

# Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda

Ardytha Luthfiarta<sup>1</sup>, Aris Febriyanto<sup>2</sup>, Heru Lestiawan<sup>3</sup>, Wibowo Wicaksono<sup>4</sup>

Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

e-mail: <sup>1</sup>ardytha.luthfiarta@dsn.dinus.ac.id, <sup>2</sup>111201408353@mhs.dinus.ac.id,

<sup>3</sup>heru.lestiawan@dsn.dinus.ac.id, <sup>4</sup>wibowo.wicaksono@dsn.dinus.ac.id

Diterima: 15 Agustus 2019; Direvisi: 6 Desember 2019; Disetujui: 6 Mei 2020

## Abstrak

Kondisi cuaca memiliki kecenderungan berubah, untuk itu badan meteorologi bekerja memprediksi perkiraan cuaca untuk dapat digunakan sebagai peringatan dini jika muncul perubahan cuaca mendadak atau bahkan ekstrem. Dengan memprakirakan cuaca yang datang mendadak secara akurat, maka dapat mengambil langkah pencegahan agar dapat meminimalkan kerugian yang akan terjadi. Diperlukan beberapa variable atau parameter yang relevan untuk dapat memodelkan data dengan baik sehingga hasil prediksinya menjadi lebih akurat. Salah satu pendekatan pemodelan data untuk prediksi cuaca adalah supervised learning dengan teknik estimasi. Estimasi memberikan prediksi nilai pada atribut target atau class attribute yang bertipe numerical. Regresi linear berganda merupakan salah satu algoritma estimasi yang handal untuk memprediksi cuaca. Empat variable independent yakni, suhu, kelembaban, tekanan, dan kecepatan angin digunakan untuk memprakirakan curah hujan sebagai variable dependent. Data yang digunakan adalah data BMKG dari Stasiun Meteorologi Ahmad Yani Semarang tahun 2015-2017. Nilai koefisien determinasi  $R^2$  sebesar 25.5 persen menunjukkan bahwa keempat variabel yang digunakan secara bersamaan dapat menjelaskan nilai curah hujan sebagai variable dependent.

**Kata kunci:** hujan, prakiraan cuaca, regresi linear berganda.

## Abstract

Weather conditions tends to change very rapidly. For this reason, the meteorological agency works to predict weather forecasts in order to provide early warning in the event of sudden or even extreme weather changes. By accurately predicting sudden weather, preventative measures can be taken to minimize the losses that will occur. Some relevant variables or parameters are needed to perform good data modelling, hence to increase the accuracy of the prediction. One approach to perform data modeling for weather prediction is supervised learning with estimation techniques. Estimates provide predictive values on target attributes or class attributes of a numerical type. Multiple linear regression is one estimation algorithm that is reliable for predicting weather. In this research, four independent variables namely temperature, humidity, pressure, and wind speed are used to forecast rainfall as the dependent variable. The data used are BMKG data from Ahmad Yani Semarang Meteorological Station in 2015-2017. The result shows coefficient of determination  $R^2$  of 25.5 percent. This means that the four variables used together can explain the value of rainfall as a dependent variable.

**Keywords:** rainfall, weather forecast, multiple linear regression.

## 1. PENDAHULUAN

Cuaca merupakan kondisi udara yang diamati dalam periode yang relatif singkat atau pendek pada teritori yang terbatas. *World Climate Conference* menyebutkan bahwa cuaca merupakan keadaan atmosfer yang diukur dengan kompleks meliputi perubahan, perkembangan, dan muncul atau hilangnya suatu fenomena udara[1]. Cuaca mempengaruhi kehidupan manusia di berbagai aspek [2] [3] [4]. Kondisi cuaca dapat mempengaruhi kegiatan yang dilakukan diluar ruangan sampai pada kegiatan sehari-hari seperti pakaian yang sedang di jemur. Menilik potensi perubahan yang fluktuatif, akurasi prakiraan cuaca merupakan salah satu studi yang utama [5]. Atas dasar pertimbangan tersebut, para peneliti terus mengembangkan model prakiraan yang lebih akurat. Prakiraan cuaca di Indonesia diterbitkan melalui Badan Meteorologi Klimatologi dan Geofisika (BMKG) untuk memprediksi cuaca, dengan demikian tiap orang dapat mempersiapkan diri dari lebih baik menghadapi kondisi cuaca yang tidak diinginkan.

Prakiraan cuaca adalah proses dari pengumpulan data dari kondisi atmosfer, yang terdiri dari temperatur, kelembaban, hujan, kecepatan angin, dan arah[5]. Faktor-faktor tersebut kemudian diteliti dan dicocokkan dengan cuaca hari, bulan, bahkan tahun sebelumnya sehingga mendapatkan perkiraan cuaca yang paling akurat. Proses meneliti dan mencocokkan data dalam jumlah yang besar ini termasuk dalam cabang ilmu komputer yaitu *data mining*.

*Data mining* sebagai salah satu trend analisa data berperan melakukan pengenalan pola data dari sebuah *data warehouse*. Pola data ini sering disebut sebagai *knowledge* yang dapat digunakan untuk menentukan atau merumuskan berbagai kebijakan atau keputusan[6].

*Data mining* sering pula didefinisikan sebagai proses *data recognizing* dalam tumpukan data yang besar atau disebut *big data*. Beberapa disiplin ilmu lain yang memiliki irisan dengan *data mining* diantaranya *artificial intelligence*, *signal processing*, *image processing*, dan *text mining*. Semuanya membahas tentang analisa pola data untuk tujuan pendukung keputusan.7

Analisis regresi adalah salah satu teknik data mining dengan basis statistik untuk pemodelan data dengan konsep persamaan garis linier. Proses analisa regresi dilakukan dengan cara menganalisa hubungan dua atau lebih *independent variables* atau disebut variabel bebas atau variabel prediktor. Variabel ini secara umum di notasikan dengan  $x$ . Variabel bebas tersebut digunakan untuk menentukan nilai dari variabel bebas atau disebut *dependent variable*. Analisa regresi yang memiliki satu variabel bebas sering disebut dengan regresi linear sederhana. Adapun regresi dengan jumlah variabel bebas dua atau lebih untuk penyelesaiannya dengan regresi linear berganda [8].

Beberapa penelitian terkait prediksi cuaca telah dilakukan dengan menggunakan analisa regresi [9,10,11,12,13]. Beberapa saran terkait pengembangan penelitian prediksi adalah penggunaan berbagai atribut yang relevan seperti suhu [5], kelembapan udara [11], tekanan udara [13], kecepatan angin [14], sehingga diharapkan dapat menemukan pola yang lebih detail dan hasil yang lebih akurat[15]. Penelitian ini akan menganalisa data histori cuaca untuk menentukan pola terjadinya hujan yang diwakili dengan atribut curah hujan, dengan melibatkan empat atribut dependent yakni, suhu, kelembapan, kecepatan angin dan kelembapan.

## 2. METODE PENELITIAN

### 2.1 Data Penelitian

Dataset cuaca yang digunakan bertipe data kuantitatif dari BMKG sebanyak 1.096 data yang dimulai dari awal Januari 2015 hingga akhir Desember 2017. Data tersebut memiliki variable tanggal, suhu rata-rata, curah hujan, tekanan udara rata-rata, kelembapan rata-rata, kecepatan angin rata-rata dan prakiraan pada hari tersebut. Detail dataset yang digunakan dijelaskan dalam tabel 1.

### 2.2 Keterangan Variabel Dependent dan Variabel Independent

Sejumlah lima variabel digunakan meliputi: suhu udara(1), tekanan udara(2), kelembapan udara(3), kecepatan angin(4), dan curah hujan(5). Dalam hal ini, empat variabel pertama digunakan sebagai variabel independen, sedangkan variabel kelima yakni curah hujan digunakan sebagai variabel dependen atau target.

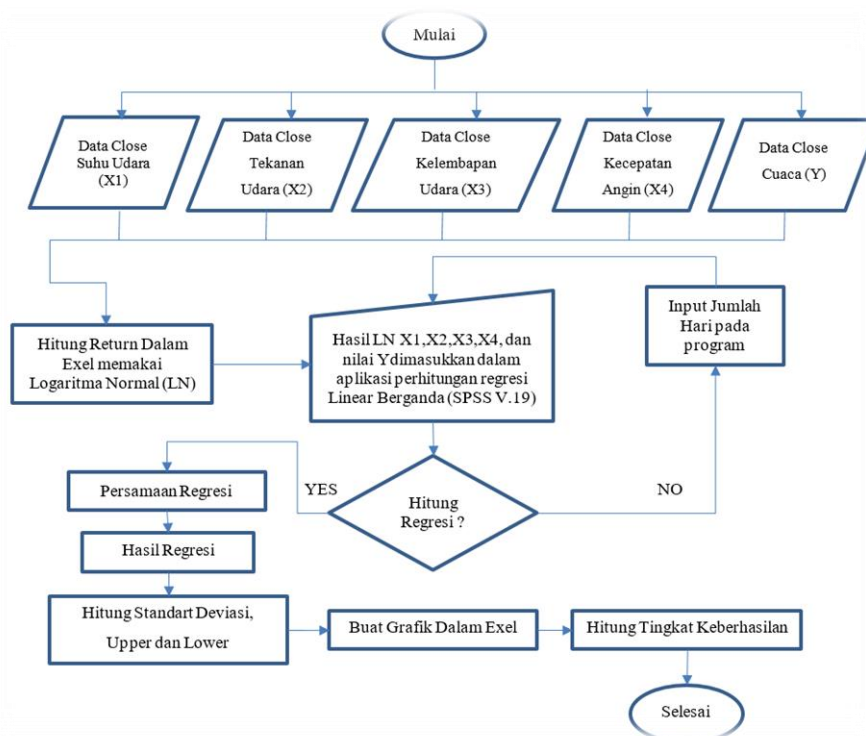
Beberapa variabel pada *raw data* tidak digunakan dalam analisa regresi karena tidak memiliki relevansi signifikan dengan curah hujan, diantaranya adalah kolom atau variabel tanggal. Sehingga pada langkah selanjutnya hanya menyisakan lima variabel. Sebelum melakukan perhitungan, data dinormalisasi.

Normalisasi dilakukan dengan mengubah nilai variabel bebas curah hujan menjadi angka. Jika label curah hujan keterangannya “Hujan” maka bernilai 0, jika label curah hujan bernilai “Mendung” maka dirubah menjadi 1, dan jika label curah hujan bernilai “Terang” maka dirubah menjadi 2. Hal ini sebagaimana ditunjukkan pada kolom terakhir pada tabel 1.

Tabel 1. Data penelitian

No.	Tanggal	Suhu Udara	Tekanan Udara	Kelembapan Udara	Kecepatan Angin	Label Curah Hujan	Curah Hujan
1	1/01/2015	26.9	1009.4	80	4	Mendung	1
2	2/01/2015	27.3	1010.1	83	5	Hujan	0
3	3/01/2015	27.0	1010.9	80	3	Mendung	1
4	4/01/2015	26.9	1009.1	81	5	Hujan	0
5	5/01/2015	27.3	1009.5	80	4	Cerah	2
6	6/01/2015	28.1	1009	78	4	Cerah	2
7	7/01/2015	27.8	1008.9	77	3	Cerah	2
8	8/01/2015	28.7	1010.5	77	4	Cerah	2
9	9/01/2015	28.2	1009.7	75	4	Cerah	2
10	10/01/2015	27.8	1009.7	69	4	Cerah	2
...	...	...	...	...	...	...	...
1096	31/12/2017	27	1008.6	84.25	3.9	Cerah	2

### 2.3 Kerangka berpikir



Gambar 1. Alur penelitian

Penelitian untuk prediksi cuaca dengan Regresi Linear Berganda memiliki alur kerja sebagaimana ditunjukkan pada gambar 1. Penelitian dimulai dengan data masukan yang telah didefinisikan pada pembahasan sebelumnya. Data tersebut diproses dengan menggunakan file excel untuk menghitung Ln dari data tersebut. Hasil perhitungan kemudian dimasukkan ke dalam SPSS. Keluaran dari SPSS adalah persamaan linear regresi berganda. Persamaan ini kemudian dievaluasi. Validasi data pada hasil akhir regresi dilakukan dengan menggunakan uji korelasi sebagai standar dalam tahap evaluasi.

### 3. HASIL PEMBAHASAN

#### 3.1. Uji Asumsi Klasik

##### A. Uji Multikolinieritas

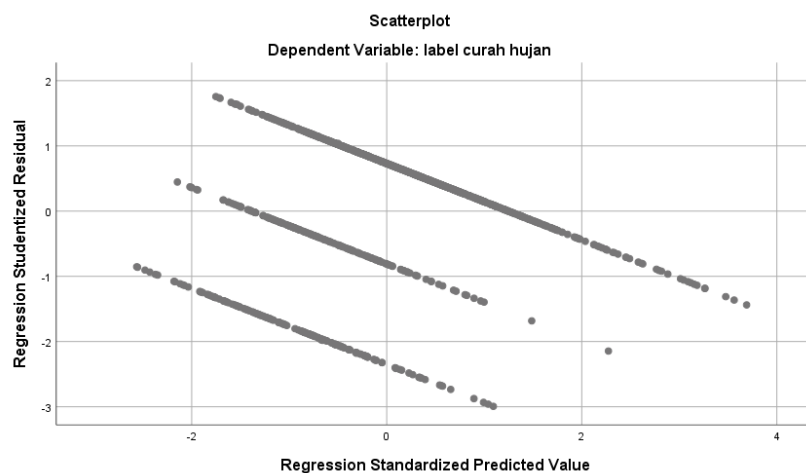
Berdasarkan tabel 2, model regresi yang dibangun cukup baik dalam uji multikolinieritas, ditunjukkan pada setiap variabel bebas memiliki nilai toleransi diatas 10 persen (0,1). Nilai VIF yang didapat memperlihatkan bahwa tiap variabel kurang dari 10. Dengan demikian dapat diambil asumsi dalam model regresi yang terbentuk tidak muncul multikolinieritas pada masing-masing variabel bebas.

Tabel 2. Uji Multikolinieritas

Variabel	Tolerance	VIF
Suhu	0,593	1,687
Tekanan	0,790	1,266
Kelembapan	0,464	2,155
Kecepatan angin	0,791	1,264

##### B. Uji Heteroskedastisitas

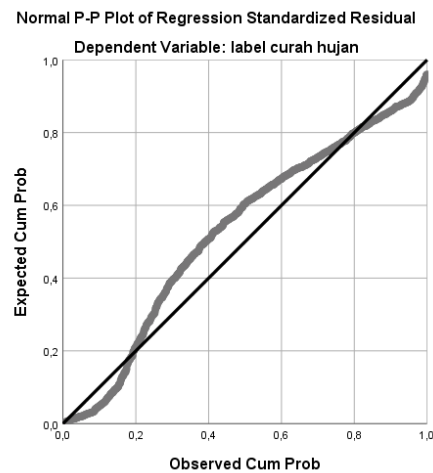
Pada gambar 2, *Scatterplot Dependent Variable* mempunyai sebaran yang terpusat, tidak nampak sebaran disekitar nilai  $Y=0$ . Hasil dari grafik *scatterplot* tersebut nampak model regresinya bebas dari gangguan heteroskedastisitas .



Gambar 2. Hasil Pengujian Heteroskedastisitas

### C. Uji Normalitas

Dari kurva dibawah yakni pada gambar 3 nampak pola grafik normal.



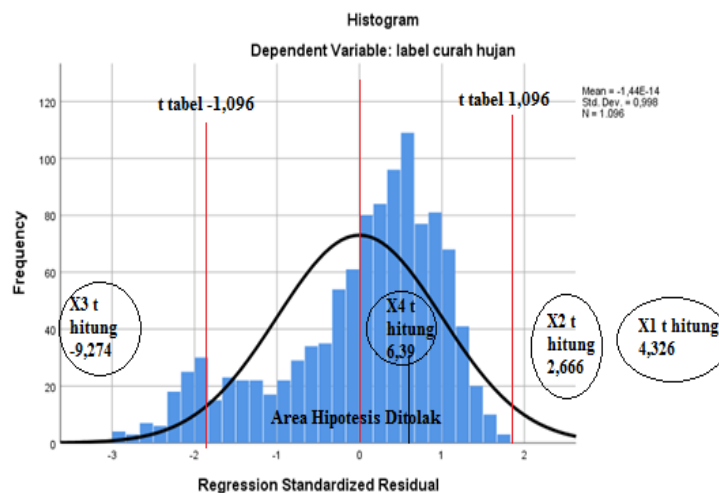
Gambar 3. Hasil Uji Normalitas

Sebaran titik dari hasil uji diatas nampak mengikuti diarea garis diagonal. Dengan demikian dapat ditarik sebuah kesimpulan pada regresi modelnya memenuhi asumsi normalitas dan dapat digunakan.

### 3.2. Uji Koefisien Regresi Linier Berganda

#### A. Uji t

Gambar 4 dibawah ini merupakan histogram hasil uji t, yang menunjukkan bahwa variabel kecepatan angin (X4) masuk dalam area hipotesis yang ditolak. Sedangkan untuk tiga variabel lainnya masuk area hipotesis yang diterima.



Gambar 4. Hasil Uji t

#### B. Uji Simultan (Uji F)

Gambar diatas menunjukkan nilai F hitung = 93,411, sementara angka signifikansi (P value) = 0,000. Derajat signifikansi mencapai 95% atau ( $\alpha=0,05$ ). Angka signifikansi (P value) sebesar  $0,000 < 0,05$ . Merujuk pada perbandingan, maka  $H_0$  ditolak, dengan demikian berarti variabel-variabel yang digunakan bersama berpengaruh secara signifikan atas variabel curah hujan.

Tabel 3. Koefisiensi Determinasi

ANOVA*						
	Model	Sum of Squares	df	Mean Square	F	Sig.
1	Regression	157,971	4	39,493	93,411	0,000 <sup>b</sup>
	Residual	461,261	1091	0,423		
	Total	619,233	1095			

### 3.3. Koefisien Determinasi

Tabel 4. Koefisiensi Determinasi

R	R Square	Adjusted R Square	Std. Error of the Estimate
0,505 <sup>a</sup>	0,255	0,252	0,650

Hasil Adjusted R<sup>2</sup> =0,255. Disimpulkan, pada keempat variabel independent dapat digunakan untuk menjelaskan variabel dependennya yakni, curah hujan dengan prosentase 25,5%.

### 3.4. Analisis Regresi Linier Berganda

Tabel 5. Hasil analisis regresi linear berganda

Model	B	Std.Error	Beta	t	Sig
suhu	0,104	0,024	0,147	4,326	0,000
tekanan	0,038	0,014	0,078	2,666	0,008
kelembapan	-0,032	0,003	-0,356	-9,274	0,000
kecepatan angin	0,011	0,018	0,019	,639	0,523

Berdasarkan tabel 4 diatas, persamaan (1) regresi linier berganda diuraikan :

$$Y = 0,147 X_1 + 0,078 X_2 - 0,356 X_3 + 0,019 X_4 \quad (1)$$

Penjelasan dari persamaan diatas adalah :

1. Variabel suhu udara ( $X_1$ ) = 0,147, memiliki nilai yang positif menunjukkan semakin tinggi suhu udara berbanding lurus terhadap curah hujan.
2. Tekanan udara ( $X_2$ ) = 0,078 juga positif, berarti semakin tinggi tekanan udara yang ada berdampak pada pengaruh curah hujan yang juga makin tinggi.
3. Kelembapan udara ( $X_3$ ) bernilai positif disimpulkan semakin tinggi kelembapan udara yang ada di BMKG Stasiun Ahmad Yani Semarang, maka tidak berarti semakin tinggi pula pengaruh curah hujan.
4. Nilai 0,019 pada variabel kecepatan angin ( $X_4$ ) adalah bernilai positif sehingga dapat dikatakan bahwa semakin tinggi kecepatan angin yang ada di BMKG Stasiun Ahmad Yani Semarang, maka akan semakin tinggi pula pengaruh curah hujan.

### 3.5. Ukuran Error (RMSE)

$$\begin{aligned}
 RMSE &= \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} & (2) \\
 &= \sqrt{\frac{14,620}{1096}} \\
 &= 0,010
 \end{aligned}$$

$\hat{y}_i$  didapat dari persamaan regresi  $Y = 0,147 X_1 + 0,078 X_2 - 0,356 X_3 + 0,019 X_4$ . Dengan memasukkan  $X_1$  (suhu),  $X_2$  (tekanan),  $X_3$  (kelembapan),  $X_4$  (kecepatan angin) ke dalam persamaan tersebut.

Hasil dari persamaan (2) dari perhitungan standar error tergolong kecil yaitu 0,010. Dikatakan kecil, karena standar error yang didapat masih di kisaran 0,0 – 1,0.

## 4. KESIMPULAN

Dari Persamaan regresi linier berganda  $Y = 0,147X_1 + 0,078X_2 - 0,356X_3 + 0,019X_4$ , didapatkan analisis regresi linier berganda, bahwa variabel suhu udara berpengaruh positif dan signifikan terhadap koefisien regresi sebesar 0,147 (14,7%). Variabel tekanan udara memiliki pengaruh positif dan signifikan terhadap koefisien regresi sebesar 0,078 (7,8 %). Untuk kelembapan udara berpengaruh negatif namun masih signifikan terhadap koefisien regresi sebesar -0,356 (-35,6 %). Sedangkan variabel kecepatan angin meski positif namun tidak signifikan terhadap curah hujan. Dapat dilihat dari  $P \text{ value} = 0,523 > 0,05$ . Uji Hipotesis menggunakan uji F (secara simultan), dapat ditarik kesimpulan empat variable yaitu suhu udara ( $X_1$ ), tekanan udara ( $X_2$ ), kelembapan udara ( $X_3$ ), dan kecepatan angin ( $X_4$ ) terhadap variabel terikat curah hujan ( $Y$ ) secara bersama-sama berpengaruh terhadap curah hujan. Berdasarkan  $F \text{ hitung} = 93,411$  dimana  $P \text{ Value}$  sebesar  $0,000 < 0,05$ , didapatkan nilai koefisien determinasi ( $R^2$ ) = 0,255. Dengan demikian 25,5% perubahan variabel curah hujan dapat dijelaskan keepmat variabel independen secara bersama-sama, sedangkan 74,5 persen lainnya dijelaskan oleh variabel lainnya.

## 5. SARAN

Untuk perbaikan penelitian ini, ada beberapa saran diantaranya penggunaan data yang lebih banyak dalam hal jumlah record sehingga diharapkan memperoleh hasil regresi yang lebih baik dan dengan asumsi untuk dapat meminimalkan error sehingga prediksi cuaca lebih akurat. Selain jumlah data, penggunaan variable lain yang relevan yang belum dimasukkan ke dalam analisa regresi ini dapat menghasilkan persamaan yang lebih detail.

## DAFTAR PUSTAKA

- [1] Fadholi, A. (2013). Persamaan Regresi Prediksi Curah Hujan Bulanan Menggunakan Data Suhu dan Kelembapan Udara di Ternate, *13*(1), 7–16.
- [2] Badhiye S. S., Wakode B. V., and Chatur P. N., "Analysis of Temperature and Humidity Data for Future value prediction," *International Journal of Computer Science and Information Technologies*, vol. 3, no. 1, pp. 3012-3014, 2012.

- 
- [3] Gaurav J and Dr. Sunil R.Gupta Sawale, "Use of Artificial Neural Network in Data Mining for Weather Forecasting," *International Journal of Computer Science And Applications*, vol. 6, no. 2, pp. 283-286, April 2013.
- [4] Meghali A. Kalyankar and Prof. S. J. Alaspurkar, "Data Mining Technique to Analyse the Metrological Data," *International Journal of Advanced Research in Computer and Software Engineering*, vol. 3, no. 2, pp. 114-118, February 2013.
- [5] Nurmahaludin. (2014). Analisis Perbandingan Metode Jaringan Syaraf Tiruan Dan Regresi Linier Berganda Pada Prakiraan Cuacah. *Jurnal INTEKNA*, Tahun XIV, No. 2 : 102-209.
- [6] Kamus Besar Bahasa Indonesia (KBBI). [Online]. <http://kbbi.web.id/lembap>
- [7] (2016, January) Ilmu Sosial - Kumpulan Ilmu Pengetahuan Sosial. [Online]. <http://www.ilmusocial.com/unsur-cuaca-dan-iklim/>
- [8] Kamus Besar Bahasa Indonesia (KBBI). [Online]. <http://kbbi.web.id/tekan>
- [9] Apriantoro, Y. (2010). Analisa Pengaruh Cuaca Terhadap Perubahan Elektris PLN Untuk Perkiraan Peyediaan Beban Harian Pada Wilayah Jakarta Banten. *Skripsi*, 35.
- [10] Budiman Dan Artesya. (2015). Aplikasi Data Mining Menggunakan Multiple Linear Regression Untuk Pengenalan Pola Curah Hujan. Kalimantan Selatan: Kumpulan Jurnal, Ilmu Komputer (KLIK). Vol. 02, No.01
- [11] Rachman, A. (2014). Model Peramalan Konsumsi Bahan Bakar Jenis Premium Di Indonesia Dengan, 166–176.
- [12] Fadholi, A. (2011). Pemanfaatan Suhu Udara Dan Kelembapan Udara Persamaan Regresi Untuk Simulasi Prediksi Total Hujan Bulanan Di Pangkalpinang.
- [13] Putri, A., Syafrialdi, Y., Mining, D., & Berganda, R. L. (2017). Analisa Pengaruh Temperatur Terhadap Titik Embun , Jarak Pandang , Kecepatan Angin , dan, 18–19.
- [14] Matematika, S., Pradipta, N. S., Sembiring, P., & Bangun, P. (2013). Analisis Pengaruh Curah Hujan, *1*(5), 459–468.
- [15] Sanjay D. Sawaitul, Prof. K. P. Wagh, and Dr. P. N. Chatur, "Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, no. 1, pp. 110-113, January 2012.
-