

Komparasi Algoritma Naïve Bayes Dan K-Nearest Neighbor Untuk Menentukan Klasifikasi Produk Terlaris (Studi Kasus: Perusahaan Frozen Food XYZ)

Sena Melianta Aji*, Raka Yusuf

Informatika, Universitas Mercu Buana, Jakarta

*Senaaji2000@gmail.com

Abstrak— Penjualan pada frozen food di kalangan masyarakat semakin meluas karena kehidupan masyarakat saat lebih memilih makanan yang praktis dan cepat saji. Salah satu yang dapat kita manfaatkan adalah teknologi dalam menggali informasi yang bermanfaat dari gudang data perusahaan penjualan frozen food. Strategis yang dapat dilakukan adalah dengan mempelajari pola perilaku konsumen. Pola tersebut dapat diketahui dengan memanfaatkan data transaksi penjualan di perusahaan PT. Frozen Food XYZ. Demikian besar transaksi harian yang terus bertambah. Transaksi yang banyak tersebut akan mempersulit pelaku perusahaan dalam mengolah data mereka. Tujuan penelitian ini adalah mencoba menerapkan teknik metode algoritma naïve bayes dan k-nearest neighbors memberikan informasi berupa klasifikasi penjualan produk frozen food yang paling laris di kalangan masyarakat dan tidak laris di kalangan masyarakat (laris dan tidak laris). Dari rata – rata accuracy yang ada dapat dilihat bahwa nilai rata rata accuracy algoritma Naïve Bayes adalah 77%, sedangkan algoritma K-Nearest Neighbor nilai rata – rata accuracy sebesar 99%.

Kata Kunci— Frozen Food; Klasifikasi; KNN; Naïve Bayes; Transaksi.

DOI: 10.22441/jitkom.v9i1.009

Article History:

Received: July 7, 2022

Revised: Aug 13, 2023

Accepted: Sept 30, 2023

Published: Oct 24, 2024

I. PENDAHULUAN

Perkembangan frozen food di pusat perbelanjaan bermunculan dengan berbagai macam bentuk dan ukuran yang menyebabkan persaingan dalam dunia ritel semakin ketat. Bisnis atau biasa disebut dengan perusahaan semakin terasa keberadaannya dalam kehidupan masyarakat sehari-hari. Beberapa contoh bentuk pusat perbelanjaan yang meramaikan dunia ritel seperti minimarket, supermarket. Perusahaan yang bergerak dalam bidang ritel dengan konsep menjual produk makanan (*food*) dari segi pemasaran memiliki peran penting sebagai salah satu lembaga pemasaran dalam pemasaran produk. Dalam fungsi-fungsi inti pemasaran ini dilakukan upaya-upaya yang meliputi peningkatan penjualan hasil produk dan menciptakan semakin banyak pelanggan dan pemasaran sangat penting untuk menjamin perkembangan perusahaan dimasa yang akan datang.

Salah satu strategis yang dapat dilakukan adalah dengan mempelajari pola perilaku konsumen. [1] Pola tersebut dapat diketahui dengan memanfaatkan data transaksi penjualan di perusahaan PT. Frozen Food XYZ. Demikian besar transaksi harian yang terus bertambah. Transaksi yang banyak tersebut akan mempersulit pelaku perusahaan dalam mengolah data mereka.

Prioritas pada peningkatan penjualan dengan menarik konsumen baru atau berorientasi pada upaya mempertahankan pelanggan yang telah ada. Namun, perlu diketahui bahwa konsumen baru tersebut tidak bisa menjanjikan akan terus menggunakan produk perusahaan dengan demikian pelaku bisnis ritel lebih baik mempertahankan pelanggan yang telah ada

karena mampu memberikan keuntungan yang lebih besar dari perusahaan. [2]

Dalam proses penelitian ini akan menggunakan Algoritma Naïve Bayes dan [3] [4] K-Nearest Neighbor [5] [6]. Algoritma Naïve Bayes merupakan suatu metode independen yang membahas mengenai klasifikasi sederhana berdasarkan teori pada bayes [7] Algoritma ini tergolong cepat dan sederhana dengan menggunakan fitur yang dapat berdiri sendiri. Sedangkan K-Nearest Neighbor merupakan klasifikasi yang menyimpan semua *data training* dan melakukan klasifikasi dengan cara membandingkan antara atribut yang paling cocok dengan atribut *record* dan *data training*. [8][9][10] Dengan keunggulan masing-masing algoritma diatas penelitian ini menggunakan kedua algoritma untuk implementasi dalam mengklasifikasi data penjualan *frozen food*.

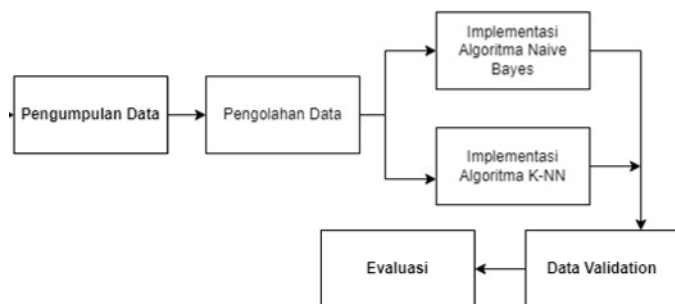
II. LITERATURE REVIEW

Penerapan algoritma klasifikasi dalam *data mining* telah terbukti efektif untuk memprediksi kepuasan pelanggan serta mengoptimalkan manajemen persediaan barang di berbagai sektor industri. Dalam analisis kepuasan dan perilaku konsumen, metode Naïve Bayes sering menjadi pilihan utama. Penelitian [1] menerapkan metode Naïve Bayes untuk memprediksi tingkat kepuasan pelanggan pada transportasi umum inter-city PO. XYZ, yang berhasil menghasilkan tingkat akurasi tinggi sebesar 94,00%. Selaras dengan hal tersebut, berikutnya [2] juga memanfaatkan algoritma Naïve Bayes pada data penjualan kantin untuk mengidentifikasi pola pembelian konsumen, mengoptimalkan strategi menu, serta meningkatkan pengalaman pelanggan secara keseluruhan.

Di sisi lain, teknik klasifikasi juga memiliki peran krusial dalam perencanaan dan pengendalian stok produk. Kemudian [7] menggunakan algoritma *Naïve Bayes* untuk mengklasifikasikan penjualan produk vitamin menjadi kategori "LAKU" atau "TIDAK LAKU" sebagai pedoman bagi apotek dalam menentukan jumlah stok gudang. Sementara itu, [8] memilih pendekatan berbeda menggunakan algoritma *K-Nearest Neighbor* (K-NN) untuk memprediksi penjualan produk Unilever terlaris di Toko Rizky Barokah Nganjuk, dengan hasil akurasi tertinggi mencapai 86,66%. Secara keseluruhan, literatur ini menunjukkan bahwa baik *Naïve Bayes* maupun K-NN memberikan kontribusi data-driven yang signifikan dalam pengambilan keputusan bisnis.

III. METODOLOGI PENELITIAN

Metode penelitian yang digunakan pada penelitian ini sebagai berikut pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

A. Pengumpulan Data

Metode pengumpulan data yang digunakan pada penelitian ini merupakan transaksi Frozen Food Tahun 2021 yang menjadi acuan dalam penentuan penjualan tahun. Jumlah data ini sebanyak 1213 data yang tersebar ke dalam 9 kolom. Setiap kolom yang ada di dalam "data_penjualan" tidak memiliki nilai NaN atau nilai kosong di dalam kolom-kolom data. Tipe data yang ada di dalam "data_penjualan" adalah tipe data datetime (tipe data yang menyatakan entitas atau keterangan waktu), float, integer, dan object (string). Masing-masing keempat tipe data tersebut memiliki kolom yang berbeda-beda seperti kolom "TANGGAL" yang bertipe data datetime. Kemudian pada tipe data float yang dimiliki oleh kolom "HARGA (1 KG)". Tipe data integer yang dimiliki oleh keempat kolom yakni "NO", "JML", "QTY (KG)", dan "TOTAL HARGA". Sedangkan tipe data object atau string dimiliki oleh keempat kolom yang ada di dalam "data_penjualan" diantaranya adalah kolom "NAMA CUSTOMER", "ITEM", "SATUAN", dan "CATEGORY".

Tabel 1. Basis Data Terpadu

ATRIBUT	TIPE	DATA
NAMA CUSTOMER	OBJEK	Gracia Catring
TANGGAL	DATETIME	2021-09-29
ITEM	OBJECT	Dori
JML	INT	14
SATUAN	OBJECT	Pack

ATRIBUT	TIPE	DATA
CATEGORY	OBJECT	Ikan
HARGA (1KG)	FLOAT	37.000
QTY (KG)	INT	14
TOTAL HARGA	INT	518.000

B. Pengolahan Awal Data (Pre-Processing)

Data yang telah di peroleh di lakukan pre-processing. Preprocessing data adalah hal yang harus dilakukan dalam proses data mining, karena tidak semua data atau atribut data dalam data digunakan dalam proses data mining. Proses ini dilakukan agar data yang akan digunakan sesuai dengan kebutuhan. Terdapat beberapa jenis preprocessing yang akan digunakan dalam penelitian ini, antara lain:

Data Cleaning

Pada tahap data cleaning dilakukan pembersihan data dari *missing values* dan duplikat yang ada pada data penjualan di PT.Frozen Food XYZ. *Missing values* adalah kondisi dimana suatu atribut memiliki nilai yang tidak valid *Processing/Cleaning* merupakan proses pembersihan yang membuang duplikat data, memeriksa data yang konsisten, dan memperbaiki kesalahan pada data. Pengecekan *missing values* dilakukan pada keseluruhan dataset, sedangkan pengecekan duplikat dilakukan pada data pertahun.

Data Reduction

Pada tahap ini penelitian dilakukan reduksi data dari sejumlah dataset dengan menghilangkan kolom pada penelitian ini tidak semua atribut digunakan untuk kolom No, Satuan pada penelitian ini dihapus.

Data Transformation

Pada tahap ini dilakukan untuk mengubah tipe data pada kolom date dari yang semula string (object) menjadi tipe data pada kolom date. Setelah mengubah tipe data pada kolom status produk dari yang semula string (object) menjadi integer. Melakukan tahapan mengubah tipe data pada kolom "Harga (1 Kg)" dari yang semula float menjadi integer dan juga kolom "Status Item" dari yang semula object menjadi integer.

Tabel 2. Hasil Pre-processing

Atribut	Keterangan	Tipe Data
1	Nama Customer	Object
2	Tanggal	Datetime
3	Nama Item	Object
4	Jumlah Item Terjual	Int
5	Kategori	Object
6	Harga (1 Kg)	Int
7	Qty (1 Kg)	Int
8	Total Harga	Int

C. Implementasi Algoritma *Naïve Bayes* Dan K-NN

Implementasikan menggunakan 2 algoritma yaitu algoritma *Naïve Bayes* dan K-NN. Dalam implementasikan nilai x dan y ditentukan yaitu variabel independen dan dependen .Nilai x

variabel yang digunakan adalah Jumlah Item Terjual, Harga (1 Kg), dan Total Harga dan nilai y yang digunakan Status item.

Algoritma *Naïve Bayes*

Proses yang dilakukan dalam algoritma *Naïve Bayes* merupakan data yang dimasukan akan di hitung probabilitas dari setiap kelas. Setelah didapat maka hasil kelas akan dibandingkan. Jika nilai yang lebih besar maka itu Terlaris dan jika nilai lebih kecil maka tidak laris

Algoritma *K-NN*

Proses yang dilakukan dalam algoritma *K-Nearest Neighbor* merupakan dataset yang digunakan akan dijadikan sebagai data pembelajaran dimana terdapat label ya atau tidak. Lalu tentukan jumlah tetangga terdekat. Setelah menemukan K maka data testing akan dihitung jarak terhadap data training tersebut. Lalu dilihat jarak paling dekat dengan K yang telah ditentukan, lalu kelompok data testing tersebut berdasarkan label mayoritas pada

D. Data Validation

Validasi model digunakan teknik *Split Validation* yaitu teknik validasi yang membagi data menjadi dua bagian secara acak, sebagai *data training* dan sebagai lainnya secara *data testing*

E. Evaluasi

Evaluasi diperlukan untuk menganalisis dan mengukur sejauh mana peninjauan model mengenai hasil evaluasi yang telah dikerjakan. Akan di lakukan perbandingan kuantitatif dengan pertimbangan nilai komparasi. Sedang untuk model digunakan teknik *confusion Matrix*. Lalu hasil evaluasi akan diberikan dengan *classification report* yang terdiri dari Accuracy, Precision dan Recall.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$precision = \frac{TP}{TP + FP} \quad (2)$$

$$recall = \frac{TP}{TP + FN} \quad (3)$$

F. K-Means Clustering

Pada tahapan ini menglompokan (mengklaster) penjualan frozen food pada penelitian bertujuan membantu penelitian dalam menentukan label dari klasifikasi terlaris dan tidak laris dalam penjualan frozen food berdasarkan “Jml” dan “Total Harga” kemudian dimasukan dalam kolom status item sebagai berikut pada Gambar 2.

NO	NAMA CUSTOMER	TANGGAL	ITEM	JML	SATUAN	CATEGORY	HARGA (1 KG)	QTY (KG)	TOTAL HARGA	STATUS ITEM
1	BP JAFAR	2021-01-04	BACKRIB	10	PACK	DAGING	65000.0	10	650000	Tidak Laris
2	IBU AISYAH	2021-01-04	DAGING GILING	2	PACK	DAGING	105000.0	2	210000	Tidak Laris
3	TOKO ANIK	2021-01-04	HATI	2	PACK	DAGING	81000.0	2	162000	Tidak Laris
4	ERICA CATERING	2021-01-04	KNUCKLE	5	PACK	DAGING	105000.0	5	525000	Tidak Laris
5	SOP IGA BINTARO	2021-01-04	NECK BONE	8	PACK	DAGING	38000.0	8	304000	Tidak Laris
6	SUKA SUKI RESTO	2021-01-04	SUICE SHORTPLATE	30	PACK	DAGING	86000.0	30	2580000	Laris
7	BP HAFIDZ	2021-01-04	SUICE SHORTPLATE	3	PACK	DAGING	78000.0	3	234000	Tidak Laris
8	KEBULI UMMIR BERKAH	2021-01-04	SUICE SHORTPLATE	30	PACK	DAGING	86000.0	30	2580000	Laris

Gambar 2. Pengelompokan Data

Dari 1.213 data dari pengelompokan data dapat diketahui bahwasanya jumlah data Tidak Laris jauh lebih banyak dari pada data Laris dengan jumlah data yang mencapai 872 sedangkan data Laris hanya mencapai nilai total sebanyak 341 data pada penjualan frozen food.

IV. HASIL DAN ANALISA

Pada bagian ini menampilkan hasil semua eksperimen yang telah dilakukan. Eksperimen yang dilakukan pada penelitian ini menggunakan *split validation* pada pembagian data yaitu data training 70% dan data testing 30%.

A. Algoritma *Naïve Bayes*

Tabel 3 dan 4 menunjukan hasil *accuracy*, *precision* dan *recall* Algoritma *Naïve Bayes* spada *training* dan *testing* dipenelitian ini.

Tabel 3. Hasil Algoritma *Naïve Bayes Training*

	Accuracy	Precision	Recall
1	77%	76%	100%
2	77%	100%	14%

Tabel 4. Hasil Algoritma *Naïve Bayes Testing*

	Accuracy	Precision	Recall
1	73%	72%	100%
2	73%	100%	12%

Confusion Matrix

-----Nilai Perbandingan Performa Naive Bayes terhadap Data Training-----

```
[[618  0]
 [198 31]]
```

Berikut adalah penjelasan dari hasil output mengenai nilai perbandingan TP, TN, FP, dan FN di dalam confusion matrix dari algoritma *Naive Bayes*:

- True positive (TP) merupakan data positif yang berhasil diklasifikasikan dengan tepat. Jumlah true positive pada pengujian kali ini adalah 618.
- True negative (TN) merupakan data negatif yang berhasil diklasifikasikan dengan benar. Jumlah true negative pada algoritma naive bayes yang berhasil diklasifikasikan dengan benar mencapai 31.
- False positive (FP) merupakan data negatif namun diklasifikasikan sebagai data positif. Jumlah nilai false positive pada pengujian kali ini adalah 0.
- False negative (FN) merupakan data positif namun diklasifikasikan sebagai data negatif. Jumlah false negative di algoritma naive bayes yaitu mencapai 198.

B. Algoritma K-Nearest Neighbor

Tabel 5 menunjukkan hasil *accuracy*, *precision* dan *recall* B. Algoritma K-Nearest neighbor dipenelitian ini.

Tabel 5. Hasil Algoritma KNN

	Accuracy	Precision	Recall
1	99%	99%	100%
2	99%	99%	98%

Confusion Matrix

```
-----Nilai Perbandingan Performa KNN terhadap Data
Training-----
[[615  3]
 [ 4 225]]
```

Berikut adalah penjelasan dari hasil output diatas terkait nilai perbandingan TP, TN, FP, dan FN di dalam confusion matrix dari algoritma K-Nearest Neighbor:

- True positive (TP) merupakan data positif yang berhasil diklasifikasikan dengan tepat. Jumlah true positive pada pengujian ini adalah 615.
- True negative (TN) merupakan data negatif yang berhasil diklasifikasikan dengan benar. Jumlah true negative pada algoritma KNN yang berhasil diklasifikasikan dengan benar mencapai 225.
- False positive (FP) merupakan data negatif namun diklasifikasikan sebagai data positif. Jumlah nilai false positive pada pengujian ini adalah 3.
- False negative (FN) merupakan data positif namun diklasifikasikan sebagai data negatif. Jumlah false negative di algoritma KNN yaitu mencapai 4.

C. Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbors

Berdasarkan perbandingan tingkat akurasi data training melalui grafik batang diatas, saya dapat mengetahui dan mengambil kesimpulan diantaranya adalah sebagai berikut:

- Nilai akurasi di dalam data training terhadap algoritma Naive Bayes adalah sebesar 0.77 atau 77% yang menjadikan algoritma Naive Bayes di data training ini memiliki hasil klasifikasi/prediksi yang tepat tetapi jika dibandingkan dengan grafik batang di hasil akurasi algoritma K-Nearest Neighbors, hasil akurasi naive bayes terkhusus untuk data training masih tidak cukup tinggi dan kurang tepat.
- Sedangkan nilai akurasi di dalam data training terhadap algoritma K-Nearest Neighbors (KNN) sebesar 0.99 atau 99%. Menurut saya akurasi tersebut sudah sangat tinggi dan tepat dalam mengklasifikasikan item laris dan tidak laris selama bulan Januari - September 2021.

V. KESIMPULAN

Berdasarkan hasil penelitian ini semua percobaan penelitian yang sudah dilakukan mengklasifikasasn dari dua algoritma. Menggunakan data yang memanfaatkan data transaksi penjualan di perusahaan PT. *Frozen Food XYZ*. Demikian besar transaksi harian yang terus bertambah, Sedangkan algoritma naive bayes nilai rata-rata accuracynya 77% lebih rendah membuat performa naive bayes di bawah algoritma KNN. Naive bayes akan berjalan baik pada data yang bersifat independen. Sedangkan data yang diolah memiliki variabel yang dependent. Sedangkan, Algoritma K-NN nilai rata – rata accuracynya 99% lebih tinggi dibandingkan algoritma Naive Bayes hal ini karena dari karakteristik data, algoritma K-NN dapat bekerja dengan baik setelah data melalui preprocessing dan menghasilkan data nominal. Keterkaitan data satu sama lain membuat performa K-NN meningkat. Dengan dilakukannya penelitian ini dapat memberikan informasi bermanfaat mengenai penjualan frozen food dan algoritma yang lebih unggul KNN dari algoritma Naive Bayes.

DAFTAR PUSTAKA

- [1] Wiyanto, T. Ngudi, and A. Saefulloh, "Analisa Tingkat Kepuasan Pelanggan Terhadap Pelayanan Perusahaan Otobus XYZ Menggunakan Metode Naïve Bayes," *Pelita Teknologi*, vol. 15, no. 1, pp. 56–67, Apr. 2020, doi: 10.37366/pelitatekno.v15i1.243.
- [2] R. P. M. Rosidi and K. Setiawan, "Implementasi Algoritma Naïve Bayes Terhadap Data Penjualan untuk Mengetahui Pola Pembelian Konsumen pada Kantin," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 5, no. 1, pp. 120–126, Jan. 2024, doi: 10.35870/jimik.v5i1.407.
- [3] N. A. H. Muchlis, D. Novaldi, Andestan Wirayuda, and I. M. Dwintara, "Penerapan Algoritma Naive Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Voucher Kuota Di Edi Cell," *Innovative: Journal Of Social Science Research*, vol. 3, no. 2, pp. 14902–14914, 2023, Accessed: July 18, 2026. [Online]. Available: <http://j-innovative.org/index.php/Innovative/article/view/1293>
- [4] I. Romli, E. Pusnawati, and A. Siswandi, "Penentuan Tingkat Penjualan Mobil Di Indonesia Dengan Menggunakan Algoritma Naive Bayes," *e-Prosiding SNASTekS*, vol. 1, no. 1, pp. 367–380, 2019, Accessed: July 18, 2026. [Online]. Available: <https://journal.unusida.ac.id/index.php/snts/article/view/106>
- [5] D. Handoko, H. S. Tambunan, and J. T. Hardinata, "Analisis Penjualan Produk Paket Kuota Internet Dengan Metode K-Nearest Neighbor," *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, vol. 6, no. 1, p. 111, Feb. 2021, doi: 10.30645/jurasik.v6i1.275.
- [6] V. Putra Virza, G. Tri Pranot, and F. E. Putra, "Klasifikasi Kebutuhan Sparepart Dengan Algoritma K-Nearest Neighbor Untuk Meningkatkan Penjualan Sparepart," *Bulletin of Information Technology (BIT)*, vol. 4, no. 3, pp. 287–293, Sept. 2023, doi: 10.47065/bit.v4i3.729.
- [7] H. D. Wijaya and Saruni Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," vol. 7, no. 1, pp. 1–7, Apr. 2020, doi: 10.31294/ji.v7i1.6203.
- [8] Aisha Alfani W.P.R, Fahrur Rozi, and Farid Sukmana, "Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor," *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 6, no. 1, pp. 155–160, June 2021, doi: 10.29100/jipi.v6i1.1910.
- [9] A. S. Miha Djami, N. W. Utami, and A. A. I. I. Paramitha, "The Prediction Of Product Sales Level Using K-Nearest Neighbor and Naive Bayes Algorithms (Case Study : PT Kotamas Bali)," *Jurnal Pilar Nusa*

- Mandiri*, vol. 19, no. 2, pp. 77–84, Sept. 2023, doi: 10.33480/pilar.v19i2.4420.
- [10] H. P. Herlambang, F. Saputra, M. H. Prasetyo, D. Puspitasari, and D. Nurlaela, “Perbandingan Klasifikasi Tingkat Penjualan Buah di Supermarket dengan Pendekatan Algoritma Decision Tree, Naive Bayes dan K-Nearest Neighbor,” *Jurnal INSAN - Journal of Information System Management Innovation*, vol. 3, no. 1, pp. 21–28, July 2023, doi: 10.31294/jinsan.v3i1.2097.