

DETEKSI MINAT BELI PELANGGAN TERHADAP PRODUK PAKET INTERNET MENGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER

Ahmad Syahrul Mubarak

Program Studi S1 Teknik Informatika, Fakultas Teknologi Informasi
Universitas Hasyim Asy'ari Jombang
ahmadsyahrullft9@gmail.com

Dedy Rahman Prehanto

Fakultas Teknologi Informasi, Universitas Hasyim Asy'ari Tebuireng Jombang
dedyrpr@yahoo.com

Mahrus Ali, S.Kom, M.Pd

Fakultas Teknologi Informasi, Universitas Hasyim Asy'ari Tebuireng Jombang
macus.alay@gmail.com

Abstrak

Kebutuhan akan teknologi komunikasi yang dapat mengetahui minat beli pelanggan merupakan sesuatu yang dibutuhkan perusahaan, tujuannya yaitu supaya perusahaan dapat memperoleh pertimbangan mengenai penjualan produk dan dengan tepat mengambil kebijakan mengenai penentuan harga terhadap suatu produk, besarnya suplai produk dan sebagainya. Metode naïve bayes classifier dirasa cukup untuk dapat mengetahui seberapa besar minat beli pelanggan terhadap produk paket internet dengan menganalisa hasil dari penjualan di masa lampau, artinya algoritma naïve bayes dapat memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya, dengan mendefinisikan masing-masing kelas dari seluruh atribut. Klasifikasi naïve bayes diasumsikan bahwa ada atau tidak ciri tertentu dan sebuah kelas tidak ada hubungannya dengan ciri dan kelas lainnya. Penelitian ini mendefinisikan pada beberapa ciri yakni operator, kuota, masa aktif dan harga produk. Dari hasil penelitian ini adalah penulis dapat mencari nilai keakuratan dari hasil perhitungan minat dan tidak dengan cara membandingkan hasil prediksi dengan hasil yang sebenarnya diluar data training. Berdasarkan data penjualan produk paket internet yang dijadikan data training, metode naïve bayes berhasil mengklasifikasikan 10 data produk dengan 858 data transaksi sebagai data training dan 10 data produk dengan 115 data transaksi sebagai data uji. Dari analisa seluruh data uji dan data training, perbandingan kebenarannya adalah 8/10, yakni 8 dari 10 produk dengan nilai prediksi yang bernilai benar. Sehingga nilai akurasi sebesar 80% dan untuk mengetahui minat beli pelanggan dengan memkalkulasi dari data penjualan yang telah lampau dengan ketentuan dan sesuai dengan kelas (label) yg ditetapkan, dengan begitu hasil dari prediksi ini diharapkan dapat menjadi pertimbangan kebijakan yang lebih baik bagi usaha penjualannya..

Kata Kunci: *naïve bayes classifier*, penjualan produk paket internet.

Abstract

The need for communication technology that can find out the customer's buying interest is something that is needed by the company, the goal is that the company can obtain consideration regarding product sales and appropriately take a policy regarding pricing of a product, the amount of product supply and so on. The naïve bayes classifier method is sufficient to be able to find out how much the customer's buying interest in an internet package product is by analyzing the results of past sales, meaning that the Naive Bayes algorithm can predict future opportunities based on past experience, by defining each class from all attribute. The naïve bayes classification is assumed that there is or not a certain feature and a class has nothing to do with other characteristics and classes. This study defines several characteristics, namely operator, quota, active period and product price. From the results of this study is the author can find the value of accuracy from the results of calculating interest and not by comparing the results of predictions with actual results outside the training data. Based on data on internet package product sales as training data, the Naive Bayes method successfully classifies 10 product data with 858 transaction data as training data and 10 product data with 115 transaction data as test data. From the analysis of all test data and training data, the truth ratio is 8/10, which is 8 out of 10 products with predictive values that are of true value. So that the accuracy value is 80% and to find out the customer's buying interest by calculating the sales data that has been past with the provisions and in accordance with the specified label, the results of this prediction are expected to be a better policy consideration for the sales business.

Keywords: *naïve bayes classifier*, sales of internet package products.

PENDAHULUAN

Persaingan dalam dunia usaha mendorong kepada para pelaku usaha untuk terus melakukan terobosan demi kelangsungan bisnis mereka. Saat ini perusahaan usaha Abadi cell tersebut telah menawarkan berbagai jenis paket internet dengan berbagai macam jenis dan harga, namun tanpa mengetahui besarnya minat konsumen terhadap paket tersebut. Hal ini tentu akan membuat konsumen beralih kepada perusahaan sejenis dengan paket yang dirasa konsumen paling menguntungkannya.

Metode klasifikasi ditujukan untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data untuk dapat memperkirakan kelas dan suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Algoritma *naive bayes classifier* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. Penerapan klasifikasi berupa algoritma *naive bayes classifier* diharapkan dapat memprediksi atau mendeteksi dini potensi minat beli pelanggan berdasarkan pada transaksi yang terjadi pada masa lampau. Dengan demikian perusahaan dapat lebih mengerti paket-paket penjualan mana saja yang akan diminati pelanggannya.

Secara umum, tujuan penelitian ini adalah menyajikan gambaran dan prediksi mengenai minat beli konsumen terhadap pulsa paket Internet. Adapun rincian dari tujuan penelitian ini antara lain untuk dapat mengklasifikasikan paket pulsa internet dengan algoritma *naive bayes classifier* juga untuk mengetahui sejauh mana minat beli pelanggan terhadap minat beli pelanggan pada produk paket internet. Sehingga dalam penelitian ini, dapat diambil rumusan masalahnya antara lain membuat deteksi dengan algoritma *naive bayes classifier* pada usaha penjualan paket Internet dan klasifikasi terkait deteksi minat beli pelanggan terhadap penjualan paket Internet.

METODE

Dalam program prediksi minat beli pelanggan terhadap paket internet baru, pendekatan dalam penelitian ini adalah pendekatan kuantitatif, karena penelitian ini disajikan dengan angka-angka. Menurut (Arikunto, 2002) yang mengemukakan bahwa penelitian kuantitatif merupakan penelitian yang banyak dituntut mengemukakan angka-angka, mulai dari pengumpulan data, pengolahan data, penafsiran terhadap data tersebut serta penampilan terhadap hasilnya. Penelitian ini melalui empat tahapan antara lain.

1. Pengumpulan data

Prosedur sistematis yang digunakan untuk mengumpulkan data yaitu dengan mengumpulkan data produk beserta data transaksi penjualan.

2. Praproses data

Pada tahap praproses data, dilakukan seleksi data. Pemilihan data dalam proses seleksi terdiri dari 4 atribut yang antara lain data operator paket internet, data kuota utama paket internet, masa aktif paket dan harga jual paket internet. Apabila terdapat data numerik seperti pada data kuota, harga dan masa aktif paket internet maka yang harus dilakukan adalah mencari 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.

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \text{ atau } \mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Gambar 3.1 Rumus nilai rata-rata

Keterangan :

μ = rata-rata hitung (mean)

x_i = nilai sample ke -i

n = jumlah sampel

3. Pembagian Data

Pada tahapan ini, data dibagi menjadi 2, yaitu data latihan (training data) dan data uji. Data latihan dan data uji menggunakan data produk paket internet yang dijual beserta data penjualan paket internet yang terjadi pada bulan Mei 2015 sampai Februari 2016. Dalam penelitian, metode yang digunakan dalam proses pemisahan data latihan dan data uji ini adalah dengan pembagian 75 banding 25 yang artinya 75% data pertama ditetapkan sebagai data latihan (training data) sedangkan 25% data diakhir digunakan sebagai data uji.

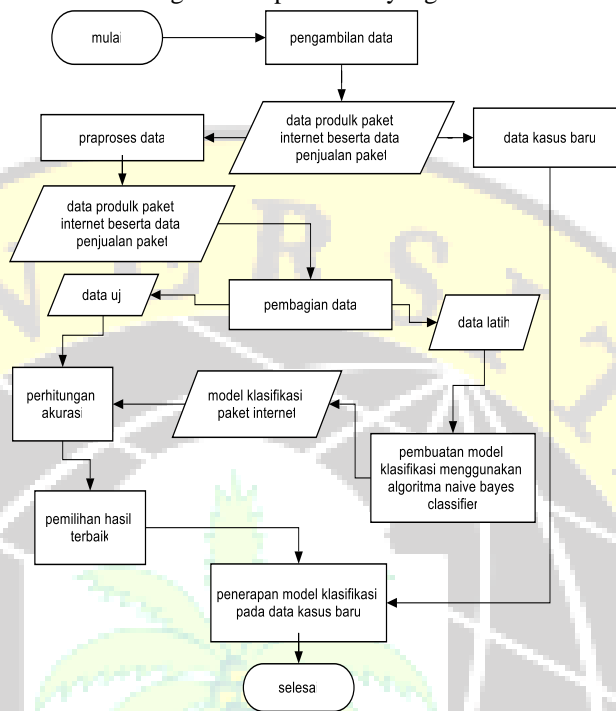
4. Membuat model klasifikasi

Dalam pembuatan model klasifikasi, diperlukan penentuan atribut. Dalam penelitian ini atribut telah ditentukan yang antara lain : operator atau penyedia paket internet, kuota paket internet, masa aktif paket internet dan

harga jual paket internet. Setelah atribut berhasil ditetapkan, maka tahap selanjutnya yakni menentukan klasifikasi dari produk-produk tersebut, dengan begitu klasifikasi sudah dapat dilakukan.

5. Penerapan menggunakan data kasus baru

Pada tahapan ini dilakukan penggunaan model pada data baru, yaitu data penjualan paket internet pada bulan Januari 2016 sampai Februari 2016. Berikut ini diagram alir penelitian yang dilakukan



Gambar 3.2 Diagram alir tahap penelitian

Adapun bahan yang dibutuhkan dari penelitian ini adalah data, yang antara lain data transaksi pembelian, data produk yang dijual beserta rincian produk tersebut. Sedangkan untuk kebutuhan lainnya dalam penelitian ini antara lain studi pustaka untuk mencari referensi lain mengenai penelitian ini, perangkat lunak untuk mengimplementasikan penelitian ini kedalam suatu aplikasi, dan perangkat kerasnya. Perangkat lunak yang digunakan dalam penelitian ini antara lain:

1. Sistem operasi Windows 10 Pro.
2. Bahasa pemrograman java (jdk1.8.0_151).
3. Netbeans IDE 8.02.
4. DB Browser for SQLite untuk mengelola database sqlite.

Perangkat keras yang digunakan dalam penelitian ini adalah laptop dengan spesifikasi:

1. Prosesor Intel Core i3 @1.8 GHz
2. RAM 2 GB
3. Harddisk 500 GB HDD

HASIL DAN PEMBAHASAN

1. Penerapan naïve bayes classifier

Naive bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai teorema bayes(Kusrini & Emha, 2009). Semakin besar jumlah data latih, semakin akurat juga hasilnya. Apabila terdapat data numerik seperti pada data kuota, harga dan masa aktif paket internet yang semua data tersebut bersifat kontinyu, maka yang harus dilakukan adalah mencari nilai *mean* dan standar deviasi dari masing-masing parameter yang merupakan data numerik.

a. Data latih

Pada penerapan naïve bayes classifier, hal pertama yang harus dilakukan adalah menentukan data latih dan data uji, berikut ini merupakan data latih dari penelitian ini :

Deteksi Minat Beli Pelanggan Terhadap Produk Paket Internet Menggunakan Algoritma *Naïve Bayes Classifier*

#	ID Barang	Nama Barang	Keterangan	Masa Aktif	Kuota Utama	Operator	Harga	Jumlah Transaksi
<input type="checkbox"/>	ID1	INDOSAT DATA 1GB		30 Hari	1024 MB	Indosat	Rp. 29.000,00	172 Transaksi
<input type="checkbox"/>	ID2	INDOSAT DATA 2GB		30 Hari	2048 MB	Indosat	Rp. 36.500,00	158 Transaksi
<input type="checkbox"/>	ID3	INDOSAT DATA 3GB		30 Hari	3584 MB	Indosat	Rp. 39.000,00	112 Transaksi
<input type="checkbox"/>	ID30	INDOSAT DATA 3.5GB		30 Hari	3584 MB	Indosat	Rp. 42.500,00	79 Transaksi
<input type="checkbox"/>	TD1	TELKOMSEL DATA 1GB		30 Hari	1024 MB	Telkomsel	Rp. 42.000,00	59 Transaksi
<input type="checkbox"/>	TD2	TELKOMSEL DATA 2 GB		30 Hari	2048 MB	Telkomsel	Rp. 51.000,00	107 Transaksi
<input type="checkbox"/>	TD5	TELKOMSEL DATA 5GB		30 Hari	5120 MB	Telkomsel	Rp. 78.000,00	31 Transaksi
<input type="checkbox"/>	XD1	XL 1GB		30 Hari	1024 MB	XL	Rp. 29.000,00	69 Transaksi
<input type="checkbox"/>	XD2	XL 2GB MASA AKTIF 2 BULAN		60 Hari	2048 MB	XL	Rp. 41.000,00	60 Transaksi
<input type="checkbox"/>	XD5	XL DATA 5GB		30 Hari	5120 MB	XL	Rp. 66.000,00	65 Transaksi
<input type="checkbox"/>		Jumlah Transaksi						912 Transaksi

Gambar 3.1 Data latihan

Langkah pertama adalah mengklasifikasikan produk yang sudah dipasarkan sehingga memperoleh kelas “diminati” atau “tidak”, dengan cara menentukan nilai rata-rata dari keseluruhan penjualan pada data sampel. Penelitian ini menggunakan total data latihan sebanyak 912 transaksi, dari jumlah tersebut dapat diambil rata-rata transaksi dengan cara total transaksi dibagi jumlah produk, yang akan menghasilkan nilai berikut :

$$\frac{\sum x}{n} = \frac{912}{10} = 91.2$$

Sehingga setiap produk yang memiliki total transaksi lebih dari atau sama dengan 91.2 dapat diberi kelas “diminati” begitu juga dengan sebaliknya mendapat label “tidak diminati” maka data sampel yang telah diklasifikasi akan menjadi sebagai berikut :

#	ID Barang	Nama Barang	Operator	Masa Aktif	Kuota Utama	Harga	Minat
<input type="checkbox"/>	ID1	INDOSAT DATA 1GB	Indosat	rendah	rendah	rendah	DIMINATI
<input type="checkbox"/>	ID2	INDOSAT DATA 2GB	Indosat	rendah	rendah	rendah	DIMINATI
<input type="checkbox"/>	ID3	INDOSAT DATA 3GB	Indosat	rendah	tinggi	tinggi	DIMINATI
<input type="checkbox"/>	ID30	INDOSAT DATA 3.5GB	Indosat	rendah	tinggi	tinggi	TIDAK DIMINATI
<input type="checkbox"/>	TD1	TELKOMSEL DATA 1GB	Telkomsel	rendah	rendah	tinggi	TIDAK DIMINATI
<input type="checkbox"/>	TD2	TELKOMSEL DATA 2 GB	Telkomsel	rendah	rendah	tinggi	DIMINATI
<input type="checkbox"/>	TD5	TELKOMSEL DATA 5GB	Telkomsel	rendah	tinggi	tinggi	TIDAK DIMINATI
<input type="checkbox"/>	XD1	XL 1GB	XL	rendah	rendah	rendah	TIDAK DIMINATI
<input type="checkbox"/>	XD2	XL 2GB MASA AKTIF 2 BULAN	XL	tinggi	rendah	tinggi	TIDAK DIMINATI
<input type="checkbox"/>	XD5	XL DATA 5GB	XL	rendah	tinggi	tinggi	TIDAK DIMINATI

Gambar 3.2 Pengklasifikasian data latihan

b. Kriteria dan probabilitas

1) Probabilitas kriteria operator paket internet

Operator dalam penelitian ini merupakan variabel yang menjelaskan pihak yang berperan sebagai penyedia layanan internet atau yang biasa disebut “operator”.

Operator	Label/kelas		Probabilitas	
	Minat	Tidak	Tidak	Tidak
Indosat	3	1	0.75	0.17
Telkomsel	1	2	0.25	0.33
XI	0	3	0.00	0.5
Jumlah	4	6	0.4	0.6

2) Probabilitas kriteria kuota utama paket internet

Dalam penelitian ini variabel kuota merupakan variabel yang dikelompokkan dalam dua kategori, yakni “tinggi” dan “rendah”. Kuota “tinggi” mencakup semua paket yang memiliki kuota lebih dari 2,5 GB, adapun kuota dengan nilai 2,5 GB ataupun kuota yang lebih rendah dari 2,5 GB masuk pada kategori kuota “rendah”.

Kuota	Label/kelas		Probabilitas	
	Minat	Tidak	Tidak	Tidak
Tinggi	1	3	0.25	0.5
Rendah	3	3	0.75	0.5
Jumlah	4	6	0.4	0.6

3) Probabilitas kriteria masa aktif paket internet

Merupakan variabel lamanya masa pakai paket internet yang dapat dikategorikan dalam 2 kategori yaitu “rendah” yang merupakan masa aktif dengan lama kurang dari atau sama dengan 30 hari, dan kategori “tinggi” untuk masa aktif dengan lama diatas 30 hari.

Masa aktif	Label/kelas		Probabilitas	
	Minat	Tidak	Tidak	Tidak
Tinggi	0	1	0.0	0.17
Rendah	4	5	1	0.83
Jumlah	4	6	0.4	0.6

4) Probabilitas kriteria harga jual paket internet

Variabel ini merupakan variabel besarnya harga jual dari paket internet yang dikategorikan dalam 2 kategori yaitu tinggi dan rendah. Untuk menentukan kategori dari masing2 harga, yang pertama dilakukan adalah menghitung rata-rata dari harga :

$$\frac{\sum x}{n} = \frac{454000}{10} = 45400$$

Kategori “tinggi” yaitu harga yang lebih dari rata rata harga dari semua produk, adapun nilai rata-rata harga produk adalah Rp.45.400. Kategori “rendah” yaitu harga yang kurang dari atau sama dengan rata rata harga dari semua produk, adapun nilai rata-rata harga produk adalah Rp.45.400.

Harga	Label/kelas		Probabilitas	
	Minat	Tidak	Tidak	Tidak
Tinggi	2	5	0.5	0.83
Rendah	2	1	0.5	0.17
Jumlah	4	6	0.4	0.6

c. Pengujian

Penelitian ini saya mulai dengan memasukkan data baru, disini penulis contohkan dengan memasukkan data produk paket internet baru dengan ketentuan sebagai berikut

Data sampel	Nama	Kuota utama	Masa aktif	Operator	Harga
	Indosat Data 2gb	2 Gb	30 Hari	Indosat	Rp.36..500

Dari data uji diatas yang telah dimasukkan, tahap kedua yakni klasifikasi data. Klasifikasi ini sesuai dengan ketentuan data produk paket tersebut dengan ketentuan yang telah dijelaskan pada pembahasan sebelumnya

Data sampel	Nama	Kuota utama	Masa aktif	Operator	Harga
	Indosat Data 2gb	Rendah	Rendah	Indosat	Rendah

Mulai Perhitungan Klasifikasi

a. Menghitung P(X|Ci) untuk setiap Kelas (label)

$$p(\text{operator}|\text{minat}) = p(\text{Indosat}|\text{minat}) = \frac{3}{4} = 0.75$$

$$p(\text{operator}|\text{tidak}) = p(\text{Indosat}|\text{tidak}) = \frac{1}{4} = 0.25$$

$$p(\text{masa aktif}|\text{minat}) = p(\text{Rendah}|\text{minat}) = \frac{4}{9} = 0.4444444444444444$$

$$p(\text{masa aktif}|\text{tidak}) = p(\text{Rendah}|\text{tidak}) = \frac{5}{9} = 0.5555555555555556$$

$$p(\text{kuota}|\text{minat}) = p(\text{Rendah}|\text{minat}) = \frac{3}{6} = 0.5$$

$$p(\text{kuota}|\text{tidak}) = p(\text{Rendah}|\text{tidak}) = \frac{3}{6} = 0.5$$

$$p(\text{harga}|\text{minat}) = p(\text{Rendah}|\text{minat}) = \frac{2}{3} = 0.6666666666666666$$

$$p(\text{harga}|\text{tidak}) = p(\text{Rendah}|\text{tidak}) = \frac{1}{3} = 0.3333333333333333$$

b. Perhitungan Kelas Minat/Tidak

$$\begin{aligned} \rho(\chi|minat) &= 0.75 \times 0.4444444444444444 \times 0.5 \times 0.6666666666666666 \\ &= 0.1111111111111111 \end{aligned}$$

$$\begin{aligned} \rho(\chi|tidak) &= 0.25 \times 0.5555555555555556 \times 0.5 \times 0.3333333333333333 \\ &= 0.023148148148148147 \end{aligned}$$

c. Perhitungan Kelas Minat/Tidak dengan Akumulasi Kelas Label

$$\rho(X|C_i). \rho(C_i) = \rho(X|minat). \rho(minat)$$

$$= 0.1111111111111111 \times \frac{4}{10}$$

$$= 0.1111111111111111 \times 0.4$$

$$= 0.044444444444444446$$

$$\rho(X|C_i). \rho(C_i) = \rho(X|tidak). \rho(tidak)$$

$$= 0.023148148148148147 \times \frac{6}{10}$$

$$= 0.009259259259259259$$

d. Perbandingan Masing Masing Kelas

Berdasarkan perhitungan akhir dengan mengalikan nilai peluang dari kasus yang di angkat, kita melihat bahwa nilai $\rho(X|minat). \rho(minat)$ lebih tinggi dari $\rho(X|tidak). \rho(tidak)$ yakni sebesar 0.04444444446 dengan 0.009259259259 sehingga dapat disimpulkan bahwa Kartu Internet tersebut masuk dalam klasifikasi “Diminati”.

PENUTUP

Simpulan

Berdasarkan penelitian tentang memprediksi penggunaan listrik rumah tangga dapat ditarik beberapa kesimpulan sebagai berikut

- a. Nilai keakuratan dari hasil perhitungan minat dan tidak dengan cara membandingkan hasil prediksi dengan hasil yang sebenarnya diluar data training. Berdasarkan data penjualan produk paket internet yang dijadikan data training, metode naive bayes berhasil mengklasifikasikan 10 data produk dengan 858 data transaksi sebagai data training dan 10 data produk dengan 115 data transaksi sebagai data uji. Dari analisa seluruh data uji dan data training, perbandingan kebenarannya adalah 8/10, yakni 8 dari 10 produk dengan nilai prediksi yang bernilai benar Sehingga metode naïve bayes classifier berhasil memprediksi besarnya minat beli pelanggan terhadap produk paket internet dengan persentase keakuratan sebesar 80%.
- b. Pengguna dapat mengklasifikasikan produk paket internet berdasarkan pada penjualan yang terjadi sebelumnya, dengan mengklasifikasikan berdasarkan masing-masing kelas dari seluruh atribut yang ada. Semakin banyak data penjualan yang terjadi sebelumnya, maka pengklasifikasian ini juga semakin akurat.

Saran

Banyaknya data yang digunakan untuk penelitian ini yang dapat diuji coba disarankan lebih banyak, karena jumlah data juga berpengaruh terhadap keakuratan hasil perhitungan. Adapun hasil penelitian ini diharapkan dapat digunakan sebagai satu gambaran pengambilan keputusan serta penggunaan nilai *support* dan *confidence* yang bervariasi, sehingga dapat menghasilkan lebih banyak aturan asosiasi antar data yang mengandung informasi penting yang lebih bermanfaat.

DAFTAR PUSTAKA

- Alfa Saleh. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga.
- Arikunto, S. (2002). Prosedur Penelitian Suatu Pendekatan Praktik. Jakarta: Rineka Cipta.
- Han. (2012). Data Mining Concepts and Techniques 2nd Edition. San Fransisco: Morgan Kaufmann Publishers.
- Han, J., Pei, J., & Kamber, M. (2006). Data mining: Concepts and techniques. All rights.
- Kusrini, & Emha, T. L. (2009). Algoritma data mining. Yogyakarta: Andi Offset.