

Penerapan Model Regresi Logistik Umum dan Dengan Metode Bayesian untuk Memprediksi Fraktur Tulang

Vika Valencia Susanto

2440062123

Teknik Informatika dan Statistika

vika.susanto@binus.ac.id

Abstrak

Beriringan dengan bertambahnya usia manusia di dunia, osteoporosis menjadi masalah kesehatan global terutama untuk individu lanjut usia. Osteoporosis menjadi faktor utama yang menyebabkan terjadinya fraktur pada tulang. Oleh karena itu pada *paper* ini akan dibuat sebuah model untuk melihat faktor biologis apa saja yang berperan terhadap resiko seseorang untuk mengalami fraktur agar masyarakat bisa lebih waspada akan kesehatan tulangnya dengan membuat model regresi logistik biasa dan model regresi logistik dengan metode bayesian.

KATA KUNCI: Fraktur – Regresi Logistik – Regresi Logistik dengan Metode Bayesian

Latar Belakang

Seiring dengan pergantian tahun, populasi manusia di dunia juga ikut menua. Osteoporosis merupakan masalah kesehatan masyarakat yang terjadi secara global dan umumnya menyerang individu dengan usia lanjut. Osteoporosis ditandai dengan berkurangnya kepadatan mineral tulang (*Bone Mineral Density*) yang bisa menyebabkan peningkatan resiko fraktur/patah tulang.

Fraktur pada tulang dapat disebabkan oleh berbagai faktor internal dan eksternal. Salah satu contoh faktor internalnya adalah faktor biologis penyebab osteoporosis seperti umur, jenis kelamin, kepadatan tulang, dan lain. Untuk faktor eksternal biasanya terjadi karena kecelakaan atau kesalahan obat-obatan.

Menurut National Osteoporosis Foundation, wanita cenderung lebih beresiko mengalami osteoporosis dan mengalami fraktur dengan perkiraan dari 10 juta orang Amerika yang terkena osteoporosis, 80% berjenis kelamin wanita. Kemudian, menurut Yayasan Osteoporosis Internasional, antara 12% sampai 20% dari penderita fraktur dan patah tulang akan meninggal dalam jangka waktu satu tahun. Indonesia sendiri merupakan negara dengan tingkat fraktur terbesar di Asia Tenggara dengan estimasi penderita sebanyak 1.3 juta orang setiap tahunnya. Oleh karena itu diperlukan sebuah model untuk melihat faktor biologis apa saja yang berperan terhadap

resiko seseorang untuk mengalami fraktur agar masyarakat bisa lebih waspada akan kesehatan tulangnya.

Penelitian Sebelumnya

Beberapa peneliti sudah membuat penelitian tentang memprediksi fraktur pada tulang. Salah satunya berjudul “Radiographic Texture Analysis of Densitometric Calcaneal Images: Relationship to Clinical Characteristics and to Bone Fragility” oleh Tamara Vokes. Penelitian ini meneliti tentang resiko fraktur dengan menggunakan RTA dengan menggunakan variabel BMD, umur, riwayat pengobatan, kerentanan fraktur, dan FRAX. Hasilnya menunjukkan bahwa orang dengan BMD tinggi dan lanjut usia menunjukkan resiko fraktur yang lebih tinggi. Bertolak belakang dengan mayoritas osteoporosis yang berjenis kelamin wanita, ditunjukkan bahwa pria memiliki tingkat kerentanan fraktur lebih tinggi dibandingkan wanita, namun dengan nilai BMD yang lebih rendah.

Kemudian ada juga penelitian dengan judul “Fracture Prediction, Imaging, and Screening in Osteoporosis” oleh Nicholas R, Elizabeth M, Kate A, Nicholas C, Elaine M, dan Cyrus Cooper. Penelitian ini meneliti hubungan osteoporosis dan fraktur dengan melihat kerentanan tulang dan BMD dan dikombinasikan dengan FRAX (*Fracture Risk Assessment Tool*). Kesimpulan dari penelitian ini adalah BMD, tingkat resiko, dan FRAX bisa digunakan untuk memprediksi apakah individual memiliki resiko tinggi terkena fraktur.

Metodologi

- Model Regresi Logistik

Model regresi logistik masih termasuk ke dalam model linear umum:

$$Y = a + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Dengan:

Y = variabel respon

$a = \beta_0$ = intercept

x_k = variabel independen

β_k = konstan mewakili parameter yang tidak diketahui

Model linear tadi disubstitusi untuk membuat fungsi peluang logistiknya:

$$P(Y_i \leq g|x_i) = f(Y) = \frac{1}{1 + e^{-Y}} = \frac{1}{1 + e^{-(a + \sum_{i=1}^k \beta_i X_i)}} = \frac{\exp(a_g + \sum_{i=1}^k \beta_i X_i)}{1 + \exp(a_g + \sum_{i=1}^k \beta_i X_i)}$$

Sehingga, model regresi logistik ordinal dapat ditulis dengan:

$$\text{logit}[P(Y_i < g|x_i)] = \ln \left[\frac{P(Y_i \leq g|x_i)}{1 - P(Y_i \leq g|x_i)} \right] = a_g + x_i \beta_i$$

Dengan,

Y = variabel respon/dependen

g = banyaknya kategori berskala ordinal pada variabel respon

x_i = vector variabel independen/prediktor pada pengamatan ke-i

- Estimasi Parameter

Maximum Likelihood Estimation

Estimasi parameter dilakukan dengan metode MLE karena dapat digunakan untuk model yang tidak linear seperti regresi logistik. Pertama, ambil n sampel random saling bebas yang berdistribusi Bernolli dengan fungsi peluang:

$$f(Y_i) = \pi(x_i)^{y_i}$$

Berikut adalah fungsi likelihood untuk 4 kategori respon:

$$l(\beta) = \prod_{i=1}^g [\pi_i(x_i)^{y_{ii}}]$$

Kemudian buat fungsi ln-likelihood atau logaritma naturalnya:

$$L(\beta) = \ln l(\beta) = \ln \left(\prod_{i=1}^n [\pi_i(x_i)^{y_{ii}}] \right) = \sum_{i=1}^n (y_{ii} \ln(\pi_i x_i))$$

Maksimumkan nilai ln-likelihood dengan menurunkan $L(\beta)$ terhadap β dan menyamadengkan dengan 0. Karena turunannya tidak linear, maka digunakan iterasi Newton Raphson untuk mendapatkan estimasi parameternya dengan rumus:

$$\beta^{t+1} = \beta^t - (H^t)^{-1} q^t$$

Dengan

β = parameter regresi

q^t = matriks turunan pertama terhadap parameter = $q^t = \left(\frac{\partial L(\beta)}{\partial \beta_0} \frac{\partial L(\beta)}{\partial \beta_1} \frac{\partial L(\beta)}{\partial \beta_2} \right)^T$

H^t = matriks turunan kedua terhadap parameternya

t = banyaknya iterasi

Iterasi dilakukan sampai konvergen dan berhenti apabila $||\beta^{t+1} - \beta^t|| \leq \epsilon$. Nantinya, hasil estimasi pada iterasi terakhir yang diperoleh adalah β^{t+1} .

Metode Bayesian

Metode Bayesian biasa digunakan untuk melakukan estimasi parameter regresi linear berganda dengan mempertimbangkan prior atau informasi awalnya. Prior ini bisa didapatkan dari penelitian-penelitian terdahulunya. Distribusi prior akan dikombinasikan dengan fungsi likelihood untuk menghasilkan distribusi posterior yang akan menjadi dasar inferensi.

Rumus umum metode bayes adalah:

$$P(\mu|y) = \frac{P(y|\mu)P(\mu)}{P(y)}$$

Dimana:

y = data

μ = parameter dalam model

$P(y)$ = bukti

$P(y|\mu)$ = Likelihood

$P(\mu) = \text{Prior}$

$P(\mu|y) = \text{Posterior}$

- Uji Signifikansi Parameter

Uji signifikansi parameter serentak

Likelihood Ratio Test

Metode yang biasanya dipanggil dengan *Likelihood Ratio Chi-square test* merupakan sebuah metode untuk memilih model terbaik dari dua model beranak. Model akan dipecah menjadi dua, yaitu model di bawah H_0 dengan model di bawah H_1 .

Hipotesis

$$H_0 = \beta_1 = \beta_2 = \dots = \beta_i = 0$$

$$H_1 = \text{Paling tidak ada satu } \beta \text{ yang tidak sama dengan } 0$$

H_0 akan ditolak jika nilai dari P-value kurang dari alpha yaitu 0.05.

Uji signifikansi parameter parsial

T-test

Merupakan sebuah tes statistik untuk membandingkan rata-rata dari dua grup. Tes ini digunakan untuk menguji hipotesis untuk melihat apakah variabel independen berpengaruh secara signifikan pada variabel dependen.

Hipotesis

$$H_0 = \beta_i = 0$$

$$H_1 = \beta_i \neq 0$$

Hasil dari *t-test* pada R akan menghasilkan dua nilai, yaitu t-value dan p-value. *T-value* merupakan cara untuk mengukur perbedaan rata-rata populasi. Nilai *T-value* akan dibandingkan dengan *T-table* dengan alpha sebesar 0.05 dan H_0 akan ditolak jika nilai *T-value* lebih besar dari pada *T-table*. Sedangkan *P-value* merupakan probabilitas untuk memperoleh *T-value* dengan nilai absolut dan akan menolak H_0 jika P-value bernilai kurang dari alpha yaitu 0.05.

Credible Interval

Credible interval merupakan interval di mana nilai parameter yang tidak teramati jatuh dengan probabilitas tertentu. *Credible interval* akan memprediksi bahwa sebuah koefisien berada di atas nilai *CredibleInterval_low* dan di bawah *CredibleInterval_high*. Sebuah variabel independen bisa dibilang berpengaruh secara signifikan terhadap variabel dependen jika dalam *credible interval*-nya tidak mencakup nilai 0.

- *Backward Elimination Method*

Metode *backward elimination* merupakan salah satu metode yang digunakan untuk variable selection, di mana variabel independen yang tidak mempengaruhi variabel dependen secara signifikan akan dieliminasi.

Cara dari metode ini adalah dengan mengeliminasi variabel yang paling tidak berpengaruh satu per satu dengan melihat nilai *P-value* terbesar dari setiap variabel. Satu per satu variabel akan dieliminasi sampai mendapatkan model yang signifikan.

- Uji kelayakan (Hosmer-Lemeshow Test)
Merupakan uji kelayakan yang digunakan untuk model regresi logistik, dan khusus untuk variabel independen dalam bentuk *binary*. Tes ini menyatakan seberapa baik data sesuai dengan modelnya.

Hipotesis

H0: model yang digunakan sudah sesuai dengan data (tidak ada perbedaan signifikan)

H1: model yang digunakan belum sesuai dengan data (ada perbedaan signifikan)

Tolak H0 jika nilai $p\text{-value} < \alpha (0.05)$

- Evaluasi Model

Confusion Matrix dan Akurasi

Umumnya *confusion matrix* berwujud tabel dengan 4 kombinasi berbeda dari nilai prediksi dan aktual. Ada 4 interpretasi dari *confusion matrix*.

	Positif	Negatif
Positif	<i>True positive</i>	Error 1: <i>false positive</i>
Negatif	Error 2: <i>false negative</i>	<i>True negative</i>

True positive: prediksi positif dan aktual positif

True negative: prediksi negatif dan aktual negatif

False positive (Error tipe 1): prediksi positif dan aktual negatif

False negative (Error tipe 2): prediksi negatif dan aktual positif

Dari *confusion matrix* ini nantinya dapat dihitung akurasi dari model dalam memprediksi.

Root Mean Square Error

Merupakan metode pengukuran untuk mengukur perbedaan nilai prediksi sebuah model dengan nilai aktual observasinya. RMSE merupakan akar kuadrat dari MSE. Semakin kecil nilai RMSE, maka model prediksi bisa dibilang lebih akurat.

Mean Absolute Percentage Error

MAPE merupakan salah satu model untuk mengukur keakuratan sebuah model dalam melakukan prediksi. Nilai MAPE menyatakan persentase kesalahan hasil pendugaan. Nilai MAPE sebesar 15% menyatakan bahwa selisih rata-rata nilai prediksi dan aktual adalah 15%. Berikut adalah interpretasi dari persentase MAPE:

Nilai MAPE	Interpretasi
<10	Model sangat akurat
10-20	Model melakukan prediksi dengan baik
20-50	Model melakukan prediksi dengan layak
>50	Model tidak akurat

Semakin rendah nilai MAPE, maka model akan dianggap semakin akurat dalam memprediksi.

Dataset

Dataset diambil dari “BIO STATISTICS COLLABORATION OF AUSTRALIA; Machine Learning for Biostatistics”

Link: https://bookdown.org/tpinto_home/Introduction2022/#datasets-used-in-the-examples

Atau pada Kaggle <https://www.kaggle.com/datasets/amarsharma768/bmd-data>

id	age	sex	fracture	weight_kg	height_cm	medication	waiting_tin	bmd
469	57.05277	F	no fracture	64	155.5	Anticonvul:	18	0.8793
8724	75.74122	F	no fracture	78	162	No medical	56	0.7946
6736	70.7789	M	no fracture	73	170.5	No medical	10	0.9067
24180	78.24718	F	no fracture	60	148	No medical	14	0.7112
17072	54.19188	M	no fracture	55	161	No medical	20	0.7909
3806	77.17775	M	no fracture	65	168	No medical	7	0.7301
17106	56.18062	M	no fracture	77	159	No medical	26	1.0096
23834	49.91614	F	no fracture	59	150	No medical	9	0.731
2454	68.4084	M	no fracture	64	167	Glucocortic	6	0.6893
2088	66.25665	M	no fracture	72	159.5	No medical	10	0.9466
5364	45.86658	M	no fracture	62	169	No medical	12	0.8015
8922	73.97047	F	no fracture	68	164	No medical	5	0.5793
23890	60.55542	F	no fracture	76	155	No medical	11	0.976
3047	64.2131	M	no fracture	90	175	Glucocortic	28	0.9184
2179	53.39544	M	no fracture	70	162.5	No medical	73	0.802
3800	66.82546	M	no fracture	76	171	No medical	13	0.8033
7528	57.93459	M	no fracture	67	160	Glucocortic	5	0.7978
5288	40.23237	M	no fracture	66	165	No medical	8	1.039
109	69.04818	F	no fracture	72	154	No medical	11	0.7861

Dataset berisi 169 sampel yang didapat dari tes kepadatan mineral tulang atau pemeriksaan BMD. Terdapat 9 variabel :

No.	Variabel	Deskripsi
1	id	nomor pasien (unique)
2	age	umur pasien
3	sex	jenis kelamin pasien (F: Female, M: Male)
4	fracture	apakah pasien memiliki fraktur panggul atau tidak (fracture/no fracture)
5	weight_kg	berat badan pasien dalam satuan kilogram
6	height_cm	tinggi badan pasien dalam satuan sentimeter
7	medication	pengobatan untuk pasien
8	waiting_time	lamanya waktu pasien menunggu untuk pemeriksaan BMD
9	bmd	bone mineral density atau kepadatan mineral tulang di panggul

```
> summary(fracture)
```

id	age	sex	fracture	weight_kg	height_cm	medication
Min. : 35	Min. :35.81	F:83	fracture : 50	Min. :36.00	Min. :142.0	Anticonvulsant : 9
1st Qu.: 2018	1st Qu.:54.42	M:86	no fracture:119	1st Qu.:56.00	1st Qu.:154.0	Glucocorticoids: 24
Median : 6702	Median :63.49			Median :64.50	Median :160.5	No medication :136
Mean : 9103	Mean :63.63			Mean :64.67	Mean :160.3	
3rd Qu.:17100	3rd Qu.:72.08			3rd Qu.:73.00	3rd Qu.:166.0	
Max. :24208	Max. :88.75			Max. :96.00	Max. :177.0	
	waiting_time					
	bmd					
Min. : 5.00	Min. :0.4076					
1st Qu.: 9.00	1st Qu.:0.6708					
Median :14.00	Median :0.7861					
Mean :19.74	Mean :0.7831					
3rd Qu.:24.00	3rd Qu.:0.8888					
Max. :96.00	Max. :1.3624					

Variabel numerik: age, weight_kg, height_cm, waiting_time, bmd

Variabel kategorik: id, sex, fracture, medication

Hasil dan Pembahasan

Komputasi akan dilakukan dengan menggunakan software R. Variabel “id” hanya merupakan pengenal untuk setiap pasien dan variabel “waiting_time” merupakan lamanya waktu pasien menunggu untuk pemeriksaan sehingga tidak akan dipakai di dalam model. Data akan dicek *missing value*nya sebelum diproses dan kebetulan dataset ini tidak memiliki *missing value*. Sebelum membuat model, variabel “weight_kg” dan “height_cm” akan dikonversi ke dalam bentuk BMI dengan rumus:

$$BMI = \frac{Berat(kg)}{Tinggi(m)^2}$$

Model akan dibuat untuk memprediksi hasil fracture, di mana kategori “fracture” akan dianggap belabel 1 dan kategori “no fracture” akan berlabel 0. Sehingga variabel yang akan digunakan adalah:

Variabel independen	age, sex, medication, bmi, bmd
Variabel dependen	Fracture

Model yang akan dibuat ada 2, yaitu model regresi logistik biasa dan model regresi logistik dengan metode bayes. Dua model pertama yang terbuat adalah:

Model regresi logistik biasa:

$$\begin{aligned}
 \text{logit}[P] &= \ln \left[\frac{P}{1-P} \right] \\
 &= 10.52793 + 0.01582x_{age} + 0.98824x_{sexM} \\
 &\quad - 1.25121x_{medicationGlucocorticoids} + 0.60293x_{nomedication} \\
 &\quad - 0.03766x_{bmi} - 16.93851x_{bmd}
 \end{aligned}$$

Model regresi logistik dengan bayes

$$\begin{aligned}
\text{logit}[P] &= \ln \left[\frac{P}{1-P} \right] \\
&= 10.67902657 + 0.01802256x_{age} + 0.94562599x_{sexM} \\
&\quad - 1.18543136x_{medicationGlucocorticoids} + 0.65567013x_{nomedication} \\
&\quad - 0.0446487x_{bmi} - 17.20173583x_{bmd}
\end{aligned}$$

Uji signifikansi akan dilakukan pada setiap model untuk melihat variabel independen yang tidak berpengaruh signifikan terhadap variabel dependen. Likelihood Ratio Test dan *T-test* akan digunakan untuk model regresi logistik biasa dan *credible interval* akan digunakan untuk model regresi logistik dengan metode bayesian.

Berikut adalah nilai *p-value* dan *credible interval* untuk setiap variabel pada model pertama:

Variabel	Metode		
	Regresi Logistik biasa		Regresi Logistik dengan metode bayes
	LRTest	T-test	T-test
age	<2.2e-16	0.469872	-0.01767562 - 0.05477485
sexM		0.064416	0.12946967 - 1.80930654
medicationGlucocorticoids		0.296785	-2.88749586 - 0.49883759
medicationNo medication		0.532495	-0.80237761 - 2.06890405
bmi		0.531455	-0.14895137 - 0.05888765
bmd		6.8e-08	-22.63514624 – (-12.51017897)

Variabel dengan nilai *P-value* kurang dari alpha (0.05) dan *credible interval* yang mencakup 0 ditunjukkan pada angka berwarna merah pada tabel di atas. Sebelum dilanjutkan ke metode *backward elimination*, setiap variabel kategorik dengan lebih dari 2 kategori yang faktornya tidak berpengaruh signifikan akan digabung menjadi satu kategori. Pada kasus ini kategori “Glucocorticoids” dan “Anticonvulsant” pada variabel “medication: akan digabung menjadi kategori “med”, sehingga ketika diuji kembali, dihasilkan:

Variabel	Metode	
	Regresi Logistik biasa	Regresi Logistik dengan metode bayes
	T-test	Credible interval
age	0.38636	-0.01534010 - 0.05594332
sexM	0.07830	0.07009733 - 1.75562864
medicationNo medication	0.04326	0.26610938 - 2.38562839
bmi	0.38994	-0.15722230 - 0.04545287
bmd	5.72e-08	-21.84152562 – (-12.21241965)

Setelah dilakukan kalkulasi ulang, metode *backward elimination* akan dilakukan dengan mengeliminasi variabel dengan nilai *P-value* tertinggi yaitu “bmi”. Setelah dilakukan *backward elimination method* yang pertama, dilakukan lagi uji signifikansi yang menghasilkan nilai:

Variabel	Metode	
	Regresi Logistik biasa	Regresi Logistik dengan metode bayes
	T-test	Credible interval
age	0.478009	-0.01793241 - 0.05250693
sexM	0.047969	0.16165411 - 1.82126427
medicationNo medication	0.043739	0.26970626 - 2.39402722
bmd	6.17e-09	-22.34816694 – (-13.00231519)

Hasil uji signifikansi dari model *backward elimination* yang pertama menunjukkan bahwa masih ada satu variabel yang perlu dieliminasi yaitu variabel “age”. Oleh karena itu, akan dilakukan kembali metode *backward elimination* kedua untuk mengeliminasi variabel “age”. Sehingga uji signifikansi final menghasilkan nilai p-value dan credible interval sebagai berikut:

Variabel	Metode	
	Regresi Logistik biasa	Regresi Logistik dengan metode bayes
	T-test	Credible interval
sexM	0.0439	0.2292865 - 1.791816
medicationNo medication	0.0426	0.2195824 - 2.344897
bmd	3.25e-10	-22.8304264 – (-13.758504)

Karena hasil dari uji signifikansi parameter sudah menunjukkan hasil bahwa semua variabel independen berpengaruh secara signifikan terhadap variabel dependen, maka kedua model sudah dianggap sebagai model terbaik. Kedua model tersebut dapat dijabarkan sebagai berikut:

Model final regresi logistik biasa:

$$\begin{aligned} \text{logit}[P] &= \ln \left[\frac{P}{1-P} \right] \\ &= 10.5854 + 1.0402x_{\text{sexM}} + 1.3468x_{\text{nomedication}} - 17.9534 x_{\text{bmd}} \end{aligned}$$

Model final regresi logistik dengan metode bayes:

$$\begin{aligned} \text{logit}[P] &= \ln \left[\frac{P}{1-P} \right] \\ &= 10.600370 + 1.009095x_{\text{sexM}} + 1.266994x_{\text{nomedication}} \\ &\quad - 17.857314 x_{\text{bmd}} \end{aligned}$$

Setelah model final terbuat, sekarang tahap selanjutnya adalah untuk menguji kelayakan model dengan menggunakan Hosmer-Lemeshow Test. Tes ini mengharapkan *P-value* bernilai lebih dari alpha (0.05) artinya gagal tolak H_0 yang menyatakan model yang dibuat sudah sesuai dengan data. Nilai *P-value* dari setiap model dari Hosmer-Lemeshow Test ditunjukkan pada tabel berikut:

Model	Hosmer-Lemeshow Test.
Model final regresi logistik biasa	0.539
Model final regresi logistik dengan metode bayes	0.5437

Hasil dari Hosmer-Lemeshow Test menunjukkan bahwa kedua model memiliki nilai *P-value* lebih besar dari pada alpha (0.05), sehingga kedua model dinyatakan layak dan sudah sesuai dengan data.

Masuk ke tahap terakhir yaitu evaluasi model. Model akan dievaluasi dengan melihat akurasi, nilai RMSE, dan juga MAPEnya. Ketiga nilai yang didapatkan dijabarkan pada tabel berikut:

Metode evaluasi	Regresi Logistik biasa	Regresi Logistik dengan metode bayes
Akurasi	88.17%	88.17%
RMSE	1.047168	1.046903
MAPE	0.8030928	0.802657

Kesimpulan

Setelah dilakukan uji signifikansi, kedua model terbaik dibuat dengan menggunakan variabel “sex”, “medication”, dan “bmd” tanpa menggunakan variabel “age” dan “bmi”.

Metode evaluasi	Regresi Logistik biasa	Regresi Logistik dengan metode bayes
Akurasi	88.17%	88.17%
RMSE	1.047168	1.046903
MAPE	0.8030928	0.802657

Dari nilai akurasi, RMSE, dan MAPE, dapat disimpulkan bahwa kedua model memiliki akurasi yang sama yaitu 88.17%, namun nilai RMSE dan MAPE menunjukkan bahwa model regresi logistik dengan metode bayes lebih akurat dalam memprediksi fraktur tulang dibandingkan dengan model regresi logistik biasanya.

Model:

$$\begin{aligned} \text{logit}[Fracture] \\ = 10.600370 + 1.009095x_{sexM} + 1.266994x_{nomedication} - 17.857314 x_{bmd} \end{aligned}$$

Kedepannya, pemecahan masalah prediksi fraktur tulang mungkin akan dikembangkan dengan menggunakan metode baru, atau dengan penambahan variabel lainnya. Dengan demikian, paper ini bisa digunakan sebagai referensi atau dasar untuk memprediksi fraktur pada tulang berdasarkan aspek umur, jenis kelamin, pengobatan, bmi, dan bmd dan bisa digunakan untuk membantu mengantisipasi fraktur tulang beserta menjadi acuan medis untuk kesehatan tulang.

Referensi

Tamara Vokes, Diane Lauderdale, Siu-Ling Ma, Mike Chinander, Keona Childs, and Maryellen Giger. Radiographic Texture Analysis of Densitometric Calcaneal Images: Relationship to Clinical Characteristics and to Bone Fragility. 10.1359/jbmr.090714.

Nicholas R. Fuggle, Elizabeth M. Curtis, Kate A. Ward, Nicholas C. Harvey, Elaine M. Dennison & Cyrus Cooper. Fracture prediction, imaging and screening in osteoporosis. Nature Reviews Endocrinology volume 15, hal535–547 (2019).

Cooper, C., Campion, G. & Melton, L. J. 3rd Hip fractures in the elderly: a world-wide projection. Osteoporos. Int. 2, 285–289 (1992).

Dr. H. L. Wong. Patah Tulang Pinggul pada Manula. [Online]; [diakses 2023 Januari 29]. Ref: <https://www21.ha.org.hk/smartpatient/EM/MediaLibraries/EM/EMMedia/Geriatic-Hip-Fracture-Indonesian.pdf?ext=.pdf>

CNN Indonesia. Alasan Wanita Lebih Rentan Terkena Osteoporosis. [Online]; 2021 [diakses 2023 Januari 29]. Ref: <https://www.cnnindonesia.com/gaya-hidup/20211015100824-255-708137/alasan-wanita-lebih-rentan-terkena-osteoporosis>

Code

Read dan import dataset

```
> getwd()
[1] "C:/users/62821/Documents/Binus Semester 5/Bayesian Data Analysis/Bayesian"
> data <- read.csv(file = "bmd.csv", header = TRUE)
> summary(data)
```

id	age	sex	fracture	weight_kg	height_cm	medication	waiting_time	bmd
Min. : 35	Min. :35.81	F:83	fracture : 50	Min. :36.00	Min. :142.0	Anticonvulsant : 9	Min. : 5.00	Min. :0.4076
1st Qu.: 2018	1st Qu.:54.42	M:86	no fracture:119	1st Qu.:56.00	1st Qu.:154.0	Glucocorticoids: 24	1st Qu.: 9.00	1st Qu.:0.6708
Median : 6702	Median :63.49			Median :64.50	Median :160.5	No medication :136	Median :14.00	Median :0.7861
Mean : 9103	Mean :63.63			Mean :64.67	Mean :160.3		Mean :19.74	Mean :0.7831
3rd Qu.:17100	3rd Qu.:72.08			3rd Qu.:73.00	3rd Qu.:166.0		3rd Qu.:24.00	3rd Qu.:0.8888
Max. :24208	Max. :88.75			Max. :96.00	Max. :177.0		Max. :96.00	Max. :1.3624

Lihat missing value dan hitung BMI

```
> sum(is.na(data))
[1] 0
>
> data$bmi <- data$weight_kg / (data$height_cm/100)^2
> summary(data)
```

id	age	sex	fracture	weight_kg	height_cm	medication	waiting_time	bmd
Min. : 35	Min. :35.81	F:83	fracture : 50	Min. :36.00	Min. :142.0	Anticonvulsant : 9	Min. : 5.00	Min. :0.4076
1st Qu.: 2018	1st Qu.:54.42	M:86	no fracture:119	1st Qu.:56.00	1st Qu.:154.0	Glucocorticoids: 24	1st Qu.: 9.00	1st Qu.:0.6708
Median : 6702	Median :63.49			Median :64.50	Median :160.5	No medication :136	Median :14.00	Median :0.7861
Mean : 9103	Mean :63.63			Mean :64.67	Mean :160.3		Mean :19.74	Mean :0.7831
3rd Qu.:17100	3rd Qu.:72.08			3rd Qu.:73.00	3rd Qu.:166.0		3rd Qu.:24.00	3rd Qu.:0.8888
Max. :24208	Max. :88.75			Max. :96.00	Max. :177.0		Max. :96.00	Max. :1.3624

bmi
Min. :15.43
1st Qu.:22.15
Median :24.96
Mean :25.20
3rd Qu.:27.55
Max. :38.54

Buat model pertama dengan semua variabel. Set seed agar hasil acak tidak berubah-ubah setiap kali dijalankan.

```
> modelLogistik <- glm(fracture=="fracture"~age + sex + medication + bmi + bmd,
+                       data = data, family = "binomial")
> summary(modelLogistik)
```

Call:

```
glm(formula = fracture == "fracture" ~ age + sex + medication +
    bmi + bmd, family = "binomial", data = data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.5709	-0.4920	-0.1432	0.3113	2.6146

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	10.52793	2.95531	3.562	0.000368	***
age	0.01582	0.02189	0.723	0.469872	
sexM	0.98824	0.53439	1.849	0.064416	.
medicationGlucocorticoids	-1.25121	1.19923	-1.043	0.296785	
medicationNo medication	0.60293	0.96591	0.624	0.532495	
bmi	-0.03766	0.06018	-0.626	0.531455	
bmd	-16.93851	3.13891	-5.396	6.8e-08	***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 205.27 on 168 degrees of freedom
 Residual deviance: 105.23 on 162 degrees of freedom
 AIC: 119.23

Number of Fisher Scoring iterations: 6

```
> set.seed(42)
> library(rstanarm)
> modelBayes <- stan_glm(fracture=="fracture"~age + sex + medication + bmi + bmd,
+                       data = data, family = "binomial", prior = normal(),
+                       prior_intercept = normal())
```

SAMPLING FOR MODEL 'bernoulli' NOW (CHAIN 1).

chain 1:

chain 1: Gradient evaluation took 0.001 seconds

chain 1: 1000 transitions using 10 leapfrog steps per transition would take 10 seconds.

chain 1: Adjust your expectations accordingly!

chain 1:

chain 1:

chain 1: Iteration: 1 / 2000 [0%] (warmup)

chain 1: Iteration: 200 / 2000 [10%] (warmup)

chain 1: Iteration: 400 / 2000 [20%] (warmup)

chain 1: Iteration: 600 / 2000 [30%] (warmup)

chain 1: Iteration: 800 / 2000 [40%] (warmup)

chain 1: Iteration: 1000 / 2000 [50%] (warmup)

chain 1: Iteration: 1001 / 2000 [50%] (sampling)

chain 1: Iteration: 1200 / 2000 [60%] (sampling)

chain 1: Iteration: 1400 / 2000 [70%] (sampling)

chain 1: Iteration: 1600 / 2000 [80%] (sampling)

chain 1: Iteration: 1800 / 2000 [90%] (sampling)

chain 1: Iteration: 2000 / 2000 [100%] (sampling)

chain 1:

chain 1: Elapsed Time: 0.721 seconds (warm-up)

chain 1: 0.696 seconds (sampling)

chain 1: 1.417 seconds (Total)

chain 1:

```
> summary(modelBayes)
```

Model Info:

```
function:      stan_glm
family:        binomial [logit]
formula:       fracture == "fracture" ~ age + sex + medication + bmi + bmd
algorithm:     sampling
priors:        see help('prior_summary')
sample:        4000 (posterior sample size)
observations:  169
predictors:    7
```

Estimates:

	mean	sd	2.5%	25%	50%	75%	97.5%
(Intercept)	10.8	2.9	5.4	8.8	10.7	12.6	16.8
age	0.0	0.0	0.0	0.0	0.0	0.0	0.1
sexM	1.0	0.5	0.0	0.6	0.9	1.3	2.0
medicationGlucocorticoids	-1.2	1.0	-3.2	-1.9	-1.2	-0.5	0.8
medicationNo medication	0.6	0.9	-1.1	0.1	0.7	1.2	2.4
bmi	0.0	0.1	-0.2	-0.1	0.0	0.0	0.1
bmd	-17.3	3.1	-23.7	-19.3	-17.2	-15.2	-11.7
mean_PPD	0.3	0.0	0.2	0.3	0.3	0.3	0.4
log-posterior	-65.8	1.9	-70.2	-66.8	-65.5	-64.4	-63.1

Diagnostics:

	mcse	Rhat	n_eff
(Intercept)	0.1	1.0	2757
age	0.0	1.0	2873
sexM	0.0	1.0	3366
medicationGlucocorticoids	0.0	1.0	2877
medicationNo medication	0.0	1.0	2934
bmi	0.0	1.0	3124
bmd	0.1	1.0	2354
mean_PPD	0.0	1.0	3884
log-posterior	0.0	1.0	1758

For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor on split chains (at convergence Rhat=1).

```
> modelBayes$coefficients
```

	age	sexM	medicationGlucocorticoids	medicationNo medication
(Intercept)	10.67902657	0.01802256	0.94562599	-1.18543136
bmi	-0.04464871	-17.20173583		0.65567013

```
> posterior_interval(modelBayes)
```

	5%	95%
(Intercept)	6.11617103	15.83478025
age	-0.01767562	0.05477485
sexM	0.12946967	1.80930654
medicationGlucocorticoids	-2.88749586	0.49883759
medicationNo medication	-0.80237761	2.06890405
bmi	-0.14895137	0.05888765
bmd	-22.63514624	-12.51017897

Gabungkan factor pada “medication”

```
> data$medication <- as.character(data$medication)
> data$medication <- as.array(data$medication)
> c <- ("No medication")
> data$medication[!(data$medication %in% c)] <- 'med'
> data$medication <- as.factor(data$medication)
> summary(data)
```

id	age	sex	fracture	weight_kg	height_cm	medication	waiting_time	bmd
Min. : 35	Min. :35.81	F:83	fracture : 50	Min. :36.00	Min. :142.0	med : 33	Min. : 5.00	Min. :0.4076
1st Qu.: 2018	1st Qu.:54.42	M:86	no fracture:119	1st Qu.:56.00	1st Qu.:154.0	No medication:136	1st Qu.: 9.00	1st Qu.:0.6708
Median : 6702	Median :63.49			Median :64.50	Median :160.5		Median :14.00	Median :0.7861
Mean : 9103	Mean :63.63			Mean :64.67	Mean :160.3		Mean :19.74	Mean :0.7831
3rd Qu.:17100	3rd Qu.:72.08			3rd Qu.:73.00	3rd Qu.:166.0		3rd Qu.:24.00	3rd Qu.:0.8888
Max. :24208	Max. :88.75			Max. :96.00	Max. :177.0		Max. :96.00	Max. :1.3624

```
bmi
Min. :15.43
1st Qu.:22.15
Median :24.96
Mean :25.20
3rd Qu.:27.55
Max. :38.54
```

Buat model baru dengan variabel “medication” yang baru

```
> modelLogistik2 <- glm(fracture=="fracture"~age + sex + medication + bmi + bmd,
+                        data = data, family = "binomial")
> summary(modelLogistik2)
```

```
Call:
glm(formula = fracture == "fracture" ~ age + sex + medication +
    bmi + bmd, family = "binomial", data = data)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5812  -0.5008  -0.1590   0.3283   2.5810
```

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      9.66428    2.80116   3.450  0.00056 ***
age               0.01864    0.02152   0.866  0.38636
sexM             0.93150    0.52907   1.761  0.07830 .
medicationNo medication 1.34076    0.66336   2.021  0.04326 *
bmi              -0.05087    0.05917  -0.860  0.38994
bmd             -16.53305    3.04620  -5.427  5.72e-08 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 205.27 on 168 degrees of freedom
Residual deviance: 106.35 on 163 degrees of freedom
AIC: 118.35
```

```
Number of Fisher Scoring iterations: 6
```

```
> modelBayes2 <- stan_glm(fracture=="fracture"~age + sex + medication + bmi + bmd,
+                          data = data, family = "binomial", prior = normal(),
+                          prior_intercept = normal())
```

```
SAMPLING FOR MODEL 'bernoulli' NOW (CHAIN 1).
```

```
Chain 1:
```

```
Chain 1: Gradient evaluation took 0 seconds
```

```
Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 0 seconds.
```

```
Chain 1: Adjust your expectations accordingly!
```

```
Chain 1:
```

```
Chain 1:
```

```
Chain 1: Iteration: 1 / 2000 [ 0%] (warmup)
```

```
Chain 1: Iteration: 200 / 2000 [ 10%] (warmup)
```

```
Chain 1: Iteration: 400 / 2000 [ 20%] (warmup)
```

```
Chain 1: Iteration: 600 / 2000 [ 30%] (warmup)
```

```
Chain 1: Iteration: 800 / 2000 [ 40%] (warmup)
```

```
Chain 1: Iteration: 1000 / 2000 [ 50%] (warmup)
```

```
Chain 1: Iteration: 1001 / 2000 [ 50%] (sampling)
```

```
Chain 1: Iteration: 1200 / 2000 [ 60%] (sampling)
```

```
Chain 1: Iteration: 1400 / 2000 [ 70%] (sampling)
```

```
Chain 1: Iteration: 1600 / 2000 [ 80%] (sampling)
```

```
Chain 1: Iteration: 1800 / 2000 [ 90%] (sampling)
```

```
Chain 1: Iteration: 2000 / 2000 [100%] (sampling)
```

```
Chain 1:
```

```
Chain 1: Elapsed Time: 0.85 seconds (warm-up)
```

```
Chain 1: 0.52 seconds (sampling)
```

```
Chain 1: 1.37 seconds (Total)
```

```
Chain 1:
```

```
> summary(modelBayes2)

Model Info:

function:      stan_glm
family:        binomial [logit]
formula:       fracture == "fracture" ~ age + sex + medication + bmi + bmd
algorithm:     sampling
priors:        see help('prior_summary')
sample:        4000 (posterior sample size)
observations:  169
predictors:    6

Estimates:
              mean   sd   2.5%   25%   50%   75%   97.5%
(Intercept)    9.9   2.8   4.8    7.9   9.8  11.7  15.6
age             0.0   0.0   0.0    0.0   0.0   0.0   0.1
sexM            0.9   0.5  -0.1    0.6   0.9   1.2   1.9
medicationNo medication 1.3   0.6   0.1    0.8   1.2   1.7   2.6
bmi            -0.1   0.1  -0.2   -0.1  -0.1   0.0   0.1
bmd           -16.9   2.9 -23.0  -18.8 -16.7 -14.9 -11.5
mean_PPD        0.3   0.0   0.2    0.3   0.3   0.3   0.4
log-posterior -64.9   1.8 -69.2  -65.8 -64.5 -63.6 -62.4

Diagnostics:
              mcse Rhat n_eff
(Intercept)  0.1 1.0 3068
age           0.0 1.0 3383
sexM          0.0 1.0 3403
medicationNo medication 0.0 1.0 3576
bmi           0.0 1.0 3376
bmd           0.1 1.0 2287
mean_PPD      0.0 1.0 4519
log-posterior 0.0 1.0 1724

For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor on split chains (at convergence Rhat=1).
> modelBayes2$coefficients
              (Intercept)              age              sexM medicationNo medication              bmi              bmd
              9.77785651              0.02085819              0.89932211              1.24190903              -0.05446538              -16.68106913
>
> posterior_interval(modelBayes2)
              5%              95%
(Intercept)  5.50839179 14.77926461
age          -0.01534010 0.05594332
sexM         0.07009733 1.75562864
medicationNo medication 0.26610938 2.38562839
bmi          -0.15722230 0.04545287
bmd          -21.84152562 -12.21241965
```

Backward elimination pertama

```
> #backward elimination 1
> modelLogistik3 <- glm(fracture=="fracture"~age + sex + medication + bmd,
+                        data = data, family = "binomial")
> summary(modelLogistik3)
```

Call:

```
glm(formula = fracture == "fracture" ~ age + sex + medication +
    bmd, family = "binomial", data = data)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-2.4703	-0.5079	-0.1572	0.3054	2.6259

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	9.13203	2.70093	3.381	0.000722	***
age	0.01479	0.02085	0.710	0.478009	
sexM	1.02536	0.51848	1.978	0.047969	*
medicationNo medication	1.34855	0.66873	2.017	0.043739	*
bmd	-17.23386	2.96520	-5.812	6.17e-09	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 205.27 on 168 degrees of freedom
 Residual deviance: 107.08 on 164 degrees of freedom
 AIC: 117.08

Number of Fisher scoring iterations: 6

```

<
> modelBayes3 <- stan_glm(fracture=="fracture"~age + sex + medication + bmd,
+                           data = data, family = "binomial", prior = normal(),
+                           prior_intercept = normal())

SAMPLING FOR MODEL 'bernoulli' NOW (CHAIN 1).
Chain 1:
Chain 1: Gradient evaluation took 0 seconds
Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 0 seconds.
Chain 1: Adjust your expectations accordingly!
Chain 1:
Chain 1:
Chain 1: Iteration:    1 / 2000 [  0%] (warmup)
Chain 1: Iteration:   200 / 2000 [ 10%] (warmup)
Chain 1: Iteration:   400 / 2000 [ 20%] (warmup)
Chain 1: Iteration:   600 / 2000 [ 30%] (warmup)
Chain 1: Iteration:   800 / 2000 [ 40%] (warmup)
Chain 1: Iteration:  1000 / 2000 [ 50%] (warmup)
Chain 1: Iteration:  1001 / 2000 [ 50%] (sampling)
Chain 1: Iteration:  1200 / 2000 [ 60%] (sampling)
Chain 1: Iteration:  1400 / 2000 [ 70%] (sampling)
Chain 1: Iteration:  1600 / 2000 [ 80%] (sampling)
Chain 1: Iteration:  1800 / 2000 [ 90%] (sampling)
Chain 1: Iteration:  2000 / 2000 [100%] (sampling)
Chain 1:
Chain 1: Elapsed Time: 0.498 seconds (warm-up)
Chain 1:                0.478 seconds (sampling)
Chain 1:                0.976 seconds (Total)
Chain 1:

> posterior_interval(modelBayes3)
              5%              95%
(Intercept)    4.95299610  13.71915414
age            -0.01793241   0.05250693
sexM           0.16165411   1.82126427
medicationNo medication  0.26970626   2.39402722
bmd            -22.34816694 -13.00231519

```

Backward elimination kedua


```
> modelLogistik4 <- glm(fracture=="fracture"~sex + medication + bmd,
+                        data = data, family = "binomial")
> summary(modelLogistik4)
```

Call:

```
glm(formula = fracture == "fracture" ~ sex + medication + bmd,
    family = "binomial", data = data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.4972	-0.4947	-0.1521	0.3286	2.7015

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	10.5854	1.8535	5.711	1.12e-08	***
sexM	1.0402	0.5162	2.015	0.0439	*
medicationNo medication	1.3468	0.6641	2.028	0.0426	*
bmd	-17.9534	2.8560	-6.286	3.25e-10	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 205.27 on 168 degrees of freedom
 Residual deviance: 107.59 on 165 degrees of freedom
 AIC: 115.59

Number of Fisher Scoring iterations: 6

```
> modelBayes4 <- stan_glm(fracture=="fracture"~sex + medication + bmd,
+                          data = data, family = "binomial", prior = normal(),
+                          prior_intercept = normal())
```

SAMPLING FOR MODEL 'bernoulli' NOW (CHAIN 1).

Chain 1:

chain 1: Gradient evaluation took 0 seconds

chain 1: 1000 transitions using 10 leapfrog steps per transition would take 0 seconds.

chain 1: Adjust your expectations accordingly!

Chain 1:

chain 1:

chain 1: Iteration: 1 / 2000 [0%] (warmup)

chain 1: Iteration: 200 / 2000 [10%] (warmup)

chain 1: Iteration: 400 / 2000 [20%] (warmup)

chain 1: Iteration: 600 / 2000 [30%] (warmup)

chain 1: Iteration: 800 / 2000 [40%] (warmup)

chain 1: Iteration: 1000 / 2000 [50%] (warmup)

chain 1: Iteration: 1001 / 2000 [50%] (sampling)

chain 1: Iteration: 1200 / 2000 [60%] (sampling)

chain 1: Iteration: 1400 / 2000 [70%] (sampling)

chain 1: Iteration: 1600 / 2000 [80%] (sampling)

chain 1: Iteration: 1800 / 2000 [90%] (sampling)

chain 1: Iteration: 2000 / 2000 [100%] (sampling)

chain 1:

chain 1: Elapsed Time: 0.496 seconds (warm-up)

chain 1: 0.462 seconds (sampling)

chain 1: 0.958 seconds (Total)

Chain 1:

```
> summary(modelBayes4)

Model Info:

function:      stan_glm
family:        binomial [logit]
formula:       fracture == "fracture" ~ sex + medication + bmd
algorithm:     sampling
priors:        see help('prior_summary')
sample:        4000 (posterior sample size)
observations:  169
predictors:    4

Estimates:
              mean      sd    2.5%   25%   50%   75%   97.5%
(Intercept)   10.7     1.8     7.3    9.4  10.6  11.9  14.5
sexM           1.0     0.5     0.1    0.7   1.0   1.3   2.0
medicationNo medication 1.3     0.6     0.0    0.8   1.3   1.7   2.6
bmd          -18.0     2.8   -23.8  -19.7 -17.9 -16.1 -12.9
mean_PPD       0.3     0.0     0.2    0.3   0.3   0.3   0.4
log-posterior -62.7     1.4   -66.4  -63.5 -62.4 -61.7 -61.0

Diagnostics:
              mcse  Rhat  n_eff
(Intercept)  0.0  1.0  2676
sexM         0.0  1.0  3098
medicationNo medication 0.0  1.0  2780
bmd          0.1  1.0  1976
mean_PPD     0.0  1.0  4393
log-posterior 0.0  1.0  1679

For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor on split chains (at convergence Rhat=1).
> modelBayes4$coefficients
              (Intercept)          sexM medicationNo medication          bmd
              10.600370          1.009095          1.266994         -17.857314
> posterior_interval(modelBayes4)
              5%          95%
(Intercept)  7.7702616  13.836937
sexM         0.2292865   1.791816
medicationNo medication 0.2195824   2.344897
bmd         -22.8304264 -13.758504
```

Uji kelayakan model

```
> library(ResourceSelection)
> aktual <- ifelse(data$fracture=="no fracture",0,1)
> hoslem.test(aktual, fitted(modelLogistik4))
```

Hosmer and Lemeshow goodness of fit (GOF) test

```
data: aktual, fitted(modelLogistik4)
X-squared = 6.9784, df = 8, p-value = 0.539
```

```
> hoslem.test(aktual, fitted(modelBayes4))
```

Hosmer and Lemeshow goodness of fit (GOF) test

```
data: aktual, fitted(modelBayes4)
X-squared = 6.9342, df = 8, p-value = 0.5437
```

Evaluasi model

```
> yhatBayes <- fitted(modelBayes4)
> head(yhatBayes)
      1      2      3      4      5      6
0.006050377 0.089312566 0.035068047 0.303062353 0.223239009 0.459795774
> predictBayess <- as.factor(ifelse(round(yhatBayes)==1, "fracture", "no fracture"))
> head(predictBayess)
      1      2      3      4      5      6
no fracture no fracture no fracture no fracture no fracture no fracture
Levels: fracture no fracture
>
> yhatBiner <- fitted(modelLogistik4)
> predictBiner <- as.factor(ifelse(round(yhatBiner)==1, "fracture", "no fracture"))
> head(predictBiner)
      1      2      3      4      5      6
no fracture no fracture no fracture no fracture no fracture no fracture
Levels: fracture no fracture
```

```
> library(caret)
> confusionMatrix(data$fracture, predictBayess)
Confusion Matrix and Statistics
```

	Reference	
Prediction	fracture	no fracture
fracture	38	12
no fracture	8	111

Accuracy : 0.8817
 95% CI : (0.8232, 0.9262)
 No Information Rate : 0.7278
 P-Value [Acc > NIR] : 9.688e-07

 Kappa : 0.7092
 McNemar's Test P-Value : 0.5023

 Sensitivity : 0.8261
 Specificity : 0.9024
 Pos Pred Value : 0.7600
 Neg Pred Value : 0.9328
 Prevalence : 0.2722
 Detection Rate : 0.2249
 Detection Prevalence : 0.2959
 Balanced Accuracy : 0.8643

 'Positive' Class : fracture

```
> confusionMatrix(data$fracture, predictBiner)
Confusion Matrix and Statistics
```

	Reference	
Prediction	fracture	no fracture
fracture	37	13
no fracture	7	112

Accuracy : 0.8817
 95% CI : (0.8232, 0.9262)
 No Information Rate : 0.7396
 P-Value [Acc > NIR] : 4.548e-06

 Kappa : 0.7057
 McNemar's Test P-Value : 0.2636

 Sensitivity : 0.8409
 Specificity : 0.8960
 Pos Pred Value : 0.7400
 Neg Pred Value : 0.9412
 Prevalence : 0.2604
 Detection Rate : 0.2189
 Detection Prevalence : 0.2959
 Balanced Accuracy : 0.8685

 'Positive' Class : fracture

```
> aktual1 <- as.factor(aktual)
> library(Metrics)
> rmse(as.numeric(aktual1), yhatBayes)
[1] 1.046903
> rmse(as.numeric(aktual1), yhatBiner)
[1] 1.047168
> mape(as.numeric(aktual1), yhatBayes)
[1] 0.802657
> mape(as.numeric(aktual1), yhatBiner)
[1] 0.8030928
```