

Algoritma Regresi Linear Sederhana Dalam Mengestimasi Kedatangan dan Kepindahan Penduduk Di Kabupaten Simalungun

Randi Ardiansyah Darmadi¹, Surya Dharma², M. Safii³

^{1,2,3}STIKOM Tunas Bangsa, Pematang Siantar, Indonesia

e-mail: ¹randiardiansyah1802@gmail.com, ²suryadarmaatb@gmail.com,

³m.safii@amiktunasbangsa.ac.id

Abstrak

Sebagai salah satu wilayah di Indonesia, Kabupaten Simalungun turut mengalami perubahan yang signifikan terhadap jumlah kedatangan dan kepindahan penduduk. Pada tahun 2022, tercatat sebanyak 14.998 jiwa pendatang baru dan 16.795 jiwa yang pindah ke Kabupaten Simalungun. Tingginya tingkat pergerakan penduduk di Kabupaten Simalungun menunjukkan adanya permasalahan dalam manajemen populasi di wilayah ini. Maka dari itu dibutuhkan suatu metode yang dapat membantu dalam memahami hubungan antara variabel-variabel yang mempengaruhi kedatangan dan kepindahan penduduk. Data mining adalah salah satu metode yang mampu melakukan analisis data dan mengidentifikasi hubungan statistik dalam data yang relevan. Dalam konteks ini proses analisa data mining akan menggunakan algoritma regresi linear sederhana dengan tujuan estimasi dengan tingkat akurasi yang lebih baik. Hasil dari penelitian ini berupa angka yang menunjukkan hasil estimasi dari kedatangan dan kepindahan penduduk di Kabupaten Simalungun untuk bulan-bulan berikutnya.

Kata kunci: Data mining, Regresi Linear Sederhana, Estimasi

Abstract

As one of the regions in Indonesia, Simalungun Regency has also experienced significant changes in the number of arrivals and population movements. In 2022, 14,998 newcomers and 16,795 people will settle in Simalungun regent. The high level of population mobility in Simalungun Regency indicates a population management problem in this area. Therefore, we need a method that can help understand the relationship between variables that influence the arrival and mobility of residents. Data mining is a method capable of analyzing data and identifying statistical relationships in related data. In this context, data mining analysis will use a simple linear regression algorithm to estimate with a better level of accuracy. The results of this study are presented in the form of estimated arrivals and movements of residents at Simalungun Regency in the following months.

Keywords: Data mining, Simple Linear Regression, Estimation

1. Pendahuluan

Kedatangan dan kepindahan adalah dua fenomena umum yang terjadi di masyarakat. Kedatangan merupakan bentuk datangnya seseorang ke suatu tempat dengan maksud dan tujuan tertentu. Sementara itu kepindahan bisa dimaknakan sebagai bentuk perpindahan dari satu tempat ke tempat lain tanpa mempertimbangkan tempat asal atau tujuan akhir. Sekilas pengertian dari keduanya tampak mirip namun keduanya adalah dua hal yang berbeda.

Dalam pelaksanaannya, kedatangan dan kepindahan melibatkan sejumlah proses dan biasanya akan dilakukan pendataan oleh petugas yang bernaung di bawah Lembaga Pemerintahan yang bertanggung jawab dalam hal Administrasi Penduduk dan Statistik sebagai bentuk pelaporan. Sehingga dari pendataan tersebut dapat diketahui hasil yang akurat dari pergerakan penduduk dari suatu tempat ke tempat lain. Hal tersebut tentu sangat berpengaruh bagi Pemerintah yang mana data pelaporan tersebut bisa digunakan sebagai bahan pertimbangan dalam menentukan kebijakan serta menjadi referensi dalam melakukan estimasi terhadap jumlah pelaporan kedatangan dan kepindahan penduduk di masa yang akan datang.

Sebagai salah satu wilayah di Indonesia, Kabupaten Simalungun turut mengalami perubahan yang signifikan terhadap jumlah kedatangan dan kepindahan penduduk. Pada tahun 2022, tercatat sebanyak 14.998 jiwa pendatang baru dan 16.795 jiwa yang pindah ke Kabupaten Simalungun. Tingginya tingkat pergerakan penduduk di Kabupaten Simalungun menunjukkan adanya permasalahan dalam manajemen populasi di wilayah ini. Hal tersebut dapat berdampak pada akurasi dan keandalan data yang dihasilkan. Maka dari itu dibutuhkan suatu metode yang dapat membantu dalam memahami hubungan antara variabel-variabel yang mempengaruhi kedatangan dan kepindahan penduduk.

Data mining merupakan proses menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* dalam mengekstraksi dan proses identifikasi informasi yang bermanfaat dan pengetahuan yang terikat dari berbagai basis data besar [1]. Data mining memiliki banyak metode atau algoritma dalam pencarian informasinya. Salah satunya adalah algoritma Regresi Linear Sederhana yang menjelaskan tentang hubungan sebab akibat antara faktor penyebab (x) terhadap variabel akibatnya [2]. Dalam kasus ini, Regresi Linear Sederhana digunakan untuk memprediksi jumlah kepindahan (y) berdasarkan angka estimasi terhadap kedatangan (x).

2. Metode Penelitian

2.1. Data Mining

Data mining atau yang dikenal juga sebagai *Knowledge Discovery in Database* (KDD) merupakan aktivitas yang meliputi pengumpulan, pemakaian data historis untuk menemukan pola dalam set data yang besar [3]. Para ahli berpendapat bahwa data mining merupakan bentuk analisa atas observasi data dalam jumlah besar untuk menemukan hubungan yang tidak diketahui dan metode-metode baru untuk meringkas data agar lebih mudah untuk dipahami. Beberapa pengelompokan tugas dalam data mining antara lain:

1. Klasifikasi (*Classification*)

Klasifikasi adalah teknik dalam menilai objek data untuk memasukkannya ke dalam kelas tertentu dari jumlah kelas yang tersedia. Klasifikasi membangun model berdasarkan data latih kemudian menggunakan model tersebut untuk mengklasifikasi data yang baru [1]. Contohnya pada pengelompokan bermain badminton yang mana variabelnya yaitu cuaca, kelembapan, dan berangin. Algoritma yang biasanya digunakan dalam klasifikasi adalah C4.5, *Naïve Bayes*, *Cart*, dan lainnya.

2. Deskripsi (*Description*)

Deskripsi biasanya digunakan dalam pencarian cara dalam menemukan pola dan kecenderungan dalam data [4]. Contohnya terdapat pada supermarket, yang mana pelanggan membeli produk X dan Y di tempat yang sama atau berdekatan sehingga pelanggan tidak kesulitan saat kembali membeli produk tersebut (Buulolo, 2020). Algoritma yang biasanya digunakan dalam deskripsi adalah algoritma Apriori.

3. Prediksi (*Prediction*)

Dalam data mining, prediksi digunakan untuk memperkirakan suatu nilai yang akan datang berdasarkan data-data sebelumnya [5]. Contohnya pada harga gula pasir dalam lima tahun mendatang. Algoritma yang biasa digunakan dalam prediksi diantaranya Nearest Neighbor, *Naïve Bayes*, C4.5, dan lainnya.

4. Klusterisasi (*Clustering*)

Klasterisasi merupakan pengelompokan sejumlah data ke dalam kluster agar berisi data yang semirip mungkin dan berbeda dengan objek kluster lainnya [6]. Contohnya pada pengelompokan pada mahasiswa penerima beasiswa berdasarkan keaktifan, prestasi, kondisi keuangan, dan IPK. Algoritma yang biasanya dipakai dalam klasterisasi diantaranya *K-Means*, *K-Medoids*, dan *K-Nearest Neighbor*.

5. Asosiasi (*Association*)

Asosiasi bertujuan untuk mendapatkan pengelompokan atribut ke dalam segmen berdasarkan similaritas [5]. Biasanya asosiasi digunakan untuk menemukan barang yang dibeli dan tidak pernah dibeli. Algoritma yang biasanya digunakan dalam asosiasi yakni *Association Rule*.

6. Estimasi (*Estimation*)

Estimasi digunakan untuk menerka suatu nilai yang belum diketahui. Contohnya ketika menerka penghasilan seseorang. Algoritma yang biasa digunakan adalah *Regresi Linear*.

2.2. Regresi Linear Sederhana

Regresi Linear Sederhana bertujuan untuk membuat estimasi berdasarkan nilai variabel yang telah diketahui. Analisis regresi melibatkan satu variabel independen (x) dan satu variabel dependen (y) [7]. Regresi linear sederhana umumnya menggunakan rumus:

$$Y = a + bX \quad (1)$$

Keterangan:

- Y = Variabel dependen
- X = Variabel independen
- a = Konstanta
- b = Gradien garis

Penentuan konstanta a dan gradien garis b dalam regresi linear sederhana dapat dilakukan dengan menggunakan metode kuadrat terkecil. Sehingga penentuan a dan b dapat dirumuskan:

$$a = \frac{\sum y}{n} - b \frac{\sum x}{n} \quad (2)$$

Keterangan:

- a = Konstanta
- $\sum y$ = Total y
- $\sum x$ = Total x
- n = Jumlah indeks
- b = Gradien garis

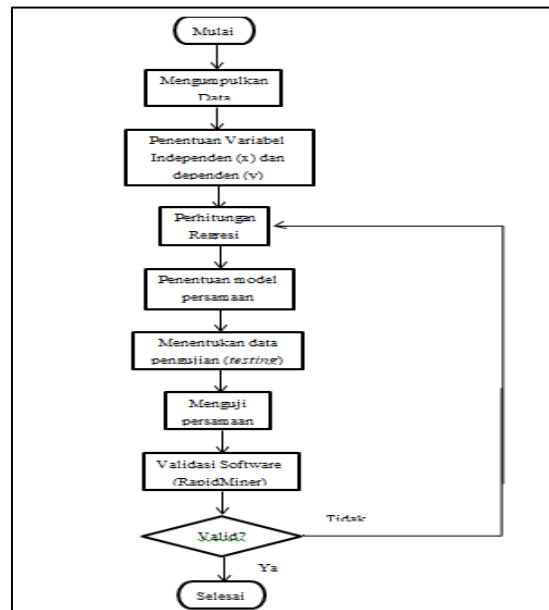
$$b = \frac{n \sum xy - (\sum x)(\sum y)}{(n \sum x^2) - (\sum x)^2} \quad (3)$$

Keterangan:

- b = Gradien garis
- n = Jumlah indeks
- $\sum xy$ = Total xy
- $\sum x$ = Total x
- $\sum y$ = Total y
- $\sum x^2$ = Total x^2

2.3. Rancangan Penelitian

Penelitian ini bertujuan untuk menerapkan algoritma regresi linear sederhana pada data mining untuk menganalisis hubungan antara variabel independen (x) dan variabel independen (y) melalui data set yang telah dikumpulkan. Berikut adalah kerangka penelitian yang akan diikuti:



Gambar 1. Flowchart Regresi Linear Sederhana

3. Hasil dan Pembahasan

3.1. Perhitungan Manual

Data menjadi salah satu faktor penting dalam penelitian ini. Data yang digunakan adalah data pelaporan jumlah kedatangan dan perpindahan penduduk di Kabupaten Simalungun untuk periode tahun 2022 dari bulan Januari sampai Desember yang mana data tersebut diperoleh dari situs BPS Kabupaten Simalungun. Berikut dilampirkan data set tersebut pada tabel 1 di bawah.

Tabel 1. Data pelaporan kedatangan dan perpindahan penduduk di Kabupaten Simalungun

Bulan	Kedatangan (x)	Kepindahan (y)
Januari	1288	989
Februari	1168	814
Maret	1313	886
April	1114	523
Mei	881	1369
Juni	1301	1837
Juli	1446	1948
Agustus	1412	1799
September	1473	1926
Oktober	1254	1640
November	1230	1592
Desember	1118	1472

Melalui tabel di atas maka akan ditentukan data kedatangan sebagai variabel independen (x) dan data perpindahan sebagai variabel dependen (y). Kemudian dengan data di atas maka dilakukan perhitungan untuk mencari nilai x^2 , y^2 , xy , serta total dari masing-masing variabel.

Tabel 2. Perhitungan x^2 , y^2 , xy

Bulan	Kedatangan (x)	Kepindahan (y)	xy	x^2	y^2
Januari	1288	989	1273832	1658944	978121
Februari	1168	814	950752	1364224	662596
Maret	1313	886	1163318	1723969	784996
April	1114	523	582622	1240996	273529
Mei	881	1369	1206089	776161	1874161
Juni	1301	1837	2389937	1692601	3374569
Juli	1446	1948	2816808	2090916	3794704
Agustus	1412	1799	2540188	1993744	3236401
September	1473	1926	2836998	2169729	3709476
Oktober	1254	1640	2056560	1572516	2689600
November	1230	1592	1958160	1512900	2534464
Desember	1118	1472	1645696	1249924	2166784
Total	14998	16795	21420960	19046624	26079401

Maka langkah selanjutnya adalah melakukan perhitungan nilai a dan b berdasarkan rumus di atas. Alur perhitungan dimulai dari penentuan nilai b (gadien garis), kemudian dilanjutkan dengan menentukan nilai a (nilai konstanta). Setelah kedua nilai didapatkan, maka perhitungan estimasi regresi linear sederhana dapat dilakukan. Berikut adalah proses perhitungannya.

1. Menghitung nilai b:

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

$$b = \frac{(12 \times 21420960) - (14998)(16795)}{(12 \times 19046624) - (14998)^2}$$

$$b = \frac{5160110}{3619484}$$

$$b = 1,426$$

2. Menghitung nilai a:

$$a = \frac{\sum y}{n} - b \frac{\sum x}{n}$$

$$a = \frac{16795}{12} - 1,426 \frac{14998}{12}$$

$$a = -382,239$$

3. Setelah mendapatkan nilai a dan b, maka dapat ditentukan persamaan regresi linear.

$$Y = a + bX$$

$$Y = -382,239 + 1,426 X$$

Melalui persamaan di atas maka hasil estimasi terhadap jumlah kepindahan dapat diperoleh. Langkah berikutnya adalah melakukan pengujian dengan memasukkan data sampel

pada nilai X. Data sampel berisi data jumlah kedatangan yang nilainya dipilih secara acak untuk periode tahun 2023 dari bulan Januari hingga Juli. Berikut adalah data sampel yang digunakan termuat pada tabel 3 di bawah.

Tabel 3. Data sampel

Bulan	Kedatangan (x)	Kepindahan (y)
Januari	1562	?
Februari	947	?
Maret	1331	?
April	1092	?
Mei	1802	?
Juni	1544	?

Dengan menggunakan persamaan dan data sampel di atas, nilai estimasi yang didapatkan untuk jumlah kepindahan adalah:

- Januari

$$Y = a + bX$$

$$Y = -382,239 + 1,426 (1562)$$

$$Y = 1844,623$$

$$Y = 1845$$
- Februari

$$Y = a + bX$$

$$Y = -382,239 + 1,426 (947)$$

$$Y = 967,85$$

$$Y = 968$$
- Maret

$$Y = a + bX$$

$$Y = -382,239 + 1,426 (1331)$$

$$Y = 1515,298$$

$$Y = 1515$$
- April

$$Y = a + bX$$

$$Y = -382,239 + 1,426 (1092)$$

$$Y = 1174,569$$

$$Y = 1175$$
- Mei

$$Y = a + bX$$

$$Y = -382,239 + 1,426 (1802)$$

$$Y = 2186,779$$

$$Y = 2187$$
- Juni

$$Y = a + bX$$

$$Y = -382,239 + 1,426 (1544)$$

$$Y = 1818,961$$

$$Y = 1819$$

Hasil perhitungan tersebut dapat dijabarkan pada tabel 4 di bawah.

Tabel 4. Hasil estimasi

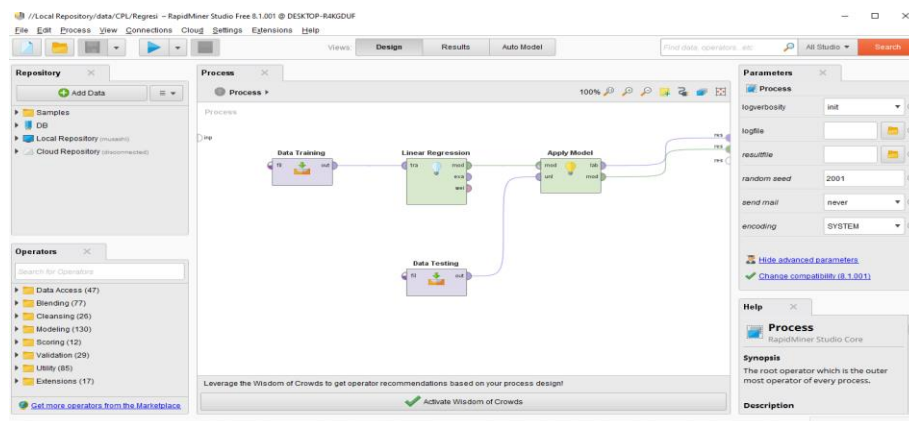
Bulan	Kedatangan	Kepindahan
Januari	1562	1845
Februari	947	968
Maret	1331	1515
April	1092	1175
Mei	1802	2187
Juni	1544	1819

3.2. Validasi Sistem (RapidMiner Studio)

Keakuratan perhitungan menjadi hal yang penting bagi setiap orang demi mendapatkan kepercayaan akan perhitungan yang telah dibuat. Tidak sedikit orang yang rela melakukan validasi berulang kali dan dengan berbagai cara demi mendapatkan nilai yang akurat tersebut. Seiring berkembangnya teknologi komputer tentunya mendapatkan keakuratan data menjadi lebih mudah. Contohnya dengan menggunakan suatu *software* berbasis analisis data. Demi mendapatkan keakuratan perhitungan di atas penulis menggunakan aplikasi *RapidMiner Studio* versi 8.1.001 yang mana aplikasi tersebut cukup terkenal dan paling sering digunakan untuk memecahkan masalah algoritma dalam data mining.

Dalam memproses hasil estimasi algoritma regresi linear sederhana, diperlukan operator-operator dalam *RapidMiner* guna memperoleh hasil. Ada pun operator yang digunakan dalam menjalankan algoritma regresi linear sederhana antara lain:

- Read Excel untuk data *training*
- Read Excel untuk data *testing*
- Linear Regression
- Apply Model



Gambar 2. Operator-operator yang digunakan

Dengan menggunakan operator di atas, maka hasil estimasi diperoleh. Berikut adalah hasil yang diberikan oleh RapidMiner.



Gambar 3. Hasil description dari operator Linear Regression

Row No.	Bulan	Kepindahan	prediction(K...	Kedatangan
1	Januari	?	1844.623	1562
2	Februari	?	967.850	947
3	Maret	?	1515.298	1331
4	April	?	1174.569	1092
5	Mei	?	2186.779	1802
6	Juni	?	1818.961	1544

Gambar 4. Hasil dari operator Apply Model yang memuat hasil estimasi

Setelah melakukan serangkaian validasi dengan menggunakan aplikasi *RapidMiner Studio* 8.1.001 maka dapat disimpulkan bahwa perhitungan manual adalah benar dikarenakan telah sesuai dengan yang telah divalidasi oleh sistem. Tampak jelas bahwa persamaan regresi linear dan hasil estimasi dari data *testing* telah sesuai dengan perhitungan manual. Maka dengan ini dinyatakan bahwa algoritma regresi linear sederhana dapat digunakan untuk mengestimasi terhadap kedatangan dan kepindahan penduduk di Kabupaten Simalungun.

4. Kesimpulan

Setelah melalui serangkaian perhitungan di atas maka dapat ditarik kesimpulan bahwasanya algoritma regresi linear sederhana dapat digunakan untuk mengestimasi jumlah kedatangan dan kepindahan penduduk di Kabupaten Simalungun. Setelah melalui serangkaian perhitungan berdasarkan data set kedatangan dan kepindahan penduduk di Kabupaten Simalungun periode tahun 2022 maka ditetapkan $Y = -389,239 + 1,426 X$ sebagai bentuk persamaan dalam melakukan estimasi terhadap kedatangan dan kepindahan penduduk di Kabupaten Simalungun untuk periode bulan yang datang. Berdasarkan angka pada data *testing* maka dapat diestimasi jika jumlah kedatangan di bulan Januari tahun 2023 adalah sebesar 1562 maka dapat diestimasi jumlah kepindahannya sebesar 1845 jiwa. Jika jumlah kedatangan di bulan Februari tahun 2023 adalah sebesar 947 maka jumlah kepindahannya sebesar 968 jiwa dan seterusnya. Penelitian ini diharapkan dapat memberikan pemahaman bagaimana algoritma dalam Data mining, yakni regresi linear sederhana dapat melakukan estimasi khususnya terhadap jumlah kedatangan dan kepindahan penduduk. Penelitian ini diharapkan dapat digunakan sebagai acuan dalam melakukan studi kasus yang berbeda dan lebih bermanfaat.

Daftar Pustaka

- [1] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [2] A. S. Lubis, Tugiono, and Hafizah, "Jurnal Sistem Informasi Jurnal Sistem Informasi," *J. Sist. Inf. TGD*, vol. 1, no. 6, pp. 888–897, 2022.
- [3] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Eeccis*, vol. 7, no. 1, pp. 59–64, 2013, doi: 10.1038/hdy.2009.180.
- [4] Yuli Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.
- [5] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review," *Fakt. Exacta*, vol. 13, no. 1, p. 35, 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [6] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi*, vol. 7, no. 3, p. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
- [7] F. Ginting, E. Buulolo, and E. R. Siagian, "Implementasi Algoritma Regresi Linear Sederhana Dalam Memprediksi Besaran Pendapatan Daerah (Studi Kasus: Dinas Pendapatan Kab. Deli Serdang)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 274–279, 2019, doi: 10.30865/komik.v3i1.1602.