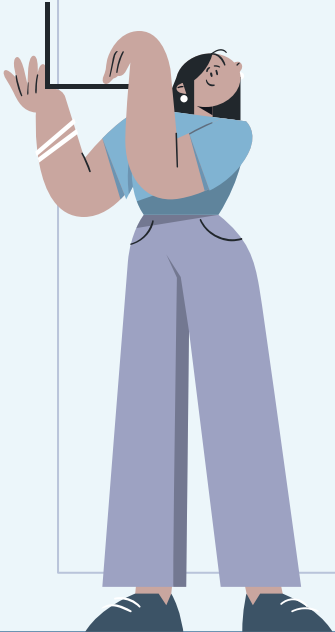
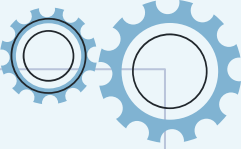


**Analisis Hubungan antara GRE Score, TOEFL Score, University Rating, Statement of Purpose, Letter of Recommendation, Undergraduate GPA, dan Research Experience terhadap Kemungkinan Mahasiswa Sarjana Diterima Program Magister**

Project 2 - Model Linier PTA 2022/2023





# Anggota Kelompok

**01**

**Safira Raissa Rahmi**

2006568891

**04**

**Daniel Lie**

2206820610

**02**

**Muhammad Daffa**

2006568626

**05**

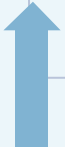
**Caturangga Argya  
Nuriski**

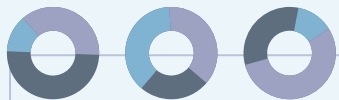
2206813800

**03**

**Yoel Klamedia Sinurat**

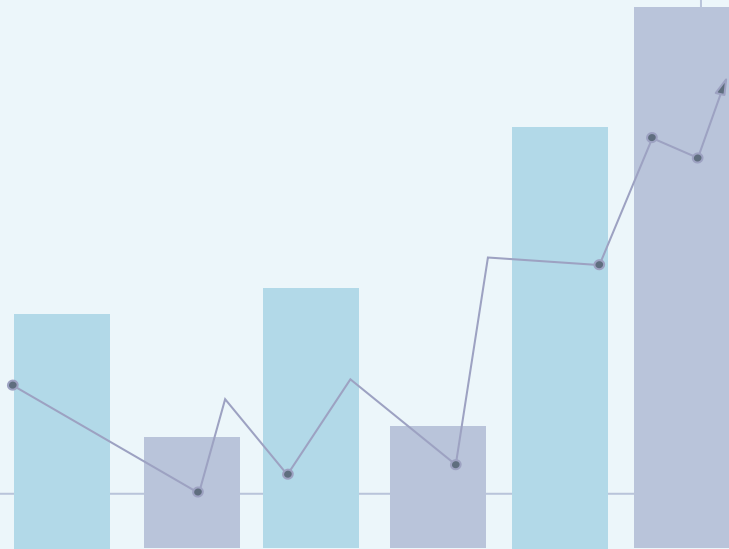
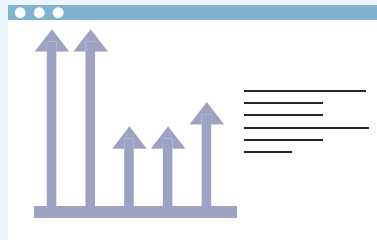
2206051802





01

# Pendahuluan

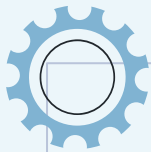


# Apa masalah yang akan dibahas?

Bagaimana **memprediksi kemungkinan mahasiswa sarjana diterima di program magister** berdasarkan GRE score, TOEFL score, *university rating*, *statement of purpose* (SOP), *letter of recommendation* (LOR), *undergraduate GPA* (CGPA), dan *research experience*?

**Variabel apa saja yang berpengaruh signifikan** terhadap besar kemungkinan mahasiswa sarjana diterima di program magister?





# Data yang digunakan

Penelitian oleh **Mohan S. Acharya, dkk**

Sumber data:

<https://www.kaggle.com/datasets/mohansacharya/graduate-admissions/data>

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	337	118	4	4.5	4.5	9.65	1	0.92
1	324	107	4	4.0	4.5	8.87	1	0.76
2	316	104	3	3.0	3.5	8.00	1	0.72
3	322	110	3	3.5	2.5	8.67	1	0.80
4	314	103	2	2.0	3.0	8.21	0	0.65
...	...	...	...	...	...	...	...	...
495	332	108	5	4.5	4.0	9.02	1	0.87
496	337	117	5	5.0	5.0	9.87	1	0.96
497	330	120	5	4.5	5.0	9.56	1	0.93
498	312	103	4	4.0	5.0	8.43	0	0.73
499	327	113	4	4.5	4.5	9.04	0	0.84

gambaran dari beberapa data  
yang ditampilkan



# Data yang digunakan



## Ukuran data

500 data



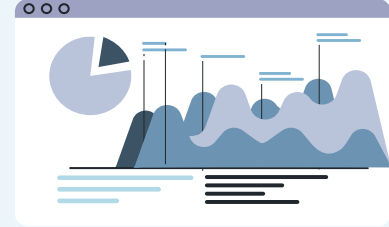
## Kolom data

8 kolom



## Skala/tipe data

- GRE Score : Skala Rasio
- TOEFL Score : Skala Rasio
- University Rating : Skala Rasio
- Statement of Purpose (SOP) : Skala Rasio
- Letter of Recommendation (LOR) : Skala Rasio
- Research Experience : Skala Ordinal
- Chance of Admittance : Skala Rasio





# Penjelasan masing-masing variabel

**GRE Score** (0-340): Skor yang diperoleh mahasiswa pada tes yang digunakan untuk menilai kesiapan mahasiswa dalam melanjutkan pendidikan ke program magister.

**TOEFL Score** (0-120): Skor yang diperoleh mahasiswa dalam tes untuk menilai kemampuan berbahasa Inggris TOEFL.

**University Rating** (0-5): Akreditasi universitas yang didasarkan pada berbagai faktor.

**Statement of Purpose** (0-5): Nilai evaluasi terhadap esai yang ditulis oleh mahasiswa berisi penjelasan alasan ingin melanjutkan studi ke program magister.

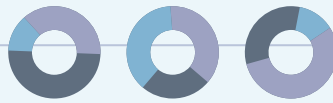
**Letter of Recommendation** (0-5): Nilai evaluasi terhadap surat rekomendasi/referensi oleh pihak lain yang mengenal mahasiswa dengan baik mengenai wawasan, kinerja, profesionalisme, dan karakter. Biasanya ditulis oleh dosen, profesor, atau atasan.

**Undergraduate GPA** (0-10): Rata-rata nilai akademis mahasiswa yang diperoleh selama studi sarjana.

**Research Experience** (0 atau 1): Pengalaman penelitian yang dimiliki mahasiswa. Nilai 0 berarti mahasiswa tidak memiliki pengalaman melakukan penelitian, sedangkan nilai 1 berarti mahasiswa memiliki pengalaman melakukan penelitian.

**Chance of Admittance** (0-1): Nilai probabilitas dari mahasiswa diterima di program magister



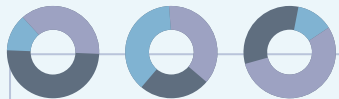


“Tujuan utama dari penelitian ini adalah untuk **menggambarkan hubungan antara variabel-variabel** seperti GRE score, TOEFL score, university rating, statement of purpose (SOP), letter of recommendation (LOR), dan undergraduate GPA (CGPA), serta untuk **memprediksi kemungkinan seorang mahasiswa sarjana diterima program magister** berdasarkan kontribusi masing-masing variabelnya.”

### — Tujuan Penelitian



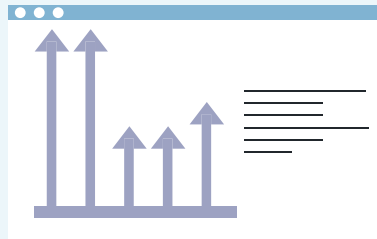




02

# *Preprocessing &* **Analisis Deskriptif**

Notebook: [Project 1\\_Molin\\_Kelompok D](#)



# Pemeriksaan *Missing Values*

```
df.isnull().sum()
```

Serial No.	0
GRE Score	0
TOEFL Score	0
University Rating	0
SOP	0
LOR	0
CGPA	0
Research	0
Chance of Admit	0
dtype: int64	

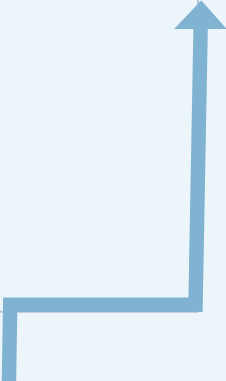
# *Encoding Data* Kategorik

```
df['Research'].value_counts()
```

```
1      280
```

```
0      220
```

```
Name: Research, dtype: int64
```



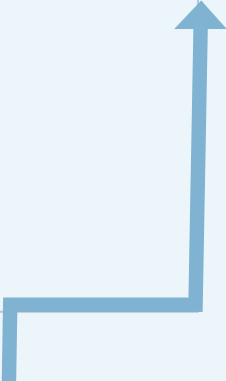
# *Encoding Data* Kategorik

```
df['Research'].value_counts()
```

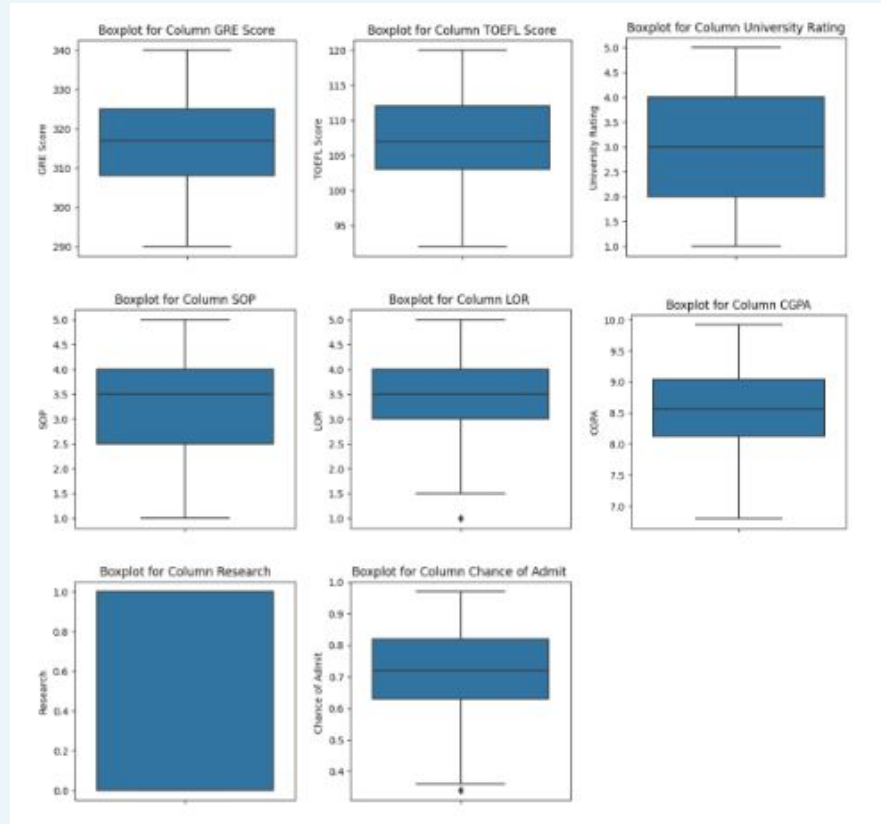
```
1      280
```

```
0      220
```

```
Name: Research, dtype: int64
```



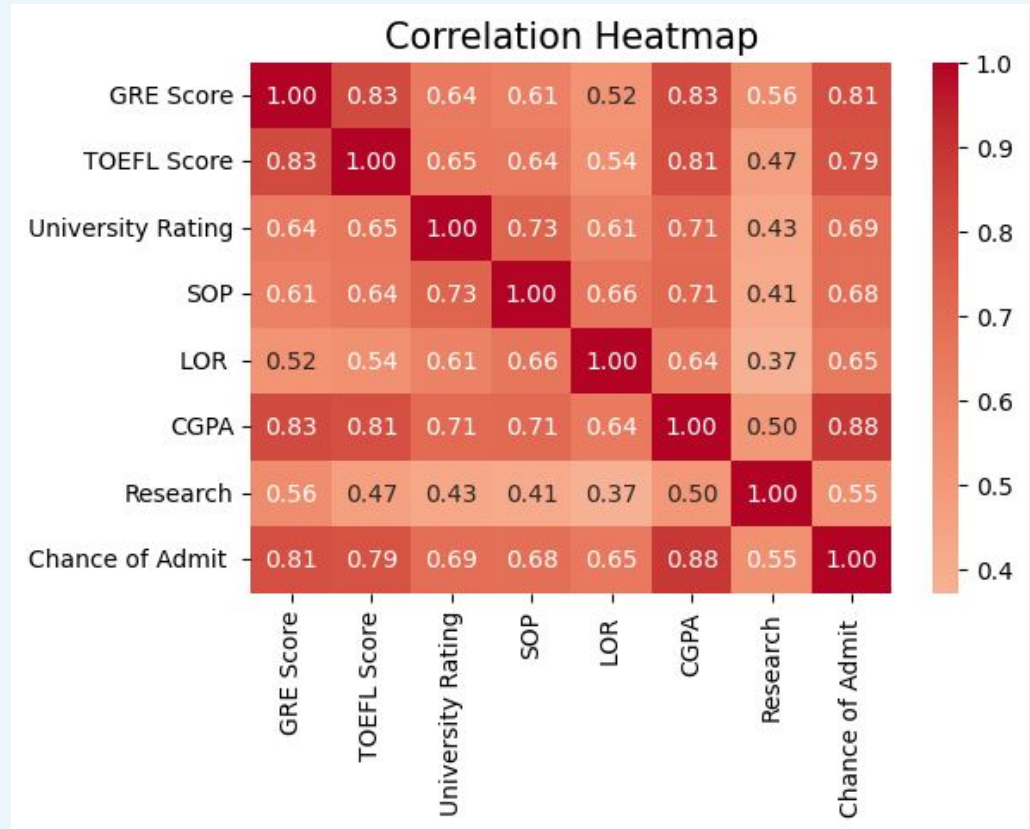
# Pemeriksaan Outliers



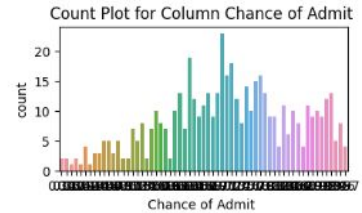
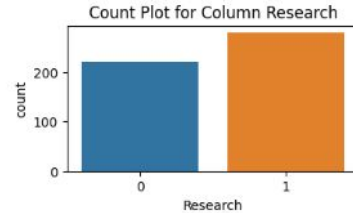
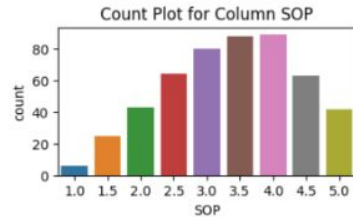
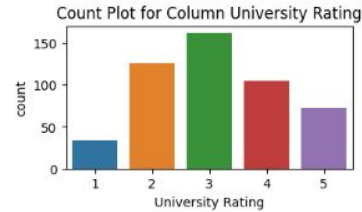
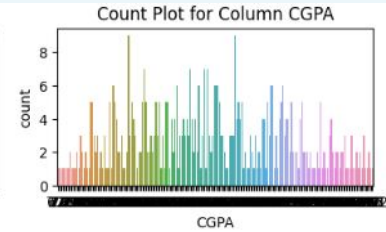
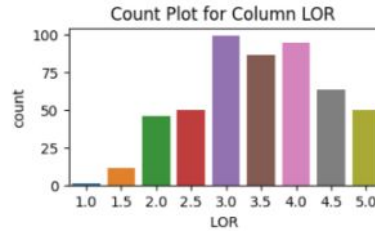
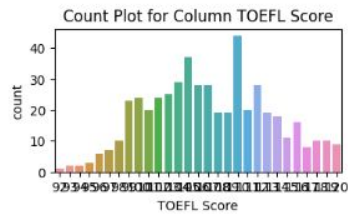
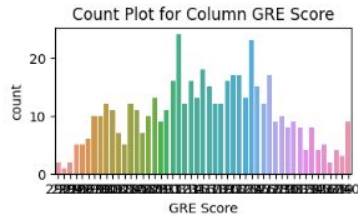
# Analisis Dataset

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	316.472000	107.192000	3.114000	3.374000	3.48400	8.576440	0.560000	0.72174
std	11.295148	6.081868	1.143512	0.991004	0.92545	0.604813	0.496884	0.14114
min	290.000000	92.000000	1.000000	1.000000	1.00000	6.800000	0.000000	0.34000
25%	308.000000	103.000000	2.000000	2.500000	3.00000	8.127500	0.000000	0.63000
50%	317.000000	107.000000	3.000000	3.500000	3.50000	8.560000	1.000000	0.72000
75%	325.000000	112.000000	4.000000	4.000000	4.00000	9.040000	1.000000	0.82000
max	340.000000	120.000000	5.000000	5.000000	5.00000	9.920000	1.000000	0.97000

# Koefisien Korelasi Antar Variabel

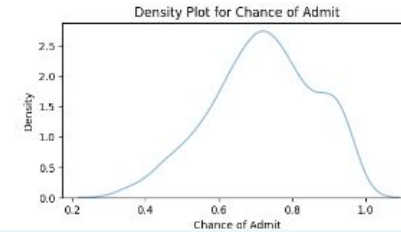
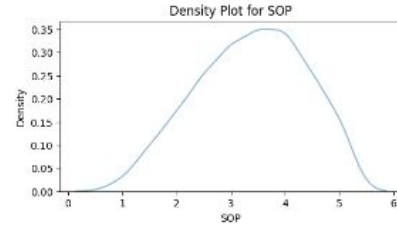
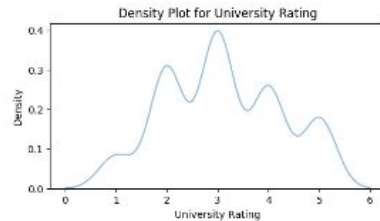
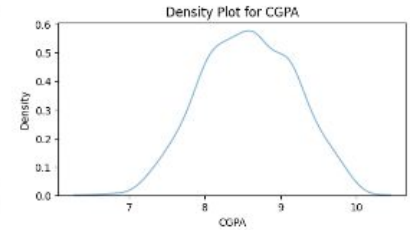
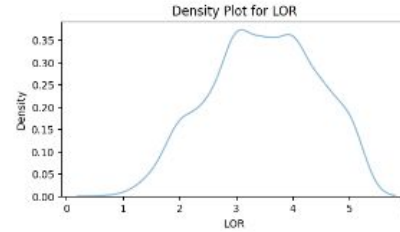
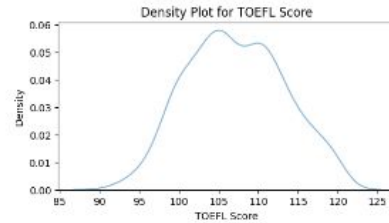
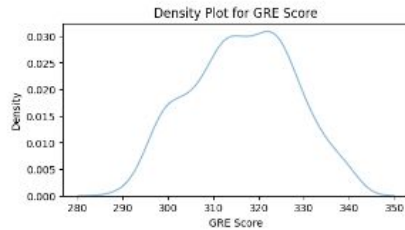


# Count Plot Masing-masing Variabel

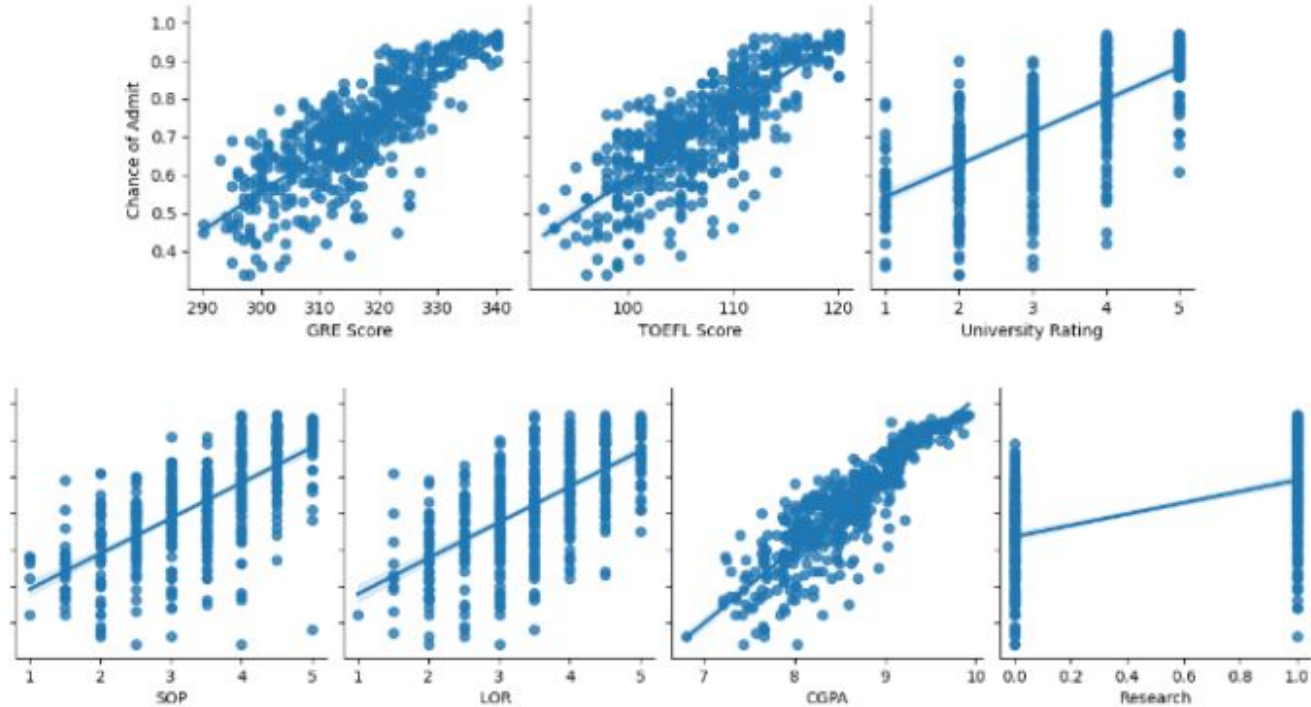




# Density Plot Masing-masing Variabel



# Linearity Plot Masing-masing Variabel



# Feature Scaling - Standardisasi

$$X' = \frac{X - \mu}{\sigma}$$

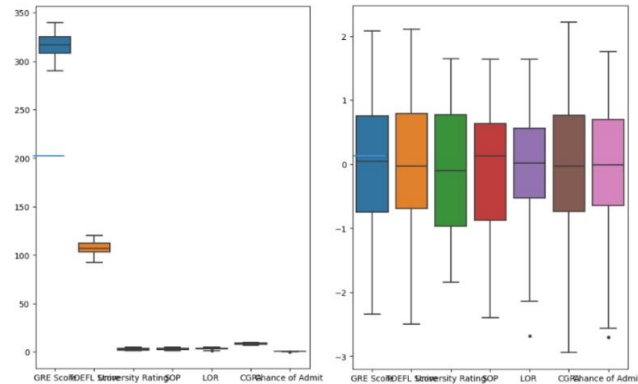
$X'$ : Data yang terstandarisasi

$X$ : Data awal

$\mu$ : Rata-rata nilai variabel

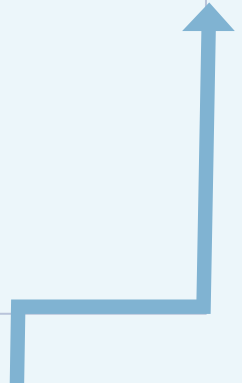
$\sigma$ : Standar deviasi nilai variabel

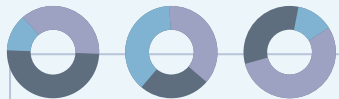
Sehingga persebaran data dalam grafik *box plot* sebelum dan sesudah standarisasi adalah sebagai berikut:



# Hipotesis

Hipotesis yang akan diteliti adalah GRE score, TOEFL score, university rating, statement of purpose (SOP), letter of recommendation (LOR), undergraduate GPA (CGPA), dan Research Experience berpengaruh terhadap kemungkinan mahasiswa sarjana diterima program magister (Chance of admittance).

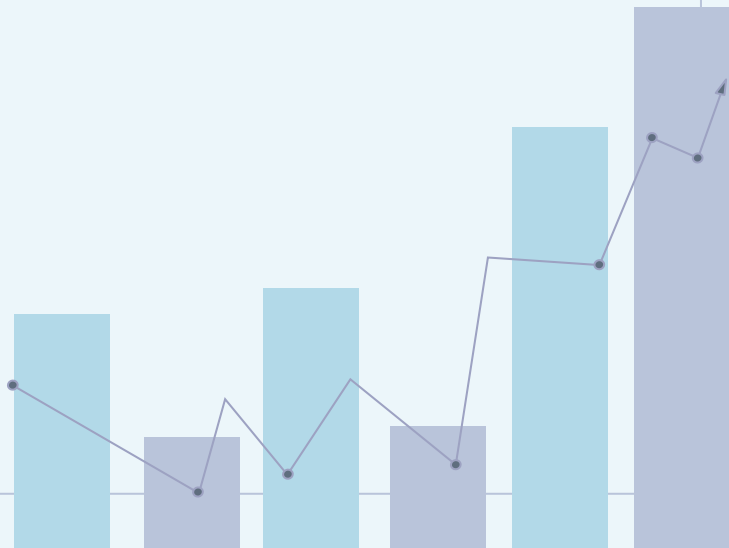
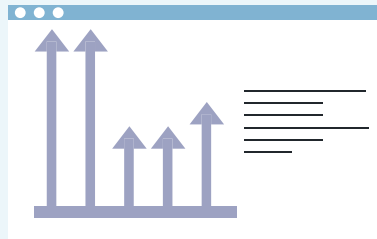




03

# Pemodelan

Notebook: [Project 1\\_Molin\\_Kelompok D](#)





$\alpha = 0.05.$

## Asumsi

### Variabel Signifikan

Menerima uji parsial untuk masing-masing variabel nilai  $p > \alpha = 0.05$  dan memiliki nilai  $VIF < 10$ .



# Model 1

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \varepsilon$$

$y$  := *Chance of admittance*

$x_1$  := *GRE score*

$x_2$  := *TOEFL score*

$x_3$  := *University Rating*

$x_4$  := *Statement of Purpose*

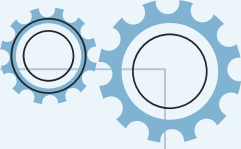
$x_5$  := *Letter of Recommendation*

$x_6$  := *Undergraduate GPA*

$x_7$  := *Research Experience*

$\varepsilon$  := *Error/Galat*





OLS Regression Results						
=====						
Dep. Variable:	Chance of Admit	R-squared:	0.822			
Model:	OLS	Adj. R-squared:	0.819			
Method:	Least Squares	F-statistic:	324.4			
Date:	Tue, 31 Oct 2023	Prob (F-statistic):	8.21e-180			
Time:	13:33:42	Log-Likelihood:	701.38			
No. Observations:	500	AIC:	-1387.			
Df Residuals:	492	BIC:	-1353.			
Df Model:	7					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	-1.2757	0.104	-12.232	0.000	-1.481	-1.071
GRE Score	0.0019	0.001	3.700	0.000	0.001	0.003
TOEFL Score	0.0028	0.001	3.184	0.002	0.001	0.004
University Rating	0.0059	0.004	1.563	0.119	-0.002	0.013
SOP	0.0016	0.005	0.348	0.728	-0.007	0.011
LOR	0.0169	0.004	4.074	0.000	0.009	0.025
CGPA	0.1184	0.010	12.198	0.000	0.099	0.137
Research	0.0243	0.007	3.680	0.000	0.011	0.037
=====						
Omnibus:	112.770	Durbin-Watson:	0.796			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	262.104			
Skew:	-1.160	Prob(JB):	1.22e-57			
Kurtosis:	5.684	Cond. No.	1.30e+04			
=====						

	Feature	VIF
0	GRE Score	4.225203
1	TOEFL Score	3.898373
2	University Rating	2.615772
3	SOP	2.834454
4	LOR	2.029486
5	CGPA	4.776161
6	Research	1.170262

$R^2$   
82.2%



Drop Variabel



- University Rating memiliki p-value = 0.119 >  $\alpha = 0.05$
- SOP memiliki p-value = 0.728 >  $\alpha = 0.05$ .





## Model 2

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \varepsilon$$

$y$  := *Chance of admittance*

$x_1$  := *GRE score*

$x_2$  := *TOEFL score*

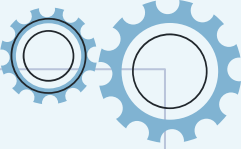
$x_5$  := *Letter of Recommendation*

$x_6$  := *Undergraduate GPA*

$x_7$  := *Research Experience*

$\varepsilon$  := *Error/Galat*





### OLS Regression Results

```
=====
Dep. Variable:    Chance of Admit    R-squared:            0.821
Model:            OLS                Adj. R-squared:       0.819
Method:           Least Squares      F-statistic:         452.1
Date:            Tue, 31 Oct 2023    Prob (F-statistic):   9.97e-182
Time:            14:14:12            Log-Likelihood:       699.65
No. Observations: 500                AIC:                 -1387.
Df Residuals:     494                BIC:                 -1362.
Df Model:         5
Covariance Type:  nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	-1.3357	0.099	-13.482	0.000	-1.530	-1.141
GRE Score	0.0019	0.001	3.760	0.000	0.001	0.003
TOEFL Score	0.0030	0.001	3.501	0.001	0.001	0.005
LOR	0.0193	0.004	5.092	0.000	0.012	0.027
CGPA	0.1230	0.009	13.221	0.000	0.105	0.141
Research	0.0252	0.007	3.814	0.000	0.012	0.038

```
=====
Omnibus:            109.027    Durbin-Watson:           0.800
Prob(Omnibus):      0.000     Jarque-Bera (JB):        248.874
Skew:               -1.130    Prob(JB):                9.07e-55
Kurtosis:           5.615     Cond. No.                1.23e+04
=====
```

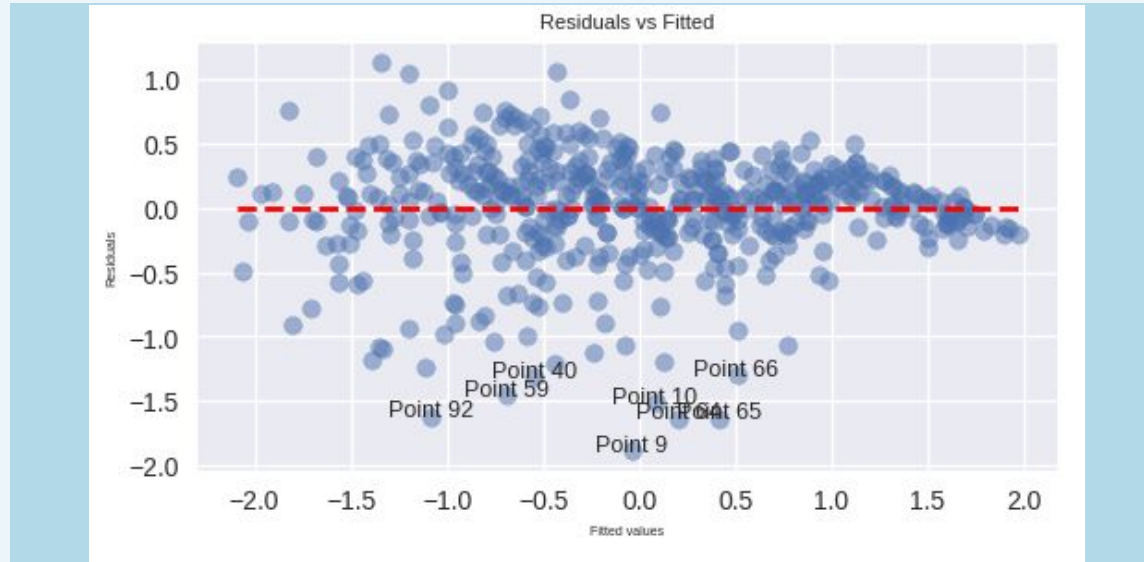
	Feature	VIF
0	GRE Score	4.209225
1	TOEFL Score	3.795962
2	LOR	1.696170
3	CGPA	4.370664
4	Research	1.168252

$R^2$

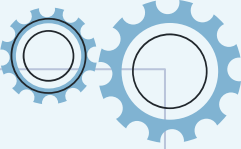
82.1%



# Dropping Outlier



- **Indeks *Outlier*** : 9, 10, 40, 59, 64, 65, 66, dan 92
- **Value *Outlier*** :  
-1.8843977798251839, -1.513368720554275, -1.311025594378516, -1.4525315842865516,  
-1.634919495326382, -1.6354360693732906, -1.298544056161342, dan -1.6186940339113074.



### OLS Regression Results

```
=====
Dep. Variable:      Chance of Admit    R-squared:          0.853
Model:              OLS                Adj. R-squared:     0.852
Method:             Least Squares      F-statistic:        564.8
Date:               Mon, 18 Dec 2023   Prob (F-statistic): 7.88e-200
Time:               17:51:55           Log-Likelihood:     -216.71
No. Observations:   492               AIC:                445.4
Df Residuals:       486               BIC:                470.6
Df Model:           5
Covariance Type:    nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	-0.0444	0.029	-1.512	0.131	-0.102	0.013
GRE Score	0.1788	0.036	4.950	0.000	0.108	0.250
TOEFL Score	0.1371	0.033	4.122	0.000	0.072	0.203
LOR	0.1157	0.022	5.207	0.000	0.072	0.159
CGPA	0.5158	0.036	14.475	0.000	0.446	0.586
Research	0.1249	0.042	2.953	0.003	0.042	0.208

```
=====
Omnibus:            46.541    Durbin-Watson:      0.935
Prob(Omnibus):      0.000    Jarque-Bera (JB):    68.065
Skew:               -0.670    Prob(JB):            1.66e-15
Kurtosis:           4.235    Cond. No.            5.47
=====
```

	Features	VIF
0	const	2.964192
1	GRE Score	4.494472
2	TOEFL Score	3.822170
3	LOR	1.709610
4	CGPA	4.403566
5	Research	1.511538

$R^2$

85.3%



**BEST Model**





# Model Ketiga (Terbaik)

$$y = -0.0444 + 0.1788x_1 + 0.1371x_2 + 0.1157x_5 + 0.5158x_6 + 0.1249x_7 + \varepsilon$$

$y$  := *Chance of admittance*

$x_1$  := *GRE score*

$x_2$  := *TOEFL score*

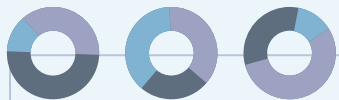
$x_5$  := *Letter of Recommendation*

$x_6$  := *Undergraduate GPA*

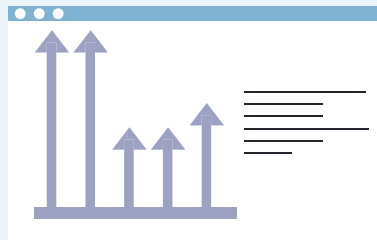
$x_7$  := *Research Experience*

$\varepsilon$  := *Error/Galat*





04



# Pengolahan Data dan Analisis Hasil

Notebook: [Project 1\\_Molin\\_Kelompok D](#)



# Analisis Residual

**Residu dari regresi adalah selisih dari nilai variabel asli dengan hasil prediksi.**

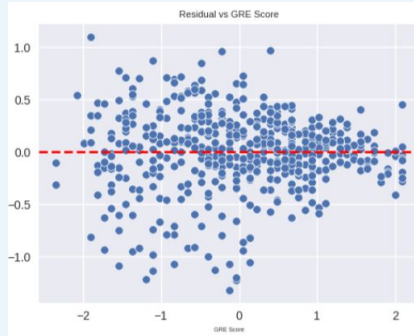
$$\hat{\varepsilon} = y - \hat{y} = y - (\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k)$$

Model regresi yang baik harus memenuhi asumsi:

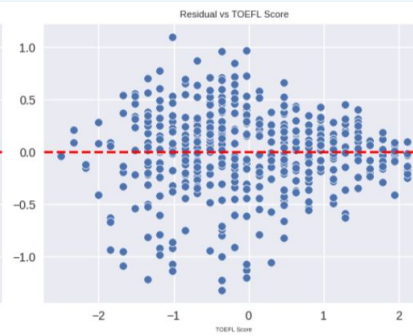
- Rata-rata residual nol
- Homoskedastisitas
- Residu berdistribusi normal
- Independensi

# Residual vs Variabel Independen

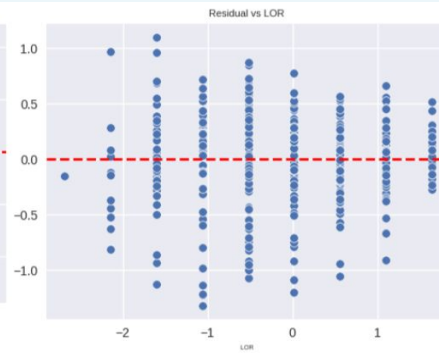
GRE



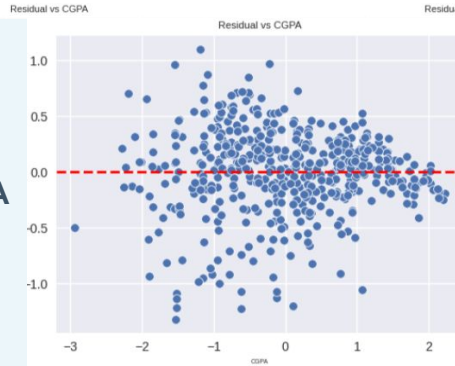
TOEFL



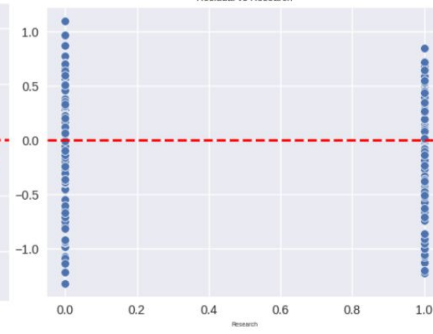
LOR



CGPA



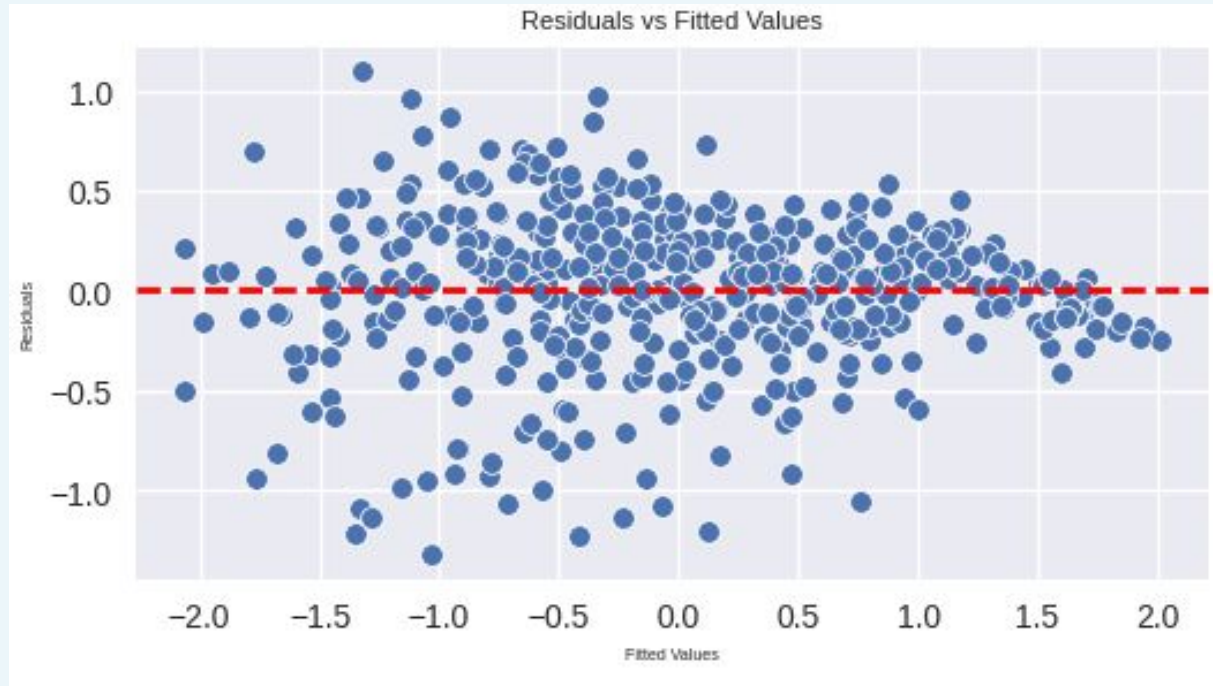
Residual vs Research



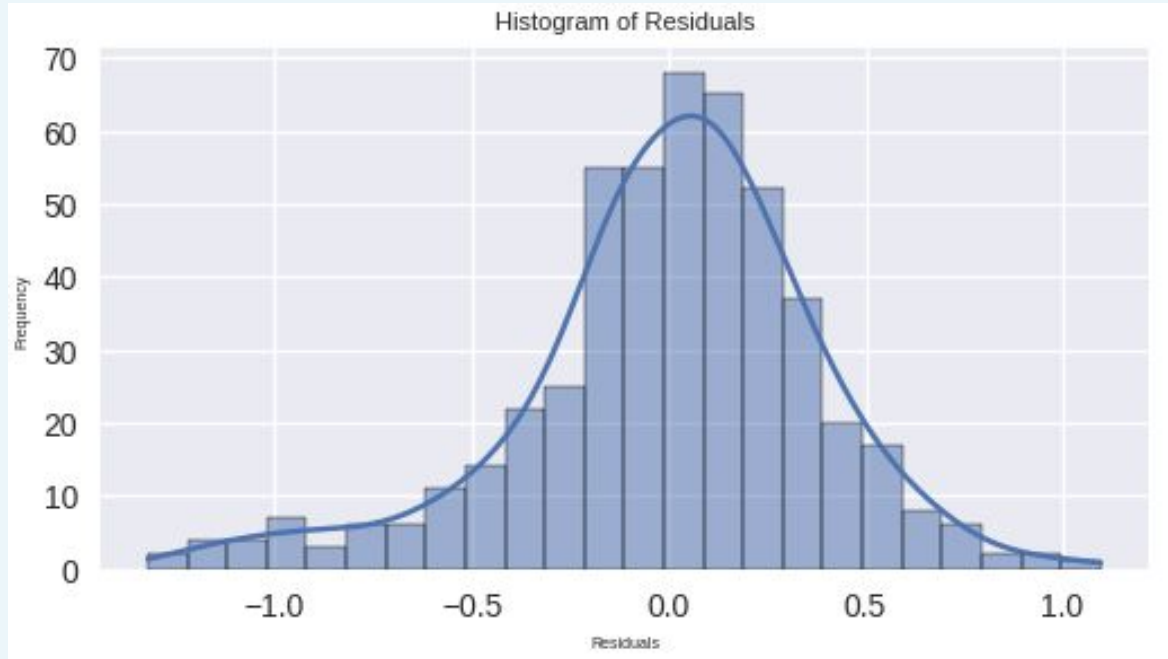
Research



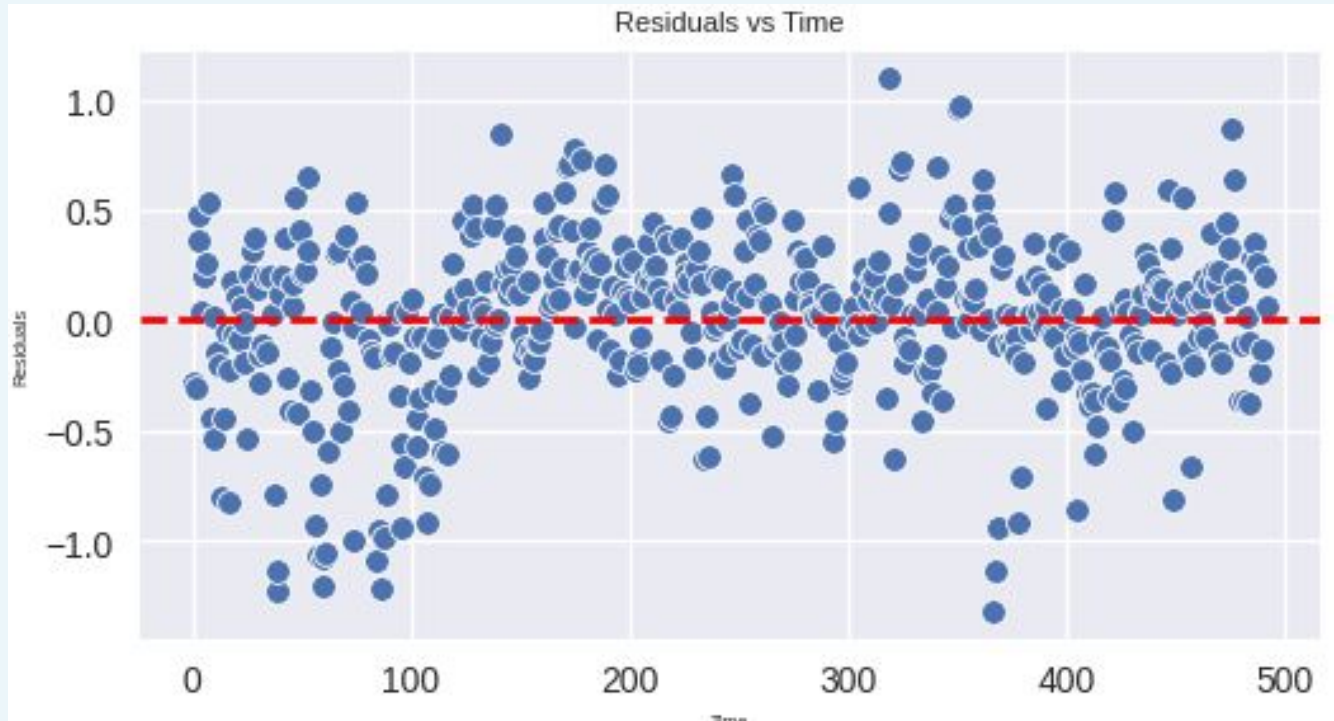
# Residual vs Hasil Prediksi



# Histogram Residual



# Residual vs Time





# Kesimpulan

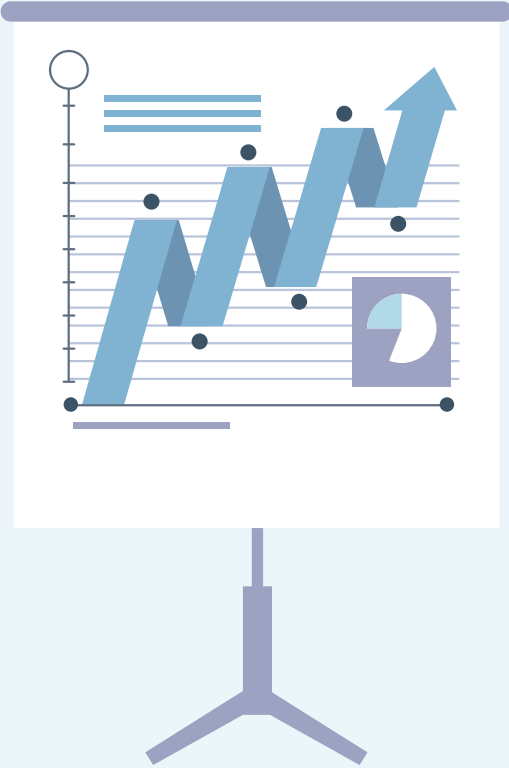
**Model ketiga memiliki  $R^2$  yang baik** yakni 0.853 yang berarti 85.3% variabilitas dari Chance of Admittance dapat diprediksi oleh variabel-variabel yang diusulkan pada model kedua ini. Model ini juga **memiliki global p-value yang baik** sebesar  $7.78 \times 10^{-200} < \alpha = 0.05$ . Selain itu, **model ini telah memenuhi asumsi normalitas, homoskedastisitas, linearitas, dan tidak terdapat multikolinearitas variabel prediktor.**

Dengan demikian, **model kedua yang telah dilakukan *drooping outlier* dapat digunakan untuk memprediksi kemungkinan mahasiswa sarjana diterima di program magister** berdasarkan GRE score, TOEFL score), letter of recommendation (LOR), undergraduate GPA (CGPA), dan research experience dengan baik.



# Daftar pustaka

- Chauhan, N. S. (n.d.). From data pre-processing to optimizing a regression model performance - KDnuggets. KDnuggets. <https://www.kdnuggets.com/2019/07/data-pre-processing-optimizing-regression-model-performance.html>
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (eds.). (2013). An introduction to statistical learning: with applications in R. New York: Springer
- Knief, U., & Forstmeier, W. (2021, May 7). Violating the normality assumption may be the lesser of two evils - behavior research methods. SpringerLink. Retrieved December 28, 2021, from <https://link.springer.com/article/10.3758/s13428-021-01587-5>
- Mendenhall, W., & Sincich, T. (2012b). A second course in statistics: Regression Analysis. Pearson College Division.
- M. S. Acharya, A. Armaan and A. S. Antony, "A Comparison of Regression Models for Prediction of Graduate Admissions," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862140.
- Pan, Y. and Jackson, R.T. (2008) Ethnic Difference in the Relationship between Acute Inflammation and Serum Ferritin in US Adult Males.



# Terima Kasih!

