

Prediksi Pertumbuhan Penduduk Dengan Model *Clustering* Metode Regresi Linear

Aprilia Nur Sa'adah^{1*}, Aswan Supriyadi Sunge², Ahmad Turmudi Zy³
^{1*,2,3}Universitas Pelita Bangsa

*Email: aprilianursaadah7@gmail.com

Abstract

Based on population data from the Central Bureau of Statistics for the City of Cirebon from 2019 to 2022, the birth rate has increased from 2,867 to 3,713. This study aims to determine the population growth rate with the linear regression method. The linear regression algorithm was chosen because it is able to estimate by utilizing old data. So that it can produce a pattern of relationships between attributes that affect the rate of population growth. This study uses prediction techniques with stages in data mining, namely rapidminer tools to predict population growth. The data used is the 2019-2022 dataset which consists of the number of births and deaths in the city of Cirebon. Processing of the data that is used as a dataset is divided into two, namely 90% training data and 10% test data. The results of the tests that have been carried out indicate that the variables or attributes used in this study (births and deaths) have a significant effect on this study. It is proven that the use of linear regression algorithm can give good results with Root Mean Squared Error: 0.998 +/- 0.000 and Squared Error: 0.996 +/- 0.707. The conclusion from the research conducted by applying the linear regression algorithm can be made a prediction based on the functional relationship to the variables or attributes in the data.

Kerywords: Population growth, linear regression, rapidminer

Abstrak

Berdasarkan data kependudukan dari Badan Pusat Statistik Kota Cirebon dari tahun 2019 hingga 2022 tercatat jumlah angka kelahiran meningkat dari 2.867 jiwa menjadi 3.713 jiwa. Penelitian ini bertujuan untuk mengetahui laju pertumbuhan penduduk dengan metode regresi linear. Algoritma regresi linear dipilih karena mampu membuat suatu estimasi dengan memanfaatkan data-data lama. Sehingga dapat dihasilkan suatu pola hubungan antara atribut-atribut yang memengaruhi laju pertumbuhan penduduk. Penelitian ini menggunakan teknik prediksi dengan tahapan pada data mining yaitu *tools* rapidminer untuk memprediksi pertumbuhan penduduk. Data yang digunakan adalah dataset tahun 2019-2022 yang terdiri dari jumlah kelahiran dan kematian di Kota Cirebon. Pengolahan data yang dijadikan data set dibagi menjadi dua yaitu 90% data training dan 10% data testing. Hasil pengujian yang telah dilakukan bahwa variabel atau atribut yang digunakan dalam penelitian ini (kelahiran dan kematian) berpengaruh signifikan terhadap penelitian ini terbukti dengan menggunakan algoritma regresi linear mampu memberikan hasil yang baik dengan nilai *Root Mean Squared Error*: 0.998 +/- 0.000 dan *Squared Error*: 0.996 +/- 0.707. Kesimpulan dari penelitian yang dilakukan dengan menerapkan algoritma regresi linear dapat dilakukan suatu prediksi berdasarkan hubungan fungsional pada variabel atau atribut didalam data tersebut.

Kata Kunci: Pertumbuhan penduduk, regresi linear, *rapidminer*

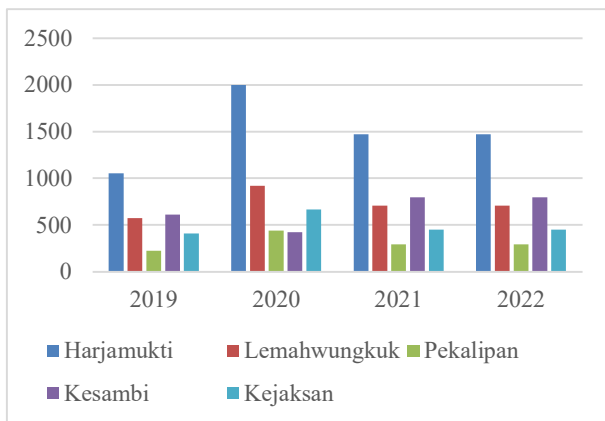
1. Pendahuluan

Badan Pusat Statistik (BPS) adalah Lembaga Pemerintah Non Kementerian yang bertanggung jawab langsung kepada Presiden. Sebelumnya, BPS merupakan Biro Pusat Statistik, yang dibentuk berdasarkan UU Nomor 6 Tahun 1960 tentang Sensus dan UU Nomer 7 Tahun 1960 tentang Statistik. Sebagai pengganti kedua UU tersebut ditetapkan UU Nomor 16 Tahun 1997 tentang Statistik. Berdasarkan UU ini yang ditindaklanjuti dengan peraturan perundangan dibawahnya, secara formal nama Biro Pusat Statistik diganti menjadi Badan Pusat Statistik [1].

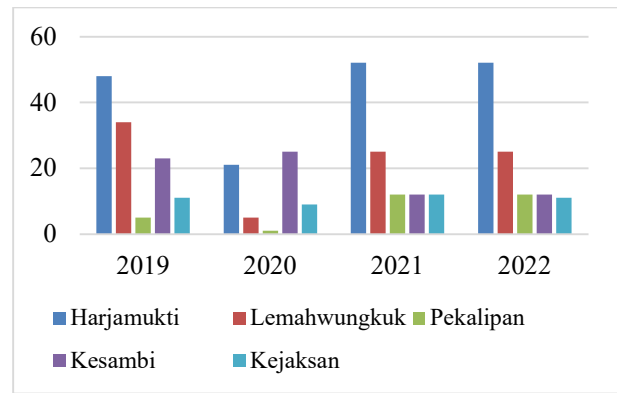
Pertumbuhan penduduk merupakan keseimbangan yang dinamis antara kekuatan-kekuatan yang menambah dan kekuatan yang mengurangi jumlah penduduk. Secara terus menerus penduduk akan dipengaruhi oleh jumlah bayi yang lahir tetapi secara bersamaan akan dikurangi oleh jumlah kematian yang terjadi pada semua golongan umur [2].

Sebagaimana diketahui laju pertumbuhan penduduk di suatu wilayah dipengaruhi oleh besar kecilnya komponen pertumbuhan penduduk. Salah satu pertumbuhan penduduk adalah pertumbuhan penduduk alamiah, dimana perubahan angka pertumbuhan penduduk alamiah disebabkan oleh kelahiran dan kematian [3].

Untuk mengetahui pertumbuhan penduduk di suatu wilayah diperlukan suatu data yang diperoleh dari pencatatan penduduk yakni Badan Pusat Statistik Kota setempat, sebagaimana ditunjukkan pada Gambar 1 dan 2.



Gambar 1. Diagram data Kelahiran tahun 2019-2022



Gambar 2. Diagram data kematian tahun 2019-2022

Berdasarkan Tabel 1 dan Tabel 2 data kependudukan dari Badan Pusat Statistik Kota Cirebon dari tahun 2019 hingga 2022 tercatat jumlah angka kelahiran meningkat dari 2.867 jiwa menjadi 3.713 jiwa.

Dengan demikian laju pertumbuhan penduduk adalah angka yang menunjukkan tingkat penambahan penduduk pertahun dalam jangka tertentu. Sehingga untuk memprediksi jumlah penduduk yang menggunakan perhitungan yang didasarkan variabel yang berpengaruh terhadap pertumbuhan penduduk dimasa yang akan datang.

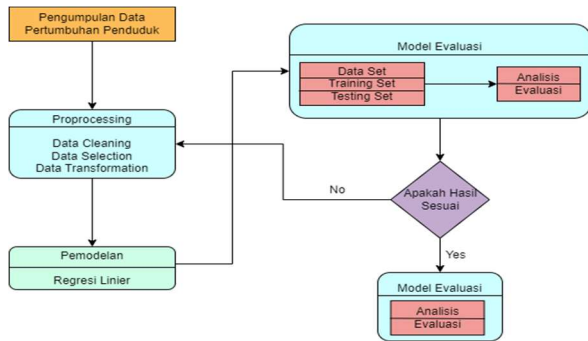
Seiring bertambahnya penduduk hal tersebut memengaruhi kepadatan penduduk sehingga menimbulkan masalah kependudukan seperti kemiskinan, perumahan dan lapangan pekerjaan. Dari beberapa permasalahan yang timbul akibat pertumbuhan penduduk, diperlukan suatu penelitian untuk mengestimasi laju pertumbuhan penduduk pada tahun berikutnya. Diharapkan setelah mengetahui laju pertumbuhan penduduk membantu pemerintah untuk merencanakan strategi peningkatan ekonomi, infrastruktur dll yang berdampak pada kesejahteraan masyarakat. Salah satu metode untuk mengestimasi pertumbuhan penduduk adalah *data mining*.

2. Metode Penelitian

2.1. Tahapan Penelitian

Dalam melakukan prediksi data pertumbuhan penduduk 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 3. Tahapan Penelitian

Berdasarkan Gambar 3, maka tahapan penelitian dapat diuraikan sebagai berikut:

1. Pengumpulan Data: Mengumpulkan data yang diperlukan, seperti data pertumbuhan penduduk dari periode sebelumnya dan data variabel independen yang dapat mempengaruhi pertumbuhan penduduk (misalnya, tingkat kelahiran, tingkat kematian dll.).
2. Pembersihan dan Persiapan Data: Melakukan pemeriksaan kualitas data untuk mengidentifikasi dan menangani nilai-nilai yang hilang, outlier, atau tidak konsisten. Menstandarisasi dan mengubah format data agar siap digunakan dalam analisis.
3. Pemodelan Regresi Linear: Memodelkan hubungan antara variabel dependen (pertumbuhan penduduk) dan variabel independen dengan metode regresi linear. Menggunakan teknik seperti regresi linier sederhana atau regresi linier berganda, tergantung pada kompleksitas hubungan dan jumlah variabel independen.
4. Analisis: terhadap data untuk mendapatkan gambaran awal tentang distribusi variabel-variabel yang akan digunakan dalam analisis regresi.
5. Pemilihan Variabel Independen: Memilih variabel-variabel independen yang paling relevan dan memiliki potensi kuat dalam mempengaruhi pertumbuhan penduduk berdasarkan pengetahuan domain dan analisis korelasi awal.

2.2. Pertumbuhan Penduduk

Pertumbuhan penduduk adalah perubahan jumlah penduduk yang dipengaruhi oleh kelahiran, kematian dan migrasi penduduk. Penduduk suatu wilayah akan bertambah apabila terdapat kelahiran dan penduduk yang datang ke wilayah tersebut sedangkan penduduk suatu wilayah akan berkurang apabila terdapat kematian dan terdapat penduduk yang meninggalkan wilayah tersebut [4].

2.3. Kelahiran

Menurut Mulyadi, Kelahiran merupakan hasil reproduksi yang nyata dari seorang wanita atau sekelompok wanita. Dengan kata lain fertilitas ini menyangkut banyaknya bayi yang lahir hidup. Fertilitas dalam pengertian demografi adalah kemampuan riil seorang wanita untuk melahirkan, yang dicerminkan dalam jumlah bayi yang dilahirkan. Tingkat kelahiran ditentukan oleh jumlah penduduk wanita berada pada usia reproduksi. Semakin banyak jumlah penduduk wanita usia reproduksi, maka diasumsikan jumlah kelahiran semakin banyak pula [5].

2.4. Kematian

Hilangnya kehidupan manusia secara permanen dan sifatnya dapat mengurangi jumlah penduduk disebut kematian. Indikator ini memiliki ikatan kuat karena jika semakin tinggi angka kematian maka semakin rendah angka jumlah penduduk [6].

2.5. Algoritma Regresi Linear

Metode Kuadrat terkecil (least square method): metode yang paling populer untuk menetapkan persamaan regresi linear sederhana. Bentuk Umum Regresi Linear Sederhana:

Penetapan Persamaan Regresi Linear Sederhana nilai a dan b

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

$$a = \bar{y} - b\bar{x} \text{ sehingga } a = \frac{\sum_{i=1}^n y_i}{n} - b \frac{\sum_{i=1}^n x_i}{n}$$

n: banyaknya pasangan data

yi: nilai peubah tak bebas Y ke-i

xi: nilai peubah bebas X ke-i

Menghitung persamaan regresi linearnya

$$Y = a + bX$$

Y: peubah tak-bebas

X: peubah bebas

a: konstanta

b: kemiringan

2.6. Clustering

Clustering merupakan suatu metode pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki 18 kedekatan (kemiripan). *Cluster* adalah kumpulan data yang memiliki kemiripan berdasarkan objek-objek yang sama dalam *cluster* atau *cluster* lainnya. Akan tetapi, algoritma *clustering* mencoba untuk melakukan pembagian terhadap semua data menjadi beberapa kelompok yang memiliki kedekatan/kemiripan, yang mana kemiripan *record* dalam satu kelompok akan bernilai lebih besar, sedangkan kemiripan dengan *record* di kelompok lain akan bernilai lebih kecil [7].

2.7. Data Mining

Data Mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara membedakan data sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data [8].

Secara umum, kegunaan data Mining dapat dibagi menjadi dua: deskriptif berarti digunakan untuk mencari pola-pola yang dapat dipahami manusia yang menjelaskan karakteristik data dan prediktif digunakan untuk membentuk sebuah model pengetahuan yang akan digunakan untuk prediksi. Berdasarkan fungsionalitasnya, tugas-tugas data mining dapat dikelompokkan kedalam enam kelompok berikut ini:

1. Klasifikasi: men-generalisasi struktur yang diketahui untuk diaplikasikan pada data-data baru.
2. Klasterisasi: mengelompokkan data yang tidak diketahui label kelasnya ke dalam

sejumlah kelompok tertentu sesuai dengan ukuran kemiripannya.

3. Regresi: menemukan suatu fungsi yang memodelkan data dengan galat (kesalahan prediksi) seminimal mungkin.
4. Deteksi anomali: mengidentifikasi data yang tidak umum bisa berupa outlier, perubahan atau deviasi yang mungkin sangat penting dan perlu investigasi lebih lanjut.
5. Pembelajaran aturan asosiasi atau pemodelan kebergantungan: mencari relasi antar variabel.
6. Perangkuman: menyediakan representasi data yang lebih sederhana meliputi visualisasi dan pembuatan laporan.

2.8. Machine Learning

Machine Learning dapat didefinisikan sebagai aplikasi komputer dan algoritma matematika yang diadopsi dengan cara pembelajaran yang berasal dari data dan menghasilkan prediksi di masa yang akan datang [9].

Berdasarkan teknik pembelajarannya, tipe-tipe *machine learning* dapat dibedakan menjadi *supervised learning*, *unsupervised learning*, *semi supervised learning* dan *reinforcement learning*. *Supervised learning* merupakan salah satu teknik *machine learning* yang menggunakan dataset (data training) yang sudah berlabel (labeled data) untuk melakukan pembelajaran pada mesin, sehingga mesin mampu mengidentifikasi label input dengan menggunakan fitur yang dimiliki untuk selanjutnya melakukan prediksi maupun klasifikasi, sedangkan *unsupervised learning* adalah teknik dengan menarik kesimpulan berdasarkan dataset yang merupakan input data labeled response [10].

2.9. Rapidminer

Rapidminer adalah perangkat lunak yang bersifat terbuka (*opensource*). *Rapidminer* merupakan solusi untuk melakukan suatu analisis terhadap data mining, text mining, dan analisis prediksi. *Rapidminer* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada 22 pengguna sehingga dapat membuat keputusan yang paling

baik. Rapidminer memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, *data preprocessing* dan visualisasi. *Rapidminer* merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. *Rapidminer* ditulis dengan menggunakan bahasa java sehingga dapat bekerja pada semua sistem operasi [11].

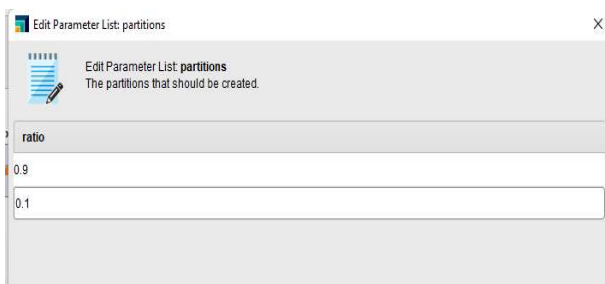
3. Hasil Penelitian

3.1. Data Uji

Penelitian ini menggunakan algoritma Regresi Linear, untuk mengidentifikasi pertumbuhan penduduk dan akan mendapatkan hasil *Root Mean Square Error* (RMSE) serta prediksi yang dapat digunakan dalam pengambilan keputusan ketika terjadi laju pertumbuhan penduduk atau tidak. Sumber data sebagai objek pada penelitian ini adalah data historis yang diambil dari situs resmi Badan Pusat Statistik Kota Cirebon. Data yang digunakan dalam penelitian ini terdiri dari atribut atau variabel seperti kelahiran dan kematian.

3.2. Split Validation

Teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian sebagai *data training* dan sebagian lainnya sebagai *data testing*. Dengan menggunakan *Split Validation* akan dilakukan percobaan training berdasarkan *split ratio* yang telah ditentukan sebelumnya, untuk kemudian sisa dari *split ratio data training* akan dianggap sebagai *data testing*. Proses *split ratio* ditunjukkan pada Gambar 4.



Gambar 4. Proses Split Ratio

Berdasarkan Gambar 2. Proses *Split Ratio* dengan nilai 0.9 sebagai Data Uji dan nilai 0.1 sebagai Data Testing.

Tabel 1. Dataset

Kelahiran	Kematian	Kecamatan
1053	26	1
571	20	2
224	3	3
611	12	4
408	6	5
1999	14	1
918	5	2
438	1	3
421	15	4
665	9	5
1472	30	1
706	15	2
293	7	3
794	6	4
448	12	5
1472	30	1
706	15	2
293	7	3
794	6	4
448	11	5

3.3. Perhitungan Regresi Linear

Untuk memperoleh koefisien regresi a, b₁ dan b₂ dapat diperoleh cara sebagai berikut:

$$b_1 = \frac{n \sum(x_1y) - (\sum x_1)(\sum y)}{n(\sum x_1)^2 - (\sum x_1)^2}$$

$$b_1 = \frac{20(35867) - (14734)(60)}{20(217090756) - (14734)^2}$$

$$b_1 = \frac{717340 - 884040}{4341815120 - 217090756 - 166700}$$

$$b_1 = \frac{4124724364}{-166700}$$

$$b_1 = -0,0000404148$$

$$b_2 = \frac{n \sum(x_2y) - (\sum x_2)(\sum y)}{n(\sum x_2)^2 - (\sum x_2)^2}$$

$$b_2 = \frac{20(610) - (250)(60)}{20(62500) - (250)^2}$$

$$b_2 = \frac{12200 - 15000}{1250000 - 625000 - 2800}$$

$$b_2 = \frac{-2800}{625000}$$

$$b_2 = -0,00448$$

$$a = \frac{\sum y - b_1 \sum x_1 + b_2 \sum x_2}{n}$$

$$a = \frac{60 - ((-0,0000404148) * (14734)) + ((-0,00448) * (250))}{20}$$

$$a = \frac{60 - (-0,5954716632) + (-1,12)}{20}$$

$$a = \frac{60 - (-1,7154716632)}{20}$$

$$a = \frac{61,715476632}{20}$$

$$a = 3,0857738316$$

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

b₁ = -0,0000404148
 b₂ = -0,00448
 a = 3,0857738316

Berikut dataset yang akan digunakan sebagai data uji:

Kelahiran(x1)	Kematian(x2)	Kecamatan(y)
1053	26	?
571	20	?
224	3	?

Menghitung Persaman Regresi Linear

$$Y = a + b_1X_1 + b_2X_2$$

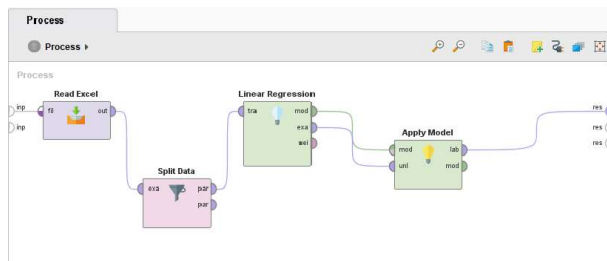
$$y = 3,0857738316 + (-0,0000404148 \times 1053) + (0,00448 \times 26) = 3,1596970472$$

$$y = 3,0857738316 + (-0,0000404148 \times 571) + (0,00448 \times 20) = 3,1522969808$$

$$y = 3,0857738316 + (-0,0000404148 \times 224) + (0,00448 \times 3) = 3,0901609164$$

3.4. Proses Pengujian Data (Rapidminer)

Melakukan *select attributes* yaitu untuk mengetahui hasil prediksi dari rapidminer, hasil perhitungan manual dan hasil uji di rapidminer. Dapat dilihat pada Gambar 5.



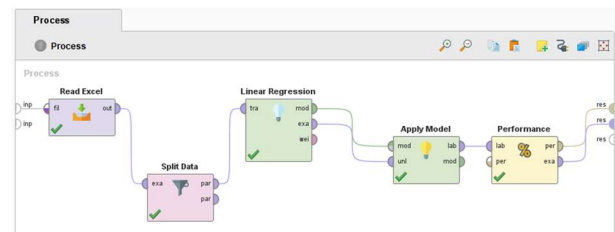
Gambar 5. Proses Rapidminer

Pada proses ini untuk memasukan *data training* dan *data testing* yang akan diuji untuk menghasilkan prediksi pada atribut *class* yang ditampilkan pada Gambar 6.

Row No.	Kecamatan(y)	prediction(Kecamatan(y))	Kelahiran (x1)	Kematian(x2)
3	4	3,298	611	12
4	5	4,018	408	6
5	1	0,920	1999	14
6	2	3,257	918	5
7	3	4,294	438	1
8	4	3,411	421	15
9	5	3,406	665	9
10	1	0,732	1472	30
11	2	2,949	706	15
12	3	4,139	293	7
13	4	3,392	794	6
14	5	3,562	448	12
15	2	2,949	706	15
16	3	4,139	293	7
17	4	3,392	794	6
18	5	3,627	448	11

Gambar 6. Hasil Prediksi

Ketika prediksi sudah dicari langkah selanjutnya yaitu mengukur seberapa akurat hasil prediksi yang telah kita buat pada Gambar 7.



Gambar 7. Proses Pencarian *Root Mean Squared Error* dan *Squared Error*

Untuk mempermudah dalam pembacaan data pertumbuhan penduduk, maka perlu di masukan *tools performance* untuk mencari *Root Mean Squared Error* dan *Squared Error*. Berikut ini hasilnya pada Gambar 8.

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 0.998 +/- 0.000
squared_error: 0.996 +/- 0.707
```

Gambar 8. Hasil Pengujian *Root Mean Squared Error* dan *Squared Error*

Langkah kedua, dilakukan implementasi algoritma regresi linear dengan menggunakan *tools rapidminer*. Berikut adalah tahapan dalam penerapan algoritma regresi linear:

1. Menentukan prediksi data test yang dilakukan oleh rapidminer dan menghasilkan nilai *confidence* yang telah diprediksi.
2. Menentukan performance dengan output untuk mencari *Root Mean Squared Error* dan *Squared Error*.

Pada permodelan *split validation* didalamnya terdapat dua bagian, yaitu bagian *training* (digunakan untuk algoritma klasifikasi) dan bagian *testing* (menggunakan fitur *Apply Model* untuk mengaplikasikan model pada data testing dan fitur *Performance* untuk menampilkan *Root Mean Squared Error* dan *Squared Error*).

3.5. Analisa Hasil

Berdasarkan hasil pengujian yang telah dilakukan bahwa variabel atau atribut yang digunakan dalam penelitian ini (kelahiran dan kematian) berpengaruh signifikan terhadap penelitian ini terbukti dengan menggunakan algoritma regresi linear mampu memberikan hasil yang baik dengan nilai *Root Mean Squared Error*: 0.998 +/- 0.000 dan *Squared Error*: 0.996 +/- 0.707. Hal ini dikarenakan adanya korelasi atau hubungan fungsional (sebab – akibat) antara variabel yang satu (dependen atau kriteria) dengan variabel yang lain (independen atau predictor). Proses pengujian ini dilakukan untuk mengidentifikasi pertumbuhan penduduk dengan algoritma regresi linear.

4. Kesimpulan

Pada penelitian ini dengan memanfaatkan beberapa data pertumbuhan penduduk yang telah tersimpan dalam basis data menggunakan atribut diantaranya kelahiran dan kematian. Sehingga dengan menerapkan algoritma regresi linear dapat dilakukan suatu prediksi berdasarkan hubungan fungsional pada variable atau atribut didalam data tersebut.

Mengolah data pertumbuhan penduduk menggunakan algoritma regresi linear dimulai dari tahap seleksi data (atribut yang digunakan dan penentuan data training serta data testing), tahap pengujian algoritma (regresi linear), dan

tahap uji akurasi (menggunakan *split validation*).

Proses pengujian data pada penelitian ini menggunakan algoritma regresi linear mampu memberikan hasil yang baik dengan nilai *Root Mean Squared Error*: 0.998 +/- 0.000 dan *Squared Error*: 0.996 +/- 0.707.

5. Saran

Perlu memaksimalkan atau menambah atribut yang lebih spesifik dan lebih banyak dalam menentukan prediksi pertumbuhan penduduk.

Perlu adanya penelitian lebih lanjut dalam melakukan pengujian dengan metode atau algoritma lain agar memperoleh perbandingan dengan tingkat akurasi yang paling baik dalam mengidentifikasi data pertumbuhan penduduk. Serta penambahan jumlah dataset atau sampel pada penelitian agar mendapatkan nilai akurasi yang lebih baik.

6. Daftar Pustaka

- [1] M. Fardi, "Sistem Informasi Perjalanan Dinas Pada Badan Pusat Statistik (BPS) Provinsi Lampung," *J. Ilmu Data*, vol. 2, no. 1, pp. 1–11, 2022, [Online]. Available: <http://ilmudata.org/index.php/ilmudata/article/view/59>
- [2] S. Sapri and H. Hermawansa, "Sistem Pelaporan Jumlah Penduduk Desa Kembangseri Kecamatan Bermani Ilir Kabupaten Kepahiang Provinsi Bengkulu," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 263–269, 2020, doi: 10.36294/jurti.v4i2.1440.
- [3] L. Belakang, "Latar Belakang Demografi," pp. 1–8, 2010, [Online]. Available: <http://siat.ung.ac.id/files/wisuda/2012-1-87202-451408136-bab1-16082012073224>
- [4] K. Risnawati, "Analisis Kebijakan Spasial Dan Peran Kabupaten Puncak Jaya Dalam Wilayah Provinsi Papua," vol. 8, no. 1, pp. 19–30, 2023.
- [5] I. Setyawati, "Implementasi Data Mining Dengan Algoritma Naïve Bayes Untuk Memprediksi Tkp Kriminalitas Di Kabupaten Ponorogo," *Pap. Knowl.*

- Towar. a Media Hist. Doc.*, vol. 7, no. 1, pp. 1–33, 2020, [Online]. Available: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/MT_Globalization_Report_2018.pdfhttp://eprints.lse.ac.uk/43447/1/India_globalisation%2C_society_and_inequalities%28Isero%29.pdf<https://www.quora.com/What-is-the>
- [6] D. watik, A. Trisiana, and F. Novitasari, “Analisis Peran Pemerintahan Dalam Pengendalian Pertumbuhan Penduduk,” *Res. Fair Unisri*, vol. 6, no. 1, pp. 45–56, 2022, doi: 10.33061/rsfu.v6i1.6853.
- [7] K. Annisa, B. S. Ginting, and M. A. Syari, “Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada PDAM Langkat,” *J. Ilmu Komput. dan Inform. Vol.*, vol. 6, no. 1, pp. 113–129, 2022.
- [8] E. Manurung and P. S. Hasugian, “Data Mining Tingkat Pesanan Inventaris Kantor Menggunakan Algoritma Apriori pada Kepolisian Daerah Sumatera Utara,” *J. Inform. Pelita Nusant.*, vol. 4, no. 2, pp. 8–13, 2019.
- [9] A. Roihan, P. A. Sunarya, and A. S. Rafika, “Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper,” *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.
- [10] E. Retnoningsih and R. Pramudita, “Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python,” *Bina Insa. Ict J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.
- [11] R. Rustam, S. Rahmatullah, S. Supriyato, and S. Wahyuni, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Triplek Pada Pt Puncak Menara Hijau Mas,” *J. Inf. dan Komput.*, vol. 8, no. 2, pp. 75–86, 2020, doi: 10.35959/jik.v8i2.186.