

PREDIKSI PERSEDIAAN BAHAN BAKU PEMBUATAN PEMPEK MENGGUNAKAN METODE NAIVE BAYES (STUDY KASUS : HOME INDUSTRY PEMPEK YURA)

Yudha Prahara^{1*}, Sewaka¹

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan, Banten 15310, Indonesia

Email: ^{1*}yudhprahara77@gmail.com, ²dosen00120@unpam.ac.id

(* : coresponding author)

Abstrak– Pempek Yura merupakan salah satu Home Industri yang memproduksi beberapa jenis pempek yaitu Pempek Kapal Selam, Pempek Lenjer, Pempek Kulit, dan Tekwan beralamat Jelupang, Tangerang selatan, Banten. Sering terjadinya kehabisan stok bahan baku Ikan Tenggiri, Sagu, dan Telur yang membuat tidak bisa memproduksi barang dan mengecewakan konsumen berakibat pembatalan, maka dibutuhkan prediksi bahan baku untuk menentukan perencanaan penyediaan stok bahan baku. Untuk mengetahui kebutuhan bahan baku yang akan datang. Penelitian ini menitik beratkan pada prediksi penentuan stok bahan baku pempek agar produksi tetap jalan, dan bahan baku selalu tersedia. Analisa prediksi stok disesuaikan dari data produksi dan penjualan, dari data produksi dan penjualan dilakukan pengolahan data menggunakan tahapan Knowledge Discovery in Database yaitu dengan menseleksi data yang dibutuhkan dalam data mining, kemudian memproses data tersebut dengan cara menghapus data penjualan yang kosong, merubah data numerik yang tidak dapat diolah oleh data mining naïve bayes, kemudian data mining menggunakan Aplikasi Rapid Miner Versi 9.10 dengan metode naïve bayes. Operator yang dipakai dalam rapidminer yaitu Retrive, Cross Validation, Naïve Bayes, Apply Model dan Performance. Hasil Akurasi dengan rapid miner pada penelitian ini menunjukkan 99.57 % dengan rincian Hasil Prediksi bahan baku Cukup 207 data sedangkan bahan baku Kurang 22 data.

Kata Kunci: Data Mining, Algoritma Naïve Bayes, Prediksi Bahan Baku

Abstract– Pempek Yura is one of the home industries that produces several types of pempek, namely Pempek Kapal Selam, Pempek Lenjer, Pempek Kulit, and Tekwan having the addresses at Jelupang, Tangerang Selatan, Banten. There are frequent stockouts of raw material for Mackerel, Sago and Eggs which prevent the production of goods and disappoint consumers resulting in cancellations, so raw material predictions are needed to determine the plan for supplying raw material stocks. To determine the future raw material requirements. This research focuses on predicting the stock of pempek raw materials so that production continues, and raw materials are always available. Stock prediction analysis is adjusted from production and sales data, from production and sales data data processing is carried out using the Knowledge Discovery in Database stage, namely by selecting the data needed in data mining, then processing the data by deleting empty sales data, changing the numerical data cannot be processed by naïve Bayes data mining, then data mining uses the Rapid Miner Application Version 9.10 with the naïve Bayes method. The operators used in rapidminer are Retrive, Cross Validation, Naïve Bayes, Apply Model and Performance. Accuracy results with the rapid miner in this study showed 99.57% with details of Prediction Results for raw materials Enough 207 data while raw materials lacked 22 data.

Keywords: Data Mining, Naïve Bayes Algorithm, Raw Material Prediction.

1. PENDAHULUAN

Pempek Yura merupakan *Home* Industri yang memproduksi beberapa jenis pempek yaitu Pempek Kapal Selam, Pempek Lenjer, Pempek Kulit, dan Tekwan, yang berbahan baku Ikan Tenggiri, Sagu, dan Telur Pempek Yura berada di Pintu Masuk Lapangan Gelora Hasanryka Jelupang, kecamatan Serpong Utara, Kota Tangerang Selatan. Pempek Yura didirikan pada Desember 2019. Dalam pengamatan kondisi yang terjadi pada Pempek Yura selama ini, masih terdapat masalah dalam penyediaan bahan baku. Diantaranya adalah masih sering terjadi kehabisan bahan baku sebelum waktu yang sudah ditentukan yaitu selama satu bulan. Hal ini memunculkan kemungkinan pembeli (konsumen)

kecewa bahkan beralih ke Warung Pempek lainnya, di karenakan ketersediaan bahan baku yang butuhkan sudah habis.

Data mining adalah suatu metode pengolahan data untuk menentukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode data mining ini dapat di gunakan untuk mengambil keputusan dimasa depan. salah satu teknik yang dibuat dalam data mining adalah bagaimana menelusuri data yang ada untuk membangun sebuah model. Kebutuhan untuk prediksi juga dapat memanfaatkan teknik ini. Dalam data mining, pengelompokan data juga bisa dilakukan. tujuannya adalah agar kita dapat mengetahui pola universal data-data yang ada (Yahya & Puspita Hidayanti, 2020).

Naïve Bayes Classifier (NBC) yang merupakan sebuah pengklasifikasi probabilitas sederhana yang mengaplikasikan *teorema bayes* dengan asumsi ketidaktergantungan (*independent*) yang tinggi. Keuntungan menggunakan NBC bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matrix kovarians. (Simanjuntak, 2021)

Berdasarkan latar belakang tersebut, maka penulis tergerak untuk membuat prediksi yang memuat informasi tentang penyediaan bahan baku dari bulan lalu yaitu agustus 2021. Pada latar belakang diatas, maka penulis mengajukan judul penelitian “**PREDIKSI PERSEDIAAN BAHAN BAKU PEMBUATAN PEMPEK DENGAN MENGGUNAKAN METODE NAVIE BAYES (STUDY KASUS : HOME INDUSTRY PEMPEK YURA)**”.

2. METODOLOGI PENELITIAN

2.1 Definisi Penerapan

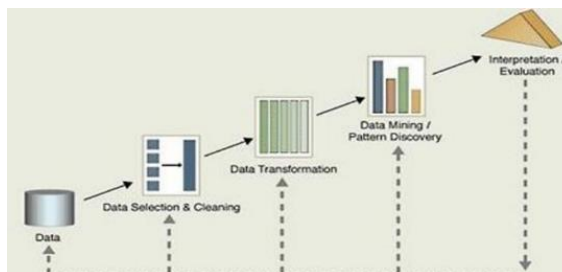
Pengertian Penerapan Menurut J.S Badudu dan Sutan Mohammad Zain, penerapan adalah hal, cara atau hasil (Badudu & Zain, 1996:1487). Adapun menurut Lukman Ali, penerapan adalah mempraktekkan, memasang (Ali, 1995:1044). Wahab, 1990:45 Berdasarkan pengertian tersebut dapat disimpulkan bahwa penerapan merupakan sebuah tindakan yang dilakukan baik secara individu maupun kelompok dengan maksud untuk mencapai tujuan yang telah dirumuskan. Adapun unsur-unsur penerapan meliputi:

- a. Adanya program yang dilaksanakan
- b. Adanya kelompok target, yaitu masyarakat yang menjadi sasaran dan diharapkan akan menerima manfaat dari program tersebut.
- c. Adanya pelaksanaan, baik organisasi atau perorangan yang bertanggung jawab dalam pengelolaan, pelaksanaan maupun pengawasan dari proses penerapan tersebut (Monata et al., 2022).

2.2 Data Mining

Menurut Suntoro, J., (2019), di dunia ini skala volume data yang kita buat setiap hari sebesar 2,5 exabyte ($2,5 \times 10^{18}$ byte), dimana data 90% data berada di dunia ini telah kita buat sejak dari tahun lalu. Diprediksi sampai tahun 2020, skala volume data akan meningkat drastis menjadi 40 zeeabyte (40×10^{21} byte). Skala volume data yang berjumlah besar tersebut akan menjadi sampah pada penyimpanan apabila tidak diolah menjadi informasi yang berguna (Yulianto, 2019). *Data mining* merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Data mining mulai ada sejak 1990-an sebagai cara yang benar dan tepat untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antara data untuk melakukan pengelompokkan ke dalam satu atau lebih cluster sehingga objek-objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. *Data mining* merupakan bagian dari proses penemuan pengetahuan dari basis data *Knowledge Discovery in Databases* (Mai et al., 2022).

2.3 Knowledge Discovery in Database



Gambar 1. Knowledge Discovery In Database Process

Knowledge Discovery in Database (KDD) digunakan untuk mengekstrak model yang menggambarkan data kelas yang penting dilakukan beberapa tahapan yang disebut preprocessing. Hal tersebut disebabkan karena teknik preprocessing memiliki dampak signifikan terhadap kinerja pada algoritma machine learning. Desain database yang baik dan analisis yang baik dapat mereduksi permasalahan missing data melalui preprocessing. Dengan menggunakan teknik visualisasi preprocessing dapat diperoleh beberapa pengetahuan mengenai data. Berikut merupakan beberapa tahapan awal hingga hasil dari penambangan data (Kokom Komariyah et al., 2020; dan Niar et al., 2022)

a. Pemilihan (*Data Selection*)

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.

b. Pemrosesan Awal (*Preprocessing*)

Sebelum proses penambangan data dapat dilaksanakan, perlu dilakukan proses cleaning dengan tujuan untuk membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

c. Transformasi

Proses coding pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses penambangan data. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam database.

d. Penambangan Data (*Data Mining*)

Proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam penambangan data sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

e. Interpretasi/Evaluasi

Pola informasi yang dihasilkan dari proses penambangan data perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya atau tidak.

2.4 Definisi Prediksi

Pengertian Prediksi sama dengan ramalan atau perkiraan. Menurut kamus besar bahasa Indonesia, prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan merupakan input bagi proses perencanaan dan pengambilan keputusan. (Rustam et al., 2020).

2.5 Algoritma Naive Bayes

Algoritma Naive Bayes (NB) merupakan metode yang sederhana dalam klasifikasi berdasarkan teori probabilitas yang dikemukakan pertama kali oleh ilmuwan Inggris bernama Thomas Bayes. Disebut naif karena menyederhanakan masalah yang bergantung pada dua asumsi penting (Osmanbegović & Suljić, 2012). Keuntungan dari klasifikasi *Naive Bayes* adalah bahwa algoritma ini tidak membutuhkan data pelatihan dalam jumlah yang besar dalam proses klasifikasi (Bhardwaj & Pal, 2011).

2.6 Cross Validation

Klasifikasi dengan 10 fold cross-validation, dimana total data trainingnya ada (satu) bagian dijadikan data uji coba (testing) dan ada (sembilan) bagian yang lain dijadikan sebagai data latihan (training), dan proses tersebut dilakukan berulang secara bergantian hingga mendapat nilai hasil secara keseluruhan sehingga dapat di ambil nilai rata-ratanya. Hasil dari model klasifikasi yang dibangun akan dilakukan proses pengujian dengan beberapa sampel yang telah dibagi sebagai testing. Dari sembilan model akan dilakukan perbandingan dengan mengukur tingkat keakuratannya (Rachmatika & Bisri, 2020).

3. ANALISA DAN PEMBAHASAN

3.1 Analisis Data Knowledge Discovery in Database

Tahapan yang akan dilakukan pada pengolahan data mining ini akan mengacu pada tahapan pengolahan Knowledge Discovery in Database (KDD), akan menghasilkan informasi yang telah diolah sesuai dengan urutan tahapan yang sudah ditentukan, berikut data tabel yang akan diproses berdasarkan tahapan-tahapannya:

3.1.1 Data Selection

Data yang dipilih adalah data sekunder yaitu produksi dan data penjualan pada bulan Agustus 2021, dari data tersebut yang digunakan adalah Jenis Bahan Baku, Nama Barang, Stok BB, Jumlah Pakai BB, Sisa Pakai BB, Jual, dan Bahan Baku yang akan diprediksi.

Tabel 1. Data Produksi dan Penjualan Pempek Yura

No	Jenis BB	Nama Barang	Stok BB/Kg	Jumlah Pakai BB/Kg	Sisa BB/Kg	Jual
1	Ikan	Pempek K.Selam	20	1.3	18.7	19
2	Sagu	Pempek K.Selam	20	1.3	18.7	19
3	Telur	Pempek K.Selam	2	0.3	1.7	19
4	Ikan	Pempek Kulit	18.7	0.3	18.5	15
5	Sagu	Pempek Kulit	18.7	0.3	18.5	15
6	Ikan	Pempek Lenjer	18.5	1.3	17.1	20
7	Sagu	Pempek Lenjer	18.5	1.3	17.1	20
8	Ikan	Tekwan	17.1	0.9	16.3	13
9	Sagu	Tekwan	17.1	0.9	16.3	13

3.1.2 Preprocessing

Tahapan selanjutnya pada KDD yaitu tahap preprocessing yang merupakan tahap yang harus dilakukan sebelum melakukan data mining. Tahapan ini meliputi proses integrasi data untuk menggabungkan data dari database yang berbeda-beda, selanjutnya agar menghasilkan data yang bersih untuk melakukan tahapan mining maka data harus terlebih dahulu melakukan proses cleaning.

3.1.3 Transformation

Data transformation dilakukan dengan memberikan inisialisasi terhadap data yang memiliki nilai akan disesuaikan type data yang dibutuhkan pada algoritma Naïve Bayse. Type data yang dibutuhkan yaitu Nominal dan Real, memberikan dua kategori pada kolom Stok BB yaitu dengan merubah Habis jika Stok sudah “0” dan tiga kategori pada kolom Jual Kategori yaitu sedikit, sedang, banyak Berikut prosesnya :

1. Menghitung rentang pada setiap nama barang dengan (jumlah maksimal jumlah minimal)
 - Pempek Kpl.Selam : $30 - 9 = 21$
 - Pempek Lenjer : $30 - 6 = 24$
 - Pempek Kulit : $26 - 2 = 24$
 - Tekwan : $13 - 1 = 12$
2. Lalu hitung rentang dibagi jumlah kategori
 - Rentang : 3
 - Pempek Kpl.Selam : $21 : 3 = 7$
 - Pempek Lenjer : $24 : 3 = 8$
 - Pempek Kulit : $24 : 3 = 8$
 - Tekwan : $12 : 3 = 4$
3. Menentukan kategori 1 dengan menjumlahkan, jumlah minimal + rentang
 - Pempek Kpl.Selam : $9 + 7 = 16$
 - Pempek Lenjer : $6 + 8 = 14$
 - Pempek Kulit : $2 + 8 = 10$
 - Tekwan : $1 + 4 = 5$
4. Menentukan kategori 2 dengan menjumlahkan, kategori 1 + rentang
 - Pempek Kpl.Selam : $16 + 7 = 23$
 - Pempek Lenjer : $14 + 8 = 22$
 - Pempek Kulit : $10 + 8 = 18$
 - Tekwan : $5 + 4 = 9$
5. Menentukan kategori 3 dengan menjumlahkan, kategori 2 + rentang
 - Pempek Kpl.Selam : $23 + 7 = 30$
 - Pempek Lenjer : $22 + 8 = 30$
 - Pempek Kulit : $18 + 8 = 26$
 - Tekwan : $9 + 4 = 13$
6. Maka didapat table kategori sebagai berikut:

Tabel 2. Kategori Untuk Atribut Jual

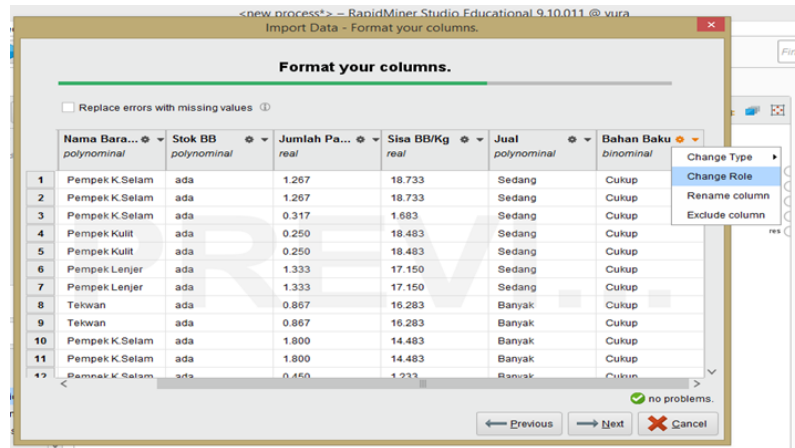
Pempek Kpl.Selam		Pempek Lenjer	
Jumlah	Kategori	Jumlah	Kategori
≤ 16	Sedikit	≤ 14	Sedikit
$> 16 \leq 23$	Sedang	$> 14 \leq 22$	Sedang
$> 23 \leq 30$	Banyak	$> 22 \leq 30$	Banyak
Pempek Kulit		Tekwan	
Jumlah	Kategori	Jumlah	Kategori
≤ 10	Sedikit	≤ 5	Sedikit
$> 10 \leq 18$	Sedang	$> 5 \leq 9$	Sedang
$> 18 \leq 26$	Banyak	$> 9 \leq 13$	Banyak

Tabel 3. Data yang sudah *Preprocessing, Transformation*

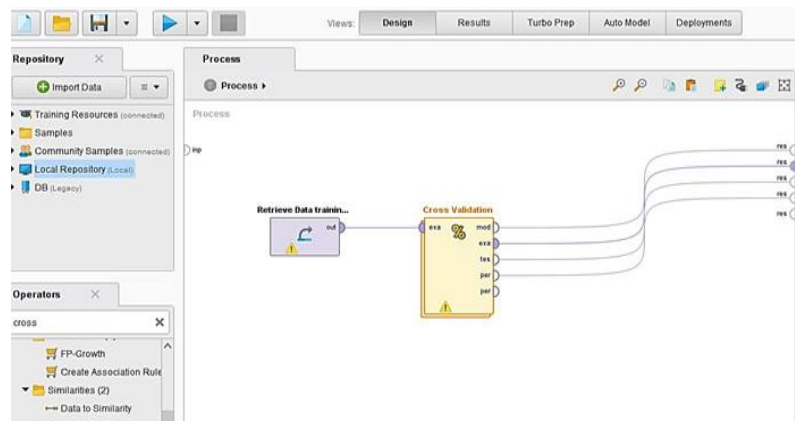
No	Jenis BB	Nama Barang	Stok BB	Jumlah Pakai BB/Kg	Sisa BB/Kg	Jual	Bahan Baku
1	Ikan	Pempek K.Selam	ada	1.3	18.7	Sedang	Cukup
2	Sagu	Pempek K.Selam	ada	1.3	18.7	Sedang	Cukup
3	Telur	Pempek K.Selam	ada	0.3	1.7	Sedang	Cukup
4	Ikan	Pempek Kulit	ada	0.3	18.5	Sedang	Cukup
5	Sagu	Pempek Kulit	ada	0.3	18.5	Sedang	Cukup
6	Ikan	Pempek Lenjer	ada	1.3	17.1	Sedang	Cukup
7	Sagu	Pempek Lenjer	ada	1.3	17.1	Sedang	Cukup
8	Ikan	Tekwan	ada	0.9	16.3	Banyak	Cukup
9	Sagu	Tekwan	ada	0.9	16.3	Banyak	Cukup

3.1.4 Data Mining

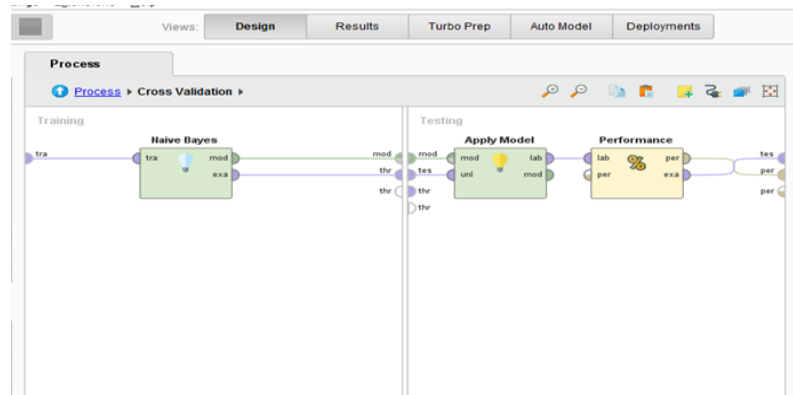
Tahapan ini merupakan tahapan untuk mencari kecocokan dari pola dan informasi menarik dalam data yang terpilih dengan menggunakan metode tertentu berdasarkan proses KDD ini secara keseluruhan, pada penelitian ini menggunakan metode Naïve Baye pada rapidminer dengan : Operator Retrieve > Cross Validation > Naïve Bayes > Apply Model > Performance berikut cara penerapannya:



Gambar 2. Dataset Pempek Yura



Gambar 3. Retrieve dan Cross Validation



Gambar 4. Operator lain yang digunakan

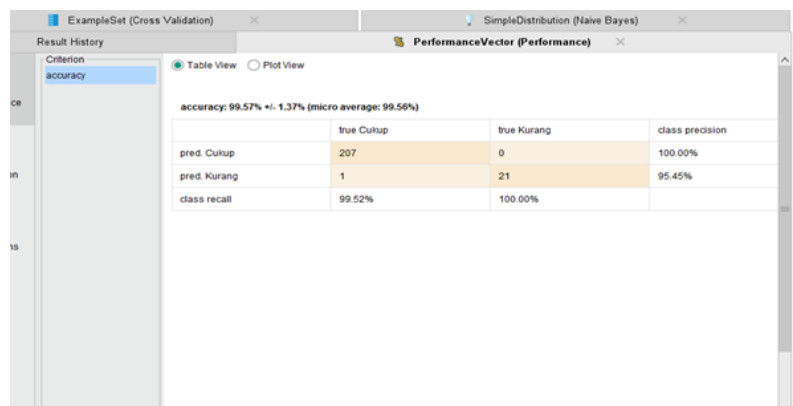
Berdasarkan gambar diatas menjelaskan bahwa dalam menerapkan model algoritma naïve bayes dengan pengujian Cross validation menggunakan aplikasi rapidminer versi 9.10 dengan operator kegunaan yaitu retriye, cross validation, naïve bayes, apply model, dan performance.

4. IMPLEMENTASI

4.1 Interpretation/Evaluasi

Pada pola informasi yang sudah dihasilkan setelah proses data mining harus ditampilkan dalam bentuk yang dapat dengan mudah di pahami oleh pihak yang berkepentingan. Tahap ini akan memeriksa apakah pola dari informasi yang telah di temukan sesuai atau bertentangan dengan fakta atau hipotesis yang ada sebelumnya. Pada tahap ini akan di dapatkan pola informasi dari prediksi persediaan bahan baku pempek dengan metode *Naïve Bayes*, pola informasi yang dihasilkan dari proses data mining berupa rules yang berasal dari perhitungan *Naïve Bayes*.

Berdasarkan penerapan diatas menggunakan pengujian cross validation dengan hasil akurasi berikut :



The screenshot shows the 'PerformanceVector (Performance)' window. It displays the following data:

accuracy: 99.57% +/- 1.37% (micro average: 99.56%)			
	true Cukup	true Kurang	class precision
pred. Cukup	207	0	100.00%
pred. Kurang	1	21	95.45%
class recall	99.52%	100.00%	

Gambar 5. Hasil Pengujian *Cross Validation*

Hasil penelitian dari Prediksi Persediaan Bahan Baku Pembuatan Pempek Menggunakan Metode Naïve Bayes pada Home Industry Pempek Yura menggunakan tahapan Knowledge Discovery in Database dengan metode Naïve Bayes pada aplikasi RapidMiner dengan pengujian Cross Validation menghasilkan akurasi 99.57% dengan rincian dari 229 data didapat 207 data prediksi Cukup, sedangkan 22 data Kurang.

5. KESIMPULAN

Berdasarkan kegiatan penelitian yang telah dilakukan penulis implementasi dan pengujian didapatkan hasil bahwa:

- a. Dengan menggunakan metode *naïve bayes* yaitu melakukan prediksi persediaan bahan baku pempek pada Pempek Yura, dapat memudahkan pemilik toko untuk melakukan proses prediksi bahan baku secara sistematis dan efisien.
- b. Dengan menggunakan metode *naïve bayes* pada toko, pemilik memperoleh informasi persediaan bahan baku untuk menentukan persediaan bahan baku secara efisien dan memiliki acuan untuk persediaan.

REFERENCES

- Kokom Komariyah, Rahaditya Dasuki, Dias Bayu Saputra, Saeful Anwar, & Gifthera Dwilestari. (2020). Klasifikasi Stok Barang Menggunakan Algoritma Naïve Bayes Pada Pt.Dharma Electrindo Manufacturing. *KOPERTIP : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 4(2), 35–41. <https://doi.org/10.32485/kopertip.v4i2.117>
- Mai, P., Tarigan, S., & Hardinata, J. T. (2022). *Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap) Implementation of Data Mining Using Apriori Algorithm in Determining Inventory (Case Study : Sinar Harahap Store)*. 2(1), 9–19.
- Monata, R. S., Yazid, M. F., & Carolina, V. (2022). Analisis Penerapan Material Lokal Elemen Interior Homestay Peraja Coffee & Bungalow Di Desa Loyok, Lombok Timur. *Jurnal Ilmiah Desain & Konstruksi*, 21(2), 280–295. <https://doi.org/10.35760/dk.2022.v21i2.7329>
- Niar, Y., Komariah, K., Surip, A., Saputra, R., & Ali, I. (2022). Implementasi Algoritma Naïve Bayes Untuk Prediksi Persediaan Barang Rotan. *KOPERTIP : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 4(1), 28–34. <https://doi.org/10.32485/kopertip.v4i1.112>
- Rachmatika, R., & Bisri, A. (2020). Perbandingan Model Klasifikasi untuk Evaluasi Kinerja Akademik Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 6(3), 417. <https://doi.org/10.26418/jp.v6i3.43097>
- Rustam, R., Rahmatullah, S., Supriyanto, S., & Wahyuni, S. (2020). pre. *Jurnal Informasi Dan Komputer*, 8(2), 75–86. <https://doi.org/10.35959/jik.v8i2.186>
- Simanjuntak, S. F. I. (2021). *Analisa Data Mining Menggunakan Metode Bayes Untuk Mengukur Tingkat Kerusakan Mesin Motor*. <http://repository.upbatam.ac.id/977/>
- Yahya, Y., & Puspita Hidayanti, W. (2020). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada “Lombok Vape On.” *Infotek : Jurnal Informatika Dan Teknologi*, 3(2), 104–114. <https://doi.org/10.29408/jit.v3i2.2279>
- Yulianto, T. (2019). Prediksi Penjualan Produk Menggunakan Algoritma Naive Bayes. *Journal Teknologi Yogyakarta*, 1(2), 3–10.