

# Kontribusi Keluarga Dalam Prediksi Mahasiswa Lulus Tepat Waktu Menggunakan Model *Support Vector Machine*

<sup>1\*</sup>Wijiyanto, <sup>2</sup>Sopingi

<sup>1,2</sup> Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta

<sup>1\*</sup>wijiyanto@udb.ac.id, <sup>2</sup>sopingi@udb.ac.id

## ABSTRAK

*Keluarga mempunyai peranan penting terhadap keberhasilan anggotanya yang menempuh perkuliahan di kampus dalam rangka dapat menyelesaikan pendidikan secara tepat waktu. Kontribusi data keluarga ini perlu untuk di tindaklanjuti lebih mendalam lagi dalam rangka untuk mengetahui seberapa besar pengaruhnya terhadap prediksi yang dilakukan. Data yang digunakan berasal dari lulusan mahasiswa tahun 2023 sebanyak 365 record, data dipisah untuk pelatihan dan pengujian dengan perbandingan 50%. Selanjutnya data pelatihan diinputkan ke dalam tahap pelatihan menggunakan model support vector machine (SVM) menggunakan kernel "linier". Setelah model terbentuk data pengujian dimasukkan ke dalam tahap pengujian yang evaluasinya menggunakan confusion matrix untuk memperoleh nilai akurasi. Dari hasil evaluasi yang dilakukan diperoleh nilai accuracy sebesar 0.86, nilai precision sebesar 0.86, nilai recall sebesar 0.99 dan nilai F1-score sebesar 0.92. Hasil evaluasi menunjukkan bahwa model SVM mempunyai keakuratan yang baik sehingga, Fakultas Ilmu Komputer Universitas Duta Bangsa Surakarta dapat menerapkannya dalam meramalkan apakah seorang mahasiswa akan lulus tepat waktu atau tidak, sehingga memungkinkan untuk melakukan antisipasi yang lebih tepat.*

**Kata Kunci :** Accuracy, Precision, Prediksi, Recall, SVM

## PENDAHULUAN

Lingkungan keluarga memiliki kemampuan untuk memengaruhi keberhasilan mahasiswa untuk lulus tepat waktu sesuai dengan jenjang pendidikannya. Dalam mengukur keberhasilan mahasiswa akan lebih cenderung melihat data akademik selama mengikuti perkuliahan di kampus, padahal keluarga juga memiliki kontribusi yang besar terhadap keberhasilan seorang mahasiswa dalam rangka dapat menyelesaikan perkuliahannya secara tepat waktu. Masih ada beberapa perguruan tinggi di Indonesia yang mengalami masalah ketepatan waktu mahasiswa sebagai isu krusial, diantaranya adalah kesulitan dalam memprediksi tingkat kelulusan secara tepat (Darmawan et al., 2023) karena tidak semua mahasiswa dapat menyelesaikan perkuliahannya dalam rentang waktu yang telah ditetapkan (Bangun et al., 2022).

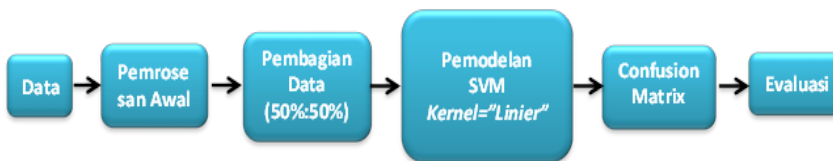
*Machine learning* merupakan pembelajaran dengan menganalisa data yang dapat berjalan dengan sendirinya tanpa suatu perintah dan salah satu tugas pentingnya adalah melakukan pengklasifikasian yang merupakan implementasi dalam *data mining* serta dapat digunakan dalam peramalan kelulusan menggunakan parameter tertentu (Rolansa et al., 2020). Banyak algoritma yang digunakan dalam model *machine learning*, diantaranya adalah *support vector machine* (SVM) dengan berbagai kategori kernel yang dimiliki karena dapat melakukan peramalan lebih tepat jika dibandingkan dengan algoritma *Decision Tree* (Wiyono & Abidin, 2018).

Pada penelitian sebelumnya yang mengimplementasikan algoritma SVM yang dilakukan oleh (Mailana et al., 2021) menghasilkan nilai akurasi 85% dalam memprediksi kelulusan, demikian juga yang dilakukan oleh (Nasution et al., 2023) diperoleh model terbaik menggunakan variabel masa studi adalah SVM yang menghasilkan akurasi sebesar 0.70. SVM juga dipakai oleh (Novianto et al., 2024) dalam memprediksi penerima beasiswa yang menghasilkan akurasi sebesar 85.84%, dengan menggunakan kernel linear memberikan akurasi terbaik 84,37% untuk dataset yang dipilih (Pande, 2023). Selain itu implementasi SVM juga digunakan dalam meramalkan kemampuan akademik mahasiswa dalam mengklasifikasikan dua kelas potensial dan tidak potensial (Nugroho et al., 2023).

Dalam eksperimen ini, algoritma SVM yang menggunakan kernel “linier” dapat digunakan dalam memprediksi mahasiswa lulus tepat waktu menggunakan data keluarga dengan harapan dapat menjadi model pengujian algoritma SVM untuk mengetahui keakuratan atau ketepatannya.

## **METODE**

Penelitian ini menerapkan pendekatan kuantitatif dalam metodologinya dimana salah satunya adalah menggunakan kategori eksperimen (Utami et al., 2021) pada gambar 1 terlihat dimulai dari mengumpulkan data, *preprocessing*, *split data*, pemodelan menggunakan SVM, *confusion matrix* dan evaluasi.



Gambar 1. Tahapan eksperimen

## A. Data

Data mining adalah langkah-langkah yang dilakukan untuk menggali informasi dari sumber data besar, sering kali dalam bentuk gudang basis data. Proses ini menghasilkan pemahaman baru yang berguna untuk pengambilan keputusan, dengan menganalisis data dalam skala yang besar (Nugroho et al., 2023). Data dikumpulkan dari mahasiswa yang lulus tahun 2023 pada Fakultas Ilmu Komputer Universitas Duta Bangsa Surakarta sebanyak 365 data dan memiliki 10 fitur dan 1 class sebagai target prediksi dimana 0 adalah tidak tepat waktu dan 1 adalah lulus tepat waktu.

## B. Pemrosesan Awal (*preprocessing*)

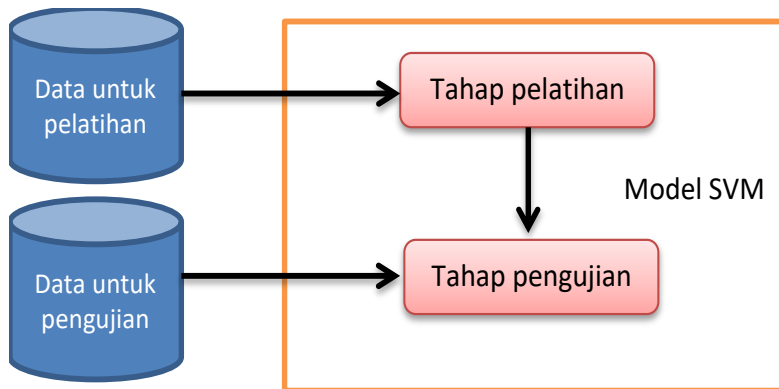
Tahap pemrosesan awal merupakan persiapan data sebelum diolah menggunakan model yang ditetapkan. Pemrosesan awal memiliki signifikansi yang tinggi dalam proses data mining karena kualitas input data yang baik akan menghasilkan analisis output yang berkualitas juga (Junaedi, 2022). Yang dilakukan pada tahap ini adalah memeriksa data yang berbentuk pilihan dibuat menjadi kategorikal, menghapus fitur yang tidak berhubungan dengan data keluarga dan memeriksa seluruh distribusi fitur yang digunakan.

## C. Pembagian Data (*Split Data*)

Pembagian dilakukan dengan memisahkan data yang digunakan untuk data training dan data untuk data testing. Pembagian tersebut disusun sedemikian rupa sehingga perbandingan antara kelas minoritas dan kelas mayoritas pada data pelatihan dan data pengujian memiliki proporsi yang relatif sama dengan data aslinya, dimana pembagian data menggunakan perbandingan 50%, dimana 182 data digunakan sebagai data pelatihan dan 183 data digunakan untuk data pengujian.

## D. Pemodelan *Support Vector Machine*

SVM (*Support Vector Machine*) merupakan sebuah metode klasifikasi yang memanfaatkan hyperplane sebagai pembatas antara kelas dalam melakukan proses klasifikasi (Amelia et al., 2018). SVM bekerja didasarkan pada *Structural Risk Minimization* yang bertujuan dalam memproses data dengan membuat *Hyperplane* supaya mampu mengklasifikasi ruang masukkan ke dalam 2 kelas. SVM dimulai dari mengelompokkan pada masalah linier yang dapat dipisah oleh *hyperplane* dan membaginya berdasarkan kelas-nya. Data yang sudah di pisah, data untuk pelatihan diinputkan kedalam tahap pelatihan pada model SVM menggunakan parameter kernel="linier". Setelah mendapatkan modelnya, data untuk pengujian dimasukkan ke dalam tahap pengujian (Haryatmi & Pramita Hervianti, 2021).



Gambar 2. Tahap pemodelan SVM

### E. Confusion Matrix

*Confusion matrix* merupakan tabel yang dapat difungsikan dalam memvisualisasikan kinerja model klasifikasi pada dataset pengujian di mana nilai actual-nya sudah diketahui. Ini membantu dalam mengidentifikasi kebingungan antar kelas, seperti salah satu kelas umumnya disalahartikan sebagai kelas lain. Jumlah prediksi yang benar dan salah dihitung dan diperinci berdasarkan kelas. *Confusion matrix* memberikan wawasan tentang bagaimana klasifikasi bingung dalam membuat prediksi, memberikan informasi detail tentang jenis kesalahan yang dilakukan oleh classifier (Sutoyo & Almaarif, 2020).

		Actual	
		Positif (1)	Negatif(0)
Predicted	Positif (1)	TP	FP
	Negatif(0)	FN	TN

Gambar 3. Model confusion matrix

## F. Evaluasi

Tahap ini dilakukan untuk mengevaluasi hasil prediksi dari algoritma dengan membandingkan nilai relatifnya terhadap klasifikasi data aktual. Metode *confusion matrix* digunakan sebagai alat evaluasi utama.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

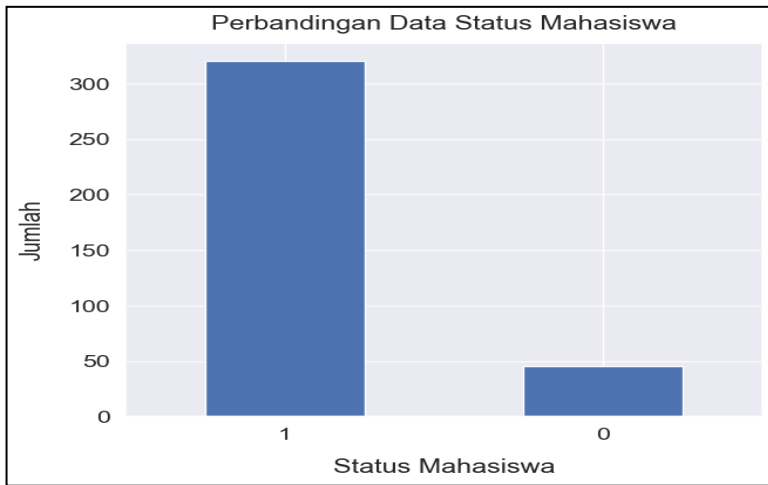
Dalam evaluasi ini, nilai-nilai seperti akurasi, *precision*, *recall* dan *F1-score* dievaluasi untuk memberikan pemahaman tentang seberapa baik algoritma melakukan prediksi terhadap data aktual..

## HASIL DAN PEMBAHASAN

Hasil eksperimen menggunakan model machine learning algoritma support vector machine dilakukan menggunakan jupyter notebook versi 6.5.4 bahasa pemrograman python.

### A. Persiapan Awal

Dari data yang dikumpulkan, mahasiswa yang dapat menyelesaikan pendidikannya secara tepat waktu sebanyak 320 mahasiswa dan yang tidak tepat waktu sebanyak 45 mahasiswa. Data tersebut sebagaimana terlihat pada gambar 4 terlihat bahwa tidak seimbang (*imbalanced data*) antara yang tepat waktu dengan tidak tepat waktu menggunakan perbandingan 1:7 namun *imbalanced data* tersebut tidak dibahas dalam penelitian ini karena salah satu tantangan dalam masalah prediksi mahasiswa adalah data yang tidak seimbang, yang mengurangi efisiensi pengklasifikasi Machine Learning (Masood & Begum, 2022).



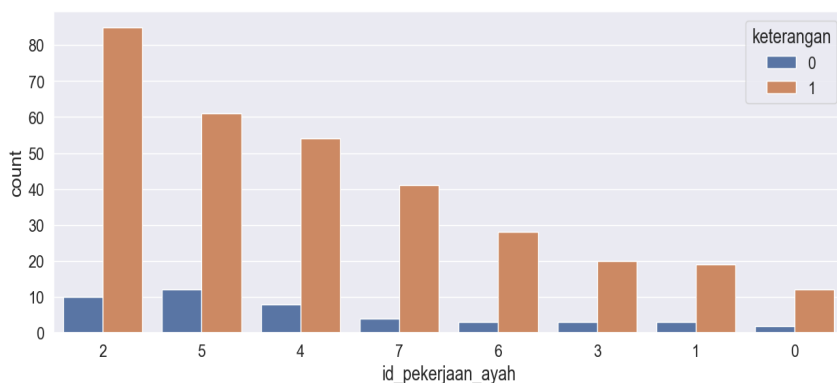
Gambar 4. Jumlah perbandingan status mahasiswa

Sementara dari tabel 1 terlihat penjelasan deskripsi data dari masing masing fitur yang digunakan sebagai panduan untuk melakukan proses selanjutnya.

Tabel 1. Deskripsi fitur

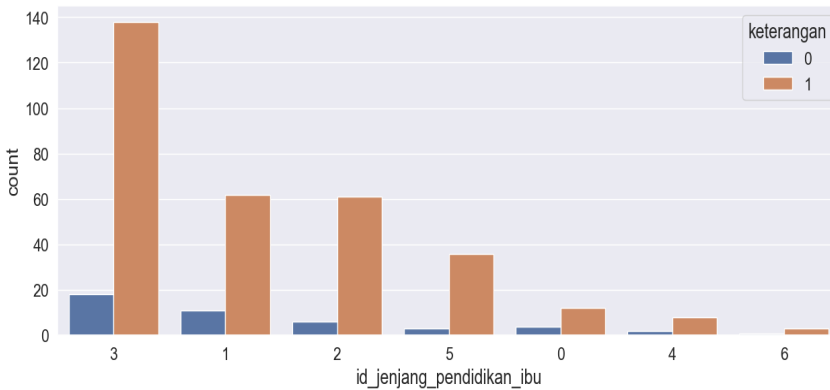
No	Fitur	Keterangan
1	jk	0= P/Perempuan, 1=L/Laki-laki
2	asal	Asal mahasiswa 0=Luar Kota, 1=Dalam Kota
3	umur	Umur mahasiswa pada saat lulus
4	Pendidikan_ayah	0=Tidak berpendidikan, 1=SD dan yang sederajat, 2=SMP dan yang sederajat 3=SMA dan yang sederajat, 4=D4/D3/D2/D1, 5=Sarjana (S1), 6=Magister (S2), 7=Doktor (S3)
5	id_pekerjaan_ayah	0=Tidak bekerja, 1=ASN/TNI/Polri, 2=Wirausaha/Wirawasta, 3="Pedagang Kecil 4=Buruh, 5=Karyawan Swasta, 6=Petani, 7=Lainnya
6	id_penghasilan_ayah	0=dibawah 500.000,- 1= 500.000,- s.d 999.999,- 2= 1.000.000,- s.d 1.999.999,- 3= 2.000.000,- s.d 4.999.999,- 4= 5.000.000,- s.d 20.000.000,-
7	id_jenjang_pendidikan_ibu	0=Tidak berpendidikan, 1=SD dan yang sederajat, 2=SMP dan yang sederajat 3=SMA dan yang sederajat, 4=D4/D3/D2/D1, 5=Sarjana (S1), 6=Magister (S2), 7=Doktor (S3)
8	id_pekerjaan_ibu	0=Tidak bekerja, 1=PNS/TNI/Polri, 2=Wirausaha/Wirawasta, 3="Pedagang Kecil 4=Buruh, 5=Karyawan Swasta, 6=Petani, 7=Lainnya
9	id_penghasilan_ibu	0=dibawah 500.000,- 1= 500.000,- s.d 999.999,- 2= 1.000.000,- s.d 1.999.999,- 3= 2.000.000,- s.d 4.999.999,- 4= 5.000.000,- s.d 20.000.000,-
10	ipk	Indeks prestasi kumulatif mahasiswa
11	keterangan	Target 0=Tidak tepat waktu, 1=tepat waktu

Fitur yang digunakan, masing masing telah di cek dalam rangka untuk mendalami distribusi setiap fiturnya. Dari gambar 5 terlihat distribusi fitur pekerjaan ayah dimana, semua mahasiswa yang lulus memiliki ayah yang bekerja. Bahkan mahasiswa yang ayahnya tidak bekerja (0) juga ada yang lulus tepat waktu, sedangkan mahasiswa yang paling banyak lulus dengan tepat waktu yaitu yang ayahnya memiliki pekerjaan nomor 2 yaitu Wirausaha atau Wirawasta.



Gambar 5. Distribusi fitur pekerjaan ayah

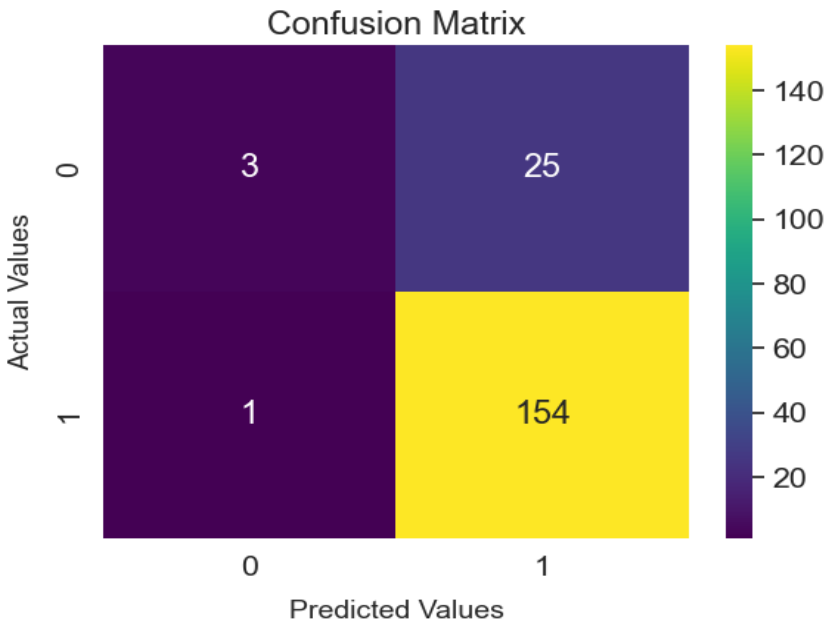
Distribusi fitur pendidikan ibu pada gambar 6 terlihat bahwa mahasiswa lulus tepat waktu paling banyak adalah yang memiliki ibu dengan pendidikan 3 (SMA dan yang sederajat), bahkan mahasiswa yang memiliki ibu tidak sekolah (0) dapat lulus tepat waktu. Dan temuan lainnya bahwa terdapat mahasiswa yang memiliki ibu dengan Pendidikan S2 (6) tidak dapat lulus tepat waktu.



Gambar 6. Distribusi fitur Pendidikan Ibu

## B. Modelling

Sebelum akan dilakukan pemodelan data dipisah ke dalam data pelatihan dan data untuk pengujian menggunakan perbandingan 50%:50% atau 182:183. Selanjutnya dilaksanakan pemodelan algoritma SVM (*support vector machine*) menggunakan kernel=linier. Hasil evaluasi menggunakan *confusion matrix* pada gambar 7 terlihat nilai nilai actual dan nilai prediksi menggunakan SVM.



Gambar 7. Confusion Matrix

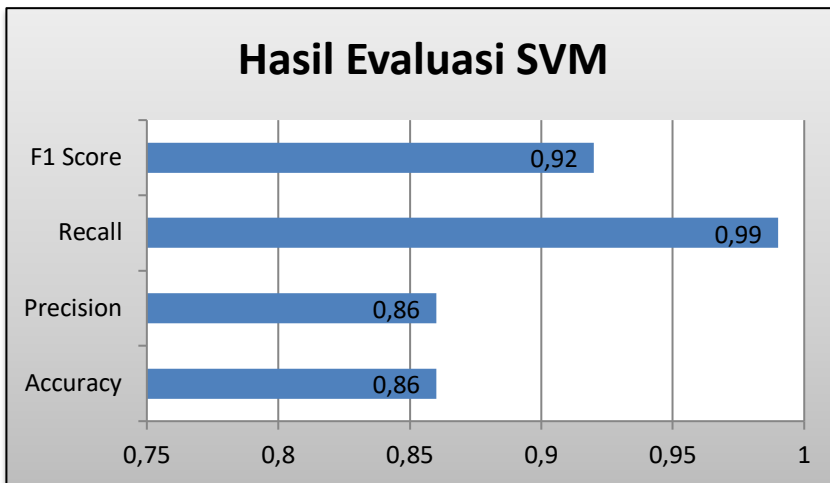
### C. Evaluasi

Dari hasil evaluasi Tabel 2 didapatkan kinerja klasifikasi algoritma SVM dengan nilai accuracy sebesar 0.86. Selain accuracy juga didapatkan nilai precision, recall dan F1-Score.

Tabel 2. Hasil kinerja klasifikasi SVM

Model	Accuracy	Precision	Recall	F1 Score
SVM Kernel="linier"	0.86	0.86	0.99	0.92

Selain itu, algoritma SVM menghasilkan nilai precision sebesar 0.86 yang sama dengan nilai accuracy, nilai recall sebesar 0.99 dan F1-Score sebesar 0.92.



Gambar 8. Grafik hasil evaluasi SVM

Dari gambar 8 terlihat bahwa evaluasi tertinggi terdapat pada nilai recall sebesar 0.99 yang memiliki perbedaan sebesar 0.07 dengan nilai F1-score dan perbedaan sebesar 0.13 dengan nilai precision dan accuracy.

## KESIMPULAN

Berdasarkan hasil penelitian, algoritma SVM (*support vector machine*) dapat digunakan dalam melakukan prediksi mahasiswa lulus tepat waktu berdasarkan data keluarga seperti asal kota, pendidikan orang tua (ayah dan ibu), pekerjaan orang tua (ayah dan ibu), penghasilan orang tua (ayah dan ibu) yang memiliki kontribusi terhadap keberhasilan perkuliahan. Setelah data di bagi dengan perbandingan 50%:50% untuk data pelatihan dan data pengujian diperoleh nilai *accuracy* sebesar 0.86, nilai *precision* sebesar 0.86, nilai *recall* sebesar 0.99 dan nilai *F1-score* sebesar 0.92. dengan demikian SVM dapat diimplementasikan oleh pihak yang berkepentingan dalam rangka untuk menghindari kegagalan pendidikan oleh seorang mahasiswa. Pada penelitian selanjutnya dapat dilakukan preprocessing untuk menangani imbalanced data dan menggunakan model machine learning lain untuk memperoleh hasil akurasi yang lebih baik.

## DAFTAR PUSTAKA

- Amelia, O. D., Soleh, A. M., & Rahardiantoro, S. (2018). Pemodelan Support Vector Machine Data Tidak Seimbang Keberhasilan Studi Mahasiswa Magister IPB. *Xplore: Journal of Statistics*, 2(1), 33–40. <https://doi.org/10.29244/xplore.v2i1.76>
- Bangun, O., Mawengkang, H., & Efendi, S. (2022). Metode Algoritma Support Vector Machine (SVM) Linier Dalam Memprediksi Kelulusan Mahasiswa. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(4), 2006. <https://doi.org/10.30865/mib.v6i4.4572>
- Darmawan, A., Yudhisari, I., Anwari, A., & Makruf, M. (2023). Pola Prediksi Kelulusan Siswa Madrasah Aliyah Swasta dengan Support Vector Machine dan Random Forest. *Jurnal Minfo Polgan*, 12(1), 387–400. <https://doi.org/10.33395/jmp.v12i1.12388>
- Haryatmi, E., & Pramita Hervianti, S. (2021). Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(2), 386–392. <https://doi.org/10.29207/resti.v5i2.3007>
- Junaedi, H. (2022). Prediksi Kelulusan Mahasiswa Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. *SMARTICS Journal*, 8(1), 1–7. <https://doi.org/10.21067/smartics.v8i1.6879>
- Mailana, A., Putra, A. A., Hidayat, S., & Wibowo, A. (2021). Comparison of C4.5 Algorithm and Support Vector Machine in Predicting the Student Graduation Timeliness. *Jurnal Online Informatika*, 6(1), 11. <https://doi.org/10.15575/join.v6i1.608>
- Masood, S. W., & Begum, S. A. (2022). Comparison of Resampling Techniques for Imbalanced Datasets in Student Dropout Prediction. *2022 IEEE Silchar Subsection Conference (SILCON)*, 1–7. <https://doi.org/10.1109/SILCON55242.2022.10028915>
- Nasution, N. B., Hartanto, D., Silitonga, D. J., Lasimin, & Mardhiyana, D. (2023). Prediksi Lama Studi dan Predikat Kelulusan Mahasiswa Menggunakan Algoritma Supervised Learning. *G-Tech: Jurnal Teknologi Terapan*, 7(2), 386–395. <https://doi.org/10.33379/gtech.v7i2.2077>

- Novianto, E., Hermawan, A., & Avianto, D. (2024). Perbandingan Metode K-Nearest Neighbor dan Support Vector Machine Untuk Memprediksi Penerima Beasiswa Keringanan UKT. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(1), Page 654-662. <http://dx.doi.org/10.30865/mib.v8i1.6913>
- Nugroho, B. I., Santoso, N. A., & Murtopo, A. A. (2023). Prediksi Kemampuan Akademik Mahasiswa dengan Metode Support Vector Machine. *remik*, 7(1), 177–188. <https://doi.org/10.33395/remik.v7i1.12010>
- Pande, S. M. (2023). Machine Learning Models for Student Performance Prediction. *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, 27–32. <https://doi.org/10.1109/ICIDCA56705.2023.10099503>
- Rolansa, F., Yunita, Y., & Suheri, S. (2020). Sistem prediksi dan evaluasi prestasi akademik mahasiswa di Program Studi Teknik Informatika menggunakan data mining. *Jurnal Pendidikan Informatika Dan Sains*, 9(1), 75. <https://doi.org/10.31571/saintek.v9i1.1696>
- Sutoyo, E., & Almaarif, A. (2020). Educational Data Mining for Predicting Student Graduation Using the Naïve Bayes Classifier Algorithm. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(1), 95–101. <https://doi.org/10.29207/resti.v4i1.1502>
- Utami, D. Y., Nurlelah, E., & Hasan, F. N. (2021). Comparison of Neural Network Algorithms, Naive Bayes and Logistic Regression to predict diabetes. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, 5(1), 53–64. <https://doi.org/10.31289/jite.v5i1.5201>
- Wiyono, S., & Abidin, T. (2018). Perbandingan Algoritma Machine Learning SVM dan Decision Tree untuk Prediksi Keaktifan Mahasiswa. *Sinkron*, 3(1), 105–108.