Indonesian Journal of Computer Science
ISSN 2302-4364 (*print*) dan 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam, No. 1, Padang, Indonesia, Telp. (0751) 7056199, 7058325
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Prediksi Tingkat Angkatan Kerja Terhadap Pengangguran Terbuka Di Semarang Menggunakan Algoritma Regresi Linear

Febrilia Hayyu Pradaningrum[1], Ferian Fauzi Abdulloh[2]
Febriliahayyu10@students.amikom.ac.id , ferian@amikom.ac.id
Universitas Amikom Yogyakarta

Informasi Artikel

Diterima : 15 Nov 2023
Direview : 27 Jan 2024
Disetujui : 15 Feb 2024

Kata Kunci

Pengangguran, Angkatan kerja, RMSE, Kuantitatif, Regresi linier

Abstrak

Pengangguran merupakan sebuah keadaan dimana seseorang yang tidak memiliki pekerjaan yang disebabkan oleh banyak faktor, tidak hanya karena malas untuk mencari kerja namun kebanyakan di wilayah Indonesia ini pengangguran disebabkan oleh keterbatasan lapangan pekerjaan, banyaknya persaingan di dunia kerja, banyaknya jumlah angkatan kerja, kurangnya pengalaman dalam dunia kerja, dan juga terlalu memilih dalam bekerja. Pengangguran yang terjadi di Semarang ini diakibatkan karena tingginya angka angkatan kerja yang membuat tingkat pengangguran semakin banyak. Pada penelitian ini penulis memprediksi tingkat pengangguran yang berada di Semarang. Penelitian ini merupakan penelitian kuantitatif yang datanya diambil dari BPS Semarang. Pada penelitian ini penulis menggunakan algoritma regresi linier. Algoritma tersebut banyak digunakan dalam kasus untuk memprediksi suatu masalah, pada penelitian ini menghasilkan sebuah nilai RMSE (Root Mean Square Error) sebesar 0.200 dengan nilai R Square sebesar 0.57. Hasil yang didapatkan dapat digunakan sebagai acuan pemerintah untuk melihat tingginya angka pengangguran yang terdapat di Semarang.

Keywords

Unemployment, Labor force, RMSE, Quantitative, Linear regression

Abstrak

Unemployment is a situation where a person who does not have a job is caused by many factors, not only because they are lazy to look for work but mostly in this region of Indonesia unemployment is caused by limited employment opportunities, a lot of competition in the world of work, the large number of the labor force, lack of experience in the world of work, and also too choosy in working. The unemployment that occurs in Semarang is caused by the high number of labor force that makes the unemployment rate more and more. In this research, the author predicts the level of unemployment in Semarang. This research is a quantitative research whose data is taken from BPS Semarang. In this research, the author uses linear regression algorithm. The algorithm is widely used in cases to predict a problem, this research produces an RMSE (Root Mean Square Error) value of 0.07 with an R Square value of 91%. The results obtained can be used as a reference for the government to see the high unemployment rate in Semarang.

Attribution-ShareAlike 4.0 International License

A. Pendahuluan

Indonesia merupakan negara yang kaya akan budaya dan keunikan dari beberapa daerah. Indonesia juga menjadi pusat perhatian dari beberapa negara maju dan berkembang di luar sana. Negara berkembang seperti Indonesia memiliki masalah seperti kemiskinan, pengangguran, bencana alam, perceraian, dan banyak masalah – masalah yang lain. Fenomena tersebut tidak hanya terjadi di negara berkembang namun juga terjadi di negara maju. Fenomena tersebut di berbagai negara sudah menjadi permasalahan yang umum di seluruh dunia [1].

Pesatnya perkembangan teknologi digital sejak adanya Covid – 19 menjadikan banyaknya pengagguran di Indonesia. Perkembangan teknologi digital menjadi salah satu faktor banyaknya pengangguran di Indonesia karena dengan berkembangnya teknologi digital banyak perusahaan menggunakan sistem AI. Sistem tersebut dapat membantu tugas manusia dengan hanya mengoperasikan melalui alat digital. Hal tersebut menjadikan beberapa orang mengalami masalah kehilangan pekerjaan [2].

Masalah pengangguran ini menurut beberapa penelitian pendahulu [3] bahwa pengangguran dipengaruhi oleh semakin meningkatnya tingkat angkatan kerja yang tidak sebanding dengan adanya lapangan pekerjaan dan tidak sebanding dengan pengalaman yang miliki. Tidak hanya seseorang yang sudah lulus dari perguruan tinggi namun yang juga dari seeorang yang sudah lulus SMA, SMP dan juga SD.

Menurut [4] bahwa pengangguran sudah menjadi masalah umum diberbagai negara, tidak hanya negara berkembang namun juga negara maju. Disetiap daerah tingkat pengangguran yang berbeda – beda, semakin banyak tingkat pengangguran dapat menyebabkan tingkat kesejahteraan di daerah dan di negara sendiri menjadi kurang. Hal tersebut dapat disimpulkan jika pemerintah belum maksimal dalam menangani masalah pegangguran yang terjadi hampir terdapat di berbagai daerah di Indonesia.

Menurut [5] Semakin banyaknya tingkat pengangguran di suatu daerah semakin daerah tersebut tidak memiliki keseimbangan antara angkatan kerja dan penyedia lapangan pekerjaan dan juga pengangguran termasuk dengan angkatan kerja yang terlalu pemilih dalam menentukan pekerjaan sehingga angka pengangguran meningkat.

Menurut [6] Tingkat pengangguran terbuka berpengaruh terhadap kemiskinan di Provinsi Jawa Tengah. Pada hasil uji menggunakan metode regresi linear berganda peneliti mendapatkan hasil R-Square sebesar 0.9445 (94,45%). Pada penelitian ini peneliti menjelaskan bahwa tingkat Indeks Pembangunan Manusia (IPM) yang tidak berpengaruh terhadap tingkat kemiskinn di Jawa Tengah.

Menurut [7] Bertambahnya jumlah penduduk di kota Bekasi dari tahun 2011 – 2020, peneliti melakukan analisis menggunakan algoritma regresi linear terhadap jumlah angkatan kerja dengan bertambahnya jumlah penduduk di kota Bekasi. Pada penelitian ini peneliti mendapatkan nilai korelasi sebesar 0,942 (94%), pada penelitian tersebut menunjukkan bahwa bertambahnya jumlah penduduk di kota Bekasi maka angka angkatan kerja meningkat.

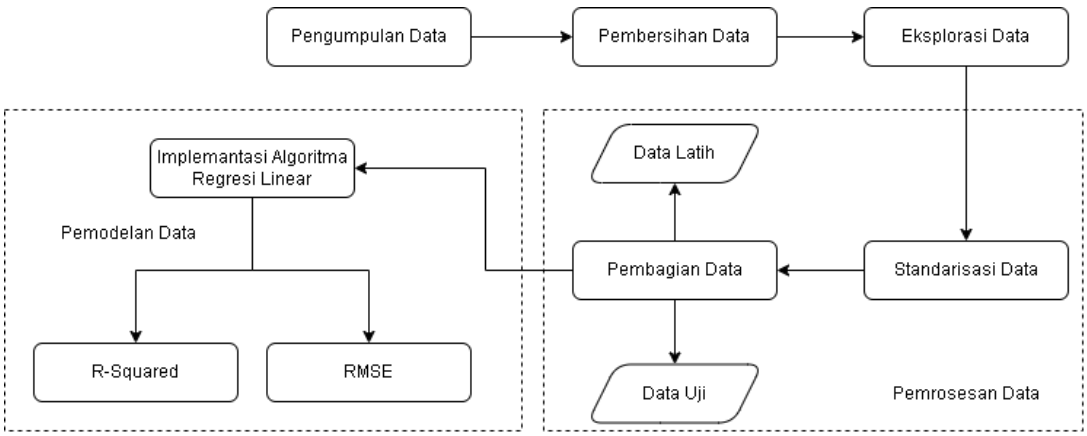
Pada penelitian ini penulis memprediksi pengangguran terbuka terhadap banyaknya tingkat angkatan kerja dan bekerja di Semarang. Pada penelitian ini menggunakan sebuah algoritma regresi. Algoritma regresi yang digunakan adalah Regresi Linear. Regresi Linear [8] adalah sebuah algoritma sederhana yang dipelajari oleh mesin untuk menganalisis sebuah prediksi dari sebuah peristiwa. Beberapa peneliti terdahulu sudah bannyak membuktikan dalam memprediksi sebuah kasus seperti prediksi kemiskinan, keuangan, cuaca, gender dan yang lainnya yang dibuktikan menggunakan regresi linear [9].

Prediksi digunakan untuk mengurangi bertambahnya kejadian yang dapat merugikan seseorang pada kemudian hari. Beberapa hal yang dapat diprediksi seperti memprediksi bencana alam, ekonomi, sosial dan budaya, ketenagakerjaan. Prediksi terhadap pengangguran tidak hanya ada di Indonesia namun juga terdapat di luar negeri. Semakin banyaknya tingkat pengangguran di satu wilayah semakin wilayah tersebut dianggap sebagai wilayah yang tidak memiliki kesejahteraan rakyat. Karena pemerintah di wilayah tersebut tidak memperhatikan kesejahteraan rakyatnya [10].

Pada penelitian ini penulis mengharapkan dapat membantu upaya pemerintah untuk lebih gencar dalam menangani masalah pengangguran, kemiskinana, atau masalah keuangan di Indonesia.

B. Metode Penelitian

Penelitian ini merupakan penelitian kuantitatif yang mana data ini berupa tingkat pengangguran terbuka di Semarang dari rentan waktu 2022- 2017 yang merupakan data *time series* . Alur penelitian dapat dilihat pada **Gambar 1.**



Gambar 1. Alur Penelitian

Pengumpulan Data

Data ini merupakan data sekunder yang diambil dari Badan Pusat Statistik Semarang. Data ini berupa Tingkat Pengangguran Terbuka di Semarang rentan waktu 2017 – 2022. Data ini merupakan data tahunan yang didalamnya terdapat Tahun, Total Angkatan Kerja, Bekerja, dan Pengangguran Terbuka dalam jumlah jiwa. Data tersebut diolah menggunakan Google Colab yang digunakan untuk menetukan prediksi menggunakan algoritma Regresi Linier.

Tahun	Total Angkatan Kerja	Pengangguran Terbuka	Bekerja
2017	963496	63700	899796
2018	921551	48724	872827
2019	951135	43198	907937
2020	1023964	98001	925963
2021	1034794	98718	936076
2022	1075827	81736	994091

Tabel 1. Data Tingkat Pengangguran Terbuka

Sumber : semarangkota.bps.go.id/

Pencarian Data Kosong

Pada tahap ini dilakukan pencarian data yang kosong pada dataset sebelum dilakukan pemrosesan data dan dilakukan sebuah permodelan data menggunakan algoritma yang sudah ditentukan. Pencarian data yang kosong juga bertujuan untuk mencegah terjadinya hasil akhir yang kurang baik.

Eksplorasi Data

Dari data diatas penulis mengambil data dari pengangguran terbuka untuk mevisualisasikan dengan menggunakan diagram pie. Dimana pada diagram tersebut menunjukkan bahwa tingkat pengangguran terbuka terbanyak terdapat berdasarkan pada tahun. Dapat dilihat di **Gambar 2.**

Gambar 2. Presentase Tingkat Pengangguran

Pada tahun 2021 di Semarang memiliki tingkat presentase pengangguran yang paling tinggi sebesar 22,7% dan hampir sebanding dengan tahun 2020 sebesar 22,6%. Pada dua tahun tersebut di Indonesia masih terdapat wabah penyakit yang bernama Covid-19 yang mana pada kejadian tersebut susahnya

untuk mencari sebuah pekerjaan dan banyaknya orang yang dikeluarkan dari tempat kerja yang dikarenakan masalah ekonomi yang tidak stabil dan kondisi di Indonesia sedang tidak baik karena adanya musibah tersebut.

C. Hasil dan Pembahasan

Pada hasil pembahasan yang dihasilkan dari penelitian tentang prediksi pengangguran terbuka yang terjadi di salah satu Provinsi Jawa Tengah ini dapat menjadi acuan pemerintahan untuk lebih memperhatikan masalah pengangguran yang terjadi di daerah tersebut agar tidak bertambah banyak. Data yang sudah ada selanjutnya dilakukan pemrosesan data.

a. Standarisasi Data

Standarisasi data merupakan proses preprocessing yang dilakukan sebelum melakukan pemodelan data. Standarisasi mengubah nilai dengan rata – rata 0 dan standar deviasi 1, pada proses ini lebih dianjurkan dalam melakukan pengolahan data dalam mencari sebuah prediksi dari beberapa kasus. Langkah – langkah standarisasi [11], [12]:

1. **Memilih Data / Variabel**

   Pada proses ini dilakukan pemilihan data / variabel yang akan dilakukan standarisasi dengan membuat dataframe baru seperti **Gambar 3.**

```
# Pilih kolom yang ingin di-standarisasi
columns_to_scale = ['Total Angkatan Kerja','Pengangguran Terbuka']
```

Gambar 3. Pemilihan Variabel

   Tujuan dari pemilihan variabel dengan membuat dataframe baru yaitu untuk memudahkan proses standarisasi sebelum dilakukan pemodelan data.

2. **Menghitung rata – rata dan standar deviasi**

   Sebelum melakuka proses standarisasi dilakukan instalasi library dari standarisasi. Setelah itu pada proses ini dilakukan perhitungan rata – rata dan standar deviasi dari masing – masing variabel. Dengan persamaan seperti dibawah.

$$\boldsymbol{Rata - rata} = \frac{\sum \boldsymbol{X}}{\boldsymbol{n}} \qquad \textbf{(1)}$$

Total Angkatan Kerja

$$= \frac{963496 + 921551 + 951135 + 1023964 + 1034794 + 1075827}{6}$$

$$= 995127.8$$

Pengangguran Terbuka

$$= \frac{63700 + 48724 + 43198 + 98001 + 98718 + 81736}{6}$$

$$= 72346.2$$

$$\boldsymbol{Standar\ Deviasi} = \sqrt{\frac{\sum(\boldsymbol{X - Mean\ X})^2}{\boldsymbol{n}}} \qquad \textbf{(2)}$$

Total Angkatan Kerja

$$= \sqrt{\frac{(1000570771 + 5413545498 + 1935366452 + 831524508 + 1573407422 + 6512360881)}{6}}$$
$$= \sqrt{\frac{17266775532}{6}}$$
$$= 53645.1$$

Pengangguran Terbuka

$$= \sqrt{\frac{(74756198.03 + 558006758 + 849615620 + 658170473.4 + 695473593.4 + 88168970.03)}{6}}$$
$$= \sqrt{\frac{2924191613}{6}}$$
$$= 22076.4$$

3. Standarisasi Variabel
Pada proses ini adalah proses setelah beberapa variabel dihitung rata – rata dan standar deviasi, selanjutnya proses standarisasi dengan persamaan.

$$\boldsymbol{Standarisasi\ Data} = \frac{\boldsymbol{Nilai\ Asli - Rata - rata(Variabel)}}{\boldsymbol{Standar\ Deviasi}} \qquad \textbf{(3)}$$

Pada persamaan diatas dijelaskan bahwa nilai asli merupakan nilai dari masing- masing variabel yang dikurangi dengan rata – rata dari variabel itu sendiri lalu di bagi dengan standar deviasi dari masing – masing variabel. Yang menghasilkan seperti **Gambar 3**.

Total Angkatan Kerja	Pengangguran Terbuka
-0.538274	-0.357524
-1.252046	-0.976791
-0.748620	-1.205295
0.490701	1.060843
0.674993	1.090492
1.373246	0.388275

Gambar 3. Hasil Standarisasi Data

b. Split data

Pada proses ini dilakukan pemilihan data yang akan diolah untuk mendapatkan sebuah nilai yang nantinya akan menjadi sebuah acuan untuk memprediksi sebuah kasus atau masalah. Dalam proses ini dipilih untuk nilai X adalah Total Angkatan Kerja dan nilai Y adalah Pengangguran Terbuka.

```
# Memisahkan variabel X dan Y
X = df['Total Angkatan Kerja'].values.reshape(-1,1)
Y = df['Pengangguran Terbuka']
```

Gambar 4. Split Data

c. Menentukan Data Latih dan Data Uji

Pada proses ini yaitu menentukan data latih dan data uji dengan menggunakan library *" train_test_split () "* dari *Sklearn*. Data dibagi menjadi 80 : 20 yaitu 80 data latih dan 20 data uji dengan penentuan pembagian data secara acak sebanyak 42.

d. Menentukan Model

Setelah menentukan data latih dan uji, selanjutnya dilakukan pemodelan data. Pada kasus ini menggunakan model regresi linear sederhana. Algoritma regresi linear merupakan sebuah algoritma yang terdapat pada machine learning, algoritma ini digunakan untuk memprediksi sebuah masalah yang terdiri dari variabel yang merupakan sebab permasalahan dan variabel yang merupakan akibat dari permasalahan tersebut[13]. Terdapat tiga jenis regresi yaitu regresi linear sederhana, regresi linear berganda dan juga regresi polinomial.

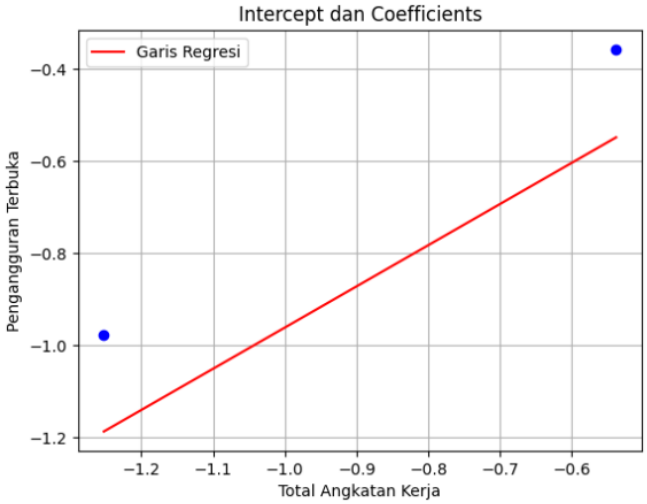
e. Menghitung Kemiringan dan Intersep

Pada tahap ini setelah dilakukan proses standarisasi dan pembagian data bisa dicari kemiringan dan intersep dari variabel X dan Y. Dengan persamaan seperti dibawah ini.

$$\boldsymbol{Kemiringan\ (m)} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad \textbf{(4)}$$

$$\boldsymbol{\alpha\ (intersep)} = \boldsymbol{Rata-rata\ Y} - (\boldsymbol{b} \times \boldsymbol{Rata-rata\ X}) \qquad \textbf{(5)}$$

Hasil dari menghitung kemiringan dan intersep variabel X dan Y menghasilkan nilai ***m (kemiringan) 0.8947587631701406*** dan nilai ***α (intersep) -0.06689737491162567.*** Dapat dilihat pada grafik dibawah.



Gambar 5. Intersep dan Kemiringan

f. Evaluasi Model

Pada tahap ini algoritma yang digunakan dalam permodelan data diuji untuk mengetahui baik atau buruknya kerja dari algoritma tersebut. Pada tahap ini dilakukan perhitungan untuk menentukan Root Mean Squared Error (RMSE) dan R-Squared.

R-Squared bertujuan untuk mengukur besar variasi variabel Y terhadap variabel X. Semakin mendekati 1 maka model yang dihasilkan baik, namun tidak menutup kemungkinan jika mendekati 1 model yang dihasilkan baik sesuai dengan kasus dan jumlah dari variabel yang dilakukan perhitungan.

RMSE bertujuan untuk mengukur hasil prediksi dari model yang digunakan dalam model regresi. Semakin kecil nilai RMSE maka kemampuan prediksi model baik. Pada penelitian ini hasil perhitungan RMSE mendapatkan nilai 0.201 dan dengan nilai R square sebesar 0.57.

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \acute{y}_i)^2}{n}} \qquad \textbf{(6)}$$

```
# Menghitung R-squared
r_squared = r2_score(Y_test, predictions)

# Menghitung RMSE
rmse = np.sqrt(mean_squared_error(Y_test, predictions))

print("R-squared:", r_squared)
print("RMSE:", rmse)
```

```
R-squared: 0.5789081915418874
RMSE: 0.20092608439887094
```

Gambar 5. Hasil Evaluasi

D. Simpulan

Pada penelitian ini dapat disimpulkan bahwa nilai RMSE sebesar 0.200 . Nilai yang dihasilkan dari sebuah prediksi pengangguran terbuka terhadap jumlah angkatan kerja di Semarang merupakan hasil yang baik. Namun tidak menutup kemungkinan jika terdapat data baru hasilnya akan tetap sama. Hasil dari RMSE bergantung dengan data yang diprediksi. Pada Banyaknya angka angkatan kerja membuat tingkat pengangguran lebih banyak dan juga karena kurangnya lapangan pekerjaan dan keahlian yang dimiliki oleh beberapa orang. Namun pengangguran juga disebabkan oleh di PHK dari kerjaan lama, karena perusahaan mengurangi jumlah karyawan. Penelitian ini dibuat untuk membantu pemerintah dalam menangani pengangguran di wilayah Semarang atau di daerah yang lainnya.

E. Ucapan Terima Kasih

Demikian jurnal yang saya buat, terimakasih kepada Bapak Theopilus Bayu Sasongko, S.Kom, M,Eng., selaku dosen data mining, Bapak Ferian Fauzi Abdulloh, M.Kom. sebagai pembimbing kolaborasi riset dan juga keluarga, teman – teman saya yang sudah mendukung saya untuk memilih riset kolaborasi.

F. Referensi

[1] J. K. Ekonomi *dkk.*, "Analisis Kausalitas Indeks Pembangunan Manusia Pengangguran Pertumbuhan Ekonomi dan Ketimpangan Pendapatan di Indonesia." [Daring]. Tersedia pada: http://ejournal.unp.ac.id/students/index.php/epb/index

[2] Sherty Veronika dan A. Y. Mafruhat, "Pengaruh Pertumbuhan Ekonomi, Investasi dan Inflasi terhadap Pengangguran Terdidik di Provinsi Jawa Barat," *Jurnal Riset Ilmu Ekonomi dan Bisnis*, hlm. 139–146, Des 2022, doi: 10.29313/jrieb.vi.1290.

[3] D. Provinsi, K. Selatan, S. Fajriansyah, I. Chandriyanti, P. Studi, dan E. Pembangunan, "Analisis Pengaruh Pertumbuhan Ekonomi, Upah Minimum Provinsi (UMP) dan Tingkat Pengangguran terhadap Tingkat Kemiskinan," *JIEP: Jurnal Ilmu Ekonomi dan Pembangunan*, vol. 5, no. 2, hlm. 558–570, 2022.

[4] I. Azhami, E. Irawan, D. Suhendro, dan G. Artikel, "Implementasi Algoritma One-step Secant untuk Peramalan Pengangguran Terbuka Menurut Tamatan Pendidikan Tertinggi Implementation of One-step Secant Algorithm for Forecasting Open Unemployment by Highest Educational Graduate," *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, vol. 1, no. 3, hlm. 2828–9099, 2022, doi: 10.55123/jomlai.v1i3.946.

[5] K. Setiawan, M. Haikal, A. G. Wicaksana, dan Dermawan. Deris, "ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT PENGANGGURAN TERBUKA DI PROVINSI BANTEN 2017-2021," *Jurnal Riset Rumpun Ilmu Ekonomi (JURRIE)*, vol. 2, no. No.1, hlm. 107–120, 2023.

[6] I. Purboningtyas, R. Sari, T. Guretno, A. Dirgantara, A. Dwi, dan M. Al Haris, "ANALISIS PENGARUH TINGKAT PENGANGGURAN TERBUKA DAN INDEKS PEMBANGUNAN MANUSIA TERHADAP KEMISKINAN DI PROVINSI JAWA TENGAH ANALYSIS OF THE INFLUENCE OF OPEN UNEMPLOYMENT LEVELS AND HUMAN DEVELOPMENT INDEX ON POVERTY IN CENTRAL JAVA PROVINCE," 2020. [Daring]. Tersedia pada: www.jateng.bps.go.id

[7] K. Nisa, "Analisa Pengaruh Jumlah Penduduk terhadap Jumlah Angkatan Kerja pada Kota Bekasi Menggunakan Metode Regresi Linear," *Jl Jatiwaringin Raya*, vol. 5, no. 2, hlm. 28534471, 2021.

[8] D. Maulud dan A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, hlm. 140–147, Des 2020, doi: 10.38094/jastt1457.

[9] A. Novebrian Maharadja, I. Maulana, dan B. Arif Dermawan, "Penerapan Metode Regresi Linear Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi," 2021. [Daring]. Tersedia pada: http://jurnal.polibatam.ac.id/index.php/JAIC

[10] J. Khatib Sulaiman Dalam No, F. Jari Harianto, F. Fauzi Abdulloh, dan U. Amikom Yogyakarta, "Linear Regression Algorithm Analysis to Predict the Effect of Inflation on the Indonesian Economy," *Indonesian Journal of Computer Science Attribution*, vol. 12, no. 4, hlm. 1673.

[11] Y. Goktua Siadari dan D. Saripuna, "DATA MINING UNTUK MENGESTIMASI JUMLAH PENUMPANG PADA PT. PINEM LAU GUNA MEDAN DENGAN

MENGGUNAKAN METODERE GRESI LINEAR BERGANDA," *Jurnal CyberTech*, vol. x. No.x, 2020.

[12] N. Ariyani dan A. Z. Arifin, "PREDIKSI TINGKAT PENGANGGURAN DI KABUPATEN TUBAN TAHUN 2020 MENGGUNAKAN METODE REGRESI LINEAR SEDERHANA," vol. 03, no. 01, hlm. 6–13, 2021.

[13] D. Y. Perdana dan M. A. I. Pakereng, "Prediksi Tingkat Pengangguran Berdasarkan Data Time Series Menggunakan Regresi Linear (Studi Kasus : Kota Salatiga)," *Jurnal EMT KITA*, vol. 6, no. 2, hlm. 361–367, Jul 2022, doi: 10.35870/emt.v6i2.702.