai berikut:

$$\begin{aligned} P(X | Y) &= P(X_1, X_2, X_3 | Y) \\ &= P(X_1 | Y) P(X_2 | Y, X_1) P(X_3 | Y, X_1, X_2) \end{aligned} \quad (2)$$

Menurut Novita (2014), karena asumsi independen bersyarat memegang peranan maka dianggap bahwa untuk setiap atribut  $X_i$  terkondisi saling independen terhadap setiap atribut  $X_j$  untuk  $i \neq j$ . Apabila asumsi independen bersyarat dipergunakan, maka Persamaan (2) dapat dituliskan menjadi:

$$P(X_1, X_2, X_3 | Y) = P(X_1 | Y) P(X_2 | Y) P(X_3 | Y) \quad (3)$$

Apabila  $X$  mengandung  $n$  atribut yang bersifat independen bersyarat antara satu dengan lainnya yang diberikan oleh  $Y$ , secara umum dapat dituliskan sebagai berikut:

$$P(X_1, \dots, X_n | Y) = \prod_{i=1}^n P(X_i | Y) \quad (4)$$

Secara umum,  $X_1, \dots, X_n$  dan  $Y$  adalah variable diskrit. Tujuan dari metode ini adalah untuk melatih sebuah *classifier* yang akan menghasilkan *output* berupa probabilitas atas kemungkinan nilai  $Y$ , untuk setiap  $X$  baru yang akan diklasifikasikan. Pernyataan untuk peluang  $Y$  akan diambil nilai ke- $k$  yang mungkin sesuai dengan aturan *Bayes*, dapat dituliskan sebagai berikut:

$$P(Y = y_k | X_1, \dots, X_n) = \frac{P(Y = y_k) P(X_1, \dots, X_n | Y = y_k)}{\sum_j P(Y = y_j) P(X_1, \dots, X_n | Y = y_j)} \quad (5)$$

Pada penyebut terlihat bahwa jumlah yang dimaksud adalah pengambilan setiap nilai  $y_j$  dari  $Y$ . apabila diasumsikan bahwa untuk setiap atribut  $X_i$  adalah independen bersyarat yang diberikan oleh  $Y$ , Persamaan (5) dapat ditulis kembali sebagai berikut:

$$P(Y = y_k | X_1, \dots, X_n) = \frac{P(Y = y_k) \prod_{i=1}^n P(X_i | Y = y_k)}{\sum_j P(Y = y_j) \prod_{i=1}^n P(X_i | Y = y_j)} \quad (6)$$

Persamaan (6) merupakan persamaan dasar untuk metode *Naïve Bayes*. Menurut Novita (2014) apabila diberikan nilai  $X' = (X_1, \dots, X_n)$  yang baru, persamaan tersebut menunjukkan bagaimana cara untuk menghitung probabilitas bahwa  $Y$  akan terjadi untuk setiap atribut  $X'$  dan diberikan distribusi  $P(Y)$  dan  $P(X_i | Y)$  yang diestimasi berdasarkan data *training*. Dalam menentukan peluang  $X'$  menjadi kategori dari kelas  $Y$ , maka aturan klasifikasi dalam *Naïve Bayes* dapat dituliskan sebagai berikut:

$$Y \leftarrow \arg \max_{y_k} \frac{P(Y = y_k) \prod_{i=1}^n P(X_i | Y = y_k)}{\sum_j P(Y = y_j) \prod_{i=1}^n P(X_i | Y = y_j)} \quad (7)$$

Aturan keputusan dalam klasifikasi *Naïve Bayes* sering disebut dengan aturan keputusan *Maximum A Posteriori* (MAP) karena pada saat klasifikasi, pendekatan *Bayes* akan menghasilkan label kategori yang paling tinggi probabilitasnya. Pada Persamaan (7) terlihat bahwa penyebut tidak bergantung pada  $y_k$ , maka Persamaan (7) dapat disederhanakan menjadi:

$$Y \leftarrow \arg \max_{y_k} P(Y = y_k) \prod_{i=1}^n P(X_i | Y = y_k) \quad (8)$$

## 2.4 Ketepatan Klasifikasi

Pada penelitian ini, ketepatan klasifikasi diukur oleh *Apparent Error Rate* (APER). Untuk menghitung nilai APER beberapa prosedur menyarankan untuk membagi total sampel yang digunakan kedalam dua kelompok, yaitu *Analysis Sample (Data Training)* dan *Holdout Sample (Data Testing)*. Pada kelompok *Analysis Sample* digunakan untuk membuat model klasifikasi, sedangkan pada kelompok *Holdout Sample* digunakan untuk menguji ketepatan klasifikasi yang dilakukan (Hair, et al, 1998).

APER dihitung dengan terlebih dahulu dengan membuat tabel klasifikasi seperti pada Tabel 1:

**Tabel 1.** Klasifikasi Untuk *Actual Group* dan *Predicted Group*

<i>Actual Group</i>	<i>Predicted Group</i>	
	$\pi_1$	$\pi_2$
$\pi_1$	$n_{11}$	$n_{12}$
$\pi_2$	$n_{21}$	$n_{22}$

$n_{11}$  = jumlah pengamatan dari  $\pi_1$  tepat diklasifikasikan sebagai  $\pi_1$

$n_{12}$  = jumlah pengamatan dari  $\pi_1$  tepat diklasifikasikan sebagai  $\pi_2$

$n_{21}$  = jumlah pengamatan dari  $\pi_2$  tepat diklasifikasikan sebagai  $\pi_1$

$n_{22}$  = jumlah pengamatan dari  $\pi_2$  tepat diklasifikasikan sebagai  $\pi_2$

$$APER = \left( \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}} \right) \times 100\% \quad (9)$$

## 2.5 Langkah Penyelesaian *Naïve Bayes*

Langkah-langkah penyelesaian pada metode *Naïve Bayes* adalah sebagai berikut:

- 1) Bagi data menjadi 2 bagian, yaitu data *Training* dan data *Testing*
- 2) Bentuk model dengan *Data Training*, menurut Zhang J., et al. (2007), dalam memilih pembagian *Data Training* dan *Data Testing* dapat berdasarkan 90% berbanding 10%, 80% berbanding 20%, dan 70% berbanding 30%. Dalam penelitian ini, perbandingan data *Training* dan data *Testing* yang digunakan sebesar 80% berbanding 20%
- 3) Evaluasi seberapa tepat klasifikasi dilakukan dari *Data Training* dan *Data Testing*
- 4) Buatlah model klasifikasinya.

## 2.7 Variabel Penelitian

Variabel-variabel yang terlibat dalam penelitian ini adalah :

a) Variabel Dependen

Variabel dependen dalam penelitian ini yaitu merk sepeda motor sebelumnya meliputi merk T, merk C, merk B, merk A, merk lain dan konsumen yang belum pernah memiliki sepeda motor. Variabel dependen untuk karakter *Repeat Order* yaitu konsumen dengan sepeda motor sebelumnya merk T. Sedangkan Variabel dependen untuk karakter *New Customer* yaitu konsumen dengan merk sepeda motor sebelumnya merk C, merk B, merk A.

b) Variabel independen pada penelitian ini terdiri dari tujuh variabel, yaitu:

- 1)  $X_1$  merupakan jenis sepeda motor yang dibeli (Bebek, *Automatic* dan *Sport*)
- 2)  $X_2$  merupakan pekerjaan konsumen yang diamati. Terdiri dari sebelas jenis yaitu, pegawai negeri, pegawai swasta, ojek, wiraswasta/pedagang, mahasiswa/ pelajar, guru/dosen, TNI/Polri, ibu rumah tangga, petani/nelayan, profesional (dokter/pengacara, dll), dan lain-lain.
- 3)  $X_3$  merupakan besarnya pengeluaran konsumen dalam satu bulan yang terdiri dari 13 jenis yaitu, <Rp.700.000, Rp. 600.001 s/d Rp. 900.000, Rp.700.001 s/d Rp.1.000.000, Rp.1.000.001 s/d Rp.1.500.000, Rp. 1.250.001 s/d Rp. 1.750.000, Rp. 1.500.001 s/d Rp. 2.000.000, Rp. 1.750.000 s/d Rp.2.000.000, Rp.2.000.001 s/d Rp.3.000.000, Rp.2.500.001 s/d Rp.3.500.000 Rp.3.000.001 s/d Rp.4.000.000, >Rp. 3.500.000, >Rp.4.000.000.
- 4)  $X_4$  merupakan jenis sepeda motor sebelumnya (bebek, automatic dan sport)
- 5)  $X_5$  merupakan pemakai sepeda motor yang diamati. Terdiri dari empat jenis yaitu, saya sendiri, anak, pasangan (suami/ istri), dan lain-lain.
- 6)  $X_6$  merupakan kota dimana konsumen membeli sepeda motor merk T. Terdiri dari 18 kota yaitu, Bandung, Bekasi, Bogor, Ciamis, Cianjur, Cikarang, Cirebon, Depok, Garut, Indramayu, Karawang, Kuningan,



Majalengka, Purwakarta, Subang, Sukabumi, Sumedang, dan Tasikmalaya.

- 7)  $X_7$  merupakan tahun pembelian sepeda motor (2011, 2012, dan 2013).

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil Klasifikasi

**Tabel 2.** Karakteristik Konsumen Sepeda Motor Merk T di Jawa Barat

T Variabel	Karakteristik Konsumen	
	<i>New Customer</i>	<i>Repeat Order</i>
Jenis Sepeda Motor Yang Dibeli	AT ( <i>Automatic</i> )	AT ( <i>Automatic</i> )
Pekerjaan Konsumen	Pegawai Swasta	Pegawai Swasta dan Wiraswasta/Pedagang
Pengeluaran Konsumen Dalam Satu Bulan	Rp. 1.500.001 s/d Rp. 2.000.000	Rp. 1.500.001 s/d Rp. 2.000.000
Jenis Sepeda Motor Sebelumnya	Belum Pernah Memiliki	CUB (Bebek)
Pemakai Sepeda Motor	Sendiri	Sendiri
Kota	Bekasi	Bekasi
Tahun Pembelian	2013	2013

Tabel 2 menunjukkan hasil klasifikasi karakteristik konsumen *New Customer* sepeda motor merk T yaitu jenis sepeda motor yang dibeli adalah AT (*Automatic*), bekerja sebagai pegawai swasta, pengeluaran per-bulannya Rp. 1.500.001 s/d Rp. 2.000.000, sebelumnya tidak memiliki sepeda motor, sepeda motor yang dibeli untuk digunakan sendiri, berdomisili di kota Bekasi, membeli sepeda motor tersebut pada tahun 2013. Sebaliknya, klasifikasi karakteristik pada konsumen *Repeat Order* sepeda motor merk T yaitu jenis sepeda motor yang dibelinya adalah AT (*Automatic*), bekerja sebagai pegawai swasta atau wirausahawan/pedagang, pengeluaran per-bulannya Rp. 1.500.001 s/d Rp. 2.000.000, sebelumnya memiliki sepeda motor berjenis CUB (Bebek), sepeda motor yang dibeli untuk digunakan sendiri, berdomisili di kota Bekasi, membeli sepeda motor tersebut pada tahun 2013.

Dari hasil klasifikasi karakteristik konsumen sepeda motor merk T di Jawa Barat tersebut terlihat bahwa sepeda motor dengan jenis AT (*Automatic*) sangat

diminati oleh masyarakat dengan pengeluaran per-bulan Rp. 1.500.001 s/d Rp. 2.000.000 dan sepeda motor tersebut akan digunakan oleh sendiri. PT. JKL mencatat pada tahun 2013 penjualan tertinggi sepeda motor merk T di Jawa Barat terdapat di Kota Bekasi.

### 3.2 Model Klasifikasi

Model-model yang digunakan dalam penelitian ini adalah :

- 1) Model Klasifikasi *New Customer* (NC)

$$P(NC | X_1, X_2, \dots, X_7) = P(X_1 | NC) \cdot P(X_2 | NC) \cdot P(X_3 | NC) \cdot P(X_4 | NC) \\ \cdot P(X_5 | NC) \cdot P(X_6 | NC) \cdot P(X_7 | NC) \cdot P(NC)$$

- 2) Model Klasifikasi *Repeat Order* (RO)

$$P(RO | X_1, X_2, \dots, X_7) = P(X_1 | RO) \cdot P(X_2 | RO) \cdot P(X_3 | RO) \cdot P(X_4 | RO) \\ \cdot P(X_5 | RO) \cdot P(X_6 | RO) \cdot P(X_7 | RO) \cdot P(RO)$$

Peluang *prior* untuk konsumen dengan karakteristik *New Customer* sebesar 0,634 dan peluang *prior* untuk konsumen dengan karakteristik *Repeat Order* sebesar 0,365.

### 3.3 Ketepatan Klasifikasi

**Tabel 3.** Klasifikasi Untuk Actual Group dan Predicted Group

<i>Actual Group</i>	<i>Predicted Group</i>	
	<i>New Customer</i>	<i>Repeat Order</i>
<i>New Customer</i>	219752	68312
<i>Repeat Order</i>	7850	158145

$$APER = \left( \frac{68312 + 7850}{219752 + 7850 + 68312 + 158145} \right) = 0,167 \times 100\%$$

Hasil perhitungan APER menunjukkan bahwa kesalahan klasifikasi karakteristik *New Customer* dan *Repeat Order* konsumen sepeda motor merk T untuk data *Testing* sebesar  $(0,167 \times 100\%) = 16,77\%$  dengan ketepatan klasifikasi

sebesar  $(1 - 0.167) \times 100\% = 83,23\%$ . Hasil tersebut menunjukkan bahwa ketepatan klasifikasi pada penelitian ini cukup tinggi.

#### 4. KESIMPULAN DAN SARAN

Dengan menggunakan metode *Naïve Bayes Classifier* diperoleh peluang *prior* untuk konsumen dengan karakteristik *New Customer* sebesar 0,634 dan peluang *prior* untuk konsumen dengan karakteristik *Repeat Order* sebesar 0,365. Untuk pengklasifikasian karakteristik konsumen sepeda motor merk T menunjukkan bahwa konsumen yang menggunakan sepeda motornya sendiri dan tidak pernah memiliki sepeda motor sebelumnya memiliki peluang lebih dari 0,65 untuk diklasifikasikan sebagai konsumen dengan karakteristik *New Customer*. Sedangkan untuk konsumen yang membeli sepeda motor merk T jenis AT (*Automatic*), pernah memiliki sepeda motor jenis CUB (Bebek), dan menggunakan sepeda motornya sendiri memiliki peluang di atas 0,65 untuk diklasifikasikan sebagai konsumen dengan karakteristik *Repeat Order*. Hasil pengklasifikasian tersebut memiliki kekeliruan klasifikasi sebesar 0,167.

Berdasarkan hasil evaluasi yang telah dilakukan pada penelitian ini, peneliti mengemukakan beberapa saran sebagai berikut:

- 1) Keberhasilan dari suatu penelitian data mining sangat bergantung kepada database. Artinya *knowledge discovery* dari penelitian *Data Mining* dilandasi oleh tidak hanya kualitas dan kuantitas data tetapi juga berbagai format yang digunakan (Olson dan Shi, 2007). Penulis menyarankan khususnya di dunia bisnis atau industri, pembangunan database betul-betul dirancang sedemikian rupa sehingga memudahkan untuk ditindaklanjuti oleh aplikasi *Data Mining*.
- 2) Agar memperoleh hasil klasifikasi yang lebih mendetail penulis menyarankan untuk menambahkan variabel prediktor pada penelitian selanjutnya.

#### DAFTAR PUSTAKA

Berson A., Smith, S.J., dan Thearling, K. *Building Data Mining Applications for CRM (1st Edition)*, USA: McGraw-Hill Companies, Inc., 2001.

- Domingos, P. dan Pazzani, M., *On the Optimality of the Simple Bayesian Classifier Under Zero-one loss*, Machine Learning, 1997.
- Giudici, P., *Applied Data Mining: Statistical Methods for Business and Industry*, 2003.
- Hair, J. F., *Multivariate Data Analysis*, Prentice Hall, Upper Saddle River, N. J., 1998.
- Mitchell, T. M., *Machine Learning*, McGraw-Hill Companies, Inc., USA, 2010.
- Novita, E., *Penerapan Analisis Sentimen Dengan Metode Naïve Bayes pada Klasifikasi Data Teks*, Skripsi, Departemen Statistika FMIPA Universitas Padjadjaran, Bandung, 2014.
- Olson, D. and Shi, Y., *Introduction to Business Data Mining*. McGraw-Hill Companies Inc., Singapore., 2007.
- Zhang, Y.-C., Rossow, W.B., Stackhouse, P. W., Romanou, A., dan Wielicki, B. A., *Decadal Variations of Global Energy and Ocean Heat Budget and Meridional Energy Transports Inferred from Recent Global Data Sets*, J. Geophys. Res., D22101, 2010.
- Zulhanif. *Analisis Credit Scoring dengan Bayesian Klasifikasi*, Prosiding Seminar Nasional Statistika : Peranan Statistika di Bidang Pemasaran dalam Penyusunan Strategi Bisnis, Universitas Padjadjaran, Bandung, 2012.