

Penerapan Model Regression Untuk Prediksi Cuaca Wilayah seberang Ulu 1 Palembang

Moh Fajri Al Amin¹, Yesi Novaria Kunang², Susan Dian Purnamasari³

^{1,2,3}Universitas Bina Darma

^{1,2,3} Jln. Jenderal Ahmad Yani No.02. Palembang

¹mohfajri97@gmail.com, ²yesinovariakunang@binadarma.ac.id, ³susandian@binadarma.ac.id

ABSTRACT

Weather is a short air condition. Therefore, the role of weather forecasting is very much needed considering the number of activities and activities of humans that depend on the weather. Because of this, the processing of weather data must be done quickly and accurately. Therefore, the use of data mining is needed to solve this problem, one of them is a regression model. Regression models are a type of method that predicts predictive data through straight lines as a correlation between 2 variables or more. Linear regression algorithm, used as a technique to learn how to relate variables to the data forecasting process with accurate results. The dataset is taken from AWS LAPAN Palembang. The results of this study are regression models that can later be used as a reference to predict rainy weather or not rain for one day.

Keywords: data mining, linear regression, regression models, linear algorithms

ABSTRAK

Cuaca adalah kondisi udara yang berlangsung secara singkat. Oleh karena itu, peran prakiraan cuaca sangatlah dibutuhkan mengingat banyaknya aktifitas maupun kegiatan manusia yang bergantung dengan cuaca. Karena nya, pengolahan data cuaca haruslah dilakukan secara cepat serta akurat. Oleh karena itu dibutuhkan pemanfaatan data mining untuk menyelesaikan masalah ini, salah satunya dengan model regresi. Model regresi adalah jenis metode yang mengukur data prediksi melalui garis lurus sebagai gambaran hubungan korelasi diantara 2 variabel atau lebih. Algoritma regresi linier, digunakan sebagai teknik mempelajari bagaimana hubungan variabel-variabel pada proses peramalan data dengan hasil yang akurat. Dataset diambil dari AWS LAPAN Palembang. Hasil dari penelitian ini adalah model-model regresi yang nantinya dapat dijadikan acuan untuk memprediksi cuaca hujan atau tidak hujan untuk suatu hari kedepan.

Kata Kunci: data mining, *linear regression*, model regresi, Algoritma linier

I. PENDAHULUAN

Indonesia merupakan negara *maritime-continent* yang kaya uap air karena berada antara dua samudera, yaitu samudera Pasifik dan samudera Hindia. Letak geografis Indonesia yang berada pada lintang rendah dan ekuator membuat Indonesia mempunyai *heat energy* dan insolasi yang besar untuk mengangkat uap air tersebut ke atmosfer. Keadaan ini memungkinkan Indonesia mempunyai karakteristik cuaca yang beragam di berbagai daerah. Informasi tentang prakiraan cuaca yang cepat dan tepat menjadi suatu hal yang penting karena cuaca menjadi bagian yang tidak dapat dipisahkan dari aktifitas manusia dan mempengaruhi berbagai bidang kehidupan, seperti penentuan masa tanam (pertanian) dan kelayakan keberangkatan pesawat udara (transportasi). Menyikapi hal tersebut maka diperlukan kemampuan dan metode yang efektif dalam prakiraan cuaca, khususnya prakiraan cuaca jangka pendek.

Pengetahuan tentang meteorologi telah menarik perhatian para ilmuwan untuk mempelajari proses dinamika meteorologi yang terjadi di atmosfer. Berbagai penelitian untuk menemukan ide-ide baru berkaitan dengan dinamika meteorologi telah mulai dilakukan beberapa dekade sebelumnya. Salah satu tema penelitian yang sampai sekarang menjadi perhatian adalah mengembangkan metode yang efektif dalam prakiraan cuaca. Berbagai persamaan matematis yang berkaitan dengan permasalahan dalam prakiraan cuaca dipelajari dalam dinamika meteorologi, fisika atmosfer dan komputasi meteorologi.

Prakiraan cuaca memerlukan waktu yang lama dan basis data yang kompleks sehingga perlu dicari suatu metode untuk memprediksikan cuaca dengan cepat dan tepat. Metode prakiraan cuaca melalui pendekatan statistik dan berdasarkan data yang telah direkam dapat menjadi alternatif untuk mengatasi permasalahan di atas.

Cuaca adalah kondisi udara yang berlangsung secara singkat. Oleh karena itu, peran prakiraan cuaca sangatlah dibutuhkan mengingat banyaknya aktifitas maupun kegiatan manusia yang bergantung dengan cuaca. Karena nya, pengolahan data cuaca haruslah dilakukan secara cepat serta akurat. Kondisi suatu cuaca dipengaruhi oleh berbagai faktor, di dalam data cuaca terdapat data suhu, rata kelembaman, rata kecepatan angin, arah angin, dan curah hujan yang masing-masing di anggap sebuah item. Karena dari data inilah seorang prakiraan cuaca dapat memprediksi kondisi suatu cuaca untuk keesokan harinya. Oleh karena itu penelitian ini menggunakan model *regression* untuk mengolah hasil rekaman cuaca oleh Pusat Sainst dan Teknologi Atmosfer Lembaga Penerbangan dan Antariksa Nasional (PSTA LAPAN) yang dilakukan di Universitas Bina Darma ,dimana dengan penelitian ini diharapkan dapat memberikan hasil berupa prakiraan cuaca yang dapat berguna bagi masyarakat atau bahkan mungkin dapat diimplementasikan ke bentuk web sehingga lebih mudah diakses masyarakat

Penelitian ini menggunakan model *regression* dengan algoritma linear. Algoritma ini mengukur data prediksi melalui garis lurus sebagai gambaran hubungan korelasi diantara 2 variabel atau lebih. Prediksi regresi linier, digunakan sebagai teknik mempelajari bagaimana hubungan variabel-variabel pada proses peramalan data.

Berdasarkan permasalahan tersebut, maka penulis tertarik untuk melakukan penelitian dan menjadikan sebagai judul skripsi dengan judul “**Penerapan Model *Regression* untuk Prediksi Cuaca Wilayah Seberang Ulu 1 Palembang**”.

II. LANDASAN TEORI

2.1 Cuaca

Menurut Gibbs (1987:3), Cuaca adalah keadaan atmosfer yang dinyatakan dengan nilai berbagai parameter, antara lain suhu, tekanan, angin, kelembaban, dan berbagai fenomena hujan, di suatu tempat atau wilayah selama kurun waktu yang pendek.

2.1.2 Curah Hujan

Curah hujan merupakan jumlah air yang jatuh dipermukaan tanah datar selama periode tertentu yang diukur dengan satuan tinggi (mm) di atas permukaan horizontal bila tidak terjadi evaporasi, runoff, dan infiltrasi.

Menurut Asid Ur Rehman (2018) ,Curah hujan terbagi atas beberapa kategori antara lain ,*light rain*, *moderate rain*, *heavy rain*, *very heavy rain* , dan *Extreme rain* . Setiap kategori memiliki nilai ukur sendiri yang dapat dilihat pada gambar 2.1.

SN	Rain Category	Threshold (mm/h)
1	Light rain	<1
2	Moderate rain	1-4
3	Heavy rain	4-16
4	Very heavy rain	16-50
5	Extreme rain	>50

Gambar 2.1 kategori curah hujan

Dapat dilihat dari gambar diatas dapat dilihat bahwa kategori *light rain* memiliki curah hujan sebesar 1mm, untuk *moderate rain* sebesar 1 hingga 4 mm, kemudian untuk *heavy rain* kisaran 4-16 mm, untuk *Very heavy rain* kisaran 16-50mm, dan *Extreme rain* lebih dari 50mm.

2.1.3 Data

Menurut Abdul Kadir (2009:3), Data adalah fakta fakta mentah yang mewakili kejadian kejadian yang berlangsung dalam organisasi atau lingkungan fisik sebelum ditata dan diatur ke dalam bentuk yang dapat dipahami dan digunakan.

2.1.4 Database

Database adalah kumpulan informasi yang disimpan di dalam komputer secara sistematis sehingga dapat diperiksa menggunakan suatu program komputer untuk memperoleh informasi dari basis data tersebut. Perangkat lunak yang digunakan untuk mengelola dan memanggil kueri (*query*) basis data disebut sistem manajemen basis data (*database management system*, DB MS).

Menurut Anhar (2010:45), Database atau basisdata adalah sekumpulan tabel-tabel yang berisi data dan merupakan kumpulan dari field atau kolom. Struktur file yang menyusun sebuah database adalah Data Record dan Field

Sedangkan menurut Raharjo (2011:3), Database atau basisdata adalah kumpulan data yang terintegrasi dan diatur sedemikian rupa sehingga data tersebut dapat dimanipulasi, diambil, dan dicari secara cepat

Kemudian menurut Bambang Hariyanto (2004), Basis data merupakan sebuah kumpulan data-data yang secara logik yang berkaitan dengan merepresentasikan fenomena atau fakta secara terstruktur dalam domain tertentu untuk mendukung aplikasi pada suatu sistem tertentu. Pengertian lain dari Database yaitu kumpulan data yang saling berhubungan yang merefleksikan fakta-fakta yang terdapat di suatu organisasi.

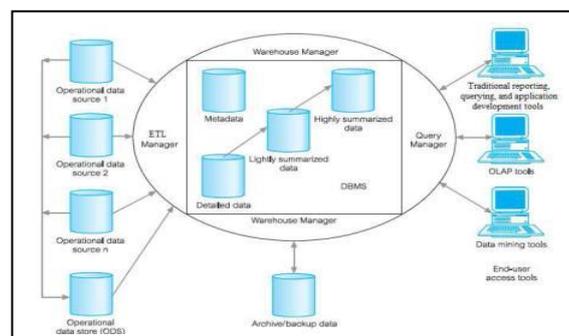
2.1.5 Data Warehouse

Data Warehouse dapat memiliki bermacam-macam pengertian namun mempunyai inti yang sama, seperti pendapat para beberapa ahli berikut:

Menurut Thomas Connolly dan Carolyn Begg (2010:1197), *data warehouse* adalah sebuah koleksi data yang terdiri dari *subject-oriented, integrated, time variant*, dan *non volatile* yang mendukung manajemen dalam proses pengambilan keputusan.

Menurut Vidette Poe (1998), *data warehouse* adalah database yang read-only analisis dan digunakan sebagai dasar sistem pendukung keputusan.

Dari pendapat diatas dapat disimpulkan *data warehouse* adalah koleksi data yang terdiri dari *subject-orientef, integrated, time variant*, dan *non volatile* untuk mendukung sebuah sistem penarik kesimpulan.



Gambar 2.2 Arsitektur Data Warehouse

2.1.6 Data Mining

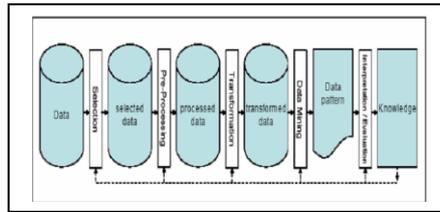
Menurut Sid Adelman (2000, p145) *data mining* adalah proses pencarian pola data yang tidak diketahui atau tidak diperkirakan sebelumnya.

Sedangkan menurut Larose (2005), *Data mining* merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari database yang besar

Kemudian menurut Maimon dan Last (2010), *Data mining* adalah bagian dari proses KDD (*Knowledge Discovery in Databases*) yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan, transformasi, *data mining*, dan evaluasi hasil.

2.1.7 KDD(Knowledge Discovery in Database)

Adapun metode yang digunakan di dalam mengolah data di dalam penerapan data mining menggunakan metode knowledge discovery in database yang terdiri dari proses data selection, pre processing, transformation, data mining, dan evaluation yang akan diilustrasikan di Gambar 2.3.



Gambar 2.3 Tahapan Data Mining

Tahap-tahap data mining ada 5 yaitu:

1. **Pemilihan data**
Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.
2. **Pra Pengolahan**
Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.
3. **Transformasi**
Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.
4. **Data Mining**
Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
5. **Evaluasi hasil**
Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.1.8 Model Regresi

Model regresi adalah suatu metode analisis statistik yang digunakan untuk melihat pengaruh antara dua atau lebih variabel. Hubungan variabel tersebut bersifat fungsional yang diwujudkan dalam suatu model matematis. Pada analisis regresi, variabel dibedakan menjadi dua bagian, yaitu variabel respons (*response variable*) atau biasa juga disebut variabel bergantung (*dependent variable*) dan variabel *explanatory* atau biasa disebut penduga (*predictor variable*) atau disebut juga variabel bebas (*independent variabel*).

Menurut **David Olson & Yong Shi (2008)**, dalam bukunya yang berjudul Pengantar Ilmu Pengalian Data Bisnis mengemukakan bahwa regresi ialah salah satu perangkat dasar untuk analisis yang bisa digunakan untuk membuat model prediktif untuk berbagai jenis data.

Sedangkan menurut **Hasan (2008)**, regresi merupakan suatu alat ukur yang juga digunakan untuk mengukur ada tidaknya korelasi antarvariabel. Istilah regresi yang berarti ramalan atau taksiran. Analisis regresi lebih akurat dalam melakukan analisis korelasi, karena pada analisis itu kesulitan dalam menunjukkan slop (tingkat perubahan suatu variabel terhadap variabel lainnya dapat ditentukan). Analisis regresi dapat meramal atau memperkirakan nilai variabel bebas lebih akurat. Regresi linier adalah regresi yang variabel bebasnya (variabel x) berpangkat paling tinggi satu.

2.1.9 Algoritma Linear

Regresi linear adalah metode statistik yang berfungsi untuk menguji sejauh mana hubungan sebab akibat antara variabel faktor penyebab (X) terhadap variabel akibatnya. Faktor penyebab pada umumnya dilambangkan dengan X atau disebut juga dengan *predictor* sedangkan Variabel Akibat dilambangkan dengan Y atau disebut juga dengan *response*.

Regresi Linear merupakan salah satu Metode Statistik yang dipergunakan dalam produksi untuk melakukan peramalan ataupun prediksi tentang karakteristik kualitas maupun Kuantitas.

Langkah-langkah di dalam algoritma linear adalah:

1. Tentukan Tujuan dari melakukan Analisis Regresi Linear Sederhana
2. Identifikasikan Variabel Faktor Penyebab (Predictor) dan Variabel Akibat (Response)
3. Lakukan Pengumpulan Data
4. Hitung X^2 , Y^2 , XY dan total dari masing-masingnya
5. Hitung a dan b berdasarkan rumus berikut,:

$$a = \frac{(\sum y) (\sum x^2) - (\sum x) (\sum xy)}{(\sum x) (\sum xy)}$$

Dan untuk rumus b sebagai berikut:

$$b = \frac{n(\sum xy) - (\sum x) (\sum y)}{n(\sum x^2) - (\sum x)^2}$$

6. Buat Model Persamaan Regresi Linear Sederhana.
7. Lakukan Prediksi atau Peramalan terhadap Variabel Faktor Penyebab atau Variabel Akibat.

$$Y = a + bX$$

Dimana:

Y = Variabel Response atau Variabel Akibat (Dependent)

X = Variabel Predictor atau Variabel Faktor Penyebab (Independent)

a = Konstanta

b = Koefisien regresi (kemiringan); besaran Response yang ditimbulkan oleh Predictor.

n = Banyak data

2.1.10 Regresi Logistik

Regresi logistik adalah salah satu jenis regresi yang menghubungkan antara satu atau beberapa variabel independen (variabel bebas) dengan variabel dependen yang beruoa kategori ,biasanya) dan 1. Jenis variabel inilah yang membedakannya dengan regresi linear.

Nilai kategori bisa saja tertulis 0 dan 1 tetapi bisa juga digunakan untuk kondisi “ya” ataupun “tidak”. Dapat dilihat rumus persamaan regresi logistik adalah sebagai berikut:

$$\text{logit}(p) = \log \left(\frac{p}{1-p} \right) = \ln \left(\frac{p}{1-p} \right)$$

Dan untuk model regresi logistik adalah sebagai berikut:

$$\text{Log} \left(\frac{P}{1-p} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Dimana:

P = memiliki nilai antara 0-1

X1,X2,...n= Variabel Independen

B = Koefisien regresi

2.1.11 R

R adalah sebuah bahasa komputer dan merupakan lingkungan pemrograman yang interaktif untuk analisa data dan grafik. Ketika kita menginputkan suatu fungsi atau perintah, maka R akan langsung memprosesnya serta mengevaluasikannya sehingga kita dapat melihat langsung hasil yang kita kerjakan tadi. R juga bahasa pemrograman berorientasi objek,di mana R adalah bahasa yang case sensitive yakni penggunaan huruf besar dan kecil memiliki makna yang berbeda.seperti objek ABC tidak sama dengan abc. Tipe objek di dalam R bisa berupa vektor,matriks,dan array.dalam penggunaan yang lebih kompleks bisa berupa list ataupun fungsi. Cara kerja R ,antara lain:

1. R dapat digunakan secara interaktif
2. Hasil analisa R yang langsung dapat dilihat
3. Dapat melakukan perhitungan kompleks,dengan menggunakan perintah R di dalam text editor, kemudian dijalankan
4. Bahasa R berorientasi objek, artinya semua data,fungsi,hasil, dan sebagainya disimpan di dalam memori aktif komputer dalam bentuk objek atau memiliki nama
5. Penamaan objek bersifat *case sensitive* dengan kata lain huruf besar dan kecil berpengaruh
6. R memiliki banyak fungsi ,yang biasa di sebut package yang dapat di install sesuai kebutuhan.

Tujuan utama R yaitu memungkinkan dan mendorong terciptanya analisa data yang baik. Oleh karena itu R,

1. Memberikan banyak fungsi dan fitur yang dapat digunakan untuk mengolah,organisasi,menyimpan,ataupun memanggil data
2. Fungsi fungsi yang beragam dengan *metoda numerik* yang sesuai dengan pengguna
3. Memberikan cara yang *interaktif*, sederhana, *informatif*, maupun *fleksibel* di dalam memandang data. R juga dapat digunakan untuk berbagai hal seperti analisa statistika, keuangan, cuaca, akademis, matematika, dan analisis data lain nya.

Didalam penggunaannya R menggunakan package sebagai alat ataupun tools di dalam pengolahan datanya. Package R dapat diinstall langsung melalui software R . Adapun package yang digunakan di dalam penelitian ini antara lain: *knitr*, *caret*, *gmodels*, *lattice*, *ggplot*, *gridExtra*, *Kmisc*, *ROCR*, dan *corrplot*.

3. PERANCANGAN SISTEM

3.1 Seleksi Data (Data Selection)

Seleksi data dilakukan dengan menganalisa sebuah data yang relevan karena tidak semua data akan digunakan di dalam proses *data mining*. Sumber data yang digunakan di dalam penelitian ini berasal dari data AWS Pusat Saint dan Teknologi Atmosfer Lembaga Penerbangan dan Antariksa Nasional (PSTA LAPAN) berlokasi di binadarma pada bulan juni pada tahun 2010 hingga agustus tahun 2016. Berikut proses seleksi data dapat dilihat pada gambar 3.1.

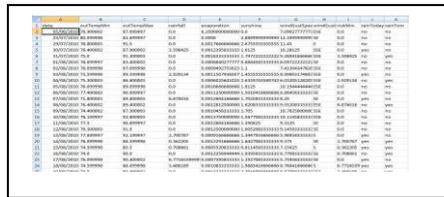
```
data.frame(idTime, waktu, tempat, baromet, tekanan, altitud, ktrop, outTemp, suhuudara, kelembapan, hal, suhu, heading, diras, huan)
#>   idTime waktu tempat baromet tekanan altitud ktrop outTemp suhuudara kelembapan hal suhu heading diras huan
#> 1  175642000 1  S 20.76000  NaN NaN 276.80000 276.80000 276.80000 276.80000  NaN NaN NaN
#> 2  175642000 1  S 20.84000  NaN NaN 77.00000  80.00000  80.00000  70.00000  NaN NaN NaN
#> 3  175642000 1  S 20.84000  NaN NaN 77.00000  80.00000  87.00000  87.00000  70.00000  NaN NaN NaN
#> 4  175642000 1  S 20.84000  NaN NaN 77.00000  80.00000  87.00000  87.00000  70.00000  NaN NaN NaN
#> 5  175642000 1  S 20.84000  NaN NaN 77.00000  80.00000  80.00000  80.00000  70.00000  NaN NaN NaN
#> 6  175642000 1  S 20.84000  NaN NaN 77.00000  80.00000  80.00000  80.00000  70.00000  NaN NaN NaN
#> 7  175642000 1  S 20.84000  NaN NaN 77.00000  80.00000  80.00000  80.00000  70.00000  NaN NaN NaN
#> 8  175642000 1  S 20.84000  NaN NaN 77.00000  80.00000  80.00000  80.00000  70.00000  NaN NaN NaN
#> 9  175642000 1  S 20.84000  NaN NaN 77.00000  80.00000  80.00000  80.00000  70.00000  NaN NaN NaN
#> 10 175642000 1  S 20.84000  NaN NaN 77.00000  80.00000  80.00000  80.00000  70.00000  NaN NaN NaN
```

Gambar 3.1 Proses Seleksi Data

3.2 Prapengolahan Data

Setelah tahap seleksi data, maka tahap selanjutnya adalah tahap prapengolahan data.pada tahap ini, jumlah record data mentah yang akan digunakan adalah sebanyak 300.000 data. Data tersebut terbagi atas dua data yang diambil dengan cara yang berbeda,yaitu data yang direkam oleh alat dan dengan *software*. Data yang akan digunakan di dalam penelitian ini adalah data dari software dimana selanjutnya data tersebut yang semula terekam per 5 menit akan di cari rata-rata nilai nya per hari. Kemudian dari data tersebut di bagi kembali atas dua bagian, yaitu data pada jam 9 pagi dan jam 3 sore. Data akan di *cleaning* agar tidak ada data NA (data kosong), data yang terduplikasi kemudian memeriksa serta memperbaiki apabila ada kesalahan pada data sehingga data yang sudah di *cleaning* siap untuk digunakan di dalam proses data mining yang dapat dilihat pada gambar 3.2.

Kemudian dataset yang sudah melalui tahapan transformasi ini di simpan dengan format “csv” (comma delimited) yang akan di import ke dalam program R untuk selanjutnya dilajuti ke tahap data mining.



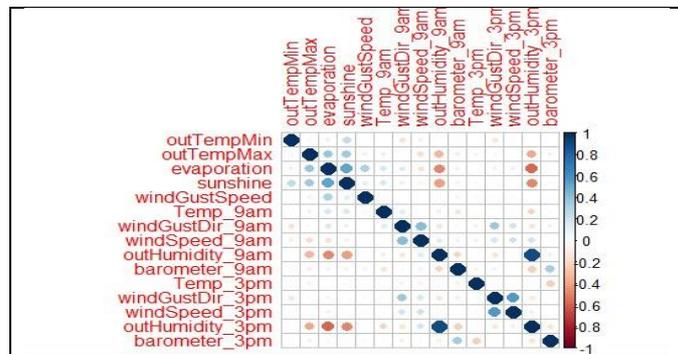
Gambar 3.7 Dataset dengan format csv

4. HASIL

4.1 Data Mining

4.1.1 Analisis Data

Data yang diinputkan selanjutnya akan di analisa menggunakan R untuk dianalisa. Data yang digunakan di dalam penelitian sebanyak 1306 data dengan variabel sebanyak 24 variabel. Data yang diinputkan selanjutnya akan di analisa tingkat korelasi dari antar variabel atau yang disebut qualitative analysis. Dimana apabila indikator menunjukkan warna biru maka akan berkorelasi positif tetapi sebaliknya apabila merah maka negatif.



Gambar 4.1 Hasil *Qualitative analysis*

4.1.2 Pengembangan Model Multiple Logistic

Setelah melalui tahapan analysis, data tersebut akan dibagi atas beberapa model sebagai indikator didalam menentukan variabel yang paling berpengaruh dan memiliki nilai prediksi terbesar yang dapat digunakan didalam proses *reporting* pada akhirnya.

Model yang akan digunakan pada tahap ini adalah model 9am, model 3pm, model evening c2, dan model evening c3.

1. Model 9am

Pada faktor jam 9 ada lima variabel yang digunakan antara lain: *Temp_9am*, *outHumidity_9am*, *barometer_9am*, *windGustDir_9am*, dan *windSpeed_9am*. dan hasil akurasi dari model tersebut menunjukkan hasil yang cukup yakni sebesar 70 persen.

```

predictors_9am_c1 <- c("Temp_9am", "outHumidity_9am",
"barometer_9am", "windGustDir_9am", "windSpeed_9am")
> formula_9am_c1 <- as.formula(paste("raintom",
paste(predictors_9am_c1, collapse="+"), sep="-"))
> mod9am_c1_fit <- train(formula_9am_c1, data=training,
method="glm",
family="binomial", trcontrol =
trcontrol, metric = 'Accuracy')
> mod9am_c1_fit$results$Accuracy
[1] 0.7016493
    
```

Gambar 4.2 Model 9am

2. Model 3pm

Pada analisa di jam 3 sore variabel yang digunakan di dalam model ini adalah *temp_3pm*, *outhumidity_3pm*, *barometer_3pm*, *windgustdir_3pm*, *windspeed_3pm*, dan *winddir_3pm*

```
predictors_3pm_c1 <- c("temp_3pm", "outHumidity_3pm",  
"barometer_3pm", "windgustDir_3pm", "windspeed_3pm",  
"windDir_3pm")  
> formula_3pm_c1 <- as.formula(paste("rainTom",  
paste(predictors_3pm_c1, collapse="+"), sep="~"))  
> mod3pm_c1_fit <- train(formula_3pm_c1,  
data = training, method = "glm", family = "binomial",  
trControl = trControl,  
metric = 'Accuracy')  
> mod3pm_c1_fit$results$Accuracy  
[1] 0.6863547
```

Gambar 4.3 Model 3pm

Dapat dilihat bahwa model 3pm memperoleh nilai akurasi sebesar 0,68 atau 68 persen. kemudian model selanjutnya adalah evening c2

3. Model Ev c2

```
formula_evening_c2 <- as.formula(paste("rainTom",  
paste(predictors_evening_c2, collapse="+"), sep="~"))  
> mod_ev_c2_fit <- train(formula_evening_c2, data=training,  
method="glm", family="binomial", trControl = trControl, metric  
= 'Accuracy')  
> mod_ev_c2_fit$results$Accuracy  
[1] 0.7009999
```

Gambar 4.4 Model ev 2

Terlihat pada gambar di atas bahwa model ev 2 mendapatkan nilai akurasi sebesar 70%. Sedikit lebih tinggi daripada model 3pm dan untuk model selanjutnya adalah model ev 3.

4. Model ev 3

kemudian sebagai model terakhir dari evening, model ini menggunakan variabel *barometer_3pm* dan *sunshine* sebagai pembanding untuk melihat apakah model ini akan mendapatkan nilai yang lebih baik ataupun tidak.

```
predictors_evening_c3 <- c("barometer_3pm", "sunshine")  
> formula_evening_c3 <- as.formula(paste("rainTom",  
paste(predictors_evening_c3, collapse="+"), sep="~"))  
> mod_ev_c3_fit <- train(formula_evening_c3, data=training,  
method="glm", family="binomial", trControl = trControl, metric  
= 'Accuracy')  
> mod_ev_c3_fit$results$Accuracy  
[1] 0.6996913
```

Gambar 4.5 Model ev 3

Dapat dilihat pada model evening c3 bahwa model ini mendapatkan nilai yang cukup dibanding dengan beberapa model yang di atas yaitu dengan nilai 0,69 atau sekitar 69%. Kemudian dari keseluruhan model ini dapat di simpulkan bahwa setiap variabel pembentuk model sangat berpengaruh terhadap hasil akurasi dari setiap model. Serta di dapat juga hasil resume variabel pembentuk ke empat model utama adalah:

```
1. mod9am_c1_fit: rainTom ~ temp_9am + outHumidity_9am +  
barometer_9am + windgustDir_9am + windSpeed_9am  
2. mod3pm_c1_fit: rainTom ~ windgustDir_3pm + outHumidity_3pm +  
barometer_3pm + Temp_3pm + windSpeed_3pm + windDir_3pm  
3. mod_ev_c2_fit: rainTom ~ sunshine + outHumidity_3pm +  
barometer_3pm + Temp3pm + WindSpeed_3pm  
4. mod_ev_c3_fit: rainTom ~ Barometer_3pm + Sunshine
```

Gambar 4.6 Multiple regression model

Dapat dilihat pada gambar di atas bahwa setiap model mempunya variabel prediktor tersendiri dan memiliki nilai prediksi yang berbeda-beda.

4.1.3 Pengembangan Nilai Optimal (cut off)

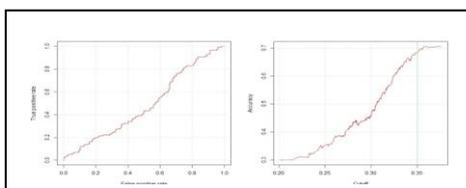
Pada tahap ini ke empat model tadi akan di jabarkan ke dalam bentuk ROCR analysis. Tujuannya untuk melihat titik potong di antara model model tersebut.

1. Model 9 pm

```
mod9am_pred_prob <- predict(mod9am_c1_fit, testing,
                             type="prob")
> mod9am_pred_resp <- prediction(mod9am_pred_prob$yes,
                                testing$rainTom)
> glm.perf.plot(mod9am_pred_resp, 0.35)
[1] 0.4862125
```

Gambar 4.7 Cut off model 9 am

Dapat dilihat model 9 am menghasilkan nilai sekitar 0.48 dan untuk hasil visualisasi dari model 9 am dapat dilihat pada gambar 4.8



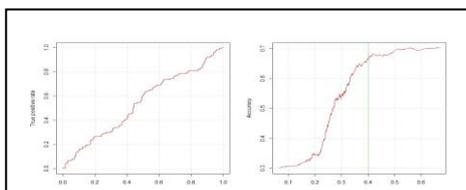
Gambar 4.8 Visualisasi ROCR model 9 am

2. Model 3 pm

```
mod3pm_pred_prob <- predict(mod3pm_c1_fit, testing,
                             type="prob")
> mod3pm_pred_resp <- prediction(mod3pm_pred_prob$yes,
                                testing$rainTom)
> glm.perf.plot(mod3pm_pred_resp, 0.4)
[1] 0.5406451
```

Gambar 4.9 Cut off model 3 pm

Dapat dilihat pada gambar bahwa model 3pm memperoleh nilai sebesar 0.54 dan untuk visualisasi ROCR dapat dilihat pada gambar 4.10



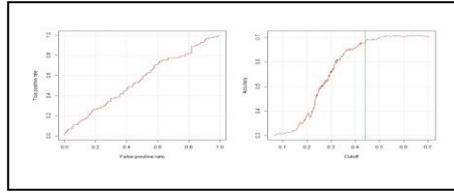
Gambar 4.10 Visualisasi ROCR model 3pm

3. Model ev 2

```
mod_ev_c2_prob <- predict(mod_ev_c2_fit,
                           testing, type="prob")
> mod_ev_c2_pred_resp <- prediction
(mod_ev_c2_prob$yes,
 testing$rainTom)
> glm.perf.plot(mod_ev_c2_pred_resp, 0.44)
[1] 0.5568345
```

Gambar 4.11 Cut off model ev 2

Dapat dilihat model ev 2 mendapatkan nilai sebesar 0.55 dan untuk hasil visual ROCR dapat dilihat pada gambar 4.12.



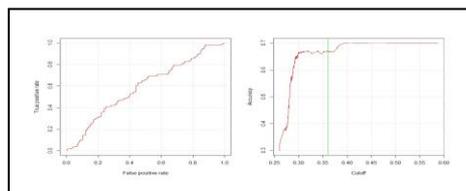
Gambar 4.12 Visualisasi ROCR model ev 2

4. Model ev 3

```
mod_ev_c3_prob <- predict(mod_ev_c3_fit, testing, type="prob")
mod_ev_c3_pred_resp <- prediction(mod_ev_c3_prob$yes,
testing$rainfall)
> glm.perf.plot(mod_ev_c3_pred_resp, 0.36)
[1] 0.5884023
```

Gambar 4.13 Proses penentuan nilai *cut off*

Dapat dilihat pada gambar diatas menunjukkan bahwa model di atas mempunyai nilai sebesar 0.58 ,maka untuk tampilan visual nya dapat dilihat pada gambar 4.14



Gambar 4.14 Visualisasi ROCR model ev 3

Berdasarkan hasil dari keempat model di atas, dapat dipastikan model ev 3 memiliki nilai tertinggi yaitu sebesar 0.58 .Model ini nantinya akan digunakan sebagai model prediktor di dalam proses reporting.

4.1.4 Reporting

Setelah melalui serangkaian proses analisis,multiple logistic regresi, serta nilai optimal, langkah terakhir adalah melakukan laporan reporting hasil prakiraan cuaca. Dalam pembentukan proses reporting dilakukan serangkaian pengujian pada variabel *rainfall,outtempmax,humidity,windgustspeed,mintemp,dan maxtemp* dimana variabel tersebut. Untuk proses reporting dapat dilihat pada gambar 4.15

```
weather_report <- function(today_record, rain_tomorrow_model,
cutoff) {
+ # RainTomorrow prediction
+ rain_tomorrow_prob <- predict(rain_tomorrow_model,
today_record, type="prob")
+ rain_tomorrow_pred = ifelse(rain_tomorrow_prob$yes >=
cutoff, "yes", "no")
+
+ # Rainfall prediction iff RainTomorrow prediction is
yes;
+ chance_of_rain_probability
+ rainfall_pred <- NA
+ chance_of_rain <- NA
+ if (rain_tomorrow_pred == "yes") {
+ rainfall_pred <- round(predict(rf_fit,
today_record), 1)
+ chance_of_rain <- round(rain_tomorrow_prob$yes*100)
+ }
+
+ # WindGustSpeed prediction
+ wgs_pred <- round(predict(wgs_fit, today_record), 1)
+
+ # Humidity3pm prediction
+ h3pm_pred <- round(predict(h3pm_fit, today_record), 1)
+
+ # sunshine prediction is used to fit Cloud9am and
Cloud3pm
+ sun_pred <- predict(sun_fit, today_record)
+
+ # MinTemp prediction
+ mintemp_pred <- round(predict(mintemp_fit,
today_record), 1)
+
+ # MaxTemp prediction
+ maxtemp_pred <- round(predict(maxtemp_fit,
today_record), 1)
+
+ # converting all numeric predictions to strings
+ if (is.na(rainfall_pred)) {
+ rainfall_pred_str <- "< 0.25 mm"
+ } else {
+ rainfall_pred_str <- paste(rainfall_pred, "mm", sep
= " ")
+ }
```

Dan untuk hasil visual dari reporting dapat dilihat pada gambar 4.16

```
> (tomorrow_report <- weather_report(data_skrripsi[1098,], mod_ev_c3_fit, 0.36))
Rainfall ChanceofRain windgustspeed humidity Mintemp Maxtemp
1 3.8 mm 36% 8.8 km/h 0.1% 75.5°C 91.5°C
> (tomorrow_report <- weather_report(data_skrripsi[1099,], mod_ev_c3_fit, 0.36))
Rainfall ChanceofRain windgustspeed humidity Mintemp Maxtemp
1 4 mm 38% 4.3 km/h 0% 76.2°C 91.4°C
> (tomorrow_report <- weather_report(data_skrripsi[1100,], mod_ev_c3_fit, 0.36))
Rainfall ChanceofRain windgustspeed humidity Mintemp Maxtemp
1 < 0.25 mm 3.8 km/h 0.1% 75.3°C 92°C
> (tomorrow_report <- weather_report(data_skrripsi[56,], mod_ev_c3_fit, 0.36))
Rainfall ChanceofRain windgustspeed humidity Mintemp Maxtemp
1 < 0.25 mm 6.4 km/h 95.9% 76°C 90.1°C
> (tomorrow_report <- weather_report(data_skrripsi[900,], mod_ev_c3_fit, 0.36))
Rainfall ChanceofRain windgustspeed humidity Mintemp Maxtemp
1 < 0.25 mm 5.1 km/h 0.3% 75.6°C 92.4°C
> (tomorrow_report <- weather_report(data_skrripsi[46,], mod_ev_c3_fit, 0.36))
Rainfall ChanceofRain windgustspeed humidity Mintemp Maxtemp
1 < 0.25 mm 5.9 km/h 90.9% 76.3°C 90.9°C
> (tomorrow_report <- weather_report(data_skrripsi[1100,], mod_ev_c3_fit, 0.36))
Rainfall ChanceofRain windgustspeed humidity Mintemp Maxtemp
1 < 0.25 mm 3.8 km/h 0.1% 75.3°C 92°C
> (tomorrow_report <- weather_report(data_skrripsi[1123,], mod_ev_c3_fit, 0.36))
Rainfall ChanceofRain windgustspeed humidity Mintemp Maxtemp
1 3.8 mm 36% 8.5 km/h 0.1% 76.6°C 92°C
> (tomorrow_report <- weather_report(data_skrripsi[1124,], mod_ev_c3_fit, 0.36))
Rainfall ChanceofRain windgustspeed humidity Mintemp Maxtemp
1 < 0.25 mm 8.1 km/h 0.3% 76.2°C 90.9°C
> (tomorrow_report <- weather_report(data_skrripsi[1125,], mod_ev_c3_fit, 0.36))
Rainfall ChanceofRain windgustspeed humidity Mintemp Maxtemp
1 3.7 mm 36% 10.9 km/h 0.1% 76.8°C 91.4°C
> (tomorrow_report <- weather_report(data_skrripsi[1130,], mod_ev_c3_fit, 0.36))
Rainfall ChanceofRain windgustspeed humidity Mintemp Maxtemp
1 < 0.25 mm 5.3 km/h 0.1% 75.7°C 90.5°C
> (tomorrow_report <- weather_report(data_skrripsi[1156,], mod_ev_c3_fit, 0.36))
Rainfall ChanceofRain windgustspeed humidity Mintemp Maxtemp
1 < 0.25 mm 11.4 km/h 0.1% 77.6°C 92.4°C
> (tomorrow_report <- weather_report(data_skrripsi[1154,], mod_ev_c3_fit, 0.36))
```

Gambar 4.16 Visualisasi proses Reporting

5. KESIMPULAN

Berdasarkan hasil analisis data mining prakiraan cuaca di daerah seberang ulu 1 Palembang, maka diperoleh beberapa kesimpulan sebagai berikut:

1. Metode analisis menggunakan regresi linear yang telah dilakukan menghasilkan suatu informasi berupa laporan data prakiraan cuaca apakah hari esok hujan ataupun tidak, dengan total record data sebanyak 1306.
2. Hasil dari pengujian regresi linear terhadap data cuaca yang terekam untuk daerah seberang ulu 1 Palembang dengan metode regresi linear mendapatkan nilai akurasi sebesar 70%. Sebuah nilai yang cukup tapi tidak terlalu tinggi.
3. Berdasarkan hasil reporting, apabila nilai curah hujan dibawah 0.25 mm, maka dipastikan akan menghasilkan kondisi raintom "no" dan apabila nilai nya diatas 0,25 dan semakin tinggi nilai nya, maka semakin besar pula persentase untuk kemungkinan hujan di hari esok.

DAFTAR PUSTAKA

- Gibbs.1987. World Survey of Climatology. Chicago: Elsevier Science
- Ur Rehman Asid.2018. Evaluation of Three-Hourly TMPA Rainfall Products Using Telemetric Rain Gauge Observation at Lain Nullah Basin in Islamabad, Pakistan. Artikel terakhir diakses tanggal 20 Februari 2018.
- Anhar.2010. panduan menguasai php & mysql. Jakarta: Media Kita
- Kristanto, Hariyanto. 1994. Konsep dan Perancangan database. Yogyakarta: Andi
- Ullmar, Molina. 2002. Database System: The Complete Book. New Jersey : Prentice Hall.
- Connolly, Thomas and Begg, Carolyn. 2010. Database Systems A Practical Approach to Design, Implementation, and Management Sixth Edition. Boston: Pearson Education.
- Dyche, Jill. 2000. E-Data : Turning data into information with data warehousing. Singapore: Addison wesley.
- Connolly, Thomas and Begg, Carolyn. 2005. Database Systems A Practical Approach to Design, Implementation, and Management Forth Edition. Boston: Pearson Education.
- Han, Jiawei and Kamber, M. 2006. Data Mining – Concepts and Techniques 2nd Edition. San Francisco: Morgan Kaufmann Publisher
- Maimon and last. 2010. Knowledge Discovery and Data Mining. Dordrecht : Kluwer.
- Olson, David dan Shi, Yong. 2008. Pengantar Ilmu Penggalan Data Bisnis. Jakarta : Salemba empat.
- Ali, Hasan. 2008. Marketing. Yogyakarta : Media Utama.
- Fadholi, Akhmad. 2013. *Persamaan Regresi Prediksi Hujan Bulanan di Pontianak dengan Prediktor Suhu dan Kelembapan Udara*. Jurnal Forum Ilmiah. Vol.10, No.2.

Data Science +, 2017. *Weather Forecast With Regression Models- part 1*.
<https://datascienceplus.com/weather-forecast-with-regression-models-part-1/> ,terakhir diakses pada
20 Februari 2019 pukul 20.05 WIB.

Ledolter, Johannes. 2013. *Data Mining and Business Analytics With R*. Canada: John Wiley & Sons.