

ANALISA TINGKAT KEPUASAN PELANGGAN TERHADAP PENJUALAN *BEAUTY* PRODUK PADA *ONLINE SHOP* DENGAN MENGGUNAKAN METODE *NAIVE BAYES*

Donny Maulana¹⁾, Emi Lia Nurjanah²⁾

Program Studi Teknik Informatika Fakultas Teknik
Universitas Pelita Bangsa
donny.maulana@pelitabangsa.ac.id

Disetujui, 25 September 2019

Abstraksi

Perkembangan teknologi saat ini semakin memudahkan dan memanjakan manusia, salah satu contohnya yang sudah kita kenal sejak lama yaitu berbelanja secara *online*. Disamping itu kepuasan pelanggan menjadi target utama dalam dunia bisnis, karena pelanggan adalah pondasi utama dalam berdirinya suatu perusahaan dan menjadi tujuan utama. Upaya dalam meningkatkan pelayanan kepuasan pelanggan pada salah satu toko yang ada di *marketplace* terbesar di *shopee* yaitu *DeBeautyHouse* peneliti melakukan analisa tingkat kepuasan dalam *online shop* tersebut melalui data penilaian yang ada, adapun atribut yang mempengaruhi kepuasan pelanggan yaitu, nama produk, jenis produk, kualitas produk, Harga produk dan Pengiriman produk. Metode yang digunakan dalam penelitian ini yaitu *Naive Bayes* untuk menentukan tingkat akurasi data. Hasil dari penelitian ini diperoleh tingkat akurasi sebesar 96,00% dengan nilai *recall* sebesar 95,24% dan nilai *precision* 100,00%

Kata Kunci : Kepuasan Pelanggan, *Online Shop*, *Naive Bayes*

Abstract

The development of technology is now making it easier and pampering humans, one example that we have known for a long time is shopping online. Besides that customer satisfaction becomes the main target in the business world, because customers are the main foundation in the establishment of a company and become the main goal. Efforts to improve customer satisfaction services in one of the stores in the biggest marketplace in shopee namely DeBeautyHouse, the authors analyze the level of satisfaction in the online shop through existing assessment data, as for the attributes that affect customer satisfaction, namely, product name, product type, product quality, product prices and product shipments. The method used in this study is Naive Bayes to determine the level of data accuracy. The results of this study obtained an accuracy rate of 96.00% with a recall value of 95,24% and a percision value of 100,00%.

Keywords : Customer Satisfaction, *Online Shop*, *Naive Bayes*

1. Pendahuluan

Disiplin ilmu yang mempelajari data dengan menggunakan metode adalah *data mining*, *data mining* merupakan gabungan dari beberapa ilmu yaitu seperti matematika, statistika, analisis dan lainnya. Singkatnya *data mining* yaitu sebuah ilmu yang digunakan untuk mencari sebuah informasi yang belum pernah diketahui dari sebuah *dataset* yang besar dengan cara menguji metode yang terdapat dalam *data mining* yang telah dikembangkan dengan sebuah *dataset*, sehingga dapat menghasilkan solusi yang mampu menyelesaikan suatu permasalahan yang ada. Adapun metode dalam *data mining* yaitu, *estimasi*, *prediksi*, *klasifikasi*, *klastering* dan *asosiasi*. Dengan *data mining* kita bisa merubah data yang sudah tidak terpakai menjadi sebuah informasi yang berguna baik untuk kepentingan bisnis maupun ilmu dan pengetahuan dalam dunia pendidikan, kesehatan, pemerintahan dan lainnya. Seperti pada contoh kasus dalam data mining sebagai berikut, dengan metode estimasi kita bisa mengetahui waktu yang dibutuhkan kurir untuk mengantar paket, dan kita bisa mengetahui pola perilaku pelanggan dan kebiasaannya dan apa saja produk sering dibeli.

Dari berbagai contoh kasus diatas, *data mining* bisa menjadi solusi alternatif dari permasalahan yang dihadapi oleh sebuah *online shop* dari *market place* terbesar seperti *shopee*. Adapaun permasalahan yang dihadapi yaitu kepuasan pelanggan yang mempengaruhi tingkat tinggi dan rendahnya penjualan produk pada *online shop* tersebut, dengan memprediksi mengenai produk apa saja yang paling diminati dan jenis produk apa saja yang paling sering dicari *make up* atau *skincare*, kemudian bagaimana kepuasan pelanggan mengenai kualitas produk, Harga produk dan Pengiriman produk pada *online shop* tersebut melalui penilaian yang telah dilakukan oleh pelanggan, serta metode data mining dengan algoritma apa yang cocok untuk diterapkan dalam perhitungan penelitian mengenai kepuasan pelanggan yang mampu menyelesaikan permasalahan pada *online shop* tersebut. Berdasarkan permasalahan yang ada, maka peneliti memutuskan untuk menggunakan metode klasifikasi *naive bayes* sebagai algoritma perhitungannya. Dengan demikian peneliti menetapkan penelitian penelitian dengan judul Diharapkan dengan menerapkan *data mining* ke dalam pengelolaan data penilaian penjualan tersebut dapat mengurangi permasalahan yang ada.

2. Tinjauan Studi

Tahapan proses pada analisis tingkat kepuasan pelanggan terhadap pelayanan terhadap penjualan *beauty* produk pada *online shop* dengan metode KDD dengan algoritma *naive bayes* antara lain menghitung probabilitas kelas, menghitung probabilitas masing-masing kelas pengujian dengan *k-folds cross validation*, prediksi data dan pengujian dan validasi dengan aplikasi *rapidminer*. Knowledge Discovery In Database (KDD) adalah proses yang bertujuan untuk menggali dan menganalisis sejumlah besar himpunan data dan mengekstrak informasi serta pengetahuan yang berguna [14]. Berikut langkah penting dalam proses KDD yaitu, Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Pattern Evaluation, Knowledge Persentation setelah melalui proses *cleaning* data sampai data *selectin* kemudian menerapkan algoritma *naive bayes* sebagai data miningnya dengan pengujian menggunakan *k-folds cross validation* yang kemudian hasilnya di uji dengan aplikasi *rapidminer*.

Pada penelitian ini data yang didapat adalah data penilaian penjualan pada bulan Mei 2019-Juli 2019. Data yang didapat bersifat data public, yaitu data yang terdapat dan diambil secara langsung pada penilaian pelanggan di satu toko yang berada di *market place shopee* yaitu toko *DeBeautyHouse*. Data yang diperoleh sebanyak 486 data penilaian. Namun setelah melalui proses *cleaning* data dan terdapat adanya duplikasi data sehingga data tersebut di saring dan dikelola dan data diperoleh sebanyak 102 data. Adapun atribut yang terdapat pada data penilaian penjualan pada *online shop DeBeautyHouse* yaitu nama produk, jenis, kualitas produk, Harga produk, dan Pengiriman produk.

Tabel 2.1 Kriteria Yang Digunakan

NO	Jenis	Kualitas Produk	Harga Produk	Pengiriman Produk	Potensi Penilaian
1	2	Sangat Baik	Sangat Baik	Sangat Baik	Puas
2	1	Sangat Baik	Sangat Baik	Sangat Baik	Puas
3	2	Sangat Baik	Sangat Baik	Sangat Baik	Puas
4	2	Sangat Baik	Sangat Baik	Sangat Baik	Puas
5	1	Baik	Standar	Standar	Tidak puas
6	2	Standar	Standar	Baik	Tidak puas
7	1	Baik	Baik	Baik	Puas
8	2	Sangat Baik	Sangat Baik	Sangat Baik	Puas
9	1	Sangat Baik	Sangat Baik	Sangat Baik	Puas
10	2	Sangat Baik	Sangat Baik	Sangat Baik	Puas
...
...
102	1	Sangat Baik	Sangat Baik	Sangat Baik	Puas

2.1. Naive Bayes

Naive Bayes adalah teknik memprediksi probabilitas anggota suatu *class* yang menggunakan pengelompokan klasifikasi statistik. Model *Naive Bayes* mempunyai tingkat ketelitian yang tinggi pada saat diimplementasikan pada data yang besar didalam *database* yang ditemukan oleh Thomas Bayes di abad ke-18. Konsep dasarnya adalah Teorema *Bayes* yaitu menghitung probabilitas untuk melakukan pengklasifikasian [19].

Rumus *Naive Bayes* ditunjukkan seperti persamaan (1).

$$(1) = \frac{P(A) \cdot P(B|A)}{P(A) \cdot P(B|A) + P(B) \cdot P(A|B)}$$

Keterangan :
 X : Bukti
 H : Hipotesis
 $P(H|X)$: Probabilitas posterior H dengan syarat X
 $P(X|H)$: Probabilitas posterior X dengan syarat H
 $P(H)$: Probabilitas prior hipotesis H
 $P(X)$: Probabilitas prior bukti X

2.2. K-folds Cross Validation

K-Fold Cross Validation berguna untuk membagi kumpulan data menjadi data latih dan data uji. *K-Fold Cross Validation* digunakan karena merupakan salah satu metode yang terbaik untuk memvalidasi data yang akan digunakan. Salah satu contoh dari *K-Fold Cross Validation* adalah *10-Fold Cross Validation*. Teknik ini akan membagi kumpulan data menjadi 10 subset dengan ukuran yang sama, sembilan dari 10 subset data digunakan untuk pelatihan, sementara satu subset yang tertinggal digunakan untuk pengujian. Proses diulang selama sepuluh kali, dan hasil akhirnya diperkirakan sebagai tingkat kesalahan rata-rata pada contoh uji. Cara kerja *K-folds cross validation*, yaitu total data dibagi menjadi n bagian, iterasi atau *fold* ke 1, yaitu bagian ke 1 menjadi *testing*, bagian sisanya menjadi data *training*.

3 Kerangka Konsep

3.1. Confusion Matrix

Fungsi dari *confusion matrix* adalah untuk mengukur tingkat *accuracy*, *precision*, dan nilai *recall* dari suatu model algoritma yang dievaluasi. Nilai *accuracy* merupakan tingkat ketepatan presentase antara nilai prediksi dan nilai sebenarnya, kemudian nilai *precision* ialah nilai akurasi dengan *class* yang telah diprediksi. Sedangkan nilai *recall* merupakan presentase nilai kinerja keberhasilan algoritma yang dipakai. Berikut ini merupakan tabel *confusion table* dapat dilihat pada tabel :

Confusion Matrix		Nilai Prediksi	
		Positive	Negative
Nilai Sebenarnya	Positive	TP	FP
	Negative	FN	TN

Tabel 3.1 Confusion Table Rumus *Accuracy* :

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Rumus Recall : Recall = \frac{TP}{TP + FN}$$

$$Rumus precision : \frac{TP}{TP + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

Keterangan :
 TP (*True Poritive*) FP (*False Positive*) TN (*True Negative*) FN (*False Negative*)

3.2. Rapidminer

RapidMiner merupakan *software* perangkat lunak untuk pengolahan data. Dengan menggunakan prinsip dan algoritma data mining, *RapidMiner* mengekstrak pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database. *RapidMiner* memudahkan penggunaanya dalam melakukan perhitungan data yang sangat banyak dengan menggunakan operator-operator. Operator ini berfungsi untuk memodifikasi data. Data dihubungkan dengan node-node pada operator kemudian kita hanya tinggal menghubungkannya ke node hasil untuk melihat hasilnya. Hasil yang diperlihatkan *RapidMiner* pun dapat ditampilkan secara visual dengan grafik. Menjadikan *RapidMiner* adalah salah satu *software* pilihan untuk melakukan ekstraksi data dengan metode-metode data mining [30].

4. Metodologi Penelitian

4.1. Menghitung Probabilitas Kelas

Perhitungan ini untuk menentukan keputusan dengan metode *naïve bayes* untuk mencari probabilitas dari masing-masing kelas. Dalam menentukan keputusan akan ditentukan kelas “Bermasalah” dan “Tidak Bermasalah”. Perhitungannya dimana mencari jumlah data bermasalah dan tidak bermasalah dari total keseluruhan data training dan membaginya dengan total keseluruhan data.

Tabel 4.1 Probabilitas Kelas

Probabilitas Kelas	
Puas	Tidak puas
84	18
Total	102

4.2. Probabilitas Masing-Masing Atribut

Mencari probabilitas suatu atribut dengan membandingkan atribut dari data testing dengan atribut dari data *training*. Jumlah atribut dari kelas “Bermasalah” yang ada pada data *training* kemudian dibagikan dengan probabilitas kelas “Bermasalah”. Begitu juga mencari probabilitas kelas “Tidak Bermasalah”

a. Penilaian Jenis Produk

Tabel 4.2 Tabel Atribut Jenis Produk

Penilaian Jenis Produk		
Jenis Produk	Puas	Tidak Puas
1	29	12
2	55	6
	84	18

b. Penilaian Kualitas Produk

Tabel 4.3 Tabel Atribut Kualitas Produk

Penilaian Kualitas Produk		
Kualitas Produk	Puas	Tidak Puas
Sangat Baik	65	0
Baik	19	4
Standar	0	14
	84	18

c. Penilaian Harga Produk

Tabel 4.4 Tabel Atribut Harga

Penilaian Harga Produk		
Harga Produk	Puas	Tidak Puas
Sangat Baik	64	0
Baik	18	4
Standar	2	14
	84	18

Produk

d. Penilaian Pengiriman Produk

Tabel 4.5 Tabel Atribut Pengiriman Produk

Penilaian Pengiriman Produk		
Pengiriman Produk	Puas	Tidak Puas
Sangat Baik	64	0
Baik	17	1
Standar	3	17
	84	18

4.3 Kasus Perhitungan Cross Validation

a. Perhitungan Cross Validation K-Fold 1

Tabel 4.6 Cross Validation K-Fold 1

No	Jenis	Kualitas Produk	Harga Produk	Pengiriman Produk	Potensi Penilaian
1	2	Sangat Baik	Sangat Baik	Sangat Baik	Puas
2	1	Sangat Baik	Sangat Baik	Sangat Baik	Puas
...
...
9	1	Sangat Baik	Sangat Baik	Sangat Baik	Puas
10	2	Sangat Baik	Sangat Baik	Sangat Baik	Puas
Perhitungan Cross Validation Confusion Matrix Fold 1					
Prediksi		Puas		Tidak puas	
Puas		8		0	
Tidak puas		0		2	

b. Perhitungan Cross Validation K-Fold 10

Tabel 4.7 Cross Validation K-Fold 10

No	Jenis	Kualitas Produk	Harga Produk	Pengiriman Produk	Potensi Penilaian
93	1	Standar	Baik	Standar	Tidak puas
94	2	Sangat Baik	Sangat Baik	Sangat Baik	Puas
...
...
101	2	Sangat Baik	Sangat Baik	Sangat Baik	Puas
102	1	Sangat Baik	Sangat Baik	Sangat Baik	Puas
Perhitungan Cross Validation Confusion Matrix Fold 10					
Prediksi		Puas		Tidak puas	
Puas		6		0	
Tidak puas		0		4	

3.4. Confusion Matrix Total

Tabel 4.8 Confusion Matrix

Confusion Matrix Total	Prediksi	Puas	Tidak puas
	Puas	80	0
	Tidak puas	4	18
Accuracy	96%		

Total

5. Hasil dan Pembahasan

5.1 Implementasi Klasifikasi Naïve Bayes Pada Rapidminer

Perhitungan yang telah dilakukan diatas untuk menentukan potensi Penilaian Pelanggan dengan metode *naïve bayes*. Setelah melakukan perhitungan manual terhadap data training dan testing. Langkah kedua, dilakukan implementasi algoritma *naïve bayes* dengan menggunakan *tools rapidminer*.

Berikut adalah tahapan dalam penerapan algoritma *naïve bayes*: Menentukan nilai *accuracy*, *recal* dan *percision* dan Menentukan performance. Pada tahap akhir ini dilakukan evaluasi hasil pengolahan data dengan program yang telah dikembangkan dan mengevaluasi hasil prediksi mining dengan data real yang sudah terjadi, sehingga terlihat seberapa akurat proses mining pada tools rapid miner.

a. *Accuracy*

Hasil *accuracy* menggunakan algoritma *naïve bayes* dengan *cross validation* dalam pengujian yaitu sebesar 96.00%, dapat dilihat pada gambar berikut :

accuracy: 96.00% +/- 4.90% (micro average: 96.00%)

	true Puas	true Tidak puas	class precision
pred. Puas	80	0	100.00%
pred. Tidak puas	4	18	81.82%
class recall	95.24%	100.00%	

Gambar 5.1 Hasil *accuracy*

Hasil analisa antara data *training* dan data *test* pada Rapidminer dapat dilihat pada lampiran. Untuk menghitung akurasi sebagai berikut :

Diketahui :

Jumlah data yang diuji : 102

Jumlah data yg di prediksi benar : 98

Jumlah data yang di prediksi salah : 4

b. *Recall*

Nilai *recall* dihitung dengan cara membagi data benar yang bernilai positive (*True Positive*) dengan hasil penjumlahan dari data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai negatif (*False Negative*).

recall: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Tidak puas)

	true Puas	true Tidak puas	class precision
pred. Puas	80	0	100.00%
pred. Tidak puas	4	18	81.82%
class recall	95.24%	100.00%	

Gambar 5.2 Hasil *Recall*

c. *Precision*

Nilai *precision* dihitung dengan cara membagi jumlah data benar yang bernilai positif (*True Positive*) dibagi dengan jumlah data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai positif (*False Positive*).

precision: 83.33% +/- 21.08% (micro average: 81.82%) (positive class: Tidak puas)

	true Puas	true Tidak puas	class precision
pred. Puas	80	0	100.00%
pred. Tidak puas	4	18	81.82%
class recall	95.24%	100.00%	

Gambar 5.3 Hasil *Precision*

d. Tampilan Depenelitian *Performance Vector*

```

PerformanceVector
PerformanceVector:
accuracy: 96.00% +/- 4.90% (micro average: 96.08%)
ConfusionMatrix:
True: Puas Tidak puas
Puas: 80 0
Tidak puas: 4 18
precision: 83.33% +/- 21.05% (micro average: 81.82%) (positive class: Tidak puas)
ConfusionMatrix:
True: Puas Tidak puas
Puas: 80 0
Tidak puas: 4 18
recall: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Tidak puas)
ConfusionMatrix:
True: Puas Tidak puas
Puas: 80 0
Tidak puas: 4 18
    
```

Gambar 5.4 Depenelitian *Performance Vector*

7	Tidak puas	0.000	1.000	Happy Alex V...	2	Standar	Standar	Standar
7	Puas	1.000	0.000	The Ordinar...	2	Sangat Baik	Sangat Baik	Sangat Baik
7	Puas	0.994	0.006	Medhead L.L.	2	Baik	Baik	Baik
7	Tidak puas	0.000	1.000	OL Maxcara...	1	Standar	Standar	Standar
9	Tidak puas	0.000	1.000	Etude House...	1	Baik	Standar	Standar

Gambar 5.5 Prediksi *RapidMiner*

Gambar 4.3. merupakan hasil prediksi dari data *testing* dengan menggunakan *sample* penilaian data yang digunakan dalam penelitian ini.

Prediction	Polynomial	0	Label	Tidak puas (6)	Ya	Puas (32)	Nilai	Puas (32)
Confidence_Puas	Real	0	Min	0.000	Max	1.000	Average	0.848
Confidence_Tidak_puas	Real	0	Min	0.000	Max	1.000	Average	0.152

Gambar 5.6 Tingkat Kepuasan Pada Hasil Prediksi

6. Kesimpulan

Dengan adanya sistem pengujian pada data penilaian penjualan dengan metode *Naïve Bayes*, dapat diambil suatu kesimpulan sebagai berikut:

1. Berdasarkan data dari status penilaian penjualan menghasilkan tingkat kepuasan pelanggan sebesar 82% yaitu dengan nilai akurasi sebesar 96,00% dengan nilai *recall* 95,24% dan nilai *precision* 100,00%, sehingga proses klasifikasi dapat berjalan dengan baik dan akurat.
2. Dengan metode *Naïve Bayes* yang digunakan, jenis produk yang paling diminati oleh pelanggan yaitu merupakan dari produk *skincare* (pada atribut jenis 2) dengan 55 (65,48%) nilai puas dari 84 total nilai puas pada penilaian Jenis Produk dari total 102 dataset penilaian pelanggan.
3. Tingkat kepuasan pelanggan dari segi kualitas produk, Harga produk dan Pengiriman produk pada *online shop DeBeautyHouse* ini sebesar 82,35%. Diperoleh dengan cara perhitungan dengan menggunakan metode *Naïve Bayes* yaitu dari rata-rata nilai puas 84 dari total 102 dataset penilaian penjualan.
4. Prediksi 38 data dengan berdasarkan klasifikasi dari data training 102 sebelumnya dengan menggunakan metode *Naïve Bayes*.
5. *Bayes* melalui perhitungan dengan menggunakan *tools rapidminer* ini menghasilkan prediksi jumlah puas sebanyak 32 dan tidak puas 6 dengan tingkat rata-rata *confidence* (Puas) sebesar 0.848 dan *confidence* (Tidak Puas) sebesar 0.152.

Daftar Pustaka

- Siregar and Mkd. Puspabhuana, DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner. CV. Kekata Group, 2017.
- Widiastuti, N.A., Santosa, S., Supriyanto, "Algoritma Klasifikasi Data Mining Naïve Bayes Berbasis Particle Swarm Optimization Untuk Deteksi Penyakit Jantung," J. Pseudocode, 1(1), 11-14. ISSN 2355-5920., vol. 1(1), 2014.
- B. R. C.T.I. et al., "Implemetasi kmeans clustering pada rapidminer untuk analisis daerah rawan kecelakaan," Semin. Nas. Ris. Kuantitatif Terap. 2017, no. April, pp. 58-60, 2017