

## **Forecasting Model Penyakit Demam Berdarah *Dengue* Di Provinsi DKI Jakarta Menggunakan Algoritma Regresi Linier Untuk Mengetahui Kecenderungan Nilai Variabel Prediktor Terhadap Peningkatan Kasus**

**Aji Rahmat Muhajir<sup>1)\*</sup>, Edi Sutoyo<sup>2)</sup>, Irfan Darmawan<sup>3)</sup>**

*Prodi Sistem Informasi, Fakultas Rekayasa Industri, Universitas Telkom<sup>1),2),3)</sup>  
[ajirahmat@student.telkomuniversity.ac.id](mailto:ajirahmat@student.telkomuniversity.ac.id)<sup>1)</sup>\*, [edisutoyo@telkomuniversity.ac.id](mailto:edisutoyo@telkomuniversity.ac.id)<sup>2)</sup>,  
[irfandarmawan@telkomuniversity.co.id](mailto:irfandarmawan@telkomuniversity.co.id)<sup>3)</sup>*

### **Abstrak**

*Di Indonesia khususnya di Provinsi DKI Jakarta sampai saat ini Demam Berdarah Dengue masih merupakan masalah kesehatan masyarakat yang utama. Meski sudah ada beberapa langkah untuk mengatasi penyebaran penyakit Demam Berdarah Dengue (DBD), namun harus ada metode analisis untuk melakukan peramalan terhadap kasus DBD menggunakan serangkaian data yang ada, dan memperkirakan nilai data dimasa yang akan datang. Penelitian ini bertujuan untuk membuat model forecasting peningkatan jumlah kasus Demam Berdarah Dengue menggunakan algoritma regresi linear dan melakukan analisis pengaruh dari temperatur, kelembapan dan curah hujan dalam kenaikan kasus penyakit Demam Berdarah Dengue di Provinsi DKI Jakarta dari model regresi yang dibuat. Data DBD yang digunakan merupakan dataset pemantauan penyakit endemik yang diperoleh dari Dinas Kesehatan DKI Jakarta sedangkan data cuaca merupakan dataset yang didapat dari Dinas Lingkungan Hidup DKI Jakarta. Dari model regresi yang dibuat diperoleh nilai  $R^2$  sebesar 0.3622, hal tersebut menunjukkan presentase pengaruh variabel predictor terhadap kasus demam berdarah sebesar 36.22%, sedangkan 63.78% dipengaruhi oleh faktor lain diluar variabel independen tersebut. Setelah melakukan uji simultan, dapat disimpulkan bahwa temperatur, kelembapan, dan curah hujan secara bersama-sama berpengaruh terhadap kenaikan jumlah kasus demam berdarah di Provinsi DKI Jakarta. Selanjutnya uji parsial membuktikan bahwa, kelembapan dan curah hujan memiliki pengaruh signifikan terhadap kenaikan kasus demam berdarah, sedangkan untuk variabel bebas, temperatur terbukti tidak memiliki pengaruh yang signifikan terhadap kenaikan kasus demam berdarah dengue di Provinsi DKI Jakarta.*

**Kata kunci:** *Data Mining, Predictive Mining, Regresi Linier, Demam Berdarah Dengue*

### **Abstract**

*[Forecasting Model of Dengue Hemorrhagic Fever in DKI Jakarta Using Linear Regression Algorithm to Know Trends of Predictor Variable Value for Case Increasing] In Indonesia specifically in DKI Jakarta Province, Dengue fever is still the main public health problem. Although there are already several steps to overcome the spread of Dengue Fever (DHF), there still needs to be an analytical method to forecast the increase dengue cases using and estimated data values in the future. This study aims to make a forecasting model for increasing the number of cases of Dengue Fever using a linear regression algorithm and analyzing the effect of temperature, humidity and rainfall in the case of Dengue Hemorrhagic Fever in DKI Jakarta Province from a regression model made. The DHF data used is an endemic disease monitoring dataset obtained from the DKI Jakarta Health Office while the weather data is a dataset obtained from the DKI Jakarta Environmental Service. From the regression model made, the value of  $R^2$  is 0.3622, it shows the percentage of the influence of temperature, humidity and rainfall on cases of dengue fever is 36.22%, while 63.78% is influenced by other factors outside the independent variable. After conducting a simultaneous test, it can be concluded that temperature, humidity and rainfall together, influence the increase in the number of dengue cases in DKI Jakarta Province. Then the partial test proves that humidity and rainfall have a significant influence on the increase in dengue cases, whereas for temperature independent variables proved that no significant effect on the increase in cases of dengue hemorrhagic fever in DKI Jakarta Province.*

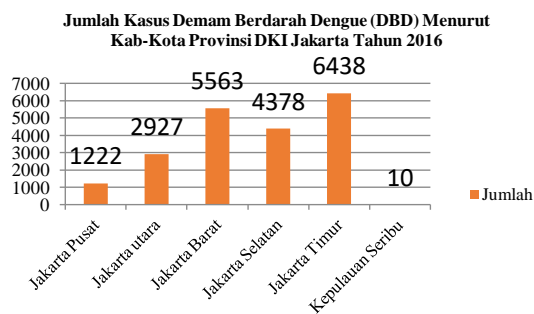
**Keywords:** *Data Mining, Predictive Mining, Linear Regression, Dengue Fever (DHF)*

## 1. PENDAHULUAN

Demam Berdarah *Dengue* (DBD) adalah penyakit infeksi yang disebabkan oleh virus *dengue* ditularkan melalui gigitan nyamuk *Aedes Aegypti* dan merupakan penyakit yang paling cepat menyebar di dunia dalam 50 tahun terakhir. Asia Tenggara dan wilayah Pasifik Barat merupakan wilayah yang paling banyak terjadi kasus demam berdarah di dunia, menurut WHO negara yang berada di Asia Tenggara dan wilayah Pasifik Barat merupakan negara-negara yang menanggung hampir 75% dari total penyakit DBD saat ini secara global, dengan sekitar 1,8 miliar populasi yang beresiko terkena DBD[1].

Di Indonesia sampai saat ini Demam Berdarah *Dengue* masih merupakan masalah kesehatan masyarakat yang utama, hal tersebut dikarenakan meningkatnya mobilitas dan kepadatan penduduk, jumlah penderita, dan luas daerah penyebarannya semakin bertambah. Penyakit DBD merupakan penyakit menular yang pada umumnya menyerang pada usia anak-anak umur 15 tahun dan juga bisa menyerang pada orang dewasa[2].

Pada tahun 2016 menjadi tahun dengan jumlah penderita DBD terbanyak di Provinsi DKI Jakarta sebanyak 22.692, jauh lebih besar dari tahun 2015 sebanyak 4194 kasus. Angka tertinggi penderita DBD di provinsi DKI Jakarta pada tahun 2016 berada di wilayah Jakarta Timur dengan jumlah kasus sebesar 6.438 kasus, dan jumlah kasus terendah sebanyak 10 di wilayah Kepulauan Seribu Dan terjadinya peningkatan *Incidence Rate* (IR) sebagai indikator frekuensi serangan, atau angka kejadian kasus penderita Demam Dengue dalam populasi 100 ribu penduduk di Provinsi DKI Jakarta, sebesar 81,7 per-100 ribu penduduk pada tahun 2015, dan pada tahun 2016 meingkat pesat sebesar 220 per-100 ribu penduduk dibandingkan *Incidence Rate* DBD tahun sebelumnya[3]. Berikut merupakan perbandingan jumlah kasus DBD di Provinsi DKI Jakarta berdasarkan wilayah Kabupaten-Kota pada tahun 2016 dalam diagram.



**Gambar 1.** Kasus DBD 2016 (Jakarta Provincial Health Agencies, 2016 [3])

Meski sudah ada beberapa langkah untuk mengatasi penyebaran penyakit DBD, namun harus ada metode analisis untuk melakukan peramalan terhadap peningkatan kasus DBD menggunakan serangkaian data yang ada, dan melakukan prediksi (*forecasting*)

memperkirakan nilai data dimasa yang akan datang menggunakan algoritma regresi linier. Regresi linier merupakan salah satu algoritma data mining yang digunakan untuk melakukan forecasting berdasarkan histori data yang pernah terjadi. Algoritma regresi linier telah digunakan dibeberapa penelitian sebelumnya, pada penerapannya di Provinsi Yogyakarta untuk melakukan forecasting menggunakan time series data dan analisa terhadap resiko wilayah endemik [4], degan begitu diharapkan prediksi yang dilakukan dapat memberikan masukan kepada Dinas Kesehatan Provinsi DKI Jakarta dalam mengambil tindakan dengan berkaca pada kejadian yang akan terjadi lewat teknik peramalan menggunakan regresi ini, selain itu analisis menggunakan regresi diharapkan dapat membantu Dinas Kesehatan Provinsi DKI Jakarta dalam melakukan antisipasi kenaikan kasus DBD dan potensi wilayah endemik di Provinsi DK Jakarta dengan memperhatikan gejala-gejala alam yang telah ditentukan sebagai variable yang berpengaruh terhadap kenaikan kasus demam berdarah.

Penelitian ini bertujuan untuk melakukan prediksi (*forecasting*) peningkatan kasus penyakit DBD di Provinsi DKI Jakarta, variable regresi yang digunakan merupakan indikator yang berkaitan dengan DBD dan persebarannya di Provinsi DKI Jakarta. Prediksi dilakukan di Provinsi DKI Jakarta dengan mengambil data dari 4 kota administratif, Jakarta Pusat, Jakarta Timur, Jakarta Selatan dan Jakarta Utara. Penelitian ini bertujuan untuk membuat model prediksi (*forecasting model*) kasus penyakit DBD di Provinsi DKI Jakarta, atribut yangdigunakan merupakan faktor cuaca yang berkaitan dengan DBD, pada penelitian sebelumnya mengenai demam berdarah yang menggunakan metode *data mining*, digunakan beberapa atribut cuaca seperti temperatur dan curah hujan [5], sedangkanpada penelitian di Cagayan de Oro, Filipina variabel temperatur, kelembapan, dan curah hujan digunakan untuk memprediksi DBD menggunakan regresi linear [6].

## 2. BAHAN DAN METODE

### 2.1 Demam Berdarah *Dengue*

Demam Berdarah *Dengue* (DBD) adalah penyakit infeksi yang disebabkan oleh virus *dengue* dan mengakibatkan *spectrum* manifestasi klinis yang bervariasi antara yang paling ringan, demam *dengue*, DBD dan demam *dengue* yang disertai renjaan atau *dengue shock syndrome* (DSS), virus ditularkan melalui nyamuk *Aedes aegypti* dan *Aedes albopictus* yang telah terinfeksi sebelumnya.

Faktor utama yang berperan dalam penyebaran penyakit DBD adalah nyamuk *Aedes aegypti*. Nyamuk ini tersebar luas di daerah tropik dan subtropik. Nyamuk *Aedes aegypti* hidup di sekitar pemukiman manusia, di dalam dan di luar rumah terutama di daerah perkotaan dan berkembang biak dalam

Berikut merupakan pengaruh lingkungan fisik yang terkait dalam perkembangan nyamuk adalah :

1. Macam tempat penampungan air
2. Ketinggian tempat
3. Curah hujan
4. Hari hujan
5. Kecepatan angin
6. Suhu udara
7. Tata guna tanah
8. Pestisida
9. Kelembaban udara [7]

Maka dalam membuat model regresi untuk memperkirakan kasus demam berdarah, berikut merupakan variabel indikator yang akan digunakan untuk melakukan prediksi (*forecasting*):

1. Jumlah Kasus
2. Temperatur
3. Kelembapan
4. Curah Hujan

## 2.2 Data Mining

*Data mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar, sumber data dapat mencakup *database*, *data warehouse*, *website*, atau repositori informasi lainnya[8]. Hasil keluaran dari data mining dapat membantu dalam pengambilan keputusan di masa depan. *Data mining* merupakan bagian dari proses *Knowledge Discovery in Database* (KDD) untuk menemukan informasi yang berguna dalam data. Dalam *data mining* terdapat dua model yaitu *supervised learning*, dan *unsupervised learning*, *supervised learning* merupakan model dari *data mining* untuk sebuah data set yang memiliki label dan sudah memiliki target output sebelumnya, sedangkan *unsupervised learning* merupakan model dari data mining yang tidak memiliki label dan target output[9]. *Data mining* telah diimplementasikan dalam berbagai bidang ilmu, teori maupun permasalahan dunia nyata [10]–[14].

Tugas *data mining* secara general dapat diklasifikasikan menjadi dua kategori, yaitu :

1. *Descriptive mining* yang bertugas untuk mengkarakterisasi properti dari data dalam kumpulan data target, berikut merupakan daftar dari fungsi deskripsi pada data mining: *concept description*, *Mining of frequent pattern*, *Mining of associatios*, *Mining of correlations*, *Mining of clutsers*
2. *Predictive mining* yang bertugas melakukan interferensi data saat ini untuk membuat prediksi, *predictive mining* melibatkan penggunaan beberapa *variable* atau *fields* dalam *data set* untuk memprediksi nilai yang tidak diketahui atau nilai mendatang dari *variable* lain [9].

## 2.3 Regresi Linier

Regresi linear adalah metode khusus yang digunakan untuk menghitung data perkiraan waktu seri dalam periode tertentu. Regresi linear dibagi

menjadi dua, yaitu regresi linier dengan satu variabel dan regresi linier dengan multivariable [15].

### 1. Regresi linear dengan variabel berganda

Model regresi ini mencoba untuk memodelkan hubungan antara dua atau lebih variable penjelas dan variable respon dengan memasang persamaan linear untuk data yang diamati. Setiap nilai dari variable independen  $x$  dikaitkan dengan nilai variable dependen  $y$ . Berikut merupakan persamaan regresi linear berganda [4]

$$y_t = \beta_0 + \beta_1 x_1 + \dots + \beta_i x_i + \varepsilon_t \quad (1)$$

### 2. Koefisien Korelasi Berganda

Koefisien Korelasi Berganda merupakan sebuah ukuran relative antara dua variable, Tujuan dari koefisien korelasi adalah mengetahui secara bersama-sama kuatnya hubungan antara variable dependen dengan variabel independen, dalam regresi berganda besaran  $r$  bernilai positif (antara 0-1) [16]. Berikut merupakan rumus yang digunakan untuk menghitung Koefisien Korelasi Berganda [16] :

$$R = \sqrt{a_1 \frac{\sum X_1 Y + a_2 \sum X_2 Y + \dots + a_i \sum X_i Y}{\sum Y^2}} \quad (2)$$

### 3. Koefisien Determinasi

Koefisien Determinasi atau dikenal dengan  $R$  *Square* digunakan untuk mengukur kebaikan dari persamaan regres (*goodness of fit*), yang merupakan proporsi atau presentasi variasi total dalam variable terikat yang dijelaskan oleh variable bebas [16]. Nilai  $R$  *Square* bernilai antara 0-1, semakin mendekati 1 maka dapat dikatakan model memiliki kecocokan yang lebih baik, sebaliknya apabila mendekati 0 maka variable independen tidak dapat menjelaskan variable dependen secara keseluruhan. Berikut merupakan rumus yang digunakan untuk mendapatkan koefisien determinasi [16] :

$$R^2 = \sqrt{a_1 \frac{\sum X_1 Y + a_2 \sum X_2 Y + \dots + a_i \sum X_i Y^2}{\sum Y^2}} \quad (3)$$

### 4. Uji Multikolinearitas

Uji multikolinearitas bertujuan untuk mengetahui penyimpangan, yaitu adanya hubungan linear antara variable independen dalam model regresi, hubungan linear antara variable independen tersebut dapat menyebabkan gagal estimasi. Untuk mengetahui hasil uji multikolinearitas salah satunya dapat dengan melihat nilai *varians inflation factor* (VIF) [17].

### 5. Uji Normalitas Residual

Uji normalitas residual bertujuan untuk menguji sebuah model regresi memiliki distribusi normal. Salah satu uji yang dapat menentukan normalitas residual, adalah dengan melakukan uji Kolmogorov-Smirnov, uji ini dilakukan untuk membandingkan distribusi kumulatif dan distribusi data empiric dengan distribusi normal [17].

### 6. Uji Signifikansi Simultan

Uji F merupakan pengujian model regresi yang bertujuan untuk menguji hubungan linier variabel-variabel bebas  $X$  terkait dengan variabel terikat  $Y$ , atau salah satu variabel bebas  $X$  benar benar memiliki

hubungan dengan variabel terikat Y. pada uji ini pengambilan keputusan dilakukan dengan membandingkan nilai  $F_{hitung}$  dan  $F_{tabel}$  dengan tingkat signifikansi alpha ( $\alpha$ ) sebesar 5% (0.05), berikut merupakan rumus untuk menghitung  $F_{hitung}$ :

$$F_{hitung} = \frac{R^2(N-k-1)}{k(1-R) |2|} \quad (4)$$

#### 7. Uji Signifikansi Parsial

Uji Parsial merupakan sebuah uji yang digunakan untuk melihat kontribusi sebuah variabel bebas terhadap variabel terkait secara parsial.

Dalam uji-T dirumuskan hipotesa awal dan hipotesa alternatif. Hipotesa tersebut sebagai berikut:

- $H_0$ : Secara parsial tidak ada pengaruh atau tidak memiliki hubungan linier antara variabel bebas dengan variabel terikat
- $H_1$ : Secara parsial variabel bebas berpengaruh atau memiliki hubungan linear dengan variabel terikat [18].

#### 2.4 Perhitungan Akurasi

Untuk mengetahui akurasi dari model yang telah dibuat adalah dengan beberapa cara yang digunakan yaitu: *Mean Absolute Error*, *Mean Absolute Percentage Error*, *Root Mean Square Error*

##### 1. Mean Absolute Error

*Mean Absolute Percentage Error* (MAPE) menghasilkan rata-rata *error* prediksi dalam persen, Rumus MAPE adalah sebagai berikut [19]:

$$MAPE = \frac{\sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} * 100}{n} \quad (5)$$

##### 2. Mean Absolute Percentage Error

*Root Mean Absolute Percentage Error* (RMSE) digunakan untuk menghitung akurasi dari prediksi dengan mengkuadratkan rata-rata *error* dan mengakarkannya. Rumus dari RMSE adalah sebagai berikut [19]:

$$MAPE = \frac{\sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} * 100}{n} \quad (6)$$

##### 3. Uji Root Mean Square Error

*Root Mean Absolute Percentage Error* (RMSE) digunakan untuk menghitung akurasi dari prediksi dengan mengkuadratkan rata-rata *error* dan mengakarkannya. Rumus dari RMSE adalah sebagai berikut [19]:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}} \quad (7)$$

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Data Preprocessing

Data yang dikumpulkan dari dua instansi pemerintah provinsi DKI Jakarta akan diolah untuk mendapatkan model regresi, namun perlu dipastikan

bahwa data-data yang digunakan untuk membuat model regresi :

1. Tidak memiliki *missing value*
2. Tidak memuat *noisy data*: memuat *error* atau memuat *outliers* (data yang secara nyata berbeda dengan data-data yang lain)
3. Konsisten [20]

Adapun tujuan utama dalam data *preprocessing* ini adalah sebagai berikut :

- Pembersihan Data

Proses pembersihan data pada penelitian ini bertujuan untuk mengisi nilai yang hilang, mengidentifikasi dan menghapus *outliers*. Pada data cuaca yang didapatkan dari Dinas Lingkungan Hidup DKI Jakarta ditemukan beberapa nilai yang hilang dan ditemukan *noisy* dimana terdapat ketidaksesuaian atau *error* pada nilai pada atribut-atribut tersebut berbeda dengan nilai yang sewajarnya

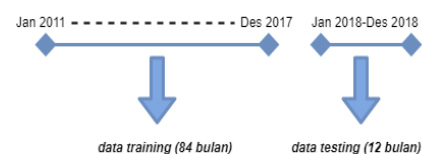
- Transformasi Data (Agregasi dan Normalisasi)

Normalisasi data dilakukan agar jangkauan suatu data dengan data yang tidak terlalu jauh, pada penelitian ini metode normalisasi yang digunakan adalah *minmax*, metode normalisasi *minmax* melakukan transformasi linier terhadap data asli sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses [21], berikut merupakan formula dari metode normalisasi *minmax* :

$$X'n = \left( X - \min x \times \frac{1-0}{\max x - \min x} + 0 \right) \quad (8)$$

- Data training dan testing

Metode yang digunakan dalam membagi data menjadi dua subset data (data training dan data testing) menggunakan *simple split*. Data training digunakan untuk pembelajaran pola-pola oleh algoritma sehingga menjadi sebuah model yang akan digunakan untuk melakukan prediksi. Dari total 96 jumlah *dataset*, sebanyak 84 *dataset* digunakan untuk membuat model prediksi, dan sebanyak 12 *dataset* digunakan pengujian performa dari model yang telah dibuat, 12 *dataset* tersebut dianggap sebagai representatif dataset keseluruhan, jumlah tersebut merupakan total dataset dari setiap kota 4 administratif selama 1 tahun, karena penelitian ini mengamati kejadian kasus DBD pada 4 kota administratif tersebut setiap bulan



Gambar 2. Ilustrasi pembagian dataset

#### 3.2 Pembuatan Model dan Analisis

Pada tahap ini akan dilakukan pembuatan model regresi dan analisis regresi. Model regresi yang dibuat

pada penelitian ini menggunakan variabel bebas (*independent variable*) temperatur, kelembapan, curah hujan, jumlah hari hujan, sedangkan untuk variabel terikat yaitu jumlah kasus. Berikut merupakan persamaan regresi yang didapat :

$$y_t = -0.1267 + 0.1490x_1 + 0.6563x_2 + 0.2847x_3 \quad (9)$$

Nilai koefisien untuk temperatur ( $x_1$ ) bernilai positif sebesar 0.1490, diasumsikan bahwa apabila terjadi peningkatan terhadap temperatur sebanyak 1%, dengan asumsi mengabaikan variabel bebas lain maka jumlah kasus demam berdarah akan bertambah sebanyak 0.1490.

Nilai koefisien untuk kelembapan ( $x_2$ ) bernilai positif sebesar 0.6563, diasumsikan bahwa apabila terjadi peningkatan terhadap kelembapan sebanyak 1%, dengan asumsi mengabaikan variabel bebas lain maka jumlah kasus demam berdarah akan meningkat sebanyak 0.6563.

Nilai koefisien untuk curah hujan ( $x_3$ ) bernilai positif sebesar 0.2847, diasumsikan bahwa apabila terjadi peningkatan terhadap temperature sebanyak 1%, dengan asumsi mengabaikan variabel bebas lain maka jumlah kasus demam berdarah akan meningkat sebanyak 0.2847.

Adapun hasil Analisis terhadap model tersebut sebagai berikut :

- **Koefisien Korelasi Ganda dan Koefisien Determinasi**

Uji koefisien korelasi berganda dan koefisien determinasi memiliki tujuan untuk mengetahui hubungan seluruh variabel independen secara bersamaan dengan variabel dependen. Nilai R ada di rentang antara 0-1, apabila nilai mendekati 1 maka hubungan semakin kuat, sebaliknya maka hubungan semakin lemah. Hasil uji koefisien korelasi ganda dan koefisien determinasi pada tabel 1

**Tabel 1.** Koefisien korelasi dan Determinasi

Regression Statistics	
R	0.6018
$R^2$	0.3622
Adjusted $R^2$	0.3382

Tabel 1 menunjukkan data statistik yang dihasilkan oleh model regresi yang telah dibuat, dapat diketahui nilai koefisien korelasi sebesar 0.6018, hal ini menunjukkan hubungan antara temperature, kelembapan, curah hujan terhadap kasus demam berdarah sedang. Diketahui dalam tabel 1 nilai koefisien determinasi sebesar 0.3622, hal tersebut menunjukkan presentase pengaruh dari temperature, kelembapan dan curah hujan terhadap kasus demam berdarah sebesar 36.22%, sedangkan 63.78% dipengaruhi oleh faktor lain diluar variabel independen tersebut

- **Uji Multikolinierias**

Uji Multikolinieritas pada penelitian ini ditujukan untuk melihat keterkaitan hubungan linier antara variabel bebas  $x_1$  (temperatur),  $x_2$  (kelembapan),  $x_3$  (curah hujan) dalam model regresi tersebut

Untuk mengetahui hasil uji multikolinieritas dari variabel-variabel diperlihatkan oleh tabel 2 berikut ini

**Tabel 2.** Uji Multikolinieritas

Model	Collinearity Statistic (VIF)
$x_1$	1.901469
$x_2$	1.985987
$x_3$	1.065097

Tabel 2 menunjukkan *output* uji asumsi multikolinieritas dengan menggunakan statistik VIF, diketahui bahwa dari tabel tersebut nilai VIF semua variable dependen  $< 10$ , maka dapat di simpulkan bahwa antar variabel yang diuji tidak memiliki multikolinieritas

- **Uji Normalitas Residual**

Uji normalitas dilakukan untuk menguji model regresi, apakah dalam model regresi tersebut, variabel residual memiliki distribusi normal, *Kolmogorov-Smirnov normality test* merupakan statistik uji yang sering digunakan dalam melakukan uji normalitas.

Hasil *p-value* yang didapat adalah 0.6971 dan lebih besar jika dibandingkan dengan  $\alpha = 0.05$ , dengan demikian dapat disimpulkan  $H_0$  diterima, dan asumsi kenormalan *error* tidak dilanggar

- **Uji Simultan Model Regresi**

Dalam uji F dirumuskan hipotesa awal dan hipotesa alternatif, dimana hipotesa tersebut sebagai berikut :

1.  $H_0$  : tidak ada pengaruh signifikan secara bersama-sama antara variabel independen temperatur, kelembapan dan curah hujan terhadap variabel dependen jumlah kasus demam berdarah
2.  $H_1$  : terdapat pengaruh signifikan secara bersama-sama antara variabel independen temperature, kelembapan dan curah hujan terhadap variabel dependen jumlah kasus demam berdarah.

Uji F pada penelitian ini digunakan untuk mengetahui apakah variabel independen ( $x_1, x_2, x_3$ ) yaitu temperatur, kelembapan, dan curah hujan terhadap variabel dependen ( $y$ ) yaitu jumlah kasus demam berdarah.

Berikut merupakan langkah yang akan dilakukan untuk melakukan uji F dengan menentukan signifikansi (*p value*) dari model regresi dibandingkan dengan derajat kepercayaan ( $\alpha = 0.05$ ) :

1. Apabila nilai signifikansi ( $p$  value)  $< 0,05$  maka dapat disimpulkan bahwa  $H_0$  ditolak dan  $H_1$  diterima
2. Apabila nilai signifikansi ( $p$  value)  $> 0,05$  maka dapat disimpulkan bahwa  $H_0$  diterima dan  $H_1$  ditolak

Tabel 3 menunjukkan ringkasan nilai uji-F dari model regresi linear

**Tabel 3.** Uji F

Model	Df	F	p-value
1.Regression	3	15.14	0.0000000683
	80		

Diketahui dari tabel 3 bahwa nilai F yang didapat adalah 15.14 dengan angka signifikansi sebesar 0.00005384, dengan begitu dapat disimpulkan bahwa :

1. Angka signifikansi  $0.0000000683 < 0.05$
2.  $H_0$  ditolak, variabel independen mempunyai pengaruh yang signifikan secara bersama-sama terhadap variabel dependen

#### • Uji Parsial Model Regresi

Uji-T pada penelitian ini digunakan untuk mengetahui apakah variabel independen ( $x_1, x_2, x_3$ ) yaitu temperatur, kelembapan, dan curah hujan berpengaruh secara parsial terhadap variabel dependen ( $y$ ) yaitu jumlah kasus demam berdarah.

Dalam uji-t dirumuskan hipotesa awal dan hipotesa alternatif, dimana hipotesa tersebut sebagai berikut :

1.  $H_0$  : tidak ada pengaruh signifikan secara parsial antara variabel independen temperatur, kelembapan dan curah hujan terhadap variabel dependen jumlah kasus demam berdarah
2.  $H_1$  : terdapat pengaruh signifikan secara parsial antara variabel independen temperatur, kelembapan dan curah hujan terhadap variabel dependen jumlah kasus demam berdarah

**Tabel 4.** Uji T

Model		
	t	p-value
x1	0.925	0.3576
x2	4.707	0.0000104
x3	2.211	0.0299

Tabel 4 menunjukkan hasil uji T pada penelitian ini, dari tabel tersebut diketahui bahwa nilai p-value yang didapat pada setiap variabel dependen, dengan begitu dapat disimpulkan bahwa :

1.  $x_1$  yaitu sebesar  $0.3576 > 0.05$ , sehingga  $H_0$  diterima. Dengan demikian temperatur secara

parsial tidak berpengaruh terhadap kejadian demam berdarah. Dari ketiga variabel yang diuji nilai  $p$ -value yang dihasilkan adalah yang tertinggi dibanding dengan variabel lain, dengan begitu kecenderungan penambahan nilai temperatur pengaruhnya menjadi kurang signifikan terhadap jumlah kasus demam berdarah *dengue*

2.  $x_2$  yaitu sebesar  $0.00001 > 0.05$ , sehingga  $H_0$  ditolak. Dengan demikian kelembapan secara parsial berpengaruh terhadap kejadian demam berdarah. Dari ketiga variabel yang diuji nilai  $p$ -value yang dihasilkan adalah yang terkecil dibanding dengan variabel lain, dengan begitu kecenderungan penambahan nilai kelembapan pengaruhnya menjadi paling signifikan terhadap jumlah kasus demam berdarah *dengue*.
3.  $x_3$  yaitu sebesar  $0.0299 > 0.05$ , sehingga  $H_0$  ditolak. Dengan demikian curah hujan secara parsial berpengaruh terhadap kejadian demam berdarah. Sedangkan nilai  $p$ -value curah hujan merupakan nilai terkecil setelah kelembapan, maka dengan begitu dapat disimpulkan bahwa kecenderungan penambahan nilai curah hujan memiliki pengaruh setelah kelembapan.

### 3.3 Pengujian Model

Bagian ini membahas pengujian model yang telah dibuat berdasarkan algoritma regresi linier berganda dengan menggunakan menggunakan 84 *data training* untuk Provinsi DKI Jakarta. Berikut merupakan hasil perbandingan antara data sebelumnya dengan data prediksi dan selisih *error* terjadi dalam bentuk normalisasi :

**Tabel 5.** Prediksi dalam interval normalisasi

Tahun-Bulan	Real	Prediksi	Selisih Error
2018-1	0.066491689	0.10912815	0.04264
2018-2	0.069116360	0.19680244	0.12769
2018-3	0.114610674	0.17690555	0.06229
2018-4	0.188101487	0.21872755	0.03063
2018-5	0.249343832	0.27564978	0.02631
2018-6	0.125984252	0.24013532	0.11415
2018-7	0.110236220	0.09474943	0.015487
2018-8	0.052493438	0.06915602	0.01666
2018-9	0.008748906	0.03954632	0.0308
2018-10	0.003499563	0.16407907	0.16058
2018-11	0.020997375	0.14880540	0.12781
2018-12	0.117235346	0.14745110	0.03022

Data denormalisasi hasil prediksi menggunakan model regresi pada objek linear model akan ditampung dalam objek denorm. validasi, denormalisasi dilakukan

menggunakan fungsi denormalize yang telah di definisikan sebelumnya, data yang digunakan ditampung dalam objek validasi.linear, dengan nilai maksimum diambil dari objek maxval, dan nilai minimum diambil dari objek minval. Berikut merupakan hasil denormalisasi data ditunjukkan oleh tabel 6:

**Tabel 6.** Prediksi setelah Denormalisasi

Tahun-Bulan	Real	Prediksi	Selisih error
2018-1	136	184.7335	48.7335
2018-2	139	284.9452	145.945
2018-3	191	262.2030	71.203
2018-4	275	310.0056	35.0056
2018-5	345	375.0677	30.0677
2018-6	204	334.4747	130.475
2018-7	186	168.2986	17.7014
2018-8	120	139.0453	19.0453
2018-9	70	105.2014	35.2014
2018-10	64	247.5424	183.542
2018-11	84	230.0846	146.085
2018-12	194	228.5366	34.5366

Untuk mengukur performa ketepatan dan keakuratan dari model prediksi kasus demam berdarah di Provinsi DKI Jakarta, diukur dengan menggunakan

- *Mean Absolute Error (MSE)*
- *Mean Absolute Percentage Error (MAPE)*
- *Root Mean Square Error (RMSE)*

**Tabel 7.** Evaluasi

MSE	MAPE	RMSE
74.79511613	0.681404268	94.00106913

Pada table 7 nilai MSE menunjukkan rata-rata selisih *error* kasus demam berdarah dari data asli dengan data prediksi menggunakan model regresi yang telah dibuat, sedangkan nilai MAPE menunjukkan presentasi nilai selisih *error* dari data validasi yang prediksi pada table 6

#### 4. KESIMPULAN

Dari semua model regresi yang telah dibuat dapat dijelaskan bahwa nilai koefisien untuk variabel bebas temperatur memiliki nilai terkecil dibanding koefisien dengan nilai terbesar dibandingkan temperatur dan curah hujan, sedangkan nilai koefisien curah hujan lebih besar dibandingkan dengan koefisien temperatur namun lebih kecil dari kelembapan.

Dari uji simultan kepada semua model regresi yang dibuat, dapat disimpulkan bahwa temperatur, kelembapan, dan curah hujan secara bersama-sama berpengaruh terhadap jumlah kasus demam berdarah di Provinsi DKI Jakarta, selanjutnya uji parsial terhadap semua model membuktikan bahwa

kelembapan dengan curah hujan memiliki pengaruh signifikan terhadap kenaikan kasus demam berdarah secara parsial, sedangkan untuk variabel bebas temperatur terbukti secara parsial tidak memiliki pengaruh yang signifikan terhadap kenaikan kasus demam berdarah dengue di Provinsi DKI Jakarta. Dari ketiga variabel bebas tersebut, variabel yang paling berpengaruh terhadap kenaikan kasus demam berdarah secara berurutan adalah kelembapan, curah hujan dan temperatur.

Dinas Kesehatan DKI Jakarta dapat menggunakan semua model yang telah dibuat pada penelitian ini untuk melakukan prediksi dalam ruang lingkup Provinsi DKI Jakarta. Dari hasil analisa regresi yang dilakukan Dinas Kesehatan DKI Jakarta perlu mewaspadai kenaikan kasus DBD apabila kondisi cuaca di Provinsi DKI Jakarta mengalami kelembapan disertai curah hujan yang tinggi.

#### 5. DAFTAR PUSTAKA

- [1] World Health Organization, "Dengue: guidelines for diagnosis, treatment, prevention, and control," *Spec. Program. Res. Train. Trop. Dis.*, p. x, 147, 2009.
- [2] InfoDatin Kementerian Kesehatan, "Situasi DBD di Indonesia," *InfoDATIN*. p. p 12, 2016.
- [3] Jakarta Provincial Health Agencies, "Health Profile of DKI Jakarta Province 2016," p. 131, 2016.
- [4] A. QomaruddinMunir and E. Winarko, "Classification Model Disease Risk Areas Endemicity Dengue Fever Outbreak based Prediction of Patients, Death, IR and CFR using Forecasting Techniques," *Int. J. Comput. Appl.*, vol. 114, no. 2, pp. 20–25, 2015.
- [5] Y. A. Lesnussa, L. J. Sinay, and M. R. Idah, "Aplikasi Jaringan Saraf Tiruan Backpropagation untuk Penyebaran Penyakit Demam Berdarah Dengue (DBD) di Kota Ambon," *J. Mat. Integr.*, vol. 13, no. 2, p. 63, 2017.
- [6] L. C. Lubos, "Dengue Incidence using Climate Variables as Predictors," *Asian J. Heal.*, vol. 5, no. 1, pp. 1–11, 2015.
- [7] M. M. Sintorini, "Pengaruh Iklim terhadap Kasus Demam Berdarah Dengue," *Kesmas Natl. Public Heal. J.*, vol. 2, no. 1, p. 11, 2016.
- [8] J. Han and M. Kamber, *Data Mining, Southeast Asia Edition: Concepts and Techniques*. Morgan kaufmann, 2006.
- [9] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [10] E. Sutoyo, I. T. R. Yanto, R. R. Saedudin, and T. Herawan, "A soft set-based co-occurrence for clustering web user transactions," *Telkonnika (Telecommunication Comput. Electron. Control.*, vol. 15, no. 3, 2017.
- [11] E. Sutoyo, R. R. Saedudin, I. T. R. Yanto, and

- A. Apriani, "Application of adaptive neuro-fuzzy inference system and chicken swarm optimization for classifying river water quality," in *Electrical, Electronics and Information Engineering (ICEEIE), 2017 5th International Conference on*, 2017, pp. 118–122.
- [12] H. Chiroma *et al.*, *An intelligent modeling of oil consumption*, vol. 320. 2015.
- [13] R. R. Saedudin, E. Sutoyo, S. Kasim, H. Mahdin, and I. T. R. Yanto, "Attribute selection on student performance dataset using maximum dependency attribute," in *Electrical, Electronics and Information Engineering (ICEEIE), 2017 5th International Conference on*, 2017, pp. 176–179.
- [14] E. S.-F. of I. Journal and undefined 2018, "Implementasi Limited Tolerance Relation Untuk Sistem Informasi yang Tidak Lengkap Pada Data Mahasiswa," *ejournal.unida.gontor.ac.id*.
- [15] N. Singh, R. A. M. S. Raw, and R. K. Chauhan, "Data Mining With Regression Technique," vol. 3, no. 1, pp. 199–202, 2012.
- [16] K. D. Ariani, J. Teknik, I. Fik, and J. N. N. Semarang-, "Data Pabrik Gula Rendeng Kudus," no. 5.
- [17] S. Sulistyono and W. Sulistiyowati, "Peramalan Produksi dengan Metode Regresi Linier Berganda," *PROZIMA (Productivity, Optim. Manuf. Syst. Eng.*, vol. 1, no. 2, p. 82, 2018.
- [18] A. Amrin, "Data Mining Dengan Regresi Linier Berganda Untuk Peramalan Tingkat Inflasi," *J. Techno Nusa Mandiri*, vol. XIII, no. 1, pp. 74–79, 2016.
- [19] C. Eswaran and R. Logeswaran, "A Comparison of ARIMA , Neural Network and Linear Regression Models for the Prediction of Infant Mortality Rate," no. 2, 2010.
- [20] J. Han, "Data mining: concepts and techniques," *J. Chem. Inf. Model.*, pp. 101–103, 2012.
- [21] T. T. Hanifa, S. Al-faraby, F. Informatika, and U. Telkom, "Analisis Churn Prediction pada Data Pelanggan PT . Telekomunikasi dengan Logistic Regression dan Underbagging," vol. 4, no. 2, pp. 3210–3225, 2017.