IMPLEMENTASI DATA MINING DALAM OPTIMASI STOK BAHAN BAKU KUE DI PT. XYZ MENGGUNAKAN METODE K NEAREST NEIGHBOR

Vicco Dwi Arya Putra, Raditia Vindua
Teknik Informatika, Universitas Pamulang
Jl. Raya Puspitek, Buaran, Kec. Pamulang, Kota Tangerang Selatan, Banten 15310
dosen02380@unpam.ac.id

ABSTRAK

Setiap perusahaan pasti menginginkan jumlah persediaan yang cukup agar proses produksi tidak terganggu. Jumlah yang besar karena banyak mengandung resiko seperti bahan baku rusak atau hilang karena kurangnya pengawasan, biaya pemeliharaan yang tinggi, uang yang tertanam dalam persediaan terlalu besar selain itu resiko uang pada bahan baku karena terlalu lama tidak dipakai. Dalam masalah ini, data mining dapat digunakan dalam proses penggalian data secara manual dari kumpulan data berupa pengetahuan yang tidak diketahui. Salah satu metode dalam data mining adalah metode K-Nearest Neighbor (K-NN) yang merupakan metode yang digunakan untuk klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data lain. Tujuan dari penerapan metode K-NN dapat mengklasifikasikan objek baru berdasarkan attribut dan *training sample*, dimana dengan menggunakan metode K-NN, diharapkan dapat membantu perusahaan dalam mengefesiensi pembelian bahan baku kue dan mengoptimalkan tingkat persediaan sehingga mengurangi biaya pembelian bahan baku serta menghindari terjadinya kekurangan atau kelebihan stok yang dapat mempengaruhi proses produksi bahan kue perusahaan. Dengan menggunakan metode K-NN, PT. XYZ dapat mengoptimalkan bahan baku kue berdasarkan data histori. Dalam penggunaan K-NN sebagai alat prediksi menghasilkan nilai akurasi sebesar 95,00%. Dengan hasil akurasi yang cukup besar artinya metode K-NN dapat membantu perusahaan dalam perencanaan produksi dan pengadaan stok bahan baku.

Kata kunci : *Data Mining, K-Nearest Neighbor,* bahan baku

1. PENDAHULUAN

Pada perusahaan, bahan baku merupakan salah satu masalah yang cukup dominan di bidang produksi selain masalah keuangan dan kepegawaian. Setiap perusahaan pasti menginginkan jumlah persediaan yang cukup agar proses produksi tidak terganggu. Jumlah yang besar karena banyak mengandung resiko seperti bahan baku rusak atau hilang karena kurangnya pengawasan, biaya pemeliharaan yang tinggi, uang yang tertanam dalam persediaan terlalu besar selain itu resiko uang pada bahan baku karena terlalu lama tidak dipakai.

Persediaan harus berada di batas normal tidak terlalu besar tidak terlalu kecil, di mana persediaan yang terlalu kecil pun mengandung resiko seperti kehabisan persediaan bahan baku pada saat proses produksi, sehingga dapat merugikan perusahaan karena tidak memenuhi konsumen. [1]

Perkembangan dunia usaha ini sangat pesat ditandainya munculnya berbagai jenis perusahaan, baik perusahaan berskala kecil (mikro), menengah maupun yang berskala besar (makro). Perusahaan pasti memiliki tujuan untuk menghasilkan keuntungan yang optimal supaya dapat mempertahankan kelangsungan hidupnya, memajukan, serta mengembangkan usaha perusahaannya ke tingkat yang lebih tinggi [2].

Dalam bidang usaha Perdagangan barang selain manajemen yang baik hal lain yang harus diperhatikan lebih mendalam adalah tentang persediaan barang. Persediaan atau stok merupakan barang yang dibeli kemudian disimpan untuk dijual, sehingga perusahaan harus memperhatikan persediaannya dengan perhatian yang besar. Persediaan mempunyai arti yang sangat strategis bagi perusahaan, baik perusahaan dengan maupun perusahaan perusahaan industri. Masalah yang sering terjadi dalam perusahaan salah satunya adalah kekurangan persediaan saat adanya permintaan.

Perusahaan yang kurang efektif dalam menentukan kuantitas persediaan barang dagang mengakibatkan perusahaan kerap kali melakukan pembelian barang secara berlebihan. Pembelian barang yang berlebihan mengakibatkan penumpukan barang sehingga arus keuangan tidak berjalan dengan baik dan lancar. Atas dasar hal di atas maka peneliti ingin meneliti lebih lanjut terkait persediaan serta membantu perusahaan dalam menentukan jumlah persediaan agar tidak memunculkan penumpukan barang dan komplain dari pelanggan. [3]

Data mining merupakan proses untuk menggali informasi yang berguna dari banyaknya data yang disimpan di dalam basis data. Salah satu metode yang ada dalam data mining adalah bagaimana menggali data yang disimpan untuk membangun sebuah model, selanjutnya dengan model tersebut digunakan agar dapat membuat pola data yang lain yang tidak berada dalam basis data yang tersimpan. [4]

Menurut Mujiasi Data mining atau sering disebut sebagai *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. [5]

Data mining adalah pembelajaran mesin, pengenalan pola, *database*, statistik, dan teknik

visualisasi yang digunakan untuk memecahkan masalah penggalian informasi dari repositori database besar. Data mining sangat penting dalam proses penggalian data secara manual dari kumpulan data berupa pengetahuan yang tidak diketahui. Berdasarkan pengertian Data Mining tersebut dapat disimpulkan bahwa Data Mining adalah suatu proses pencarian data secara otomatis dapat mendapatkan sebuah model dari *database* yang besar.

Metode *K-Nearest Neighbor* (K-NN) adalah algoritma pembelajaran berbasis contoh. Metode K-NN termasuk klasifikasi sangat baik. Dalam klasifikasi data terdapat dua proses yaitu *learning* dan *classification*. Teori ini digunakan untuk mengklasifikasikan berdasarkan jumlah dari tetangga terdekat. Nilai K tergantung data yang dipilih dengan optimasi parameter.

Menurut [6] metode K-NN merupakan metode yang melakukan klasifikasi berdasarkan jarak (kedekatan lokasi) suatu data dengan data lain, metode K-NN merupakan metode yang cukup sederhana namun memiliki tingkat akurasi yang tinggi. K-NN termasuk algoritma *supervised learning*, yang mana hasi dari *query instance* baru, diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN.

Algoritma K-NN bekerja dengan cara mengklasifikasikan data baru dengan memilih sejumlah data yang letaknya terdekat dari data baru yang belum diketahui *class*-nya. Metode K-NN merupakan metode yang cukup sederhana namun memiliki tingkat akurasi yang tinggi. Tujuan dari implementasi algoritma K-NN yaitu dapat mengklasifikasikan data baru berdasarkan attribut dan *training samples*. Sehingga dengan menggunakan algoritma K-NN ini dapat mempermudah PT. XYZ pada pengoptimalan bahan baku kue dengan mengambil objek baru berdasarkan data yang letaknya terdekat dari data baru tersebut.

Optimasi stok bahan baku kue ini menggunakan metode K-NN. K-NN adalah metode klasifikasi dengan mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (*neighbor*) terdekatnya dalam data pelatihan. Algoritma K-NN adalah salah satu metode dalam data mining untuk mengklasifikasi objek yang baru berdasarkan mayoritas dari kategori tetangga yang terdekat. [7]

PT. XYZ ini adalah merupakan produsen produk bakery dan pastry ingredients yang berfokus pada segmen pasar antara lain industry makanan, *bakery* dan *pastry shop*, *bakery* dan *pastry* ritel, dan *food catering*. PT. XYZ juga menyediakan *Baking Centre* sebagai tempat pelatihan dan selalu membuka diri untuk semua *customer* yang ingin belajar maupun bertukar pengalaman. Berdasarkan latar belakang yang telah dibahas diatas penulis tertarik untuk melakukan penelitian yang berjudul "Implementasi Data Mining Dalam Optimasi Stok Bahan Baku Kue Di PT. XYZ Menggunakan Metode *K-Nearest Neighbor*".

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian terkait dengan metode k nearest neighbord dengan penelitian sebelumnya yaitu:

a. Penelitian oleh Dewi dan kawan-kawan yang berjudul "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode k-Nearst Neighbor (k-NN)". Penelitian ini membahas tentang perhitungan dengan teknik data mining. Algoritma k nearest neighbor menghasilkan hasil prediksi dengan nilai akurasi yang tinggi, sehingga teknik data mining dan metode algoritma k nearest neighbor ini dapat diimplementasikan untuk memprediksi penjualan produk terlaris pada UD Andar. [8]
b. Penelitian oleh Mujilahwati & Windasari yang berjudul "Implementasi Metode k-Nearest Neighbor (k-NN) untuk Memprediksi Penjualan Buah di Indonesia berbasis Website". Penelitian ini membahas tentang prediksi penjualan buah di Indonesia, metode yang digunakan untuk prediksi ini adalah K- Nearest Neighbor (K-NN). [9]

2.2. Data Mining

Data mining merupakan suatu proses pengumpulan informasi penting dari suatu data yang sangat besar. Proses data mining biasanya menggunakan metode matematika, statistika, hingga memanfaatkan teknologi artificial intelligence. Data mining adalah proses analisis dari peninjauan kumpulan data untuk menemukan pola atau hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami sehingga bermanfaat bagi pemilik data. Data mining adalah bidang ilmu dari beberapa bidang keilmuan yang menyatukan teknik dari pengenalan pola, pembelajaran mesin, statistik, database dan visualisasi untuk menangani permasalahan dalam pengambilan informasi dari database yang besar. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terikat dari berbagai database besar. [10]

Berdasarkan pengertian data mining yang telah dijabarkan sebelumnya, maka data mining adalah ilmu atau pengetahuan yang tersembunyi dalam bentuk informasi di dalam database yang di proses untuk menemukan pola dengan cara teknik statistik matematika, kecerdasan buatan, dan machine learning untuk menggali dan mengidentifikasi informasi pengetahuan dari database tersebut.

2.3. K-Nearest Neighbor

Algoritma K-Nearest Neighbor (K-NN) adalah algoritma yang digunakan untuk mengklasifikasi suatu objek, berdasarkan k buah data latih yang jaraknya paling dekat dengan objek tersebut. Dengan syarat

nilai k harus ganjil dan lebih dari satu serta nilai k tidak boleh lebih besar dari jumlah data latih.

K-Nearest Neighbor (K-NN) adalah suatu teknik yang menggunakan algoritma supervised dimana hasil dari jarak terdekat yang baru diklasifikasikan berdasarkan mayoritas dari label *class* pada K-NN. Metode K-NN merupakan sebuah metode untuk mengklasifikasikan objek baru berdasarkan K tetangga terdekatnya. Yang akan menjadi kelas hasil klasifikasi adalah kelas yang paling banyak muncul . [8]

Tujuan dari algoritma K-NN adalah melakukan klasifikasi terhadap objek baru berdasarkan atribut dan data training. Sehingga dengan menerapkan algoritma K-NN dapat mempermudah penjualan produk dengan mengambil objek baru berdasarkan data yang jaraknya terdekat dari data baru tersebut. Algoritma K-NN melakukan klasifikasi ketetanggaan sebagai nilai prediksi dari training sample yang baru. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak Eucledian.

Ada beberapa tahapan dalam perhitungan yang ada di dalam algoritma K-NN yaitu:

a. Menentukan Parameter K
b. Menentukan jarak antara data training dan data testing
c. Mengurutkan jarak yang terbentuk
d. Menentukan jarak terdekat
e. Menentukan kelompok data hasil uji
f. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi

Langkah berikutnya adalah penentuan atau pengambilan keputusan terhadap knowledge yang dihasilkan Model persamaan algoritma K-NN [11]. Penelitian ini penulis hanya menggunakan jarak Euclidean, maka rumus perhitungan jarak dengan Euclidean berikut ini:

$$d_e = \sqrt{\sum_{k=1}^{m} |fd_{i,k-} k_j|^2} \qquad (1)$$

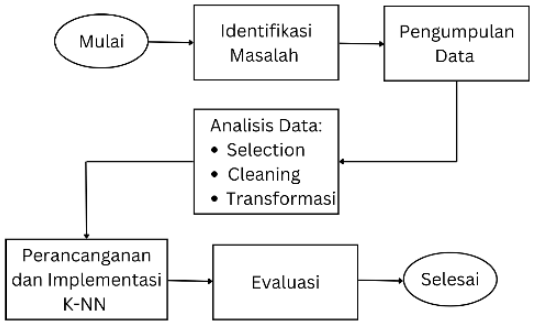
Keterangan
d_e = Jarak
fd_i = Data training (Data Latihan)
k = Data testing (Data Uji)
m = Jumlah data dalam penelitian

Langkah-langkah dalam perhitungan algoritma K-NN dengan perhitungan jarak Euclidean sebagai berikut:

a. Hitung jarak Euclidean antara titik data yang ingin diprediksi dan semua titik data dalam set pelatihan.
b. Pilih k titik terdekat berdasarkan jarak Euclidean
c. Hitung mayoritas label (kelas) dari k titik data terdekat
d. Label dari titik data yang ingin diprediksi adalah label mayoritas dari k tetangga terdekat.

3. METODE PENELITIAN

Penelitian ini merupakan penelitian untuk optimasi stok bahan baku kue dalam data mining menggunakan metode K-NN. Dalam metodologi penelitian, ada beberapa langkah yang digunakan dalam melaksanakan penelitian. Berikut Gambaran umum dalam penelitian ini:



Gambar 1. Tahapan penelitian

Adapaun langkah-langkah dalam penelitian sebagai berikut:

3.1. Identifikasi Masalah

Mengidentifikasi masalah yang signifikan untuk dipecahkan ke dalam perumusan masalah.

3.2. Pengumpulan Data

Mengumpulkan data dan memilih data untuk diuji yang berkaitan dengan landasan penelitian. Pengumpuan data yang dilakukan dengan menggunakan beberapa cara diantara lainnya adalah studi literatur, wawancara dan observasi

a. Studi Literatur
   Pengumpulan data dengan cara mencari peneliti yang sama dengan apa yang kita teliti bisa berupa jurnal atau buku.
b. Wawancara
   Pengumpulan data dengan cara mengajukan pertanyaan lisan kepada subjek tempat penelitian atau responden narasumber. Secara umum data yang didapatkan data yang bersifat komplek atau masalah tertentu.
c. Observasi
   Teknik pengumpulan data yang dilakukan secara langsung atau aktivitas yang dilakukan didalam kegiatan yang sedang diamati dengan tujuan merasakan proses tersebut.

3.3. Analisis Data

Analisis data dilakukan dengan menerapkan *Knowledge Discovery in Database* (KDD) dalam data mining. KDD adalah keseluruhan proses untuk mengkonversi data mentah menjadi suatu pengetahuan yang bermanfaat. Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

a. Data Selection
   Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses

data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

b. Data Cleaning
Sebelum proses data mining dapat dilaksanakan, maka perlu dilakukan proses cleaning pada data yang akan menjadi focus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

c. Transformasi
Tahap ini dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data. Data diubah atau digabung ke dalam format khusus sebelum bisa diaplikasikan, atau juga bisa disebut pengelompokan data.

3.4. Perancangan dan Implementasi K-Nearest Neighbor (K-NN)

Perancangan dan implementasi *K-Nearest Neighbor* (K-NN) untuk menghasilkan output yang dapat dipahami dari proses pola informasi data mining dimana dilakukan untuk mengekstrak pola-pola potensial menghasilkan data yang berguna.

3.5. Evaluasi Sistem

Pada tahap ini dilakukan setelah data diolah dengan algoritma yang dipilih menggunakan tools rapidminer dan tujuannya adalah untuk melakukan evaluasi terhadap hasil proses data mining.

4. HASIL DAN PEMBAHASAN

Pada penelitian ini peneliti menggunakan metode K-Nearest Neighbor (K-NN). Tujuan dari algoritma K-NN adalah untuk mengklasifikasi objek baru berdasarkan atribut dan *training samples*. Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Sebelum di klasifikasi, data di analisis menggunakan *Knowledge Discovery in Database* (KDD).

4.1. *Data Selection*

Data selection merupakan pemilihan dari sekumpulan data operasional yang perlu dilakukan sebelum tahapan penggali informasi dalam KDD dimulai. Berikut adalah data seleksi yang akan diteliti:

Tabel 1. *Data Selection*

No	Nama Produk	Jumlah (Kuintal)	Bulan
1	ANT VANILA	30	JANUARI
2	BAKING POWDER	30	JANUARI
3	COKLAT ELMER 10-12%	28	JANUARI
4	GARAM	20	JANUARI
5	GULA PASIR	40	JANUARI
…	…	…	…
120	LAKTOSA	20	JUNI

Tabel 1 merupakan hasil pengumpulan data berupa stok bahan baku dalam jumlah kuintal yang dikumpulkan dari bulan januari sampai bulan juni

4.2. *Data Cleaning*

Tahapan selanjutnya pada metode KDD yaitu preprocessing yang merupakan proses pembersihan data, proses *preprocessing* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Berikut adalah data yang sudah di kelompokkan dan diurutkan sesuai dengan bulan.

Tabel 2. *Data Cleaning*

No	Nama Produk	Jan	Feb	Mar	Apr	Mei	Jun
1	ANT VANILA	30	26	23	40	33	40
2	BAKING POWDER	30	24	20	25	28	38
3	COKLAT ELMER 10-12%	28	25	16,5	27,5	10	24
4	GARAM	20	25	17	16	24	30
5	GULA PASIR	40	45	25	40,5	40	47
…	…	…	…	…	…	…	…
20	LAKTOSA	25	19	23	19	25	20

Tabel 2 merupakan hasil permbersihan data yang sebelumnya terdapat 120 data, di ubah menjadi 20 data per bulan

4.3. *Data Transformation*

Selanjutnya dari data diatas akan di transformasi dengan mengubah menjadi 3 kategori yaitu Banyak Digunakan, Sedang Digunakan dan Sedikit Digunakan. Berikut ini adalah proses transformasi data:

a. Menghitung rentan dengan jumlah maksimal-jumlah minimal 250-70=180
b. Kemudian hitung rentan dibagi jumlah kategorinya Rentan 3
180/3 = 60
c. Lalu menentukan kategori 1 dengan menjumlahkan jumlah minimal:
70+60=130 (Jumlah $\leq$ 130 = Sedikit Digunakan)
d. Menentukan kategori ke 2 dengan menjumlahkan kategori 1 + rentan
130+60=190 (130 < Jumlah $\leq$ 190 =Sedang Digunakan)
e. Menentukan kategori 3 dengan menjumlahkan kategori 2 + rentan
190+ 60= 250 (190 < Jumlah $\leq$ 250 = Banyak Digunakan)
f. Berikut ini yaitu table kategori yang sudah didapat:

Setelah melakukan perhitungan diatas maka akan didapatkan hasil dari pembagian 3 kategori tersebut lalu dapat dilihat pada tabel berikut ini:

Tabel 3. Data Transformation

No	Nama Produk	Jumlah	Kategori
1	ANT VANILA	192	Banyak digunakan
2	BAKING POWDER	165	Sedang digunakan
3	COKLAT ELMER 10-12%	131	Sedang digunakan
4	GARAM	132	Sedang digunakan
5	GULA PASIR	237,5	Banyak digunakan
…	…	…	
20	LAKTOSA	131	Sedang digunakan

Tabel 3 merupakan hasil data transformasi jumlah stok bahan baku selama 6 bulan menjadi 3 kategori

4.4. *Implementasi K-Nearest Neighbor* (K-NN)

Setelah data sudah siap selanjutnya akan diklasifikasika menjadi dua data yaitu data *training* dan data *testing*, perbandingan dari data tersebut yaitu 80:20. Dimana pembagian dari data tersebut dilakukan secara random menggunakan *Sampling Split data* yang terdapat pada *rapidminer*.

Pada penerapan metode K-NN, ada beberapa tahapan dalam melakukan perhitungan menggunakan algoritma K-NN yaitu:

a. Menentukan Parameter K, disini peneliti menggunakan k=3
b. Menentukan jarak antara data *training* dan data *testing*, langkah berikutnya adalah penentuan atau pengambilan keputusan terhadap *knowledge* yang dihasilkan Model persamaan algoritma K-NN. Pada penelitian ini penulis hanya menggunakan jarak *Euclidean*. Berikut hasil dari perhitungan *Euclidean Distance*:

Tabel 4. Perhitungan *Euclidean Distance*

Data	d1	d2	d3	d4
1	45,0888	147,153	51,7107	66,468
2	32,5883	158,458	66,5507	59,6909
3	23,2271	172,115	73,5085	54,8407
4	21,5523	193,508	94,1037	32,8862
5	24,8042	197,153	94,0705	34,0184
6	24,347	190,176	91,3619	42,1307
7	21,8403	172,528	78,5239	45,4533
8	24,0936	196,966	94,5807	36,0902
9	61,6462	130,151	47,5316	81,4325
10	18,5472	171,817	75,7694	48,775

Selanjutnya menggurutkan data dari hasil yang sudah di hitung kemudian akan diurutkan dengan nilai jarak yang terdekat ke jarak yang terjauh (*ascending*). Setelah itu menentukan kelompok dari data hasil uji berdasarkan mayoritas nilai k terdekat. Dalam perhitungan ini peneliti menentukan nilai k = 3 maka yang akan diambil 3 data terkecil dari hasil perhitungan yang sudah di hitung. Maka berikut adalah penentuan kelompoknya.

Tabel 5. Hasil Kategori *Ascending* Data 1

Data	d1	Kategori
10	18,5472	Sedang Digunakan
3	23,2271	Sedang Digunakan
4	21,5523	Sedikit Digunakan

Tabel diatas yaitu prediksi dari perhitungan data 1 yang Dimana hasil dari prediksi menunjukan jumlah tetangga yang muncul terbanyak yaitu "Sedang Digunakan".

Tabel 6. Hasil Kategori *Ascending* Data 2

Data	d1	Kategori
9	130,151	Banyak Digunakan
1	147,153	Banyak Digunakan
2	158,458	Sedang Digunakan

Tabel diatas yaitu prediksi dari perhitungan data 2 yang dimana hasil dari prediksi menunjukan jumlah tetangga yang muncul terbanyak yaitu "Banyak Digunakan".

Tabel 7. Hasil Kategori *Ascending* Data 3

Data	d1	Kategori
9	47,5316	Banyak Digunakan
1	51,7107	Banyak Digunakan
2	66,5507	Sedang Digunakan

Tabel diatas yaitu prediksi dari perhitungan data 2 yang dimana hasil dari prediksi menunjukan jumlah tetangga yang muncul terbanyak yaitu "Banyak Digunakan".

Tabel 8. Hasil Kategori Ascending Data 4

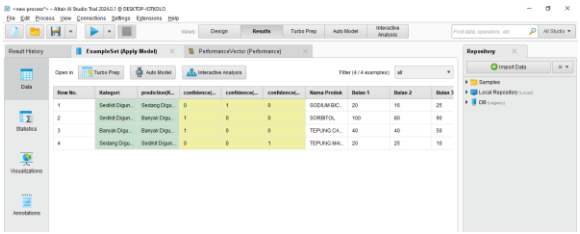
Data	d1	Kategori
4	32,8862	Sedikit Digunakan
5	34,0184	Sedikit Digunakan
8	36,0902	Sedikit Digunakan

Tabel diatas yaitu prediksi dari perhitungan data 2 yang dimana hasil dari prediksi menunjukan jumlah tetangga yang muncul terbanyak yaitu "Sedikit Digunakan". Pada tabel data 5-8 data *prediction* pada aplikasi *RapidMiner*, dapat disimpulkan kategori yang paling banyak muncul yaitu:

Tabel 9. Hasil Pengkategorian Data Training

Data	Nama Produk	Kategori
1	SODIUM BICARBONATE	Sedang Digunakan
2	SORBITOL	Banyak Digunakan
3	TEPUNG CAKRA	Banyak Digunakan
4	TEPUNG MAIZENA	Sedikit Digunakan

Selanjutnya tabel 9 dibandingkan dengan gambar di bawah ini:

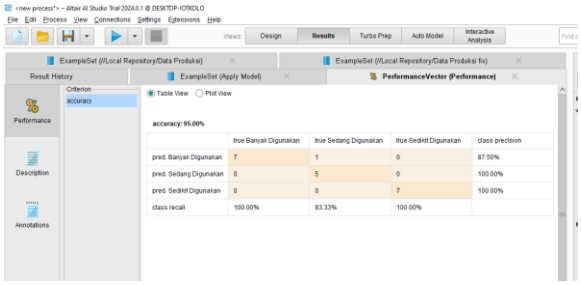


Gambar 2. Hasil kategori pada RapidMiner

Pada tabel 9 yaitu prediksi menggunakan perhitungan manual, sedangkan gambar 2 yaitu prediksi pada RapidMiner. Pada kedua prediksi tersebut menghasilkan prediksi yang sama.

4.5. Evaluasi

Pengujian ini dilakukan untuk mengetahui besar akurasi yang sudah didapatkan dalam prediksi yang sudah dilakukan sebelumnya, pada penelitian ini terdapat dua pengujian yaitu *Confusion Matrix* dan *Cross Validation*. Berikut adalah hasil pengujian *Confusion Matrix*:



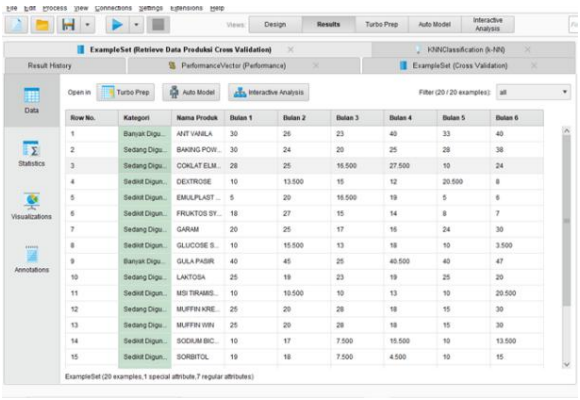
Gambar 3. Pengujian *confusion matrix*

Gambar 3 merupakan hasil dari *confusion matrix* yang menunjukan nilai akurasi yang didapat sebesar accuracy: 95,00%. dengan masing masing kategori yang sudah dibagi sebagai berikut:

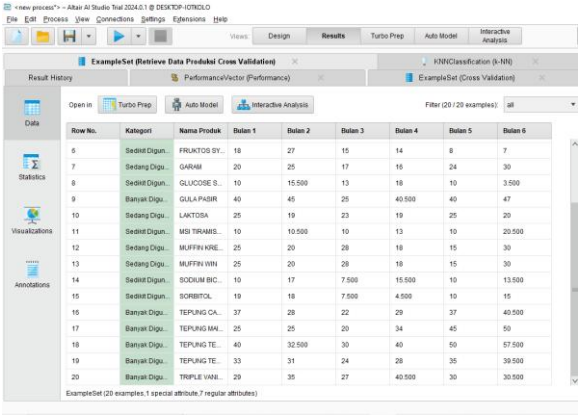
a. *Class Precision*
- Prediksi Sedikit Digunakan, terdapat 7 data yang sedikit digunakan, 1 data sedang digunakan, dan 0 data banyak digunakan, dengan *class precision* 87,50%.
- Prediksi Sedang Digunakan, terdapat 0 data sedikit digunakan, 5 data sedang digunakan, dan 0 data banyak digunakan, dengan class precision 100%.
- Prediksi Banyak Digunakan, terdapat 0 data sedikit digunakan, 0 data sedang digunakan, dan 7 data yang banyk digunakan, denga *class precision* 100%.

b. *Class Recall*
- True Sedikit Digunakan terdapat 7 data yang sedikit digunakan, 0 data sedang digunakan, dan 0 data banyak digunakan, dengan *class recall* 100%.
- True Sedang Digunakan terdapat 1 data sedikit digunakan, 5 data sedang digunakan, dan 1 data banyak digunakan, dengan class recall 83,33%.
- True Banyak Digunakan terdapat 0 data sedikit digunakan, 0 data sedang digunakan, dan 7 data yang banyk digunakan, denga *class recall* 100%.

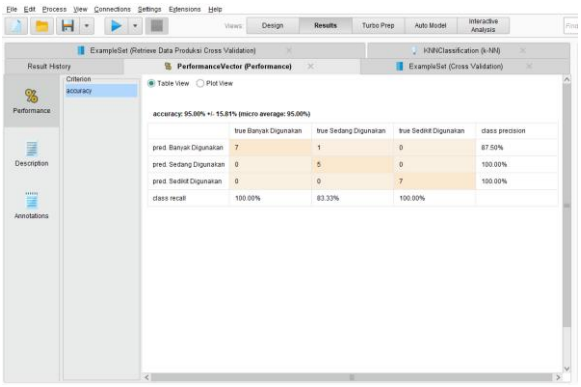
Sedangkan untuk hasil *Cross Validation* sebagai berikut :



Gambar 4. Hasil pengujian *cross validation* 1



Gambar 5. Hasil pengujian *cross validation* 2



Gambar 6. Hasil akurasi *cross validation*

Dari gambar 4, 5 dan 6 diatas dihasilkan nilai akurasi dari pengujian *cross validation* sebesar *accuracy*: 95,00% dengan standar devisi +/- sebesar 15,81% (*micro average*: 95,00%).

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang sudah dilakukan penulis, dari implementasi dan pengujian adapun hasil

yang didapatkan yaitu dengan menggunakan metode K-Nearest Neighbor, PT. XYZ dapat mengoptimalkan bahan baku kue berdasarkan data histori. Dalam penggunaan K-NN sebagai alat prediksi menghasilkan nilai akurasi sebesar 95,00%. Dengan hasil akurasi yang cukup besar artinya metode K-NN dapat membantu perusahaan dalam perencanaan produksi dan pengadaan stok bahan baku dan menggurangi resiko kekurangan atau kelebihan persediaan bahan baku kue yang tidak diinginkan yang dapat mempengaruhi proses produksi bahan kue di perusahaan.

DAFTAR PUSTAKA

[1] S. Wahyuni and A. Fitriyani, "Penerapan Algoritma K-Nearest Neighbour Untuk Memprediksi Pembelian Bahan Baku Power Transformer Pada PT Bando Electronics Indonesia," *J. Inform. SIMANTIK*, vol. 6, no. 1, pp. 12–16, 2021.

[2] M. Makalalag and H. Tjodi, "Penerapan Akuntansi Barang Dagang Sofa Di Cv. Terena Manado," *J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 10, no. 3, p. 932, 2022, doi: 10.35794/emba.v10i3.43569.

[3] N. D. Anggraeni and A. Octaviano, "Penerapan Data Mining Untuk Klasifikasi Persediaan Barang Prediksi Persediaan Barang Menggunakan Metode Safety Stock & ROP (Studi Kasus : PT . Macro Jaya Agung)," *OKTAL J. Ilmu Komput. dan Sci.*, vol. 2, no. 7, pp. 1980–1992, 2023.

[4] J. Suntoro, *Data Mining Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: Elex Media Komputindo, 2019.

[5] S. Mujiasih, "Pemanfatan Data Mining Untuk Prakiraan Cuaca," *J. Meteorol. dan Geofis.*, vol. 12, no. 2, pp. 189–195, 2011, doi: 10.31172/jmg.v12i2.100.

[6] I. Yolanda and H. Fahmi, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT . Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor," vol. 3, no. 3, pp. 9–15, 2021.

[7] E. Bu'ulolo, I. S. Tampubolon, C. V. Nababan, and L. N. Nasution, "Implementasi Algoritma K-Nearest Neighbor(K-NN) Dalam Klasifikasi Kredit Motor," *Bull. Inf. Syst. Res.*, vol. 1, no. 1, pp. 18–22, 2022, [Online]. Available: https://journal.grahamitra.id/index.php/bios

[8] S. P. Dewi, N. Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.

[9] S. Mujilahwati and L. D. Windasari, "Implementasi Metode k-Nearest Neighbor (k-NN) untuk Memprediksi Penjualan Buah di Indonesia berbasis Website," *Semin. Nas. Teknol. Sains*, vol. 3, no. 1, pp. 7–14, 2024, doi: 10.29407/stains.v3i1.4077.

[10] W. Irmayani and K. Kunci, "VISUALISASI DATA PADA DATA MINING MENGGUNAKAN METODE KLASIFIKASI Diterima : Diterbitkan :," vol. IX, no. I, pp. 68–72, 2021.

[11] D. Desyanti and D. Wulandari, "Implementasi Algoritma K-Nearest Neighbour dalam Memprediksi Stok Sepeda Motor," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1576–1581, 2022, doi: 10.47065/bits.v4i3.2579.