

Prediksi Laju Pertumbuhan Ekonomi di Indonesia Menggunakan Metode Regresi Linear

1st Alif Rakasha Pratama Putra
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
1197050012@student.uinsgd.ac.id

2nd Harry Akbar Fauzan
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
harryakbarfz@gmail.com

3rd Muhammad Rizkon Fithri Wikarta
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
mrizkonfw@gmail.com

4th Rifqi Fadhur Rahman
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
rifqi09fadhur@gmail.com

5th Yunita Minarni
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
yunitaminarni12@gmail.com

6th Dian Saadillah Maylawati
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
diansm@uinsgd.ac.id

Abstract— Ekonomi merupakan bagian terpenting dari suatu negara, bahkan menjadi perhatian dunia dalam misi pengembangan SDGs. Banyak faktor yang mempengaruhi ekonomi Indonesia, seperti tingkat pengangguran terbuka, tingkat kemiskinan, rasio gini, indeks pembangunan manusia, dan jumlah provinsi yang ada. Penelitian ini ditujukan untuk melakukan prediksi terhadap nilai laju pertumbuhan ekonomi Indonesia dengan menggunakan pembelajaran mesin, yaitu algoritma regresi linear sederhana dan berganda. Pada regresi linear sederhana dengan komposisi 75 : 25, akurasi yang diperoleh beragam untuk masing-masing provinsi, tetapi dapat disimpulkan belum cukup baik. Sedangkan, pada regresi linear berganda, telah diperoleh nilai koefisien determinasi (R^2) sebesar 11.9% dengan rasio pembagian data latih dan uji adalah 80 : 20 dalam persen. Dengan adanya penelitian ini, diharapkan dapat menjadi sumber rujukan atau ide pengembangan baru untuk penelitian berikutnya guna memperoleh hasil yang lebih baik.

Keywords—ekonomi, Indonesia, prediksi, regresi linear, SDGs

I. PENDAHULUAN

Ekonomi merupakan salah satu sektor terpenting pada suatu negara, bahkan seluruh dunia. Dengan adanya pembangunan ekonomi, suatu negara dapat dengan mudah mendapatkan ketersediaan kebutuhan hidup untuk masyarakat, antara lain kebutuhan primer seperti sandang, pangan, dan papan, serta kebutuhan sekunder dan tersier untuk mengangkat derajat serta memenuhi kebahagiaan diri.

Ekonomi juga menjadi salah satu titik fokus pemerintahan dunia pada target pembangunan berkelanjutan, atau yang lebih sering dikenal sebagai *Sustainable Development Goals* (SDGs). Tepatnya pada SDGs ke-8, yang berjudul *Decent Work and Economic Growth* (Pekerjaan Layak dan Pertumbuhan Ekonomi). Banyak negara yang terlibat dalam mewujudkan misi SDGs ini, terutama Indonesia.

Indonesia memiliki beberapa aspek yang berkaitan dengan angka ekonomi, di antaranya adalah Tingkat Pengangguran Terbuka (persentase jumlah pengangguran terhadap jumlah angkatan kerja), Tingkat Kemiskinan (persentase jumlah penduduk yang berada di bawah garis kemiskinan), Rasio Gini (tingkat ketimpangan pengeluaran dan pendapatan), Indeks Pembangunan Manusia (indikator keberhasilan pembangunan kualitas hidup manusia), dan Laju

Pertumbuhan Ekonomi (indikator keberhasilan pembangunan suatu daerah).

Berdasarkan data dari Badan Pusat Statistik, pada bulan Agustus tahun 2022, Indonesia memiliki persentase tingkat pengangguran terbuka lebih besar bila dibandingkan dengan negara lain di Asia Tenggara. Nilai pengangguran tersebut dapat disebabkan oleh faktor-faktor seperti kurangnya lowongan kerja yang layak, keengganan beberapa golongan masyarakat untuk bekerja secara layak, pengetahuan dan keterampilan tentang dunia kerja yang belum memenuhi standar, atau pengaruh pandemi yang sempat melanda dunia selama dua tahun terakhir. Faktor lain dari laju ekonomi Indonesia yang masih tergolong rendah dan angka pengangguran yang cukup tinggi adalah banyaknya jumlah provinsi yang terpisah pada beberapa pulau, berakibat kepada pembangunan ekonomi yang tidak merata kepada seluruh bagian negeri Indonesia.

Dengan adanya teknologi informasi dan komunikasi, manusia dapat mempercepat aktivitas komputasi, perhitungan, dan berbagai hal-hal rumit lainnya, terutama dalam melakukan prediksi terhadap suatu data. Pemanfaatan teknologi *Machine Learning* dapat dijadikan sebagai salah satu solusi. Dengan bantuan algoritma pada model pembelajaran regresi, komputer dapat dengan cepat mengolah dan memberikan hasil kalkulasi yang dilakukan. Terdapat cukup banyak algoritma populer yang digunakan dalam penyelesaian masalah mengenai regresi, salah satunya adalah Regresi Linear.

Pada saat ini, terdapat cukup banyak penelitian yang menggunakan pembelajaran mesin untuk memprediksi suatu data tertentu dengan objek kajian yang berbeda-beda menggunakan regresi linear atau algoritma model regresi lainnya, di antaranya penelitian prediksi penyakit demam berdarah dengue di DKI Jakarta [1], prediksi penjualan produk elektronik [2], estimasi penjualan mobil [3], prediksi pemesanan bibit pohon [4], estimasi jumlah perceraian di pengadilan [5], prediksi keuntungan dan kerugian kelapa sawit [6], prediksi harga *crude palm oil* pasar domestik [7], prediksi kurs rupiah terhadap dollar [8], prediksi besaran pendapatan daerah [9], memprediksi produksi padi [10], [11], dan masih banyak lagi.

Tujuan dilaksanakan penelitian yang berjudul “Prediksi Laju Pertumbuhan Ekonomi di Indonesia Menggunakan Metode Regresi Linear” ini adalah untuk mengetahui,

menguji, dan menganalisis performa serta akurasi dari algoritma Regresi Linear dalam memperhitungkan dan memprediksi data laju pertumbuhan ekonomi di Indonesia pada tahun-tahun berikutnya.

II. METODE PENELITIAN

A. Dataset

Dataset yang digunakan sebagai data latih dan data uji pada penelitian ini diperoleh dari Badan Perencanaan Pembangunan Nasional (Bapennas), dengan rincian data meliputi tingkat kemiskinan, tingkat pengangguran terbuka, indeks pembangunan manusia (IPM), rasio gini, dan laju pertumbuhan ekonomi dari 34 provinsi di Indonesia selama sembilan tahun, yaitu dari tahun 2013-2021.

Data pada tahun 2022 tidak digunakan, karena terdapat beberapa data aspek ekonomi yang masih belum lengkap dan dipublikasikan oleh Bapennas, sehubungan dengan tahun 2022 yang belum berakhir, sehingga masih belum bisa dilakukan perhitungan. Jumlah provinsi yang dijadikan sebagai data pun masih berjumlah 34, karena perubahan jumlah provinsi menjadi 38 mulai berlaku sejak tahun 2022.

B. Preprocessing

Pada awalnya, keseluruhan data masih dalam keadaan terpisah satu sama lain, sehingga diperlukan tahap penggabungan, penyeragaman, dan integrasi data. Data yang sudah terintegrasi memiliki sebanyak 306 data, yang memiliki kolom-kolom meliputi: provinsi, tahun, tingkat kemiskinan, tingkat pengangguran terbuka, rasio gini, IPM, dan laju pertumbuhan ekonomi. Data tersusun sebagaimana pada TABEL I., dengan thn adalah tahun, t_k adalah tingkat kemiskinan, t_p adalah tingkat pengangguran terbuka, r_g adalah rasio gini, ipm adalah indeks pembangunan manusia, dan l_e adalah laju pertumbuhan ekonomi.

TABEL I. TAMPILAN DATASET

#	provinsi	thn	t_k	t_p	r_g	ipm	l_e
0	Aceh	2013	17.60	8.34	0.331	68.30	3.04
1	Sumatera Utara	2013	10.06	6.09	0.328	68.36	5.67
2	Sumatera Barat	2013	8.14	6.39	0.351	68.91	7.02
3	Riau	2013	7.72	4.19	0.393	69.91	1.05
4	Jambi	2013	8.07	2.89	0.327	67.76	6.39
...							
301	Sulawesi Barat	2021	11.85	3.13	0.366	66.36	2.56
302	Maluku	2021	16.30	6.93	0.316	69.71	3.04
303	Maluku Utara	2021	6.38	4.71	0.278	68.76	16.4
304	Papua Barat	2021	21.82	5.84	0.374	65.26	-0.51
305	Papua	2021	27.38	3.33	0.396	60.62	15.11

Pada Gambar 1, yang merupakan keterangan mengenai data yang sudah terintegrasi, tidak terdapat data yang bersifat *null*, sehingga tidak diperlukan tahap pembersihan (*data cleaning*).

```

RangeIndex: 306 entries, 0 to 305
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   provinsi              306 non-null   object  
1   tahun                 306 non-null   int64   
2   tingkat_kemiskinan    306 non-null   float64  
3   tingkat_pengangguran  306 non-null   float64  
4   rasio_gini            306 non-null   float64  
5   ipm                   306 non-null   float64  
6   laju_ekonomi          306 non-null   float64  
dtypes: float64(5), int64(1), object(1)

```

Gambar 1. Informasi Data(Sumber: Penulis)

Pada data awal, diketahui bahwa setiap kolom data memiliki format, batasan, atau skala yang berbeda. Maka

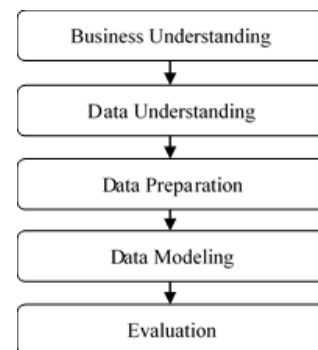
diperlukan tahap transformasi data (*data transformation*), berupa normalisasi. Hal ini ditujukan untuk menyeragamkan skala setiap data numerik, terutama pada nilai Rasio Gini yang memiliki skala 0-1, sangat berbeda dengan aspek lainnya yang nilainya berada di atas angka satu. Proses normalisasi dilakukan dengan fungsi *MinMaxScaler* pada *library sklearn*, yaitu pengubahan data menjadi rentang 0-1. Selain itu, supaya memudahkan sistem dalam mengolah data kategori, dilakukan pengubahan menjadi suatu label, menggunakan fungsi *LabelEncoder*. Hasil setelah dilakukannya *encoding* dan normalisasi disajikan pada TABEL II.

TABEL II. TAMPILAN DATASET SETELAH NORMALISASI

#	provinsi	tahun	t_k	t_p	r_g	ipm	l_e
0	0	0	0.5653	0.7616	0.7211	0.4847	0.3769
1	32	0	0.3231	0.5561	0.7145	0.4871	0.4297
2	30	0	0.2614	0.5835	0.7647	0.509	0.4568
3	24	0	0.2479	0.3826	0.8562	0.5494	0.3370
4	7	0	0.2592	0.2639	0.7124	0.4629	0.4441
...							
301	25	8	0.3806	0.2858	0.7973	0.4066	0.3673
302	18	8	0.5236	0.6328	0.6884	0.5414	0.3769
303	19	8	0.2049	0.4301	0.6056	0.5032	0.6451
304	23	8	0.7009	0.5333	0.8148	0.3624	0.3057
305	22	8	0.8795	0.3041	0.8627	0.1757	0.6192

C. Metode CRISP-DM

Cross-Industry Standard Process for Data Mining atau yang lebih dikenal dengan metode CRISP-DM, merupakan suatu standar tahapan pengerjaan dalam penelitian yang berfokus pada *Data Mining*. Meskipun begitu, standar ini pun relevan bila digunakan pada lingkup pengolahan data lain seperti pengembangan *Machine Learning* [12]. Metode CRISP-DM terdiri dari beberapa proses sebagaimana digambarkan pada Gambar 2.



Gambar 2. Tahapan Metode CRISP-DM (Sumber: Penulis)

Tahapan *Business Understanding* berisi penjelasan pemahaman dari masalah yang diteliti. Tersusun dari topik penelitian, latar belakang, rumusan masalah, tujuan penelitian, dan batasan penelitian. Tahapan *Data Understanding* berisi pemahaman mengenai data yang digunakan, membahas sumber diperolehnya data, data yang dibutuhkan, informasi mengenai data, dan sebagainya. Tahapan *Data Preparation* membahas tentang metode pengolahan data sebelum dilakukan pemodelan, atau yang sering disebut dengan *preprocessing*. Tahapan *Data Modeling* adalah pengimplementasian model atau metode kepada data untuk mendapatkan hasil yang sesuai dengan tujuan penelitian. Tahapan *Evaluation* berisi informasi mengenai akurasi, nilai kegagalan, koefisien tertentu, dan hal-hal lainnya yang berkaitan dengan pengujian [12].

D. Regresi Linear (Linear Regression)

Regresi Linear adalah salah satu algoritma regresi, yang merupakan model pembelajaran pada *Machine Learning* yang termasuk pada *supervised learning* (pembelajaran yang terstruktur dan terorganisir). Biasanya, metode regresi digunakan dalam perhitungan suatu data tertentu yang berkaitan dengan perkiraan waktu dan prediksi dalam suatu periode tertentu [1].

Dalam penyelesaian masalah, metode regresi linear melakukan pemodelan antara beberapa variabel, yang meliputi variabel bebas atau dependen sebagai x , dan variabel terikat atau independen sebagai y . Pemodelan dilakukan berdasarkan data yang telah ada sebelumnya, dengan keseluruhan data tersebut akan dianalisis dan ditemukan keterkaitan antar variabel sehingga dapat dilakukan proses prediksi suatu nilai y pada suatu nilai x tertentu [13].

Bila diperhatikan dari jumlah variabel bebas yang digunakan, regresi linear dibedakan menjadi dua jenis, meliputi regresi linear sederhana dan regresi linear berganda. Regresi linear sederhana hanya melibatkan satu variabel bebas dan satu variabel terikat. Sedangkan, regresi linear berganda memiliki lebih dari satu variabel bebas untuk melakukan prediksi pada variabel terikat [2]

$$y_t = ax + b \quad (1)$$

Persamaan (2) merupakan *formula* atau rumus dari regresi linear berganda [1], dengan keterangan:

y_t : hasil prediksi
 a : koefisien variabel bebas
 x : nilai variabel bebas
 b : konstanta

$$y_t = \beta_0 + \beta_1 x_1 + \dots + \beta_i x_i + \varepsilon \quad (2)$$

Persamaan (2) merupakan *formula* atau rumus dari regresi linear berganda [1], dengan keterangan:

y_t : hasil prediksi
 β_0 : konstanta
 β_i : koefisien variabel bebas
 x_i : variabel bebas
 i : jumlah/penomoran variabel bebas
 ε : *error term*

Untuk mengukur akurasi dan performa dari algoritma Regresi Linear, terdapat beberapa istilah khusus, seperti *R-squared* (R^2) atau koefisien determinasi, *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Root Mean Squared Error* (RMSE) [1]. Dalam analisis akurasi, semakin rendah nilai MAE, MAPE, dan RMSE maka semakin tinggi akurasinya. Tetapi, untuk nilai R^2 berbanding lurus dengan akurasi.

E. Ordinary Least Square (OLS)

Ordinary Least Square atau OLS merupakan suatu teknik analisis data yang merupakan bagian dari regresi linear berganda. Metode ini melakukan perhitungan untuk menentukan suatu garis lurus terbaik di antara sebaran data yang ada, dibentuk dengan kecenderungan mendekati garis lurus atau persamaan linear [14]. Selain itu, metode ini dapat meminimumkan jumlah kesalahan kuadrat pada proses regresi dan melakukan estimasi parameter dengan menduga koefisien regresi dari parameter tersebut [15].

Metode *Ordinary Least Square* dianggap sebagai salah satu metode terbaik dengan nilai penyimpangan atau kesalahan terkecil bila dilakukan perbandingan dengan metode analisis data lainnya [14].

III. HASIL DAN PEMBAHASAN

A. Regresi Linear Sederhana

Pada tahap pengujian dengan algoritma regresi linear sederhana, data yang digunakan tidak mengalami *preprocessing*, dikarenakan hanya menggunakan satu variabel bebas dan satu variabel terikat saja. Sehingga, tidak perlu dilakukan penyeragaman antara keduanya.

Pengujian dilakukan dengan membagi komposisi data latih dan data uji menjadi 75% : 25%. Kemudian, pengaplikasian proses regresi menggunakan fungsi *LinearRegression* yang berasal dari *library sklearn.linear_model*. Selanjutnya, dilakukan perulangan proses regresi untuk masing-masing provinsi, sehingga keseluruhan berjumlah 34 kali pengulangan. Tahun yang dilakukan prediksi adalah pada tahun 2023.

TABEL III. HASIL REGRESI LINEAR SEDERHANA

Provinsi	Hasil Prediksi Laju Ekonomi
Aceh	6.18
Sumatera Utara	4.69
Sumatera Barat	3.26
Riau	3.65
Jambi	2.99
Sumatera Selatan	7.84
Bengkulu	4.41
Lampung	4.08
Bangka-Belitung	3.55
Kepulauan Riau	2.15
DKI Jakarta	5.98
Jawa Barat	4.41
Jawa Tengah	5.46
Yogyakarta	6.76
Jawa Timur	5.22
Banten	4.62
Bali	6.08
Nusa Tenggara Barat	-16.72
Nusa Tenggara Timur	3.96
Kalimantan Barat	6.65
Kalimantan Tengah	3.53
Kalimantan Selatan	5.43
Kalimantan Timur	2.6
Kalimantan Utara	0.75
Sulawesi Utara	5.85
Sulawesi Tengah	30.55
Sulawesi Selatan	4.83
Sulawesi Tenggara	6.78
Gorontalo	4.57
Sulawesi Barat	3.14
Maluku	6.12
Maluku Utara	9.8
Papua Barat	4.45
Papua	9.79

Pada TABEL III., disajikan hasil prediksi laju pertumbuhan ekonomi pada setiap provinsi di tahun 2023. Bila dilakukan pengamatan terhadap tahun-tahun sebelumnya, data ini masih dapat dikatakan sesuai dengan pola dan tidak terlalu jauh dengan nilai pada tahun sebelumnya.



Gambar 3. Hasil Evaluasi menggunakan model OLS (Sumber: Penulis)

Untuk mempermudah pembacaan data, telah dibuat bentuk visualisasi data, sebagaimana disajikan pada Gambar 3., dapat diketahui provinsi yang memiliki laju pertumbuhan ekonomi tertinggi adalah Sulawesi Tengah, sedangkan provinsi yang memiliki laju pertumbuhan ekonomi terendah adalah Nusa Tenggara Barat.

```

PROVINSI : Jawa Barat
-----
Akurasi Laju Pertumbuhan Ekonomi :
Mean Absolute Error: 1.8702857142856861
Mean Squared Error: 9.3150708571428
Root Mean Squared Error: 3.052060100512898
R2 Score: -0.2817136223666492

```

Gambar 4. Hasil Evaluasi menggunakan model OLS (Sumber: Penulis)

Gambar 4 menampilkan salah satu perhitungan akurasi pada regresi linear sederhana yang dilakukan. Dikarenakan, pada kasus ini, terdapat perulangan pengujian sebanyak 34 kali sesuai dengan jumlah provinsi. Maka, nilai dari MAE, MSE, RMSE, dan R^2 sangat beragam.

Tetapi, bila dilakukan analisis dan penarikan kesimpulan. Akurasi dari regresi linear sederhana ini masih belum cukup baik, sebab terdapat R^2 yang bernilai negatif pada beberapa provinsi, meskipun bernilai positif tetapi masih jauh untuk mendekati satu. Selain itu, pada nilai MAE, MSE, dan RMSE masih terbilang cukup besar pada sebagian besar provinsi.

Hal ini mungkin disebabkan karena periode pada dataset yang kurang banyak, yaitu hanya sebanyak 9 tahun. Kemungkinan lainnya adalah data pada setiap tahunnya mengalami perubahan yang sembarang atau tidak berpola, sehingga model regresi yang didapatkan tidak begitu akurat.

B. Regresi Linear Berganda

Untuk tahap pengujian dan analisis data, sebelumnya telah dilakukan terlebih dahulu tahap pengolahan data *preprocessing* untuk mengurangi *noise* pada data. Pada tahapan pengujian, dilakukan berbagai skenario berdasarkan pembagian data dari komposisi data latih dan data uji, yaitu 50:50, 60:40, 70:30, 80:20, dan 90:10 (dalam persen).

Setelah itu, dilakukan eliminasi variabel menggunakan *Recursive Feature Elimination* untuk melakukan *ranking* fitur terhadap tingkat pentingnya atau besar pengaruhnya terhadap proses prediksi. Diperoleh hasil sebagaimana ditampilkan pada Gambar 4.

```

[('provinsi', False, 4),
 ('tahun', False, 2),
 ('tingkat_kemiskinan', False, 3),
 ('tingkat_pengangguran', True, 1),
 ('rasio_gini', True, 1),
 ('ipm', True, 1)]

```

Gambar 5. Hasil Proses Recursive Feature Elimination (Sumber: Penulis)

Berdasarkan hasil eliminasi pada Gambar 5, didapatkan kolom atau aspek ekonomi yang memiliki nilai *True* adalah tingkat pengangguran terbuka, rasio gini, dan indeks pembangunan manusia. Sehingga, hanya tiga variabel bebas tersebut yang akan digunakan pada proses pemodelan dan prediksi.

TABEL IV. PERBANDINGAN KOMPOSISI DATA

No	Rasio Data Latih : Data Uji	R-Squared	Adj. R-Squared
1	50% : 50%	0.116	0.098
2	60% : 40%	0.111	0.096
3	70% : 30%	0.104	0.091
4	80% : 20%	0.119	0.108
5	90% : 10%	0.074	0.064

Dari seluruh tahap pengujian, diperoleh hasil pada TABEL IV., diketahui komposisi data terbaik diperoleh rasio 80:20. Pada tahap evaluasi, dilakukan pengecekan dan analisis menggunakan *Ordinary Least Square*, menghasilkan seluruh informasi pada Gambar 6.

OLS Regression Results						
=====						
Dep. Variable:	laju_ekonomi	R-squared:	0.119			
Model:	OLS	Adj. R-squared:	0.108			
Method:	Least Squares	F-statistic:	10.77			
Date:	Thu, 22 Dec 2022	Prob (F-statistic):	1.15e-06			
Time:	11:52:38	Log-likelihood:	246.39			
No. Observations:	244	AIC:	-484.8			
Df Residuals:	240	BIC:	-470.8			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	0.2875	0.024	12.169	0.000	0.241	0.334
tingkat_pengangguran	-0.0849	0.028	-3.010	0.003	-0.140	-0.029
rasio_gini	0.0732	0.027	2.722	0.007	0.020	0.126
ipm	-0.0938	0.036	-2.591	0.010	-0.165	-0.022
=====						
Omnibus:	168.422	Durbin-Watson:	1.969			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4579.328			
Skew:	2.241	Prob(JB):	0.00			
Kurtosis:	23.745	Cond. No.	9.27			

Gambar 6. Hasil Evaluasi menggunakan model OLS (Sumber: Penulis)

Pada Gambar 6, didapatkan nilai koefisien determinasi (*R squared*) sebesar 11.9%. Dengan nilai *adjusted R squared* sebesar 10.8%. Selain itu, didapatkan pula rumus regresi dari laju ekonomi yaitu pada Persamaan (3).

$$y_t = 0.2875 + (-0.0849)x_1 + (0.0732)x_2 + (-0.0938)x_3 \quad (3)$$

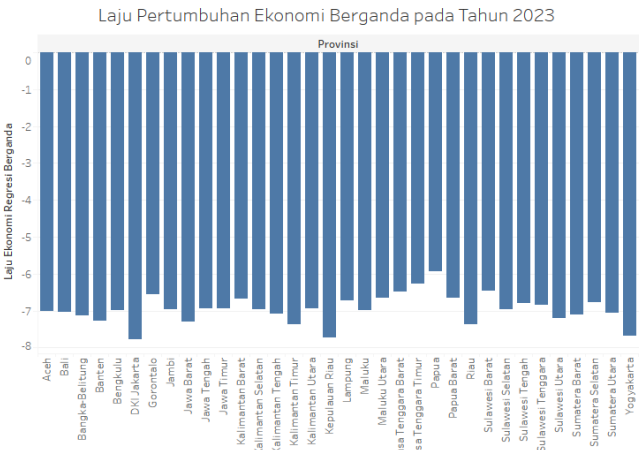
y_t : hasil prediksi
 x_1 : nilai tingkat pengangguran terbuka
 x_2 : nilai rasio gini
 x_3 : nilai indeks pembangunan manusia

Untuk mendapatkan nilai laju pertumbuhan ekonomi pada tahun berikutnya, perlu diketahui juga nilai-nilai pada variabel bebas pada tahun tersebut. Sehingga, telah dilakukan regresi linear sederhana untuk mencari nilai dari variabel-variabel tersebut sesuai dengan masing-masing provinsi. Kemudian, dilakukan substitusi nilai tersebut kepada rumus regresi yang telah didapatkan pada Persamaan (3), sehingga menghasilkan nilai prediksi sebagaimana disajikan pada TABEL V.

TABEL V. HASIL REGRESI LINEAR BERGANDA

Provinsi	Hasil Prediksi Laju Ekonomi
Aceh	-7.01
Sumatera Utara	-7.05
Sumatera Barat	-7.10
Riau	-7.38
Jambi	-6.97
Sumatera Selatan	-6.76
Bengkulu	-6.99
Lampung	-6.72
Bangka-Belitung	-7.14
Kepulauan Riau	-7.73
DKI Jakarta	-7.77
Jawa Barat	-7.30
Jawa Tengah	-6.94
Yogyakarta	-7.68
Jawa Timur	-6.95
Banten	-7.27
Bali	-7.03
Nusa Tenggara Barat	-6.48
Nusa Tenggara Timur	-6.28
Kalimantan Barat	-6.67
Kalimantan Tengah	-7.09
Kalimantan Selatan	-6.96
Kalimantan Timur	-7.38
Kalimantan Utara	-6.94
Sulawesi Utara	-7.20
Sulawesi Tengah	-6.79
Sulawesi Selatan	-6.96
Sulawesi Tenggara	-6.84
Gorontalo	-6.56
Sulawesi Barat	-6.46
Maluku	-6.99
Maluku Utara	-6.64
Papua Barat	-6.65
Papua	-5.93

Bila data prediksi pada TABEL V dibandingkan dengan data pada tahun sebelumnya, dapat dikatakan bahwa hasil tersebut tidak sesuai dan cukup berbeda dari tahun sebelumnya. Selain itu, dapat diperhatikan juga bahwa seluruh nilai tersebut bernilai negatif dan berada pada kisaran -5 sampai dengan -7, yang mana tidak mungkin bila seluruh provinsi memiliki nilai laju pertumbuhan ekonomi yang berdekatan, sedangkan pada tahun sebelumnya memiliki nilai yang beragam dan berada pada jangkauan data yang luas. Hal tersebut berkemungkinan disebabkan karena penggunaan dataset secara keseluruhan untuk membuat persamaan regresi, yang seharusnya dilakukan sebanyak 34 kali untuk masing-masing provinsi [2].



Gambar 7. Hasil Evaluasi menggunakan model OLS (Sumber: Penulis)

Gambar 7 merupakan bentuk visualisasi data dari hasil regresi pada TABEL V. Dapat diketahui provinsi yang memiliki laju pertumbuhan ekonomi tertinggi adalah Papua, sedangkan provinsi yang memiliki laju pertumbuhan ekonomi terendah adalah DKI Jakarta.

```
PROVINSI : Jawa Barat
-----
Akurasi Tingkat Pengangguran :
Mean Absolute Error: 0.7214285714285721
Mean Squared Error: 1.4577000000000349
Root Mean Squared Error: 1.2073524754602671
R2 Score: -1.2175642195853946
-----
Akurasi Rasio Gini :
Mean Absolute Error: 0.008400000000000003
Mean Squared Error: 0.0001042514285714285
Root Mean Squared Error: 0.010210358885535243
R2 Score: -0.02207282913165365
-----
Akurasi Indeks Pembangunan Manusia :
Mean Absolute Error: 0.19285714285718875
Mean Squared Error: 0.09530000000004457
Root Mean Squared Error: 0.3087069808087348
R2 Score: 0.9451582292603499
```

Gambar 8. Hasil Evaluasi Regresi Linear Sederhana (Sumber: Penulis)

Pada Gambar 8, disajikan akurasi dari perolehan nilai variabel bebas untuk Persamaan (3) untuk provinsi Jawa Barat. Tetapi, setiap provinsi memiliki nilai akurasi yang berbeda-beda. Bila disimpulkan akurasi untuk mendapatkan nilai indeks pembangunan manusia (IPM) cukup bagus, dikarenakan nilai dari R^2 mendekati satu, dan nilai dari MAE, MSE, dan RMSE cukup kecil. Sedangkan, untuk akurasi dari variabel rasio gini dan tingkat pengangguran masih belum cukup baik karena sebagian besar provinsi memiliki R^2 yang bernilai negatif atau kecil, dan nilai dari MAE, MSE, dan RMSE cukup besar.

Dapat dikatakan bahwa proses pemodelan dan prediksi yang dilakukan pada model regresi linear berganda, masih jauh dari kata sempurna. Dikarenakan nilai akurasi hanya mencapai 11.9%, dan hasil prediksi yang diperoleh tidak begitu beragam dan cukup jauh bila dibandingkan dengan data tahun sebelumnya. Kekurangan ini dapat dipengaruhi oleh beberapa penyebab, salah satunya adalah persamaan regresi yang tidak seharusnya diperoleh dari keseluruhan data, melainkan dari masing-masing provinsi [2]. Faktor lainnya antara lain kurangnya variabel ekonomi atau lainnya dari negara Indonesia pun yang mungkin dapat mempengaruhi proses regresi lebih baik lagi [3]. Selanjutnya, kurangnya metode penyempurnaan data pada tahap persiapan

ataupun pemodelan, ataupun terdapat bagian yang tidak seharusnya dilibatkan pada saat pemodelan [13]. Adapun faktor terakhir, yaitu periode waktu yang digunakan terlalu sedikit untuk dilakukan prediksi, yaitu hanya 9 tahun.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan mengenai penerapan algoritma regresi linear sederhana dan berganda menggunakan metode CRISP-DM pada kasus prediksi nilai laju pertumbuhan ekonomi di Indonesia pada tahun 2023 dan seterusnya, didapatkan akurasi yang belum terlalu baik, secara berturut-turut memiliki komposisi data latih dan data uji memiliki rasio 75% : 25% dan 80% : 20%. Akurasi pada regresi linear sederhana memiliki nilai yang beragam pada setiap provinsinya. Sedangkan, pada persamaan regresi linear berganda memiliki akurasi R^2 sebesar 11.9%. Hal ini menunjukkan bahwa algoritma regresi linear dapat digunakan untuk memprediksi laju ekonomi Indonesia, namun perlu dilakukan penyempurnaan model dan data terlebih dahulu.

Kegagalan dalam mencapai akurasi yang tinggi pada penelitian ini berkemungkinan disebabkan oleh beberapa faktor seperti kesalahan dalam pengelompokkan dan penggunaan dataset [2], yang seharusnya berfokus pada masing-masing provinsi di Indonesia. Kurangnya variabel ekonomi lain yang dapat mempengaruhi proses regresi [3]. Selain itu, diperlukannya metode-metode lain untuk mengolah data lebih baik disertai dengan pemahaman yang lebih baik [13]. Kemudian, jumlah data yang terlalu sedikit, terutama pada periode waktu.

Dengan memerhatikan faktor-faktor tersebut, diharapkan pada penelitian selanjutnya dapat dilakukan pemodelan regresi pada objek kajian yang sama ataupun berbeda, dengan algoritma regresi linear yang lebih baik dari segi akurasi, performa, dan output proses prediksi.

REFERENCES

- [1] A. R. Muhajir, E. Sutoyo, and I. Darmawan, "Forecasting Model Penyakit Demam Berdarah Dengue Di Provinsi DKI Jakarta Menggunakan Algoritma Regresi Linier Untuk Mengetahui Kecenderungan Nilai Variabel Prediktor Terhadap Peningkatan Kasus," *Fountain Informatics J.*, vol. 4, no. 2, p. 33, 2019, doi: 10.21111/fij.v4i2.3199.
- [2] R. Dewi Putri and Andri, "Prediksi Penjualan Produk Elektronik yang Terlaris pada CV. Istana Komputer Palembang menggunakan Algoritma Regresi Linear Sederhana," *J. Mantik*, vol. 6, no. 2, pp. 2254–2263, 2022.
- [3] A. Al-Fadhilah Nur Wahyudin, A. Primajaya, and A. Susilo Yuda Irawan, "Penerapan Algoritma Regresi Linear Berganda Pada Estimasi Penjualan Mobil Astra Isuzu Implementation of Double Linear Regression Algorithm On Sales Estimation of Astra Isuzu Car," vol. 19, no. 4, pp. 364–374, 2020.
- [4] D. S. O. Panggabean, E. Buulolo, and N. Silalahi, "Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, p. 56, 2020, doi: 10.30865/jurikom.v7i1.1947.
- [5] R. A. Samosir, M. F. Rozy, and A. P. Windarto, "Penerapan Algoritma Regresi Linier Berganda dalam Mengestimasi Jumlah Perceraian di Pengadilan Agama Simalungun," *TIN Terap. Inform. Nasant.*, vol. 2, no. 1, pp. 16–20, 2021, [Online]. Available: <https://ejurnal.seminar-id.com/index.php/tin>
- [6] C. E. Simbolon, "Penerapan Algoritma Regresi Linier Sederhana Dalam Memprediksi Keuntungan dan Kerugian Kelapa Sawit Pt . Sri Ulina," *J. Inf. Syst. Res.*, vol. 2, no. 2, pp. 169–172, 2021.
- [7] A. Fitri Boy, "Implementasi Data Mining Dalam Memprediksi Harga Crude Palm Oil (CPO) Pasar Domestik Menggunakan Algoritma Regresi Linier Berganda (Studi Kasus Dinas Perkebunan Provinsi Sumatera Utara)," *J. Sci. Soc. Res.*, vol. 4307, no. 2, pp. 78–85, 2020, [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [8] N. Kusumawati, F. Marisa, and I. D. Wijaya, "Prediksi Kurs Rupiah Terhadap Dollar Amerika Dengan Menggunakan Metode Regresi Linear," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 2, no. 3, pp. 45–56, 2017, doi: 10.37438/jimp.v2i3.79.
- [9] 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.
- [10] E. Triyanto, H. Sismoro, and A. D. Laksito, "Implementasi Algoritma Regresi Linear Berganda Untuk Memprediksi Produksi Padi Di Kabupaten Bantul," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 4, no. 2, pp. 66–75, 2019, doi: 10.36341/rabit.v4i2.666.
- [11] A. A. Suryanto, "Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi," *Saintekbu*, vol. 11, no. 1, pp. 78–83, 2019, doi: 10.32764/saintekbu.v11i1.298.
- [12] A. Maulana, M. Fattah, A. Voutama, N. Heryana, and N. Sulistiyowati, "Pengembangan Model Machine Learning Regresi sebagai Web Service untuk Prediksi Harga Pembelian Mobil dengan Metode CRISP-DM," *J. Ris. Komput.*, vol. 9, no. 5, pp. 2407–389, 2022, doi: 10.30865/jurikom.v9i5.5021.
- [13] N. R. Setyoningrum and P. J. Rahimma, "Implementasi Algoritma Regresi Linear Dalam Sistem Prediksi Pendaftar Mahasiswa Baru Sekolah Tinggi Teknologi Indonesia Tanjungpinang," *Pros. Semin. Nas. Ilmu Sos. dan Teknol.*, no. 4, pp. 13–18, 2022.
- [14] P. K. A. Sanjaya and I. P. J. Saputra, "ANALISIS BEBERAPA FAKTOR YANG BERPENGARUH TERHADAP KINERJA KARYAWAN BADAN USAHA MILIK DESA (BUMDes) GENTHA PERSADA TIBUBENENG KABUPATEN BADUNG: ORDINARY LEAST SQUARE MODEL," *Prosiding*, no. 2, pp. 19–36, 2020, [Online]. Available: <https://ojs.stimihandayani.ac.id/index.php/PROSIDI NG/article/view/366>
- [15] I. D. Syahputra and A. Karim, "Perbandingan Metode Ordinary Least Square (OLS) dan Regresi Robust," *Semin. Nas. Pendidikan, Sains dan Teknol. Fak. Mat. dan Ilmu Pengetah. Alam Univ. Muhammadiyah Semarang*, pp. 127–131, 2016.