**JURNAL INFORMATIKA**

 1

**Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat**

Herry Derajad Wijaya1 , Saruni Dwiasnati2

1Universitas Mercu Buana Jakarta

Fakultas Ilmu Komputer

Jl. Meruya Selatan 1 No. 1, Kembangan, 11650, Jakarta Barat

Herry.derajad@mercubuana.ac.id

2Universitas Mercu Buana Jakarta

Fakultas Ilmu Komputer

Jl. Meruya Selatan 1 No. 1, Kembangan, 11650, Jakarta Barat

Saruni.Dwiasnati@mercubuana.ac.id

**Abstrak**

*Jenis obat yang makin lama makin bervariatif, dari obat yang berharga murah sampai harga yang kalau dilihat sangat kurang masuk akal namun fungsinya sangat bagus. Meningkatnya peredaran jenis obat terutama vitamin, hal ini mendorong penulis untuk melakukan penelitian untuk menentukan produk vitamin mana yang LAKU atau TIDAK LAKU yang bisa di gunakan sebagai pedoman sebuah apotek menentukan jumlah stok barang yang harus ada pada gudang apotek tersebut. Penelitian ini bertujuan untuk mendapatkan nilai accuracy untuk data penjualan obat terutama jenis-jenis vitamin dengan menggunakan algoritma klasifikasi data mining yaitu algoritma Naïve Bayes. Penelitian ini menggunakan tools Rapidminner versi 8 sebagai media untuk menguji data yang akan diolah untuk mendapatkan hasil accuracy dan ROC.*

***Kata kunci : Datamining, Penjualan Obat, Naïve Bayes, Rapidminner.***

***Abstract***

 *The types of drugs are increasingly varied, from low-priced drugs to prices that when seen are very less reasonable but the function is very good. Increased circulation of types of drugs, especially vitamins, this prompted the authors to conduct research to determine which vitamin products are LAKU or TIDAK LAKU that can be used as a guideline for a pharmacy to determine the amount of stock that must be in the warehouse of the pharmacy. This study aims to obtain the value of accuracy for data on drug sales, especially types of vitamins using the data mining classification algorithm, namely the Naïve Bayes algorithm. This study uses Rapidminner version 8 tools as a medium to test the data to be processed to obtain accuracy and ROC results.*

***Keywords: Datamining, Drug Sales, Naïve Bayes, Rapidminner.***

-------------------------------------------------------------------------------------------------------

1. **Pendahuluan**

Dalam dunia kesehatan khususnya dunia obat-obatan di seluruh indonesia sudah sampai pada titik yang mengesankan dan sangat menarik untuk terus ditelusuri dalam kuantity jenis dari berbagai merk yang ada. Pertumbuhan bisnis kesehatan kususnya apotek-apotek semakin berkembang dari waktu ke waktu, hal ini dapat dilihat dari banyaknya apotek yang tersebar baik di perkotaan sampai ke pedalaman Indonesia. Dijaman sekarang, apotek tidak hanya menyediakan obat-obatan saja namun sudah sampai menyediakan alat-alat kosmetik, Karena hal tersebut, banyak apotek yang berlomba-lomba dalam menyediakan puluhan ragam varian obat-obatan untuk kesehatan serta vitamin agar dapat memudahkan pelanggannya dalam memenuhi kebutuhannya dengan variasi harga yang berbeda-beda.

Dalam pemberian informasi mengenai data LAKU atau TIDAK LAKU sebuah apotek terhadap barang yang ada di apotek tersebut, seperti vitami sering mengalami kesulitan untuk menentukan barang mana yang LAKU atau TIDAK LAKU yang diminati oleh para konsumennya. Hal ini diakibatkan oleh menumpuknya stok vitamin yang kurang diminati di gudang, namun untuk vitamin yang diminati sama sekali tidak ada dalam persedian pada apotek tersebut.

Untuk mengatasi masalah yang dijelaskan di atas, maka dilakukan pengklasifikasian obat yang LAKU dan TIDAK LAKU terjual di Apotek tersebut berdasarkan variabel-variabel yang di dapatkan. Algoritma yang digunakan dalam pengklasifian obat yang LAKU atau TIDAK LAKU menggunakan algoritma Naïve Bayes yang mampu memberikan informasi mengenai vitamn yang LAKU atau TIDAK LAKU pada apotek tersebut untuk meminimalisasikan stok yang menumpuk untuk vitamin yang di kategorikan TIDAK LAKU.

 Ketersediaan data yang banyak dalam sebuah bidang dan kebutuhan untuk sebuah informasi atau pengetahuan dalam bidang tertentu dapat menjadi acuan sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis dan dukungan infrastruktur di segala bidang merupakan cikal- bakal dari lahirnya teknologi *data mining*. Penggunaan teknik *data mining* diharapkan dapat membantu percepatan proses pengambilan keputusan, memungkinkan perusahaan untuk mengelola informasi yang terkandung didalam data nasabah dan menjadi sebuah pengetahuan (*knowledge*) yang baru untuk menentukan keputusan yang dibutuhkan.

 Oleh karena itu, untuk benar-benar mengetahui dan menetapkan jenis vitamin ini LAKU atau TIDAK LAKU perlu dilakukan forecasting (prediksi/estimasi) sehingga jenis vitamin tersebut akan sesuai dengan jenis vitamin yang diminati oleh pelanggan dari apotek tersebut. Setelah jenis obat ditetapkan, kemudian diestimasi berapa banyak permintaan jenis vitamin tersebut pada bulan berikutnya dengan menggunakan algoritma Naive Bayes.

 Dalam melakukan estimasi (forecasting), banyak sekali metode-metode yang telah tersedia. Akan tetapi, dalam mengestimasi banyaknya penjualan pada bulan berikutnya terhadap jenis vitamin yang banyak di cari oleh pelanggan, metode forecasting yang cocok adalah algoritma Naive Bayes. Algoritma Naive Bayes cocok diterapkan pada data yang berskala ordinal. Algoritma Naive Bayes adalah suatu metode yang dapat menangani faktor prediksi pada bulan selanjutnya dan trend secara langsung. Keuntungan algoritma ini adalah memiliki kemampuan yang sangat baik dalam mengestimasi data yang memiliki pola trend dan prediksi untuk bulan selanjutnya. Oleh karena itu, algoritma Naive Bayes merupakan algoritma yang cocok dalam mengestimasi banyaknya penjualan pada bulan berikutnya terhadap jenis vitamin yang ada pada apotek yang diminati oleh pelanggan.

Berawal dari beberapa disiplin ilmu yang ada, data mining memiliki tujuan untuk memperbaiki teknik tradisional yang dapat menghasilkan sesuatu yang jauh lebih informatif sehingga bisa menangani: Jumlah data yang sangat besar, Dimensi data yang tinggi dan Data yang heterogen dan berbeda sifat [ Ratniasih, 2019 ].

Setiap metode tersebut memiliki algoritma-algoritma yang terbaik yang dapat digunakan untuk memproses dataset yang ada. Namun dalam Penelitian ini diharapkan dapat membangun sebuah sistem dalam mengestimasi jenis vitamin untuk meningkatkan penjualan pada bulan berikutnya yang sesuai dengan data-data penjualan sebelumnya sehingga jenis vitamin yang diminati oleh pelanggan benar-benar tersedia pada apotek tersebut sehingga dapat membantu apotek dalam meningkatkan omset penjualannya menggunakan algoritma Naïve Bayes.

Data mining merupakan suatu langkah dalam melakukan Knowledge Discovery in Databases (KDD). Knowledge discovery sebagai suatu proses terdiri atas pembersihan data (data cleaning), integrasi data (data integration), pemilihan data (data selection), transformasi data (data transformation), data mining, evaluasi pola (pattern evaluation) dan penyajian pengetahuan (knowledge presentation). Data mining mengacu pada proses untuk menambang (mining) pengetahuan dari sekumpulan data yang sangat besar untuk menghasilkan sebuah pengetahuan baru dari bidang tertentu [Firdaus, 2017].

 Pada Data Mining terdapat 3 bagian yaitu: Assosiation, Klasifikasi dan Clustering. Assosiation memiliki definisi sebuah proses yang digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data yang dimiiki, sedangkan Klasifikasi adalah teknik yang dilakukan untuk memprediksi class atau properti dari setiap instance data, dan Clustering sendiri memliki makna pengelompokan data tanpa berdasarkan kelas data tertentu ke dalam kelas objek yang sama sesuai dengan topik yang diangkat.

 Prediksi memiliki kemiripan definisi dengan klasifikasi, akan tetapi data yang digunakan pada klasifikasi berdasarkan perilaku atau nilai yang diperkirakan pada masa yang akan datang yang dilihat dari pola/nilai pada masa lalu. Contoh dari tugas prediksi misalnya untuk memprediksikan adanya pengurangan jumlah pelanggan dalam beberapa waktu yang lalu dan prediksi harga saham dalam empat bulan ke depan.

 Salah satu metode utama data mining adalah Supervised Learning ( Pembelajaran dengan Guru ), algoritma ini melakukan proses belajar berdasarkan nilai dari indikator target yang terhubung dengan indikator prediktor. Sedangkan metode analisis yang terdapat pada data mining salah satu nya adalah Klasifikasi. Klasifikasi sendiri memiliki definisi sebuah proses yang digunakan untuk menemukan model atau fungsi yang mendeskripsikan dan membedakan data ke dalam kelas-kelas yang bertujuan untuk menemukan pola yang lebih memiliki nilai dari data yang berukuran relatif besar hingga sangat besar. Klasifikasi melibatkan proses pemeriksaan karakteristik dari objek dan memasukkan objek ke dalam salah satu kelas yang sudah didefinisikan sebelumnya. Dalam klasifikasi hanya terdapat satu atribut dari sekian banyaknya atribut yang bisa menjadi kemungkinan yang disebut atribut target, sedangkan atribut lainnya yang terdapat disebut atribut prediktor. Tiap kemungkinan nilai yang dimiliki oleh atribut target menunjukkan class yang diprediksi berdasarkan nilai-nilai dari atribut prediktor.

Naive Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes (aturan bayes) dengan asumsi independensi ( ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam Naive Bayes model yang digunakan adalah model fitur independen (Prasetyo,2012).

 Sistem pengolahan data yang ada

pada beberapa apotek yang ada di Indonesia masih belum dapat mengetahui dengan pasti produk apa saja yang LAKU atau TIDAK LAKU pada bulan sebelumnya agar mengetahui produk mana yang harus di stok lebih banyak lagi untuk bulan selanjutnya

Data yang akan digunakan pada penelitian ini adalah 150 dataset. Penelitian ini dilakukan karena atas masalah mengetahui produk pada bulan sebelumnya agar dapat menentukan stok produk mana yang mesti di perbanyak menggunakan algoritma Naïve Bayes agar mengetahui nilai accuracy nya.

Penelitian sebelumnya mengenai data mining dengan menggunakan algoritma Naïve Bayes adalah hasil penelitian yang dilakukan oleh (Dwiasnati, Devianto, 2018) menghasilkan algoritma Naïve Bayes dapat dilakukan dalam penelitian di berbagai aspek bidang, seperti pada data set dalam memprediksi Nasabah Potensial pada sebuah perusahaan asuransi. Hasil penelitian yang dilakukan oleh ( Budiyanto, Dwiasnati, 2018) menyatakan bahwa algoritma Naïve Bayes tersebut mampu membantu penyelesaian permasalahan pemilihan produk terlaris di sebuah perusahaan yang bergerak dalam pengadaan dan layanan barang dari sistem CCTV bersama dengan sistem Kontrol Akses dengan tingkat akurasi memiliki kebenaran 81.36%. Selanjutnya penelitian yang dilakukan oleh (Prabowo, Kodar, 2019) dapat digunakan untuk memprediksi lama masa studi mahasiswa pada sebuah universitas baik fakultas eksak ataupun fakultas social.

1. **Metode Penelitian**

Penelitian ini akan mengusulkan hasil accuracy baru untuk sebuah permasalahan memprediksi Produk LAKU atau TIDAK LAKU pada bulan sebelumnya untuk mnegetahui pola minat pelanggan terhadap pembelian vitamin pada sebuah apotek. Data yang digunakan dalam penelitian ini di ambil pada tahun 2018 bulan Januari-Mei, data yang di ambil kisaran 5.000 an namun untuk data testing untuk penelitian ini penulis hanya mengambil sampel data sebanyak 137.

Model desain ini akan melakukan pemrosesan data training dan data testing untuk menguji metode algoritma yang digunakan. Tahapan yang akan dilalui dibagi menjadi 3 bagian, yaitu preprocessing, seleksi fitur (Feature Selection) dan validation yang didalamnya berisi sub proses training dan testing.

**a. Pengumpulan Data**

Data yang digunakan pada penelitian kali ini merupakan data sekunder, karena sumber data diperoleh melalui media perantara atau secara tidak langsung yang berupa buku, catatan, bukti yang telah tervalidasi, atau arsip baik yang dipublikasikan maupun yang tidak dipublikasikan secara umum yaitu 100 ribu data yang akan dipecah menjadi 80 persen data training dan 20 persen data testing. Data testing yang digunakan sekitar 150 data yang telah diklasifikasikan berdasarkan variable yang ada.

Masalah yang harus dipecahkan pada penelitian kali ini adalah, bagaimana menghasilkan nilai accuracy yang jauh lebih optimal untuk mengklasifikasikan prediksi Calon Nasabah Potensial untuk ditawari produk lainnya. Dataset yang digunakan pada penelitian ini ada pada Gambar di bawah ini:



Gambar 1. Dataset

**b. Pengolahan Data**

Dataset ini dalam tahap preprocessing harus melalui 3 proses, yaitu:

Tahap ini akan melibatkan sub proses antara lain**:**

1. Membuang duplikasi data
2. Memeriksa data yang inkonsisten
3. Memperbaiki kesalahan pada data.

Dan Hasil akhir dari data ini berupa kumpulan data yang sudah bersih atau tidak ada missing valuenya.

**c. Evaluasi dan Validasi Data**

Validasi dilakukan menggunakan 10 fold cross validation. Dimana dengan menggunakan teknik ini dengan membagi secara acak ke dalam tiap bagian dimana terdiri dari 10 bagian untuk setiap bagian akan dilakukan proses klasifikasi terlebih dahulu, Sedangkan pengukuran akurasi diukur dengan confusion matrix dan kurva ROC (Receiver Operating Characteristics) untuk mengukur nilai AUC.

AUC digunakan untuk mengukur kinerja diskriminatif dengan memperkirakan probabilitas output yang sudah di dapatkan hasilnya dari sampel yang dipilih secara acak dari populasi positif atau negatif, semakin besar nilai AUC, semakin kuat klasifikasi yang dihasilkan. Karena AUC merupakan bagian dari daerah unit persegi, nilainya yang dihasilkan akan selalu sama yang dihasilkannya, antara 0,0 dan 1,0.

Tabel 1. Confusion Matrix



Gambar 2. Sumber : (Gorunescy, 2011)

1. **Hasil dan Pembahasan**

Data yang digunakan dalam pengklasifikasiaan Calon Nasabah Potensial terdiri dari 5.000 dataset,137 data yang digunakan untuk data testing berdasarkan variabel yang tersedia.

Klasifikasi dilakukan dengan menggunakan Software Rapidminer dengan versi 8 untuk mengolah data yang sudah di tentukan

3.1. Confusion Matrix algoritma Naïve Bayes

Gambar.3 merupakan perhitungan akurasi data training menggunakan algoritma Naïve Bayes yang menghasilkan accuracy 88.00%. Diketahui data training terdiri dari 150 record data, 10 data di klasifikasikan LAKU ternyata TIDAK LAKU, 2 data diprediksi TIDAK LAKU dan benar-benar TIDAK LAKU, 129 di prediksikan LAKU ternyata benar-benar LAKU serta 8 data diprediksikan TIDAK LAKU ternyata LAKU. 

Gambar 3. Model Confusion Matrix algoritma Naïve Bayes

PerformanceVector:

accuracy: 88.00% +/- 10.67% (mikro: 87.92%)

ConfusionMatrix:

True: Tidak Ya

Tidak: 2 8

Ya: 10 129

precision: 92.74% +/- 3.26% (mikro: 92.81%) (positive class: Ya)

ConfusionMatrix:

True: Tidak Ya

Tidak: 2 8

Ya: 10 129

recall: 94.01% +/- 11.73% (mikro: 94.16%) (positive class: Ya)

ConfusionMatrix:

True: Tidak Ya

Tidak: 2 8

Ya: 10 129

AUC (optimistic): 0.514 +/- 0.330 (mikro: 0.514) (positive class: Ya)

AUC: 0.494 +/- 0.329 (mikro: 0.494) (positive class: Ya)

AUC (pessimistic): 0.478 +/- 0.328 (mikro: 0.478) (positive class: Ya)

Gambar 4 .Text view model Confusion Matrix algoritma Naïve Bayes

Gambar 4. merupakan perhitungan accuracy data menggunakan algoritma Naïve Bayes. Diketahui data training terdiri dari 150 record data, 10 data di klasifikasikan LAKU ternyata TIDAK LAKU, 2 data diprediksi TIDAK LAKU dan benar-benar TIDAK LAKU, 129 di prediksikan LAKU ternyata benar-benar LAKU serta 8 data diprediksikan TIDAK LAKU ternyta LAKU.

Model yang dihasilkan pada algoritma Naïve Bayes yang dihasilkan menggunakan Rapidminner, yaitu:

Gambar 5. Model algoritma Naïve Bayes

**3.2 Algoritma Naïve Bayes**

Algoritma Naïve Bayes Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediski probabilitas keanggotaan suatu class. Bayesian classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decesion tree dan neural network. Bayesian classification terbukti memiliki akurasai dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar[Supriyanto, 2013].

Metode Bayes merupakan pendekatan statistic untuk melakukan inferensi induksi pada persoalan klasifikasi. Pertama kali dibahas terlebih dahulu tentang konsep dasar dan definisi pada Teorema Bayes, kemudian menggunkan teorema ini untuk melakukan klasifikasi dalam Data Mining.Teorema Bayes memiliki bentuk umum sebagai berikut :



Keterangan :

X = Data dengan class yang belum diketahui H = Hipotesis data X merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.) P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

**3.3 Penjualan**

Penjualan merupakan sebuah proses dimana kebutuhan pembeli dan kebutuhan penjual dipenuhi, melalui antar pertukaran informasi dan kepentingan. Dalam proses penjualan, penjual atau penyedia barang dan jasa memberikan kepemilikan suatu komoditas kepada pembeli untuk suatu harga tertentu. Secara umum pengertian penjualan adalah sejumlah total yang dikenakan kepada pelanggan untuk barang atau jasa yang di jual.

**4. Kesimpulan**

1. Algoritma Naïve Bayes memiliki tingkat accuracy dengan nilai 88.00%.

2. Rule yang dihasilkan algoritma Naïve Bayes diterapkan dalam prototype prediksi penjualan obat di sebuah apotek yang ada di Jakarta selatan. Berdasarkan akurasi yang dihasilkan prototype menunjukkan bahwa algoritma dan prototype yang diterapkan untuk pencarian hasil dapat diimplementasikan dengan baik.

**Referensi**

 Arif, B., Saruni, D. (2018). The Clasification Product Using Naïve Bayes Algorithm

(A Case Study at PT Putradabo Perkasa). International Journal of Computer Techniques, 68-74

Fajar, E.P., Achmad, K.(2019). Analisis Prediksi Masa Studi Mahasiswa

Menggunakan Algoritma Naïve Bayes. Jurnal Ilmu Teknik dan Komputer, 147-151.

Firdaus,Diky (2017). “Penggunaan Data

Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer”. Jurnal FORMAT, Vol 6 Nomor 2, ISSN : 2089-5615

Prasetyo, E, 2012, Data Mining : Konsep dan Aplikasi menggunakan MATLAB, Yogyakarta: Andi

Ratniasih, Ni Luh, 2019. Optimasi Data Mining Menggunakan Algoritma Naïve Bayes dan C4.5 Untuk Klasifikasi Kelulusan Mahasiswa. Jurnal Teknologi Informasi dan Komputer, Volume 5, Nomor 1, Januari 2019.

Saruni, D., Yudo, D. (2018). Naive Bayes Optimization Based On Particle Swarm Optimization To Predict The Decision Of Insurance Customer Candidate. International Journal of Computer Techniques, 1-7.

Supriyanto, Catur. Purnama Parida. 2013. deteksi penyakit diabetes type ii dengan naive bayes berbasis particle swarm optimization. Jurnal Teknologi Informasi, Volume 9 Nomor 2, Oktober 2013.