

Penerapan Metode Regresi Linear Berganda dalam Mengestimasi Jumlah Penduduk pada Kabupaten Samosir

Robin Pangihutan Simbolon¹, Sardo P Sipayung²

Fakultas Ilmu Komputer, Universitas Katolik Santo Thomas Medan, Indonesia

Email : simbolonrobin06@gmail.com¹, pinsarsiphom@gmail.com²

Abstrak

Penduduk adalah sekelompok individu yang tinggal di wilayah Republik Indonesia minimal enam bulan atau lebih, atau yang bertujuan menetap di suatu kawasan meskipun hanya kurang dari enam bulan, dan saling berinteraksi. Pertumbuhan populasi terus meningkat dengan cepat, memerlukan tindakan preventif untuk mengendalikan laju pertumbuhan. Pertumbuhan penduduk di suatu daerah memiliki peran vital dalam kemajuan wilayah tersebut. Penelitian ini bertujuan untuk mengestimasi jumlah penduduk Kabupaten Samosir menggunakan teknik estimasi Regresi Linier Berganda. Metode ini dipilih karena mampu membuat prediksi dengan memanfaatkan data historis tentang pertumbuhan populasi, sehingga dapat membentuk pola hubungan yang bermanfaat. Pengolahan data dilakukan dengan menggunakan perangkat lunak Rapidminer, dengan dataset dari tahun 2013-2017 yang mencakup jumlah penduduk laki-laki dan perempuan di Samosir. Data penelitian diperoleh dari Badan Pusat Statistik (BPS) Kabupaten Samosir. Hasil penelitian ini dapat memberikan masukan penting bagi BPS untuk mengantisipasi pertumbuhan penduduk, serta menegaskan keandalan Regresi Linier Berganda dalam melakukan estimasi populasi.

Kata Kunci: Penduduk; Kabupaten Samosir; Data Mining; Regresi Linier, Rapidminer

Abstract

The population consists of individuals residing in the territory of the Republic of Indonesia for at least six months or more, or those intending to settle in an area, even if for less than six months, and interact with each other. Population growth is rapidly increasing, requiring preventive measures to control its rate. Population growth in an area plays a vital role in the development of that region. This research aims to estimate the population of Samosir District using Multiple Linear Regression estimation technique. This method is chosen because it can make predictions by utilizing historical data on population growth, thus forming useful relationship patterns. Data processing is done using Rapidminer software, with a dataset from 2013-2017 including the population count of males and females in Samosir. The research data is obtained from the Central Statistics Agency (BPS) of Samosir District. The findings of this research can provide valuable input for BPS to anticipate population growth and affirm the reliability of Multiple Linear Regression in population estimation.

Keywords: Population; Samosir District; Data Mining; Linear Regression, Rapidminer

PENDAHULUAN

Penduduk merupakan sekumpulan individu yang berdomisili di wilayah geografis Republik Indonesia selama enam bulan atau lebih dan atau mereka yang berdomisili kurang dari enam bulan tetapi bertujuan untuk menetap tinggal di suatu 427ariabl dan saling berinteraksi satu sama lain. Tingkat pertumbuhan penduduk di sebuah kecamatan sangat penting demi perkembangan sebuah kecamatan dan bisa menjadi tolak ukur apakah sebuah kecamatan dapat berkembang dengan baik dari segi penduduknya. Bertambahnya jumlah penduduk suatu daerah semakin tinggi perubahan daerah tersebut dan semakin banyak pula persoalan yang di hadapi oleh suatu wilayah. Tingkat pertumbuhan penduduk yang terlalu tinggi akan beresiko menimbulkan berbagai masalah bagi daerah tersebut, seperti 427ariabl pengangguran yang tinggi, kemiskinan dan kekurangan pangan yang mengakibatkan kelaparan. Faktor yang mempengaruhi pertumbuhan penduduk antara lain: kelahiran (fertilitas), kematian (mortalitas) dan juga adanya migrasi penduduk. Secara terus menerus penduduk akan dipengaruhi oleh bertambahnya jumlah kelahiran bayi (fertilitas), tetapi secara bersamaan pula akan dikurangi oleh jumlah kematian (mortalitas) yang terjadi pada semua golongan umur.

Datamining merupakan metode pengolahan data untuk menemukan pola tersembunyi dari data lama sehingga menghasilkan data baru yang dapat dipahami dan bisa digunakan sebagai rujukan atau refererensi. Ilmu ini mempelajari tentang metode untuk mengaplikasikan dalam penganalisaan untuk memecahkan suatu masalah dengan jumlah data yang besar. Salah satu bagian dari Datamining adalah regresi linier. Regresi linier merupakan sebuah metode yang termasuk dalam data mining yang berfungsi untuk menguji, memperkirakan hubungan antara 427variabel bebas dan 427variabel terikat dengan mencari koefisien korelasinya. Regresi linier memiliki kemampuan untuk membuat suatu estimasi/prediksi dengan memanfaatkan data-data lama. Sehingga dapat dihasilkan suatu pola hubungan antara atribut-atribut yang mempengaruhi laju pertumbuhan penduduk. Data yang digunakan pada penelitian ini diakses dari situs resmi pemerintah Indonesia yakni Badan Pusat Statistik (BPS) Kabupaten Samosir melalui websitenya <https://samosirkab.bps.go.id> . Dalam hal ini, peneliti mengangkat topik Implementasi Linear Regression Dalam Memprediksi Proyeksi Penduduk Kabupaten Samosir. Penelitian yang dilakukan bertujuan untuk memberikan pengetahuan tentang penggunaan tools Rapidminer dalam proses pengujian data dengan algoritma regresi linier.

METODE PENELITIAN

Metode Regresi Linier

Regresi Linier adalah metode yang digunakan untuk menguji hubungan antara variabel bebas dan variabel terikat. Ini melibatkan prediksi menggunakan garis lurus untuk menggambarkan hubungan antara dua variabel atau lebih. Variabel adalah besaran yang nilainya berubah-ubah, terbagi menjadi dua jenis: variabel pemberi pengaruh yang dapat dianggap sebagai sebab, dan variabel terpengaruh yang merupakan akibatnya. Peramalan ini didasarkan pada asumsi bahwa pola pertumbuhan data historis bersifat linier, meskipun tidak selalu sepenuhnya linier. Model ini mendekati pola pertumbuhan dengan menggambarkan hubungan-hubungan yang terkait dalam suatu keadaan. Regresi linier berganda adalah model persamaan yang

menjelaskan hubungan antara satu variabel tak bebas (Y) dengan dua atau lebih variabel bebas (X_1, X_2, \dots, X_n), digunakan untuk memprediksi nilai variabel tak bebas saat nilai-nilai variabel bebasnya diketahui. Selain itu, model ini membantu dalam memahami arah hubungan antara variabel tak bebas dengan variabel bebasnya.

Langkah-langkah metode *Regresi Linier* Berganda.

$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$

Y = Variabel tak bebas (nilai variabel yang akan diprediksi)

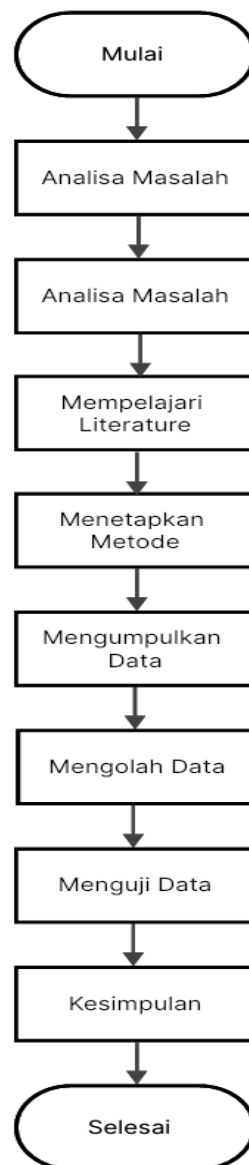
a = konstanta

b_1, b_2, \dots, b_n = nilai koefisien regresi

X_1, X_2, \dots, X_n = variabel bebas

Rancangan Penelitian

Diagram alur penelitian :



Gambar 1. Rancangan Penelitian

Bila terdapat 2 variabel bebas, yaitu X_1 dan X_2 , maka bentuk persamaan regresinya adalah:

$$Y = a + b_1X_1 + b_2X_2$$

Ada beberapa tahapan dalam membuat penyelesaian masalah dengan metode regresi linear berganda dalam yaitu :

1. Mempersiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan variabel bebas dan variabel tidak bebas.
Variabel bebas yaitu :
Jumlah Laki-Laki (X_1)
Jumlah Perempuan (X_2)
Variabel tidak bebas yaitu :
Jumlah Penduduk (Y)
3. Mencari nilai persamaan regresi linear berganda.
 $Y = a + b_1.x_1 + b_2.x_2$
4. Menentukan nilai konstanta dan koefisien regresi.

HASIL DAN PEMBAHASAN

Pengolahan Data

1. Mengetahui jumlah penduduk tahun 2013-2017. Sampel data yang digunakan dalam menentukan estimasi pada data penduduk adalah sebagai berikut :

Tabel 1. Penentuan Pusat Awal

No	Kecamatan	2013	2014	2015	2016	2017
1	Sianjur Mula-Mula	9.311	9.394	9.448	9.501	9.546
2	Harian	8.010	8.070	8.114	8.158	8.195
3	Sitiotio	7.260	7.302	7.341	7.376	7.408
4	Onan Runggu	10.525	10.624	10.687	10.748	10.800
5	Nainggolan	12.074	12.189	12.261	12.332	12.392
6	Palipi	16.392	16.550	16.648	16.744	16.826
7	Ronggur Nihuta	8.514	8.584	8.632	8.680	8.720
8	Pangururan	29.970	30.283	30.468	30.648	30.803
9	Simanindo	19.868	20.069	20.190	20.309	20.409
Jumlah Penduduk :		121.924	123.065	123.789	124.496	125.099

2. Menentukan jumlah data Penduduk Per Tahun Berdasarkan Jenis Kelamin.

Tabel 2. Akumulasi Data Per Tahun Berdasarkan Jenis Kelamin

Tahun	Laki-Laki	Perempuan	Jumlah Penduduk
2013	60,588	61,336	121,924
2014	61,080	61,985	123,065
2015	61,406	62,383	123,789
2016	61,904	62,592	124,496
2017	62,214	62,885	125,099
Jumlah	307,192	311,181	618,373

Karena dalam perhitungan regresi linier berganda banyak dilakukan perkalian dan perpangkatan maka untuk menyederhanakan bilangan akan dibagi 1000 dan tabel ini penentuan X_1 , X_2 dan Y Sehingga menghasilkan tabel 3. sebagai berikut.

Tabel 3. Penentuan X_1 , X_2 Dan Y Serta Penyederhanaan Dengan Di Bagi (1000)

Tahun	X1	X2	Y
2013	60,588	61,336	121,924
2014	61,080	61,985	123,065
2015	61,406	62,383	123,789
2016	61,904	62,592	124,496
2017	62,214	62,885	125,099
Jumlah	307,192	311,181	618,373

3. Persamaan umum algoritma regresi linier berganda, adalah :

$$Y = a_0 + a_1.X_1 + a_2.X_2 + \dots + a_n.X_n$$

Keterangan :

Y = variabel terikat / variabel dependent (nilai yang diprediksi)

a_0, a_1, a_2, a_n = koefisien regresi

X_1, X_2, \dots, X_n = variabel bebas / variabel independent.

Analisis regresi linier berganda digunakan untuk mengestimasi pertumbuhan penduduk sebagai (Y) di Kabupaten Samosir ditinjau dari 2 variabel yaitu Jumlah Laki-Laki (X_1) dan Jumlah Perempuan (X_2) yang akan dilakukan prediksi menggunakan analisis regresi linier berganda. Untuk mencari nilai konstanta dan variabel regresi setiap variabel bebas dapat diperoleh dengan menggunakan rumus regresi linear berganda seperti tabel 4 sebagai berikut :

Tabel 4. Ikhtisar Penentuan

Tahun	X1Y	X2Y	X1X2	X1 ²	X2 ²	Y ²
2013	7387,131312	7478,330464	3716,225568	3.670,906	3.762,105	14.865,462
2014	7.516,810	7628,184025	3786,0438	3.730,766	3.842,140	15.144,994
2015	7601,387334	7722,329187	3830,690498	3.770,697	3.891,639	15.323,717
2016	7706,800384	7792,453632	3874,695168	3.832,105	3.917,758	15.499,254
2017	7782,909186	7866,850615	3912,32739	3.870,582	3.954,523	15.649,760
Jumlah	37995,0384 2	38488,1479 2	19119,9824 2	18.875,05 6	19.368,16 5	76.483,18 6

Untuk memperkecil angka yang terdapat pada tabel diatas, maka dilakukan pembagian dengan bagi 10 setiap data pada tabel diatas, sehingga hasilnya menjadi seperti tabel 5. berikut ini :

Tabel 5. Ikhtisar Penentuan Yang di Sederhanakan

Tahun	X1Y	X2Y	X1X2	X1 ²	X2 ²	Y ²
2013	738,7131312	747,8330464	371,6225568	367,0905744	376,2104896	1486,546178
2014	751,68102	762,8184025	378,60438	373,07664	384,2140225	1514,499423
2015	760,1387334	772,2329187	383,0690498	377,0696836	389,1638689	1532,371652
2016	770,6800384	779,2453632	387,4695168	383,2105216	391,7758464	1549,925402
2017	778,2909186	786,6850615	391,232739	387,0581796	395,4523225	1564,97598
Jumlah	3799,503842	3848,814792	1911,998242	1887,505599	1936,81655	7648,318634

Untuk memperoleh koefisien regresi a , b_1 dan b_2 dapat diperoleh dengan cara simultan dari tiga persamaan sebagai berikut:

$$a_n + b_1 \sum X_1 + b_2 \sum X_2 = \sum Y$$

$$a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 = \sum X_1 Y$$

$$a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 = \sum X_2 Y$$

Masukkan angka yang telah didapat pada ikhtisar perhitungan tabel 4 dan nilai $\sum X_1$, $\sum X_2$ dan $\sum Y$ (dari Tabel.4) sehingga diperoleh hasil :

$$a_5 + b_1(307,192) + b_2(311,181) = 618,373$$

$$a(307,192) + b_1(1887,505599) + b_2(1911,998242) = 3799,503841$$

$$a(311,181) + b_1(1911,998242) + b_2(1936,81655) = 3848,814792$$

Ketiga persamaan diatas diselesaikan hingga memperoleh nilai pada a, b₁, dan b₂. Langkah langkah penyelesaiannya adalah sebagai berikut :

a. Persamaan 1 dan 2 di eliminasi :

$$\begin{array}{rcl} a5+ b_1(307,192) + b_2(311,181) & = & 618,373 \quad | \quad \times \\ 307,192 & & \\ \hline a(307,192) + b_1(1887,505599) + b_2(1911,998242) & = & 3799,503841 \quad | \times 5 \\ \hline 1535,96 a + 94,367 b_1 + 95592,31375 b_2 & = & 189959,2386 \\ 1535,96 a + 9437,5 b_1 + 9559,99121 b_2 & = & 18997,51921 - \\ \hline 8081,608 b_1 + 8150,733 b_2 & = & 16232,342 \end{array}$$

b. Persamaan 1 dan 3 di eliminasi :

$$\begin{array}{rcl} a5+ b_1(307,192) + b_2(311,181) & = & 618,373 \quad | \quad \times \\ 311,181 & & \\ \hline a(311,181) + b_1(1911,998242) + b_2(1936,81655) & = & 3848,814792 \quad | \times 5 \\ \hline 1555,905 a + 95592,31375 b_1 + 96833,61476 b_2 & = & 192425,9285 \\ 1555,905 a + 9559,99121 b_1 + 9684,08275 b_2 & = & 19244,07396 - \\ \hline 86032,32254 b_1 + 87149,53201 b_2 & = & 173181,8546 \end{array}$$

c. Persamaan 4 dan 5 di eliminasi :

$$\begin{array}{rcl} 8081,608 b_1 + 8150,733 b_2 & = & 16232,342 \quad | \times 86032,322542 \\ 86032,32254 b_1 + 87149,53201 b_2 & = & 173181,8546 \quad | \times 84929,396869 \\ \hline 7306673265 b_1 + 7401560522 b_2 & = & 14708233787 \\ 7306673265 b_1 + 7401557191 b_2 & = & 14708230456 - \\ \hline 3330,8609 b_2 & = & 3330,8609 \\ b_2 & = & 3330,8609 / 3330,8609 \\ b_2 & = & 1 \end{array}$$

d. Masukkan nilai b₂ kedalam persamaan 4 dengan mensubsitusikan :

$$\begin{array}{rcl} 84929,4 b_1 + 86032,32254 b_2 & = & 170961,7194 \\ 84929,4 b_1 + 86032,32254 (1) & = & 170961,7194 \\ 84929,4 b_1 + 86032,32254 & = & 170961,7194 \\ 84929,4 b_1 & = & 256994,04194 \\ 84929,4 b_1 & = & 256994,04194 \\ b_1 & = & 256994,04194 / 84929,4 \\ b_1 & = & 3,025972654 \end{array}$$

e. Masukkan nilai b₁ dan b₂ ke persamaan 1 dengan mensubsitusikan :

$$\begin{array}{rcl} 5a+ b_1(307,192) + b_2(311,181) & = & 618,373 \\ 5a + (3,025972654) * (307,192) + (1) * (311,181) & = & 618,373 \\ 5a + 929,5545915 + 311,181 & = & 618,373 \\ 5a + 1240,735592 & = & 618,373 \\ 5a & = & 618,373 - 1240,735592 \\ 5a & = & -622,362592 \\ a & = & -622,362592 / 5 \\ a & = & -124,4725184 \end{array}$$

Jadi sekarang telah didapat nilai a, b₁, b₂ dimana nilai-nilainya adalah sebagai berikut :

$$\begin{aligned} a &= -124,4725184 \\ b_1 &= 3,025972654 \\ b_2 &= 1 \end{aligned}$$

Sehingga menghasilkan persamaan regresi sebagai berikut :

$$Y = -124,4725184 + 3,025972654 X_1 + 1X_2$$

Setelah persamaan regresi linear didapatkan, maka untuk mengestimasi pertumbuhan penduduk dapat diperoleh dengan mudah. Dengan memasukkan nilai x_1 dan x_2 pada periode tahun terakhir (tahun 2017) yaitu nilai x_1 sebesar 62,214 dan nilai x_2 sebesar 62,885. Maka menghitung mengestimasi pertumbuhan penduduk adalah dengan menggunakan persamaan regresiliner berganda diatas, sehingga:

$$\begin{aligned} Y &= a + b_1.x_1 + b_2.x_2 \\ &= -124,4725184 + 3,025972654 * (62,214) + 1 * (62,885) \\ &= -124,4725184 + 188,2578627 + 62,885 \\ &= 126,670 \end{aligned}$$

Maka hasil dari estimasi (perkiraan) pertumbuhan penduduk tahun 2018 adalah sebanyak 62,938 penduduk. Diperoleh hasil untuk pertumbuhan penduduk tahun 2018 sampai 2024 pada tabel 6. sebagai berikut :

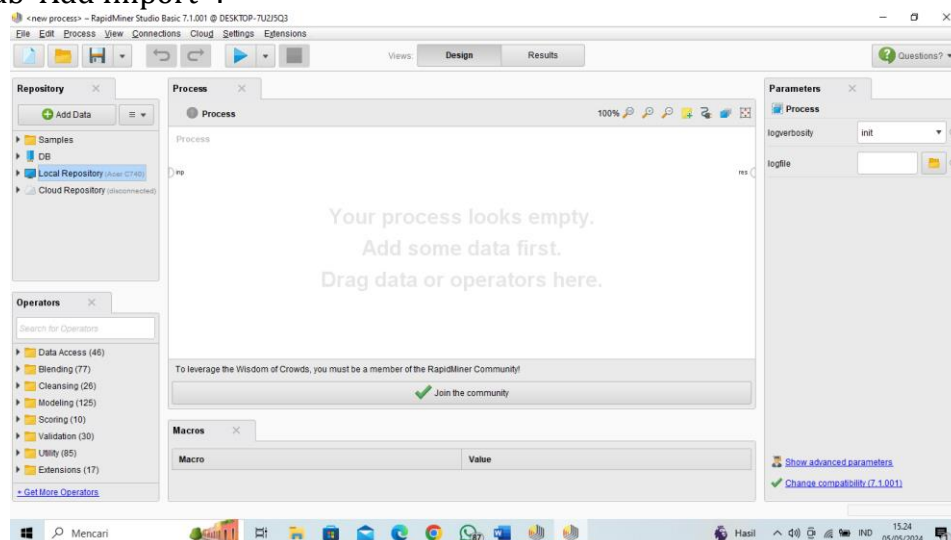
Tabel 6. Hasil Estimasi untuk Tahun 2018-2022

Tahun	X1	X2	Y
2018	62,938	63,732	126,670
2019	64,382	65,325	129,707
2020	67,330	68,341	135,671
2021	73,307	74,301	147,608
2022	85,181	86,473	171,654
2023	109,125	110,631	219,756
2024	157,006	159,361	316,367
Jumlah	619,269	628,164	1.247,433

Implementasi pada Rapid Miner

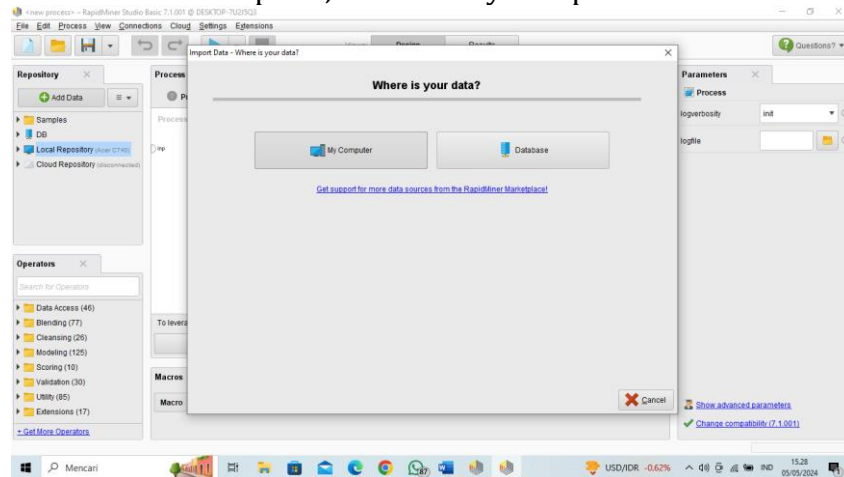
1. Import data kedalam *RapidMiner* dalam bentuk *Sheet Excel*.

Pada hal ini data yang akan dieksekusi berupa data excel. Klik pada bagian kiri tab "Add Import".



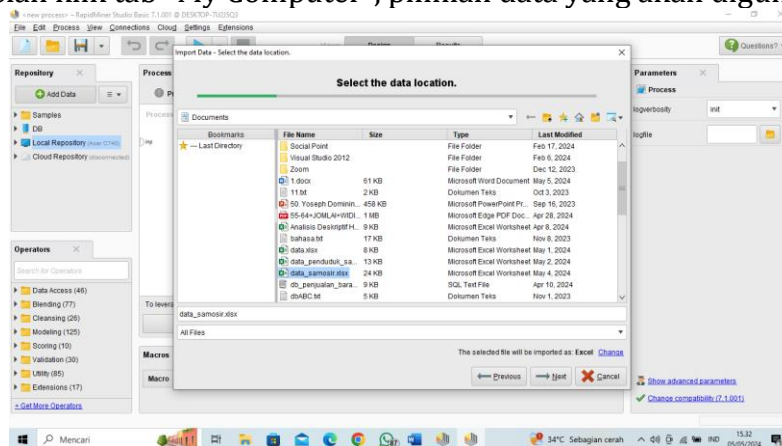
Gambar 2. Add data

Setelah klik tab “Add Import”, klik tab “My Computer” untuk memilih data.



Gambar 3. My Computer

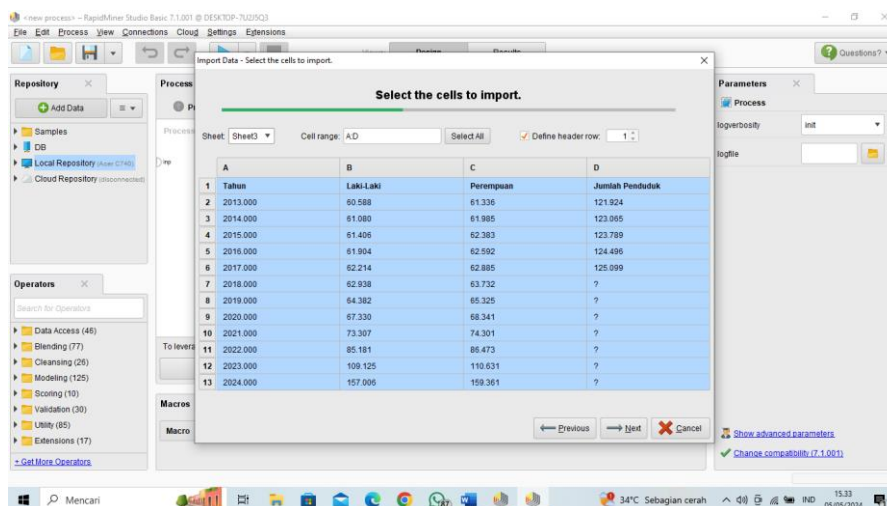
Setelah klik tab “My Computer”, pilihlah data yang akan digunakan



Gambar 4. Data yang digunakan

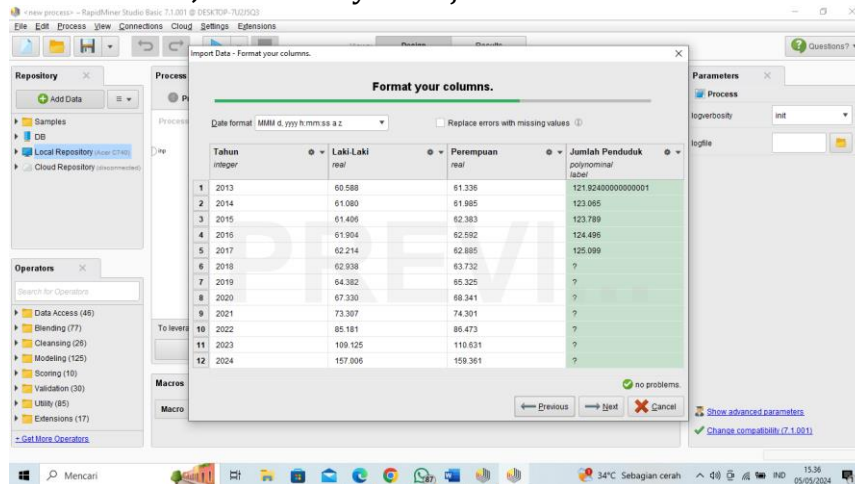
2. Proses

Pilih data pada folder yang akan di import.



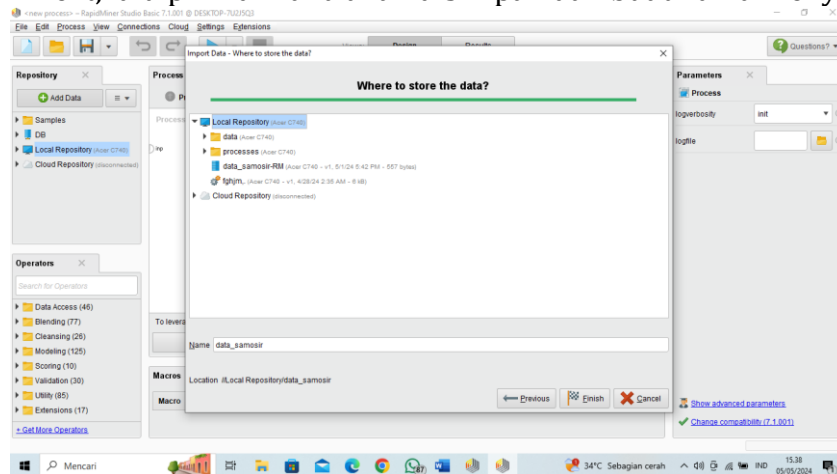
Gambar 5. Alur Proses Import Data

Setelah data yang akan digunakan dipilih, maka klik tombol “Next”. Lalu pada kolom **Jumlah Penduduk**, ubah rolenya menjadi label.



Gambar 6. Role Jumlah Penduduk

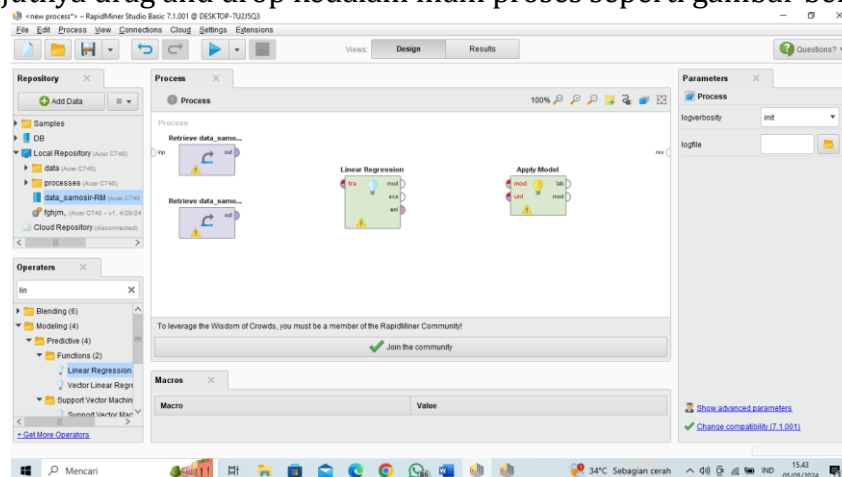
Setelah itu klik next, lalu pilih dimana akan disimpan dan buat nama filenya.



Gambar 7. Penyimpanan

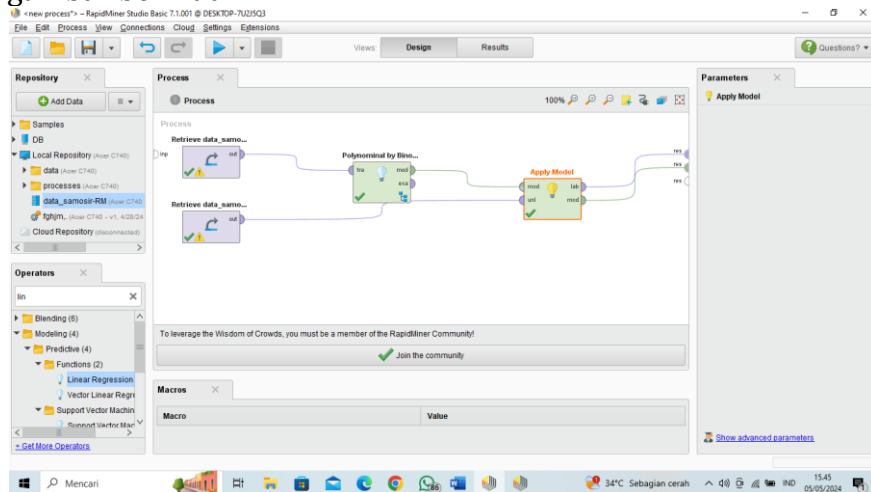
3. Keluaran (Output) Sistem

Selanjutnya drag and drop kedalam main proses seperti gambar berikut.



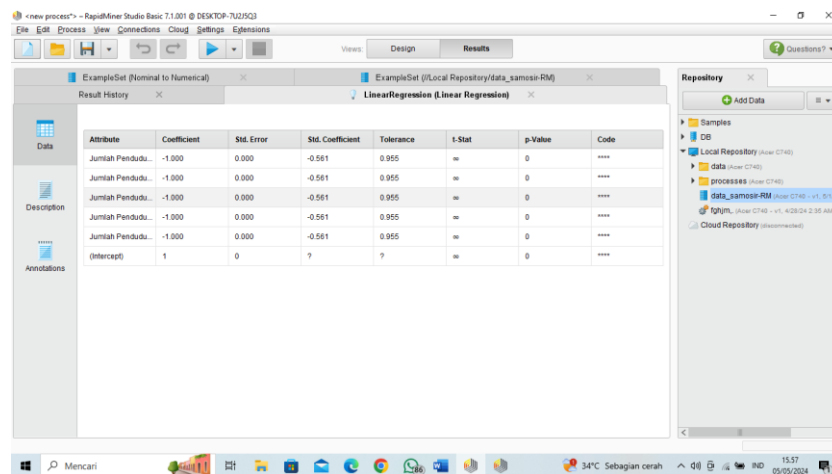
Gambar 8. Drag and Drop Data Penduduk

4. Langkah selanjutnya drag and drop operator Linier Regression dan operator Apply Model lalu hubungkan pada output. Selanjutnya klik run yang berada diposisi tengah atas seperti gambar berikut.

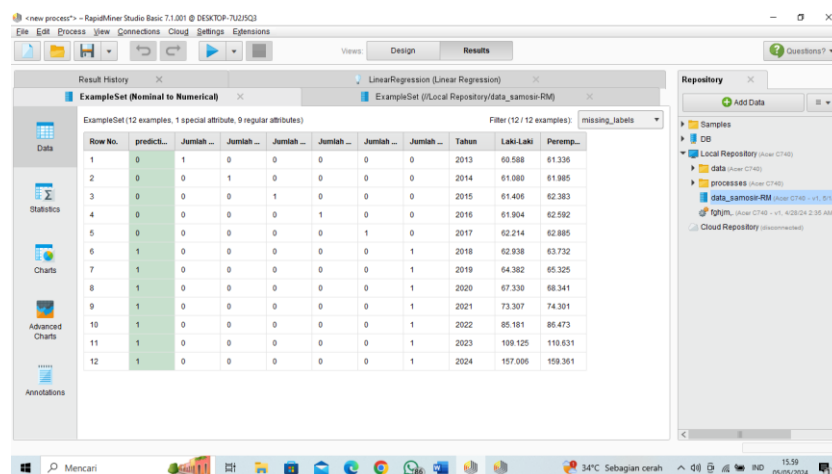


Gambar 9. Drag and Drop Data Operator Linear Regression dan Apply Model

Pada saat tool run diklik akan muncul pohon keputusan seperti pada gambar 11, sebagai berikut.



Gambar 10. Hasil Estimasi Regresi Linier Berganda



Gambar 11. Hasil Prediksi Regresi Linier Berganda

KESIMPULAN

Berdasarkan pembahasan dan hasil yang telah diuraikan diatas dapat disimpulkan bahwa data tersebut diolah menggunakan Microsoft Excel . Dan Hasil estimasi jumlah pertumbuhan penduduk menggunakan metode Regresi Linier Berganda adalah sebanyak 126,670 penduduk.

UCAPAN TERIMA KASIH

Kami ingin mengucapkan terima kasih kepada Badan Pusat Statistik Kabupaten Samosir atas kerjasama yang luar biasa dalam proses estimasi jumlah penduduk. Dukungan dan data yang diberikan telah menjadi landasan yang kuat dalam penerapan Metode Regresi Linear Berganda. Kami juga ingin mengucapkan terima kasih kepada jurnal yang digunakan sebagai referensi utama dalam penelitian ini, yang telah memberikan wawasan yang berharga dalam pengembangan model. Semoga kerjasama ini terus berbuah hasil positif bagi Kabupaten Samosir. Terima kasih atas kontribusi berarti dari Badan Pusat Statistik Kabupaten Samosir dan jurnal yang mendukung.

DAFTAR PUSTAKA

- [1] Sinaga,Widia Ayu Lestari,. Sumarno, S,. Sari, Ika Purnama. 2022. Penerapan Metode Regresi Linier Berganda Untuk Estimasi Jumlah Penduduk Pada Kecamatan Gunung Malela. **JOMLAI, 1(1)**: 55-64
- [2] Candra Adi Rahmat,. Kurniabudi,. Yudi Novianto. 2023. Penerapan Metode Regresi Linier Berganda Untuk Mengestimasi Laju Pertumbuhan Penduduk Kabupaten Musi Banyuasin. **JAKAKOM, 3(1)**: 359-369
- [3] Ayu Wulandari, Agus Perdana Windarto , Hendry Qurniawan. 2022. Penerapan Multiple Regression dalam Mengestimasi Jumlah Pertumbuhan Penduduk di Kecamatan Siantar. **JOMLAI, 1(4)**: 347-354