

**Algoritma *Linear Regression* dalam Memprediksi Pertumbuhan Jumlah Penduduk Menurut Provinsi dan Jenis Kelamin**

**Bella Aprianti<sup>1</sup>, Anisa Aulia<sup>2</sup>, Intan Purnamasari<sup>3</sup>**

<sup>1,2,3</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

**Email:** [bella.aprianti18004@student.unsika.ac.id](mailto:bella.aprianti18004@student.unsika.ac.id)

**Info Artikel**

Sejarah Artikel:

Diterima: 22 Maret 2022

Direvisi: 26 Maret 2022

Dipublikasikan: April 2022

e-ISSN: 2089-5364

p-ISSN: 2622-8327

DOI: 10.5281/zenodo.6408866

**Abstract:**

*Basically the population in a country will continue to increase every year. According to the Central Statistics Agency (BPS), Indonesia's population in 2015 continued to increase until 2020. The increasing population growth will have an impact on the economic, social, political and national defense sectors. Proper structured planning by the government is needed to avoid the possible impacts. Researchers predict population growth using the Linear Regression Algorithm in prediction techniques in data mining. The results of the Linear Regression Algorithm are almost perfect because they are very close to the original data, and the resulting accuracy is very good. The lowest RMSE result is 125,321.*

**Keywords:** *BPS, Population Data, Data Mining, Linear Regression*

**PENDAHULUAN**

Penduduk adalah orang-orang yang tinggal di suatu wilayah yang terikat oleh aturan-aturan yang berlaku dan selalu berinteraksi satu sama lain. Dalam sosiologi, penduduk adalah kumpulan orang yang menempati wilayah dan ruang geografis tertentu (Ruslan., 2016).

Pada dasarnya jumlah penduduk di suatu negara akan terus meningkat setiap tahunnya. Menurut Badan Pusat Statistik (BPS), jumlah penduduk Indonesia pada

tahun 2015 terus meningkat hingga tahun 2020. Gambar 1. merupakan informasi jumlah penduduk dari tahun 2015 hingga tahun 2021.

TH2015	TH2016	TH2017	TH2018	TH2019	TH2020
255587.9	258496.5	261355.5	264161.6	266911.9	269603.4

**Gambar 1.** Jumlah penduduk Indonesia tahun 2015-2020

Pertumbuhan jumlah penduduk yang terus meningkat akan berdampak pada sektor

ekonomi, sosial, politik dan pertahanan negara (Dewi, A. K., & Furqon, M. T., 2020).

Perencanaan terstruktur yang tepat oleh pemerintah sangat diperlukan untuk menghindari dampak yang mungkin terjadi (Seruni, D. S., Furqon, M. T., & Wihandika, R. C., 2020). Upaya yang dapat dilakukan yaitu dengan memprediksi pertumbuhan jumlah penduduk setiap tahunnya menggunakan teknik data mining. Data Mining adalah proses mencari pola atau informasi menarik dalam sebuah data menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database (KDD)* secara keseluruhan (Yuli Mardi., 2016). Salah satu teknik data mining yaitu teknik prediksi.

Teknik prediksi merupakan teknik yang dipakai pada proses data mining, teknik prediksi memberikan nilai dari hasil yang akan ada di masa mendatang. Algoritma yang dapat dipakai pada teknik prediksi yaitu *Linear Regression*.

*Linear Regression* merupakan algoritma prediksi yang menggunakan garis lurus untuk menggambarkan hubungan diantara dua variabel atau lebih. Variabel tersebut terbagi atas dua jenis yaitu variabel pemberi pengaruh dan variabel terpengaruh. Variabel pemberi pengaruh dapat dianalogikan sebab, sementara variabel terpengaruh merupakan akibat (Susanto, S., & Suryadi, D., 2010). Informasi yang dihasilkan dari data mining dengan *Linear Regression* bisa dijadikan sebagai dasar untuk pengambilan keputusan dan *Linear Regression* ini dipilih karena sederhana, cepat dan sangat akurat.

Prediksi pertumbuhan jumlah penduduk sebelumnya sudah banyak dilakukan, diantaranya yaitu yang dilakukan oleh Djajeng Sekar Seruni, dkk (2020). Penelitian tersebut memprediksi jumlah penduduk kota Malang dengan menggunakan Algoritma *K-Nearest*

*Neighbor Regression*, pengujian pada penelitian ini menghasilkan nilai *Mean Absolute Percentage Error (MAPE)* terbaik sebesar 0,02526%.

Berikutnya yaitu yang dilakukan oleh M. Ruslan (2016). Penelitian tersebut memprediksi jumlah penduduk Provinsi Kalimantan Selatan menggunakan metode *Semi Average*, dari hasil uji coba 3 tahun terakhir menunjukkan validitas metode *Semi Average* adalah 98,34%.

Berdasarkan latar belakang tersebut, peneliti memilih Algoritma *Linear Regression* dalam melakukan prediksi pertumbuhan jumlah penduduk, dengan tujuan untuk menguji kelayakan *Linear Regression* sebagai algoritma prediksi, yang diharapkan dapat membantu pemerintah dalam membuat rancangan perencanaan negara kedepannya guna menghindari dampak buruk yang mungkin terjadi.

## **METODOLOGI PENELITIAN**

Penelitian ini menggunakan metodologi *Knowledge Discovery in Database (KDD)*, dengan pendekatan kuantitatif yaitu:

### **1. Data Original**

Data Original merupakan data asli yang didapat setelah melalui tahap pemilihan data, data yang digunakan yaitu data jumlah penduduk menurut provinsi dan jenis kelamin dalam jangka waktu 5 tahun (tahun 2015 sampai tahun 2020).

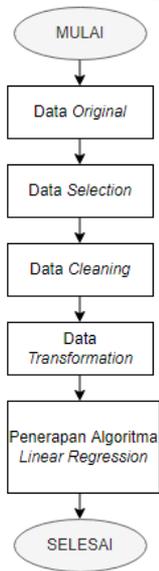
### **2. Data Selection**

Setelah memiliki data original, selanjutnya yaitu melakukan seleksi data. Tahap seleksi data ini yaitu melakukan proses seleksi atau pemilihan data yang relevan terhadap penelitian untuk diterima dari koleksi data yang ada.

### **3. Data Cleaning**

Setelah dilakukan proses pemilihan data dan seleksi data selanjutnya adalah melakukan proses pembersihan data. Pembersihan data dilakukan dengan cara membuang data yang tidak perlu dan melengkapi data yang tidak lengkap.

4. *Data Transformation*  
Setelah proses pembersihan data dilakukan selanjutnya adalah melakukan proses transformasi data. Pada tahap ini data diubah atau digabung ke dalam format yang sesuai.
5. Penerapan Algoritma *Linear Regression*  
Pada proses ini, dilakukan pengujian Algoritma *Linear Regression* dengan memanfaatkan tools *Rapid Miner*.

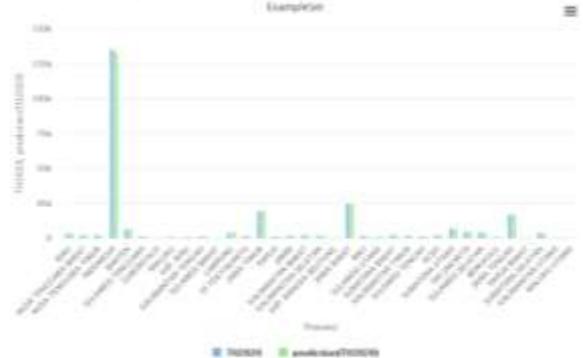


**Gambar 2.** Flowchart Metodologi Penelitian

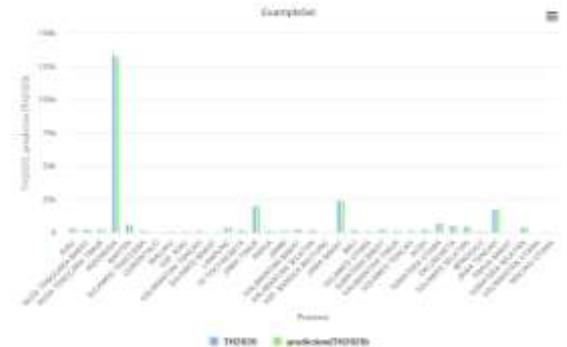
**HASIL DAN PEMBAHASAN**

Pada penelitian ini, pengujian Algoritma *Linear Regression* dilakukan dengan menggunakan *Cross Validation*. Peneliti melakukan pengujian data dengan 3 jumlah data yaitu data Penduduk Laki-Laki, Penduduk Perempuan dan Penduduk Total.

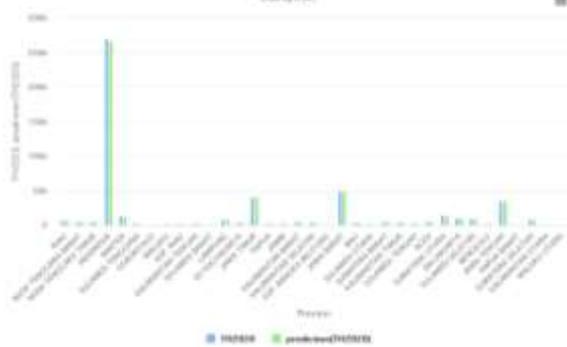
1. Hasil prediksi data Penduduk Laki-Laki



2. Hasil prediksi data Penduduk Perempuan



3. Hasil prediksi data Penduduk Total



Berdasarkan hasil prediksi pada 3 grafik bar tersebut dapat disimpulkan bahwa prediksi yang dihasilkan Algoritma *Linear Regression* sangat mendekati data asli. Tabel 1. merupakan hasil dari *Root Mean Square Error (RMSE)*.

**Tabel 1.** Hasil *Root Mean Square Error (RMSE)*

Data	RMSE
Data Penduduk Laki-Laki	125.321
Data Penduduk Perempuan	169.292
Data Penduduk Total	317.120

**KESIMPULAN**

Dari hasil penelitian tersebut dapat disimpulkan bahwa:

1. Algoritma *Linear Regression* dapat memberikan hasil prediksi yang hampir sempurna karena sangat mendekati data asli, serta akurasi yang dihasilkan sangat baik.
2. Algoritma *Linear Regression* dengan hasil *Root Mean Square Error (RMSE)* terendah yaitu 125.321.
3. *Pre-pocessing* yang dilakukan pada pengolahan data terhadap dataset yang

ada sangat berguna dalam menentukan *Root Mean Square Error (RMSE)*.

#### **DAFTAR PUSTAKA**

- Ruslan. (2016). Prediksi Jumlah Penduduk Provinsi Kalimantan Selatan Menggunakan Metode Semi Average. 2(1), 1–7.
- Dewi, A. K., & Furqon, M. T. (2020). Prediksi Laju Pertumbuhan Penduduk Menggunakan Metode Support Vector Regression (Studi Kasus: Kota Malang). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN, 2548, 964X.
- Seruni, D. S., Furqon, M. T., & Wihandika, R. C. (2020). Sistem Prediksi Pertumbuhan Jumlah Penduduk Kota Malang menggunakan Metode K-Nearest Neighbor Regression. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN, 2548, 964X.
- Yuli Mardi, “Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika,” *Jurnal Edik Informatika Penelitian Bid. Komput. Sains dan Pendidik. Inform.*, vol. 2, pp. 213–219, 2016.
- Susanto, S., & Suryadi, D. (2010). Pengantar data mining: mengagali pengetahuan dari bongkahan data.