

REGRESI LOGISTIK BINER UNTUK MENGLASIFIKASIKAN CARA BELAJAR MAHASISWA MENURUT SUMBER BELAJARNYA

BINARY LOGISTIC REGRESSION TO CLASSIFY HOW STUDENTS LEARN BASED ON THEIR LEARNING SOURCES

Nanda Arista Rizki, Mutiara Mumtaza*, Carolina Fadia Dewi, Dhira Syahlafandi
Universitas Mulawarman, Jl. Kuaro, Gn. Kelua, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur

*Penulis untuk korespondensi, e-mail: aramumtz@gmail.com

Received [23-09-2023] Revised [03-04-2024] Accepted [01-06-2024]

ABSTRAK

Penelitian ini bertujuan untuk membuat model regresi logistik biner yang dapat mengklasifikasikan cara belajar mahasiswa berdasarkan sumber belajarnya. Data diambil dari 111 mahasiswa program studi pendidikan matematika Universitas Mulawarman. Sumber belajar yang menjadi variabel prediktor merupakan pilihan ganda majemuk. Hasil penelitian menunjukkan bahwa mahasiswa yang menjadikan YouTube sebagai sumber belajarnya berpeluang untuk belajar secara mandiri sebesar 2,232 kali lebih besar dari pada belajar matematika berkelompok. Sementara mahasiswa yang menjadikan buku cetak sebagai sumber belajarnya berpeluang untuk belajar secara kelompok sebesar 1,968 kali lebih besar dari pada belajar matematika mandiri. Nilai skor F1 tertinggi terletak pada pembagian data 90:10 yaitu sebesar 0,643. Skor AUC untuk model regresi logistik biner yang digunakan adalah sebesar 0,611.

Kata kunci: Cara belajar mahasiswa; regresi logistik biner; skor AUC; sumber belajar

ABSTRACT

This study aimed to create a binary logistic regression model to classify student learning styles based on their learning resources. This research took data from 111 students of the mathematics education study program at Universitas Mulawarman. Learning resources that are predictor variables are multiple choice. The results showed that students who use YouTube as their learning source had the opportunity to study independently by 2.232 times greater than learning mathematics in groups. Meanwhile, students who used printed books as their learning source had the opportunity to study in groups 1.968 times greater than independent mathematics learning. The highest F1 score lies in the data division of 90:10, which was 0.643. The AUC score for the binary logistic regression model used was 0.611.

Keywords: Student way of learning; Binary logistic regression; AUC score; Learning resources

PENDAHULUAN

Saat mendalami materi matematika, mahasiswa dapat melakukannya secara mandiri atau berkelompok. Seorang mahasiswa yang belajar mandiri biasanya memiliki sifat *self efficacy* yaitu seseorang yang memiliki keyakinan telah menguasai materi matematika, keyakinan untuk dapat membantu kesulitan orang lain selama proses pembelajaran, keyakinan telah memiliki keterampilan yang cukup untuk dapat mengajar matematika secara efektif, dan keyakinan untuk dapat meningkatkan pencapaian orang lain saat belajar matematika (Zuya dkk., 2016). Mahasiswa calon guru matematika harus dapat meningkatkan 2 aspek terakhir tersebut dari 4 aspek *self efficacy* (Sumartini, 2020). Sementara selama perkuliahan, dosen memberi kesempatan mahasiswanya untuk menerapkan ide-ide mereka sendiri dan membangun sendiri pengetahuan mereka (Trianto, 2011). Hal ini mengakibatkan mayoritas mahasiswa yang belajar matematika secara berkelompok memiliki tujuan yang sama yaitu menyatukan pemikiran mengenai materi yang mereka pelajari (Rizki dkk., 2023). Selain itu, belajar kelompok dapat membentuk keakraban dan kerjasama dalam kelas, meningkatkan kemampuan akademis, rasa percaya diri, dan melatih seseorang untuk mengeluarkan aspirasinya secara terstruktur (Setiawan, 2015).

Salah satu hal yang penting dalam melaksanakan proses pembelajaran baik di dalam kelas maupun di luar kelas adalah sumber belajar. Beberapa sumber belajar yang dapat dimanfaatkan oleh mahasiswa selama belajar di luar kelas antara lain buku, jurnal, YouTube, podcast, atau les/kursus (baik secara *offline* maupun secara *online* seperti Zenius). Mahasiswa harus memiliki lebih dari 1 sumber belajar agar dapat mengkonstruksi pemahaman konsep matematika berdasarkan pengetahuan yang telah mereka miliki sebelumnya.

Dalam penelitian ini, cara belajar mahasiswa adalah data *supervised learning* atau data yang memiliki label yaitu belajar mandiri atau belajar berkelompok. Oleh karena itu, variabel ini dapat diklasifikasikan dengan memanfaatkan *machine learning*. Beberapa algoritma *machine learning* yang dapat mengklasifikasikan data *supervised learning* antara lain *Support Vector Machines* (SVM), Regresi Logistik, Naïve Bayes, Analisis Diskriminan, Pohon keputusan, dan *k-Nearest Neighbor* (k-NN). Penelitian ini merupakan penelitian lanjutan untuk mengeksplorasi cara belajar mahasiswa ketika ditinjau dari sumber belajarnya. Hasil penelitian sebelumnya yang menggunakan pohon keputusan, menunjukkan bahwa YouTube merupakan faktor utama yang dapat menentukan cara belajar seorang mahasiswa. Jika mahasiswa menggunakan sumber belajar YouTube maka diklasifikasikan sebagai

mahasiswa yang belajar secara mandiri. Namun jika seorang mahasiswa tidak menggunakan YouTube sebagai sumber belajarnya, maka diklasifikasikan sebagai mahasiswa yang belajar secara berkelompok (Rizki dkk., 2023).

Model regresi logistik adalah salah satu model statistika yang tidak mengasumsikan adanya distribusi Normal. Model yang dihasilkan juga mengkombinasi fitur linier dan fungsi non linear (sigmoid). Selain itu, regresi logistik juga menghasilkan interpretasi probabilitas yang bagus yang akan mempermudah untuk memperbarui model dan mengambil data baru.

Regresi logistik bukan merupakan metode yang baru yang diterapkan pada data mahasiswa. Imaslihkah dkk. (2013) mengklasifikasi predikat kelulusan mahasiswa S1 di Institut Teknologi Sepuluh Nopember berdasarkan fakultas, jenis kelamin, asal daerah, jalur masuk ITS, status SMA, pekerjaan ayah, pekerjaan ibu, dan pendapatan orang tua. Daruyani dkk. (2013) memprediksi IPK mahasiswa semester pertama berdasarkan nilai rapor, nilai UN, jalur masuk, pilihan jurusan, tempat tinggal, metode belajar, biaya hidup perbulan, hubungan mahasiswa dengan teman, hubungan mahasiswa dengan keluarga serta motivasi belajar. Setyawati dkk. (2020) memprediksi nilai IPK mahasiswa program studi Pendidikan Matematika Fakultas Pendidikan MIPA Universitas Pendidikan Mandalika berdasarkan jurusan saat SMA dan asal daerah. Susetyoko dkk. (2022) mengklasifikasi status mahasiswa yang terdaftar melalui jalur masuk SBMPTN, apakah mahasiswa tersebut memilih jalur Kartu Indonesia Pintar atau tidak berdasarkan status orang tua, penghasilan orang tua, status rumah, jumlah rumah, jumlah motor, jumlah mobil, dan daya listrik rumah. Nurmalitasari & Purwanto (2022) mengklasifikasi indikator performa akademik apakah lulus kuliah atau tidak untuk data mahasiswa Universitas Duta Bangsa Surakarta (UDB) berdasarkan nilai IPK, pendapatan orang tua, status pernikahan, status pekerjaan mahasiswa, dan minat terhadap program studinya. Kebaharuan dalam penelitian ini adalah memanfaatkan regresi logistik untuk memprediksi cara belajar Mahasiswa berdasarkan sumber belajar yang merupakan pilihan ganda majemuk. Sumber belajar yang dipilih menjamin fleksibilitas dalam penentuan pilihan. Mahasiswa dapat memilih salah satu atau lebih pilihan sebagai sumber belajarnya. Sehingga model regresi logistik yang digunakan dapat mengatasi fleksibilitas pada pilihan ganda majemuk tersebut.

Penelitian ini bertujuan untuk membuat model regresi logistik biner yang dapat memprediksi cara belajar mahasiswa berdasarkan sumber belajarnya

METODE

Bahan dan Alat

Penelitian ini menggunakan bahasa pemrograman Python yang dikompilasi melalui Google Colab. Dengan memanfaatkan Google Colab, maka peneliti memperoleh kemudahan akses ke sumber daya komputasi berbasis cloud dan berkolaborasi antar peneliti.

Subjek dan Jenis Data Penelitian

Populasi yang digunakan dalam penelitian ini adalah semua mahasiswa yang masih mengambil mata kuliah. Sementara sampel penelitiannya adalah data dari 111 Mahasiswa program studi pendidikan matematika Universitas Mulawarman dari angkatan 2019, 2020, 2021, dan 2022. Penelitian ini menggunakan teknik sampling acak. Data berupa kuesioner dan dibagikan ketika mahasiswa bersedia menjadi responden.

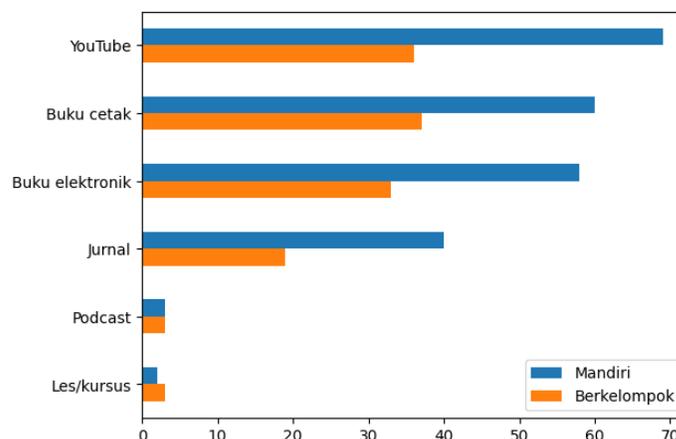
Isian data kuesioner mencakup cara belajar dan sumber belajar mahasiswa. Untuk cara belajar, inputan yang dapat dipilih adalah belajar mandiri atau berkelompok. Isian cara belajar ini adalah pilihan yang diyakini selama berperan menjadi mahasiswa. Sumber belajar yang menjadi variabel prediktor merupakan pilihan ganda majemuk. Hal ini memungkinkan bahwa seorang Mahasiswa dapat memilih lebih dari 1 sumber belajar. Setelah memperoleh model regresi logistik, selanjutnya menginterpretasi model tersebut dengan menggunakan *odd ratio* dan peluang dari model logit. Cara belajar Mahasiswa menjadi variabel respon biner dengan pemberian nilai 1 ketika Mahasiswa dominan belajar secara berkelompok dan variabel ini diberi nilai 0 ketika Mahasiswa dominan belajar secara mandiri.

Agar model regresi logistik yang diperoleh telah bekerja dengan benar, maka dilakukan pembagian data *training* dan data *testing* dengan proporsi 60:40, 70:30, 80:20, dan 90:10. Setelah dilakukan pembagian data ini, selanjutnya dihitung nilai akurasi, nilai presisi, nilai sensitivitas, dan nilai skor F1 untuk masing-masing proporsi. Lalu diakhiri dengan pembuatan kurva *Receiver Operating Characteristic* (ROC) dan penghitungan nilai *Area Under Curve* (AUC).

HASIL

Analisis statistika deskriptif dilakukan sebelum menganalisis data menggunakan model regresi logistik biner. Hasil analisis ini divisualisasikan melalui diagram batang (Gambar 1). Mahasiswa bebas memilih kanal YouTube, buku cetak, buku elektronik, jurnal, *podcast*,

les/kursus, atau kombinasinya sebagai sumber belajar yang digunakan selama belajar memahami materi perkuliahan. Sementara mahasiswa hanya boleh memilih salah satu cara belajar yang dominan yang diterapkan selama menjadi mahasiswa. Berdasarkan olahan data dalam Gambar 1, dapat ditunjukkan bahwa banyaknya mahasiswa yang belajar matematika secara mandiri lebih banyak dari pada mahasiswa yang belajar berkelompok. Sementara sumber belajar yang dipilih oleh mahasiswa sebagian besar adalah YouTube, lalu buku cetak, kemudian buku elektronik.



Gambar 1. Diagram batang data jumlah mahasiswa berdasarkan cara belajar dan sumber belajar

Selanjutnya adalah membuat model regresi logistik biner. Berdasarkan Tabel 1, sumber belajar yang memiliki koefisien negatif adalah Jurnal dan YouTube. Hal ini berarti mahasiswa yang menggunakan sumber belajar ini cenderung belajar materi matematika secara mandiri, karena variabel respon biner bernilai 0 jika mahasiswa tersebut belajar secara mandiri dan bernilai 1 jika mahasiswa tersebut belajar matematika berkelompok. Nilai *odd ratio* diperoleh dengan menghitung fungsi eksponensial dari koefisien suatu variabel. Berdasarkan nilai $1/odd\ ratio$, bahwa seorang mahasiswa yang menjadikan YouTube sebagai sumber belajarnya memiliki peluang sebagai mahasiswa yang belajar secara mandiri 2,232 kali lebih besar dari pada peluang sebagai Mahasiswa yang belajar berkelompok. Hal ini sesuai dengan penelitian sebelumnya bahwa sumber belajar YouTube merupakan faktor utama yang menentukan cara mahasiswa belajar. Mahasiswa yang tidak menjadikan YouTube sebagai sumber belajar diklasifikasikan sebagai mahasiswa yang belajar berkelompok. Sementara mahasiswa yang menggunakan YouTube sebagai sumber belajar diklasifikasikan sebagai mahasiswa yang belajar secara individu (Rizki dkk., 2023). Sementara berdasarkan nilai *odd ratio*, bahwa seorang mahasiswa yang menggunakan buku cetak sebagai sumber belajarnya memiliki peluang sebagai mahasiswa yang belajar secara kelompok 1,968 kali lebih besar dari pada peluang sebagai mahasiswa yang belajar mandiri. Buku cetak

yang relevan dengan materi yang didiskusikan masih menjadi sumber belajar yang baik ketika mahasiswa belajar secara berkelompok. Buku cetak menjadi bahan diskusi ketika mereka berkumpul untuk belajar bersama. Berkaitan dengan pemanfaatan aneka sumber belajar perlu disesuaikan dengan kebutuhan, efisiensi, dan efektivitas penggunaannya. Sumber yang dipilih sebaiknya dapat dipergunakan dan kapan pun dengan peralatan yang tersedia di sekitarnya yang tersedia mudah dipindah ke mana-mana (Supriadi, 2017).

Tabel 1. Koefisien dan *odd ratio* untuk model regresi logistik biner

Sumber belajar	Variabel	Koefisien	<i>Odd Ratio</i>	<i>1 / Odd Ratio</i>
	Konstanta	-0.323	0.724	1,381
Buku cetak	x_1	0.677	1.968	0,508
Les/Kursus	x_2	0.611	1.842	0,543
Podcast	x_3	0.358	1.431	0,699
Buku Elektronik	x_4	0.089	1.093	0,915
Jurnal	x_5	-0.412	0.662	1,511
YouTube	x_6	-0.803	0.448	2,232

Berdasarkan Tabel 1, maka diperoleh model logit sebagai berikut.

$$\text{logit}(P) = -0.323 + 0.677x_1 + 0.611x_2 + 0.358x_3 + 0.089x_4 - 0.412x_5 - 0.803x_6$$

Atau dapat ditulis menjadi peluang seseorang belajar secara berkelompok berikut.

$$P(\text{belajar berkelompok}) = \frac{e^{-0.323+0.677x_1+0.611x_2+0.358x_3+0.089x_4-0.412x_5-0.803x_6}}{1 + e^{-0.323+0.677x_1+0.611x_2+0.358x_3+0.089x_4-0.412x_5-0.803x_6}}$$

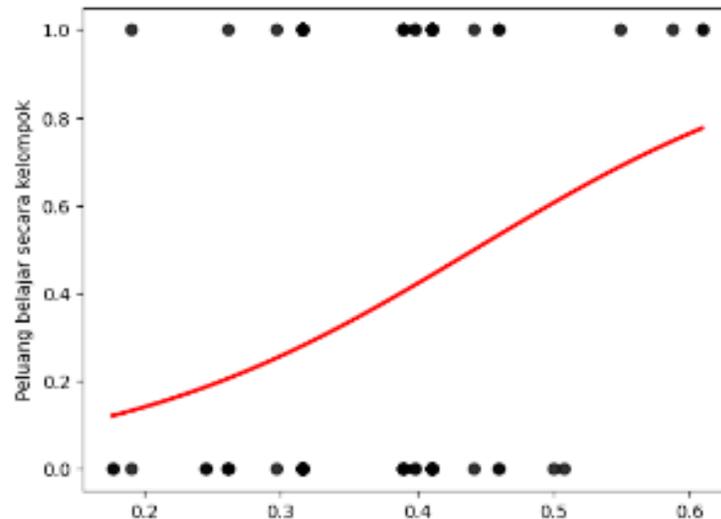
Artinya jika mahasiswa menggunakan buku cetak dan YouTube sebagai sumber belajar, maka peluang mahasiswa tersebut belajar secara berkelompok adalah

$$P(\text{belajar berkelompok} \mid \text{menggunakan sumber belajar buku cetak dan YouTube}) = \frac{e^{-0.323+0.677(1)+0.611(0)+0.358(0)+0.089(0)-0.412(0)-0.803(1)}}{1 + e^{-0.323+0.677(1)+0.611(0)+0.358(0)+0.089(0)-0.412(0)-0.803(1)}} = \frac{e^{-0.449}}{1 + e^{-0.449}} = 0.389 < 0.5$$

Karena peluangnya kurang dari 50%, maka mahasiswa dengan sumber belajarnya berupa buku cetak dan YouTube, diklasifikasikan sebagai mahasiswa yang belajar secara mandiri (individu).

Fungsi sigmoid untuk model regresi logistik yang diperoleh disajikan ke dalam kurva dalam Gambar 2. Kurva dari fungsi ini bernilai antara 0 dan 1 untuk variabel respon yang menginterpretasikan nilai peluang

seorang Mahasiswa belajar secara berkelompok. Jika nilai prediksi di atas 50%, maka diklasifikasikan sebagai mahasiswa yang belajar secara berkelompok. Kurva ini sedikit menyerupai berbentuk S.



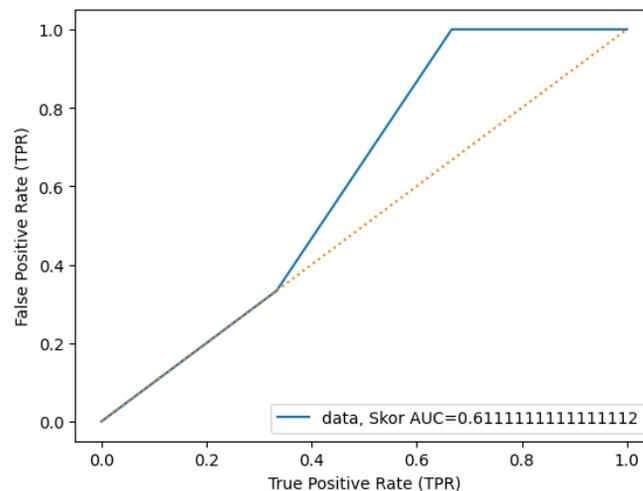
Gambar 2. Kurva regresi logistik

Performa model regresi logistik untuk beberapa pembagian data *training* dan data *testing* disajikan ke dalam Tabel 2. Nilai akurasi model meningkat seiring bertambahnya proporsi data *training*. Nilai akurasi tertinggi berada pada pembagian data 90:10 yaitu 0,750, yang berarti persentase data *testing* yang diprediksi benar adalah sebesar 75%. Nilai presisi tertinggi terletak pada pembagian data 70:30, yaitu 0,765, yang berarti jika mahasiswa yang ditaruh dalam data *testing* yang diklasifikasikan bahwa mahasiswa tersebut belajar secara berkelompok maka 76,5% mahasiswa tersebut memang belajar berkelompok (reliabel). Seberapa sensitif model mendeteksi data berlabel "belajar secara berkelompok" dijawab oleh nilai sensitivitas yang meningkat seiring bertambahnya proporsi data *training*. Nilai sensitivitas tertinggi berada pada pembagian data 90:10 yaitu 0,750, yang berarti dari semua mahasiswa yang belajar secara berkelompok bahwa 75% mahasiswa tersebut diklasifikasi secara benar (belajar secara berkelompok). Sementara nilai skor F1 merangkum kesimpulan dari nilai presisi dan sensitivitas. Nilai skor F1 bertambah ketika proporsi data *training* ditambah. Nilai skor F1 tertinggi terletak pada pembagian data 90:10 yang berarti pembagian ini merupakan proporsi yang tinggi untuk nilai presisi dan sensitivitas. Hal ini berarti 64,3% data diprediksi dengan tepat oleh model regresi logistik untuk pembagian data 90:10.

Tabel 2. Metrik model klasifikasi

Pembagian data	Akurasi	Presisi	Sensitivitas	Skor F1
60:40	0,600	0,563	0,600	0,486
70:30	0,706	0,795	0,706	0,610
80:20	0,739	0,546	0,739	0,628
90:10	0,750	0,562	0,750	0,643

Informasi metrik untuk model regresi logistik biner berdasarkan pembagian data lalu disampaikan secara lengkap melalui kurva *Receiver Operating Characteristic* (ROC) dan nilai skor luas *Area Under Curve* (AUC) pada Gambar 3. Kurva ROC dibentuk dari perhitungan *False Positive Rate* (FPR) dan *True Positive Rate* (TPR). Sementara AUC adalah luas daerah di bawah kurva ROC. Berdasarkan skor nilai AUC, bahwa peluang data diklasifikasi dengan tepat oleh model regresi logistik biner adalah sebesar 61,11%. Hal ini berarti hasil klasifikasi yang dihasilkan oleh model ini masih kurang sempurna karena kurang mendekati 100% (Han dkk., 2022).



Gambar 3. Kurva ROC

KESIMPULAN

Berdasarkan hasil dan pembahasan, maka disimpulkan bahwa mahasiswa yang menjadikan YouTube sebagai sumber belajarnya memiliki kecenderungan sebagai Mahasiswa yang belajar secara mandiri. Sedangkan mahasiswa yang menggunakan buku cetak sebagai sumber belajarnya cenderung menjadi mahasiswa yang belajar secara kelompok. Model regresi logistik biner yang diperoleh menghasilkan

akurasi, sensitivitas, skor F1 tertinggi pada pembagian data 90:10, berturut-turut yaitu 0,750, 0,750, dan 0,643. Nilai presisi pada model ini terletak pada pembagian data 70:30 yaitu sebesar 0,795. Skor AUC untuk model regresi logistik biner ini adalah sebesar 0,611.

DAFTAR PUSTAKA

- Daruyani, S., Wilandari, Y., & Yasin, H. (2013). Faktor-faktor yang Mempengaruhi Indeks Prestasi Mahasiswa FSM Universitas Diponegoro Semester Pertama dengan Metode Logistik Biner. *Prosiding Seminar Nasional Statistika Universitas Diponegoro 2013*, 185–194.
- Han, J., Pei, J., & Tong, H. (2022). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- Imaslihkah, S., Ratna, M., & Ratnasari, V. (2013). Analisis Regresi Logistik Ordinal terhadap Faktor-faktor yang Mempengaruhi Predikat Kelulusan Mahasiswa S1 di ITS Surabaya. *Jurnal Sains dan Seni ITS*, 2(2), D177–D182. <https://doi.org/10.12962/j23373520.v2i2.4959>
- Nurmalitasari, & Purwanto, E. (2022). Prediksi Performa Mahasiswa Menggunakan Model Regresi Logistik. *Jurnal Derivat: Jurnal Matematika Dan Pendidikan Matematika*, 9(2), 145–152. <https://doi.org/10.31316/jderivat.v9i2.2639>
- Rizki, N. A., Mumtaza, M., Dewi, C. F., Syahlafandi, D., & Fendiyanto, P. (2023). Implementasi Pohon Keputusan untuk Klasifikasi Cara Belajar Mahasiswa Mandiri dan Berkelompok Berdasarkan Sumber Belajarnya. *Prosiding Seminar Nasional Pendidikan Matematika, Universitas Mulawarman*, 3, 101–106.
- Setiawan, A. (2015). *Penerapan Belajar Kelompok Untuk Meningkatkan Minat Dan Prestasi Belajar Mata Pelajaran Bahasa Indonesia Siswa Kelas III SD Negeri Kepek, Pengasih, Kulon Progo Tahun Pelajaran 2014/2015* [Skripsi, Universitas Negeri Yogyakarta]. <https://eprints.uny.ac.id/17975/>
- Setyawati, D. U., Korida, B. D., & Febrilia, B. R. A. (2020). Analisis Regresi Logistik Ordinal Faktor-Faktor yang Mempengaruhi IPK Mahasiswa. *Jurnal Varian*, 3(2), 65–72. <https://doi.org/10.30812/varian.v3i2.615>
- Sumartini, T. S. (2020). Self Efficacy Calon Guru Matematika. *Mosharafa: Jurnal Pendidikan Matematika*, 9(3), 419–428. <https://doi.org/10.31980/mosharafa.v9i3.797>
- Supriadi. (2017). Pemanfaatan Sumber Belajar Dalam Proses Pembelajaran. *Lantanida Journal*, 3(2), 127–139. <http://dx.doi.org/10.22373/lj.v3i2.1654>
- Susetyoko, R., Yuwono, W., & Purwantini, E. (2022). Model Klasifikasi Pada Seleksi Mahasiswa Baru Penerima KIP Kuliah Menggunakan Regresi Logistik Biner. *Jurnal Informatika Polinema*, 8(4), 31–40. <https://doi.org/10.33795/jip.v8i4.914>
- Trianto. (2011). *Model-model pembelajaran inovatif berorientasi konstruktivistik: Konsep, landasan teoretis-praktis dan implementasinya*. Prestasi Pustaka Publisher.
- Zuya, H. E., Kwalat, S. K., & Attah, B. G. (2016). Pre-Service Teachers' Mathematics Self-Efficacy and Mathematics Teaching Self-Efficacy. *Journal of Education and Practice*, 7(14), 93–98.