Prediksi Curah Hujan Menggunakan Metode Regresi Bayesian di Kota Bandung pada Tahun 2022

Gilang Utama Auliarahman*, Suwanda

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*gilangrahman321@gmail.com, idris1000358@gmail.com

Abstract. Regression analysis is an analysis that predicts the functional relationship between the dependent variable and the independent variable and is then expressed in the form of a mathematical equation. Multiple regression analysis explains the relationship between one dependent variable and two or more independent variables. To find out how much the relationship requires parameter estimation which can be explained using the Ordinary Least Squares (OLS) method or the Least Squares Method. Then in addition to using these two methods, another method that can be used is the Bayesian Regression method which is based on the Bayes Theorem. The Bayesian Regression method is a method required to estimate the parameters to be estimated by utilizing prior information from a population. This information is then combined with information from the sample used in estimating population parameters. In this method, the researcher needs to determine the prior distribution of the estimated parameters. This distribution is based on the researcher's intuition so it is objective. After the information from the data obtained from sampling is combined with the prior information of the parameter, the posterior distribution of the parameter will be obtained. Then this Bayesian method treats all unknown parameters as random variables and has a distribution. In this case it is very helpful in predicting weather, one of which is predicting rainfall. In this thesis, the method is applied to predict rainfall in Bandung City in 2022. The results show that the prediction of rainfall is greater in January, May, June, July, August, and December. But for the months of February, April, September, and November, the predicted data is close to the actual data.

Keywords: *Regression Analysis, Bayesian Regression Method, Rainfall*.

Abstrak. Analisis regresi merupakan analisis yang memprediksi hubungan fungsional antara variabel terikat dengan variabel bebas lalu dinyatakan dalam bentuk persamaan matematik. Analisis regresi berganda menjelaskan adanya hubungan antara satu variabel terikat dengan dua atau lebih variabel bebas. Untuk mengetahui seberapa besarnya hubungan tersebut memerlukan estimasi parameter yang dapat dijelaskan menggunakan metode *Ordinary Least Squares* (OLS) atau Metode Kuadat Terkecil. Lalu selain menggunakan kedua metode tersebut, metode lainnya yang dapat dipakai adalah metode Regresi Bayesian yang didasari dari Teorema Bayes. Metode Regresi Bayesian merupakan metode yang diperlukan untuk menaksir parameter yang akan diestimasi dengan memanfaatkan informasi awal (*prior*) dari suatu populasi. Informasi ini kemudian digabungkan dengan informasi dari sampel yang digunakan dalam mengestimasi parameter populasi. Pada metode ini, peneliti perlu menentukan distribusi *prior* dari parameter yang ditaksir. Distribusi ini berdasarkan intuisi peneliti sehingga bersifat objektif. Setelah informasi dari data yang didapat dari pengambilan sampel digabungkan dengan infromasi awal (*prior*) dari parameter, akan didapat distribusi *posterior* dari parameter. Lalu metode Bayesian ini memperlakukan semua parameter yang tidak diketahui sebagai variabel *random* dan memiliki distribusi. Dalam hal tersebut sangat membantu dalam memprediksi cuaca, salah satunya yaitu memprediksi curah hujan. Pada skripsi ini menerapkan metode tersebut untuk memprediksi curah hujan di Kota Bandung pada tahun 2022. Didapatkan hasilnya bahwa prediksi curah hujan lebih besar pada bulan Januari, Mei, Juni, Juli, Agustus, dan Desember. Tetapi untuk bulan Februari, April, September, dan November memperlihatkan data hasil prediksi mendekati data aktual.

Kata Kunci: *Analisis Regresi, Metode Regresi Bayesian, Curah Hujan*.

Corresponding Author
Email: idris1000358@gmail.com

A. Pendahuluan

Analisis regresi adalah analisis yang memprediksi hubungan fungsional antara variabel terikat dengan variabel bebas lalu dinyatakan dalam bentuk persamaan matematik. Analisis regresi merupakan suatu proses statistik untuk mengestimasi hubungan antara variabel-variabel, yakni berupa teknik-teknik memodelkan dan melakukan analisis beberapa variabel atas dasar bentuk hubungan antara satu variabel terikat dan satu atau lebih variabel bebas (prediktor) (Amstrong, 2012). Lalu pada analisis regresi dibagi berdasarkan jumlah variabel bebasnya, yaitu analisis regresi linier, dan analisis regresi berganda. Analisis regresi linier merupakan analisis yang menjelaskan adanya hubungan antara satu variabel terikat dengan satu variabel bebas. Sementara untuk analisis regresi berganda merupakan analisis yang menjelaskan adanya hubungan antara satu variabel terikat dengan dua atau lebih variabel bebas. Lalu untuk mengetahui besarnya hubungan antara variabel terikat dengan variabel bebas diperlukan estimasi parameter, salah satunya menggunakan metode Ordinary Least Squares (OLS). Metode OLS ini merupakan suatu metode penaksiran parameter model regresi yang meminimumkan jumlah kuadrat error. Metode ini juga merupakan metode paling umum yang digunakan untuk estimasi parameter pada analisis regresi berganda (Aisha & Suliadi, 2023).

Seiring berkembangnya zaman, metode penaksiran parameter terus dikembangkan hingga ditemukan beberapa metode baru yang mempermudah penelitian, salah satunya yaitu metode regresi bayesian. Metode regresi bayesian merupakan metode yang memperlakukan semua parameter yang tidak diketahui sebagai variabel random dan memiliki distribusi. Metode ini tentu berbeda dengan teori statistik klasik (frequentist). Metode regresi bayesian merupakan metode yang diperlukan untuk menaksir parameter yang akan diestimasi dengan memanfaatkan informasi awal (prior) dari suatu populasi. Prior tersebut ditentukan berdasarkan intuisi peneliti sehingga bersifat objektif. Informasi ini kemudian digabungkan dengan informasi dari sampel yang digunakan dalam mengestimasi parameter populasi, sehingga akan didapatkan posterior dari parameter.

Metode regresi bayesian adalah sebuah metode yang menggabungkan informasi dari data sampel dan informasi data sebelumnya (prior). Terdapat beberapa jenis dari distribusi prior, tapi pada penelitian ini akan menggunakan prior konjugat. Kemudian informasi awal tersebut dikombinasikan dengan informasi data sampel, sehingga didapatkan posterior. Untuk mendapatkan posterior, dibutuhkan salah satu pendekatan yaitu Markov Chain Monte Carlo (MCMC) dengan menggunakan algoritma Gibbs Sampler. Pendekatan ini dilakukan karena pada perhitungan secara analitik di metode regresi bayesian sangat sulit untuk dilakukan. MCMC merupakan alat komputasi yang sangat penting dalam metode regresi bayesian, ini dikarenakan pada MCMC dapat menerapkan teknik integrasi kompleks yang tidak dapat dihitung secara analitik. Gagasan yang mendasari metode ini menghasilkan markov chain melalui iterasi pada simulasi monte carlo. Dalam MCMC terdapat beberapa algoritma, yaitu gibbs sampler dan Metrapolis-Hastings. Pada penelitian ini akan menggunakan algoritma Metrapolis-Hastings.

Musim di Indonesia dibagi menjadi dua jenis, yaitu musim hujan dan musim kemarau. Pada tiap musim tersebut biasanya berlangsung selama enam bulan, contohnya untuk musim hujan terjadi mulai bulan Oktober dan berakhir pada bulan Maret. Indonesia merupakan negara yang dilewati garis ekuator dan termasuk kedalam negara tropis, inilah mengapa curah hujan di Indonesia bisa dikatakan cukup tinggi. Rata-rata curah hujan yang terjadi berkisar antara 2.000-3.000 mm per tahun. Curah hujan merupakan ketinggian air hujan yang terkumpul dalam tempat yang datar, tidak menguap, tidak meresap, dan tidak mengalir. Satuan curah hujan dinyatakan dalam satuan milimeter untuk Indonesia. Ada beberapa faktor yang mempengaruhi curah hujan, seperti suhu udara, kelembapan udara, tekanan udara, kecepatan angin, penyinaran sinar matahari, dan sebagainya. Rata-rata curah hujan yang tercatat pada tahun 2022 yaitu sebesar 2.898 mm. Provinsi yang sering dilanda yaitu Sulawesi Utara yang berkisar 256 hari, dan provinsi yang paling rendah yaitu Sumatra Selatan yang berkisar 219 Hari. Untuk provinsi Jawa Barat berada pada posisi keempat yang curah hujannya berkisar 239 hari. Lalu pada data curah hujan di Kota Bandung di tahun 2022 sebesar 2311.4 mm. Jumlah tersebut meningkat dari tahun sebelumnya, tetapi masih rendah dibandingkan pada tahun 2020 yang sebesar 2418.6 mm. Data tersebut bersumber dari Badan Pusat Statistik (BPS). Karena pada kota Bandung tiap pertahunnya memiliki curah hujan yang berbeda-beda, ini berarti ada beberapa faktor yang mempengaruhi hal tersebut tetapi belum diketahui. Maka dari itu metode regresi bayesian dapat digunakan dalam hal tersebut karena memperlakukan semua parameter yang tidak diketahui menjadi variabel random dan memiliki distribusi. Oleh karena itu penulis tertarik untuk melakukan penelitian dengan menggunakan metode Regresi Bayesian pada prediksi curah hujan di Kota Bandung pada tahun 2022. Apakah metode tersebut akurat atau tidak dalam memprediksi curah hujan.

B. Metodologi Penelitian

Data yang digunakan merupakan data sekunder yang merujuk pada data yang diperoleh dari sumber yang sudah ada dan dikumpulkan dan dicatat oleh pihak lain. Data yang digunakan pada skripsi ini adalah data yang berasal dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) di Kota Bandung pada tahun 2020 hingga 2021. Data yang diambil berjumlah 24 yang datanya berupa jumlah curah hujan, tekanan udara, rata-rata temperatur, rata-rata kelembapan udara, dan rata-rata kecepatan angin menurut bulan di Kota Bandung.

Metode Analisis Data

Analisis Regresi Linier Berganda

Analisis regresi linier merupakan analisis yang bertujuan untuk mengetahui seberapa besar pengaruh dari variabel bebas terhadap variabel terikat (Kutner et al., 2004). Pada penelitian ini teknik analisis yang digunakan yaitu analisis regresi linier berganda. Analisis regresi linier berganda merupakan analisis yang digunakan untuk mencari suatu hubungan dalam bentuk persamaan antara variabel terikat dengan satu atau lebih variabel bebas. Lalu untuk persamaan dari analisis regresi linier berganda dinyatakan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i, i = 1, 2, \ldots, n \quad (1)$$

Dimana:

Y_i = Variabel terikat (Dependen) individu ke i

$\beta_0, \ldots, \beta_k$ = Parameter koefisien regresi

$x_{1i}, \ldots, x_{ki}$ = Variabel bebas (Independen) individu ke i

Metode *Ordinary Least Square* (OLS)

Ordinary Least Square (OLS) atau metode kuadrat terkecil merupakan suatu metode yang digunakan untuk menduga parameter dengan cara meminimumkan jumlah kuadrat residual. Dalam OLS hanya terdapat satu variabel terikat, sedangkan untuk variabel bebas terdapat satu atau lebih. Jika variabel bebas yang digunakan hanya satu dapat disebut dengan analisis regresi linier sederhana, sedangkan jika variabel bebas yang digunakan lebih dari satu disebut dengan analisis regresi linier berganda. Pada OLS bertujuan untuk meminimumkan jumlah kuadrat residual. Residual itu sendiri merupakan selisih antara nilai pengamatan (y) dengan nilai estimasinya $(\hat{y})$. Fungsi dari metode OLS dapat dituliskan sebagai berikut:

$$\min \sum_{i=1}^{n} \varepsilon_i^2 = \min \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \quad (2)$$

Adapun untuk menaksir parameternya adalah sebagai berikut (Makarti & Karim, 2017):

$$\hat{\beta} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y} \quad (3)$$

Nilai $\hat{\beta}$ merupakan penaksir OLS dari vektor parameter $\boldsymbol{\beta}$, $\boldsymbol{X}$ merupakakan matriks dari variabel bebas berukuran $n \times (k+1)$ dan $\boldsymbol{Y}$ merupakan matriks dari variabel terikat berukuran $n \times 1$. Analisis regresi linier berganda dengan metode OLS ini memiliki asumsi yang harus dipenuhi supaya memperoleh model regresi yang bersifat *Best Linear Unbiased Estimation (BLUE)* yaitu:

$$E(\hat{\beta}) = E((\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y}) = \boldsymbol{\beta} \quad (4)$$

Dan

$$Var(\hat{\beta}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma^2 \tag{5}$$

Jika σ^2 tidak diketahui, maka dapat ditaksir oleh:

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n} e_i^2 \tag{6}$$

Asumsi yang harus dipenuhi diantaranya galat menyebar secara normal dengan rata-rata nol ($\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$), ragam sisaan homogen untuk setiap variabel bebas ($Var\ (\varepsilon_i) = \sigma^2$), sisaan saling bebas atau tidak ada autokorelasi jika data merupakan runtun waktu $\left(Cov\ (\varepsilon_i, \varepsilon_j) \right) = 0; i \neq j$, dan tidak ada multikolinieritas diantara variabel bebas.

Metode Regresi Bayesian

Metode penaksiran Bayes akan digunakan pada penaksiran koefisien regresi linier berganda. Dari asumsi kekeliruan pada persamaan 1 bahwa berdistribusi normal n variat, $\varepsilon \sim N_n(0, \sigma^2\boldsymbol{I})$, berakibat bahwa vektor variabel terikat juga berdistribusi n variat, yaitu $Y \sim N_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$. Oleh karena itu fungsi kemungkinan dari $Y_1, Y_2, \dots, Y_n$ adalah:

$$f(y_1, y_2, \dots, y_n|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} f(y_i|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) \tag{7}$$

Asumsikan bahwa vektor parameter regresi ($\boldsymbol{B}$) merupakan vektor acak yang berdistribusi normal $(k+1)$ variat dan $d = k + 1$ dengan vektor rata-rata ($\boldsymbol{m}$) dan matriks kovarian ($\boldsymbol{V}$), dituliskan $\boldsymbol{B} \sim N_{k+1}(\boldsymbol{m}, \boldsymbol{V})$ sehingga fungsi densitas bersama dari $B_0, B_1, \dots, B_k$ adalah sebagai berikut:

$$f(\boldsymbol{\beta}|\sigma^2, \boldsymbol{m}, \boldsymbol{V}) = (2\pi\sigma^2)^{-\frac{d}{2}}|\boldsymbol{V}|^{-\frac{d}{2}} \times \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{m})^T\boldsymbol{V}^{-1}(\boldsymbol{\beta} - \boldsymbol{m})\right\} \tag{8}$$

Persamaan (8) merupakan fungsi densitas bersama dari distribusi prior $\boldsymbol{B}$. Untuk distribusi prior variabel acak varian ($\sum^2$) akan digunakan distribusi *Inverse Gamma* dengan fungsi densitas yang dinotasikan $I\mathcal{G}(\alpha, \beta)$ maka persamaannya dapat dinyatakan sebagai berikut:

$$f(\sigma^2|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{\alpha-1}\exp\left\{-\frac{\beta}{\sigma^2}\right\} \tag{9}$$

Dimana untuk $\alpha > 0$ dan $\beta > 0$. Selanjutnya adalah menentukan distribusi posterior dari parameter $\boldsymbol{B}$ dan $\sum^2$. Berdasarkan metode bayesian diperoleh fungsi densitas posteriornya sebagai berikut:

$$\begin{aligned} f(\boldsymbol{\beta}, \sigma^2|y, \boldsymbol{X}) &= \frac{f_y(y|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2)f(\boldsymbol{\beta}|\sigma^2, \boldsymbol{m}, \boldsymbol{V})f(\sigma^2|\alpha, \beta)}{f_y(y)} \\ &\propto f_y(y|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2)f(\boldsymbol{\beta}|\sigma^2, \boldsymbol{m}, \boldsymbol{V})f(\sigma^2|\alpha, \beta) \\ &\propto (\sigma^2)^{-\frac{n}{2}-\frac{d}{2}+a-1}\exp\left\{-\frac{A}{2\sigma^2}\right\} \end{aligned} \tag{10}$$

Maka akan didapatkan distribusi *posterior* gabungannya untuk β dan σ^2 yaitu sebagai berikut:

$$\begin{aligned} f(\beta, \sigma^2|y, X) &\propto f_y(y|X, \beta, \sigma^2)f(\beta|\sigma^2, m, V)f(\sigma^2|a, b) \\ &\propto (\sigma^2)^{-\frac{d}{2}}\exp\left\{-\frac{(\beta-\mu)^T\Lambda^{-1}(\beta-\mu)}{2\sigma^2}\right\} \\ &\times (\sigma^2)^{-\frac{n}{2}+a-1}\exp\left\{-\frac{m^TV^{-1}m - \mu^T\Lambda^{-1}\mu + 2b + y^Ty}{2\sigma^2}\right\} \end{aligned} \tag{11}$$

Oleh karena itu persamaan (11) menunjukkan distribusi *posterior* $f(\beta, \sigma^2|y, X)$

sebanding dengan perkalian kernel dari $N_n(\mu, \sigma^2\Lambda)$ dan $IG\left(a^* = -\frac{n}{2} + a, b^* = b + \frac{m^T V^{-1} m - \mu^T \Lambda^{-1} \mu + y^T y}{2}\right)$.

Karena distribusi posterior sebanding dengan perkalian normal multivariat dan invers gamma, maka distribusi prior pada persamaan (8) dan (9) disebut sebagai distribusi prior konjugat.

Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) adalah metode yang banyak digunakan dalam pemodelan dengan pendekatan Bayes untuk memperoleh solusi karena distribusi *posterior* yang ada memerlukan proses integrasi dengan dimensi yang besar sehingga proses perhitungannya menjadi sangat rumit (Carlin dan Chib 1995). MCMC juga dapat memberikan solusi dalam estimasi parameter suatu model dengan tingkat kesulitan yang tinggi (Gelman dkk, 2004). Dasar dari MCMC adalah dengan membangkitkan data parameter sesuai proses *Markov Chain* dengan menggunakan simulasi *Monte Carlo* secara iteratif sehingga akan diperoleh distribusi *posterior* yang stasioner. MCMC lebih bersifat umum dan fleksibel dibandingkan dengan teknik simulasi langsung. Selain *Markov Chain*, pada *Monte Carlo* digunakan untuk simulasi. Simulasi ini diperlukan untuk sampel acak dalam jumlah yang besar. Pada dasarnya pendekatan Monte Carlo digunakan ketika perhitungan secara analitik sulit untuk dilakukan seperti integral kompleks. Maka dari itu terdapat dua algoritma yang digunakan dalam MCMC, yaitu *Gibbs Sampler* dan Metropolis-Hastings.

C. Hasil Penelitian dan Pembahasan

Model Regresi Linier Berganda

Model regresi digunakan untuk mencari hubungan antara curah hujan (Y) dengan tekanan udara (X_1), rata-rata temperatur (X_2), rata-rata kelembapan udara (X_3), dan rata-rata kecepatan angin (X_4). Dengan bantuan software R Studio, didapatkan model regresi berganda sebagai berikut:

$$\hat{Y} = 40711.25 - 45.98X_1 + 12.36X_2 + 20.85X_3 - 2.42X_4$$

Uji Asumsi Klasik

Uji asumsi klasik dilakukan untuk mengetahui apakah model regresi sudah bagus atau tidak. Tujuan dari asumsi klasik ini untuk memberikan kepastian bahwa model regresi memiliki ketepatan dalam estimasi, tidak bias, dan konsisten. Pada penelitian ini akan menggunakan taraf nyata sebesar 5%. Beberapa pengujian asumsi klasik sebagai berikut:

Tabel 1. Uji Shapiro-Wilk

$\boldsymbol{Shapiro - Wilk\ normality\ test}$			
$\boldsymbol{data}$: $\boldsymbol{model\_freq\$residual}$			
W	0.97633	$P.Value$	0.82

Berikut merupakan pemeriksaan asumsi normalitas menggunakan taraf nyata sebesar 5% dengan uji Shapiro-Wilk yang disajikan pada tabel 1 didapatkan $p.value$ sebesar 0.82. Maka $p, value > 5\%$ atau $0.82 > 0.05$ yang artinya H_0 diterima didapatkan kesimpulan bahwa residual berdistribusi normal.

Tabel 2 Uji Glejser

$\boldsymbol{Uji\ Glejser}$			
$\boldsymbol{Statistic}$	$\boldsymbol{p.value}$	$\boldsymbol{parameter}$	$\boldsymbol{alternative}$
1.78	0.776	4	$greater$

Berikut merupakan pemeriksaan asumsi heterokedastisitas menggunakan taraf nyata sebesar 5% dengan uji Glejser yang disajikan pada tabel 2 didapatkan $p.value$ sebesar 0.776.

Maka $p.value > 5\%$ yaitu $0.776 > 0.05$ yang artinya H_0 diterima didapatkan kesimpulan bahwa tidak terdapat masalah heteroskedastistitas (residual homogen).

Tabel 3. Uji Durbin Watson

$\boldsymbol{Durbin\ Watson\ Test}$			
$\boldsymbol{data: Data\_lm}$			
DW	1.9983	$p.value$	0.4941

Berikut merupakan pemeriksaan asumsi autokorelasi dengan uji Durbin Watson yang disajikan pada tabel 3 didapatkan DW sebesar 1.9983. Maka $dU < d < 4 - dU$ yaitu $1.7753 < 1.9983 < 2.2247$ yang artinya H_0 diterima didapatkan kesimpulan bahwa tidak terdapat autokorelasi.

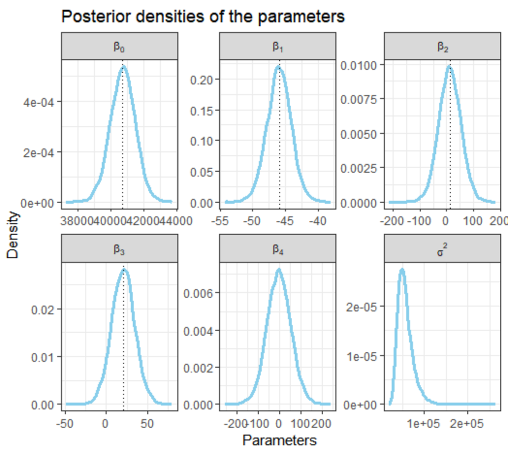
Tabel 4. Uji Multikolinieritas

$\boldsymbol{VIF}$			
$\boldsymbol{X_1}$	$\boldsymbol{X_2}$	$\boldsymbol{X_3}$	$\boldsymbol{X_4}$
1.289402	1.121067	1.179831	1.459492

Berikut merupakan pemeriksaan asumsi multikolinieritas dengan melihat nilai VIF yang disajikan pada tabel 4 didapatkan nilai VIF sebesar 1.289402 untuk X_1, 1.121067 untuk X_2, 1.179831 untuk X_3, dan 1.459492 untuk X_4. Pada nilai VIF tersebut terlihat bahwa semua nilai $VIF < 10$ yang artinya tidak terjadi gejala multikolinieritas.

Model Regresi Bayesian

Untuk mencari model regresi bayesian dibutuhkan distribusi posterior yang merupakan perpaduan distribusi prior konjugat yang diasumsikan $N_d(m, \sigma^2 V)$ dari β dan σ^2. Proses ini menggunakan metode MCMC dengan iterasi sebanyak 10.000 kali melalui proses algoritma Metrapolis-Hastings. Lalu untuk melihat apakah setiap parameter itu optimal. Dapat dilihat dari plot densitas posterior apakah mengikuti distribusi normal atau tidak yang dapat dinyatakan sebagai berikut:



Gambar 1. Distribusi Posterior

Berdasarkan Gambar 1 tersebut dapat dilihat bahwa plot pada parameter β_0 berdistribusi normal, plot pada parameter β_1 berdistribusi normal, plot pada parameter β_2 berdistribusi normal, plot pada parameter β_3 berdistribusi normal, dan plot pada parameter β_4 berdistribusi normal. Tetapi untuk σ^2 terlihat tidak berdistribusi normal. Karena prior yang digunakan berdistribusi normal, maka pada posterior akan didapatkan distribusi normal.

Karena dibutuhkan model regresi bayesian dengan parameter yang optimal, sehingga dapat memprediksi lebih akurat dibutuhkan fungsi kerugian. Fungsi kerugian yang digunakan untuk β dan σ sehingga didapatkan model regresi bayesian sebagai berikut:

$$\hat{Y} = 40714 - 45.960X_1 + 12.136X_2 + 20.590X_3 - 1.424X_4$$

Model tersebut mengartikan bahwa terjadi hubungan yang searah antara curah hujan (Y) dengan rata-rata temperatur (X_2), dan rata-rata kelembapan udara (X_3), sedangkan untuk tekanan udara (X_1), dan rata-rata kecepatan angin (X_4) memiliki hubungan yang sebaliknya. Dan σ^2 yang didapat sebesar 10.

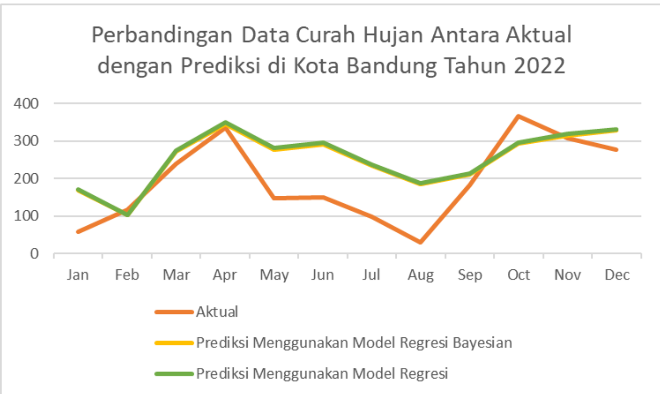
Prediksi Curah Hujan Menggunakan Metode Regresi Bayesian

Pada tahap ini untuk memprediksi curah hujan menggunakan model regresi bayesian, akan menggunakan model regresi bayesian dan variabel bebasnya yaitu tekanan udara (X_1), rata-rata temperatur (X_2), rata-rata kelembapan udara (X_3), dan rata-rata kecepatan angin (X_4) akan menggunakan data di Kota Bandung pada tahun 2022. Sehingga perbedaan pada prediksi menggunakan regresi bayesian dengan data aktual yang diperoleh dilapangan, sebagai berikut:

Tabel 5. Perbandingan Data Curah Hujan Antara Aktual dan Prediksi

Tanggal	Aktual	Prediksi Model Regresi Bayesian	Prediksi Model Regresi
Jan-22	59.5	169.0408	170.96
Feb-22	117.1	102.432	103.41
Mar-22	238.9	272.486	275.514
Apr-22	336.2	344.8152	349.574
May-22	146.9	278.138	281.804
Jun-22	150.6	291.5364	295.846
Jul-22	98.5	234.5572	237.77
Aug-22	29.9	185.5148	188.214
Sep-22	182.2	210.6592	214.222
Oct-22	366.7	293.0084	297.428
Nov-22	307.2	315.7036	319.934
Dec-22	277.7	328.1356	332.25

Dapat dilihat dari tabel 5 yaitu hasil prediksi curah hujan di Kota Bandung pada tahun 2022 menunjukkan pada bulan Februari, Maret, April, September, dan November yang mendekati data aktual. Tetapi untuk bulan lainnya menunjukkan hasil prediksi lebih besar (over estimate) dari data aktual.



Gambar 2. Perbandingan Grafik Curah Hujan Antara Aktual dengan Prediksi di Kota Bandung Pada Tahun 2022

Pada gambar 2 menjelaskan bahwa perbandingan data curah hujan antara aktual dengan prediksi menggunakan metode regresi bayesian pada Kota Bandung tahun 2022. Grafik tersebut memperlihatkan data hasil perdiksi lebih besar pada bulan Januari, Mei, Juni, Juli, Agustus, dan Desember. Tetapi untuk bulan oktober memperlihatkan data hasil prediksi lebih kecil. Lalu untuk bulan Februari, Maret, April, September, dan November memperlihatkan data hasil perdiksi mendekati data aktual.

D. Kesimpulan

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa model regresi bayesian yang didapatkan yaitu sebagai berikut:

$$\hat{Y} = 40714 - 45.960X_1 + 12.136X_2 + 20.590X_3 - 1.424X_4$$

Sehingga untuk prediksi curah hujan menggunakan metode regresi bayesian di Kota Bandung pada tahun 2022 dapat diterapkan. Lalu hasil yang didapatkan yaitu untuk hasil prediksi lebih besar pada bulan Januari, Mei, Juni, Juli, Agustus, dan Desember. Tetapi untuk bulan oktober memperlihatkan data hasil prediksi lebih kecil. Lalu untuk bulan Februari, Maret, April, September, dan November memperlihatkan data hasil perdiksi mendekati data aktual.

Acknowledge

Puji dan syukur penulis panjatkan kehadirat Allah SWT karena berkat rahmat dan hidayah-Nya penulis dapat menyelesaikan penelitian ini. Terimakasih kepada ayah, ibu, kakak-kakak dan keluarga yang selalu mendo'akan dan memberi dukungan baik moral maupun materi kepada penulis. Terimakasih kepada Bapak Dr. Suwanda, Drs., M.S. yang telah memberi bimbingan kepada penulis hingga penelitian ini selesai. Dosen-dosen Program Studi Statistika Universitas Islam Bandung yang telah banyak memberikan ilmu pengetahuan. Sahabat dan teman-teman serta semua pihak yang telah hingga penelitian ini selesai.

Daftar Pustaka

[1] Ali, A., Inglis, A. n., Prado, E., Wundervald, B. *Bayesian Linear Regression.*

[2] Amstrong, S. J. (2012). Illusion in Regression Analysis. *International Journal Forecasting*, Vol. 28, 689-693.

[3] Carlin, B. P., & Chib, S. (1995). *Bayesian Model Choice via Markov Chain Monte Carlo Methods*. Journal of the Royal Statistical Society. Series B (Methodological), 473-484.

[4] Gelman, Andrew, dkk. (2004). *Bayesian Data Analysis Second Edition.* Florida: CRC Press LLC

[5] Kutner, Nachtsheim, & Neter. (2004). *Applied Linear Regression Models Fourth Edition*. New York City: The McGraw-Hill Company, Inc.

[6] Makarti, p., & Karim, A. (2017). Perbandingan Metode Ordinary Least Square (Ols) Dan Metode Regresi Robust Pada Hasil Produksi Padi Di Kabupaten Indramay. Jurnal Unimus.

[7] Aisha Kusuma Putri, & Suliadi. (2023). Rekomendasi Destinasi Wisata di Indonesia Menggunakan Metode Item2Vec. *Jurnal Riset Statistika*, 11–18. https://doi.org/10.29313/jrs.v3i1.1770