

Penerapan Data Mining Analisa Data Penjualan Obat Menggunakan Metode Random Forest

Cut Reka

Fakultas Ilmu Komputer dan Teknologi Informasi, Program Studi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia
Email: cutrekha11@gmail.com
Email Penulis Korespondensi: cutrekha11@gmail.com

Abstrak-Data Mining adalah suatu metode yang digunakan untuk mencari dan menggali data tertentu yang tersembunyi dari sebuah data yang besar. Sebagai contoh, data mining dapat digunakan untuk mencari informasi kombinasi item dalam suatu penjualan. Dengan menggunakan data mining, perusahaan akan mampu menganalisa secara tepat, cepat dan akurat dibandingkan dengan menganalisa secara manual. Apotek merupakan salah satu perusahaan yang bisa memanfaatkan metode data mining ini, karena di Apotek transaksi penjualan berlangsung setiap hari sehingga semakin lama data penjualan yang tersimpan sangatlah besar. Penerapan Data Mining analisa data penjualan Menggunakan Metode Random Forest pada metode ini data maupun atribut diambil secara acak sehingga dimungkinkan akan menghasilkan sebagai model pohon keputusan, manfaatnya mempermudah analisis data yang besar dan membantu memberikan informasi data penjualan yang diolah.

Kata Kunci: Penjualan Obat; Data Mining; Random Forest

Abstract-Data Mining is a method used to find and explore certain hidden data from big data. For example, data mining can be used to find information on the combination of items in a sale. By using data mining, companies will be able to analyze precisely, quickly and accurately compared to analyzing manually. The pharmacy is one of the companies that can take advantage of this data mining method, because in the pharmacy sales transactions take place every day so that the longer the sales data stored is very large. Application of Data Mining sales data analysis Using the Random Forest Method in this method the data and attributes are taken at random so that it is possible to produce a decision tree model, the benefits of which are to facilitate the analysis of large data and help provide information on processed sales data.

Keywords: Drug Sales; Data Mining; Random Forest

1. PENDAHULUAN

Penjualan adalah merupakan kegiatan yang terpadu untuk mengembangkan rencana – rencana strategis yang diarahkan kepada usaha pemuasan kebutuhan serta keinginan pembeli/konsumen, gunanya untuk mendapat penjualan yang dihasilkan laba atau keuntungan. Atau suatu kegiatan transaksi yang dilakukan 2 orang atau lebih dengan memakai alat pembayaran yang sah.

Apotek Mitha Farma bergerak dalam bidang kesehatan yang berlokasi di jalan Amaliun no 14 f Medan. Dimana terjadi proses penjualan obat, apotek Mitha Farma merupakan sebuah apotek yang sedang berkembang dalam membantu masyarakat dengan memberikan solusi dari berbagai macam penyakit – penyakit yang beredar di kehidupan sehari – hari. Apotek bukan menjadi hal asing lagi ditelinga masyarakat Indonesia, karena keberadaannya sudah semakin dekat dan ada disekitar kita. Kalau dulu dalam suatu wilayah tertentu kita baru menjumpai satu atau dua apotek, maka sekarang kita bisa melihat banyak apotek. Hal ini dikarenakan apotek memiliki fungsi sebagai pusat obat masyarakat semakin diperlukan. Dalam dunia bisnis yang selalu dinamis dan penuh dengan persaingan bagi para pelaku dibidang yang sama harus selalu memikirkan cara untuk terus mengembangkan skala bisnis tersebut.

Sistem yang digunakan pada saat ini masih bersifat manual, dimana dalam penjualan obat dilakukan dengan pencatatan di buku setiap diadakan transaksi penjualan obat. setiap kekurangan dalam pelayanan kefarmasian yang dilakukan oleh petugas membuat banyak pasien yang merasa tidak puas. dimana seseorang dapat memberikan tanggapan mengenai baik buruknya pelayanan suatu apotek. Dikarenakan hal tersebutlah menimbulkan masalah bagi Apotek Mitha Farma yang dimana ketika kekurangan *stock* dan tertimbunnya *stock* akan berpengaruh pada pemasukan bagi Apotek Mitha Farma. Selain itu jika obat – obatan yang diperlukan oleh masyarakat tidak tersedia akan merusak *image*. Padahal hal tersebut dapat diatasi dengan mengetahui penjualan obat – obatan yang dilakukan oleh Apotek Mitha Farma selama ini. Untuk mengetahui dan melakukan pengolahan data penjualan dapat dilakukan dengan sebuah teknik yang dinamakan Data Mining.

Data mining merupakan sebuah analisis dan peninjauan dari kumpulan data untuk menemukan sebuah informasi yang tidak terduga dan meringkas data dengan cara yang berbeda dengan data sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. Data mining merupakan sebuah teknik yang merupakan gabungan dari beberapa bidang keilmuan untuk melakukan *visualisasi*, pengenalan pola dan statistik untuk penanganan permasalahan pengambilan informasi dari kumpulan-kumpulan data [1], [2]. Dengan memanfaatkan teknik data mining diharapkan dapat mengatasi permasalahan yang dialami oleh Apotek Mitha Farma. Kiranya melakukan pengolahan pada data penjualan Apotek Mitha Farma dapat mengetahui obat – obatan yang diperlukan oleh masyarakat umum, dengan mengetahui obat – obatan yang diperlukan tidak terjadi lagi kekosongan *stock* ataupun penimbunan *stock* pada Apotek Mitha Farma. Salah satu metode yang digunakan untuk menyelesaikan masalah menggunakan teknik Data Mining tersebut adalah Metode *Random Forest*.

Metode *Random Forest* merupakan sebuah teknik klasifikasi yang digunakan pada kumpulan – kumpulan data berskala besar yang merupakan sebuah penggabungan dari beberapa pohon dan pengembangan dari metode *Classification and Regression Tree* (CART). Metode ini didasarkan pada teknik pohon keputusan sehingga mampu menyelesaikan

permasalahan yang bersifat non-linear, penggabungan pohon pada Metode *Random Forest* akan mempengaruhi tingkat akurasi hasil yang didapatkan menjadi lebih baik dan dapat digunakan bagi pemilik data. Metode ini untuk membangun pohon keputusan yang terdiri dari *root node*, *internal node*, dan *leaf node* dengan mengambil atribut dan data secara acak sesuai ketentuan yang diberlakukan. *Root node* merupakan simpul yang terletak paling atas. *Internal node* adalah simpul percabangan, dimana *node* ini mempunyai *output* minimal dua *input*. Sedangkan *leaf node* merupakan simpul terakhir yang memiliki satu *input* dan tidak mempunyai *output*. Pohon keputusan dimulai dengan cara menghitung nilai *entropy* sebagai penentu tingkat ketidakhomogenan atribut dan nilai *information gain*.

Berdasarkan penelitian yang dilakukan oleh penelitian lain menunjukkan bahwa akurasi yang dihasilkan dari proses klasifikasi sangatlah tinggi sehingga hasil yang didapatkan adalah Akurat dan penggunaan metode *random forest* sangat baik.

Penerapan metode *Random Forest* ini diharapkan dapat mengolah data Penjualan obat dan memberikan informasi yang akurat juga bisa dijadikan bahan pertimbangan instansi dalam *stock* Penjualan obat yang diperlukan, agar kiranya tidak terjadi penumpukan barang yang membuat instansi rugi.

2. METODOLOGI PENELITIAN

2.1 Data Mining

Data *mining* adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. Menurut Tan data *mining* sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data *mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data *mining* kadang disebut juga *knowledge discovery*. Data *mining* berisi pencarian trend atau pola yang diinginkan dalam database yang besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya [1].

2.2 Metode Random Forest

Metode *random forest* adalah pengembangan dari metode *CART*, yaitu dengan menerapkan metode *bootstrap aggregating* (*bagging*) dan *random feature selection*. Metode ini merupakan metode pohon gabungan. Dalam *random forest*, banyak pohon ditumbuhkan sehingga terbentuk suatu hutan (*forest*), kemudian analisis dilakukan pada kumpulan pohon tersebut. Pada gugus data yang terdiri atas n amatan dan p peubah penjelas, prosedur untuk melakukan *random forest* adalah [3], [4]:

1. Lakukan penarikan contoh acak berukuran n dengan pemulihan pada gugus data. Tahap ini adalah tahapan *bootstrap*.
2. Dengan menggunakan contoh *bootstrap*, pohon dibangun sampai mencapai ukuran maksimum (tanpa pemangkasan). Pembangunan pohon dilakukan dengan menerapkan *random feature selection* pada setiap proses pemilihan pemilah, yaitu m peubah penjelas dipilih secara acak dimana $m \ll p$, lalu pemilah terbaik dipilih berdasarkan m peubah penjelas tersebut.
3. Ulangi langkah 1 dan 2 sebanyak k kali, sehingga terbentuk sebuah hutan yang terdiri atas k pohon.

Random forest memprediksi respons suatu amatan dengan cara menggabungkan (*aggregating*) hasil prediksi k pohon [5]. Untuk masalah klasifikasi, pohon yang dibangun adalah pohon klasifikasi dan hasil prediksi *random forest* adalah berdasarkan *majority vote* (suara terbanyak), Nilai salah klasifikasi *random forest* diduga melalui *error OOB* yang diperoleh dengan cara:

1. Lakukan prediksi terhadap setiap data OOB pada pohon yang bersesuaian. Data OOB (*out of bag*) adalah data yang tidak termuat dalam contoh *bootstrap*.
2. Secara rata-rata, setiap amatan gugus data asli akan menjadi data OOB sebanyak sekitar 36% dari banyak pohon. Oleh karena itu, pada langkah 1, masing-masing amatan gugus data asli mengalami prediksi sebanyak sepertiga kali dari banyaknya pohon. Jika a adalah sebuah amatan dari gugus data asli, maka hasil prediksi *random forest* terhadap a adalah gabungan dari hasil prediksi setiap kali a menjadi data OOB.
3. *Error OOB* dihitung dari proporsi misklasifikasi hasil prediksi *random forest* dari seluruh amatan gugus data asli.

Breiman dan *Cutler* menyarankan untuk mengamati error OOB saat k kecil, lalu memilih m yang menghasilkan error OOB terkecil. Jika *random forest* dilakukan dengan menghasilkan variable importance, disarankan untuk menggunakan banyak pohon, misalnya 1000 pohon atau lebih. Jika peubah penjelas yang dianalisis sangat banyak, nilai tersebut dapat lebih besar agar *variable importance* yang dihasilkan semakin stabil. Dengan diterapkannya metode *Random Forest* ini pada permasalahan data mining untuk prediksi *stock* bahan kimia yang paling dibutuhkan dan bahan kimia yang memang akan di *stock* berlebih sebagai cadangan persediaan, diharapkan dapat membantu dalam menyusun rencana kebutuhan bahan kimia dengan strategi yang tepat untuk menekan dan mengurangi penumpukan bahan kimia yang tidak terpakai.

3. HASIL DAN PEMBAHASAN

Analisis apotek Mitha Farma merupakan suatu organisasi bisnis yang berkembang di Indonesia. Apotek bergerak dalam bidang bisnis khususnya menjual obat-obatan dan alat alat kesehatan. Pada Apotek Mitha Farma terdapat beberapa permasalahan yang kerap muncul mengenai penjualan obat Dalam pengolahan data perusahaan menggunakan sistem manual. Analisis data pengendalian *stock* obat yang mudah dan efektif agar tidak dapat terjadi penumpukan obat dan kekurangan *stock* obat sehingga pihak apotek dalam menganalisa jenis obat mana yang paling diminati atau tidak diminati konsumen.

3.1 Penerapan Metode Random Forest

Metode ini merupakan metode pohon gabungan. Dalam *random forest*, banyak pohon ditumbuhkan sehingga terbentuk suatu hutan (*forest*), kemudian analisis dilakukan pada kumpulan pohon tersebut. Pada gugus data yang terdiri atas n amatan dan p peubah penjelas, prosedur untuk melakukan *random forest* sebagai berikut

1. Lakukan penarikan contoh acak berukuran n dengan pemulihan pada gugus data. Tahap ini adalah tahapan *bootstrap*.
2. Dengan menggunakan contoh *bootstrap*, pohon dibangun sampai mencapai ukuran maksimum (tanpa pemangkasan). Pembangunan pohon dilakukan dengan menerapkan *random feature selection* pada setiap proses pemilihan pemilah, yaitu m peubah penjelas dipilih secara acak dimana $m \ll p$, lalu pemilah terbaik dipilih berdasarkan m peubah penjelas tersebut.
3. Ulangi langkah 1 dan 2 sebanyak k kali, sehingga terbentuk sebuah hutan yang terdiri atas k pohon *Randomforest* memprediksi respons suatu amatan dengan cara menggabungkan (*aggregating*) hasil prediksi k pohon. Untuk masalah klasifikasi, pohon yang dibangun adalah pohon klasifikasi dan hasil prediksi *random forest* adalah berdasarkan *majority vote* (suara terbanyak), yaitu kategori atau kelas yang paling sering muncul sebagai hasil prediksi dari k pohon klasifikasi.

Data diperoleh dari Apotek Mitha Farma Medan yang tersedia. Dalam penelitian ini data penjualan obat yang akan diolah pada sistem data mining dengan menggunakan metode *Random Forest*, terdapat 10 jenis obat yang digunakan di Apotek Mitha Farma Medan. Adapun data penjualan obat pada Apotek Mitha Farma Medan seperti berikut ini:

Tabel 1. Data obat Pada Apotek Mitha Farma Medan

No	Nama Obat	Satuan	Masuk	Keluar	Total
1	Viks F44 Syrup	botol	60	60	60
2	Bodrexin Tablet	box	4	4	4
3	Buscopan Tablet	box	3	3	3
4	Paracetamol Tablet	box	5	5	5
5	Mylanta Syrup	Botol	12	12	12
6	Nerobion Putih Tablet	box	15	15	15
7	Wodss Exp Syrup	Botol	6	6	6
8	Panadol Tablet	box	12	12	12
9	Decolgen Tablet	box	5	5	5
10	Ambroxol Syrup	Botol	12	12	12

Berikut ini adalah penyelesaian dengan kasus berdasarkan data yang sudah disediakan pada tabel 4.1. Proses pembentukan C1 atau disebut dengan 1 itemset dengan jumlah minimum support = 100%, dengan rumus sebagai berikut:

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}} * 100\%$$

Maka dari data diatas dapat dihitung untuk nilai suport setiap item seperti di bawah ini :

$$S (\text{Viks F44}) = \frac{60}{134} * 100 = 44\%$$

$$S (\text{Bodrexing}) = \frac{4}{134} * 100 = 6\%$$

$$S (\text{Buscopan}) = \frac{3}{134} * 100 = 7\%$$

Proses diteruskan hingga didapatkan tabel seperti berikut

Tabel 2. Nilai Koefisien

No	Nama Obat	Perhitungan	Support
1	Viks F44 Syrup	$60/134 * 100\%$	0,44%
2	Bodrexin Tablet	$4/134 * 100\%$	6%
3	Buscopan Tablet	$4/134 * 100\%$	7%
4	Paracetamol Tablet	$5/134 * 100\%$	3%
5	Mylanta Syrup	$12/134 * 100\%$	4%
6	Nerobion Putih Tablet	$15/134 * 100\%$	60%
7	Wodss Exp Tablet	$6/134 * 100\%$	4%
8	Panadol Tablet	$12/134 * 100\%$	8,08%
9	Decolgen Tablet	$5/134 * 100\%$	3%

No	Nama Obat	Perhitungan	Support
10	Ambroxol Syrup	12/134*100%	8,08%

Proses pembentukan pola kombinasi 2 item yang dibentuk dari tabel di atas dan Proses Pembentukan C2 atau disebut dengan 2 item, dapat diselesaikan dengan rumus berikut:

$$\text{Support}(A,B) = P(A \cap B)$$

$$\text{Support}(A \cap B) = \frac{\text{Transaksi Mengandung A dan B}}{\text{Total Transaksi}} * 100\%$$

Tabel 3. Nilai Tingkat Koefisien

No	Nama Obat	Perhitungan	Support
1	Viks F44 Syrup	60/134*100%	0,44%
2	Paracetamol Tablet	5/134*100%	3%
3	Nerobion Putih Tablet	81/134*100%	60%
4	Panadol Tablet	12/134*100%	8,08%
5	Decolgen Tablet	5/134*100%	3%
6	Ambroxol Syrup	12/134*100%	8.08%

Setelah semua pola frekuensi tinggi ditemukan, baru dicari aturan association rules yang memenuhi syarat minimum *confidence*, dengan menghitung *confidence* aturan *asosiasi* jika A ke B, Dengan Minimal *confidence* = 2 % Nilai *Confidence* dari aturan A ke B

Untuk mencari nilai *confidence* nilai *support* ditentukan minimal 2 % maka Dari kombinasi item yang telah ditemukan, dapat dilihat besarnya nilai *support* dan *confidence* dari calon aturan asosiasi seperti pada tabel di bawah ini:

Tabel 4. Aturan metode Random Forest yang Memenuhi Confidence Minimal Confidence

No	Nama Obat	Confidence %
1	Viks f 44 syrup	60 0,44%
2	Paracetamol Tablet	5 3%
3	Nerobion putih	81 60%
4	Woods exp Syrup	6 4%
5	Panadol Tablet	12 8,08%
6	Decolgen Tablet	5 3%
7	Ambroxol Syrup	12 9.08%

Hasil dari metode random forest menunjukkan nilai dalam bentuk presentase dimana obat yang sering digunakan di Apotek Mitha Farma Medan, hasil yang ditampilkan mulai dari 2% hingga 60 %, yang artinya obat yang sering cepat habis yaitu *Neurobion putih* dengan nilai presentasi 60 % disusul dengan obat lainya seperti *Viks F44* dengan nilai presentasi yang digunakan 0.44%.

4. KESIMPULAN

Kesimpulan yang dapat diambil dari proses penelitian Data yang digunakan dalam menentukan penjualan obat di Apotek Mitha Farma berdasarkan obat yang diperlukan dan diminati oleh konsumen sehingga data tersebut dapat menghasilkan sistem yang meningkatkan penjualan obat di Apotek Mitha Farma. Algoritma *Random Forest* dapat membantu mengembangkan strategi penjualan yang sangat baik digunakan untuk menentukan *stock* di apotek Mitha Farma agar tidak terjadi kekurangan dalam penjualan obat-obatan.

REFERENCES

- [1] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Deepublish, 2020.
- [2] D. Nofriansyah, *Konsep Data Mining Vs Sistem Pendukung Keputusan*. 2015.
- [3] H. Zhu, X. Xia, J. Yao, H. Fan, Q. Wang, and Q. Gao, "Comparisons of different classification algorithms while using text mining to screen psychiatric inpatients with suicidal behaviors," *J. Psychiatr. Res.*, vol. 124, no. December 2019, pp. 123–130, 2020.
- [4] W. Apriliah, I. Kurniawan, M. Baydhowi, and T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *Sistemasi*, vol. 10, no. 1, p. 163, 2021.
- [5] A. Thakkar and R. Lohiya, "Attack classification using feature selection techniques: a comparative study," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 1, 2021.