

| Prediksi Penerimaan Calon Mahasiswa Pascasarjana

Bernadetha Emma / 2440015101

Diana Petrina / 2440015442

Felicia Ferren / 2440013071

Sumber Data

Mohan S Acharya, Asfia Armaan, Aneeta S Antony: *A Comparison of Regression Models for Prediction of Graduate Admissions*, IEEE International Conference on Computational Intelligence in Data Science 2019

Referensi:

Mohan S Acharya, “*Graduate Admissions*”, *Predicting Admission from Important Parameters*, Kaggle, April 2018. <https://www.kaggle.com/mohansacharya/datasets>.

Latar Belakang

Banyak calon mahasiswa pascasarjana yang mengalami dilema dalam memilih perguruan tinggi (S2). Setelah menentukan pilihan perguruan tinggi pun, calon mahasiswa masih perlu melewati proses admisi. Pada proses ini biasanya akan ada banyak prediktor dan konsultan yang membimbing calon mahasiswa, namun semuanya adalah berdasarkan proses-proses admisi pada tahun sebelumnya, dimana tidak menutup kemungkinan terdapat asumsi dan pandangan pribadi di dalamnya.

Sehingga diperlukan model yang dapat memprediksi tingkat probabilitas calon mahasiswa dapat diterima di suatu perguruan tinggi berdasarkan kompetensi yang dimilikinya secara objektif. Dari prediksi kemungkinan diterima tersebut, diharapkan calon mahasiswa bisa mendapatkan gambaran mengenai **kriteria dan peluang** penerimaan mahasiswa pada perguruan tinggi tersebut.

Previous Research

(Acharya)

- Membandingkan beberapa pendekatan model regresi linier dengan model dasar, seperti regresi support vector, regresi decision tree, dan regresi random forest.
- Matriks performa: Mean Squared Error (MSE) dan R-squared.
- Hasil: model regresi linier (model dasar) unggul > nilai MSE yang terendah dan nilai R-squared yang tertinggi.

(Syarifah)

- Membandingkan model regresi linier ganda dengan model bayesian linear.
- Ditemukan asumsi yang tidak dapat dipenuhi di model regresi linier ganda (model dasar) > alternatif: model dengan pendekatan Bayesian.

Struktur Data

Berikut ini adalah beberapa parameter yang dianggap penting dalam pendaftaran Program Magister:

No.	Nama Variabel	Jenis
1.	GRE Scores (out of 340)	Independen (x1)
2.	TOEFL Scores (out of 120)	Independen (x2)
3.	University Rating (out of 5)	Independen (x3)
4.	Statement of Purpose (out of 5)	Independen (x4)
5.	Letter of Recommendation Strength (out of 5)	Independen (x5)
6.	Undergraduate GPA (out of 10)	Independen (x6)
7.	Research Experience (either 0 or 1)	Independen (x7)
8.	Chance of Admit (ranging from 0 to 1)	Dependen (y)

Struktur data (dalam R)

```
> str(dataGraduate)
```

```
'data.frame':  400 obs. of  9 variables:
```

```
$ Serial.No.      : int  1 2 3 4 5 6 7 8 9 10 ...
```

```
$ GRE.Score       : int  337 324 316 322 314 330 321 308 302 323 ...
```

```
$ TOEFL.Score     : int  118 107 104 110 103 115 109 101 102 108 ...
```

```
$ University.Rating: int  4 4 3 3 2 5 3 2 1 3 ...
```

```
$ SOP             : num  4.5 4 3 3.5 2 4.5 3 3 2 3.5 ...
```

```
$ LOR             : num  4.5 4.5 3.5 2.5 3 3 4 4 1.5 3 ...
```

```
$ CGPA            : num  9.65 8.87 8 8.67 8.21 9.34 8.2 7.9 8 8.6 ...
```

```
$ Research        : int  1 1 1 1 0 1 1 0 0 0 ...
```

```
$ Chance.of.Admit : num  0.92 0.76 0.72 0.8 0.65 0.9 0.75 0.68 0.5 0.45 ...
```

Admission_Predict

Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
1	337	118	4	4.5	4.5	9.65	1	0.92
2	324	107	4	4	4.5	8.87	1	0.76
3	316	104	3	3	3.5	8	1	0.72
4	322	110	3	3.5	2.5	8.67	1	0.8
5	314	103	2	2	3	8.21	0	0.65
6	330	115	5	4.5	3	9.34	1	0.9
7	321	109	3	3	4	8.2	1	0.75
8	308	101	2	3	4	7.9	0	0.68
9	302	102	1	2	1.5	8	0	0.5
10	323	108	3	3.5	3	8.6	0	0.45

Data Pre-processing: Transformasi Data

- Mengkategorikan value pada *Chance.of.Admit*

```
> dataGraduate$Chance.of.Admit = ifelse(round(dataGraduate$Chance.of.Admit) == 1, 1, 0)
> head(dataGraduate, 10)
```

	GRE.Score	TOEFL.Score	University.Rating	SOP	LOR	CGPA	Research	Chance.of.Admit
1	337	118	4	4.5	4.5	9.65	1	1
2	324	107	4	4.0	4.5	8.87	1	1
3	316	104	3	3.0	3.5	8.00	1	1
4	322	110	3	3.5	2.5	8.67	1	1
5	314	103	2	2.0	3.0	8.21	0	1
6	330	115	5	4.5	3.0	9.34	1	1
7	321	109	3	3.0	4.0	8.20	1	1
8	308	101	2	3.0	4.0	7.90	0	1
9	302	102	1	2.0	1.5	8.00	0	0
10	323	108	3	3.5	3.0	8.60	0	0

Data Pre-processing: Transformasi Data

- **Mengubah variable *University.Rating* dan *Research* menjadi kategorik**

Variable *Research* diubah menjadi variabel kategorik dengan menggunakan function *factor()*.

```
> dataGraduate$Research = factor(dataGraduate$Research)
```

Sedangkan, variable *University.Rating* diubah ke dalam bentuk dummy:

```
> library(fastDummies)
> dataGraduate <- dummy_cols(dataGraduate, select_columns = "University.Rating", remove_first_dummy = TRUE,
+                             remove_selected_columns = TRUE)
> head(dataGraduate)
```

	GRE.Score	TOEFL.Score	SOP	LOR	CGPA	Research	Chance.of.Admit	University.Rating_2	University.Rating_3	University.Rating_4
1	337	118	4.5	4.5	9.65	1	1	0	0	1
2	324	107	4.0	4.5	8.87	1	1	0	0	1
3	316	104	3.0	3.5	8.00	1	1	0	1	0
4	322	110	3.5	2.5	8.67	1	1	0	1	0
5	314	103	2.0	3.0	8.21	0	1	1	0	0
6	330	115	4.5	3.0	9.34	1	1	0	0	0

	University.Rating_5
1	0
2	0
3	0
4	0
5	0
6	1

Hasil Transformasi Data

```
> str(dataGraduate)
'data.frame': 400 obs. of 11 variables:
 $ GRE.Score      : int  337 324 316 322 314 330 321 308 302 323 ...
 $ TOEFL.Score    : int  118 107 104 110 103 115 109 101 102 108 ...
 $ SOP            : num  4.5 4 3 3.5 2 4.5 3 3 2 3.5 ...
 $ LOR            : num  4.5 4.5 3.5 2.5 3 3 4 4 1.5 3 ...
 $ CGPA           : num  9.65 8.87 8 8.67 8.21 9.34 8.2 7.9 8 8.6 ...
 $ Research       : Factor w/ 2 levels "0","1": 2 2 2 2 1 2 2 1 1 1 ...
 $ Chance.of.Admit : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 1 1 ...
 $ University.Rating_2: int  0 0 0 0 1 0 0 1 0 0 ...
 $ University.Rating_3: int  0 0 1 1 0 0 1 0 0 1 ...
 $ University.Rating_4: int  1 1 0 0 0 0 0 0 0 0 ...
 $ University.Rating_5: int  0 0 0 0 0 1 0 0 0 0 ...
```

Membuat Model Regresi Logistik dengan Metode Bayesian

Model Bayesian dibangun dengan function ***stanglm*** pada library ***rstanarm***

```
> summary(stanModel)
```

```
Model Info:  
function: stan_glm  
family:    binomial [logit]  
formula:    Chance.of.Admit ~ .  
algorithm:  sampling  
sample:     4000 (posterior sample size)  
priors:     see help('prior_summary')  
observations: 400  
predictors: 11
```

```
Estimates:
```

	mean	sd	10%	50%	90%
(Intercept)	-51.3	10.2	-64.4	-51.0	-38.5
GRE.Score	0.0	0.0	0.0	0.0	0.1
TOEFL.Score	0.1	0.1	0.0	0.1	0.2
SOP	-0.6	0.4	-1.1	-0.6	-0.2
LOR	1.0	0.4	0.5	1.0	1.6
CGPA	3.4	0.8	2.4	3.4	4.5
Research1	0.0	0.6	-0.8	0.0	0.7
University.Rating_2	-1.1	0.6	-1.9	-1.1	-0.3
University.Rating_3	-1.1	0.8	-2.1	-1.1	0.0
University.Rating_4	-1.8	1.2	-3.2	-1.8	-0.3
University.Rating_5	0.0	2.0	-2.5	-0.1	2.8

```
Fit Diagnostics:
```

	mean	sd	10%	50%	90%
mean_PPD	0.9	0.0	0.9	0.9	0.9

The mean_ppd is the sample average posterior predictive distribution of the outcome variable (for details see help('summary.stanreg')).

```
MCMC diagnostics
```

	mcse	Rhat	n_eff
(Intercept)	0.2	1.0	3439
GRE.Score	0.0	1.0	3676
TOEFL.Score	0.0	1.0	3509
SOP	0.0	1.0	3283
LOR	0.0	1.0	3303
CGPA	0.0	1.0	2647
Research1	0.0	1.0	4104
University.Rating_2	0.0	1.0	2528
University.Rating_3	0.0	1.0	2573
University.Rating_4	0.0	1.0	2976
University.Rating_5	0.0	1.0	3146
mean_PPD	0.0	1.0	4030
log-posterior	0.1	1.0	1882

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).

Membuat Model Regresi Logistik dengan Metode Frequentist

Model Frequentist dibangun dengan
function **glm** pada library *rstanarm*

```
> base_model = glm(Chance.of.Admit~., data = dataGraduate, family="binomial")  
> summary(base_model)
```

Call:

```
glm(formula = Chance.of.Admit ~ ., family = "binomial", data = dataGraduate)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.13437	0.00701	0.08363	0.24903	1.86018

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-53.05948	10.88640	-4.874	1.09e-06	***
GRE.Score	0.03075	0.03871	0.794	0.4270	
TOEFL.Score	0.14252	0.08625	1.652	0.0984	.
SOP	-0.64637	0.37589	-1.720	0.0855	.
LOR	1.10454	0.44585	2.477	0.0132	*
CGPA	3.85219	0.90942	4.236	2.28e-05	***
Research1	-0.08854	0.61759	-0.143	0.8860	
University.Rating_2	-1.47725	0.73848	-2.000	0.0455	*
University.Rating_3	-1.53952	0.93820	-1.641	0.1008	
University.Rating_4	-2.62035	1.34733	-1.945	0.0518	.
University.Rating_5	10.85915	1100.43599	0.010	0.9921	

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 237.37 on 399 degrees of freedom
Residual deviance: 117.93 on 389 degrees of freedom
AIC: 139.93

Number of Fisher Scoring iterations: 18

Uji Signifikansi Parameter Serentak

dengan *likelihood ratio test*

$H_0: B_1 = B_2 = \dots = B_i = 0$

$H_1: \text{minimal ada satu } B_i \neq 0$

($i = 1, 2, 3, 4, 5, 6, 7$)

Tolak H_0 jika $p\text{-value} < \alpha (0.05)$

Kesimpulan:

P-value pada kedua model ($< 2.2e-16$) < 0.05 , sehingga tolak H_0 .

Dapat disimpulkan bahwa parameter tersebut signifikan terhadap masing-masing model secara keseluruhan.

MODEL FREQUENTIST

```
> lrtest(base_model)
Likelihood ratio test

Model 1: Chance.of.Admit ~ GRE.Score + TOEFL.Score + SOP + LOR + CGPA +
  Research + University.Rating_2 + University.Rating_3 + University.Rating_4 +
  University.Rating_5
Model 2: Chance.of.Admit ~ 1
#Df  LogLik  Df  Chisq Pr(>Chisq)
1   11  -58.963
2    1 -118.686 -10 119.45  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

MODEL FREQUENTIST

Uji Parameter Parsial (Wald Test)

$H_0: B_i = 0$

$H_1: B_i \neq 0$

($i = 1, 2, 3, 4, 5, 6, 7$)

Tolak H_0 jika $p\text{-value} < \alpha (0.05)$

Kesimpulan:

P-value pada LOR, CGPA, dan University.Rating_2 < 0.05 .

Dapat disimpulkan bahwa dari seluruh parameter, hanya parameter LOR, CGPA, dan University.Rating_2 yang signifikan terhadap model.

```
> base_model = glm(Chance.of.Admit~., data = dataGraduate, family="binomial")
> summary(base_model)
```

```
Call:
glm(formula = Chance.of.Admit ~ ., family = "binomial", data = dataGraduate)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.13437	0.00701	0.08363	0.24903	1.86018

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-53.05948	10.88640	-4.874	1.09e-06	***
GRE.Score	0.03075	0.03871	0.794	0.4270	
TOEFL.Score	0.14252	0.08625	1.652	0.0984	.
SOP	-0.64637	0.37589	-1.720	0.0855	.
LOR	1.10454	0.44585	2.477	0.0132	*
CGPA	3.85219	0.90942	4.236	2.28e-05	***
Research1	-0.08854	0.61759	-0.143	0.8860	
University.Rating_2	-1.47725	0.73848	-2.000	0.0455	*
University.Rating_3	-1.53952	0.93820	-1.641	0.1008	
University.Rating_4	-2.62035	1.34733	-1.945	0.0518	.
University.Rating_5	10.85915	1100.43599	0.010	0.9921	

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 237.37 on 399 degrees of freedom
Residual deviance: 117.93 on 389 degrees of freedom
AIC: 139.93

Number of Fisher Scoring iterations: 18

Uji Parameter Bayesian

Parameter dikatakan tidak signifikan apabila pada posterior intervalnya mengandung nilai 0

Kesimpulan:

Interval TOEFL.Score, SOP, LOR, CGPA, dan University.Rating_2 tidak mengandung nilai 0

Dapat disimpulkan bahwa dari seluruh parameter, hanya parameter TOEFL.Score, SOP, LOR, CGPA, dan University.Rating_2 yang signifikan terhadap model.

MODEL BAYESIAN

```
> posterior_interval(stanModel)
```

	5%	95%
(Intercept)	-68.489024841	-35.05551236
GRE.Score	-0.025150967	0.09907860
TOEFL.Score	0.005217254	0.27134011
SOP	-1.237886756	-0.04318669
LOR	0.382040735	1.74682254
CGPA	2.172194609	4.74259609
Research1	-0.957757901	0.91968193
University.Rating_2	-2.166583480	-0.10329930
University.Rating_3	-2.397375836	0.28283585
University.Rating_4	-3.605906713	0.18669754
University.Rating_5	-3.036489470	3.53425087

Uji Parameter Parsial (Wald Test) Model 2

MODEL FREQUENTIST

$H_0: B_i = 0$

$H_1: B_i \neq 0$

($i = 1, 2, 3, 4, 5, 6, 7$)

Tolak H_0 jika $p\text{-value} < \alpha (0.05)$

Kesimpulan:

P-value pada CGPA < 0.05.

Dapat disimpulkan bahwa dari ketiga parameter, hanya parameter CGPA yang signifikan terhadap model.

```
> base_model2 = glm(Chance.of.Admit~University.Rating_2+LOR+CGPA, data=dataGraduate, family = "binomial")
> summary(base_model2)

Call:
glm(formula = Chance.of.Admit ~ University.Rating_2 + LOR + CGPA,
    family = "binomial", data = dataGraduate)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.02723   0.02943   0.09380   0.27547   1.99016

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -31.4494     5.3643  -5.863 4.55e-09 ***
University.Rating_2 -0.6121     0.4611  -1.327  0.1844
LOR           0.6340     0.3530   1.796  0.0725 .
CGPA          3.9674     0.7076   5.607 2.06e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 237.37  on 399  degrees of freedom
Residual deviance: 130.74  on 396  degrees of freedom
AIC: 138.74

Number of Fisher Scoring iterations: 7
```


Uji Parameter Parsial (Wald Test) Model 2

Parameter dikatakan tidak signifikan apabila pada posterior intervalnya mengandung nilai 0

Kesimpulan:

Interval TOEFL.Score, SOP, LOR, dan CGPA tidak mengandung nilai 0

Dapat disimpulkan bahwa dari seluruh parameter, hanya parameter University.Rating_2 yang tidak signifikan terhadap model.

MODEL BAYESIAN

```
> posterior_interval(stanModel2)
```

	5%	95%
(Intercept)	-52.25407457	-29.4178664
TOEFL.Score	0.03231112	0.2623955
SOP	-1.30441230	-0.1858099
LOR	0.30718817	1.6109540
CGPA	2.13518390	4.7429402
University.Rating_2	-1.25375071	0.2915679

Modifikasi Model Regresi Logistik (Model 3) hanya dengan Parameter Signifikan

Frequentist:

$H_0: B1 = 0$

$H_1: B1 \neq 0$

Tolak H_0 jika $p\text{-value} < \alpha (0.05)$

Bayesian:

Parameter dikatakan tidak signifikan apabila pada posterior intervalnya mengandung nilai 0

Kesimpulan:

Frequentist:

P-value pada CGPA < 0.05 .

Dapat disimpulkan bahwa parameter CGPA signifikan terhadap model Frequentist.

Bayesian:

Tidak ada interval yang mengandung nilai 0.

Dapat disimpulkan bahwa semua parameter berpengaruh signifikan terhadap model Bayesian.

MODEL FREQUENTIST

```
> base_model3 = glm(Chance.of.Admit~CGPA, data=dataGraduate, family = "binomial")
> summary(base_model3)
```

Call:
glm(formula = Chance.of.Admit ~ CGPA, family = "binomial", data = dataGraduate)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.96569	0.04213	0.11535	0.28681	1.90345

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-33.9543	5.1407	-6.605	3.98e-11 ***
CGPA	4.4581	0.6498	6.861	6.84e-12 ***

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

MODEL BAYESIAN

```
> posterior_interval(stanModel3)
```

	5%	95%
(Intercept)	-52.23416799	-29.4427146
TOEFL.Score	0.03410571	0.2629598
SOP	-1.24444988	-0.1681373
LOR	0.25287529	1.6013730
CGPA	2.07255279	4.6202419

Hosmer and Lemeshow Goodness of Fit Test

Setelah ketiga model dari masing-masing metode dibangun, peneliti melakukan uji kecocokan model dengan dengan fungsi **hoslem.test()** dan hipotesis:

H0 : model sudah sesuai

H1 : model belum sesuai

Kaidah Keputusan: tolak H0 jika $P\text{-value} > \alpha(0.05)$

Seluruh memiliki nilai $P\text{-value} > \alpha(0.05)$, sehingga keputusan yang diambil adalah **gagal tolak hipotesis H0**. Artinya **seluruh model tersebut sudah sesuai** dengan data.

<i>Model</i>	<i>P-value</i>
<i>base_model3</i>	0.9977
<i>stanModel3</i>	0.609

Uji Asumsi: Multikolinearitas

MODEL FREQUENTIST

```
> vif(base_model1)
GRE.Score      TOEFL.Score      SOP      LOR      CGPA      Research
1.752911      2.185389      1.952549      1.544954      1.604826      1.172094
University.Rating_2 University.Rating_3 University.Rating_4 University.Rating_5
2.368324      2.796357      1.643700      1.000001

> vif(base_model2)
University.Rating_2      LOR      CGPA
1.018001      1.116491      1.116177
```

MODEL BAYESIAN

```
> vif(stanModel1)
GRE.Score      TOEFL.Score      SOP      LOR      CGPA      Research
1.694450      1.999146      1.939850      1.547411      1.500792      1.142692
University.Rating_2 University.Rating_3 University.Rating_4 University.Rating_5
1.792788      2.075832      1.344903      1.027013

> vif(stanModel2)
TOEFL.Score      SOP      LOR      CGPA University.Rating_2
1.486735      1.814845      1.488111      1.518280      1.015000

> vif(stanModel3)
TOEFL.Score      SOP      LOR      CGPA
1.534587      1.854124      1.526590      1.580393
```

Untuk menguji multikolinearitas, digunakan *function vif()*. Suatu variabel dikatakan memiliki multikolinearitas apabila nilai $vif > 10$. Dari hasil di atas, diketahui bahwa tidak ada nilai $VIF > 10$. Oleh karena itu, bisa disimpulkan bahwa **tidak ada multikolinearitas** pada seluruh model regresi logistik ini.

Membandingkan Performa Model

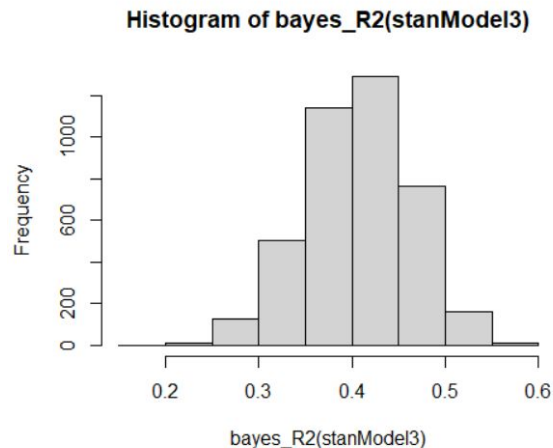
Tabel. Hasil Evaluasi Performa Model Regresi Logistik

<i>Model</i>	<i>Accuracy</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Precision</i>
<i>base_model3 (Frequentist)</i>	0.930	0.9836	0.3714	0.9423
<i>stanModel3 (Bayesian)</i>	0.935	0.9890	0.3714	0.9426

Jadi, model Bayesian memiliki performa yang lebih baik daripada model Frequentist.

Nilai R-Squared pada *stanModel3*

Pendekatan Bayesian tidak memiliki point estimate, namun yang dimiliki adalah kumpulan simulasi posterior. Oleh karena itu, R-Square pada model bayes dapat digambarkan berupa histogram seperti gambar di samping. Untuk menyimpulkan seberapa besar variabel independen dapat menjelaskan keragaman variabel dependen, maka digunakan nilai median dan rata-ratanya. Dari hasil perhitungan, dapat disimpulkan bahwa variabel independen dalam model ini hanya mampu menjelaskan 40% keragaman variabel dependen.



```
> median(bayes_R2(stanModel3))  
[1] 0.4087473  
> mean(bayes_R2(stanModel3))  
[1] 0.4066015
```

Kesimpulan

- Performa model Bayesian lebih unggul dari pada model Frequentist, walaupun perbedaan nilai performa antara keduanya tidaklah jauh signifikan. Kemiripan performa model *Frequentist* dan Bayesian bisa disebabkan jika kedua model tersebut berhasil memenuhi seluruh asumsi regresi logistik seperti multikolinearitas.
- Terdapat empat variabel yang berpengaruh signifikan terhadap penerimaan mahasiswa pascasarjana baru, yaitu *TOEFL Score*, *Statement Of Purpose*, *Letter of Recommendation Strength* dan *Undergraduate GPA*. Namun sayangnya, prediksi nilai variabel dependen ini sebenarnya kurang kuat apabila hanya bergantung pada keempat variabel independen ini mengingat hanya 40% dari keragaman variabel dependen yang dapat dijelaskan. Jadi, calon mahasiswa pascasarjana baru sebaiknya tidak hanya berfokus pada empat aspek penilaian tersebut.
- Di sisi lain, penulis juga merekomendasikan penelitian selanjutnya untuk mengatasi *imbalance data* agar dapat meningkatkan nilai *specificity* model. Selain itu, pemilihan distribusi prior juga perlu dipertimbangkan kembali dengan harapan model Bayes akan menghasilkan performa yang lebih optimal daripada penelitian ini.

Terima Kasih