

Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Classifier Untuk Mendeteksi Minat Beli Pelanggan Terhadap Produk Paket Internet

Swi Indriyaningtyas^{a1}, Aji Bagus Prasetya^{a2}, Rodhiyatu Robbi Al Marjuwah^{a3}, Saifur Rohman Cholil^{b4}

^aProdi S1 Sistem Informasi, Fakultas Teknologi Informasi Komunikasi, Universitas Semarang
Jl. Soekarno Hatta, Kota Semarang, Jawa Tengah, Indonesia

¹swidrygtys@gmail.com

²190501bagus@gmail.com

³rodhiyaturobbialmarjuwah@gmail.com

^bFakultas Teknologi Informasi dan Komunikasi, Universitas Semarang
Jl. Soekarno Hatta, Kota Semarang, Jawa Tengah, Indonesia

⁴cholil@usm.ac.id

Abstrak

AM Cell merupakan salah satu unit bisnis yang bergerak di bidang penjualan kartu perdana internet yang bertahan dalam persaingan di Kabupaten Jepara. AM Cell menawarkan berbagai jenis dan harga paket Internet, namun tidak mengetahui seberapa besarnya minat konsumen terhadap paket internet tersebut. Hal ini akan memungkinkan konsumen beralih ke perusahaan usaha serupa yang menurut konsumen paling menguntungkan. Sehingga penelitian ini untuk memberikan gambaran mengenai minat beli konsumen terhadap produk paket internet di AM Cell, mendeteksi dengan algoritma Naïve Bayes Classifier pada usaha penjualan paket Internet dan klasifikasi terkait deteksi minat beli pelanggan terhadap penjualan paket Internet. Berdasarkan data penjualan paket internet pada bulan Juni 2021, metode Naïve Bayes dapat mengklasifikasikan 52 data produk paket internet dengan 986 data transaksi sebagai data training dan 52 data produk internet dengan 422 data transaksi sebagai data uji. Sehingga untuk mengetahui minat beli pelanggan paket internet didapatkan hasil dari perhitungan Naïve Bayes menggunakan software RapidMiner dengan akurasi sebesar 78.85%. Sedangkan, perhitungan Naïve Bayes menggunakan software Microsoft Excel memperoleh hasil akurasi sebesar 88%.

Keywords: Minat Beli, Paket Internet, Naïve Bayes, RapidMiner

Abstract

AM Cell is one of the business units engaged in selling internet starter packs that survives the competition in Jepara Regency. AM Cell offers various types and prices of Internet packages, but they do not know how much consumers are interested in these Internet packages. This will allow consumers to switch to similar business companies that consumers find most profitable. So this research is to provide an overview of consumer buying interest in internet package products at AM Cell, detect using the Naïve Bayes Classifier algorithm in the Internet package sales business and classification related to detecting customer buying interest in Internet package sales. Based on internet package sales data in June 2021, the Naïve Bayes method can classify 52 internet package product data with 986 transaction data as training data and 52 internet product data with 422 transaction data as test data. So, to find out the customer's buying interest in internet packages, the results obtained from Naïve Bayes calculations using RapidMiner software with an accuracy of 78.85%. Meanwhile, the Naïve Bayes calculation using Microsoft Excel software obtained an accuracy of 88%.

Keywords: Buying Interest, Internet Packages, Naïve Bayes, RapidMiner

1. Pendahuluan

Tuntutan akan teknologi komunikasi yang sangat tinggi dan adanya kebutuhan yang semakin besar akan perkembangan penggunaan akses data internet terutama di masa pandemi seperti sekarang ini. Persaingan dalam dunia bisnis membuat para pengusaha terus berinovasi untuk bisnis mencapai kelangsungan yang revolusioner. Prediksi minat konsumen akan sangat penting bagi perusahaan untuk mengatur strategi dalam menjalankan bisnisnya. Pentingnya konsumen bagi sebuah perusahaan tergantung kepada bagaimana perusahaan menjalin hubungan dengan pelanggan [1].

AM Cell merupakan salah satu unit bisnis yang bergerak di bidang penjualan kartu perdana internet yang bertahan dalam persaingan di Kabupaten Jepara. AM Cell menawarkan berbagai jenis dan harga paket Internet, namun tidak mengetahui seberapa besarnya minat konsumen terhadap paket internet tersebut. Hal ini akan memungkinkan konsumen beralih ke perusahaan usaha serupa yang menurut konsumen paling menguntungkan.

Aplikasi pemeringkatan berupa algoritma pengklasifikasi Naïve Bayes dapat mendeteksi atau memprediksi penjualan produk paket Internet terlaris berdasarkan transaksi masa lalu, sehingga perusahaan dapat memahami penjualan produk paket internet dan menjadi peluang untuk meningkatkan kualitas produk kemasan Internet [2].

Untuk memperkirakan kelas dan objek dengan label yang tidak diketahui, gunakan metode klasifikasi untuk menemukan model dan fungsi yang mencirikan atau membedakan kelas data. Untuk mencapai hal ini, proses klasifikasi mungkin membagi data ke dalam kelas-kelas terpisah berdasarkan aturan dan fungsi tertentu, menghasilkan pembentukan model.

Perangkat lunak (*software*) pendukung Rapidminer merupakan salah satu *software* untuk pengolahan data mining yang banyak digunakan untuk analisis dan menyelesaikan masalah dalam penelitian ini. RapidMiner text mining melakukan pekerjaan yang berkisar dengan analisis teks, mengekstrak pola-pola dari dataset yang besar mengkombinasikannya dengan metode statistika, kecerdasan buatan dan database [3].

Secara umum, tujuan penelitian ini adalah untuk memberikan gambaran mengenai minat beli konsumen terhadap produk paket internet di AM Cell. Sehingga dalam penelitian ini dapat diambil rumusan masalahnya untuk mendeteksi dengan algoritma Naïve Bayes classifier pada usaha penjualan paket Internet dan klasifikasi terkait deteksi minat beli pelanggan terhadap penjualan paket Internet.

2. Metode Penelitian

Metode pada penelitian ini menggunakan Naïve Bayes yang merupakan algoritma klasifikasi dengan efektif dan efisien [4]. Metode yang digunakan untuk mendukung penelitian ini adalah metode kuantitatif, karena penelitian ini disajikan dalam bentuk angka-angka. Penelitian kuantitatif merupakan penelitian yang banyak dituntut menggunakan angka-angka, mulai dari pengumpulan data, pengolahan data, penafsiran terhadap data tersebut serta penampilan terhadap hasilnya [5].

2.1. Pengumpulan Data

Metode pengumpulan data dilakukan oleh peneliti untuk mendapatkan dan mengumpulkan data atau informasi yang dibutuhkan untuk mencapai tujuan penelitian [6]. Prosedur yang digunakan untuk mengumpulkan data yaitu dengan mengumpulkan data produk paket internet beserta data transaksi penjualan. Penelitian ini berhasil mengumpulkan data dengan total 1408 transaksi pada penjualan paket internet selama 1 bulan, yaitu bulan Juni 2021.

Tabel 1. Data Set

Operator	Jenis	Masa Aktif (Hari)	Kuota Utama (MB)	Harga	Jml Transaksi
Telkomsel	Voucher Tsel 2,5	7	1500	13000	70
Telkomsel	Voucher Tsel 10GB	30	7000	65000	14
Telkomsel	Voucher Tsel 3,5	7	2000	22000	46
Telkomsel	Voucher Tsel 1,5	3	1000	10000	15
Smartfren	Voucher Smartfren 6GB	28	6000	31000	22

Smartfren	Voucher Smartfren 14GB	28	10000	47000	7
Smartfren	Voucher Smartfren 2,5	3	500	6000	1
Smartfren	Voucher Smartfren Lite	28	500	55000	20
XL	Voucher XL 12GB	30	10000	20000	93
XL	Voucher XL 19GB	30	9000	41000	27
XL	Voucher XL 60GB	30	13000	65000	31
XL	Voucher XL 3/10	10	500	12000	22
Indosat	Voucher Indosat 1,5/5hr	5	1500	12000	63
Indosat	Voucher Indosat F2	30	2000	18000	58
Indosat	Voucher Indosat 1+u	7	1000	26000	97
Indosat	Voucher Indosat 2+u	30	2000	35000	3
Tri	Voucher 3 1+u	7	1000	26000	15
Tri	Voucher 3 1/5	5	1000	10000	136
Tri	Voucher 3 8GB	30	8000	50000	2
Tri	Voucher 3 1/3	3	1000	8000	10
Axis	Voucher Axis 1/5	5	1500	10000	61
Axis	Voucher Axis 3/bulan	30	1000	33000	4
Axis	Voucher Axis 3/15	15	3000	25000	2
Axis	Voucher Axis 2/4	4	2000	18000	4
..
..
Jumlah Transaksi					1408

2.2. Proses Data

Pada tahap praproses data yang dilakukan adalah seleksi data data produk paket internet beserta data transaksi penjualan. Pemilihan data dalam proses seleksi terdiri dari 5 atribut yang antara lain data operator paket internet, kuota utama paket internet, harga jual, jumlah transaksi dan jumlah stok paket internet. Jika terdapat data numerik seperti pada kuota utama, harga, jumlah transaksi dan jumlah stok paket internet maka harus mencari nilai mean dan standar deviasi dari masing-masing data yang berupa data numerik [5]. Adapun alur metode Naïve Bayes sebagai berikut :

- a. Mengklasifikasikan produk paket data dengan persamaan sebagai berikut

$$\mu = \frac{\sum_{i=1}^n X_i}{n} \tag{1}$$

- b. Perhitungan kelas menggunakan teorema bayes sebagai berikut.

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)} \tag{2}$$

Keterangan :

X adalah data sampel dengan kelas (label) yang tidak diketahui

H adalah hipotesa bahwa X adalah data dengan kelas (label).

P(H) adalah peluang dari hipotesa **H**.

P(X) adalah peluang data sampel yang akan diamati.

P(X|H) adalah peluang data sampel **X**, bila diasumsikan bahwa hipotesa benar (valid).

2.3. Pembagian Data

Pada tahapan ini, data dibagi menjadi 2, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih dan data uji menggunakan data produk paket internet yang dijual beserta data penjualan paket internet yang terjadi pada bulan Juni 2021.

2.4. Membuat Model Klasifikasi

Pembuatan model klasifikasi yang diperlukan adalah penentuan atribut. Dalam penelitian ini atribut telah ditentukan antara lain : operator atau penyedia paket internet, kuota paket internet, harga, jumlah transaksi dan jumlah stok. Setelah atribut ditetapkan, maka tahap selanjutnya adalah menentukan klasifikasi dari produk-produk paket internet tersebut.

3. Hasil dan Pembahasan

Naïve Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [7]. Keuntungan menggunakan metode Naïve Bayes hanya membutuhkan jumlah data (*training data*) dalam menentukan estimasi parameter yang dibutuhkan dalam proses pengklasifikasian [4].

Data yang akan dianalisa dalam penelitian ini berhubungan dengan data penjualan paket internet pada bulan Juni 2021. Atribut yang akan diprediksi adalah penjualan jenis operator Telkomsel, Smartfren, Indosat, Tri, XL dan Axis.

3.1. Data Latih

Penerapan Naïve Bayes Classifier yang harus dilakukan adalah menentukan data latih (*training data*), berikut ini merupakan data latih dari penelitian ini [8]:

Tabel 2. Data Latih

Operator	Jenis	Masa Aktif (Hari)	Kuota Utama (MB)	Harga	Jml Transaksi
Telkomsel	Voucher Tsel 10GB	30	7000	65000	10
Smartfren	Voucher Smartfren 6GB	28	6000	31000	15
XL	Voucher XL 12GB	30	10000	20000	65
Indosat	Voucher Indosat 1+u	7	1000	26000	68
Tri	Voucher 3 1/5	5	1000	10000	95
Axis	Voucher Axis 1/5	5	1500	10000	43
..
..
Jumlah Transaksi					986

Pengklasifikasian produk paket internet yang sudah dipasarkan memperoleh hasil yaitu kelas “diminati” atau “tidak diminati” dengan cara menentukan nilai rata-rata dari keseluruhan penjualan pada data sampel. Penelitian ini menggunakan data latih dengan total sebanyak 986 transaksi, dari jumlah tersebut dapat diambil nilai rata-rata dengan cara membagi total transaksi dan jumlah produk, sehingga menghasilkan nilai rata-rata transaksi sebagai berikut :

$$\frac{\sum x}{n} = \frac{986}{52} = 18,9$$

Sehingga setiap produk paket internet yang memiliki total lebih dari atau sama dengan 18,9 dapat diberi kelas “diminati” dan sebaliknya mendapat kelas “tidak diminati”. Data sampel yang telah diklasifikasi sebagai berikut :

Tabel 3. Klasifikasi Data Latih

Operator	Jenis	Masa Aktif (Hari)	Kuota Utama (MB)	Harga	Jml Transaksi
Telkomsel	Voucher Tsel 10GB	Tinggi	Tinggi	Tinggi	Tidak Diminati
Smartfren	Voucher Smartfren 6GB	Tinggi	Rendah	Rendah	Tidak Diminati
XL	Voucher XL 12GB	Tinggi	Rendah	Rendah	Diminati
Indosat	Voucher Indosat 1+u	Rendah	Tinggi	Rendah	Diminati
Tri	Voucher 3 1/5	Rendah	Tinggi	Rendah	Diminati
Axis	Voucher Axis 1/5	Rendah	Rendah	Rendah	Diminati
..
..

3.2. Kriteria dan Probabilitas

1. Kriteria operator paket internet

Operator dalam penelitian ini merupakan variabel yang berperan sebagai penyedia layanan internet [9].

Tabel 4. Kriteria Operator Paket Internet

Operator	Label/Kelas	
	Minat	Tidak
Telkomsel	2	3
Smartfren	1	6
XL	4	2
Indosat	6	7
Tri	6	7
Axis	1	7

2. Kriteria kuota utama paket internet

Variabel kuota utama dalam penelitian ini dibagi menjadi dua kriteria, yaitu “tinggi” dan “rendah”. Kuota utama “tinggi” mencakup semua paket data internet yang memiliki kuota utama lebih dari sama dengan 5 GB, sebaliknya kuota utama lebih rendah dari 5 GB masuk pada kriteria “rendah”.

Tabel 5. Kriteria Kuota Utama Paket Internet

Kuota	Label/Kelas	
	Minat	Tidak
Tinggi	8	Tinggi
Rendah	12	Rendah

3. Kriteria masa aktif paket internet

Variabel masa aktif atau lamanya masa pakai paket internet dikelompokkan mejadi 2 kriteria yaitu “tinggi” dan “rendah”. Kriteria “tinggi” merupakan masa aktif dengan lama diatas 15 hari, dan kriteria “rendah” untuk masa aktif kurang dari atau sama dengan 15 hari.

Tabel 6. Kriteria Masa Aktif Paket Internet

Masa Aktif	Label/Kelas	
	Minat	Tidak
Tinggi	11	20
Rendah	9	12

4. Kriteria harga paket internet

Variabel besarnya harga dari paket internet dikelompokkan menjadi 2 kriteria yaitu “tinggi” dan “rendah”. Penentuan kriteria dari masing-masing harga yang perlu dilakukan adalah menghitung rata-rata sebagai berikut :

$$\frac{\sum x}{n} = \frac{1634000}{52} = 31423$$

Hasil perhitungan rata-rata harga produk paket internet sebesar 31423. Kriteria “tinggi” yaitu harga yang lebih dari rata-rata harga produk paket internet, dan kriteria “rendah” dengan harga kurang dari rata-rata harga produk paket internet.

Tabel 7. Kriteria Harga Paket Internet

Harga	Label/Kelas	
	Minat	Tidak
Tinggi	6	15
Rendah	14	17

3.3. Pengujian

Pada penelitian ini setelah penerapan Naïve Bayes Classifier dengan menentukan data latih (*training data*), berikut ini merupakan data uji dari penelitian ini :

Tabel 8. Data Uji

Operator	Jenis	Masa Aktif (Hari)	Kuota Utama (MB)	Harga	Jml Transaksi
Telkomsel	Voucher Tsel 10GB	30	7000	65000	4
Smartfren	Voucher Smartfren 6GB	28	6000	31000	7
XL	Voucher XL 12GB	30	10000	20000	28
Indosat	Voucher Indosat 1+u	7	1000	26000	29
Tri	Voucher 3 1/5	5	1000	10000	41
Axis	Voucher Axis 1/5	5	1500	10000	18
..
..
Jumlah Transaksi					422

Data uji telah dimasukkan, kemudian pengklasifikasian data uji sesuai dengan ketentuan data produk paket internet.

Tabel 9. Klasifikasi Data Uji

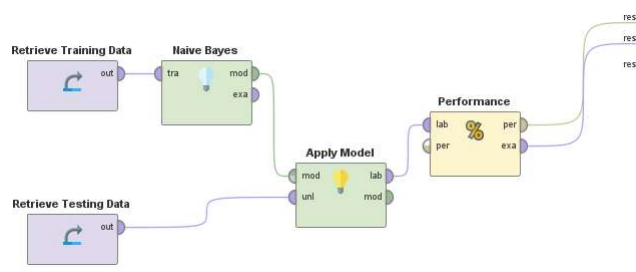
Operator	Jenis	Masa Aktif (Hari)	Kuota Utama (MB)	Harga	Jml Transaksi
Telkomsel	Voucher Tsel 10GB	Tinggi	Tinggi	Tinggi	Tidak Diminati
Smartfren	Voucher Smartfren 6GB	Tinggi	Rendah	Rendah	Tidak Diminati
XL	Voucher XL 12GB	Tinggi	Rendah	Rendah	Diminati
Indosat	Voucher Indosat 1+u	Rendah	Tinggi	Rendah	Diminati
Tri	Voucher 3 1/5	Rendah	Tinggi	Rendah	Diminati
Axis	Voucher Axis 1/5	Tinggi	Rendah	Tinggi	Tidak Diminati
..
..

- a. Perhitungan Naïve Bayes menggunakan Microsoft Excel
Microsoft Excel merupakan software yang memberikan fasilitas grafik cukup bervariasi untuk dimanfaatkan berbagai model perhitungan matematika [10]. *Microsoft Excel* digunakan dalam penelitian ini sebagai pendukung dalam mendeteksi minat beli paket internet menggunakan metode Naïve Bayes. Hasil perhitungan menggunakan Microsoft Excel menghasilkan tingkat akurasi yang tinggi yaitu 88%.

Confusion Table	Predicted	Class	
		Diminati	Tidak Diminati
	Diminati	0	0
	Tidak Diminati	6	46
Accuracy		88%	

Gambar 1. Hasil Perhitungan Microsoft Excel

- b. Perhitungan Naïve Bayes menggunakan RapidMiner
Berbeda dengan hasil perhitungan menggunakan Microsoft Excel, hasil perhitungan menggunakan RapidMiner menghasilkan tingkat akurasi 78.85%.



Gambar 2. Implementasi Naive Bayes pada RapidMiner

accuracy: 78.85%

	true Diminati	true Tidak Diminati	class precision
pred. Diminati	5	10	33.33%
pred. Tidak Diminati	1	36	97.30%
class recall	83.33%	78.26%	

Gambar 3. Hasil Perhitungan RapidMiner

4. Kesimpulan

Berdasarkan penelitian Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Classifier Untuk Mendeteksi Minat Beli Pelanggan Terhadap Produk Paket Internet di AM Cell, maka dapat ditarik beberapa kesimpulan sebagai berikut :

- a. Berdasarkan data penjualan paket internet, metode Naïve Bayes dapat mengklasifikasikan 52 data produk paket internet dengan 986 data transaksi sebagai data training dan 52 data produk internet dengan 422 data transaksi sebagai data uji. Sehingga untuk mengetahui minat beli pelanggan paket internet didapatkan hasil dari perhitungan Naïve Bayes menggunakan software RapidMiner dengan akurasi sebesar 78.85%. Sedangkan, perhitungan Naïve Bayes menggunakan software Mcrosoft Excel memperoleh hasil akurasi sebesar 88%.
- b. Metode Naïve Bayes dapat mengklasifikasikan produk paket internet berdasarkan data penjualan selama 1 bulan, yaitu pada bulan Juni 2021.

References

- [1] N. Haqqizar and T. N. Larasyanti, “Analisis Sentimen Terhadap Layanan Provider Telekomunikasi Telkomsel Di Twitter Dengan Metode Naïve Bayes,” *Pros. TAU SNAR-TEK 2019 Semin. Nas. Rekayasa dan Teknol.*, vol. 10, no. 2, pp. 1–15, 2019.
- [2] A. F. Aulia, I. Hadi, and K. Pos, “ANALISA KUALITAS LAYANAN INTERNET DAN PENGKLASIFIKASIAN Jurusan Teknik Elektro , Politeknik Negeri Sriwijaya , Jl . Srijaya Negara Bukit Besar , PENDAHULUAN Pada era globalisasi , internet menjadi sebuah kebutuhan yang wajib dipenuhi oleh semua manusia un,” vol. 6, no. 1, pp. 745–752, 2020.
- [3] D. Handoko, H. S. Tambunan, and J. T. Hardinata, “Analisis Penjualan Produk Paket Kuota Internet Dengan Metode K-Nearest Neighbor,” *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 6, no. 1, p. 111, 2021, doi: 10.30645/jurasik.v6i1.275.
- [4] S. Kusumadewi, “Klasifikasi Status Gizi Menggunakan Naïve Bayesian Classification,” *CommIT (Communication Inf. Technol. J.*, vol. 3, no. 1, p. 6, 2009, doi: 10.21512/commit.v3i1.506.
- [5] A. S. Mubarak, D. R. Prehanto, and M. Ali, “Deteksi Minat Beli Pelanggan Terhadap Produk Paket Internet Menggunakan Algoritma Naïve Bayes,” *Inovate*, vol. 03, no. 02, pp. 58–63, 2019.
- [6] R. Amalia, “Penerapan data mining untuk memprediksi hasil kelulusan siswa menggunakan metode naïve bayes,” *J. Inform. dan Sist. Inf.*, vol. 6, no. 1, pp. 33–42, 2020.
- [7] F. N. Zuhri and A. Alamsyah, “Menggunakan Naïve Bayes Classifier Di Forum Kaskus Publik Sentiment Analysis Of Smartfren Brand Using Naïve Bayes Classifier On Kaskus Forum,” *e- Proceeding Manag.*, vol. 4, no. 1, pp. 242–251, 2017.
- [8] M. F. Rifai, H. Jatnika, and B. Valentino, “Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS),” *Petir*, vol. 12, no. 2, pp. 131–144, 2019, doi: 10.33322/petir.v12i2.471.
- [9] D. Nofriansyah, K. Erwansyah, and M. Ramadhan, “Penerapan Data Mining dengan Algoritma Naïve Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi),” *J. Sainifikom*, vol. 15, no. 2, pp. 81–92, 2016.
- [10] H. Basri, “Implementasi Algoritma Naïve Bayes Untuk Mengetahui Potensi Siswa Dalam Pemahaman Ms. Excel,” *Sistemasi*, vol. 8, no. 3, p. 505, 2019, doi: 10.32520/stmsi.v8i3.569.