

Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT.Nippon Indosari Corpindo Tbk Menggunakan Metode *K-Nearest Neighbor*

Ike Yolanda¹, Hasanul Fahmi²

^{1,2} Program Studi Teknik Informatika, STMIK Pelita Nusantara Jl. St. Iskandar Muda No. 1 Medan, Indonesia

¹ ikeyolanda4@gmail.com; ² h.fahmizuhri@gmail.com

INFORMASI ARTIKEL

ABSTRAK

Kata Kunci:
Data Mining
K-Nearest Neighbor
Berbasis Destop

Penerapan data mining sangat dibutuhkan oleh PT.Nippon Indosari Corpindo Tbk dikarenakan perusahaan ini memproduksi jumlah roti yang sangat besar. Perusahaan ini memiliki 20 jenis produk roti yang diminati oleh masyarakat Indonesia. Dilihat dari banyaknya permintaan konsumen ternyata terdapat beberapa roti yang terlaris dan tidak begitu laris sehingga berdasarkan 3 bulan terakhir maka PT Nippon membutuhkan sebuah prediksi penjualan produk roti terlaris, agar mempermudah pihak perusahaan dalam memproduksi roti mana yang paling banyak diproduksi. Karena saat ini sistem yang digunakan masih manual dengan itu data yang didapat kurang akurat dan efisien. Maka untuk mengetahui penjualan produk roti terlaris dibutuhkanlah sebuah data mining untuk memecahkan sebuah masalah dengan menggunakan metode *k-nearest neighbor*. Metode pengembangan sistem yang digunakan dalam penelitian ini adalah kerangka kerja penelitian yang terdiri dari analisis kebutuhan, desain sistem menggunakan UML, implementasi sistem menggunakan *software* berbasis destop. Penerapan data mining untuk prediksi penjualan produk roti terlaris adalah sebuah sistem yang membantu untuk memprediksi penjualan produk roti terlaris pada PT.Nippon Indosari Corpindo Tbk agar data yang didapat akurat dan efisien. Dengan menggunakan kriteria kuantitas produk dan kuantitas terjual.

Keywords:
Data Mining
K-Nearest Neighbor
Desktop Based

ABSTRACT

The application of data mining is very much needed by PT Nippon Indosari Corpindo Tbk because this company produces a very large amount of bread. This company has 20 types of bakery products that are in demand by Indonesians. Judging from the large number of consumer demands, it turns out that there are some of the best-selling and not-so-selling breads so that based on the last 3 months, PT Nippon needs a prediction of the sales of the best-selling bread products, in order to make it easier for the company to produce which bread is produced the most. Because currently the system used is still manual, thus the data obtained is less accurate and efficient. So to find out the sales of the best-selling bakery products, you need a data mining to solve a problem using the k-nearest neighbor method. The system development method used in this research is a research framework consisting of needs analysis, system design using UML, system implementation using desktop-based software. The application of data mining to predict the sales of the best-selling bakery products is a system that helps predict the sales of the best-selling bakery products at PT Nippon Indosari Corpindo Tbk so that the data obtained is accurate and efficient. By using the criteria for product quantity and quantity sold.

I. Pendahuluan

PT.Nippon Indosari Corpindo Tbk merupakan sebuah perusahaan yang bergerak dibidang produksi roti dimana berusaha agar kualitas rasa tetap terjaga dan berusaha agar tetap diminati oleh masyarakat Indonesia. Dengan beragam macam varian dan rasa yang terdiri dari Roti Tawar, Roti Manis Isi, Roti Krim, Roti Sobek dan Roti Sandwich Isi. Diketahui bahwa peningkatan pada penjualan produk roti ini meningkat sangat pesat. Setelah makin pesatnya penjualan produk roti ini maka perusahaan PT.Nippon Indosari Corpindo Tbk semakin banyak penjualan hingga ribuan produk roti yang terjual. Dengan ribuan roti yang terjual PT.Nippon Indosari Corpindo Tbk mengalami kesulitan untuk menentukan produk yang mana-mana saja yang terlaris dari produk roti Sari Roti berdasarkan data tiga bulan terakhir, sehingga didapatkan sebuah permasalahan yang ada bahwa perusahaan kewalahan dalam menentukan produk roti mana yang paling diminati masyarakat Indonesia, sehingga dibutuhkan sebuah Data Mining agar pihak Sari Roti mengetahui roti mana saja yang paling laris dan diminati oleh masyarakat Indonesia.

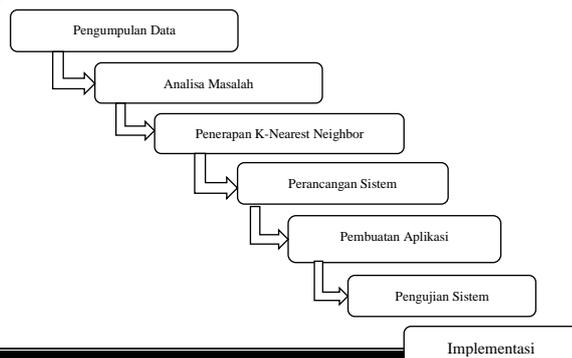
Dalam hal ini Data Mining sangat berpengaruh dalam proses menggali sebuah data dari suatu kumpulan data yang berupa pengetahuan yang selama ini tidak diketahui secara manual [8]. *Data mining* sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar [9]. Data mining proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowlwdge*) secara otomatis [12].Data mining suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database [2]. Maka data mining merupakan sebuah informasi yang sangat dibutuhkan dalam menggali data-data roti mana saja yang terlaris dan diminati oleh masyarakat Indonesia. Akan tetapi untuk menentukan produk terlaris ini dibutuhkan sebuah algoritma *K-Nearest Neighbor* agar mempermudah PT. Nippon Indosari Corpindo Tbk untuk mengetahui produk-produk mana saja dan varian-varian mana saja yang banyak diminati masyarakat Indonesia. Penerapan fuzzy mamdani dapat memprediksi gizi buruk pada balita secara lebih efektif dikarenakan kreteria yang sedikit bias mampu diselesaikan dengan lebih tepat dengan menggunakan fuzzy mamdani [5]. Akan tetapi apabila pendekatan fuzzy di gabungkan dengan *K-Nearest Neighbour* maka akan mendapatkan hasil prediksi yang lebih baik lagi.

Dengan menggunakan Algoritma *K-Nearest Neighbor* (KNN) dalam melakukan klasifikasi terhadap objek maka dibutuhkan jarak terdekat dengan objek tersebut [4]. Algoritma *K-Nearest Neighbor* berusaha mengklasifikasikan data baru yang belum diketahui class-nya dengan memilih sejumlah data yang letaknya terdekat dari data baru tersebut [6]. Menurut [11] metode *K-Nearest Neighbor* (K-NN) merupakan metode yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data lain, metode K-NN merupakan metode yang cukup sederhana namun memiliki tingkat akurasi yang tinggi. Tujuan dari penerapan algoritma *K-Nearest Neighbor* dapat mengklasifikasikan objek baru berdasarkan *attribut* dan *training sample*. Sehingga dengan penerapan algoritma *K-Nearest Neighbor* dapat mempermudah PT. Nippon Indosari Corpindo Tbk pada penjualan roti dengan mengambil objek baru berdasarkan data yang letaknya terdekat dari data baru tersebut.

Pada penelitian yang dilakukan oleh [10], menyatakan bahwa penerapan *K-Nearest Neighbor* dapat menentukan prediksi kelulusan mahasiswa STMIK-Sinar Nusantara Suarakarta. Penelitian tersebut menunjukkan bahwa penerapan algoritma *K-Nearest Neighbor* (KNN) berhasil diterapkan.

Maka berdasarkan latar belakang diatas, maka diusulkan sebuah penelitian dengan judul “Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT. Nippon Indosari Corpindo Tbk Menggunakan Metode *K-Nearest Neighbor*. Dengan Metode *K-Nearest Neighbor* dapat mempermudah dan membantu perusahaan dalam menentukan jumlah produksi roti mana saja yang harus di produksi dengan jumlah yang banyak.

II. Metode





Gambar 1. Kerangka Kerja Penelitian

Metode *K-Nearest Neighbour* (KNN) merupakan salah satu metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Tujuannya adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan data training. Klasifikasi dilakukan tanpa menggunakan model tetapi hanya berdasarkan memori. Algoritma *K-Nearest Neighbour* menggunakan klasifikasi ketetanggaan sebagai prediksi terhadap data baru [7].

Adapun langkah-langkah *K-Nearest Neighbour* :

1. Menentukan parameter K = jumlah tetangga terdekat.
2. Hitung jarak antara data baru dengan semua data training. Berikut rumus perhitungan jarak dengan *Euclidean* seperti di bawah ini [3].

$$d_i = \sqrt{\sum_{i=1}^k (x_i - x_i)^2} \tag{1}$$

Dimana :

- X1 : sampel data / data training
- X2 : data uji / testing
- d : Jarak
- k : dimensi atribut / data

3. Urutkan jarak tersebut dan tetapkan tetangga terdekat berdasarkan jarak minimum ke-K.
4. Periksa kelas dari tetangga terdekat.
5. Gunakan mayoritas sederhana dari kelas tetangga terdekat sebagai nilai prediksi data baru.

III. Hasil dan Pembahasan

Untuk mengetahui jenis roti mana-mana saja yang paling terlaris dan diminati oleh masyarakat Indonesia yaitu dengan cara mengumpulkan sebuah data penjualan dari 3 bulan terakhir.

Berdasarkan pengamatan peneliti dari Staff Pegawai data yang diambil dikumpulkan menjadi satu kemudian diperoleh data training dari sumber data yang sudah ada dan peneliti membuat data *testing* yang akan di uji nantinya.

a. Data Training

Data *Training* pada penelitian ini di dapat dari sumber data yang sudah ada. Data yang diambil adalah data penjualan produk roti dari 3 bulan terakhir yang dapat dijadikan data *training* untuk analisa dan dapat dilihat pada Tabel 1.

Tabel 1. Data *Training*

No	Nama Produk	Kuantitas Produk	Kuatitas Terjual	Bulan
1	Roti Tawar Spesial	2	3	Januari
2	Roti Tawar Double Soft	2	3	Januari
3	Roti Manis Isi Krim Coklat	3	1	Januari
4	Roti Manis Isi Krim Keju	1	4	Januari
5	Roti Manis Isi Krim Mocca	2	3	Januari
6	Roti Sandwich Isi Coklat	1	4	Januari
7	Roti Tawar Kupas	4	1	Februari
8	Roti Tawar Gandum	4	1	Februari
9	Roti Tawar Choco Chip	4	1	Februari
10	Roti Manis Isi Pisang Coklat	4	1	Februari
11	Roti Manis Isi Krim Coklat	4	1	Februari
12	Roti Manis Isi Krim Keju	1	4	Februari
13	Roti Sobek Coklat	1	4	Februari
14	Roti Sandwich Isi Coklat	1	4	Februari
15	Roti Sandwich Isi Keju	3	2	Februari
16	Roti Tawar Kupas	4	1	Maret
17	Roti Manis Isi Capucino	4	1	Maret
18	Roti Manis Isi Tiramisu	4	1	Maret
19	Roti Sandwich Isi Coklat	1	4	Maret

Dari data *training* diatas, peneliti menggunakan beberapa atribut yang dibutuhkan sebagai dasar untuk melakukan proses prediksi menggunakan metode *k-nearest neighbor* seperti Kuantitas Produk dan Kuantitas Terjual.

a. Data Selection

Tahap pemilihan atau penyeleksian terhadap data dilakukan untuk mendapatkan data yang sesuai dengan kebutuhan dalam prediksi penjualan produk roti terlaris. Dalam prediksi penjualan produk roti terlaris dibutuhkan data kuantitas produk dan data kuantitas terjual. Data yang dapat digunakan adalah data kuantitas produk dan data kuantitas terjual pada bulan Januari, Februari, dan Maret yang didapat dari pengumpulan data sebelumnya. Pada tahap pemilihan data maka diperoleh sebanyak 60 data. 20 data secara acak akan dijadikan sebagai data *training*.

b. Pre-processing/Cleaning

Tahap ini merupakan proses pembersihan terhadap data yang dilakukan untuk memastikan data yang diperoleh sebelumnya dapat digunakan serta bebas dari duplikasi dan kesalahan.

c. Transformation

Pada tahap transformasi ini hasil dari sample data diatas untuk mengelompokan data penjualan produk roti terlaris dibutuhkan jumlah penjualan tiap bulannya untuk mempermudah prediksi penjualannya, sehingga data dapat diolah dengan menggunakan Algoritma *K-Nearest Neighbor*. Maka sampel data diatas di ubah seperti bobot yang telah dibuat masing-masingnya. Dapat dilihat pada Tabel 2.

Tabel 2. Variabel Kuantitas Produk

No	Variabel	Bobot
1	5000 – 10500	4
2	10600 – 25500	3
3	25600 – 45500	2
4	45600 – 95000	1

Tabel 3. Variabel Bobot

No	Variabel	Bobot
1	Rendah	4
2	Cukup	3
3	Baik	2
4	Sangat Baik	1

Tabel 4. Variabel Kuantitas Terjual

No	Variabel	Bobot
1	3000 – 8000	1
2	9000 – 18000	2
3	19000 – 36000	3
4	37000 - 72000	4

Tabel 5. Variabel Bobot

No	Variabel	Bobot
1	Rendah	1
2	Cukup	2
3	Baik	3
4	Sangat Baik	4

Tabel 6. Proses Perubahan Data Target

No	Kriteria Target	Bobot
1	Naik	>1,5
2	Menurun	<1,5

b. Penerapan *K-Nearest Neighbor*

Untuk melakukan proses algoritma *K-Nearest Neighbor* ditentukan terlebih dahulu nilai *k* atau jumlah tetangga untuk meneliti hasil prediksi yang ingin kita lakukan.

1. Menentukan parameter *K* = jumlah tetangga terdekat. Pada penelitian ini nilai *k* yang digunakan adalah 5.
2. Hitung jarak antara data *testing* (uji) dengan semua data *training* pada tahap *transformasi* menggunakan perhitungan jarak *Euclidean Distance*, yang dipakai untuk menghitung jarak *Euclidean Distance* (1).

Tabel 7. Hasil dari Jarak *Euclidean Distance*

No	Nama Produk	Kuantitas Produk	Kuantitas Terjual	Jarak <i>Euclidean Distance</i> (2,1)
1	Roti Tawar Spesial	2	3	$\sqrt{(3-2)^2+(2-1)^2} = 2$
2	Roti Tawar Double Soft	2	3	$\sqrt{(2-2)^2+(3-1)^2} = 2$
3	Roti Manis Isi Krim Coklat	3	1	$\sqrt{(3-2)^2+(1-1)^2} = 1$
4	Roti Manis Isi Krim Keju	1	4	$\sqrt{(1-2)^2+(4-1)^2} = 3,16$
5	Roti Manis Isi Krim Mocca	2	3	$\sqrt{(2-2)^2+(3-1)^2} = 2$
6	Roti Sandwich Isi Coklat	1	4	$\sqrt{(1-2)^2+(4-1)^2} = 3,16$
7	Roti Tawar Kupas	4	1	$\sqrt{(4-2)^2+(1-1)^2} = 2$
8	Roti Tawar Gandum	4	1	$\sqrt{(4-2)^2+(1-1)^2} = 2$
9	Roti Tawar Choco Chip	4	1	$\sqrt{(4-2)^2+(1-1)^2} = 2$
10	Roti Manis Isi Pisang Coklat	4	1	$\sqrt{(4-2)^2+(1-1)^2} = 2$
11	Roti Manis Isi Krim Coklat	4	1	$\sqrt{(4-2)^2+(1-1)^2} = 2$
12	Roti Manis Isi Krim Keju	1	4	$\sqrt{(1-2)^2+(4-1)^2} = 3,16$
13	Roti Sobek Coklat	1	4	$\sqrt{(1-2)^2+(4-1)^2} = 3,16$
14	Roti Sandwich Isi Coklat	1	4	$\sqrt{(1-2)^2+(4-1)^2} = 3,16$
15	Roti Sandwich Isi Keju	3	2	$\sqrt{(3-2)^2+(2-1)^2} = 1,41$
16	Roti Tawar Kupas	4	1	$\sqrt{(4-2)^2+(1-1)^2} = 2$
17	Roti Manis Isi Capucino	4	1	$\sqrt{(4-2)^2+(1-1)^2} = 2$
18	Roti Manis Isi Tiramisu	4	1	$\sqrt{(4-2)^2+(1-1)^2} = 2$
19	Roti Sandwich Isi Coklat	1	4	$\sqrt{(1-2)^2+(4-1)^2} = 3,16$
20	Roti Sandwich Isi Keju	2	3	$\sqrt{(2-2)^2+(3-1)^2} = 2$

3. Urutkan jarak tersebut dan tetapkan tetangga terdekat berdasarkan jarak minimum ke-*K*.

Tabel 8. Hasil Berdasarkan Jarak Minimum *K*

Square Jarak ke contoh permintaan (2,1)	Jarak Minimum
$\sqrt{(2-2)^2+(3-1)^2} = 2$	3
$\sqrt{(2-2)^2+(3-1)^2} = 2$	4
$\sqrt{(3-2)^2+(1-1)^2} = 1$	1
$\sqrt{(1-2)^2+(4-1)^2} = 3,16$	15
$\sqrt{(2-2)^2+(3-1)^2} = 2$	5

$\sqrt{(1-2)^2+(4-1)^2} = 3,16$	16
$\sqrt{(4-2)^2+(1-1)^2} = 2$	6
$\sqrt{(4-2)^2+(1-1)^2} = 2$	7
$\sqrt{(4-2)^2+(1-1)^2} = 2$	8
$\sqrt{(4-2)^2+(1-1)^2} = 2$	9
Square Jarak ke contoh permintaan (2,1)	Jarak Minimum
$\sqrt{(4-2)^2+(1-1)^2} = 2$	10
$\sqrt{(1-2)^2+(4-1)^2} = 3,16$	17
$\sqrt{(1-2)^2+(4-1)^2} = 3,16$	18
$\sqrt{(1-2)^2+(4-1)^2} = 3,16$	19
$\sqrt{(3-2)^2+(2-1)^2} = 1,41$	2
$\sqrt{(4-2)^2+(1-1)^2} = 2$	11
$\sqrt{(4-2)^2+(1-1)^2} = 2$	12
$\sqrt{(4-2)^2+(1-1)^2} = 2$	13
$\sqrt{(1-2)^2+(4-1)^2} = 3,16$	20
$\sqrt{(2-2)^2+(3-1)^2} = 2$	14

4. Periksa kelas dari tetangga terdekat

Tabel 9. Penentuan parameter K

Jarak Euclidean Distance	Tetangga yang terdekat $k = 5$
1	1
1,41	2
2	3
2	4
2	5

5. Gunakan mayoritas sederhana dari kelas tetangga terdekat sebagai nilai prediksi data baru.

Tabel 10. Hasil dari Mayoritas Kelas Tetangga

Jarak Euclidean Distance	Tetangga yang terdekat $k = 10$	Kategori
1	1	Menurun
1,41	2	Menurun
2	3	Naik
2	4	Naik
2	5	Naik

Berdasarkan hasil perhitungan jarak diatas, maka akan didapatkan suatu hasil keputusan yaitu Naik = 3 dan Menurun = 2. Dapat dilihat pada Tabel 4.11 mayoritas klasifikasi yang memiliki jumlah paling banyak adalah ">1,5" dimana variable ">1,5" merupakan kategori "Naik", dan variable "<1,5" merupakan kategori "Menurun" sehingga penjualan produk roti memiliki nilai sesuai dengan data uji yang telah dihitung diprediksi masuk kedalam kategori "Naik".

IV. Kesimpulan

Dari hasil penelitian yang dilakukan, maka dapat disimpulkan sebagai berikut :

1. Penerapan Metode *K-Nearest Neighbor* ke dalam sebuah Data Mining digunakan sebagai aplikasi yang dapat memprediksi penjualan produk roti terlaris dengan menyesuaikan kriteria dan menggunakan bobot agar dapat digunakan dengan metode tersebut.
2. Aplikasi yang dirancang dapat memprediksi penjualan produk roti terlaris dengan kriteria yang sudah ditentukan.
3. Sistem ini digunakan untuk memprediksi penjualan produk roti terlaris pada PT.Nippon Indosari Corpindo Tbk dengan menggunakan algoritma *K-Nearest Neighbor* yang digunakan untuk memprediksi jenis-jenis roti yang mana saja yang paling banyak diminati oleh masyarakat Indonesia.

Daftar pustaka

- [1] Mariana. N. (2015). PENERAPAN ALGORITMA k-NN (nearest Neighbor) UNTUK DETEKSI PENYAKIT (KANKER SERVIKS). *Jurnal Dinamika Informatika*, 7(1), 26–34.
- [2] Abidin, D. Z., Nurmaini, S., & Malik, R. F. (2017). Penerapan Metode K-Nearest Neighbor dalam Memprediksi Masa Studi Mahasiswa (Studi Kasus : Mahasiswa STIKOM Dinamika Bangsa). *Prosiding Annual Research Seminar*, 3(1), 133–138.
- [3] Amalia, Y. R. (2018). Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus : PT. Bintang Multi Sarana Palembang). *Thesis*.
- [4] Bri. (2018). *DATA MINING UNTUK KLASIFIKASI STATUS GIZI DESA DI KABUPATEN MALAKA MENGGUNAKAN METODE K-NEAREST NEIGHBOR.pdf*. 6(1), 1–7.
- [5] Fahmi, H. (2017). Penerapan Sistem Pakar Untuk Diagnosa Gizi Buruk Pada Balita Menggunakan Metode Fuzzy Mamdani. *Jurnal Mantik Penusa*.
- [6] Harsemadi, G., Sudarma, M., & Pramaita, N. (2017). Implementasi Algoritma K-Nearest Neighbor pada Perangkat Lunak Pengelompokan Musik untuk Menentukan Suasana Hati. *Majalah Ilmiah Teknologi Elektro*, 16(1), 14–20. <https://doi.org/10.24843/mite.1601.03>
- [7] Ikhsanuddin, R. M. (2014). *Identifikasi Citra Pada Plat Nomor Kendaraan Mobil Pribadi Menggunakan Metode K-Nearest Neighbor*. 1–7.
- [8] Mutiara, I. dan A. (2015). Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4. *Klik - Kumpulan Jurnal Ilmu Komputer*, 2(2), 159–173. <https://doi.org/10.20527/KLIK.V2I2.26>
- [9] Pahlevi, O., Sugandi, A., & Sintawati, I. D. (2018). Penerapan Algoritma Apriori Dalam Pengendalian Kualitas Produk. *Sinkron*, 3(1), 272–278.
- [10] Panoto, A. (2017). Penerapan Algoritma K-Nearest Neighbors untuk prediksi kelulusan Mahasiswa pada STMIK Sinar Nusantara Surakarta. *TIKomSiN*, 27–31.
- [11] Prasetyo, E., Informatika, J. T., Industri, F. T., Pembangunan, U., Veteran, N., Timur, J., & Neighbor, K. (2012). *Fuzzy K-Nearest Neighbor in Every Class Untuk Klasifikasi Data. Santika*, 57–60.
- [12] Sikumbang, E. D. (2018). Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori. *Jurnal Teknik Komputer AMIK BSI (JTK)*, Vol 4, No.(September), 1–4.