

Penerapan Data Mining Untuk Estimasi Penjualan Obat Berdasarkan Pengaruh Brand Image Dengan Algoritma Expectation Maximization (Studi Kasus: PT. Pyridam Farma Tbk)

Yane Laheroi Nainel, Efori Buulolo, Ikwana Lubis

Program Studi Teknik Informatika, STMIK Budi Darma, Medan, Indonesia

Email: Yanenainel@gmail.com

Submitted 31-03-2020; Accepted 19-04-2020; Published 26-04-2020

Abstrak

Data mining, sering juga disebut knowledge discovery in database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan. PT. Pyridam Frama Tbk merupakan perusahaan multi nasional yang memproduksi farmasi. Adapun masalah yang sering terjadi pada PT Pyrida Farma adalah masalah estimasi seperti kehabisan stok barang pada distributor, kurangnya tenaga kerja, kehabisan bahan baku di pabrik. Masalah lain yang terdapat dalam instansi yaitu belum mempunyai sistem untuk memprediksi estimasi penjualan obat setiap tahunnya sehingga diperlukan suatu Algoritma yaitu Algoritma Expectation Maximization. maka dengan ini dibuatlah Penerapan Data Mining Untuk Estimasi Penjualan Obat Berdasarkan Pengaruh Brand Image. Algoritma Expectation Maximization yang merupakan metode yang mendukung dalam memperkirakan atau memprediksi estimasi target penjualan untuk periode yang akan datang. Pengujian Algoritma dilakukan dengan menggunakan software SPSS dan MYSQL. Dari hasil penelitian yang telah dilakukan maka dapat membantu pihak PT. Pyridam Farma untuk mempermudah dalam memprediksi estimasi penjualan obat dengan menggunakan Software SPSS.

Kata Kunci: Data Mining, Estimasi, Expectation, Maximization

Abstract

Data mining, often also called knowledge discovery in database (KDD), is an activity that includes the collection, use of historical data to find regularities, patterns or relationships in large data sets. The output of data mining can be used to improve decision making in the future. PT. Pyridam Frama Tbk is a multi-national company that produces pharmaceuticals. The problems that often occur at PT Pyrida Farma are estimation problems such as out of stock of goods at distributors, lack of labor, running out of raw materials at the factory. Another problem contained in the agency is that it does not yet have a system to predict the estimated drug sales each year, so we need an algorithm, the Expectation Maximization Algorithm. then with this made the Application of Data Mining for Estimating Drug Sales Based on the Effect of Brand Image. Expectation Maximization algorithm which is a method that supports in estimating or predicting sales target estimates for the coming period. Algorithm testing is done using SPSS and MYSQL software. From the results of research that has been done it can help the PT. Pyridam Farma to make it easier to predict drug sales estimates by using SPSS Software.

Keywords: Data Mining, Estimation, Expectation, Maximization

1. PENDAHULUAN

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Data Mining dapat diterapkan pada berbagai bidang yang mempunyai sejumlah data, tetapi karena wilayah penelitian dengan sejarah yang belum lamadan belum melewati masa 'remaja', maka data mining masih diperdebatkan posisi bidang pengetahuan yang memilikinya. Maka Daryl Pregibon menyatakan bahwa "Data Mining adalah campuran dari statistik, kecerdasan buatan, dan riset basis data" yang masih berkembang. Data Mining mempunyai empat akar bidang ilmu yaitu statistik, kecerdasan buatan (Artificial Intelligence) pengenalan pola dan sistem basis data.

Estimasi adalah keseluruhan proses yang memerlukan serta menggunakan estimator untuk menghasilkan sebuah estimate dari suatu parameter. Sedangkan menurut Kamus Besar Bahasa Indonesia (KBBI) Estimasi adalah perkiraan, penilaian atau pendapat. Ini menunjukkan bahwa istilah estimasi dapat digunakan secara umum untuk menyatakan perkiraan, penilaian atau pendapat mengenai sesuatu. Masalah yang sering terjadi di dalam suatu Instansi adalah masalah estimasi seperti kehabisan stok barang pada distributor, kurangnya tenaga kerja, kehabisan bahan baku di pabrik. Masalah lain yang terdapat dalam Instansi yaitu belum mempunyai sistem untuk memprediksi estimasi penjualan obat setiap tahunnya, untuk itu diperlukan suatu algoritma yaitu Algoritma Expectation Maximization yang merupakan metode yang mendukung dalam memperkirakan atau memprediksi estimasi target penjualan untuk periode yang akan datang.

Citra Merek (Brand Image) diperlukan oleh PT. Pyridam Farma dengan tujuan untuk membangun suatu daya tarik terhadap konsumen, dimana konsumen akan memilih kepercayaan terhadap produk dan jasa yang ditawarkan oleh Pyridam Farma tersebut, dan selain itu juga citra merek dapat menjadi suatu bentuk positioning dalam benak konsumen yang secara tidak langsung mendorong konsumen untuk memilih produk dari Pyridam Farma tersebut tanpa adanya pertimbangan dari produk perusahaan lainnya. Positioning ini juga dapat membangun suatu persepsi pada diri konsumen sehingga menjadi suatu keunggulan tersendiri yang hanya dimiliki oleh Pyridam Farma tersebut saja.

Kantor PT. Pyridam Farma adalah perusahaan yang memiliki lebih dari 100 produk dalam bentuk tablet, kaplet, kapsul, sirup, krim dan salep. Selain itu, pyridam juga memproduksi produk resep seperti penisilin dan non-penisilin antibiotic, anti-TBC, obat penghilang rasa sakit, serta produk non-resep prok vitamin pencegah flu dan batuk dan antipiretik. Perusahaan ini memiliki sebuah divisi khusus untuk memasarkan produk pasokan medis yang meliputi reagen laboratorium dan peralatan medis sekali pakai. Nama pyridam terinspirasi oleh fenomena Piramida salah satu tertua di dunia.

Dengan data mining diharapkan dapat membantu proses perhitungan perusahaan untuk mengelola informasi yang terkandung didalam data transaksi menjadi sebuah pengetahuan yang baru, sehingga melalui gambaran pemikiran yang ada, pihak perusahaan dapat mengambil langkah yang tepat untuk peningkatan sales target penjualan kedepannya. Dalam melakukan pencapaian tersebut, maka akan diimplementasikan sebuah metode dari data mining dengan menggunakan Algoritma Expectation Maximization.

2. METODE PENELITIAN

2.1 Data Mining

Data mining, sering juga disebut knowledge discovery in database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan [1]. Dari definisi-defenisi yang telah disampaikan, hal penting yang terkait dengan data mining adalah:

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang akan mungkin memberikan indikasi yang bermanfaat.

2.2 Penjualan

Penjualan adalah interaksi antara individu saling bertemu muka yang ditujukan untuk menciptakan, memperbaiki, menguasai atau mempertahankan hubungan pertukaran sehingga menguntungkan bagi pihak lain. Penjualan dapat diartikan juga sebagai usaha yang dilakukan manusia untuk menyampaikan barang bagi mereka yang memerlukan dengan imbalan uang menurut harga yang telah ditentukan atas persetujuan bersama [6].

2.3 Algoritma Expectation Maximization

Algoritma *Expectation Maximization* adalah salah satu algoritma yang di gunakan untuk klasifikasi atau pengelompokan data [8]. Algoritma *Expectation Maximization* memiliki kelebihan dimana mampu menyelesaikan beberapa permasalahan pada bidang statistic seperti menduga parameter bagi gabungan fungsi- fungsi serta parameter yang tidak lengkap. Algoritma *Expectation Maximization* (EM) merupakan proses yang terbagi atas dua langkah yaitu :

1. Langkah *Expectation* (E-step)
 Pencarian nilai expektasi untuk fungsi likelihood berdasarkan variable yang diamati.
2. Langkah *Maximization* (M-step)
 Pencarian MLE dari parameter-parameter dengan memaksimalkan ekspektasi likelihood yang dihasilkan dari E-step.
 Parameter-parameter yang dihasilkan dari M-step akan digunakan kembali untuk E-step. yang berikutnya, dan langkah ini akan diulang terus sampai memberikan nilai suatu parameter.

3. HASIL DAN PEMBAHASAN

Analisa dapat didefenisikan sebagai penguraian dari suatu sistem informasi yang utuh kedalam bagian-bagian komponennya dengan maksud untuk mengidentifikasi dan mengevaluasi permasalahan, kesempatan, hambatan yang terjadi dalam kebutuhan yang diharapkan sehingga dapat diusulkan perbaikan-perbaikannya. Penelitian dilakukan di PT.Pyridam Farma Tbk, dan berdasarkan penelitian brand image yang dilakukan pengolahan data.

Analisa dalam penjualan dan pemesanan sangat perlu dilakukan di *Pyridam Farma Tbk*.Analisa ini berguna untuk mendapatkan pokok-pokok permasalahan yang sebenarnya menjadi inti permasalahan yang sering terjadi di *Pyridam Farma Tbk*. Dalam hal ini perlu melakukan analisa omset penjualan pertahun, ini bertujuan agar *Pyridam Farma Tbk* dapat mengetahui langkah yang harus dilakukan dalam mendata pesanandi tahun yang akandatang ,dalam melakukan analisa pada penelitian ini digunakan data yang tercantum pada *Pyridam Farma Tbk* dari tahun 2016-2017 . Adapun data yang dimaksud adalah sudah di olah dan di analisa.

3.1 Pembahasan

Algoritma Expectation Maximization adalah salah satu algoritma yang di gunakan untuk klasifikasi atau pengelompokan data. Algoritma Expectation Maximization memiliki kelebihan dimana mampu menyelesaikan beberapa permasalahan pada bidang statistic seperti menduga parameter bagi gabungan fungsi- fungsi serta parameter yang tidak lengkap.

Tabel 1. Daftar Pemesanan Obat dan Fungsinya

Data (Y)	Kegunaan	Data obat 2016-2017 Fitur (X)	Nama Obat
Y=1	Analgesik, Antipiretik, Antiinflamasi nonsteroid, Antipirai	x1	Allopurinol
		x2	Acetosal
		x3	Fentanil
		x4	Ketoprofen
		x5	Kolkisin

Data (Y)	Kegunaan	Data obat 2016-2017	
		Fitur (X)	Nama Obat
Y=2	Anti alergi dan obat untuk Anafilaksis	x6	Morfın
		x7	Deksametason
		x8	Epinefrin
		x9	Loratadin
		x10	Klorpheniramin
		x11	Dipenhidramin
		x12	Cetirizin
		x13	Kalsium Glukonat
		x14	Mg Sulfat
		x15	Na Bikarbonat
Y=3	Antidot dan Obat lain untuk Keracunan	x16	Amikasin
		x17	Diazepam
		x18	Fenitoin
		x19	Ciprofloksasin
Y=4	Antiepilepsi – Antikonvulsi	x20	Dapson
		x21	Asiklovir
		x22	Dikloksasilin
		x23	Doksisiklin
		x24	Efavirens
		x25	Eritromisin
		x26	Bleomisin
Y=5	Anti Infeksi	x27	Cisplatin
		x28	Dakarbasin
		x29	Doksorubisin
		x30	Furosemida
Y=6	Antineoplastik, Imunosupresan dan obat untuk terapi paliatik	x31	HCT
		x32	Manitol
		x33	Spirolakton
Y=7	Diuretik	x34	Fitomenadion
		x35	Heparin
		x36	Warfarin
		x37	Traneksamat
Y=8	Obat yang mempengaruhi darah	x38	Fe Sulfat
		x39	Triheksifenidil
		x40	Levodopa + Karbidopa
Y=9	Antiparkinson	x41	Acarbose
		x42	Etinil Estradio
Y=10	Hormon, Obat endokrin lain dan Kontraseptik	x43	Glibenklamid
		x44	Antasida
		x45	Bisakodil
		x46	Cimetidin
Y=11	Saluran Cerna	x47	Dimenhidrinat
		x48	Domperidon
		x49	Lansoprazol
Y=12	Antimigrain	x50	Ergotamin
		x51	epatitis B rekombinan
		x52	Serum Antibisa ular
		x53	Serum Antidifteri
		x54	Serum Antirabies
		x55	Serum Antitetanus
		x56	Serum Imunoglobulin
Y=13	Obat yang mempengaruhi sistim imun	x57	Vaksin BCG
		x58	Vaksin Campak
		x59	Vaksin DTP
		x60	Vaksin jerap difteri tetanus
		x61	Prednison
		x62	Repaglinid
		x63	Rosiglitazon
		x64	Amlodipin
		x65	Atropin
		x66	Carvedilol
		x67	Digoksin
		x68	Dobutamin
		x69	Dopamin
		x70	ISDN
x71	KCL		
Y=14	Psikofarmaka	x72	Ambroksol

Data (Y)	Kegunaan	Data obat 2016-2017	
		Fitur (X)	Nama Obat
Y=15 Saluran Napas	Relaksan Otot Perifer dan Penghambat Kolinesterase	x73	Aminophilin
		x74	Asetil Sistein
		x75	Bromheksin
		x76	Pankuronium
		x77	Neostigmin
Y=16	Relaksan Otot Perifer dan Penghambat Kolinesterase	x78	Piridostigmin
		x79	Verapamil
		X80	Suksametonium
		X81	Glikuidon
		X82	Benorilate
Y=17 Kulit, Obat Topikal	Relaksan Otot Perifer dan Penghambat Kolinesterase	X83	Bismuth subcitrate
		X84	Carbinoxamin
		X85	Clindamicin
		X86	Dexpanthenol
		X87	Isoconaol
Y=18 Sebagai obat infeksi jamur	Relaksan Otot Perifer dan Penghambat Kolinesterase	X 88	Ketokonazole
		X89	Levamisole

Perhitungan Ekspektasi probabilitas dilakukan dengan contoh 3 kelas $Y=1$, $Y=2$ dan $Y=3$ dengan 4 fitur $X=1$, $X=2$, $X=3$, $X=4$. Data Y adalah kluster bidang obat yang akan direkomendasikan kepada mahasiswa berdasarkan fitur X . Fitur X merupakan representasi dari matakuliah dan nilainya. Perhitungan dengan algoritma ini mempunyai dua tahap utama yakni tahap perkiraan (Expectation) dan tahap maksimalisasi (Maximization), di mana tahap ini akan diulang (increment) sampai nilai data Y mencapai nilai 0 atau nilai 1. Secara teori perhitungan Expectation Maximization dijelaskan dari paper yang dibuat oleh Arthur Dempster, nan laird da Donald Rubin dalam jurnal 'Royal Statistical Society' pada tahun 1977. Algoritma EM termasuk algoritma clustering yang berbasiskan perhitungan probabilitas. Secara iteratif Algoritma EM melakukan 2 tahapan yaitu:

a. ExpectationStep

Pada tahap ini, akan menghitung probabilitas objek terhadap kluster.

b. MaximizationStep

Pada tahap ini, akan menghitung parameter-parameter yang digunakan untuk memaksimalkan nilai probabilitas.

Kedua tahap tersebut dilakukan secara berulang-ulang sampai mencapai jumlah konvergen. Yang dimaksud dengan data konvergen ini adalah threshold yang digunakan untuk menghentikan iterasi yang terus menerus dimana iterasi tersebut tidak akan berhenti jika nilainya tak nol0.

Perhitungan Ekspektasi probabilitas dilakukan dengan contoh 3 kelas $Y=1$, $Y=2$ dan $Y=3$ dengan 4 fitur $X=1$, $X=2$, $X=3$, $X=4$. Data Y adalah kluster bidang obat yang akan direkomendasikan kepada mahasiswa berdasarkan fitur X . Fitur X merupakan representasi dari matakuliah dan nilainya. Perhitungan dengan algoritma ini mempunyai dua tahap utama yakni tahap perkiraan (Expectation) dan tahap maksimalisasi (Maximization), di mana tahap ini akan diulang (increment) sampai nilai data Y mencapai nilai 0 atau nilai 1. Secara teori perhitungan Expectation Maximization dijelaskan dari paper yang dibuat oleh Arthur Dempster, nan laird da Donald Rubin dalam jurnal 'Royal Statistical Society' pada tahun 1977. Algoritma EM termasuk algoritma clustering yang berbasiskan perhitungan probabilitas.

Kedua tahap tersebut dilakukan secara berulang-ulang sampai mencapai jumlah konvergen. Yang dimaksud dengan data konvergen ini adalah threshold yang digunakan untuk menghentikan iterasi yang terus menerus dimana iterasi tersebut tidak akan berhenti jika nilainya tak nol0.

Tabel 2. Kondisi Awal Probabilitas data Y yang dimiliki setiap fitur X

Y=1	Y=2	Y=3	X1	X2	X3	X4
?	?	?	0	1	0	1
?	?	?	1	0	1	0
?	?	?	0	1	1	1
?	?	?	1	1	1	0
?	?	?	0	1	0	0
?	?	?	0	0	1	1

Pada table 2. di atas, data $Y=1$, $Y=2$ dan $Y=3$ adalah data awal yang akan dihitung probabilitas kelas yang dimiliki oleh setiap fitur $X1$, $X2$, $X3$ dan $X4$. Tabel tersebut menunjukkan bahwa data Y belum diketahui, untuk data setiap fitur dapat dilihat bernilai 1 jika obat yang diambil misalnya bernilai 1 dan 0 jika data obat tersebut bukan bernilai0.

Tabel 3. Frekuensi Fitur

Y	X1	X2	X3	X4
1	?	?	?	?
2	?	?	?	?
3	?	?	?	?

Tabel 3. di atas adalah table frekuensi yang dimiliki setiap fitur terhadap kelas $Y=1$, $Y=2$ dan $Y=3$. Pada awal kondisi nilai frekuensi setiap fitur x_1 , x_2 , x_3 dan x_4 belum diketahui. Untuk dapat menghitung nilai frekuensi setiap fitur terhadap setiap data, maka digunakan sebuah model parameter. Tabel 4. berikut ini menampung nilai model parameter.

Tabel 4. Model Parameter

Y	X1	X2	X3	X4	P(y)
1	0.1	0.3	0.8	0.8	0.7
2	0.2	0.3	0.1	0.1	0.2
3	0.7	0.4	0.1	0.1	0.1

Pada tabel 4. yang berisi model parameter masing-masing nilai X_1 , X_2 , X_3 dan X_4 diperoleh dari secara random dimana jumlah total seluruh fitur yang sama dari seluruh fitur harus berjumlah 1 (satu). Dapat dilihat bahwa $x_1 \rightarrow 0.1+0.2+0.7 = 1$, $X_2 \rightarrow 0.3+0.3+0.4 = 1$, $X_3 \rightarrow 0.8 + 0.1 + 0.1 = 1$, $X_4 \rightarrow 0.8+0.1+0.1=1$ dan $\alpha \rightarrow 0.1+0.2+0.1=1$. Dari nilai yang dihasilkan secara random ini berikutnya akan dihitung frekuensi setiap fitur X beserta nilai probabilitas terhadap kelas Y seperti yang ditampilkan pada table 5. berikut ini.

Tabel 5. Frekuensi fitur X baru dan Probabilitas data Y baru

prob Y=1					
X1	X2	X3	X4	P(y)	N
0.9	0.3	0.2	0.8	0.7	0.03024
0.1	0.7	0.8	0.2	0.7	0.00784
0.9	0.3	0.8	0.8	0.7	0.12096
0.1	0.3	0.8	0.2	0.7	0.00336
0.9	0.3	0.2	0.2	0.7	0.00756
0.9	0.7	0.8	0.8	0.7	0.28224
prob Y=2					
X1	X2	X3	X4	P(y)	N
0.8	0.3	0.9	0.1	0.2	0.00432
0.2	0.7	0.1	0.9	0.2	0.00252
0.8	0.3	0.1	0.1	0.2	0.00048
0.2	0.3	0.1	0.9	0.2	0.00108
0.8	0.3	0.9	0.9	0.2	0.03888
0.8	0.7	0.1	0.1	0.2	0.00112
prob Y=3					
X1	X2	X3	X4	P(y)	N
0.3	0.4	0.9	0.1	0.1	0.00108
0.7	0.6	0.1	0.9	0.1	0.00378
0.3	0.4	0.1	0.1	0.1	0.00012
0.7	0.4	0.1	0.9	0.1	0.00252
0.3	0.4	0.9	0.9	0.1	0.00972
0.3	0.6	0.1	0.1	0.1	0.00018

$$N=X_1 \times X_2 \times x_3 \times x_4 \times P(y);$$

Perhitungan setiap fitur X baru pada tabel 5. diperoleh dari nilai model parameter yang terdapat di tabel 4. Selanjutnya untuk mengisi frekuensi X baru pada tabel 5. tersebut akan di periksa nilai awal fitur X yang terdapat pada tabel 2. Perhitungan yang digunakan adalah jika nilai fitur X pada data awal bernilai 1, maka ambil nilai frekuensi yang dihasilkan secara random pada tabel model parameter, jika 0 maka frekuensi X baru adalah 1 dikurangi nilai frekuensi yang dihasilkan secara random pada model parameter tabel 4. Selanjutnya setelah semua nilai frekuensi fitur X baru dihasilkan, maka dapat dihitung nilai probabilitas data Y baru.

Tabel 6. Nilai probabilitas baru dari data Y

Y=1 baru	Y=2 baru	Y=3 baru
0.85	0.12	0.03
0.55	0.18	0.27
1.00	0.00	0.00
0.48	0.16	0.36
0.13	0.69	0.17
1.00	0.00	0.00

Maka, probabilitas setiap instance/ fitur terhadap sebuah data Y adalah sebagai berikut:

$$\prod (y/k) = P'(y) / \sum P'(y)$$

Di mana nilai $\prod (y/k)$ adalah nilai probabilitas setiap fitur terhadap sebuah data Y , nilai $P'(y)$ adalah nilai probabilitas yang dimiliki sebuah data Y , nilai $\sum P'(y)$ adalah jumlah seluruh probabilitas yang dimiliki seluruh kelas Y .

A. $\prod P(y/k)$ terhadap data $Y=1$

- Instance1 : $0.03024 / (0.03024 + 0.00432 + 0.00108) = 0.85$
 Instance2 : $0.00784 / (0.00784 + 0.00252 + 0.00378) = 0.55$
 Instance3 : $0.12096 / (0.12096 + 0.00048 + 0.00012) = 1.00$
 Instance4 : $0.00336 / (0.00336 + 0.00432 + 0.00252) = 0.48$
 Instance5 : $0.00756 / (0.00756 + 0.03888 + 0.00972) = 0.13$
 Instance6 : $0.28224 / (0.28224 + 0.00112 + 0.00018) = 1.00$
- B. $\prod P(y|k)$ terhadap data Y=2
 Instance1 : $0.00432 / (0.03024 + 0.00432 + 0.00108) = 0.12$
 Instance2 : $0.00252 / (0.00784 + 0.00252 + 0.00378) = 0.18$
 Instance3 : $0.00048 / (0.12096 + 0.00048 + 0.00012) = 0.00$
 Instance4 : $0.00432 / (0.00336 + 0.00432 + 0.00252) = 0.16$
 Instance5 : $0.03888 / (0.00756 + 0.03888 + 0.00972) = 0.69$
 Instance6 : $0.00112 / (0.28224 + 0.00112 + 0.00018) = 0.00$
- C. $\prod P(y|k)$ terhadap data Y=3
 Instance1 : $0.00108 / (0.03024 + 0.00432 + 0.00108) = 0.03$
 Instance2 : $0.00378 / (0.00784 + 0.00252 + 0.00378) = 0.27$
 Instance3 : $0.00012 / (0.12096 + 0.00048 + 0.00012) = 0.00$
 Instance4 : $0.00252 / (0.00336 + 0.00432 + 0.00252) = 0.36$
 Instance5 : $0.00972 / (0.00756 + 0.03888 + 0.00972) = 0.17$
 Instance6 : $0.00018 / (0.28224 + 0.00112 + 0.00018) = 0.00$

Tabel 6. di atas menunjukkan nilai probabilitas data Y yang baru. data tersebut ditunjukkan oleh data Y=1 baru, Y=2 baru dan data Y=3 baru. data tersebut diperoleh dari hasil perbandingan setiap data N dengan total jumlah N. Dengan perhitungan sebagai berikut:

$$Y_{\text{baru}} = N(\text{prob } Y1) / (N(\text{prob } Y1) + N(\text{prob } Y2) + N(\text{prob } Y3))$$

Selanjutnya nilai baru tersebut akan digunakan pada iterasi pertama sebagai tahap Maximization. Berikut tahap Maximization nilai probabilitas setiap kelas Y:

Tabel 7. Data Probabilitas yang dimiliki setiap fitur dalam iterasi -I

	Data			Fitur			
	Y=1	Y=2	Y=3	X1	X2	X3	X4
0.85	0.12	0.03	0	1	0	1	
0.55	0.18	0.27	1	0	1	0	
1.00	0.00	0.00	0	1	1	1	
0.48	0.16	0.36	1	1	1	0	
0.13	0.69	0.17	0	1	0	0	
1.00	0.00	0.00	0	0	1	1	

Tabel 8. Tabel Frekuensi jumlah dalam iterasi -I

Y	X1	X2	X3	X4	Total
1	1.04	2.46	3.03	2.84	4.01
2	0.33	0.97	0.34	0.13	1.15
3	0.63	0.57	0.63	0.03	0.83

Probabilitas seluruh data (X1.....X4) terhadap data Y={ 0,1,2}:

- a. $Y'(X)$ Untuk Y=1:
 $Y'(x) = \sum P(X|Y) = 0.85 + 0.55 + 1.00 + 0.48 + 0.13 + 1.00 = 4.01$
- b. $Y'(X)$ Untuk Y=2:
 $Y'(x) = \sum P(X|Y) = \{0.12 + 0.18 + 0.00 + 0.16 + 0.69 + 0.00\} = 1.15$
- c. $Y'(X)$ Untuk Y=3:
 $Y'(x) = \sum P(X|Y)$
 $= \{0.03 + 0.27 + 0.00 + 0.36 + 0.17 + 0.00\} = 0.83$

Jadi, masing-masing nilai fitur x adalah hasil jumlah seluruh nilai fitur x dikali dengan nilai probabilitas data Y yang dihasilkan oleh hasil tahap sebelumnya. Selanjutnya, pada iterasi ini lakukan kembali perhitungan model parameter.

Tabel 9. Model Parameter dalam iterasi -I

Y	X1	X2	X3	X4	alpha
1	0.26	0.61	0.75	0.71	0.67
2	0.29	0.84	0.30	0.11	0.19

Untuk nilai model parameter yang dihasilkan dalam iterasi I tahap Maximization ini dilakukan dengan cara membagi nilai frekuensi setiap fitur x dengan total frekuensi fitur. Hingga selanjutnya seluruh tahap yang dilakukan dalam proses

maximization ini dilakukan secara berulang terus menerus hingga nilai probabilitas data yang dihasilkan mencapai nilai 0 atau 1 yang artinya setiap konsumen dapat ditentukan akan termasuk kegunaan apa yang cocok data Y yang terpilih berdasarkan jumlah yang diperoleh (fitur x). Berikut adalah contoh nilai probabilitas kelas Y yang sudah mencapai nilai ambang 0 atau 1.

Tabel 10. Data probabilitas Iterasi terakhir dari data Y (Kelas hanya bernilai 0 atau 1)

	Data			Fitur			
	Y=0	Y=1	Y=2	X1	X2	X3	X4
0.00	1.00	0.00	0	1	0	1	
0.00	0.00	1.00	1	0	1	0	
1.00	0.00	0.00	0	1	1	1	
0.00	0.00	1.00	1	1	1	0	
0.00	1.00	0.00	0	1	0	0	
1.00	0.00	0.00	0	0	1	1	

Dari penelitian yang peneliti lakukan dapat disimpulkan bahwa secara statistik perminatan terhadap obat dapat dilihat berdasarkan jenis dan fungsi menggunakan algoritma expectation maximization clustering. Rekomendasi bidang yang diberikan adalah hasil maximization dari kumpulan yang mencapai nilai satu (1) atau nol (0) terhadap data Y (jenis obat). Nilai 1 mengindikasikan bahwa nilai yang diperoleh konsumen masuk ke data Y dan nilai nol mempunyai arti sebaliknya. Subbab berikut menjelaskan hasil persentasi rekomendasi suatu jenis obat kepada konsumen.

Setelah dilakukan secara statistik menggunakan algoritma Expectation-Maximization, seluruh hasil rekomendasi bidang data permintaan obat berdasarkan nilai mata kuliah maka dapat disimpulkan bahwa rekomendasi ini mampu memberikan persentasi hingga 100 % terhadap permintaan berdasarkan kegunaan obat dan jenis penyakit.

Tabel 11. Nilai probabilitas Iterasi terakhir dari kelas Y (data hanya bernilai 0 atau 1)

Data (Y)	jenis penyakit	Persentasi	Jumlah konsumen
Y=1	Analgesik, Antipiretik, Antiinflamasi nonsteroid, Antipirai	100%	350
		50%	170
		2%	50
Y=2	Anti alergi dan obat untuk Anafilaksis	100%	345
		75%	210
		1%	66
Y=3	Antidot dan Obat lain untuk Keracunan	100%	250
		98%	320
		3%	78
Y=4	Antiepilepsi – Antikonvulsi	100%	250
		98%	320
		3%	78
Y=5	Anti Infeksi	100%	250
		98%	320
		3%	78
Y=6	Antineoplastik, Imunosupresan dan obat untuk terapi paliatik	100%	345
		75%	210
		1%	66
Y=7	Obat yang mempengaruhi darah	100%	350
		50%	170
		2%	50
Y=8	Antiparkinson	100%	250
		98%	320
		3%	78
Y=9	Hormon, Obat endokrin lain dan Kontraseptik	100%	350
		2%	50
		100%	350
Y=10	Kontraseptik	50%	170
		100%	250
		98%	320
Y=11	Saluran Cerna	3%	78
		100%	345
		75%	210
Y=12	Antimigrain	1%	66
		100%	250
		98%	320
Y=13	Obat yang mempengaruhi sistim imun	3%	78
		100%	250
		98%	320
Y=14	Psikofarmaka	3%	78
		100%	250
		98%	320

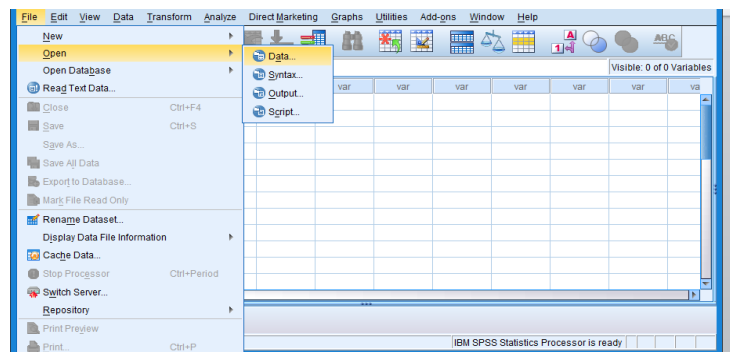
Data (Y)	jenis penyakit	Persentasi	Jumlah konsumen
Y=15	Saluran Napas	100%	250
		98%	320
		3%	78
Y=16	Relaksan Otot Perifer dan Penghambat Kolinesterase	100%	250
		98%	320
		3%	78
Y=17	Kulit, Obat Topikal	100%	345
		75%	210
		1%	66
Y=18	Sebagai obat infeksi jamur	100%	345
		75%	210
		1%	66

Tabel di atas menjelaskan hasil estimasi dari jumlah data obat sebanyak 11167 orang, di mana nilai mengestimasi terendah mencapai nilai 1 % dan nilai tertinggi 100 % serta rata-rata 67 %. Nilai 100% artinya bahwa dietimasikan suatu jenis dan fungsi obat dengan kadar 100%.

3.3 Implementasi Algoritma Expectation Maximization Pada SPSS

Dibawah ini merupakan langkah-langkah kerja pengimplementasian pada SPSS, yaitu sebagai berikut :

- Langkah pertama pembuatan format tabular pada lembar kerja *Microsoft Excel*
 Format data tabular tersebut dapat disimpan pada lembar kerja *Ms. Excel* yang menjadi *database* penyimpanan data tabulara, dengan *save as* menjadi akan dikoneksikan ke *SPSS*
- Kemudian buka *SPSS* pada laptop atau komputer yang telah terinstal .
- Untuk memulai menggunakan *SPSS* dengan memasukkan file yang ingin di proses.



Gambar 1. Tampilan data file awal

- Kemudian cari data pada folder yang ada pada *dataset*

	DataY	jenispenyakit	Persentasi	Jumlahkonsumen	va
1			100.00%	350	
2	Y=1	Analgesik, Antipiretik, Antiinflamasi nonsteroid, Antipirai	50.00%	170	
3			2.00%	50	
4		Anti alergi dan obat untuk Anafilaksis	100.00%	345	
5	Y=2		75.00%	210	
6			1.00%	66	
7		Antidot dan Obat lain untuk Keracunan	100.00%	250	
8	Y=3		98.00%	320	
9			3.00%	78	
10		Antiepilepsi - Antikonvulsi	100.00%	250	
11	Y=4		98.00%	320	
12			3.00%	78	
13		Anti Infeksi	100.00%	250	
14	Y=5		98.00%	320	
15			3.00%	78	
16		Antineoplastik, Imunosupresan dan obat untuk terapi paliatik	100.00%	345	
17	Y=6		75.00%	210	
18			1.00%	66	
19		Obat yang mempengaruhi darah	100.00%	350	
20	Y=7		50.00%	170	
21			2.00%	50	

Gambar 2. Dataset/SPSS

- Lalu tampilan selanjutnya seperti berikut ini, menjadi data editor

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	DataY	String	4	0	Data (Y)	None	None	4	Left	Nominal	Input
2	jenispenyakit	String	60	0	jenis penyakit	None	None	50	Left	Nominal	Input
3	Persentasi	Numeric	11	2		None	None	11	Right	Nominal	Input
4	Jumlahkons...	Numeric	12	0	Jumlah konsum...	None	None	12	Right	Scale	Input
5											
6											

Gambar 3. Variabel View

6. Pilih file yang diinginkan

	DataY	jenispenyakit	Persentasi	Jumlahkonsumen	var
1	Y=18		100.00%	350	
2	Y=1	Analgesik, Antipiretik, Antiinflamasi nonsteroid, Antipirai	50.00%	170	
3			2.00%	50	
4		Anti alergi dan obat untuk Anafilaksis	100.00%	345	
5	Y=2		75.00%	210	
6			1.00%	66	
7		Antidot dan Obat lain untuk Keracunan	100.00%	250	

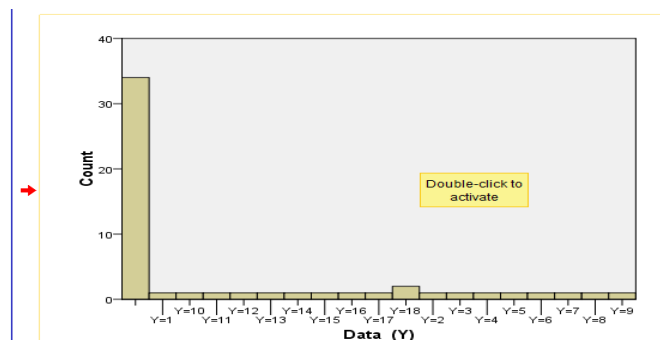
Gambar 4. Dataset/SPSS

7. Data aktivasi bulan

jenis penyakit	Count
Analgesik, Antipiretik, Antiinflamasi nonsteroid, Antipirai	1
Anti alergi dan obat untuk Anafilaksis	1
Anti infeksi	1
Antidot dan Obat lain untuk Keracunan	1
Antiepilepsi - Antikonvulsi	1
Antimigrain	1
Antineoplastik, Imunosupresan dan obat untuk terapi paliatif	1
Antiparkinson	1
Hormon, Obat endokrin lain dan Kontraseptik	2
Kulit, Obat Topikal	1
Obat yang mempengaruhi darah	1
Obat yang mempengaruhi sistem imun	1
Psikofarmaka	1
Relaksan Otot Perifer dan Penghambat Kolinesterase	1
Saluran Cerna	1
Saluran Napas	1
Sehingga obat infeksi	1

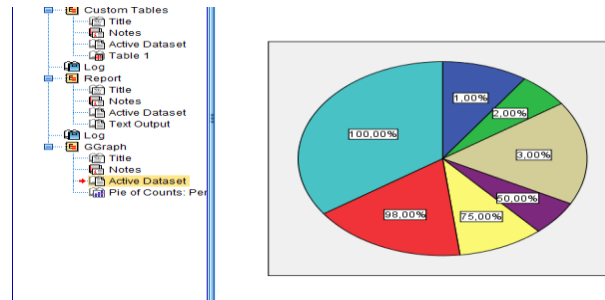
Gambar 5. Data Aktivasi Bulan

8. Tabel diagram data konsumen



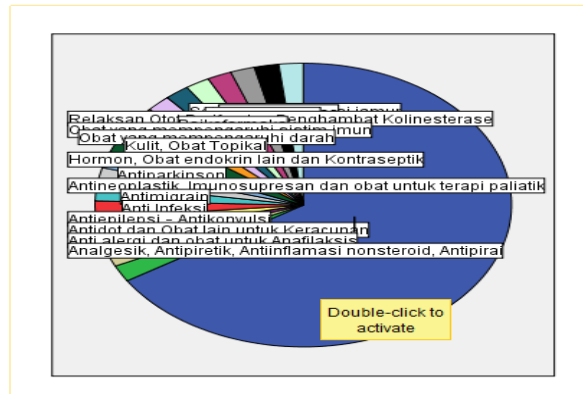
Gambar 6. Tabel Diagram Diagram Data Konsumen

9. Grafik Aktif peminat konsumen terhadap obat.



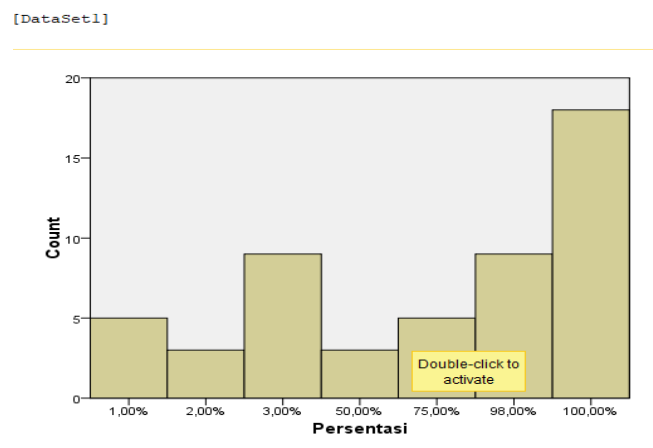
Gambar 7. Open Dataset/Pilih salah satu

10. data peminta obat yang lebih muncul.



Gambar 8. Data peminat obat yang lebih muncul

11. Data hasil tabel persentasi konsumen obat.



Gambar 9. Data Persentasi Konsumen Obat

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka penulis mengambil kesimpulan bahwa :

1. Algoritma *Expectation Maximization* dapat memiliki peranan baik pada data yang ada pada tahun-tahun sebelumnya.
2. Algoritma yang di gunakan untuk estimasi penjualan obat adalah Algoritma *Expectation Maximization*.
3. Hasil estimasi dalam pengujian SPSS adalah dalam bentuk table/grafik.

REFERENCES

- [1] Buulolo, "Implementasi Algoritma Apriori Pada Sistem Persediaan Obat (Studi Kasus : Apotik Rumah Sakit Estomihi Medan)", 2013.
- [2] Kusri dan E.T. Luthfi, Algoritma Data Mining, Yogyakarta : Andi. 2009
- [3] D.Sulistiyo, M.A. Bijaksana dan D. Darmantoro, "Data Mining Dengan Algoritma Apriori pada RDBMS ORACLE", 2003.
- [4] Budi Santosa, Data Mining, Teknik Pemanfaatan Data Untuk Keperluan Bisnis, Yogyakarta : Graha Ilmu. 2007.
- [5] Harinaldi, Prinsip-Prinsip Statistik Untuk Teknik Dan Sains, Jakarta : Erlangga. 2005
- [6] Anthony, Andeka Rocky Tanamaah, Agustinus Fritz Wijaya "Penjualan", Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK) Vol.4 No.2, Juni 2017, hlm. 136-147.

- [7] Eric Suhardi, “Pengaruh *Brand Image* Terhadap Keputusan Konsumen Untuk Membeli Pada Apotik Kharitas Bhakti Pontianak ”, Bisma, vol 1.No.5, September 2016.
- [8] Tomy Angga Kusuma,“ Algoritma *Expectation Maximization(Em)* Untuk Estimasi Distribusi Mixture“, Jurnal Konvergensi, vol. 4, No. , Oktober 2014.
- [9] Jogiyanto HM, Analisis & Desain, Yogyakarta:ANDI. 2005.
- [10] Booch, The Unifi ed Modelling Language User Guide Second Edition,United State:Addison Wesley Professional. 2005.
- [11] Hendery, Unified Modelling Language, Tangerang:Raharja Enrichment Centre(REC).2008.
- [12] Adi Nugroho, Rekayasa Perangkat Lunak berbasis Objek Dengan Metode USDP, Yogyakarta:ANDI. 2010.
- [13] Wahana Komputer, Analisis Data Penelitian Dengan SPSS 22, Yogyakarta:ANDI.2014.