

Prediksi Performa Mahasiswa Menggunakan Model Regresi Logistik

Nurmalitasari¹⁾, Eko Purwanto²⁾

^{1,2} Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta

¹nurmalitasari@udb.ac.id

²eko_purwanto@udb.ac.id

Abstrak

Prediksi performa mahasiswa merupakan suatu hal yang penting bagi sebuah Perguruan tinggi. Hal tersebut dikarenakan dapat membantu sebuah universitas untuk mencegah maupun mengatasi secara dini mahasiswa yang beresiko gagal dalam perkuliahannya. Penelitian ini bertujuan untuk memprediksi performa mahasiswa Universitas Duta Bangsa Surakarta (UDB). Model yang digunakan dalam penelitian ini adalah model regresi logistik. Regresi logistik adalah metode pemodelan matematika yang digunakan untuk menentukan hubungan antara variabel dependen biner dan satu atau lebih variabel independen. Hasil penelitian menunjukkan bahwa model regresi bisa digunakan untuk memprediksi keberhasilan mahasiswa dengan MAPE sebesar 8%.

Kata Kunci: Performa Mahasiswa, Regresi Logistik, UDB, MAPE

Abstract

Prediction of student performance is an important thing for a university. This is because it can help a university to prevent or treat students who are at risk of failing in their studies early. This study aims to predict student performance at Duta Bangsa Surakarta University (UDB). The model used in this study is a logistic regression model. Logistic regression is a mathematical modelling method used to determine the relationship between a binary dependent variable and one or more independent variables. The results showed that the logistic regression model could be used to predict student performance with MAPE by 8%.

Keyword: Student Performace, Logistic Regression, UDB, MAPE

1. PENDAHULUAN

Dewasa ini ilmu terkait dengan “Prediksi/ Peramalan” berkembang dengan sangat pesat. Ada banyak metode yang bisa digunakan untuk melakukan prediksi diantaranya machine learning (Liang, Li, & Zheng, 2016), data mining (Gaol, 2020) dan lainnya. Dalam *machine learning* ada beberapa model yang digunakan untuk melakukan prediksi, diantaranya adalah Regresi Logistik (van Smeden et al., 2019), Decicion Tree (Chen et al., 2019)(Ma & Zhou, 2018), SPV(Ma & Zhou, 2018), Neural Network, Deep Learning dan lainnya. Regresi logistik adalah metode pemodelan matematika yang digunakan untuk menentukan hubungan antara variabel dependen biner dan satu atau lebih variabel independen. Variabel independen dapat kontinu, diskrit, dan biner, atau salah satu dari variabel ini dapat dicampur kedua-duanya (van Smeden et al., 2019). Berbeda dengan regresi linier, dalam regresi logistic tidak ada syarat variabel dependen dengan variabel independent harus linier. Selain itu asumsi yang dipenuhi dari regresi logistic ini adalah jumlah sampel harus besar.

Keberhasilan mahasiswa dalam sebuah studi merupakan suatu komponen yang penting dalam sebuah universitas. Hal tersebut dikarenakan dapat membantu sebuah universitas untuk mencegah maupun mengatasi secara dini mahasiswa yang beresiko gagal dalam perkuliahannya (Nurmalitasari, Long, and Faizuddin, 2020). Begitu pula dengan Universitas Duta Bangsa Surakarta, prediksi ini sangat penting dilakukan untuk mengatasi secara dini mahasiswa yang beresiko gagal. Keberhasilm mahasiswa dalam studi merupakan point yang penting dalam sebuah universitas, terutama menjadi satu poin

penting dalam aspek akreditasi. Semakin sedikit mahasiswa yang gagal dalam bidang akademik, maka akan semakin bagus pula akreditasi sebuah perguruan tinggi, termasuk Universitas Duta Bangsa Surakarta (UDB).

Ada beberapa faktor yang mempengaruhi performa mahasiswa dalam bidang akademik seperti faktor personal, faktor akademik, faktor ekonomi, faktor sosial dan faktor Institusional (Nurmalitasari, Long, & Mohd Noor, 2021)(Nurmalitasari et al., 2020). Mengingat begitu pentingnya prediksi performa mahasiswa dalam bidang akademik, oleh karena itu dalam penelitian ini akan akan melakukan prediksi performa mahasiswa dalam bidang akademik. Model yang digunakan dalam prediksi adalah regresi logistic. Penelitian ini penting dilakukan, guna mengetahui performa mahasiswa Universitas Duta Bangsa Surakarta, sekaligus sebagai pencegahan secara dini mahasiswa yang gagal dalam kuliahnya.

2. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah metode kualitatif. Model yang digunakan dalam penelitian ini adalah model regresi logistic (Stoltzfus, 2011; Kleinbaum, 2010). Variabel dependent yang digunakan dalam penelitian ini adalah indikator keberhasilan mahasiswa yaitu lulus dan tidak lulus atau dropout. Sedangkan variabel independent yang digunakan dalam penelitian ini adalah IPK, pendapatan orangtua, status pernikahan, status pekerjaan mahasiswa, minat program studi. Populasi dalam penelitian ini adalah mahasiswa lulusan Universitas Duta Bangsa Surakarta (UDB). Sampel data yang digunakan dalam penelitian ini adalah 100 data mahasiswa lulusan tahun 2020. Data diperoleh dengan cara menyebarkan kuisioner secara online kepada 100 mahasiswa lulusan UDB tersebut. Model yang digunakan dalam penelitian ini adalah model regresi logistic.

Tabel 1. Variabel *Independent* dan *Dependent* yang digunakan dalam penelitian.

Variabel	Nilai Konstruk Variabel
Indikator Performa Akademik	0 = Lulus Kuliah
	1 = Tidak Lulus Kuliah
Indeks Prestasi Kumulatif (IPK)	1 = 0.00 - 1.99
	2 = 2.00 - 2.75
	3 = 2.76 - 3.50
	4 = 3.51 - 4.00
Pendapatan Orang Tua	1 = Tidak Berpenghasilan
	2 = Kurang dari Rp 1.000.000
	3 = Rp 1.000.001 - Rp 2.000.000
	4 = Rp 2.000.001 - Rp 5.000.000
	5 = Rp 5.000.001 - Rp 20.000.000
	6 = Lebih Besar dari Rp Rp 20.000.000
Status Pernikahan	1 = Lajang
	2 = Menikah
	3 = Berceraai
Status Pekerjaan Mahasiswa	1 = Kuliah sambil bekerja
	2 = Kuliah tidak sambil bekerja
Minat terhadap Program Study	1 = Sangat menyukai program studi yang diambil
	2 = Menyukai program studi yang diambil
	3 = Cukup menyukai program studi yang diambil
	4 = Tidak menyukai program studi yang diambil
	5 = Sangat tidak menyukai program studi yang diambil

Variabel dependent yang digunakan dalam penelitian ini adalah indicator keberhasilan mahasiswa yaitu lulus dan tidak lulus atau dropout. Sedangkan variabel independent yang digunakan dalam penelitian ini adalah IPK, pendapatan orangtua, status pernikahan, status pekerjaan mahasiswa, minat program studi. Kuisisioner meliputi 6 pertanyaan variabel yang digunakan dalam penelitian seperti yang dipaparkan dalam Table 1. Dalam penelitian ini analisis data menggunakan software SPSS versi 25.

Langkah-langkah penelitiannya adalah sebagai berikut (Brogden A K, Guthmiller M J, 2020; Srimaneekarn, Hayter, Liu, & Tantipoj, 2022):

- Step 1.** Mengitung ringkasan pemrosesan kasus
 Pada tahap ini dilakukan tahap pemrosesan pembersihan data dari missing value.
- Step 2.** Inisialisasi Variabel
 Tahap inisialisasi variabel ini diperlukan guna mengidentifikasi simbol-simbol pada konstruk variabel dependen.
- Step 3.** Melakukan uji chi square dari Hosmer dan Lemeshow
 Uji Hosmer dan Lemeshow merupakan uji kelayakan model regresi logistic. Uji ini menguji hipotesis apakah model dapat menjelaskan data dalam penelitian atau tidak.
- Step 4.** Menentukan persamaan regresi logistik
 Persamaan regresi yang bangun adala regresi logistik ganda, karena terdapat satu variabel dependen dan lima variabel independent. Persamaan regresi logistik ganda adalah sebagai berikut.

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n)}{1 + \exp(\beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n)}$$

- Step 5.** Melakukan prediksi menggunakan model regresi logistik ganda
 Dari persamaan model regresi logistik ganda yang diperoleh pada step 4, digunakan untuk menghitung nilai prediksi indikator keberhasilan mahasiswa.
- Step 6.** Menghitung tingkat keakurasian model
 Penilaian akurasi model prediksi sangat penting karena tingkat akurasi model prediksi akan menentukan kualitas prediksi yang dihasilkan.

3. HASIL DAN PEMBAHASAN

Step 1. Mengitung Ringkasan Pemrosesan Kasus

Dalam tahap ini akan dilakukan penghitungan dari kasus yang diamati, dan pendeteksian apakah seluruh data yang diamati terdapat *missing vale* atau tidak. *Missing value* didefinisikan sebagai nilai atau data yang tidak disimpan (atau tidak ada) untuk beberapa variabel dalam kumpulan data yang diberikan (Rubin & Wiley, 2014). Berdasarkan hasil analisis menggunakan SPSS, diperoleh hasil seperti pada Tabel 2.

Tabel 2. Output Case Processing Summary

Case Processing Summary			
Unweighted Cases	Included in Analysis	N	Percent
	Missing Cases	100	100
		0	0
Selected Cases	Total	100	100
Unselected Cases		0	0
Total		100	100

a. If weight is in effect, see classification table for the total number of cases.

Dari Table 2 dapat dilihat bahwa seluruh kasus atau case (100 sampel) teramati, hal ini dapat disimpulkan bahwa tidak ada data yang *missing value*. Oleh karena itu data dalam penelitian ini sudah memenuhi syarat untuk dianalisis ke proses selanjutnya yaitu prediksi.

Step 2. Inisialisasi Variabel

Dalam penelitian ini performa mahasiswa diukur dari keberhasilan mahasiswa dalam menyelesaikan studinya, yaitu lulus kuliah atau tidak. Tabel 3 menunjukkan hasil proses input data dari variabel dependent yang memiliki dua nilai yaitu 0 untuk lulus kuliah, dan 1 untuk tidak lulus kuliah. Variabel ini bersifat biner (dikotomi), yang hanya memiliki dua kategori saja, yaitu lulus kuliah dan tidak lulus kuliah.

Tabel 3. Output Dependent Variable Encoding

Dependent Variable Encoding	
Original Value	Internal Value
Lulus Kuliah	0
Tidak Lulus Kuliah	1

Proses inisialisasi merupakan suatu proses awal, dimana variabel independent belum dimasukkan kedalam model penelitian. Hal tersebut menunjukkan bahwa model persamaan regresi logistik hanya menggunakan nilai koefisien konstanta saja yang digunakan untuk memprediksi performa mahasiswa dalam bidang akademik.

Tabel 4. Variables in the Equation (step 0)

		Variable in The Equation					
		B	S.E.	Wald	df	Sig.	Exp (B)
Step 0	Constant	0.040	0.200	0.040	1	0.841	1.041

Nilai signifikansi jika hanya koefisien konstanta saja dalam model regresi logistik dilihat dari Table 4 sebesar 0.841 artinya nilai tersebut lebih besar dari nilai alpha 5% ($0.841 > 0.05$), hal ini dapat diinterpretasikan bahwa jika menggunakan persamaan model regresi logistik sederhana atau hanya konstanta saja belum bisa memberikan kesimpulan bahwa model mampu menjelaskan proporsi keberhasilan mahasiswa dalam bidang akademik.

Step 3. Melakukan Uji Chi Square dari Hosmer dan Lemeshow

Dengan memasukkan variabel dependent kedalam model regresi logistik jika dilihat dari Tabel 5, dapat disimpulkan bahwa dengan menambah variabel independent (IPK, pendapatan orangtua, status pernikahan, status pekerjaan mahasiswa, minat program studi) terdapat perubahan nilai -2 Log Likelihood sebesar 113.162.

Tabel 5. Model Summary

Model Summary			
Step	-2 Log Likelihood	Cox & Snell R Square	Nagelkerke R Square
1	113.162	0.425	0.599

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Selain itu, berdasarkan nilai Nagelkerke R Square sebesar 59,9% merupakan nilai koefisien determinasi yang memiliki nilai interpretasi pengaruh variabel independent terhadap variabel dependent atau keberhasilan mahasiswa dalam kuliahnya adalah sebesar 59,9%. Untuk memperkuat model tersebut dilakukan pengujian Hosmer dan Lemeshow. Uji Hosmer dan Lemeshow merupakan uji Goodness of fit berdasarkan value prediksi peluang. Uji ini pertama kali di perkenalkan oleh Hosmer dan Lemeshow dengan prinsip pengelompokkan nilai estimasi rata-rata peluang berkelompok, lalu mengitung nilai tes Hosmer dan Lemeshow. Hipotesis yang digunakan dalam uji Hosmer dan Lemeshow adalah sebagai berikut.

H_0 : Model dapat menjelaskan data penelitian

H_1 : Model tidak dapat menjelaskan data penelitian

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

Dengan n'_k merupakan total subjek ke-k, O_k adalah nilai banyaknya kejadian sukses dalam pengamatan kelompok ke-k dan $\bar{\pi}_k$ merupakan nilai estimasi rata-rata kelopak ke-k. Jika diambil nilai α sebesar 5%, dari Tabel 6 juga didapat nilai Sig. 0.362, nilai Signifikansi lebih besar dari 5%, artinya H_0 diterima. Keadaan tersebut dapat diinterpretasikan bahwa Model dapat menjelaskan data penelitian (*Goodness of fit*)

Tabel 6. Output Uji Hosmer dan Lemeshow

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	8.766	8	0.362

Step 4. Menentukan persamaan regresi logistik

Step terpenting dalam prediksi menggunakan regresi logistik adalah menentukan persamaan regresi logistik yang digunakan untuk melakukan prediksi. Berdasarkan Tabel 7, nilai Sig. dari semua variabel independent (Penghasilan orang tua, IPK, Status pernikahan, status pekerjaan, dan minat program studi) kurang dari nilai alpha ($\alpha = 5\%$). Nilai Signifikansi variabel “Penghasilan Orang Tua” sebesar $0.049 < 5\%$, hal tersebut dapat diinterpretasikan bahwa penghasilan orang tua memiliki pengaruh yang signifikan terhadap performa kinerja mahasiswa dalam bidang akademik dengan nilai koefisien sebesar -0.184. Nilai Signifikansi variabel “IPK” sebesar $0.002 < 5\%$, hal tersebut dapat diinterpretasikan bahwa IPK memiliki pengaruh yang signifikan terhadap performa kinerja mahasiswa dalam bidang akademik dengan nilai koefisien sebesar -1.467. Nilai Signifikansi variabel “Status Pernikahan” sebesar $0.046 < 5\%$, hal tersebut dapat diinterpretasikan bahwa status pernikahan memiliki pengaruh yang signifikan terhadap performa kinerja mahasiswa dalam bidang akademik dengan nilai koefisien sebesar 1.557. Nilai Signifikansi variabel “Status pekerjaan” sebesar $0.026 < 5\%$, hal tersebut dapat diinterpretasikan bahwa status pekerjaan memiliki pengaruh yang signifikan terhadap performa kinerja mahasiswa dalam bidang akademik dengan nilai koefisien sebesar -0.601. Nilai Signifikansi variabel “Minat Terhadap Program study” sebesar $0.003 < 5\%$, hal tersebut dapat diinterpretasikan bahwa status pekerjaan memiliki pengaruh yang signifikan terhadap performa kinerja mahasiswa dalam bidang akademik dengan nilai koefisien sebesar -0.904.

Tabel 7. Output Variabel dalam persamaan

		Variables in the Equations					
		B	S.E.	Wald	df	Sig.	Exp (B)
Step 1	Penghasilan Orang Tua	-0.184	0.270	0.467	1	0.049	0.832
	IPK	-1.467	0.481	9.317	1	0.002	0.231
	Status Pernikahan	1.557	0.780	3.978	1	0.046	4.743
	Status Pekerjaan	-0.601	0.530	1.286	1	0.026	0.548
	Minat Terhadap Program Studi	-0.904	0.307	8.652	1	0.003	0.405
	Constant	7.243	2.473	8.576	1	0.003	1398.149
	a. Variable(s) entered on step 1: Penghasilan Orang tua, IPK, Status pernikahan, status pekerjaan, minat terhadap program study						

Oleh karena itu dapat disimpulkan bahwa persamaan model regresi logistik untuk memprediksi performa mahasiswa adalah:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}$$

$$\pi(x) = \frac{\exp(7.243 - 0.184 PO - 1.467IPK + 1.557 SPn - 0.601 SPk - 0.904 MPS)}{1 + \exp(7.243 - 0.184 PO - 1.467 IPK + 1.557 SPn - 0.601 SPk - 0.904 MPS)} \quad (1)$$

Keterangan:

- PO* : Penghasilan Orangtua
- IPK* : Indeks Prestasi Kumulatif
- SPn* : Status Pernikahan
- SPk* : Status Pekerjaan
- MPS* : Minat Program Studi

Selain itu, dari Tabel 7, kita memperoleh informasi terkait dengan pengaruh variabel terhadap keberhasilan mahasiswa yang diperoleh dari nilai Exp (B), yang masing-masing akan dijelaskan sebagai berikut.

- Nilai Exp (B) pada penghasilan orangtua sebesar 0.832, dapat diinterpretasikan bahwa bagi mahasiswa yang memiliki orangtua yang tidak berpenghasilan akan cenderung memiliki resiko tidak berhasil dalam studinya dibandingkan dengan yang memiliki penghasilan lainnya.
- Nilai Exp (B) pada variabel IPK sebesar 0.231 memiliki interpretasi bahwa mahasiswa yang memiliki IPK rendah akan mengalami resiko gagal dalam kuliahnya sangat besar, dibandingkan dengan mahasiswa yang memiliki IPK tinggi.
- Nilai Exp (B) pada variabel status pernikahan sebesar 4.743, artinya mahasiswa yang memiliki status pernikahan sudah menikah akan beresiko mengalami kegagalan dalam studinya dibandingkan dengan mahasiswa yang belum menikah.
- Nilai Exp (B) pada variabel status pekerjaan sebesar 0.548, artinya mahasiswa yang kuliah sambil bekerja akan lebih beresiko gagal studinya dibandingkan dengan mahasiswa yang kuliah tidak sambil bekerja.
- Nilai Exp (B) pada variabel minat terhadap program studi sebesar 0,405, artinya mahasiswa yang sangat menyukai program studinya dapat akan berpeluang berhasil dalam studinya, dibandingkan mahasiswa yang tidak menyukai programstudi yang diambil.

Step 5. Melakukan Prediksi Menggunakan Regresi Logistik

Persamaan (1) digunakan untuk melakukan prediksi performa akademik mahasiswa, dan diperoleh hasil seperti pada Tabel 8. Nilai probabilitas prediksi yang kurang dari 0.5 dianggap lulus, sedangkan nilai prediksi yang nilainya lebih besar sama dengan 0.5 dianggap tidak lulus kuliah. Penerapan pembulatan probabilitas dari prediksi ini disesuaikan sama nilai konstruk variabel.

Tabel 8. Prediksi Performa Akademik (PA)

No	PO	IPK	SPn	SPk	MPS	PA	Probabilitas Prediksi	Hasil Prediksi
1	3	2	1	1	3	Lulus	0,08809020	Lulus
2	3	3	1	2	4	Lulus	0,27468123	Lulus
3	3	3	2	1	4	Lulus	0,76620409	Tidak Lulus
4	3	3	2	1	4	Lulus	0,76620409	Tidak Lulus
5	2	3	1	1	4	Lulus	0,4536336	Lulus
6	4	3	1	1	4	Lulus	0,36493688	Lulus
7	4	4	1	1	5	Lulus	0,05093147	Lulus
8	4	4	1	1	4	Lulus	0,11701563	Lulus
9	4	4	1	1	5	Lulus	0,05093147	Lulus
10	5	3	1	1	3	Lulus	0,05414049	Lulus
11	4	3	1	1	4	Lulus	0,36493688	Lulus
12	2	4	1	1	4	Lulus	0,1607039	Lulus
13	3	3	1	2	3	Lulus	0,48325626	Lulus
14	4	4	1	2	4	Lulus	0,06773589	Lulus
.
.
.
96	2	3	1	1	4	Tidak Lulus	0,4536336	Lulus
97	4	3	1	1	4	Tidak Lulus	0,56493688	Tidak Lulus
98	4	3	2	1	4	Tidak Lulus	0,73164801	Tidak Lulus
99	3	3	1	1	3	Tidak Lulus	0,63041556	Tidak Lulus
100	5	3	1	1	4	Tidak Lulus	0,52344164	Tidak Lulus

Step 6. Menghitung Tingkat Keakurasian Model

Penilaian akurasi model prediksi sangat penting karena tingkat akurasi model prediksi akan menentukan kualitas prediksi yang dihasilkan. Tingkat keakurasian model dalam penelitian ini menggunakan MAPE atau *Mean Absolute Percentage Error*, dengan formula sebagai berikut (Michael Gilliland, Len Tashman, 2019)(Montgomery, Jennings, & Kulahci, 2015).

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

dengan A_t nilai *actual* performa akademik, dan F_t adalah nilai prediksi. Berdasarkan Tabel 8, diperoleh nilai MAPE sebesar 18%.

$$MAPE = \frac{100\%}{100} \sum_{t=1}^{100} \left| \frac{A_t - F_t}{A_t} \right| = 8\%$$

Hal tersebut bisa diartikan bahwa jika model regresi yang dihasilkan dari step 5 digunakan untuk melakukan prediksi performa akademik mahasiswa, maka tingkat kesalahan prediksi sebesar 8%.

4. KESIMPULAN

Prediksi performa mahasiswa merupakan suatu hal yang penting bagi sebuah Perguruan tinggi. Hal tersebut dikarenakan dapat membantu sebuah universitas untuk mencegah maupun mengatasi secara dini mahasiswa yang beresiko gagal dalam

perkuliahnya. Hasil penelitian menunjukkan bahwa model regresi bisa digunakan untuk memprediksi keberhasilan mahasiswa dengan MAPE sebesar 8%.

5. REFERENSI

- Brogden A K, Guthmiller M J, T. C. E. (2020). Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-. *Ann Oncol*, (January), 2–5.
- Chen, J., Feng, J., Sun, X., Wu, N., Yang, Z., & Chen, S. (2019). MOOC Dropout Prediction Using a Hybrid Algorithm Based on Decision Tree and Extreme Learning Machine. *Mathematical Problems in Engineering*, 2019. <https://doi.org/10.1155/2019/8404653>
- Gaol, N. Y. L. (2020). Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4.5. *Jurnal Informasi & Teknologi*, 23–29. <https://doi.org/10.37034/jidt.v2i1.22>
- Kleinbaum, D. G. (2010). *Logistic Regression*. Springer. https://doi.org/10.1007/978-1-4757-4108-7_6
- Liang, J., Li, C., & Zheng, L. (2016). Machine learning application in MOOCs: Dropout prediction. *ICCSE 2016 - 11th International Conference on Computer Science and Education*, (Iccse), 52–57. <https://doi.org/10.1109/ICCSE.2016.7581554>
- Ma, X., & Zhou, Z. (2018). Student pass rates prediction using optimized support vector machine and decision tree. *2018 IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC 2018, 2018-Janua*, 209–215. <https://doi.org/10.1109/CCWC.2018.8301756>
- Michael Gilliland, Len Tashman, U. S. (2019). Business Forecasting: Practical Problems and Solutions. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). Introduction Time Series Analysis and Forecasting, 671.
- Nurmalitasari, Long, Z. A., & Mohd Noor, M. F. (2021). Reduction of Data Dimensions in the PLA Process. *Proceedings of the 2021 15th International Conference on Ubiquitous Information Management and Communication, IMCOM 2021*. <https://doi.org/10.1109/IMCOM51814.2021.9377391>
- Nurmalitasari, N., Long, Z. A., Faizuddin, M., & Noor, M. (2020). Data Preparation in Predictive Learning Analytics (PLA) for Student Dropout. *International Journal of Innovative Technology and Exploring Engineering*, 9(3S), 116–120. <https://doi.org/10.35940/ijitee.c1027.0193s20>
- Rubin, D. B., & Wiley, A. J. (2014). *Statistical Analysis with Missing Data Second Edition*.
- Srimaneekarn, N., Hayter, A., Liu, W., & Tantipoj, C. (2022). Binary Response Analysis Using Logistic Regression in Dentistry. *International Journal of Dentistry*, 2022(1), 1–7. <https://doi.org/10.1155/2022/5358602>
- Stoltzfus, J. C. (2011). Logistic regression: A brief primer. *Academic Emergency Medicine*, 18(10), 1099–1104. <https://doi.org/10.1111/j.1553-2712.2011.01185.x>
- Van Smeden, M., Moons, K. G. M., de Groot, J. A. H., Collins, G. S., Altman, D. G., Eijkemans, M. J. C., & Reitsma, J. B. (2019). Sample size for binary logistic prediction models: Beyond events per variable criteria. *Statistical Methods in Medical Research*, 28(8), 2455–2474. <https://doi.org/10.1177/0962280218784726>