

KLASIFIKASI DATA PENJUALAN ALAT TULIS KANTOR (ATK) TERLARIS UNTUK OPTIMASI STRATEGI PEMASARAN DI TOKO CITRAMEDIA MENGGUNAKAN METODE NAIVE BAYES

Edy Widodo¹⁾, Pathyatus Sa'adah²⁾

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

Disetujui, 30 Desember 2019

Abstraksi

Pada penelitian ini peneliti melakukan pengolahan data penjualan ATK untuk mengklasifikasi data ATK dan mencari nilai akurasi sebagai penentuan barang apa saja yang paling banyak diminati oleh *customer*. Tujuan dilakukannya penelitian ini yakni sebagai suatu strategi pemasaran agar dapat mengoptimalkan kerja toko dalam mengolah data stok barang yang tersedia sehingga dapat memenuhi kebutuhan *customer* dan dapat membantu meningkatkan keuntungan penjualan. Data yang ada dianalisis menggunakan metode Algoritma *Naive Bayes Classifier* dan pengujiannya diimplementasikan pada *tools RapidMiner* agar mendapatkan tingkat akurasi yang efektif dan efisien. Dari hasil pengujian yang dilakukan, didapatkan tingkat *accuracy* tertinggi sebesar 99.26% dan *accuracy* terendah sebesar 94.39% sedangkan evaluasi dengan kurva AUC hasil dari empat kali pengujian yang telah dilakukan termasuk dalam kategori klasifikasi sangat baik (*excellent classification*) yakni dengan evaluasi tertinggi sebesar 0.997.

Kata kunci: Alat Tulis Kantor, *Naive Bayes*, Klasifikasi, *RapidMiner*

Abstract

In this study the author conducted processing ATK data to classify data ATK and look for the value of accuracy as determination which are the most preferred by the customer. The purpose of this research is as a marketing strategy in order to optimize the work of the store in processing the stock data that is available so that it can meet customer needs and can help increase sales profits. Data were analyzed using the Naive Bayes Classifier Algorithm method. And testing is implemented on RapidMiner tools in order to get an effective and efficient level of accuracy. From the results of tests carried out, obtained the highest accuracy level of 99.26% and the lowest accuracy of 94.39%. While the evaluation with the AUC curve the result of four tests that have been carried out are included in the very good classification category (Excellent Classification) with the highest evaluation of 0.997.

Keywords: Office Stationery, *Naive Bayes*, Classification, *RapidMiner*

1. Pendahuluan

1.1 Latar Belakang

Penjualan merupakan suatu aktivitas atau bisnis menjual barang ataupun jasa. Dalam proses penjualan, produsen memberikan kepemilikan suatu barang atau jasa kepada konsumen dengan tujuan dapat memenuhi kebutuhan konsumen dan dapat memberikan keuntungan bagi produsen untuk harga yang telah ditentukan. Dalam suatu bisnis penjualan terdapat berbagai data mengenai transaksi jual beli barang atau jasa. Data tersebut harus diolah untuk mengetahui perkembangan bisnis yang dijalankan. Saat ini banyak pengusaha yang menyadari pentingnya peranan data dalam peningkatan bisnis mereka. Sayangnya, sebagian besar pengusaha tidak tahu bagaimana mengumpulkan data yang tepat dan sebagian lainnya tidak dapat menggunakan data tersebut secara efektif. Hingga seiring berjalannya waktu, perkembangan ilmu pengetahuan dan teknologi berkembang sangat cepat. Data dapat diolah dengan mudah menggunakan metode dan penerapan tertentu. Salah satu teknologi pengolahan data yakni dengan penerapan *data mining*. *Data mining* merupakan campuran metode-

metode analisis data dengan algoritma-algoritma untuk memproses data yang berukuran besar. *Data mining* telah banyak diaplikasikan dalam berbagai bidang, diantaranya dalam bidang bisnis dan kedokteran.

Salah satu toko yang sedang berkembang dan bergerak dalam penjualan ATK yakni Toko Citra Media, toko ini menerima pembelian barang dalam jumlah banyak ataupun sedikit. Berdasarkan pengamatan langsung yang dilakukan peneliti pada Toko Citra Media ini, pendataan barang pada gudang di toko tersebut sering terjadi kesalahan karena masih dilakukan secara manual yakni dicatat dalam satu buku lalu diinput menggunakan *Ms. Excel* satu per satu itu pun tidak secara

bertahap, sehingga tidak semua barang terinput keseluruhan ke dalam tabel *Excel*. Hal ini berpengaruh pada proses penjualan, pegawai kesulitan dalam mencari barang yang tersedia di gudang karena data belum tersipkandengan baik dan tidak tersedia laporan penjualan secara berkala. Pembuatan laporan pembelian dan laporan penjualan pun tidak optimal. Alhasil, seringkali stok barang yang di beli ternyata merupakan barang yang tidak laris terjual yang menyebabkan toko ini mengalami kerugian.

Dalam persaingan bisnis yang semakin ketat, toko Citra Media memerlukan suatu strategi pemasaran yang lebih efektif dan efisien, yang dapat menarik dan menjangkau lebih banyak pelanggan serta dapat memuaskan keinginan juga kebutuhan pelanggan. Adapun strategi pemasaran yang diterapkan oleh Citra Media saat ini dirasa kurang efektif dan efisien, dikarenakan penyediaan barang yang seringkali tidak tepat dan terdapat banyak kesalahan karena pendataan yang kurang optimal. Sehingga barang yang diperlukan pelanggan seringkali dalam keadaan tidak tersedia, dan barang yang tersedia ternyata merupakan barang yang minim peminat.

2. Tinjauan Studi

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 menarik dari data yang terdapat pada basis data. *Data Mining* digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery Database (KDD)* (Vulandari, *Data Mining Teori dan Aplikasi Rapidminer*, 2017).

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan (Ridwan, Suyono, & Sarosa, 2017) yaitu:

1. Deskripsi
Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat pada data.
2. Estimasi
Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori.
3. Prediksi
Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.
4. Klasifikasi
Dalam klasifikasi, terdapat target variabel kategori
5. Pengklusteran
Clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*)
6. Asosiasi
Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

2.2 Naive Bayes Classifier

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. *Bayesian classification* didasarkan pada teorema *Bayes* yang memiliki

kemampuan klasifikasi serupa dengan *decision tree* dan *neuralnetwork*. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan kedalam database dengan data yang besar. Metode *Bayes* merupakan pendekatan statistik untuk melakukan inferensi induksi pada persoalan klasifikasi (Annur, 2018).

Berikut ialah persamaan dari *teorema Bayes* (Saleh, 2015):

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

Keterangan :

- X = Sampel data dengan class (label) yang belum diketahui
- H = Hipotesis bahwa X adalah data suatu class (label) spesifik
- P(H|X) = Probabilitas Hipotesis H berdasar kondisi X (posterior probabilitas)
- P(H) = Probabilitas hipotesis H (prior probabilitas)
- P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) = Probabilitas X

Tahapan dari proses Algoritma

Naive Bayes adalah:

1. Menghitung jumlah kelas/label
2. Menghitung Jumlah kasus per kelas
3. Kalikan semua variabel kelas
4. Bandingkan hasil per kelas

Teori Bayesian mempunyai beberapa kelebihan yaitu:

1. Mudah untuk dipahami.
2. Hanya memerlukan pengkodean yang sederhana.
3. Lebih cepat dalam penghitungan.

2.4. Evaluasi Model

Fungsi dari *confusion matrix* ialah 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* adalah nilai akurasi dengan *class* yang telah diprediksi. Sedangkan nilai *recall* merupakan presentase nilai kinerja keberhasilan algoritma yang dipakai (Saleh, 2015). Tabel berikut ini merupakan *confusion table* :

Tabel 1. *Confusion Table*

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

Keterangan :

- TP = *True Positive*
- FP = *False Positive*
- FN = *False Negative*
- TN = *True Negative*

Rumus *Accuracy* :

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

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

$$\text{Rumus Precision : Precision} = \frac{TP}{(TP+FP)}$$

Receiver Operating Characteristic (ROC Curve) yang digunakan untuk menghasilkan evaluasi dalam bentuk grafik. Pada kurva yang dihasilkan dari ROC terdapat dua dimensi kurva yakni nilai True Positive (TP) ditempatkan pada sumbu Y dan nilai False Positive (FP) ditempatkan pada sumbu X. Klasifikasi nilai dari akurasi AUC sendiri dibagi menjadi 5 kelompok antara lain:

1. 0.90-1.00 = Excellent Classification
2. 0.80-0.90 = Good Classification
3. 0.70-0.80 = Fair Classification
4. 0.60-0.70 = Poor Classification
5. 0.50-0.60 = Failure

2.5. Rapid Miner

Rapid Miner adalah salah satu software untuk pengolahan data mining. Pekerjaan yang dilakukan oleh Rapid Miner text mining adalah berkisar dengan analisis teks, mengekstrak pola-pola dari data yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan dan database. Tujuan dari analisis teks ini adalah untuk mendapatkan informasi bermutu tertinggi dari teks yang diolah (D & Nurcahyo, 2015).

3. Desain Penelitian/ Metodologi

3.1. Metode Penelitian

Metode penelitian merupakan cara ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu. Metode penelitian juga dapat diartikan sebagai suatu cara atau prosedur yang dipergunakan untuk melakukan penelitian. Sesuai dengan tahapan penelitian dalam data mining, sebelum melakukan penelitian terlebih dahulu peneliti menyiapkan berbagai rancangan penelitian seperti rencana atau hal-hal yang harus dilakukan dan dibutuhkan sebelum, selama dan setelah penelitian selesai juga, rancangan penelitian lainnya agar penelitian dapat dilakukan sesuai dengan rumusan masalah dan tujuan penelitian. Berikut tahapan dalam melakukan penelitian data mining:



Gambar 1. Tahapan Penelitian

1. Pengumpulan data
Tahapan ini menerangkan tentang dari mana sumber data dalam penelitian ini didapatkan dan menemukan informasi yang bisa digunakan untuk penelitian.
2. Pengolahan data awal
Tahapan ini menerangkan tentang tahap awal dalam data mining. Pengolahan data awal meliputi proses input data keformat yang dibutuhkan, penggabungan data dan data training.
3. Metode yang diusulkan
Tahapan ini dijelaskan penggunaan metode algoritma naive bayes pada penelitian.
4. Pengujian dan Validasi hasil Tahapan ini menjelaskan tentang pengujian, hasil pengujian akan divalidasi dan kemudian dievaluasi.

3.2. Pengumpulan Data

Terdapat beberapa teknik dalam pengumpulan data diantaranya melalui pengamatan atau observasi,

wawancara, penelitian kepustakaan, dan juga kuesioner. Namun pada penelitian ini peneliti hanya menggunakan beberapa teknik pengumpulan data, diantaranya sebagai berikut :

1. Observasi
Peneliti melakukan pengamatan langsung terhadap proses penginputan stok barang ATK yang dilakukan oleh pegawai toko Citra Media. Stok barang yang ada pada gudang dicatat dalam satu buku lalu diinput satupersatu dalam tabel Ms. Excel. Dari data tersebut peneliti kumpulkan dan melakukan pengujian untuk menghasilkan data yang sesuai dengan tujuan penelitian.
2. Penelitian kepustakaan (*libraryresearch*)
Peneliti mengumpulkan bahan- bahan pustaka, *literature*, dan karangan ilmiah yang berkaitan dengan penelitian ini dari berbagai sumber. Diantaranyakni, pada perpustakaan kampus, internet, buku, jurnal, maupun karya ilmiah terdahulu yang bertujuan untuk membahas permasalahan yang bersifat teori dengan membaca, mengumpulkan dan menarik kesimpulan. Sehingga dapat memperluas pengetahuan dan dapat membantu memecahkan permasalahan yang ada pada penelitian ini.
3. Kuisisioner
Peneliti mengajukan beberapakolom tabel yang merupakan pertanyaan mengenai stok ATK pada para pegawai toko Citra Media untuk dapat mengklasifikasi data dengan mudah melalui jawaban yang diberikan oleh para pegawai.
4. Wawancara (*Interview*)
Peneliti melakukan wawancara kepada pemilik toko Citra Media mengenai permasalahan yang dihadapi toko terkait penjualan dan data penjualan yang akan diteliti. Juga beberapa hal seputar toko yang dibutuhkan dalam penelitian.

3.3. Pengolahan Data Awal

Pada tahap ini menjelaskan tentang tahap awal yang dilakukan pada proses data minig. Data yang didapat dari Toko Citra Media akan diolah menjadi format yang dibutuhkan dan dilakukan penentuan atribut data yang akan digunakan. Berikut tahapan dalam pengolahan data awal :

1. Select Data

Pada tahap ini akan dilakukan pemilihan data terhadap atribut atau variabel yang akan dianalisis. Karena tidak semua data yang didapat dari sumber objek penelitian akan digunakan. Data yang digunakan akan diseleksi dengan cara melihat kecenderungan data / kesesuaian data dengan topik / judul penelitian yang akan diteliti oleh peneliti. Sehingga didapat beberapa variabel yang akan dijadikan sebagai atribut.

Tabel 2. Pemilahan Data

Atribut / Variabel	Indikator	Detail Penggunaan
Kode Barang	X	-
Barcode	X	-
Nama Barang	X	-
Merek	X	-
Jenis	V	Nilai Model
Satuan	V	Nilai Model
Harga	V	Nilai Model
Kualitas	X	-
Kegunaan	X	-
Peminatan	X	Nilai Model
Klasifikasi	V	Label Target

Pada Tabel 3.1 menerangkan atribut atau variabel yang akan digunakan dan tidak digunakan menurut pandangan peneliti. Indikator "V" menerangkan bahwa atribut atau variabel tersebut digunakan, sedangkan indikator "X" menerangkan bahwa atribut atau variabel tersebut tidak diperlukan atau tereliminasi dalam tahap pengolahan data awal. Pengeliminasian beberapa atribut tersebut berdasarkan nilai model yang relatif samadan tidak mempengaruhi hasil dari proses penelitian.

2. *Pre-pocessing / Cleaning Data* Tahap *pre-processing* merupakan tahap dilakukannya pembersihan data terhadap data yang *missing value* yaitu data yang tidak konsisten atau kosong dan juga memperbaiki

data yang rusak. Proses pembersihan data dilakukan untuk memastikan bahwa data yang telah dipilih layak untuk diolah berdasarkan proses *datamining*.



Gambar 2. Proses Data *Cleaning* Pada Atribut Kualitas, Kegunaan, dan Peminatan



Gambar 3. Proses Data *Cleaning* Pada Atribut Harga

3. Transformation Data

Adapun pada tahap ini data yang diperoleh akan diubah dan digabung ke dalam format yang sesuai untuk diproses dalam data mining. Atribut yang dipakai akan diberi label sesuai dengan topik / judul penelitian yang dilakukan. Karena dalam penelitian ini akan dilakukan uji coba menggunakan tools *Rapid Miner*, maka data yang telah melalui proses sebelumnya akan ditransformasi agar dapat sesuai dengan algoritma yang dipakai yaitu algoritma *naive bayes*. Adapun label yang digunakan yakni Label / class "Klasifikasi" dengan kategori *Laris* dan *Tidak Laris*.



Gambar 4. Proses Transformasi Data

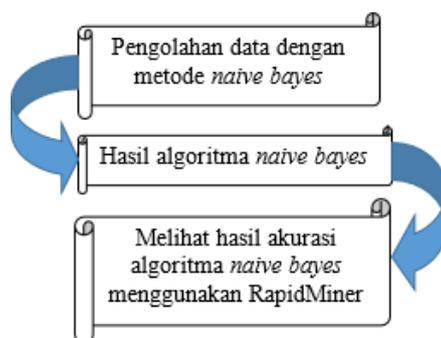
4.2 Metode yang diusulkan

Dalam penelitian ini akan dilakukan perhitungan pengklasifikasian menggunakan

metode algoritma *naive bayes* dengan kriteria-kriteria yang telah ditentukan. Data dihitung dengan menggunakan algoritma sesuai dengan metodenya kemudian dicari hasil akurasi.

4.3 Pengujian Metode

Pengujian dilakukan untuk mengetahui hasil perhitungan data yang diklasifikasi berdasarkan data yang telah dilakukan pengolahan data awal, dan juga untuk mengetahui apakah fungsi bekerja dengan baik atau tidak. Dalam tahapan ini akan dilakukan beberapa langkah pengujian data yaitu :



Gambar 5. Langkah PengujianData

Pada penelitian ini proses pengujian dilakukan dengan 2 carayakni dengan melakukan perhitungan secara manual, kemudian data diuji menggunakan *software tools RapidMiner* untuk memastikan apakah hasil perhitungan manual sesuai atau tidak dengan hasil yang diperoleh *RapidMiner*.

4.4 Evaluasi dan Validasi Hasil

Evaluasi dilakukan dengan caramenganalisa hasil dari algoritma yangdigunakan untuk memastikan bahwa hasil perhitungan dan pengujian sesuai dengan tujuan penelitian. Validasi dilakukan untuk mengukur hasil klasifikasi guna mengetahuitingkat *accuracy*, *precision*, dan *recall*.

5. Hasil Penelitian dan Pengujian

5.1 Hasil Penelitian

Hasil penelitian ini didapat dari 2 metode pengujian yakni perhitungan yang dilakukan dengan menggunakan software *Ms. Excel* yang kemudian di implementasikanpada *software tools RapidMiner* dengan total jumlah *dataset* 1303 data. Data dibagi kedalam 2 bagian yakni data *training* dan data *testing*. Masing-masing metode pengujian dilakukan sebanyak 4 kali pengujian.

Tabel 3. Hasil Penelitian

Hasil Pengujian								
Ms. Excel				RapidMiner				
Pengujian	Data training	Data testing	accuracy	Pengujian	Accuracy	Precision	Recall	AUC
1	1000	303	94%	1	94.39%	100%	85.95%	1.00
2	900	403	99%	2	99.26%	97.37%	100%	0.997
3	700	603	98%	3	98.84%	98.84%	97.16%	0.997
4	500	803	97%	4	97.76%	98.78%	94.16%	0.994

Berdasarkan hasil yang telah diperoleh dapat dilihat bahwa hasilpresentase *accuracy* dari kedua metode tidak jauh berbeda dan keduanyaapun sama-sama memiliki nilai *accuracy* tertinggi pada pengujian kedua dan nilai *accuracy* terendah pada pengujian pertama.

5.2 Hasil Pengujian

a. Perhitungan Algoritma Naive Bayes

Sesuai dengan proses perhitungan algoritma *naive bayes* tahapan- tahapan yang dilakukan dalam proses perhitungan data mining padapenelitian ini adalah sebagai berikut :

1. Baca data *training*

Untuk menentukan data yang nantinya akan diklasifikasi dengan metode *naive bayes* maka langkah

pertama yang dilakukan adalah membaca data latih (*training*). Adapun data latih yang digunakan secara keseluruhan berjumlah 1000*record*, data terlampir. Berikut sampel sebagian tabel data *training*:

Tabel 4. Data Training

NAMA BARANG	JENIS	SATUAN	HARGA	PEMINATAN	KLASIFIKASI
BK TULIS BOXY ALGOLD 36	AS	PACK	Standar	Banyak	Laris
BK TULIS DODO 58	AS	PACK	Standar	Banyak	Laris
BK TULIS KIKY 38	AS	PACK	Standar	Banyak	Laris
BK TULIS LA 38	AS	PACK	Standar	Banyak	Laris
BK TULIS MMC 42	AS	PACK	Standar	Banyak	Laris
BK TULIS SA 38	AS	PACK	Standar	Banyak	Laris
BK TULIS SIDU 32	AS	PACK	Standar	Banyak	Laris
BK TULIS SIDU 38	AS	PACK	Standar	Banyak	Laris
BK TULIS SIDU 58	AS	PACK	Standar	Banyak	Laris
BK TULIS SKOLA 58	AS	PACK	Mahal	Sedikit	Tidak Laris
BK TULIS TIARA 38	AS	PACK	Mahal	Sedikit	Tidak Laris
BK TULIS TIARA KAMPUS	AS	PACK	Standar	Banyak	Laris
BLOCK NOTE	ATK	PCS	Mahal	Sedikit	Tidak Laris
BLOGRAFI ALI BIN	ATK	PCS	Mahal	Sedikit	Tidak Laris
BLUETOOTH	OLG	PCS	Mahal	Sedikit	Tidak Laris
BOLA BASKET	OLG	PCS	Standar	Sedikit	Tidak Laris
BOLA FUTSAL	OLG	PCS	Standar	Sedikit	Tidak Laris

Keterangan :

- a. Kriteria 1 menjelaskan tentang ‘Jenis’

Tabel 5. Kriteria Jenis

Kriteria	Kategori	Keterangan
JENIS	eATK	Alat Tulis Kantor
	AS	Alat Sekolah
	OLG	Perlengkapan Olahraga

- b. Kriteria 2 menjelaskan tentang ‘Satuan’

Tabel 6. Kriteria Satuan

Kriteria	Kategori
SATUAN	Pcs
	Pack

	Unit
	Box
	Tube
	Rim
	Roll
	Set
	Lusin

	Botol
--	-------

c. Kriteria 3 menjelaskan tentang ‘Harga’

Tabel 7. Kriteria Harga

Kriteria	Kategori	Keterangan
HARGA	Murah	< Rp. 50.000
	Standar	Rp. 50,000 < Rp. 150,000
	Mahal	>Rp. 150.000

d. Kriteria 4 menjelaskan tentang ‘Peminatan’

Tabel 8. Kriteria Peminatan

Kriteria	Kategori
PEMINATAN	Banyak
	Sedikit

Seperti yang sudah dijelaskan bahwa tahap awal proses perhitungan *naive bayes* yakni dengan melakukan pengambilan data latih dari *dataset* yang diperoleh dengan variabel atau atribut yang telah dilakukan pembersihan data terlebih dahulu.

2. Menghitung probabilitas prior ($P(C_i)$)

Perhitungan ini merupakan tahap pertama dalam proses perhitungan algoritma *naive bayes*. Dengan tujuan untuk menentukan klasifikasi dengan metode *naive bayes* guna mencari probabilitas dari masing-masing kelas/label. Dalam menentukan klasifikasi data akan ditentukan probabilitas kelas C_0 “Laris” dan probabilitas kelas C_1 “Tidak Laris”. Perhitungannya yakni mengklasifikasikan produk yang dijual dengan mencari berapa jumlah yang “Laris” dan “Tidak Laris” dari total data training dan membaginya dengan total keseluruhan dataset yang diperoleh. Berikut persamaan perhitungan probabilitas:

Diketahui :

Total data *training* = 1000

Kelas C_0 “Laris” = 692

Kelas C_1 “Tidak Laris” = 308

Probabilitas kelas C_0 “Laris” adalah :

$$P(C_0) = \frac{692}{1000} = 0.692$$

Probabilitas kelas C_1 “Tidak Laris” adalah :

$$P(C_1) = \frac{308}{1000} = 0.308$$

Tabel 9. Probabilitas Klasifikasi

P(LARIS/TIDAKLARIS)	LARIS	TIDAK LARIS
	69%	31%
100%		

Berdasarkan perhitungan di atas diketahui nilai probabilitas kelas “Laris” ialah 0.692 dan probabilitas kelas “Tidak Laris” ialah 0.308 atau dalam bentuk presentase probabilitas kelas “Laris” 69% dan probabilitas kelas “Tidak Laris” 31% sesuai dengan penjelasan dalam Tabel 4.7.

3. Menghitung Probabilitas *Posterior* X bersyarat C ($P(X/C_i)$)

Perhitungan probabilitas *posterior* dilakukan pada data training sebanyak 1000 data dengan menggunakan X sebagai vector pemilihan kriteria produk penjualan ATK yaitu : X_{jenis}, X_{satuan}, X_{harga}, X_{peminatan}. Sehingga $P(X/C_i)$ dapat dijabarkan menjadi $P(X_{jenis}, X_{satuan}, X_{harga}, X_{peminatan} / C_i)$. Untuk setiap X dihitung kemungkinannya pada setiap C_i . Proses perhitungan ini merupakan tahapan kedua dalam proses algoritma naive bayes, dengan kata lain yakni tahapan menghitung jumlah kasus per kelas. Dengan cara menghitung nilai akumulasi peluang dari setiap kelas ($P(X/C_i)$). Berikut tabel-tabel perhitungan probabilitas posterior ($P(X/C_i)$) masing-masing kriteria:

Tabel 10. Probabilitas Jenis

Jenis	Klasifikasi		Probabilitas	
	Laris	Tidak Laris	Laris	Tidak Laris
ATK	603	255	0,8713873	0,827922078
AS	74	41	0,1069364	0,133116883
OLG	15	12	0,0216763	0,038961039
Total	692	308	100%	100%

Tabel 11. Probabilitas Harga

Harga	Klasifikasi		Probabilitas	
	Laris	Tidak Laris	Laris	Tidak Laris
Murah	273	134	0,3945087	0,435064935
Standar	382	128	0,5520231	0,415584416
Mahal	37	46	0,0534682	0,149350649
Total	692	308	100%	100%

Tabel 12. Probabilitas Peminatan

Peminatan	Klasifikasi		Probabilitas	
	Laris	Tidak Laris	Laris	Tidak Laris
Banyak	679	0	0,9812139	0
Sedikit	13	308	0,0187861	1
Total	692	308	100%	100%

4. Mengalikan semua variabel data testing $P(X/C_i) \cdot P(C_i)$

Data testing yang digunakan dalam pengujian ini berjumlah 303 data. Berikut ini merupakan sebagian tabel data testing. Untuk tabel data testing selengkapnya dapat dilihat pada lembar lampiran.

Tabel 13. Data Testing

NAMA BARANG	JENIS	SATUAN	HARGA	PEMINATAN	KLASIFIKASI
SOLASI KABEL	ATK	ROLL	Murah	Banyak	Laris
SOLASI KABEL	ATK	ROLL	Murah	Banyak	Laris
SOLASI SET	ATK	ROLL	Murah	Banyak	Laris
SOLASI SET FANCY	ATK	ROLL	Murah	Banyak	Laris
SOLATOP	ATK	ROLL	Murah	Banyak	Laris
SPIDOL UJASCO	ATK	PACK	Standar	Banyak	Laris
SPIDOL ABG SNOWMAN	ATK	PACK	Standar	Banyak	Laris
SPIDOL ECO PERMANEN	ATK	PACK	Standar	Banyak	Laris
SPINNER	ATK	PCS	Murah	Banyak	Laris
STABLO K HL-34	ATK	PACK	Murah	Sedikit	Tidak Laris
STABLO K HL-46	ATK	PACK	Murah	Banyak	Laris
STABLO XOYKO TWIN	ATK	PACK	Murah	Banyak	Laris
STABLO NEON	ATK	PACK	Murah	Banyak	Laris
TAS FILE SCM	ATK	PCS	Mahal	Sedikit	Tidak Laris
TAS FILE XOYKO	ATK	PCS	Mahal	Sedikit	Tidak Laris
TAS IMPORT 2 RESL L	ATK	PCS	Mahal	Sedikit	Tidak Laris
TAS IMPORT 2 RESL M	ATK	PCS	Mahal	Sedikit	Tidak Laris
TAS IMPORT BULAT M	ATK	PCS	Mahal	Sedikit	Tidak Laris

Dari contoh yang diambil secara acak diatas masing-masing kategori telah dilakukan perhitungan probabilitas di pembahasanebelumnya, berikut peneliti ulas kembali probabilitas dari masing-masing kategori berdasarkan klasifikasi laris dan tidak laris. Selanjutnya akan dilakukan perhitungan tahap mengalikan semuavariabel dari contoh data yang diambil.

Diketahui Probabilitas KriteriaLaris :

$$P(\text{Jenis} = \text{"ATK"} | \text{Laris}) = 0,8713873$$

$$P(\text{Satuan} = \text{"Pack"} | \text{Laris}) = 0,3511561$$

$$P(\text{Harga} = \text{"Murah"} | \text{Laris}) = 0,3945087$$

$$P(\text{Peminatan} = \text{"Sedikit"} | \text{Laris}) = 0,0187861$$

Diketahui Probabilitas KriteriaTidak Laris:

$$P(\text{Jenis} = \text{"ATK"} | \text{Tidak Laris}) = 0,827922078$$

$$P(\text{Satuan} = \text{"Pack"} | \text{Tidak Laris}) = 0,204545455$$

$$P(\text{Harga} = \text{"Murah"} | \text{Tidak Laris}) = 0,435064935$$

$$P(\text{Peminatan} = \text{"Sedikit"} | \text{TidakLaris}) = 1$$

Perhitungan :

- a. Mengalikan semua atribut "Laris"

$$P(X|C_0) = P(X|\text{Laris})$$

$$= P(X_{\text{jenis}} = \text{"ATK"} | \text{Laris}) \times P(X_{\text{satuan}} = \text{"Pack"} | \text{Laris}) \times P(X_{\text{harga}} = \text{"Murah"} | \text{Laris}) \times$$

$$P(X_{\text{peminatan}} = \text{"sedikit"} | \text{Laris})$$

$$= (0,8713873 \times 0,3511561 \times 0,3945087 \times 0,0187861)$$

$$= 0,00226779951$$

- b. Mengalikan semua atribut "TidakLaris"

$$P(X|C_1) = P(X|\text{Tidak Laris})$$

$$= P(X_{\text{jenis}} = \text{"ATK"} | \text{Tidak Laris}) \times P(X_{\text{satuan}} = \text{"Pack"} | \text{Tidak Laris}) \times P(X_{\text{harga}}$$

$$= \text{"Murah"} | \text{Tidak Laris}) \times P(X_{\text{peminatan}} = \text{"sedikit"} | \text{Tidak Laris})$$

$$= (0,827922078 \times 0,204545455 \times 0,435064935 \times 1)$$

$$= 0,0736112453$$

Setelah mengalikan semua variabel maka tahap selanjutnya yakni dilakukan perhitungan pemaksimalan untuk kemungkinan klasifikasi kedalam kelas "Laris" dan "Tidak Laris". Persamaan yang digunakan ialah sebagai berikut:

Persamaan kelas “Laris” : $P(X|C_i). P(C_i)$

Persamaan kelas “Tidak Laris” : $P(X|C_0). P(C_0)$

$P(X|C_1). P(C_1)$

Keterangan :

- $P(X|C_i)$: Probabilitas X berdasarkan kondisi pada hipotesis C_i
- $P(C_i)$: Probabilitas prior $P(C_0)$: Probabilitas Laris $P(C_1)$:
- Probabilitas Tidak Laris
- X : Kriteria kejadian

Diketahui :

$P(C_0) = 0.692$

$P(C_1) = 0.308$

Perhitungan :

a. Perhitungan pemaksimalan untuk kemungkinan klasifikasi kelas “Laris” ($P(C_0|x)$)

$$\begin{aligned} P(C_0|x) &= P(X|C_0). P(C_0) \\ &= 0.00226779951 \times 0.692 \\ &= 0,001569319 \end{aligned}$$

b. Perhitungan pemaksimalan untuk kemungkinan klasifikasi kelas “Tidak Laris” ($P(C_1|x)$)

$$\begin{aligned} P(C_1|x) &= P(X|C_1). P(C_1) \\ &= 0.0736112453 \times 0.308 \\ &= 0,022692591 \end{aligned}$$

Dari perhitungan diatas dihasilkannilai $P(C_0|x) = 0,001569319$ sedangkan nilai $P(C_1|x) = 0,022692591$ maka dapat disimpulkan bahwa nilai $P(C_0|x) < P(C_1|x)$ atau nilai kelas “Laris” lebih rendah daripada nilai kelas “Tidak Laris” sehingga data uji tersebut diklasifikasikan kedalam kelas “Tidak Laris” dalam penjualan ATK.

Setelah melakukan berbagai tahap perhitungan kini mulai menghitung data testing dengan mencari nilai akurasi dari setiap pengujian yang dilakukan. Untuk pengujian metode dengan *tools Ms. Excel* berikut pembahasan dan hasilnya:

Tabel 14. Akurasi Pengujian dengan

PREDICTE D	CLASS	
	LARI S	TIDAK LARIS
LARIS	182	17
TIDAK LARIS	0	104
ACCURAC Y	94%	

Sesuai dengan tabel 4.14 pada perhitungan pengujian pertama menggunakan *Ms. Excel* didapatkan nilai akurasi sebesar 94% dengan rincian perhitungan sebagai berikut:

$$\begin{aligned}
 \text{Accuracy} &= \frac{\text{class laris} + \text{class tidak laris}}{\text{(total data testing)}} = \frac{182 + 104}{303} \\
 &= 0.9416 \times 100 \\
 &= 94\%
 \end{aligned}$$

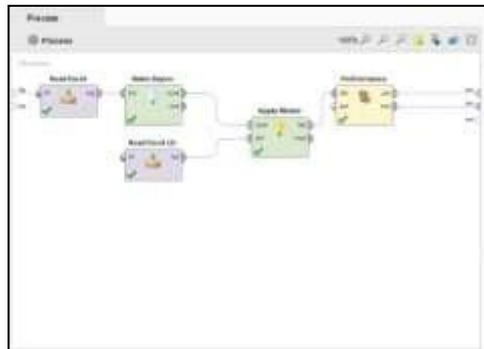
Berikut analisa hasil pengujian dengan Ms. Excel :
Tabel 4.18 Hasil Pengujian dengan

Pengujian	Data training	Data testing	accuracy
1	1000	303	94%
2	900	403	99%
3	700	603	98%
4	500	803	97%

Berdasarkan tabel 4.17 diatas dapat dilihat bahwa hasil dari keempat pengujian yang telah dilakukan perhitungan menggunakan Ms. Excel didapatkan nilai *accuracy* tertinggi yakni pada pengujian kedua sebesar 99%. Dan *accuracy* terendah yakni pada pengujian pertama yakni 94%.

b. Pengujian Metode Dengan RapidMiner

Langkah selanjutnya yang dilakukan setelah menghitung dengan Ms. Excel yakni mengimplementasikan metode algoritma *naive bayes* menggunakan tools *RapidMiner*. *RapidMiner* merupakan salah satu software pengolahan data yang mempunyai tingkat perhitungan terbaik dalam mengklasifikasi dan memprediksi data mining dengan cepat dan akurat. Berikut tahapan dalam proses implementasi metode *naive bayes* menggunakan *RapidMiner*.



Gambar 6. Proses Rapidminer

accuracy: 94.38%			
	true Tidak Laris	true Laris	class pred
pred Tidak Laris	104	0	100.00%
pred Laris	17	182	91.46%
class recall	86.87%	100.00%	

Gambar 4.2 Hasil Pengujian Rapidminer

Diketahui :
TP = 104
TN = 182
FP = 0

FN = 17

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

$$= \frac{104+182}{(104+17+182+0)} = \frac{286}{303}$$

$$= 0.94389 \times 100$$

$$= 94.39\%$$

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

$$= \frac{104}{(104+0)} = \frac{104}{104}$$

$$= 1 \times 100$$

$$= 100\%$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{104}{(104+17)}$$

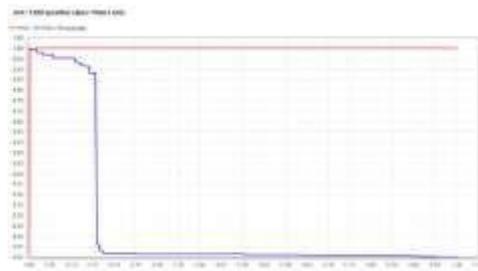
$$= \frac{104}{121} = 0.85950 \times 100$$

$$= 85.95\%$$

c. Evaluasi Hasil Akhir Klasifikasi

Evaluasi hasil akhir klasifikasidari data penjualan ATK dengan total *dataset* 1303 data terdapat 34 data erorr dengan kriteria klasifikasi

“Laris” sebanyak 5 data error dan “Tidak Laris” sebanyak 29 data eror. Dan hasil akhir pengujian



Gambar 7. Hasil Kurva ROC/AUC

Gambar 4.3 menjelaskan hasil dari pengujian pertama menggunakan *RapidMiner* yang menghasilkanprediksi kurva AUC sebesar 1.00 yang termasuk dalam kategori klasifikasi sangat baik (*Excellent Classification*). menyisakan 1296 data Ok. Maka yang termasuk dalam kategori “Laris” yakni berjumlah 870 data sedangkan kategori “Tidak Laris” sejumlah 399.

Tabel 15. Data Penjualan PalingLaris

NO	NAMA BARANG	JENIS	SATUAN	HARGA	PEMINATAN	KLASIFIKASI
1	LEM AIDON KALENG 1/2 KG	ATK	PCS	Standar	Banyak	Laris
2	AKRILIK A4	ATK	PCS	Standar	Banyak	Laris
3	ALBUM 4K 10	ATK	PCS	Standar	Banyak	Laris
4	ALBUM 4K 20	ATK	PCS	Standar	Banyak	Laris
5	ALBUM FOTO 2	ATK	PCS	Standar	Banyak	Laris
6	ALBUM FOTO M	ATK	PCS	Standar	Banyak	Laris
7	ALBUM FOTO L	ATK	PCS	Standar	Banyak	Laris
8	ALBUM FOTO K	ATK	PCS	Standar	Banyak	Laris
9	ALFALINK CD300	ATK	PCS	Standar	Banyak	Laris
10	BATERAI ALKALINE (B L&R)	ATK	PCS	Standar	Banyak	Laris
140	CRAYON FABER 12	ATK	PACK	Standar	Banyak	Laris
141	CRAYON FABER 34	ATK	PACK	Standar	Banyak	Laris
142	CRAYON FC 12 DOFF	ATK	PACK	Standar	Banyak	Laris
388	GUNTING EMIGO 165	ATK	PCS	Murah	Banyak	Laris
370	CAT AIR V-TRO	ATK	BOTOL	Murah	Banyak	Laris

6. Kesimpulan

Berikut kesimpulan dari hasil penelitian :

1. Penerapan *data mining* pada *tools RapidMiner* menggunakan *algoritma naive bayes* sangat cocok diterapkan dalam mengklasifikasi data dengan jumlah yang cukup banyak seperti data penjualan ATK. Karena dapat menentukan produk mana saja yang termasuk dalam kategori Laris maupun Tidak Laris tanpa memerlukan waktu yang lama dan lebih efisien dalam pengolahan data, sehingga *owner* Citra Media dapat lebih mudah dalam memilih produk yang mampu menunjang dan dapat meningkatkan keuntungan bisnisnya.
2. Perhitungan menggunakan metode *naive bayes* dalam klasifikasi penjualan ATK terlaris memperoleh hasil yang sangat baik. Dari pengujian yang dilakukan pada *tools RapidMiner* menghasilkan tingkat akurasi paling tinggi sebesar 99.26% dan akurasi paling rendah sebesar 94.39%. Sedangkan evaluasi dengan kurva ROC/AUC hasil dari empat kali pengujian yang telah dilakukan termasuk dalam kategori klasifikasi sangat baik (*excellent classification*) yakni dengan evaluasi tertinggi sebesar 0.997.
3. Dalam penggunaan *dataset* yang memiliki jumlah cukup besar harus lebih diperhatikan dalam penelitian huruf setiap katanya terutama pada *class* label. Karena *tools RapidMiner* tidak dapat membaca kata atau huruf yang berbeda atau tidak sesuai.
4. Penerapan *algoritma naive bayes* sebaiknya menggunakan data yang memiliki jumlah cukup banyak. Karena semakin banyak jumlah data tingkat akurasi semakin tinggi.

Daftar Pustaka

- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 1.
- D, N., & Nurcahyo, G. W. (2015). *Algoritma Data Mining dan Pengujian*. Deepublish.
- Gullo, F. (2015). For Pattern In Data To Knowledge Discovery: What Data Mining Can Do. *Physics Procedia* 62, 18-22
- Ridwan, M., Suyono, H., & Sarosa,
- M. (2017). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*, 59-64.
- S, A., & Y, A. S. (2018). Implementasi Data Mining Menggunakan WEKA. *UB Press*.
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*.
- Vulandari, R. T. (2017). *Data Mining Teori dan Aplikasi Rapidminer* (Vol. 1). Yogyakarta: Gava Media.