

Penerapan Data Mining Dalam Memprediksi Jumlah Penumpang Pada CV. Surya Mandiri Sukses Dengan Menggunakan Metode Regresi Linier

Hafizah, Tugiono, Widiarti Rista Maya

STMIK Triguna Dharma

e-mail: hafizah22isnartilyas@gmail.com

Abstrak

CV.Surya Mandiri Sukses merupakan salah satu perusahaan yang bergerak dibidang transportasi pengangkutan penumpang antar kota dalam provinsi, dimana rute pengangkutannya meliputi kota Rantau Parapat dan Padang Sidempuan. Memprediksi jumlah penumpang sangat diperlukan CV. Surya Mandiri Sukses agar dapat mempersiapkan armada mobil angkutan yang digunakan. Data Mining merupakan suatu cabang ilmu dari kecerdasan buatan (Artificial Intelligence). Dalam Data Mining terdapat beberapa jenis metode sesuai dengan pemanfaatannya diantaranya: prediksi, asosiasi, klasifikasi, klastering, dan estimasi. Regresi Linear merupakan metode statistik yang berfungsi untuk menguji sejauh mana hubungan sebab akibat antara Variabel Faktor Penyebab (X) terhadap Variabel Akibatnya (Y). Dengan penggunaan metode Regresi Linier pihak CV. Surya Mandiri Sukses dapat terbantu dalam mengatasi permasalahan di dalam memprediksi jumlah penumpang di kemudian hari.

Kata Kunci : Data Mining, Regresi Linier Berganda, Prediksi Jumlah Penumpang

1. Pendahuluan

CV.Surya Mandiri Sukses merupakan salah satu perusahaan yang bergerak dibidang transportasi pengangkutan penumpang antar kota dalam provinsi, dimana rute pengangkutannya meliputi kota Rantau parapat dan Padang Sidempuan. CV.Surya Mandiri Sukses merupakan salah satu perusahaan pengangkutan yang digemari masyarakat yang ingin berpergian, hal ini ditandai banyaknya penumpang CV.Surya Mandiri Sukses.

Memprediksi jumlah penumpang CV.Surya Mandiri Sukses sangat diperlukan, hal ini dikarenakan agar CV.Surya Mandiri Sukses dapat mempersiapkan armada mobil angkutan yang digunakan untuk mengangkut penumpang ke tujuan sehingga para penumpang tidak terlambat sampai ditujuan dikarenakan menunggu atau kekurangan armada mobil angkutan. Dalam beberapa literatur ternyata ada beberapa teknik yang dapat digunakan untuk memprediksi jumlah penumpang diantaranya adalah data mining.

Data mining merupakan suatu cabang ilmu dari kecerdasan buatan (artificial intelligence). Dalam data mining terdapat beberapa jenis metode sesuai dengan pemanfaatannya diantaranya: prediksi, asosiasi, klasifikasi, klastering, dan estimasi. Regresi Linear merupakan metode statistik yang berfungsi untuk menguji sejauh mana hubungan sebab akibat antara Variabel Faktor Penyebab (X) terhadap Variabel Akibatnya. Faktor Penyebab pada umumnya dilambangkan dengan X atau disebut juga dengan Predictor sedangkan Variabel Akibat dilambangkan dengan Y atau disebut juga dengan Response. Regresi Linier merupakan salah

satu metode statistik yang dipergunakan dalam produksi untuk melakukan peramalan ataupun prediksi tentang karakteristik kualitas maupun kuantitas.

Dengan menggunakan metode Regresi Linier ini dapat digunakan untuk memprediksi jumlah penumpang CV. Surya Mandiri Sukses kedepannya dengan melihat data-data yang ada pada sebelumnya. Dengan demikian metode ini akan dapat membantu pihak CV.Surya Mandiri Sukses didalam mengatasi permasalahan didalam memprediksi jumlah penumpang dihari kemudian.

2. Landasan Teoritis

2.1 Data Mining

Menurut Vuldari (2017:2) "Data Mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data.

"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. Knowledge Discovery in Databases (KDD) adalah penerapan metode saintifik pada data mining. Dalam konteks ini data mining merupakan satu langkah daro proses KDD".(Fajar Astuti Hermawati,2009:3).

Menurut Prasetyo (2014:2). "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 yang besar yang membantu dalam pengambilan keputusan."Kemajuan yang terus berlanjut dalam bidang data mining didorong oleh beberapa faktor antara lain :

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam data warehouse, sehingga seluruh perusahaan memiliki akses ke dalam database yang andal.
3. Adanya peningkatan akses data melalui navigasi web dan intranet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk data mining (ketersediaan teknologi).
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Istilah data mining dan knowledge discovery in database (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan keseluruhan proses KDD adalah data mining.

2.2 Regresi Linier

Regresi linier merupakan suatu cara prediksi yang menggunakan garis lurus untuk menggambarkan hubungan di antara dua variabel (atau lebih) dan memiliki hubungan secara linier antara satu variabel independen (X) dengan variabel dependen (Y). Regresi linier sederhana ini hanya memiliki satu variabel pemberi pengaruh.

Penerapan Data Mining Dalam Memprediksi Jumlah Penumpang Pada CV. Surya Mandiri Sukses Dengan Menggunakan Metode Regresi Linier

Analisis ini untuk mengetahui arah hubungan antara variabel independen dengan variabel dependen apakah positif atau negatif dan untuk memprediksi nilai dari variabel dependen apabila nilai variabel independen mengalami kenaikan atau penurunan. Data yang digunakan biasanya berskala interval atau rasio. Data hubungan antara variabel X dan Y berdasarkan pada dua hal yaitu :

1. Penentuan bentuk persamaan yang sesuai guna meramalkan rata-rata Y melalui X atau rata-rata X melalui Y dan menduga kesalahan selisih peramalan. Hal ini menitikberatkan pada observasi variabel tertentu, sedangkan variabel-variabel lain dikonstantir pada berbagai tingkat atau keadaan, hal inilah yang dinamakan Regresi.
2. Pengukuran derajat keeratan antara variabel X dan Y. Derajat ini tergantung pada pola variasi atau interelasi yang bersifat simultan dari variabel X dan Y. Pengukuran ini disebut Korelasi.

Hubungan antara variabel X dan Y kemungkinan merupakan hubungan dependen sempurna dan kemungkinan merupakan hubungan independen sempurna. Variabel X dan Y dapat dikatakan berasosiasi atau berkorelasi secara statistik jika terdapat batasan antara dependen dan independen sempurna. Metode analisis ini juga digunakan untuk mengestimasi atau menduga besarnya suatu variabel yang lain telah diketahui nilainya dengan mencari garis lurus. Tujuan mencari sebuah garis lurus yang sedekat mungkin dengan titik sehingga garis tersebut menjadi sah untuk mewakili titik-titik tersebut. Secara umum, garis tersebut dapat kita tulis dalam bentuk persamaan.

$$Y = \beta_0 + \beta_1 x$$

Dimana, Y adalah variabel terpengaruh, β_0 adalah sebuah konstanta, β_1 adalah gradien garis, dan x adalah variabel pemberi pengaruh. Untuk mendapatkan garis liner memerlukan konstanta dan gradiennya. Jadi, untuk menemukan garis Regresi Liner yang paling baik, perlu melakukan perhitungan konstanta β_0 dan gradien β_1 dengan rumus :

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\beta_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{(\sum_{i=1}^n y_i)(\sum_{i=1}^n x_i)}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}$$

3. Hasil dan Pembahasan

Dalam penyelesaian permasalahan yang ada didalam prediksi jumlah penumpang pada kemudian hari ini menggunakan dua buah variabel dan menggunakan data yang didapat dari loket. Berikut ini adalah data jumlah penumpang bulan 1 tahun 2017 s/d bulan 5 tahun 2017.

Tabel 1 Data Penumpang Bulan Januari 2017

| Tgl Berangkat | No Plat Mobil | Jenis Mobil | Nama Supir | Tujuan | Jumlah Penumpang |
|---------------|---------------|----------------|------------|------------------|------------------|
| 01-01-17 | BK 5634 AG | Minibus Avanza | Suprian | Rantau Prapat | 7 |
| | BK 4575 GA | Minibus Inova | Andi | Rantau Prapat | 7 |
| | BK 2343 UG | Minibus Kijang | Hery | Rantau Prapat | 8 |
| | BK 3321 JD | Minibus Xenia | Jony | Padang Sidempuan | 6 |

| | | | | | |
|----------|---------------|-------------------|---------|---------------------|---|
| | BK 1932 DJ | Minibus Inova | Bagus | Padang Sidempuan | 7 |
| 02-01-17 | BK 3234 AB | Minibus Avanza | Dadang | Rantau Prapat | 6 |
| | BK 4375 GB | Minibus Inova | Dany | Rantau Prapat | 8 |
| | BK 2373 GG | Minibus Kijang | Hendara | Rantau Prapat | 7 |
| | BK 1361 ED | Minibus Xenia | Sugeng | Padang Sidempuan | 5 |
| | BK 2036 UJ | Minibus Inova | Iwan | Padang Sidempuan | 6 |
| | BK 5634 AG | Minibus Avanza | Suprian | Rantau Prapat | 6 |
| 03-01-17 | BK 4575 GA | Minibus Inova | Andi | Rantau Prapat | 7 |
| | BK 2343 UG | Minibus Kijang | Hery | Rantau Prapat | 8 |
| | BK 3321 JD | Minibus Xenia | Jony | Padang Sidempuan | 5 |
| | BK 3234 AB | Minibus Avanza | Dadang | Rantau Prapat | 7 |
| 04-01-17 | BK 4375 GB | Minibus Inova | Dany | Rantau Prapat | 7 |
| | BK 2373 GG | Minibus Kijang | Hendara | Rantau Prapat | 8 |
| | BK 1361 ED | Minibus Xenia | Sugeng | Padang Sidempuan | 6 |
| | BK 5634 AG | Minibus Avanza | Suprian | Rantau Prapat | 7 |
| 05-01-17 | BK 4575 GA | Minibus Inova | Andi | Rantau Prapat | 8 |
| | BK 2343 UG | Minibus Kijang | Hery | Rantau Prapat | 8 |
| | BK 1932 DJ | Minibus Inova | Bagus | Padang Sidempuan | 7 |
| | BK 3234 AB | Minibus Avanza | Dadang | Rantau Prapat | 6 |
| 06-01-17 | BK 4375 GB | Minibus Inova | Dany | Rantau Prapat | 7 |
| | BK 2373 GG | Minibus Kijang | Hendara | Rantau Prapat | 7 |
| | BK 2036 UJ | Minibus Inova | Iwan | Padang Sidempuan | 6 |
| | BK 5634 AG | Minibus Avanza | Suprian | Rantau Prapat | 7 |
| 07-01-17 | BK 4575 GA | Minibus Inova | Andi | Rantau Prapat | 8 |
| | BK 2343 UG | Minibus Kijang | Hery | Rantau Prapat | 8 |

| | | | | | |
|---------------------------------|---------------|-------------------|---------|---------------------|------------|
| | BK 3321 JD | Minibus Xenia | Jony | Padang Sidempuan | 6 |
| | BK 1932 DJ | Minibus Inova | Bagus | Padang Sidempuan | 7 |
| 08-01-17 | BK 3234 AB | Minibus Avanza | Dadang | Rantau Prapat | 6 |
| | BK 4375 GB | Minibus Inova | Dany | Rantau Prapat | 8 |
| | BK 2373 GG | Minibus Kijang | Hendara | Rantau Prapat | 8 |
| | BK 1361 ED | Minibus Xenia | Sugeng | Padang Sidempuan | 5 |
| | BK 2036 UJ | Minibus Inova | Iwan | Padang Sidempuan | 6 |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| 31-01-17 | BK 5634 AG | Minibus Avanza | Suprian | Rantau Prapat | 7 |
| | BK 4575 GA | Minibus Inova | Andi | Rantau Prapat | 6 |
| | BK 2343 UG | Minibus Kijang | Hery | Rantau Prapat | 7 |
| | BK 1932 DJ | Minibus Inova | Bagus | Padang Sidempuan | 6 |
| Total Jumlah Keseluruhan | | | | | 921 |

Dari data penumpang, maka dapat dilakukan perhitungan untuk mendapatkan nilai $y_i x_i$ dan nilai x_1^2, y , yaitu :

1. Untuk mendapatkan nilai $y_i x_i$ pada bulan januari yaitu :

$$y_i x_i = y * x$$

$$11 * 921$$

$$10.131$$

- Untuk mendapatkan nilai x_1^2 pada bulan januari yaitu :

$$x_1^2 = x_1^2$$

$$11^2$$

$$121$$

2. Untuk mendapatkan nilai $y_i x_i$ pada bulan february yaitu :

$$y_i x_i = y * x$$

$$9 * 844$$

$$7.596$$

- Untuk mendapatkan nilai x_1^2 pada bulan february yaitu :

$$x_1^2 = x_1^2$$

$$9^2$$

$$81$$

3. Untuk mendapatkan nilai $y_i x_i$ pada bulan maret yaitu :

$$y_i x_i = y * x$$

$$10 * 921$$

9.210

Untuk mendapatkan nilai x_1^2 pada bulan maret yaitu :

$$x_1^2 = x_1^2$$

$$10^2$$

$$100$$

4. Untuk mendapatkan nilai $y_i x_i$ pada bulan april yaitu :

$$y_i x_i = y * x$$

$$8 * 895$$

$$7.160$$

Untuk mendapatkan nilai x_1^2 pada bulan april yaitu :

$$x_1^2 = x_1^2$$

$$8^2$$

$$64$$

5. Untuk mendapatkan nilai $y_i x_i$ pada bulan mei yaitu :

$$y_i x_i = y * x$$

$$11 * 942$$

$$10.362$$

Untuk mendapatkan nilai x_1^2 pada bulan mei yaitu :

$$x_1^2 = x_1^2$$

$$11^2$$

$$121$$

Berikut ini tabel hasil perincian perhitungan untuk regresi linear.

Tabel 2 Perincian Perhitungan Regresi Linier

| Bulan | x | y | $y_i x_i$ | x_1^2 |
|-----------------|------------|--------------|---------------|------------|
| Januari | 11 | 921 | 10.131 | 121 |
| Februari | 9 | 844 | 7.596 | 81 |
| Maret | 10 | 921 | 9.210 | 100 |
| April | 8 | 895 | 7.160 | 64 |
| Mei | 11 | 942 | 10.362 | 121 |
| Jumlah | 49 | 4.523 | 44.459 | 487 |
| Rat-Rata | 9,8 | 904,6 | | |

Dengan menggunakan hasil perhitungan pada tabel 3.2, maka dapat menghitung gradien β_1 dan konstanta β_0 .

Nilai β_1 , yaitu

$$\beta_1 = \frac{\sum_{i=1}^n Y_i X_i - \frac{(\sum_{i=1}^n y_i)(\sum_{i=1}^n x_i)}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}$$

$$\beta_1 = \frac{44.459 - \frac{(4.523)(49)}{5}}{487 - \frac{(49)^2}{5}} = \frac{136,6}{6,8} = 19,65$$

Sedangkan Nilai β_0 , yaitu

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\beta_0 = 904,6 - (19,65)(9,8) = 712,06$$

Dari persamaan regresi linier yang dicari berbentuk $Y = \beta_0 + \beta_1 X$, maka dari persamaan regresi linier diperoleh persamaan $Y = 712,06 + (19,65)X$. Setelah memperoleh garis regresi linier, jika jumlah hari libur pada bulan juni adalah 8 hari, maka dapat diprediksi jumlah penumpang yang akan berangkat pada bulan juni adalah :

$$\begin{aligned}
 Y &= 712,06 + (19,65) (x) \\
 &= 712,06 + (19,65) (8) \\
 &= 869 \text{ Orang}
 \end{aligned}$$

Maka diprediksi bahwa jumlah penumpang yang akan berangkat pada bulan JUNI adalah **869 Orang**

The 'Data Hari' form contains the following data table:

| No Hari | Nama Bulan | Tahun | Jumlah Hari |
|---------|------------|-------|-------------|
| HRI-001 | Januari | 2017 | 11 |
| HRI-002 | Februari | 2017 | 9 |
| HRI-003 | Maret | 2017 | 10 |
| HRI-004 | April | 2017 | 8 |
| HRI-005 | Mei | 2017 | 11 |

Gambar 1 Form Input Data Hari

The 'Data Penumpang' form contains the following data table:

| No Bulan | Nama Bulan | Tahun | Jumlah Penumpang |
|----------|------------|-------|------------------|
| BLN-001 | Januari | 2017 | 921 |
| BLN-002 | Februari | 2017 | 844 |
| BLN-003 | Maret | 2017 | 921 |
| BLN-004 | April | 2017 | 895 |
| BLN-005 | Mei | 2017 | 942 |
| BLN-006 | Juni | 2017 | 869 |

Gambar 2 Form Input Data Penumpang

The 'Data Prediksi' form displays the following regression analysis table:

| Bulan | x | y | YiXi | Xi^2 |
|----------|----|-----|-------|------|
| Januari | 11 | 921 | 10131 | 121 |
| Februari | 9 | 844 | 7596 | 81 |
| Maret | 10 | 921 | 9210 | 100 |
| April | 8 | 895 | 7160 | 64 |
| Mei | 11 | 942 | 10362 | 121 |

Summary statistics and controls:

- Jumlah: 49, 4523, 44459, 487
- Rata-Rata: 9.8, 904.6
- Proses Regresi: Nilai B1 (19.65), Nilai B0 (712.03)
- Buttons: Prediksi Penumpang, Simpan, Batal, Keluar
- Input fields: Jumlah Hari Libur Pada Bulan Berikutnya (8), Jumlah Penumpang Pada Bulan Berikutnya (869)

Gambar 3 Form Prediksi

4. Kesimpulan

Setelah melakukan berbagai macam tahapan-tahapan maka kesimpulan yang diperoleh dari adalah sebagai berikut:

1. Dengan menggunakan Sistem Prediksi jumlah penumpang dengan metode regresi linier, membantu pihak perusahaan untuk mengetahui secara detail gambaran prediksi jumlah penumpang yang akan berangkat pada bulan berikutnya. Sehingga pihak perusahaan dapat mempersiapkan unit-unit kendaraan atau armada-armada yang akan digunakan untuk mengangkut penumpang yang akan berangkat.
2. Dengan perancangan dan implementasi Sistem Prediski jumlah penumpang dengan metode regresi linier, pihak perusahaan akan terbantu untuk mengetahui berapa jumlah perkiraan penumpang yang akan berangkat pada bulan berikutnya, sehingga perusahaan dapat melakukan antisipasi dengan menyediakan jumlah armada kendaraan dengan baik dengan tujuan untuk memberikan pelayanan terbaik kepada para konsumen dengan tersedianya armada kendaraan yang cukup untuk melayani penumpang yang akan berangkat.
3. Dengan sistem prediksi jumlah penumpang dengan metode regresi linier dapat mengatasi permasalahan yang ada pada CV. Surya Mandiri Sukses, yaitu memprediksi jumlah penumpang pada kemudian hari.
4. Metode regresi linier dapat menjadi solusi pemecahan masalah prediksi jumlah penumpang

Daftar Pustaka

- [1] Amrin. (Maret 2016). Data Mining Dengan Regresi Linier Berganda Untuk Peramalan Tingkat Inflasi. Jurnal Techno Nusa Mandiri.13(1).75
- [2] Dicky Nofriansyah.,& Gunadi Widi Nurcahyo. (2015). Algoritma Data Mining Dan Pengujian.Yogyakarta: CV Budi Utama
- [3] Hendrayudi. (2011). Dasar-dasar Pemrograman Microsoft Visual Basic 2008. Bandung: CV.YRAMA WIDYA
- [4] Nana Suarna. (2010). Pedoman Panduan Praktikum Microsoft Office Access 2007. Bandung: YRAMA WIDYA
- [5] Prasetyo, (2014). Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta : ANDI.
- [6] Rosa A.S.,& M.Shalahuddin. (2016). Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek. Bandung: Informatika
- [7] Vlandari. (2017). Data Mining Teori Dan Aplikasi Rapidminer. Yogyakarta : GAVA MEDIA