

IMPELEMENTASI ALGORITMA REGRESI LINIER BERGANDA UNTUK PREDIKSI PENJUALAN

Yaya Asohi¹, Andri²,

Fakultas Ilmu Komputer, Universitas Bina Darma

Email: yayaasohi1212@gmail.com¹, andri@binadarma.ac.id²

ABSTRAK

Minimarket merupakan sebuah tempat atau toko yang menjual berbagai macam barang kebutuhan pokok masyarakat dan umumnya menggunakan teknologi atau sebuah mesin untuk melakukan transaksi pembayarannya. Permasalahan yang sering dihadapi pada minimarket yakni tidak menentunya jumlah penjualan, hal itu juga yang dihadapi oleh Minimarket Sr. Penelitian ini bertujuan untuk memprediksi jumlah penjualan barang menggunakan algoritma regresi linier berganda sebagai metode yang akan diterapkan untuk memprediksi. Hasil yang didapatkan dari penelitian ini menghasilkan jumlah prediksi penjualan barang untuk tahun 2020 sebesar 169715 barang.

Kata kunci: Minimarket, Prediction, Algoritma Regresi Linier Berganda.

ABSTRACT

Minimarket is a place or shop that sells various kinds of basic necessities of the community and generally uses technology or a machine to make payment transactions. The problem that is often faced in minimarkets is the uncertainty of the number of sales, it is also faced by Minimarket Sr. This study aims to predict the number of sales of goods using multiple linear regression algorithms as a method to be applied to predict. The results obtained from this study resulted in the predicted number of sales of goods for the year 2020 amounting to 169715 items.

Keywords: Minimarket, Prediction, Multiple Linear Regression Algorithm.

1. PENDAHULUAN

Transaksi jual beli merupakan kegiatan yang dilakukan manusia untuk memenuhi kebutuhan pokok mereka, kegiatan ini dapat dilakukan baik secara langsung bertatap muka maupun secara tidak langsung atau secara online. Kegiatan transaksi jual beli umumnya dilakukan disuatu tempat yang sering disebut sebagai toko.

Minimarket merupakan tempat atau toko yang menjual berbagai macam barang kebutuhan pokok masyarakat dan umumnya menggunakan teknologi atau sebuah mesin untuk melakukan transaksi pembayarannya [1]. Minimarket Sr merupakan salah satu perusahaan minimarket yang berada di Indonesia, dalam mengelola sebuah minimarket dibutuhkan manajemen atau tatakelola yang baik agar minimarket bisa terus berkembang maju, salah satu permasalahan yang sering dialami oleh usaha minimarket adalah belum adanya kepastian jumlah penjualan barang, akankah kedepannya terjadi kenaikan maupun penurunan jumlah penjualan barang. Salah satu solusi untuk menyelesaikan permasalahan ini yakni dengan melakukan prediksi jumlah penjualan barang.

Algoritma Regresi Linier Berganda merupakan teknik algoritma regresi linier yang memiliki jumlah variabel bebas X_1, X_2, X_n lebih dari satu dan dapat memiliki keterkaitan serta dapat mempengaruhi dari variabel Y , dimana variabel Y adalah satu variabel yang tidak bebas, tujuan dari algoritma ini yakni untuk memprediksi variabel Y [2].

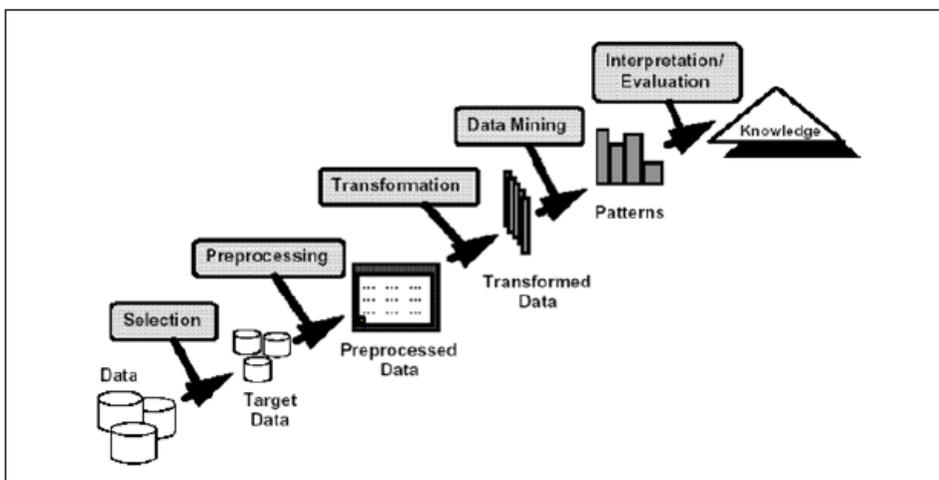
Penelitian ini akan menggunakan Algoritma Regresi Linier Berganda, hal ini dikarenakan data yang akan diolah untuk memprediksi penjualan memiliki nilai 2 variabel bebas yakni X_1 dan X_2 dan memiliki satu variabel tidak bebas Y .

2. METODOLOGI PENELITIAN

2.1. Alur Penelitian

Penelitian ini menggunakan tahapan data mining dengan metode Knowledge Discovery in Database (KDD), adapun pengertian dari KDD yaitu suatu penjelasan mengenai proses penggalian informasi yang tersembunyi di dalam suatu kumpulan data yang besar [3].

Selain itu Istilah knowledge discovery in database (KDD) juga digunakan untuk memberi penjelasan mengenai proses pencarian informasi yang belum diketahui di dalam suatu basis data yang besar [4]. Adapun tahapan prosesnya antara lain:



Gambar 1. Tahapan Dalam KDD

- Seleksi data (*selection*), merupakan kegiatan menyeleksi atau memilih data dari kumpulan data, proses ini dilakukan untuk memilah data berdasarkan kebutuhan dalam proses penelitian.
- Pra-pemrosesan (*Pre-processing*) dan pembersihan (*Cleaing*), pada proses ini dilakukan pembersihan data, adapun proses pembersihan data tersebut yakni dilakukannya pemeriksaan data yang bersifat *inconsistent*, menghapus data yang sama serta memperbaiki kesalahan pada data.
- Transformasi data, pada tahapan ini dilakukan proses penggabungan data yang dipilih berdasarkan kebutuhan penelitian, dimana data tersebut selanjutnya akan di proses pada *data mining*
- Data mining*, pada proses ini dilakukan pencarian pola atau pengetahuan yang sebelumnya belum diketahui di dalam suatu data pilihan menggunakan metode tertentu.
- Evaluasi, kegiatan mengartikan pola yang diperoleh dari tahapan *data mining*, dimana pola yang dihasilkan dari proses *data mining* tersebut perlu ditunjukkan didalam bentuk yang bisa di mengerti oleh pengguna.

2.2 Prediksi (*Forecasting*)

Prediksi merupakan suatu jenis kegiatan untuk memperkirakan sesuatu yang akan terjadi dimasa depan. Permasalahan dalam proses menentukan keputusan adalah kendala yang dihadapi,

oleh sebab itu prediksi termasuk dari permasalahan yang harus dihadapi, dikarenakan prediksi memiliki keterkaitan dengan pengambilan keputusan [5].

2.3 Data Mining

Data mining adalah penggalian data yang bertujuan untuk mendapatkan informasi yang berguna dari database. Penggalian data juga bisa diartikan sebagai pencarian informasi baru yang di dapatkan dari data dalam jumlah besar, untuk membantu dalam pengambilan keputusan [6].

Data mining adalah suatu kegiatan yang mendukung dalam pemilihan keputusan yang mana kita mencari pola informasi dalam data, proses pencarian ini dapat dilakukan oleh pengguna dengan bantuan suatu aplikasi yang secara otomatis dapat mencari pola informasi dari basis data [7].

2.4 Regresi Linier Berganda

Algoritma Regresi linier memiliki keterkaitan secara linear antara dua atau lebih variabel yang bebas (*independen*) X_1, X_2, \dots, X_n dengan variabel yang tidak bebas (*dependen*) Y . Dalam mencari nilai dari variabel regresi & konstanta, variabel bebas bisa didapatkan dari matriks determinan [8]. Adapun rumusnya antara lain:

$$A = \begin{vmatrix} N & \Sigma X_1 & \Sigma X_2 & \Sigma X_n \\ \Sigma X_1 & \Sigma X_1.X_1 & \Sigma X_2.X_1 & \Sigma X_1.X_n \\ \Sigma X_2 & \Sigma X_1.X_2 & \Sigma X_2.X_2 & \Sigma X_n.X_n \\ \Sigma X_3 & \Sigma X_1.X_3 & \Sigma X_2.X_3 & \Sigma X_n.X_n \\ \Sigma X_n & \Sigma X_n.X_n & \Sigma X_n.X_n & \Sigma X_n.X_n \end{vmatrix} \quad (1)$$

$$B = \begin{vmatrix} b_0 \\ b_1 \\ b_2 \\ b_n \end{vmatrix} \quad (2)$$

$$H = \begin{vmatrix} \Sigma Y \\ \Sigma Y.X_1 \\ \Sigma Y.X_2 \end{vmatrix} \quad (3)$$

Selanjutnya untuk menghasilkan nilai dari pada b_1, b_2 dan a , antara lain:

$$a = \frac{\text{Det}(A_0)}{\text{Det}(A)} \quad B_1 = \frac{\text{Det}(A_1)}{\text{Det}(A)} \quad B_2 = \frac{\text{Det}(A_2)}{\text{Det}(A)} \quad B_n = \frac{\text{Det}(A_n)}{\text{Det}(A)} \quad (4)$$

$$(Y = a + b_1.X_1 + b_2.X_2 \dots + b_n.X_n) \quad (5)$$

Yang mana:

Y = Value dari Prediksi

A = Value dari konstanta

X_1 = Value dari variabel bebas pertama

X_2 = Value dari variabel bebas kedua

X_n = Value dari variabel bebas n

2.5 Rapidminer

Rapidminer adalah sebuah perangkat lunak untuk mengelola *data mining*. Rapidminer dijadikan salah satu pilihan solusi untuk melakukan analisa terhadap *data mining, text mining* dan analisa prediksi. Penggunaan berbagai jenis teknis seperti teknik deskriptif serta prediksi pada

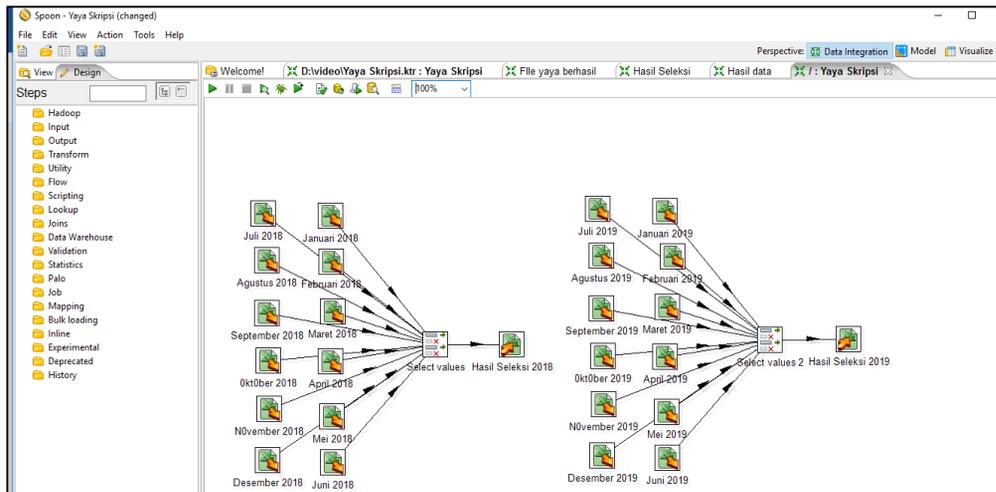
rapidminer dapat memberikan pengetahuan kepada para pengguna sehingga bisa memberikan keputusan yang paling baik [9].

3. HASIL DAN PEMBAHASAN

3.1. Proses Knowledge Discovery in Database (KDD)

3.1.1. Data Selection

Pada proses ini data diseleksi berdasarkan atribut yang dibutuhkan untuk kegiatan data mining. Dalam tahapan seleksi data, alat yang digunakan yakni pentaho data integration. Berikut ini adalah tampilan dari proses *data selection* menggunakan alat pentaho data integration:



Gambar 2. Proses *data selection*

3.1.2. Preprocessing

Dalam tahapan ini dilakukan *cleaning* dan *integration data*. Adapun prosesnya antara lain:

Table 1. Sampel Dari Data Penjualan Yang Belum Dilakukan *Cleaning Data*

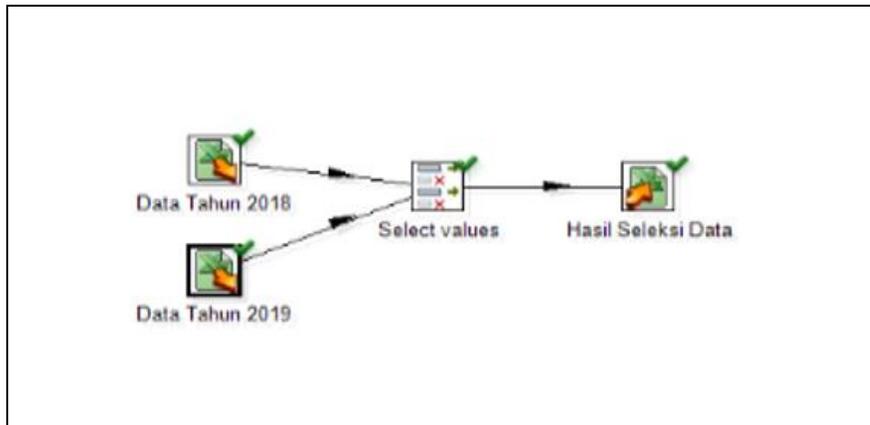
| NOMOR | NO TRANSAKSI | TANGGAL | NAMA_BARANG | QT | HARGA |
|--------|--------------|----------------|-------------------------------|----|--------|
| 1 | BR R000001 | Januari - 18 | INDOMILK LIQ MELON BTL 190ML | 2 | 3,802 |
| 2 | BR R000002 | Januari - 18 | PUCUK HARUM TEH PET 350ML/24 | 4 | 2,865 |
| 3 | BR R000003 | Januari - 18 | GARUDA ROSTA K/P BWANG 100C | 2 | 8,390 |
| 4 | BR R000004 | Januari - 18 | NN FANTASY STK RAINBOW 22.5G | 1 | 3,196 |
| 5 | BR R000005 | Januari - 18 | ROTI SOBEK 6S PCS 1 | 3 | 8,000 |
| 6 | BR R000006 | Januari - 18 | FRESTEA G.TEA MADU PET 500ML | 2 | 5,080 |
| 172 | HPYR000076 | Januari - 18 | ABC SARDINES EXT PEDAS CAN 15 | -2 | 8,000 |
| 213 | HPYR000081 | Januari - 18 | ABC KACANG HIJAU 250ML | -3 | 5,000 |
| 222 | HPYR000081 | Januari - 18 | WALL'S P.POP CHOCO MAGMA 55N | -2 | 0 |
| 236 | HPYR000082 | Januari - 18 | ALIS OREZA | -1 | 10,000 |
| 271 | HPYR000091 | Januari - 18 | GOON L30 | -1 | 52,163 |
| 301 | HPYR000101 | Januari - 18 | CUSSONS BABY COTTON EXTRA | -1 | 5,000 |
| 247192 | DR R037613 | Desember -2019 | DARK CHOCOLATE COMPOUND 25 | 2 | 18,000 |

Tahapan *cleaning data* bertujuan untuk menghapuskan data null, data yang sama serta data tidak konsisten,

Tabel 2. Sampel Data Penjualan Yang Telah Dilakukan *Cleaning*

| NOMOR | NO TRANSAKSI | TANGGAL | NAMA_BARANG | QT | HARGA |
|--------|--------------|--------------|----------------------------------|----|--------|
| 1 | BR R000001 | Januari - 18 | INDOMILK LIQ MELON BTL 190ML | 2 | 3,802 |
| 2 | BR R000002 | Januari - 18 | PUCUK HARUM TEH PET 350ML/24 | 4 | 2,865 |
| 3 | BR R000003 | Januari - 18 | GARUDA ROSTA K/P BWANG 100G | 2 | 8,390 |
| 4 | BR R000004 | Januari - 18 | NN FANTASY STK RAINBOW 22.5G/144 | 1 | 3,196 |
| 5 | BR R000005 | Januari - 18 | ROTI SOBEK 6S PCS 1 | 3 | 8,000 |
| 6 | BR R000006 | Januari - 18 | FRESTEA G.TEA MADU PET 500ML | 2 | 5,080 |
| 7 | BR R000007 | Januari - 18 | FORTUNE PP 1 L/24 | 2 | 11,500 |
| 8 | BR R000008 | Januari - 18 | PSM GULA PREMIUM 1 KG | 2 | 12,500 |
| 9 | BR R000008 | Januari - 18 | SUNLIGHT PLUS A.BAK REF 800ML | 1 | 17,075 |
| 10 | BR R000009 | Januari - 18 | SGM EKSPLOR 1+ VANILLA 600 GR | 1 | 51,300 |
| 11 | BR R000010 | Januari - 18 | BONEETO CHOCOLATE 350G | 1 | 39,165 |
| 12 | BR R000011 | Januari - 18 | CB NIPPLE JAR MIX 18S | 1 | 5,108 |
| 245244 | DS R037613 | Desember- 19 | BELFOODS SOSIS AYAM 375GR | 2 | 18,000 |

Setelah dilakukannya proses *cleaning* data, hasil yang didapatkan yakni data penjualan yang sebelumnya berjumlah 247192 *record*, setelah dilakukan *cleaning data* berubah menjadi 245244 *record*. Selanjutnya adalah proses *integration* data adapun prosesnya adalah sebagai berikut:



Gambar 3. Proses *Data Integration*

Pada tahapan dari *Data Integration* data dari berbagai sumber disatukan untuk keperluan dalam proses data mining.

3.1.3. *Data Transformation*

Dalam proses ini data diolah dan disesuaikan agar *format* dari data tersebut bisa dipakai dalam proses *data mining*.

Table 3. Sampel Data Yang Selesai Dilakukan Transformasi Data

| NOMOR | NO TRANSAK | TANGGAL | NAMA_BARANG | QT | HARGA |
|--------|------------|--------------|----------------------------------|----|--------|
| 1 | 1 | Januari - 18 | INDOMILK LIQ MELON BTL 190ML | 2 | 3,802 |
| 2 | 2 | Januari - 18 | PUCUK HARUM TEH PET 350ML/24 | 4 | 2,865 |
| 3 | 3 | Januari - 18 | GARUDA ROSTA K/P BWANG 100G | 2 | 8,390 |
| 4 | 4 | Januari - 18 | NN FANTASY STK RAINBOW 22.5G/144 | 1 | 3,196 |
| 5 | 5 | Januari - 18 | ROTI SOBEK 6S PCS 1 | 3 | 8,000 |
| 6 | 6 | Januari - 18 | FRESTEA G.TEA MADU PET 500ML | 2 | 5,080 |
| 7 | 7 | Januari - 18 | FORTUNE PP 1 L/24 | 2 | 11,500 |
| 8 | 8 | Januari - 18 | PSM GULA PREMIUM 1 KG | 2 | 12,500 |
| 9 | 8 | Januari - 18 | SUNLIGHT PLUS A.BAK REF 800ML | 1 | 17,075 |
| 10 | 9 | Januari - 18 | SGM EKSPLOK 1+ VANILLA 600 GR | 1 | 51,300 |
| 11 | 10 | Januari - 18 | BONEETO CHOCOLATE 350G | 1 | 39,165 |
| 12 | 11 | Januari - 18 | CB NIPPLE JAR MIX 18S | 1 | 5,108 |
| 13 | 11 | Januari - 18 | GLICO POCKY CHOCOLATE 47G | 1 | 6,184 |
| 247192 | 92834 | Desember- 19 | BELFOODS SOSIS AYAM 375GR | 2 | 18,000 |

Dari Proses transformasi data, menghasilkan perubahan untuk bagian data transaksi yang sebelumnya berbentuk identitas kode nomor transaksi berubah ke dalam bentuk urutan nomor transaksi.

3.1.4. Data Mining

Pada tahapan *data mining* dilakukan simulasi penghitungan data dengan penggunaan dari data penjualan dimulai sejak tahun 2018 hingga tahun 2019, berdasarkan jangka waktu prediksi untuk enam bulan kedepan. Adapun yang pertama-tama yang harus dilakukan adalah mencari nilai dari:

$$\left| \begin{array}{cccc} \sum Y & \sum X1 & \sum X1^2 & \sum X1Y \\ \sum X2 & \sum X2^2 & \sum X1.X2 & \sum X2Y \end{array} \right|$$

Tabel 4. Akumulasi Data Penjualan Per 6 Bulan Beserta Nilai Dari $\sum Y$, $\sum X2$ & $\sum X1$

| Bulan - Tahun | Jumlah Pendapatan (X2) | Jumlah Transaksi (X1) | Jumlah penjualan Barang (Y) |
|----------------------|------------------------|-----------------------|-----------------------------|
| Januari - Juni 2018 | 1245725790 | 25318 | 107044 |
| Juli - Desember 2018 | 998148367 | 23754 | 83863 |
| Januari - Juni 2019 | 1031502851 | 23029 | 81543 |
| Juli - Desember 2019 | 780362729 | 20733 | 66980 |
| Jumlah | 4055739737 | 92834 | 339430 |
| | $\sum X2$ | $\sum X1$ | $\sum Y$ |

Tabel 5. Akumulasi Nilai Dari $\sum X1.X2$, $\sum X1^2$, $\sum X2^2$, $\sum X1Y$, & $\sum X2Y$

| X1 ² | X1.X2 | X1.Y | X2 ² | X2.Y |
|-------------------|--------------------------|-------------------|-------------------------------|---------------------------|
| 641001124 | 31539285551220.00 | 2710139992.00 | 1551832743871120000.00 | 133347471464760.00 |
| 564252516 | 23710016309718.00 | 1992081702.00 | 996300162544767000.00 | 83707716501721.00 |
| 530334841 | 23754479155679.00 | 1877853747.00 | 1063998131621130000.00 | 84111836979093.00 |
| 429857289 | 16179260460357.00 | 1388696340.00 | 608965988812327000.00 | 52268695588420.00 |
| 2165445770 | 95183041476974.00 | 7968771781 | 4221097026849350000.00 | 353435720533994.00 |
| $\sum X1^2$ | $\sum X1.X2$ | $\sum X1.Y$ | $\sum X2^2$ | $\sum X2.Y$ |

Dari proses penghitungan yang telah dilakukan maka diperoleh nilai dari $\sum Y$, $\sum X1$, $\sum X1^2$, $\sum X2^2$, $\sum X1Y$, $\sum X2Y$ & $\sum X1.X2$. Berikutnya adalah mencari nilai dari A0, A1, A2 dan A3.

Tabel 6. Data A0

| | | | |
|------|------------|-------------------|------------------------|
| A0 = | n | $\sum X1$ | $\sum X2$ |
| | $\sum X1$ | $\sum X1^2$ | $\sum X1.X2$ |
| | $\sum X2$ | $\sum X1.X2$ | $\sum X2^2$ |
| A0 = | 4 | 92834 | 4055739737 |
| | 92834 | 2165445770 | 95183041476974.00 |
| | 4055739737 | 95183041476974.00 | 4221097026849350000.00 |

Tabel 7. Data A1

| | | | |
|------|-------------|-------------------|------------------------|
| A1 = | $\sum Y$ | $\sum X1$ | $\sum X2$ |
| | $\sum X1Y$ | $\sum X1^2$ | $\sum X1.X2$ |
| | $\sum X2Y$ | $\sum X1.X2$ | $\sum X2^2$ |
| A1 = | 339430 | 92834 | 4055739737 |
| | 7968771781 | 2165445770 | 95183041476974.00 |
| | 3.53436E+14 | 95183041476974.00 | 4221097026849350000.00 |

Tabel 8. Data A2

| | | | |
|------|------------|--------------------|------------------------|
| A2 = | n | $\sum Y$ | $\sum X2$ |
| | $\sum X1$ | $\sum X1Y$ | $\sum X1.X2$ |
| | $\sum X2$ | $\sum X2Y$ | $\sum X2^2$ |
| A2 = | 4 | 339430 | 4055739737 |
| | 92834 | 7968771781 | 95183041476974.00 |
| | 4055739737 | 353435720533994.00 | 4221097026849350000.00 |

Tabel 9. Data A3

| | | | |
|------|------------|-------------------|--------------------|
| A3 = | n | $\sum X1$ | $\sum Y$ |
| | $\sum X1$ | $\sum X1^2$ | $\sum X1Y$ |
| | $\sum X2$ | $\sum X1.X2$ | $\sum X2Y$ |
| A3 = | 4 | 92834 | 339430 |
| | 92834 | 2165445770 | 7968771781 |
| | 4055739737 | 95183041476974.00 | 353435720533994.00 |

Setelah proses penghitungan untuk menentukan nilai dari rumus A0, A1, A2 dan A3 berhasil didapatkan maka proses selanjutnya adalah melakukan perhitungan data tersebut untuk mendapatkan nilai A0, A1, A2 dan A3.

$$A0 = n((\sum X1^2 * \sum X2^2) - (\sum X1.X2 * \sum X1.X2)) - \sum X1((\sum X1 * \sum X2^2) - (\sum X1.X2 * \sum X2)) + \sum X2((\sum X1 * \sum X1.X2) - (\sum X1^2 * \sum X2))$$

$$A0 = 4((2165445770 * 4221097026849350000) - (95183041476974 * 95183041476974)) - 92834((92834 * 4221097026849350000) - (95183041476974 * 4055739737)) + 4055739737 + ((92834 * 95183041476974) - (2165445770 * 4221097026849350000))$$

$$A0 = 2.93365E+23$$

$$A1 = \sum Y((\sum X1^2 * \sum X2^2) - (\sum X1.X2 * \sum X1.X2)) - \sum X1((\sum X1Y * \sum X2^2) - (\sum X1.X2 * \sum X2Y)) + \sum X2((\sum X1Y * \sum X1.X2) - (\sum X1^2 * \sum X2Y))$$

$$A1 = 339430((2165445770 * 4221097026849350000) - (95183041476974 * 95183041476974)) - 92834((7968771781 * 4221097026849350000) - (95183041476974 * 353435720533994)) + 4055739737((7968771781 * 95183041476974) - (2165445770 * 353435720533994))$$

$$A1 = -7.24235E+27$$

$$A2 = n((\sum X1Y * \sum X2^2) - (\sum X1.X2 * \sum X2Y)) - \sum Y((\sum X1 - \sum X2) - (\sum X1.X2 * \sum X2)) + \sum X2((\sum X1 * \sum X2Y) - (\sum X1Y * \sum X2))$$

$$A2 = 4 ((7968771781 * 4221097026849350000) - (95183041476974 * 353435720533994)) - 339430 ((92834 * 4221097026849350000) - (95183041476974 * 4055739737)) + 4055739737 ((92834 * 353435720533994) - (7968771781 * 4055739737))$$

$$A2 = 5.0735E+23$$

$$A3 = n ((\sum X1^2 * \sum X2Y) - (\sum X1Y * \sum X1.X2)) - \sum X1 ((\sum X1 * \sum X2Y) - (\sum X1Y * \sum X2)) + \sum Y ((\sum X1 * \sum X1.X2) - (\sum X1^2 * \sum X2))$$

$$A3 = 4 ((2165445770 * 353435720533994) - (7968771781 * 95183041476974)) - 92834 ((92834 * 353435720533994) - (7968771781 * 4055739737)) + 339430 ((92834 * 95183041476974) - (2165445770 * 4055739737))$$

$$A3 = 2.00819E+19$$

Dari proses penghitungan maka didapatkanlah nilai dari A0, A1, A2 dan A3 langkah selanjutnya adalah mencari nilai dari a, b1 dan b2.

$$a = \frac{A1}{A0}$$

$$a = \frac{-7.24235E+27}{2.93365E+23}$$

$$a = -24687.1415$$

$$b1 = \frac{A2}{A0}$$

$$b1 = \frac{5.0735E+23}{2.93365E+23}$$

$$b1 = 1.729412782$$

$$b2 = \frac{A3}{A0}$$

$$b2 = \frac{2.00819E+19}{2.93365E+23}$$

$$b2 = 0.000068453667585576$$

Setelah dilakukan perhitungan maka didapatkan nilai nilai dari koefisien yaitu sebagai berikut:

$$a = -24687.142$$

$$b1 = 1.72941278$$

$$b2 = 0.000068453667585576$$

Sehingga didapatlah persamaan regresi linier berganda sebagai berikut:

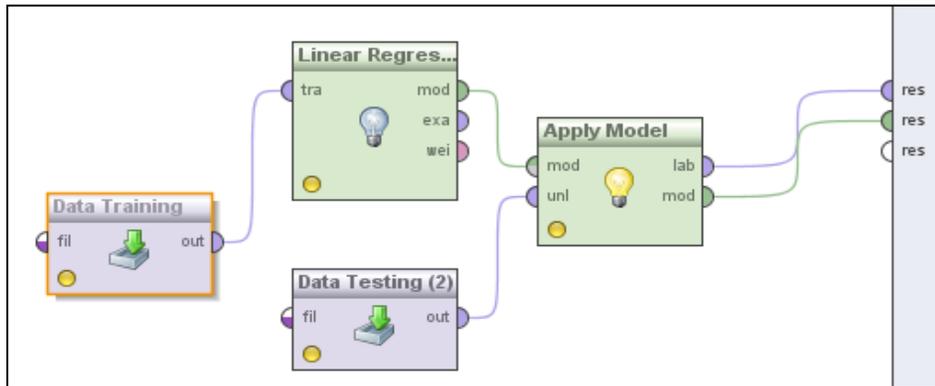
$$Y = -24687.142 + 1.72941278 * X1 + 0.000068453667585576 * X2$$

3.2. Hasil

Sebelumnya telah dilakukan proses prediksi secara manual untuk menentukan jumlah penjualan barang dalam kurun waktu 6 bulan kedepan, selanjutnya dilakukan prediksi menggunakan menggunakan aplikasi Rapidmider sebagai alat bantu dalam melakukan prediksi penjualan barang. Proses prediksi dibagi menjadi tiga yakni prediksi berdasarkan penjualan per-tahun, per-barang dan prediksi untuk mengukur tingkat kesalahan. Berikut ini adalah proses prediksi menggunakan tool Rapidminer:

3.2.1. Prediksi Per-Tahun

Proses prediksi penjualan per-tahun dilakukan untuk mencari nilai prediksi barang dalam kurun waktu satu tahun kedepan. Berikut ini adalah tampilan dari proses prediksi penjualan per-tahun menggunakan aplikasi rapidminer:



Gambar 4. Proses Prediksi Dengan Tool Rapidmir

Pada proses prediksi penjualan barang per-tahun ada 4 operator yang digukan yakni, operator *input data (training)*, *input data (testing)*, *linier regression* dan *apply model*. Adapun hasil dari prediksi sebagai berikut:

| Row No. | Tahun | Jumlah Pen... | prediction(J... | Jumlah Tra... | Total Penda... |
|---------|-------|---------------|-----------------|---------------|----------------|
| 1 | 2019 | 148523 | 169715 | 43762 | 181186558C |

Gambar 5. Hasil Dari Prediksi Per-Barang

Hasil dari proses prediksi penjualan per-barang pada priode tahun 2020 yakni, 169725 barang.

3.2.2. Prediksi Per-Barang

Proses prediksi penjuln per-barang bertujuan untuk mencari nilai penjualan berdasarkan jenis barang, adapun proses dari prediksi per-barang mempunyai kesamaan dengan proses prediksi per-barang, yang membedakan adalah dari segi data yang diinputkan kedalam operator *data (training)*, *input data (testing)*. Berikut ini merupakan hasil dari kegiatan prediksi dengan memakai tool rapidminer:

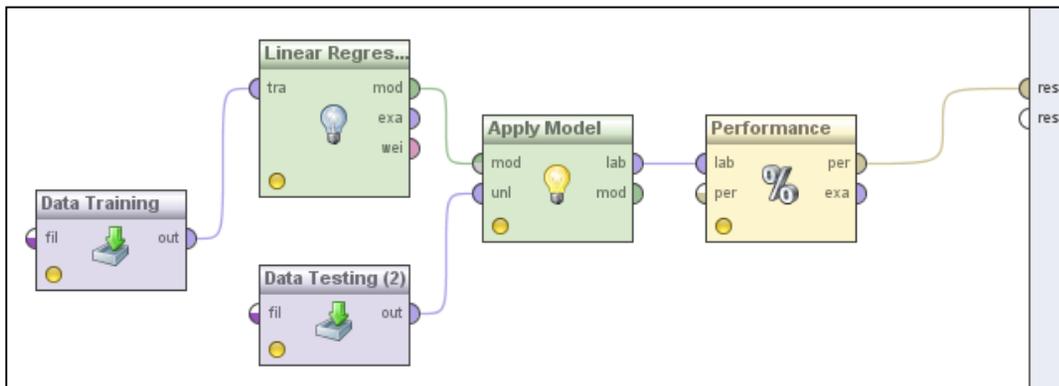
| Row No. | prediction(J... | Nama Bara... | Jumlah Tra... | Pendapata... | Jumlah pen... |
|---------|-----------------|--------------|---------------|--------------|---------------|
| 1 | 73.057 | ABC KACAN | 71 | 248500 | 131 |
| 2 | 419.494 | FORTUNE P | 233 | 2665500 | 420 |
| 3 | 306.153 | FORTUNE S | 180 | 3931200 | 246 |
| 4 | 408.801 | INDOMIE GC | 228 | 696000 | 510 |
| 5 | 154.320 | SOVIA MINY | 109 | 2445500 | 154 |
| 6 | 284.768 | INDOMIE GC | 170 | 430000 | 319 |
| 7 | 68.780 | BERAS PATI | 69 | 7728000 | 85 |
| 8 | 130.797 | SUPERMI A | 98 | 294000 | 171 |
| 9 | 177.844 | INDOMIE CE | 120 | 363000 | 254 |
| 10 | 402.386 | ALFA 330 ML | 225 | 466000 | 348 |
| 11 | 23.872 | TEH BOTOL | 48 | 144000 | 107 |
| 12 | 70.919 | OREO C/S/C | 70 | 620500 | 74 |
| 13 | 137.212 | ADEM SARI | 101 | 656500 | 119 |
| 14 | 141.489 | SEDAAP MIE | 103 | 260000 | 202 |
| 15 | 132.935 | TEH GLAS | 99 | 381500 | 123 |

Gambar 6. Hasil prediksi Per-Barang

Dari proses prediksi yang telah dilakukan menghasilkan nilai prediksi penjualan per-barang dengan jenis penjualan barang yang paling tinggi yakni FORTUNE PP 1 L/24 yang mempunyai angka prediksi 419,494 barang & menghasilkan nilai prediksi terendah untuk jenis barang TEH BOTOL KOTAK SOSRO 200 ML yang mempunyai angka prediksi 23,872 barang.

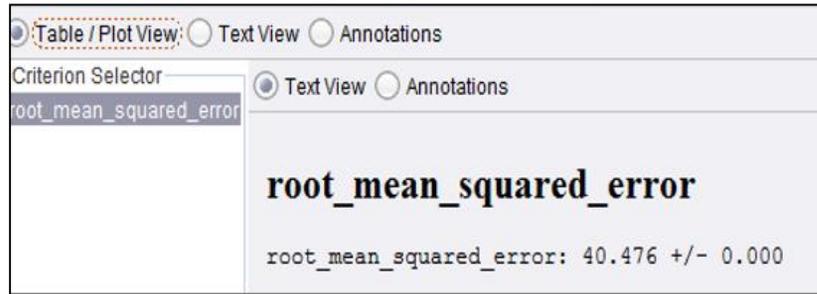
3.2.3. Pengukuran Tingkat Kesalahan (*error*)

Proses pengukuran data atau *root mean square error* (RMSE), bertujuan untuk mengetahui jarak nilai prediksi dengan kenyataannya. Adapun prosesnya antara lain:



Gambar 7. Proses Prediksi Pengukuran Tingkat Kesalahan

Tahapan dalam proses pengukuran dari tingkat kesalahan mempunyai kesamaan dengan proses prediksi penjualan barang per-barang maupun per-tahun, yang membedakannya adalah adanya penambahan 1 operator baru yakni operator performance, operator ini berguna untuk mengukur nilai RSME dari prediksi yang dilakukan. Adapun hasil dari proses ini antara lain:



Gambar 8. Nilai RMSE

Dari proses prediksi yang dilakukan untuk mengukur nilai dari RMSE, menghasilkan nilai RMSE sebesar 40,476.

4. KESIMPULAN

Prediksi penjualan barang menggunakan algoritma regresi linier dengan proses pengolahan data menggunakan tahapan dari *Knowledge Discovery In Database* (KDD), serta penggunaan tool rapidminer sebagai alat bantu dalam proses prediksi barang menghasilkan:

- a. Hasil yang diperoleh dari prediksi penjualan barang untuk periode waktu tahun 2020 sebesar 169715 barang.
- b. Dari proses prediksi penjualan berdasarkan jenis barang menghasilkan nilai penjualan barang tertinggi yaitu berjenis barang FORTUNE PP 1 L/24, yang memiliki angka prediksi penjualan barang sebesar 419,494 & menghasilkan nilai penjualan barang terendah untuk jenis barang TEH BOTOL KOTAK SOSRO 200 ML dengan angka 23,872 barang.
- c. Pengukuran tingkat *error*, diperoleh nilai dari *Root Mean Square Error* (RMSE) dengan nilai 40,476.

DAFTAR PUSTAKA

- [1] A. K. ASLAM, "Pengaruh pertumbuhan minimarket terhadap minat dan kebiasaan belanja masyarakat di kelurahan tamamaung kota makassar," 2017.
- [2] N. L. P. Wulandari, N. L. A. K. Y. Sarja, and I. G. A. D. Saryanti, "Prediksi Jumlah Pelanggan Dan Persediaan Barang Menggunakan Metode Regresi Linier Berganda Pada Bali Orchid," *Eksplora Inform.*, pp. 1–12, 2014.
- [3] I. K. Juni Arta, G. Indrawan, and G. R. Dantes, "Data Mining Rekomendasi Calon Mahasiswa Berprestasi Di Stmik Denpasar Menggunakan Metode Technique for Others Reference By Similarity To Ideal Solution," *JST (Jurnal Sains dan Teknol.*, vol. 5, no. 2, p. 792, 2017, doi: 10.23887/jst-undiksha.v5i2.8549.
- [4] N. A. Hasibuan *et al.*, "Implementasi Data Mining Untuk Pengaturan Layout," vol. 4, no. 4, pp. 6–11, 2017.
- [5] M. Irfan, L. P. Ayuningtias, and J. Jumadi, "Analisa Perbandingan Logic Fuzzy Metode Tsukamoto, Sugeno, Dan Mamdani (Studi Kasus : Prediksi Jumlah Pendaftar Mahasiswa Baru Fakultas Sains Dan Teknologi Uin Sunan Gunung Djati Bandung)," *J. Tek. Inform.*, vol. 10, no. 1, pp. 9–16, 2018, doi: 10.15408/jti.v10i1.6810.
- [6] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [7] D. S. Kusumo, M. A. Bijaksana, and D. Darmantoro, "Data Mining Dengan Algoritma Apriori Pada Rdbms Oracle," *TEKTRIKA - J. Penelit. dan Pengemb. Telekomun. Kendali, Komputer, Elektr. dan Elektron.*, vol. 8, no. 1, pp. 1–5, 2016, doi: 10.25124/tektrika.v8i1.215.

- [8] M. Syafruddin, L. Hakim, and D. Despa, "Metode Regresi Linier Untuk Prediksi Kebutuhan Energi Listrik Jangka Panjang (Studi Kasus Provinsi Lampung)," *J. Inform.*, vol. 2, no. 1, pp. 1–9, 2014, [Online]. Available: <http://journal.eng.unila.ac.id/index.php/jitet/article/download/237/228>.
- [9] A. Purwanto *et al.*, "Perbandingan Minat Siswa Smu Pada Metode Klasifikasi Menggunakan 5 Algoritma," vol. 2, no. 1, pp. 43–47, 2018.