

Prediksi Penjualan Obat Dengan Algoritma Regresi Linear

Prediction of Drug Sales With Linear Regression Algorithm

Dichi Setiawan¹, Nurhadi Surojudin², Wahyu Hadikristanto³

^{1,2,3}Program Studi Teknik Informatika, Universitas Pelita Bangsa

¹dichisetiawan@mhs.pelitabangsa.ac.id*, ²nurhadi@pelitabangsa.ac.id,

³wahyu.hadikristanto@pelitabangsa.ac.id

Abstract

Medicine is one part of the process of health services, such as hospitals, clinics or other health services. Linear regression is used to find out how the dependent variable can be predicted through independent variables or predictor variables. Prediction of total future drug sales per month can be known through the use of drug sales datasets through a predictive method approach through structured stages in analyzing the data used. The application of a simple Linear Regression algorithm model can be implemented properly and is able to show a new insight for the prediction needs of the condition of drug sales per month in the future for the next 12 months. Performance testing of the models and algorithms used in the evaluation can produce an overview that is relevant to the scenario being modeled. The RMSE value obtained when evaluating the model's performance is 0.452 +/- 0.000 through the RapidMiner Studio application.

Keywords: Data Mining, Linear Regression, Medicine

Abstrak

Obat merupakan salah satu bagian proses pelayanan kesehatan, seperti rumah sakit, klinik ataupun pelayanan kesehatan lainnya. Regresi Linear digunakan untuk mengetahui bagaimana variabel dependen dapat diprediksikan melalui variabel independen atau variabel prediktor. Prediksi total penjualan obat dimasa mendatang per bulan dapat diketahui melalui pemanfaatan dataset penjualan obat dengan melalui pendekatan metode prediksi melalui tahapan yang terstruktur dalam menganalisis data yang digunakan. Penerapan model algoritma Regresi Linear sederhana dapat diimplementasikan secara baik serta mampu menunjukkan sebuah wawasan baru bagi kebutuhan prediksi terhadap kondisi penjualan obat per bulan di masa yang akan datang selama 12 bulan kedepan. Pengujian performa terhadap model dan algoritma yang digunakan dalam evaluasi dapat menghasilkan gambaran yang relevan dengan skenario yang dimodelkan. Nilai RMSE didapat saat melakukan evaluasi performa model sebesar 0.452 +/- 0.000 melalui aplikasi RapidMiner Studio.

Kata kunci: Data Mining, Regresi Linear, Obat

Pendahuluan

Obat merupakan salah satu bagian proses pelayanan kesehatan, seperti rumah sakit, klinik ataupun pelayanan kesehatan lainnya[1]. Pengelolaan obat merupakan salah satu indikator mutu pelayanan dalam bidang kesehatan dan kefarmasian. proses pengelolaan di mulai dari seleksi, kontribusi, pengadaan terakhir adalah penggunaan obat. standar pengelolaan obat yang baik dan sesuai[2]. Bila sasaran pembelian obat tidak ditentukan secara baik, dalam arti tidak diupayakan mencari sasaran pembelian obat yang potensial, maka hanya akan menghabiskan banyak waktu dan biaya yang seharusnya bisa diminimalisir melalui pemilihan target pembelian obat yang baik[3].

Salah satu cara yang dapat diterapkan dengan menerapkan penggunaan data mining, untuk mengetahui keefektifan obat yang sering digunakan dan data mining merupakan salah satu disiplin ilmu yang dapat digunakan untuk menggali informasi dari kumpulan data obat kondisi tersebut yang berbentuk data sehingga dapat diambil intisari dari kumpulan data obat tersebut. Sekumpulan data obat yang diolah secara komputerisasi dapat membantu proses *Root Mean Square Error (Rmse)* dalam pemilihan dan prediksi obat yang tepat. *Regresi linear* digunakan untuk mengetahui bagaimana variabel dapat diprediksikan melalui variabel independen atau variabel prediktor, secara individual[4]. Dampak dari penggunaan analisis regresi

dapat digunakan untuk memutuskan apakah naik dan menurunnya variabel dependen dapat dilakukan melalui menaikkan dan menurunkan keadaan variabel independen, atau meningkatkan keadaan variabel dependen dapat dilakukan dengan meningkatkan variabel independen dan sebaliknya.

Data mining adalah kegiatan mengekstrak informasi atau pengetahuan (*knowledge*) penting dari suatu set data berukuran besar dengan menggunakan teknik tertentu[5]. Informasi atau *knowledge* yang dihasilkan dari *data mining* bisa dipakai untuk memperbaiki pengambilan keputusan. Dinamakan *data mining* atau penambangan data karena proses penemuan informasi atau *knowledge* dalam set data dilakukan seperti orang melakukan kegiatan penambangan. algoritma *linear regression* Linear Regression (*Regresi linear*) mencoba untuk memodelkan hubungan antara dua variabel dengan memasang persamaan linear untuk mengamati data[6]. Linear Regression berusaha menemukan hubungan matematis antar variabel. Didasari oleh uraian diatas maka peneliti mengangkat tema dengan judul “Prediksi Penjualan Obat Dengan Algoritma Regrsi Linier”

Metode Penelitian

Kajian Jurnal Pertama

Analisa Perbandingan Prediksi Obat Dengan Menggunakan Metode ABC Analisis Dan SVR[7]. Permasalahan rumah sakit mengenai pengolahan persediaan obat adalah kondisi obat yang habis, dikarenakan rumah sakit dalam satu tahun mengeluarkan 33% dari biaya investasi untuk investasi obat. Untuk mengatasi permasalahan diatas rumah sakit harus memiliki pengelolaan logistic dengan baik, cara pengelolaan adalah dengan melakukan perencanaan yang baik. Penulis akan memakai algoritma ABC Analysis dan SVR. ABC Analysis akan digunakan untuk proses klasifikasi obat yaitu dengan cara membagi obat menjadi tiga kelompok utama berdasarkan kepentingan yaitu kelompok A, B dan C. Penulis akan menggunakan metodo SVR untuk menghitung prediksi obat. Hasil penelitian ini adalah ABC analisis dapat membagi obat. Menjadi tiga bagian yaitu kelompok A 276 item dengan presentase 22,96% dari jumlah item keseluruhan, kelompok B sejumlah 396 item dengan presentase 33,11% dan C sejumlah 528 dengan presentase 43,94% dengan keseluruhan obat adalah 1202 item obat. Pengujian prediksi dilakukan dengan cara mengambil sample lima obat dari hasil klasifikasi. Proses perhitungan SVR adalah membandingkan metode preprocessing linier scaling dan z normalization. Hasil dari penelitian tersebut adalah MAPE menunjukkan bahwa dengan menambahkan preprocessing dengan linier scaling dapat membuat proses SVR lebih optimal dari pada menggunakan z normalization dan perhitungan dengan klasifikasi ABC analisis.

Kajian Jurnal Kedua

Prediksi Biaya Manfaat Kesehatan Pengobatan Rawat Jalan Karyawan Dengan Metode Data Mining Menggunakan Algoritma Regresi Linier [8]. Perencanaan pengelolaan anggaran menjadi hal sangat penting demi keberlangsungan bisnis dan operasional sebuah organisasi atau perusahaan. Efektifitas dan efisiensi dalam penggunaan biaya operasional maupun non operasional sudah menjadi perhatian khusus bagi manajemen untuk menjaga kestabilan alur kas keuangan. Biaya kesehatan bagi karyawan perusahaan merupakan salah satu anggaran yang perlu dialokasi dan dikelola demi kesejahteraan karyawan dalam rangka menjaga peningkatan kinerja karyawan. Salah satu biaya kesehatan adalah manfaat pengobatan rawat jalan bagi karyawan yang perlu dianggarkan tiap tahunnya. Dukungan data, informasi dan pengetahuan menjadi faktor penting dalam perencanaan yang dilakukan dalam penganggaran biaya kesehatan. Salah satu faktor yang menentukan dalam membuat rencana biaya kesehatan adalah berapa besar biaya pengobatan yang dibutuhkan untuk biaya pengobatan kesehatan rawat jalan. Untuk mendukung data, informasi dan pengetahuan berkaitan dengan perkiraan anggaran biaya kesehatan karyawan tersebut maka ingin didapatkan prediksi biaya manfaat pengobatan rawat jalan dalam satu tahun kedepan. Pada penelitian ini dibuat proses Data Mining dengan menggunakan Metode Regresi Linier untuk mendapatkan berapa total biaya dari manfaat pengobatan rawat jalan untuk biaya obat, dokter umum, dokter spesialis, lab diagnostik dan fisioterapi. Jumlah peserta menjadi variabel predictor sedangkan total biaya manfaat pengobatan rawat jalan menjadi variabel response. Dari hasil perhitungan didapat rumus persamaan yaitu $Y = -125623110 + 383795X$, untuk mendapatkan prediksi biaya obat, $Y = -8561735 + 29796 X$, untuk mendapatkan prediksi biaya dokter umum, $Y = -218247706 + 164818X$, untuk mendapatkan prediksi dokter spesialis, $Y = -2858569 + 200340X$, untuk mendapatkan prediksi

biaya lab diagnostik, serta $Y = -10912062 + 30500X$, untuk mendapatkan prediksi biaya fisioterapi.

Kajian Jurnal Ketiga

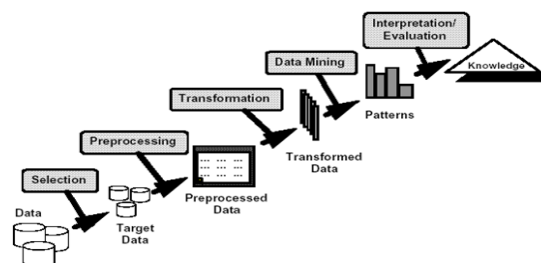
Data Mining Menggunakan Algoritma Linear Regression Berganda untuk Pengelolaan Presediaan[9]. Salah satu tugas dari bidang Kefarmasian Puskesmas adalah pengelolaan persediaan stok obat. Pada umumnya metode pengelolaan persediaan stok obat hanya bersifat standar melihat stok akan habis baru akan ditambah stoknya. Hal ini mengakibatkan kekhawatiran sendiri bagi pihak Puskesmas Sekaran didalam menjalankan pengelolaan stok obat. Oleh karena itu penelitian ini bertujuan untuk menyelesaikan masalah diatas dengan merancang aplikasi Data Mining yang berfungsi untuk memprediksi tingkat kenaikan dan penurunan suatu obat didalam stok obat pada tahun 2016. Teknik Data Mining yang diterapkan ialah Estimation, sedangkan Algoritma yang dipakai ialah Linear Regression Berganda. Adapun data yang dipakai didalam penelitian ini adalah Data Laporan Pemakaian Dan Lembar Permintaan Obat (LPLPO) tahun 2011-2015. Hasil Penelitian ini adalah pasien yang berobat di Puskesmas Sekaran sangat membutuhkan obat. Hasil dari perhitungan RMSE adalah 10340.5279, hasil tersebut masih tergolong cukup besar karena contoh data yang digunakan hanya 30 data, jika data yang digunakan lebih banyak dan dengan variabel yang lebih banyak maka standar error kemungkinan akan mengecil.

Landasan Teori

Data Mining

Basis data adalah media untuk menyimpan data agar dapat diakses dengan mudah dan cepat. Basis data adalah kumpulan data yang saling berikatan dan berhubungan satu dengan yang lainnya, tersimpan di perangkat keras komputer dan perangkat lunak digunakan untuk memanipulasinya. *Data mining* merupakan proses pengambilan data dari data *warehouse* berdasarkan prediksi variabel. *Data mining* juga sebagai penemuan pada KDD *database*[10]. Prediksi digunakan untuk menemukan data dari sekumpulan data yang ditentukan dengan berbagai *domain* seperti kecerdasan buatan, basis data, dan lain-lain. Penggunaan *data mining* memiliki dampak diberbagai bidang seperti bidang kesehatan. *Data mining* pada sistem medis dibutuhkan untuk mengekstrak informasi dari *database* sehingga dapat melakukan diagnosis penyakit [11].

Data Mining adalah suatu proses penemuan korelasi, pola dan tren baru yang bermakna dengan menyaring sejumlah besar data yang disimpan dalam repositori, menggunakan teknologi pengenalan pola serta statistik dan teknik matematika[12]. *Data Mining* merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. *Data mining* merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [13]. Dalam melakukan prediksi data obat yang akan di uji sesuai pemodelan data yang akan digunakan agar mempermudah penelitian dan berjalan sesuai diinginkan maka dibuat alur atau tahapan dalam penelitian ini sebagai berikut:



Gambar 1 Tahapan Penelitian

Obat

Obat adalah zat apa pun yang menyebabkan perubahan *fisiologi* atau *psikologi organisme* saat dikonsumsi. Obat-obatan biasanya dibedakan dari makanan dan zat yang menyediakan nutrisi. Konsumsi obat dapat dilakukan melalui *inhalasi*, *injeksi*, merokok, *ingesti*, *absorpsi* melalui kulit, atau disolusi di bawah lidah[14].

Dalam *farmakologi*, obat adalah zat kimia, biasanya struktur kimianya diketahui, yang ketika diberikan pada *organisme* hidup akan menghasilkan efek biologis. Obat farmasi, juga disebut medikasi atau obat dalam pemahaman masyarakat umum, adalah zat kimia yang digunakan untuk mengobati, menyembuhkan,

mencegah, atau mendiagnosis suatu penyakit. Secara tradisional, obat-obatan diperoleh melalui ekstraksi tumbuhan obat, tetapi baru-baru ini juga melalui sintesis organik[15].

Objek Penelitian

Pada penelitian ini, tahapan yang akan digunakan dalam melakukan prediksi terhadap data obat dan penentuan atribut pasien atau pemeriksaan untuk mempermudah penelitian sehingga penelitian dapat berjalan dengan baik dan sistematis, serta memenuhi tujuan yang diinginkan. Penelitian deskriptif ditujukan untuk mendeskripsikan suatu keadaan atau fenomena-fenomena apa adanya.

Kategori penelitian deskriptif merupakan kategori yang digunakan untuk menggambarkan atau menganalisis suatu hasil penelitian tetapi tidak digunakan untuk membuat kesimpulan yang lebih luas. Prediksi obat adalah kategori penelitian yang disusun untuk meneliti kemungkinan adanya sebab dan akibat antar variable.

Tahapan Penelitian

Dalam melakukan prediksi data obat yang akan di uji sesuai pemodelan data yang akan digunakan agar mempermudah penelitian dan berjalan sesuai diinginkan maka dibuat alur atau tahapan dalam penelitian ini sebagai berikut:

Pengumpulan Data

Pada tahap ini menjelaskan tentang bagaimana dan dari mana sumber data didapatkan, diantaranya adalah:

Sumber Data Primer

Data-data dalam penelitian ini cantumkan disini merupakan data-data yang terdapat dari berbagai macam internet. Dalam mempermudah penelitian dan berjalan sistematis dan memenuhi tujuan yang diinginkan maka dibuat langkah-langkah dalam tahapan penelitian yang akan dilakukan dan menjelaskan tentang bagaimana sumber data yang didapatkan dalam penelitian ini diantaranya data obat dan studi Pustaka untuk memperkuat bahan penelitian sebagai representasi teori.

Data Cleansing

Proses pembersihan dari data yang akan dipakai untuk penghapusan data dengan membuang *missing/ null value*, duplikasi data maupun untuk memeriksa *inkonsistensi* data dan memperbaiki kesalahan pada data. Proses pembersihan data dilakukan secara manual. Pada data ini dilakukan pembersihan data dengan menghapus *record* data untuk bagian yang memiliki *missing/ null value*.

Tabel 1 Contoh Data Cleansing

No. Registrasi	Nama Pasien	Date of Birth	Umur	Diagnosis Utama
11111	(Blank)	09/10/1995	23	Acute nasopharyngitis [common cold]
11111	(Blank)	07/01/2018	0,2	(Blank)
11111	(Blank)	09/11/1993	25	Headache
10601409	Aditya Saputra Ramadhan	05/04/1989	29	(Blank)
10632383	Yadi Setiadi	10/12/1985	32	(Blank)
10285461	Dian Heryanto	21-09-1982	36	(Blank)
10639190	Supriyadi	10/05/1992	25	(Blank)
10516566	Taofik Hidayat	15-06-1985	33	(Blank)

Data Selection

Tahap ini akan dilakukan pemilihan data dengan seleksi atribut yang diperlukan ke proses selanjutnya, sehingga atribut atribut yang tidak memiliki keterkaitan terhadap tahapan penerapan model nantinya akan direduksi sehingga dimensi data menjadi lebih ringkas.

Tabel 2 Data Selection (Atribut awal pada dataset)

Nama Atribut	Tipe	Nama Atribut	Tipe
Tgl Input Data	Date	Obat 4	Text
Tgl Berobat	Date	Jumlah 4	Numerik
No. Registrasi	Numerik	Obat 5	Text
Nama Pasien	Text	Jumlah 5	Numerik
<i>Date of Birth</i>	Date	Obat 6	Text
Umur	Numerik	Jumlah 6	Numerik
Diagnosis Utama	Text	Obat 7	Text
Diagnosa ICD Utama	Text	Jumlah 7	Numerik
Diagnosis Tambahan	Text	Obat 8	Text
Diagnosa ICD Tambahan	Text	Jumlah 8	Numerik
Diagnosis Tambahan	Text	Obat 9	Text
Diagnosa ICD Tambahan	Text	Jumlah 9	Numerik
Keadaan Umum	Text	Obat 10	Text
Kesadaran Pasien	Text	Jumlah 10	Numerik
Tekanan Darah Sistolik	Numerik	Resep 1	Text
Tekanan Darah Diastolik	Numerik	Jumlah 1	Numerik
Nadi	Numerik	Resep 2	Text
Suhu	Numerik	Jumlah 2	Numerik
Pernafasan	Numerik	Resep 3	Text
Tinggi Badan	Numerik	Jumlah 3	Numerik
Berat Badan	Numerik	Resep 4	Text
Keluhan Utama/ <i>Anamnesa</i>	Text	Jumlah 4	Numerik
Dokter	Text	Biaya Dokter	Numerik
Obat 1	Text	Biaya Obat	Numerik
Jumlah 1	Numerik	Biaya Lab/NonLab	Numerik
Obat 2	Text	Biaya Total	Numerik
Jumlah 2	Numerik	Biaya Total Manual	Numerik
Obat 3	Text	Grand Total	Numerik
Jumlah 3	Numerik		

Adapun pada dataset awal terdapat 57 atribut, dan pada proses seleksi atribut ini beberapa atribut yang tidak dibutuhkan dalam tahap pemodelan akan direduksi dan hanya ada 11 atribut yang akan dipakai diantaranya atribut Tgl Input, Jumlah 1, Jumlah 2, Jumlah 3, Jumlah 4, Jumlah 5, Jumlah 6, Jumlah 7, Jumlah 8, Jumlah 9, dan Jumlah 10, seperti terlihat pada Tabel 3 berikut :

Tabel 3 Hasil Seleksi Atribut Pada Dataset

Nama Atribut	Tipe	Nama Atribut	Tipe
Tgl Input	Date	Jumlah 6	Text
Jumlah 1	Text	Jumlah 7	Text
Jumlah 2	Text	Jumlah 8	Text
Jumlah 3	Text	Jumlah 9	Text
Jumlah 4	Text	Jumlah 10	Text
Jumlah 5	Text		

Data Transformation

Proses ini merupakan mengubah skala pengukuran data asli yang dimana data akan ditransformasikan ke dalam bentuk lain sehingga nantinya akan memudahkan dalam proses pembacaan data dengan algoritma pada program maupun *tools* yang digunakan. Salah satu proses yang ditambahkan yakni *feature engineering* dengan *feature creation*. Hal ini dilakukan dengan menambahkan 2 kolom baru untuk menampung *value* bulan transaksi dan *value* total jumlah penggunaan obat yang nantinya akan membantu dalam proses pemodelan. Kolom bulan digunakan dalam proses agregasi data dalam menjumlahkan total pada transaksi pada setiap bulannya.

Tabel 4 Feature Engineering menambahkan kolom bulan pada dataset

Tgl Input	Bulan
17/07/2021	Juli
21/07/2021	Juli
16/07/2021	Juli
07/02/2021	Februari
24/07/2021	Juli
07/12/2021	Desember
07/05/2021	Mei

Pemodelan

Pemodelan pada penelitian ini dilakukan dengan data mining teknik prediksi menggunakan algoritma regresi linear. Regresi Linear digunakan untuk mengetahui bagaimana variabel dependen/kriteria dapat diprediksikan melalui variabel independen atau variabel prediktor, secara individual. Dampak dari penggunaan analisis regresi dapat digunakan untuk memutuskan apakah naik dan menurunnya variabel dependen dapat dilakukan melalui menaikkan dan menurunkan keadaan variabel independen, atau meningkatkan keadaan variabel dependen dapat dilakukan dengan meningkatkan variabel *independen*/dan sebaliknya. Analisis regresi ganda digunakan oleh peneliti, bila peneliti bermaksud meramalkan bagaimana keadaan (naik turunnya) variabel dependen (kriterium), bila dua atau lebih variabel independen sebagai prediktor dimanipulasi (dinaik turunkan nilainya). Jadi analisis regresi ganda akan dilakukan bila jumlah variabel independennya minimal 2.

Perhitungan Nilai XX dan XY

Perhitungan terhadap Nilai XX, dan XY dilakukan pada data latih (*training*) untuk mendapatkan model persamaan regresi linier yang akan diterapkan dalam memprediksi stok barang terhadap data uji (*testing*). Perhitungan dilakukan antara lain dengan mendapatkan nilai XY yang merupakan perkalian dari kedua variabel tersebut kemudian nilai XX yang merupakan nilai variabel X yang dikuadratkan. Variabel X merupakan variabel yang mewakili atribut Kode Bulan pada *dataset training* yang digunakan sementara variabel Y merupakan atribut Total.

Perhitungan Koefisien A dan Koefisien B

Perhitungan koefisien A dan B yang dilakukan menggunakan rumus di bawah berikut : Koefisien A :
 Koefisien B :

$$\frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$\frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Metode Kuadrat Terkecil (*Least Square Method*) merupakan metode yang paling populer untuk menetapkan persamaan regresi linear sederhana.

Model sesuai dengan menghitung persamaan regresi linear menggunakan rumus berikut :

$$Y = a + bX$$

Dimana :

- Y : variabel terikat
- a : koefisien regresi *intersep* (konstanta)
- b : koefisien regresi *slop* (kemiringan)
- X : variabel bebas

Evaluasi dan Hasil

Evaluasi dilakukan dengan maksud mengetahui hasil perhitungan yang dianalisa dan mengukur metode serta algoritma regresi linier yang digunakan apakah berfungsi dengan baik atau tidak. Proses pengujian menggunakan tools *repidminer* dan melihat data apakah sesuai dengan hasil yang diperoleh melalui tools

tersebut, mengetahui bagaimana variabel *dependen/kriteria* dapat diprediksikan melalui variabel independen atau variabel prediktor, secara individual. Dampak dari penggunaan analisis regresi dapat digunakan untuk memutuskan apakah naik dan menurunnya variabel dependen dapat dilakukan melalui menaikkan dan menurunkan keadaan variabel independen, atau meningkatkan keadaan variabel dependen dapat dilakukan dengan meningkatkan variabel independen/dan sebaliknya Metode Kuadrat terkecil (*least square method*).

Root Mean Squared Error (RMSE) merupakan salah satu cara untuk mengevaluasi model regresi linear dengan mengukur tingkat akurasi hasil perkiraan suatu model. RMSE dihitung dengan mengkuadratkan *error* (prediksi – observasi) dibagi dengan jumlah data (= rata-rata), lalu diakarkan. RMSE tidak memiliki satuan. Secara matematis, rumusnya ditulis sebagai berikut :

$$RMSE = \left(\frac{\sum (y_i - \hat{y}_i)^2}{n} \right)^{1/2}$$

Keterangan :

RMSE = nilai *root mean square error*

Y = Nilai Hasil Observasi

\hat{y} = Nilai Hasil Prediksi

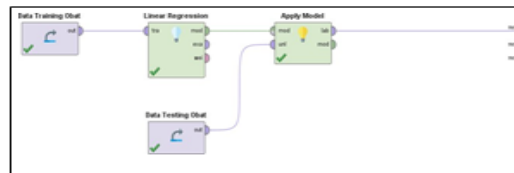
i = Urutan Data Pada Database

n = Jumlah Data

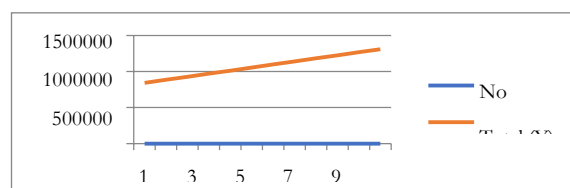
Hasil dan Pembahasan

Dalam penelitian ini penulis memanfaatkan aplikasi RapidMiner Studio dalam melakukan evaluasi terhadap data uji (*testing*) yang digunakan untuk menilai model persamaan algoritma Regresi Linier sederhana yang diterapkan dalam pembentukan prediksi dari suatu objek baru dengan keakurasian yang tepat.

Diagram proses seluruh operator fungsi dari desain yang sudah disiapkan pada aplikasi RapidMiner Studio akan membentuk alur proses seperti pada gambar 2 berikut :



Gambar 2 Implementasi Algoritma Regresi Linear pada aplikasi RapidMiner Studio.



Gambar 3 Grafik prediksi data uji perhitungan

Pengujian Performa Regresi Linear

Secara umum persamaan rumus matematis dalam menghitung nilai RMSE dapat dilihat pada persamaan berikut ini:

$$RMSE = \left(\frac{\sum (y_i - \hat{y}_i)^2}{n} \right)^{1/2}$$

Dimana keterangannya ialah sebagai berikut:

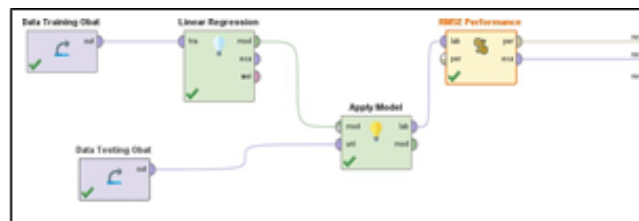
RMSE = nilai *Root Mean Square Error*

y = nilai hasil observasi

\hat{y} = nilai hasil prediksi
 i = urutan data pada list data
 n = jumlah data

Nilai RMSE (*Root Mean Square Error*) rendah menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model prakiraan mendekati variasi nilai obeservasinya. RMSE (*Root Mean Square Error*) menghitung seberapa berbedanya seperangkat nilai dari suatu objek hasil prediksi. Semakin kecil nilai RMSE (*Root Mean Square Error*), semakin dekat nilai yang diprediksi dan diamati.

Berikut merupakan diagram proses pada aplikasi RapidMiner Studio dalam mencari nilai RMSE (*Root Mean Square Error*) dari model yang diterapkan.



Gambar 4 Proses uji performa menggunakan aplikasi RapidMiner Studio

Dengan menerapkan model tersebut melalui aplikasi RapidMiner Studio menghasilkan angka RMSE sebesar **0.452 +/- 0.000**, nilai yang cukup kecil sehingga hanya kecil perbedaan yang terjadi dari proses simulasi perhitungan manual dengan aplikasi RapidMiner Studio terhadap data uji (*testing*) yang disiapkan.

Gambar 5 Nilai RMSE (*Root Mean Square Error*) pada aplikasi RapidMiner Studio

Dalam memprediksi total penggunaan obat selama 1 tahun kedepan (12 bulan) maka persamaan model Regresi Linear yang didapat kemudian secara umum diterapkan pada sejumlah data uji (*testing*) yakni sebanyak 12 baris data seperti pada tabel dibawah ini.

Table 5 Dataset Uji (Data Testing)

No.	Kode (Variabel X)	Total (Y <i>predict</i>)
1	13	?
2	14	?
3	15	?
4	16	?
5	17	?
6	18	?
7	19	?
8	20	?
9	21	?
10	22	?
11	23	?
12	24	?

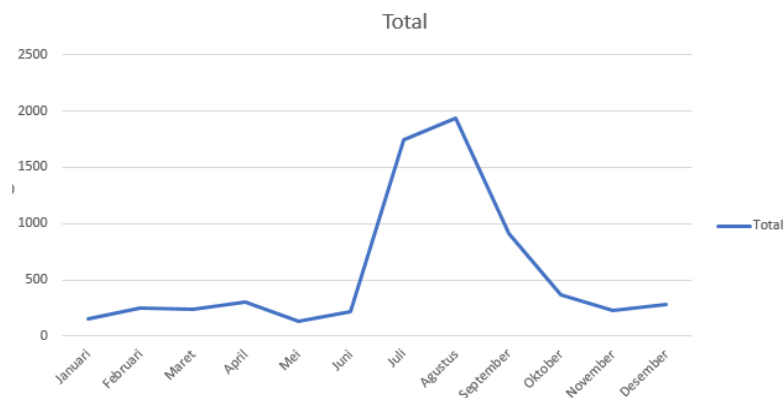
Dibawah ini adalah tabel perbandingan antara hasil perhitungan manual dan yang ada pada aplikasi Rapid Miner.

Tabel 6 Perbandingan Antara Hasil Hitung Manual dan RapidMiner

No.	Kode	Prediksi Y	Prediksi Y
-----	------	------------	------------

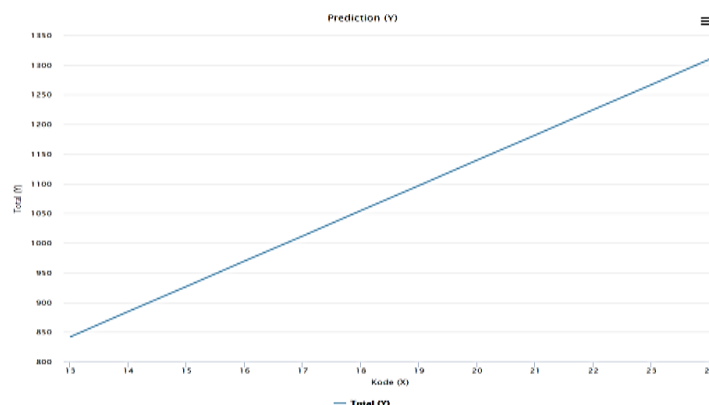
	(Variabel X)	(Manual)	(Rapid Miner)
1	13	842,000	841,894
2	14	885,000	884,390
3	15	927,000	926,887
4	16	970,000	969,383
5	17	1012,000	1011,880
6	18	1055,000	1054,376
7	19	1097,000	1096,873
8	20	1140,000	1139,369
9	21	1182,000	1181,866
10	22	1225,000	1224,362
11	23	1267,000	1266,859
12	24	1310,000	1309,355

Pada data latih secara umum pergerakan penjualan obat selama per bulan cenderung fluktuatif dan mengalami lonjakan peningkatan pada bulan ke 7 dan 8 (data ke-7 dan data ke-8) seperti terlihat pada grafik dibawah ini.



Gambar 6 Grafik Penggunaan Obat Per Bulan Pada Data Latih

Namun secara umum terlihat hasil prediksi untuk total penjualan obat dari data uji yang digunakan dalam pemodelan semakin jauh semakin bertambah jumlahnya. Hal ini juga dapat dilihat pada grafik berikut dimana dari bulan ke bulan total penjualan obat diprediksi cenderung mengalami peningkatan.



Gambar 7 Grafik prediksi data uji pada aplikasi RapidMiner Studio

Dalam pengujian performa model juga dapat dilihat hasil simulasi prediksi menggunakan model Regresi Linear tersebut memiliki nilai RMSE (*Root Mean Square Error*) melalui aplikasi RapidMiner Studio dan menghasilkan angka RMSE sebesar **0.452 +/- 0.000**, nilai yang cukup kecil sehingga hanya kecil perbedaan yang terjadi dari proses simulasi perhitungan manual dengan aplikasi RapidMiner Studio terhadap data uji (*testing*) yang disiapkan.

```

PerformanceVector
PerformanceVector:
root_mean_squared_error: 0.452 +/- 0.000
  
```

Gambar 8 Nilai evaluasi performa model Regresi Linear berdasarkan perhitungan RMSE pada aplikasi RapidMiner Studio

Kesimpulan

Berdasarkan hasil yang penelitian yang dilakukan oleh penulis, kesimpulan yang didapat yaitu sebagai berikut :

1. Prediksi total penjualan obat dimasa mendatang per bulan dapat diketahui melalui pemanfaatan dataset penjualan obat dengan melalui pendekatan metode prediksi melalui tahapan yang terstruktur dalam menganalisis data yang digunakan.
2. Penerapan model algoritma Regresi Linear sederhana dapat diimplementasikan secara baik serta mampu menunjukkan sebuah wawasan baru bagi kebutuhan prediksi terhadap kondisi penjualan obat per bulan di masa yang akan datang selama 12 bulan kedepan.
3. Memproduksi jumlah obat generik dari hasil pengujian yang di peroleh dengan tujuan menghasilkan gambaran produksi yang relevan.

Daftar Pustaka

- [1] Amalia, R., & Huda, N. Implementasi Sistem Informasi Pelayanan Kesehatan Pada Klinik Smart Medica. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 9(3), 332-338. 2020.
- [2] Aisah, N., Satibi, S., & Suryawati, S. Evaluasi Pengelolaan Obat pada Tahap Perencanaan dan Pengadaan di Dinas Kesehatan Kabupaten Pati. *Majalah Farmasentik*, 16(1), 34-42. 2020.
- [3] Yanto, R., & Khoiriah, R. Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat. *Creative Information Technology Journal*, 2(2), 102-113. 2015.
- [4] H.W. Herwanto, T. Widiyaningtyas, and P. Indriana, "Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 4, p. 364, 2019, doi: 10.22146/jnteti.v8i4.537.
- [5] Doni, B. T. R., Susanti, S., & Mubarok, A. Penerapan Data Mining Untuk Klasifikasi Penyakit Hepatocellular Carcinoma Menggunakan Algoritma Naïve Bayes. *Jurnal Responsif: Riset Sains dan Informatika*, 3(1), 12-19. 2021.
- [6] Jollyta, D., Ramdhan, W., & Zarlis, M. Konsep Data Mining Dan Penerapan. *Deepublish*. 2020.
- [7] A. Puspitawuri, E. Santoso, and C. Dewi, "Diagnosis Tingkat Risiko Penyakit Stroke Menggunakan Metode K-Nearest Neighbor dan Naïve Bayes," vol. 3, no. 4, pp. 3319–3324, 2019.
- [8] A. Byna and M. Basit, "Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 3, pp. 407–411, 2020, doi: 10.32736/sisfokom.v9i3.1023.
- [9] Gunadi, G., & Sensuse, D. I. Penerapan metode data mining market basket analysis terhadap data penjualan produk buku dengan menggunakan algoritma apriori dan frequent pattern growth (fp-growth): studi kasus percetakan pt. Gramedia. *Telematika MKOM*, 4(1), 118-132. 2016.
- [10] U. Dharmawangsa, "Artificial Neural Network Dalam Mengidentifikasi Penyakit Stroke Menggunakan Metode Backpropagation (Studi Kasus di Klinik Apotik Madya Padang)," *J. Antares, T. Infomasi*, vol. 1, no. 1, pp. 6–14, 2020.
- [11] Hernawati, K., Insani, N., Sumarno, B., & Waryanto, N. Analysis Of Relation Term Between Cummulative Performance Index And Snmptn Data Base Using Appriority Algorithm. *Jurnal Sains Dasar*, 4(2), 152-163. 2015.
- [12] A. J. P. Sibarani, "Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 262–276, 2020, doi: 10.35957/jatisi.v7i2.195.
- [13] K. Algoritma, C. Berbasis, P. S. O. Dan, R. S. Rohman, R. A. Saputra, and D. A. Firmansaha, "Komparasi algoritma c4.5 berbasis pso dan ga untuk diagnosa penyakit stroke," vol. 5, no. 1, pp. 155–161, 2020.
- [14] D. W. Nugraha, A. Y. E. Dodu, and N. Chandra, "Klasifikasi Penyakit Stroke Menggunakan Metode Naive Bayes Classifier (Studi Kasus Pada Rumah Sakit Umum Daerah Undata Palu)," *semanTIK*, vol. 3, no. 2, pp. 13–22, 2017.
- [15] S. A. J. I. Nugroho, "Naskah publikasi perbandingan metode fuzzy k-nearest neighbor dan neighbor weighted k-nearest neighbor untuk deteksi penyakit stroke," 2020.