

Pola Prediksi Kelulusan Siswa Madrasah Aliyah Swasta dengan *Support Vector Machine* dan *Random Forest*

¹Aang Kisnu Darmawan, ²Ivana Yudhisari, ³Anwari Anwari, ⁴Masdukil Makruf

^{1, 2, 3}Prodi Sistem Informasi, Universitas Islam Madura, Indonesia

⁴Prodi Teknik Informatika, Universitas Islam Madura, Indonesia

¹ak.darmawan@gmail.com, ²ivanayudhisari@gmail.com, ³anwari.uim@gmail.com,

⁴masdukil.makruf@gmail.com

ABSTRAK

Kelulusan Siswa adalah salah satu indikator penting bagi kinerja keberhasilan sekolah. Prediksi kelulusan siswa penting bagi sekolah untuk mengidentifikasi siswa yang beresiko putus sekolah dan memberi mereka intervensi dini untuk meningkatkan kinerja akademik mereka. Ini juga dapat membantu pemangku kebijakan mengembangkan kebijakan dan program untuk meningkatkan tingkat kelulusan sekolah dan mengurangi tingkat putus sekolah. **Akan tetapi** berdasarkan penelusuran pustaka terdapat berbagai permasalahan krusial terkait prediksi kelulusan siswa yaitu sulitnya memprediksi secara akurat tingkat kelulusan hanya dengan menggunakan nilai tes penerimaan, Nilai SMA merupakan prediktor yang lebih baik untuk kelulusan perguruan tinggi tepat waktu daripada nilai tes penerimaan karena kesuksesan di perguruan tinggi tidak hanya membutuhkan kemampuan kognitif tetapi juga kompetensi pengaturan diri yang lebih baik diindeks oleh nilai SMA/MA, selain itu, terdapat beberapa faktor yang dapat mempengaruhi tingkat kelulusan siswa, seperti *self-efficacy*, budaya akademik sekolah, dan harapan di masa depan dan akhirnya, ada kebutuhan untuk menyempurnakan model penerimaan selektif untuk menentukan ukuran pencapaian sebelumnya mana yang memiliki validitas prediktif terbaik untuk kesuksesan akademik di universitas. **Penelitian ini bertujuan** untuk (1) memprediksi kelulusan siswa Madrasah Aliyah Swasta (MAS) dengan Algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) dan (2) melakukan komparasi performa algoritma SVM dan RF untuk prediksi kelulusan siswa MAS. Berdasarkan **hasil penelitian** yang telah dilakukan maka didapatkan kesimpulan bahwa untuk tujuan pertama algoritma SVM dan RF sangat baik dalam memprediksi kelulusan siswa MAS terlihat dari akurasi yang sangat tinggi yaitu SVM (98,98%) dan RF (99,49%). Untuk tujuan kedua didapatkan kesimpulan bahwa algoritma RF sedikit lebih baik daripada algoritma SVM dalam hasil pengukuran parameter *Accuracy* (99,49% vs 98,98%), *Precision* (99,74% vs 99,23%), *Recall* (99,74% vs 99,74%), *F-Measure* (99,74% vs 99,48%), dan *Classification error* (0,005 vs 0,010), akan tetapi untuk parameter waktu konsumsi SVM (0,04 detik) lebih baik daripada RF (0,26 detik). Penelitian ini memberikan **kontribusi** dalam 2 (dua) hal, pertama secara saintifik dengan menguji algoritma SVM dan RF untuk prediksi kelulusan siswa MAS, kedua memberikan rekomendasi bagi sekolah atau MAS untuk mengembangkan model penerimaan selektif siswa.

Kata Kunci: *educational data mining*, madrasah aliyah swasta, prediksi kelulusan siswa, , *random forest*, *support vector machine*

PENDAHULUAN

Kelulusan Siswa adalah salah satu indikator penting bagi kinerja keberhasilan sekolah (Sanjaya & Wahyana, 2022). Prediksi kelulusan siswa SMA atau Madrasah Aliyah (MA) mengacu pada penggunaan teknik data *mining* dan metode penelitian lainnya untuk memprediksi kinerja siswa dan tingkat kelulusan SMA (Galla et al., 2019). Prediksi kinerja siswa dan tingkat kelulusan penting bagi lembaga pendidikan untuk mengidentifikasi siswa yang berisiko putus sekolah dan memberi mereka intervensi dini untuk meningkatkan kinerja akademik mereka. Selain itu, prediksi siswa dapat membantu institusi mengembangkan strategi untuk meningkatkan kinerja siswa dan tingkat kelulusan. Penggunaan data lingkungan pembelajaran virtual, algoritma pembelajaran mesin, dan teknik penambangan data pendidikan telah diterapkan untuk memprediksi keberhasilan siswa. Sistem prediksi kelulusan siswa telah dilakukan dengan menggunakan *K-Nearest Neighbour* (Sanjaya & Wahyana, 2022), *Naive Bayes* (Kodratillah & Naya, 2021; Margita, 2022), *Decision Tree C4.5* (Effendi & Setiawan, 2020), dan membandingkan *Naive Bayes* dan *Decision Tree C4.5* (Effendi & Setiawan, 2020). Identifikasi pola perilaku siswa menggunakan *K-means clustering* dan *support vector machine* juga telah digunakan untuk memprediksi kinerja siswa. Selanjutnya, studi tentang hubungan antara ciri-ciri kepribadian Lima Besar (*Big Five*) dan prestasi akademik telah dilakukan untuk memprediksi kinerja siswa. Oleh karena itu, penggunaan teknik data mining dan metode penelitian lainnya dapat membantu institusi pendidikan untuk memprediksi kinerja siswa dan mengembangkan strategi untuk meningkatkan keberhasilan siswa.

Penting untuk melakukan penelitian tentang prediksi kelulusan siswa untuk pendidikan SMA/MA karena dapat membantu institusi pendidikan untuk mengidentifikasi siswa yang berisiko putus sekolah dan memberikan mereka intervensi dini untuk meningkatkan prestasi akademik mereka. Penggunaan teknik data mining dan metode penelitian lainnya dapat membantu lembaga pendidikan untuk memprediksi kinerja siswa dan mengembangkan strategi untuk meningkatkan keberhasilan siswa. Dengan memprediksi tingkat kelulusan siswa, institusi pendidikan dapat menilai kinerja institusional dan keberhasilan siswa (Oztek, 2016). Selain itu, penelitian tentang prediksi kelulusan siswa dapat membantu pembuat kebijakan untuk mengembangkan kebijakan dan program untuk meningkatkan angka kelulusan SMA dan menurunkan angka putus sekolah.

Akan tetapi terdapat beberapa permasalahan yang terkait dengan prediksi kelulusan siswa pada pendidikan SMA/MA. Salah satu masalah utamanya adalah sulitnya memprediksi secara akurat tingkat kelulusan hanya dengan menggunakan nilai tes penerimaan (Galla et al., 2019). Nilai SMA merupakan prediktor yang lebih baik untuk kelulusan perguruan tinggi tepat waktu daripada nilai tes penerimaan karena kesuksesan di perguruan tinggi tidak hanya membutuhkan kemampuan kognitif tetapi juga kompetensi pengaturan diri yang lebih baik diindeks oleh nilai SMA (Galla et al., 2019). Masalah lainnya adalah tantangan untuk menciptakan tenaga pengajar yang beragam ras dan etnis, yang terutama terkait dengan peningkatan kelulusan SMA, pendaftaran perguruan tinggi, dan tingkat kelulusan perguruan tinggi dari siswa minoritas (Vegas et al., 2001). Selain itu, terdapat beberapa faktor yang dapat mempengaruhi tingkat kelulusan siswa, seperti *self-efficacy*, iklim sekolah, dan harapan di masa depan (Heckhausen et al., 2013; Larson et al., 2015; Worrell & Hale, 2001). Akhirnya, ada kebutuhan untuk menyempurnakan model penerimaan selektif untuk menentukan ukuran pencapaian sebelumnya mana yang memiliki validitas prediktif terbaik untuk kesuksesan akademik di universitas (Vulperhorst et al., 2018).

Penelitian ini bertujuan untuk (1) memprediksi kelulusan siswa Madrasah Aliyah Swasta (MAS) dengan Algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) dan (2) melakukan komparasi performa algoritma SVM dan RF untuk prediksi kelulusan siswa MAS. Penelitian ini memberikan kontribusi dalam 2 (dua) hal, pertama secara saintifik dengan menguji algoritma SVM dan RF untuk prediksi kelulusan siswa MAS, kedua memberikan rekomendasi bagi sekolah atau MAS untuk mengembangkan model penerimaan selektif siswa.

TINJAUAN PUSTAKA

Berbagai penelitian telah dilakukan untuk prediksi kelulusan siswa diantaranya Prediksi waktu kelulusan mahasiswa dengan menggunakan Intelligent *K-Medoids Algorithm* (Cahaya et al., 2017), paper ini menyajikan hasil penerapan Algoritma *Intelligence K-Medoids* pada data yang telah diproses sebelumnya dari 240 siswa lulusan. Algoritma mengelompokkan siswa berdasarkan skor mereka dalam 25 mata pelajaran yang berbeda dan memprediksi waktu kelulusan mereka. Hasilnya adalah 7 cluster dengan nilai siluet 0,2416, dan setiap cluster diberi label sesuai dengan rentang waktu kelulusan siswa. Akurasi prediksi algoritma adalah 99,58% saat diuji menggunakan *k-cross* dengan 5 himpunan bagian. Kemudian Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan *Artificial Neural Network* Pada Data Mining Mahasiswa STMIK Widya Cipta Dharma Samarinda (Nurhuda & Rosita, 2017), penelitian ini menggunakan pembelajaran jaringan saraf tiruan untuk memprediksi kelulusan siswa tepat waktu berdasarkan data dari tahun akademik 2009. Hasil penelitian menunjukkan bahwa jaringan saraf dengan satu lapisan input, satu lapisan tersembunyi, dan satu lapisan keluaran mampu memprediksi secara akurat siswa mana yang akan lulus tepat waktu. Informasi ini dapat digunakan untuk menambah jumlah lulusan. Kemudian Prediksi keterlambatan kelulusan berdasarkan kinerja siswa (Ojha et al., 2017), penelitian ini menganalisis dampak berbagai faktor pada tingkat kelulusan universitas menggunakan tiga model prediktif : *Support Vector Machines* (SVM), *Gaussian Processes* (GP) dan *Deep Boltzmann Machines* (DBM). Model dilatih menggunakan data siswa yang sebenarnya, dan hasilnya menunjukkan bahwa DBM berkinerja sedikit lebih baik daripada model lainnya. Kemudian Prediksi Tingkat Kelulusan Siswa dengan Algoritma *Decision Tree C4.5* (Purnamasari et al., 2019), penelitian ini menghasilkan nilai tingkat akurasi sebesar 69,79% dan tingkat kesalahan sebesar 30,21. Faktor-Faktor yang Mempengaruhi Niat Mahasiswa Sarjana untuk Menggunakan Praktik Berbasis Bukti Setelah Lulus: Pengembangan dan Validasi Model Prediksi Berbasis Teori (Ramis et al., 2019), kemudian Prediksi Waktu Kelulusan Mahasiswa Menggunakan Algoritma Klasifikasi Dua Tingkat (Tampakas et al., 2019), paper ini menyajikan serangkaian tes yang dilakukan untuk mengevaluasi kinerja skema klasifikasi dua tingkat yang diusulkan menggunakan algoritma klasifikasi populer sebagai pembelajar A-level dan B-level. Hasilnya menunjukkan bahwa algoritma yang diusulkan menunjukkan prediksi yang dapat diandalkan berdasarkan kinerja siswa dalam kursus mereka selama dua tahun pertama studi mereka. Algoritma *Naive Bayes* adalah yang paling akurat, diikuti oleh algoritma C4.5, sedangkan algoritma RIPPER memiliki akurasi terendah. Penerapan Data *mining* Untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa (Wirawan et al., 2019), penelitian ini membandingkan tiga teknik penambangan data untuk memprediksi tingkat kelulusan siswa tepat waktu dan menyarankan menggunakan metode pohon keputusan untuk akurasi tertinggi. Nilai akurasi model adalah 89,82%, dan nilai presisi metode pohon keputusan adalah 52,63%. Nilai penarikan dari metode pohon keputusan adalah 41,67%. Penelitian ini bertujuan untuk memberikan rencana strategis bagi Syarif Hidayatullah UIN Jakarta untuk menjaga kualitas pendidikan dengan mengendalikan dan memantau siswa yang lulus tepat waktu.

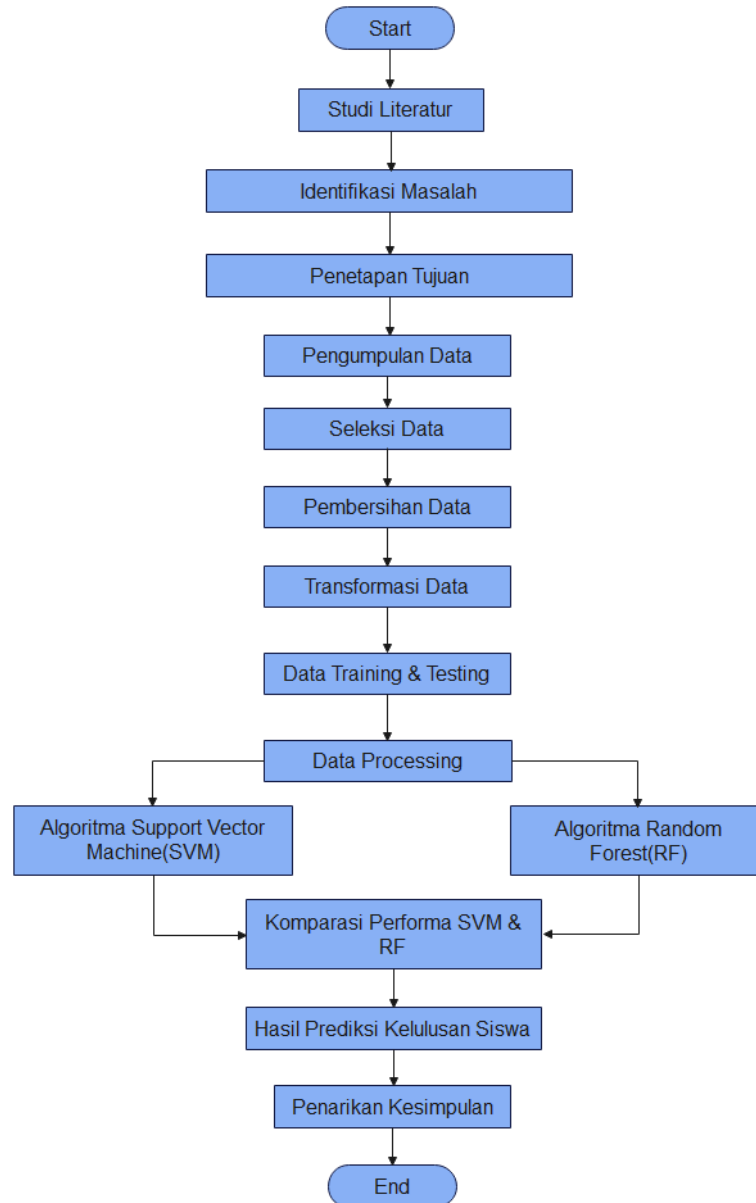
Penelitian tentang Prediksi Kelulusan Mahasiswa dengan Algoritma *Naive Bayes* (Hartatik et al., 2020), hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* dapat digunakan untuk mengembangkan model prediksi prestasi akademik siswa. Model ini menggunakan variabel seperti IPS1,2,3,4, tingkat PBB, jenis kelamin, dan status tinggal untuk memprediksi kinerja siswa dengan tingkat akurasi yang tinggi. Model ini dapat membantu agensi dan manajer membuat keputusan berdasarkan informasi tentang kemajuan siswa dan prediksi kelulusan. Kemudian penelitian tentang Peningkatan Akurasi Algoritma Klasifikasi untuk Prediksi Kelulusan Mahasiswa Menggunakan Model *Ensemble* (Lagman et al., 2020), paper ini menemukan bahwa menggabungkan prediksi beberapa algoritma pembelajaran mesin menggunakan teknik *ensemble* menghasilkan tingkat akurasi yang lebih tinggi dalam memprediksi kelulusan siswa. Secara khusus, model *Naive Bayes*, Regresi Logistik, Pohon Keputusan, dan *Neural Network* digunakan dalam kombinasi untuk mencapai hasil yang lebih baik. Kemudian Evaluasi Model Jaringan Saraf Tiruan *Backpropagation* Untuk Prediksi Awal Kelulusan Mahasiswa Universitas XYZ (Yaqin et al., 2021), penelitian ini menyajikan hasil penggunaan jaringan saraf untuk memprediksi masa studi

mahasiswa di program studi informatika di Universitas XYZ berdasarkan nilai IPK mereka. Hasil prediksi terbaik diperoleh dengan menggunakan model jaringan dengan 50 *neuron* di lapisan tersembunyi dan tingkat pembelajaran 0,01, dengan akurasi 77%. Kinerja model jaringan saraf diukur menggunakan nilai *Mean Squared Error* (MSE).

Penelitian tentang Pemodelan Berurutan Multivariat untuk Kinerja Siswa dan Prediksi Kelulusan (Kurniawati & Maulidevi, 2022), hasil penelitian menunjukkan bahwa baik LSTM maupun GRU memiliki kinerja yang luar biasa di atas 90% dalam memprediksi kinerja dan kelulusan siswa. Kinerja kedua arsitektur lebih baik daripada satu sama lain tergantung pada tugas yang sesuai. Prediksi kelulusan mahasiswa dapat memberikan kinerja yang memuaskan sejak semester pertama, meskipun memiliki *trade-off* dalam *recall*. Sedangkan dalam prediksi prestasi siswa, nilai RMSE dapat diterima sejak semester kedua. Secara keseluruhan, kinerja kinerja siswa dan prediksi kelulusan lebih baik jika menggunakan metode terpisah daripada metode gabungan. Kemudian Model Prediksi Kelulusan Mahasiswa Menggunakan *Deep Learning Convolutional Neural Network* (CNN) (Salam et al., 2022), hasilnya menunjukkan bahwa Model klasifikasi *Deep Learning* yang diusulkan yang menggunakan algoritma *Convolutional Neural Network* (CNN) mencapai nilai akurasi tinggi 87,44% dalam memprediksi kelulusan siswa. Evaluasi model mengkategorikannya sebagai model klasifikasi yang baik. Akurasi tertinggi ditemukan ketika jumlah layer antara 80-100 dalam model pelatihan. Kemudian Peningkatan Keakuratan Metode Prediksi Menggunakan SMOTE untuk Prediksi Dini Kelulusan Mahasiswa Universitas XYZ (Yaqin et al., 2022), paper ini membandingkan tiga model prediksi (ANN, K-NN, dan SVM) untuk memprediksi kelulusan awal siswa dalam program studi sistem informasi dan informatika di Universitas XYZ. Studi ini menggunakan SMOTE untuk menangani ketidakseimbangan kelas dan menemukan bahwa nilai akurasi tes terbaik adalah pada ANN dengan ketidakseimbangan data 62,5% hingga 70,5% setelah menggunakan SMOTE, dibandingkan dengan uji akurasi pada metode K-NN dengan SMOTE 69,3%, sedangkan metode SVM meningkat menjadi 69,8%. Peningkatan nilai *recall* yang paling signifikan menjadi 71,3% terjadi di ANN.

METODE PENELITIAN

Tahapan penelitian ini mengikuti langkah-langkah seperti dalam *flowchart* berikut ini:



Gambar 1. Tahapan Penelitian

Adapun tahapan penelitian dalam gambar *flowchart* diatas dijelaskan sebagai berikut:

1. Studi Literatur

Studi literatur atau kajian pustaka dilakukan dengan cara mempelajari paper penelitian terkait dengan prediksi kelulusan siswa melalui database jurnal portal garuda kemdikbud dengan kata kunci prediksi kelulusan siswa dan database jurnal yang terindeks di *scopus* melalui *software Publish or Perish(PoP)*.

2. Identifikasi Masalah

Pada tahap ini dilakukan dengan melakukan observasi di penelitian yang dilaksanakan di 3(tiga) Madrasah Aliyah Swasta yang ada di Kabupaten Pamekasan Madura yaitu MAS.

Matsaratul Huda Panempun Pamekasan, MAS. Miftahul Qulub Polagan Galis Pamekasan dan MAS. Nurul Ulum Teja Barat Pamekasan. Adapun waktu penelitian adalah mulai bulan Januari sampai dengan Maret 2023. Berdasarkan observasi didapatkan hasil bahwa permasalahan yang teridentifikasi adalah belum adanya informasi prediksi kelulusan siswa di Madrasah Aliyah Swasta tersebut.

3. Penetapan Tujuan

Pada tahapan ini dilakukan penentuan tujuan penelitian yaitu (1) melakukan prediksi kelulusan siswa dengan menggunakan algoritma SVM dan RF, (2) membandingkan performa atau kinerja algoritma SVM dan RF untuk prediksi kelulusan siswa Madrasah Aliyah Swasta.

4. Pengumpulan Data

Pada tahapan ini dilakukan dengan studi rekaman data sekunder, wawancara dengan pihak yang terkait dengan penelitian ini yaitu Kepala Madrasah dan Kepala Tata Usaha serta pengumpulan data nilai kelas XII Madrasah Aliyah Swasta. Data yang diteliti dalam penelitian ini sebanyak 493 siswa.

5. Seleksi Data

Pada proses seleksi data ini dengan menghilangkan beberapa atribut yang tidak relevan dengan tujuan penelitian. Atribut yang harus dihilangkan yaitu Nomer, Nama siswa, Kelas, Jenis kelamin, Tahun Lulus. Atribut yang dipilih hanya nilai ujian akhir madrasah, ujian praktik dan sikap karena informasi yang diberikan didalamnya sudah mewakili data yang dibutuhkan untuk dijadikan indikator penelitian. Nilai Ujian akhir Madrasah, Ujian praktik dan sikap yang dipilih menjadi atribut input dalam penelitian ini merupakan hasil rata-rata dari semua mata pelajaran kelas XII di Madrasah masing-masing .

6. Pembersihan Data

Tahapan ini bertujuan untuk memastikan tidak ada duplikasi data, mengidentifikasi data yang tidak konsisten, dan memperbaiki kesalahan pada data, seperti kesalahan pencetakan, sehingga data dapat diproses. dan digunakan untuk penambahan data. Dalam proses pembersihan data ini telah ditemukan data yang salah, data ganda dan data yang tidak konsisten sehingga penulis masih melakukan *cleaning* data agar dihasilkan data yang baik. Data yang sudah bersih terdiri dari 4(empat) atribut yaitu 3(tiga) atribut data input meliputi Ujian akhir Madrasah, Ujian praktik dan sikap serta 1(satu) atribut sebagai output yang berisi opsi tentang lulus dan tidak lulus.

7. Transformasi Data

Tahapan transformasi data ini dilakukan dengan mengubah data menjadi format yang sesuai untuk pemrosesan data *mining*. Pada penelitian ini, *software Weka* menginginkan data yang akan diolah berbentuk ARFF untuk mengatasi kompatibilitas. Data yang semula berbentuk *Microsoft Excel 2019* ditransformasi ke bentuk ARFF agar dapat menyesuaikan dengan *software Weka* tersebut.

8. Data Training dan Testing

Pada tahap ini dilakukan pembagian data menjadi *data training* dan *testing* menggunakan *software WEKA* dengan split data 20/80.

9. Pengolahan Data (Data Processing)

Pada tahap ini dilakukan *Data processing* menggunakan *percentage split* 20%. Hal ini berarti akan dilakukan pembagian data 20% *data training* dan 80% *data testing*. Dari 493 siswa, *data training* sebanyak 99 orang dan *data testing* sebanyak 394 orang. Data yang sudah dibersihkan akan diolah menggunakan *software WEKA*.

10. Analisis Algoritma SVM dan RF

Pada tahapan ini dilakukan analisis prediksi kelulusan siswa dengan algoritma SVM dan RF dengan menggunakan *software WEKA*.

11. Komparasi Performa Algoritma SVM dan RF

Pada tahapan ini dilakukan perbandingan atau komparasi performa kinerja algoritma SVM dan RF dengan menghitung dan membandingkan nilai *Accuracy*, *Precision*, *Recall*, *F-Measure*, *Classification Error* dan waktu konsumsi.

12. Hasil Prediksi Kelulusan Siswa

Pada tahapan ini diberikan hasil prediksi kelulusan siswa dengan algoritma SVM dan RF.

13. Penarikan Kesimpulan

Penarikan kesimpulan dilakukan untuk menjawab pertanyaan penelitian yaitu prediksi kelulusan siswa Madrasah Aliyah Swasta dan membandingkan kinerja algoritma SVM dan RF untuk prediksi kelulusan siswa Madrasah Aliyah Swasta(SMAS).

HASIL DAN PEMBAHASAN

Hasil penelitian prediksi kelulusan siswa Madrasah Aliyah Swasta(MAS) dengan pendekatan SVM dan RF adalah sebagai berikut:

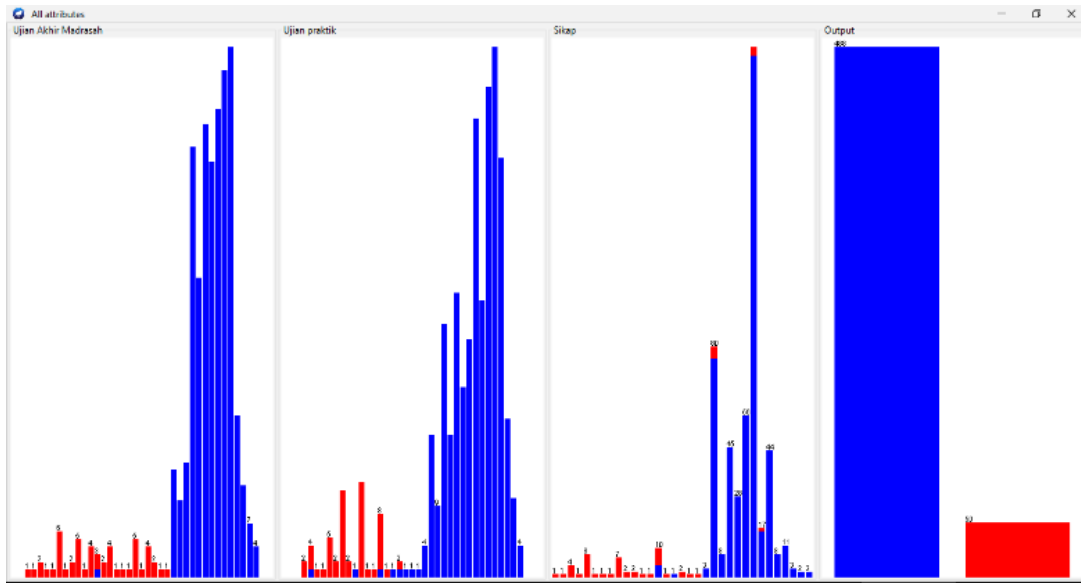
Hasil *Preprocessing* Data

Setelah melalui proses pengumpulan, pemilihan dan pembersihan data, peneliti memperoleh 493 data yang akan dipecah menjadi data latih dan data uji yang disiapkan untuk dimasukkan ke dalam perangkat lunak penelitian yaitu Weka *tools*. Hasil data yang telah dibersihkan adalah seperti di bawah ini:

Tabel 1. Data diperoleh dari Hasil *Preprocessing* Data

Ujian Akhir Madrasah	Ujian Praktek	Sikap	Output
85	83	84	Lulus
86	85	83	Lulus
87	86	82	Lulus
84	84	86	Lulus
84	83	85	Lulus
90	88	84	Lulus
85	84	85	Lulus
83	82	83	Lulus
83	84	84	Lulus
84	84	85	Lulus
...
79	80	80	Lulus
80	80	80	Lulus
82	81	82	Lulus
84	81	83	Lulus
84	50	83	Lulus
80	79	84	Lulus
81	80	82	Lulus
79	80	82	Lulus

Dari tabel tersebut dapat divisualisasikan dalam bentuk grafik pada Weka 3.9.6 seperti gambar dibawah ini:



Gambar 2. Grafik Atribut Data dengan tools WEKA

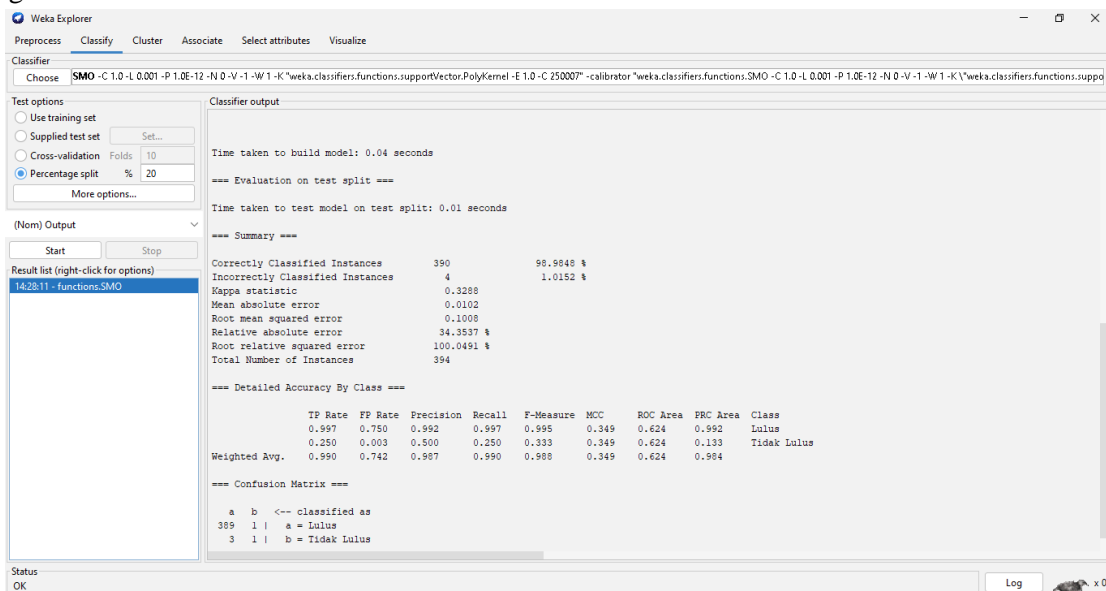
Informasi : Warna biru : Lulus, Warna merah : Tidak lulus

Hasil Processing Data

Analisis Algoritma *Support Vector Machine*(SVM) dan *Random Forest*(RF)

1. *Support Vector Machine*(SVM)

Pengolahan data pada metode ini menggunakan *percentage split* 20% dengan hasil seperti gambar dibawah ini :

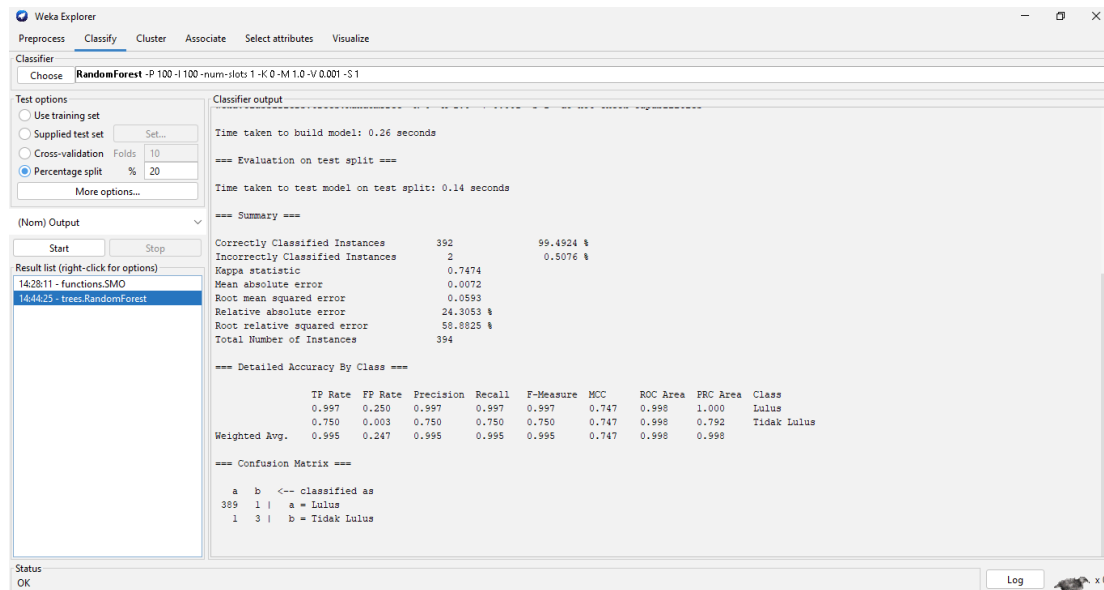


Gambar 3. Pengolahan data *Percentage split* 20% dengan SVM

Dari gambar diatas dijelaskan bahwa pada jumlah data *testing* sebanyak 394 data dihasilkan klasifikasi data yang benar sebesar 98,98% sedangkan klasifikasi data yang salah sebesar 1,02%.

2. Random Forest(RF)

Percentage split 20% pada metode ini dihasilkan seperti pada gambar dibawah ini :



Gambar 4. Pengolahan data Percentage split 20% dengan RF

Dari gambar diatas dijelaskan bahwa pada jumlah data *testing* sebanyak 394 data dihasilkan klasifikasi data yang benar sebesar 99,49% sedangkan klasifikasi data yang salah sebesar 0,51%.

Komparasi Performa SVM dan RF

Dari analisis kinerja diatas dapat dilakukan pengujian *confusion matrix* pada *percentage split* 20%. Dari pengujian tersebut dapat dilihat perbandingan antara kedua metode. Perhitungan *confusion matrix* tersebut sebagai berikut :

1. Algoritma Support Vector Machine(SVM)

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

$$= \frac{389 + 1}{389 + 1 + 3 + 1} \times 100\%$$

$$= \frac{390}{394} \times 100\% = 98,98\%$$

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

$$= \frac{389}{389 + 3}$$

$$= \frac{389}{392} = 0,9923 = 99,23\%$$

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

$$= \frac{389}{389 + 1}$$

$$= \frac{389}{390} = 0,9974 = 99,74\%$$

$$F - measure = 2x \frac{Precision \times Recall}{Precision + Recall}$$

$$= 2x \left(\frac{0,9923 \times 0,9974}{0,9923 + 0,9974} \right) = 2x \left(\frac{0,9897}{1,9897} \right)$$

$$= 2 \times 0,4974 = 0,9948 = 99,48\%$$

$$\text{Classification error} = \frac{FP + FN}{TP + TN + FP + FN}$$

$$= \frac{3 + 1}{389 + 1 + 3 + 1}$$

$$= \frac{4}{394} = 0,010$$

2. Algoritma Random Forest(RF)

Berdasarkan gambar hasil pengolahan data menggunakan *percentage split* 20% didapatkan *confusion matrix*. Adapun perhitungan *confusion matrix* dapat dilihat sebagai berikut :

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

$$= \frac{389 + 3}{389 + 3 + 1 + 1} \times 100\%$$

$$= \frac{392}{394} \times 100\% = 99,49\%$$

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

$$= \frac{389 + 1}{389 + 1}$$

$$= \frac{389}{390} = 0,9974 = 99,74\%$$

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

$$= \frac{389 + 1}{389 + 1}$$

$$= \frac{389}{390} = 0,9974 = 99,74\%$$

$$F - \text{measure} = 2x \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$= 2x \left(\frac{0,9974 \times 0,9974}{0,9974 + 0,9974} \right) = 2x \left(\frac{0,9948}{1,9948} \right)$$

$$= 2 \times 0,4987 = 0,9974 = 99,74\%$$

$$\text{Classification error} = \frac{FP + FN}{TP + TN + FP + FN}$$

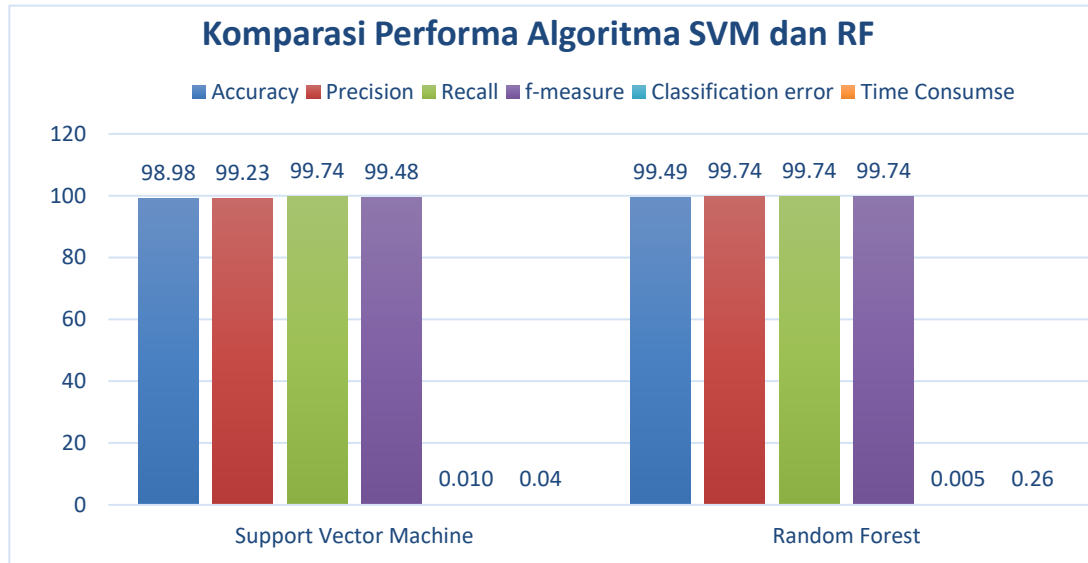
$$= \frac{1 + 1}{389 + 3 + 1 + 1}$$

$$= \frac{2}{394} = 0,005$$

Untuk mengetahui perbandingan diantara kedua metode dan menentukan metode mana yang mendapatkan hasil terbaik, berikut tabel dan grafik hasil perbandingan antara algoritma *Support Vector Machine*(SVM) dan *Random Forest*(RF) menggunakan *percentage split* 20% sebagai berikut :

Tabel 2. Perbandingan algoritma SVM dan RF

Pengolahan	Komparasi	SVM	RF
Percentage split 20%	Accuracy	98,98%	99,49%
	Presicion	99,23%	99,74%
	Recall	99,74%	99,74%
	F-measure	99,48%	99,74%
	Classification error	0,010	0,005
	Waktu konsumsi	0,04 detik	0,26 detik



Gambar 5. Perbandingan Performa Algoritma SVM dan RF

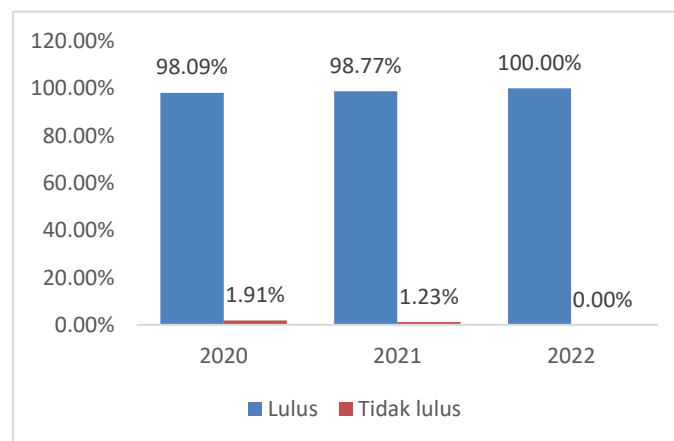
Meskipun algoritma RF sedikit lebih lambat daripada SVM dalam membangun model namun berdasarkan analisis dari aspek lainnya, algoritma RF lebih unggul maka RF merupakan metode terbaik bila dibandingkan dengan metode SVM.

Pola Prediksi Kelulusan Siswa

Berdasarkan hasil pengolahan data sebelumnya didapatkan pola prediksi kelulusan siswa sebagai berikut:

1. Pola Kelulusan Siswa

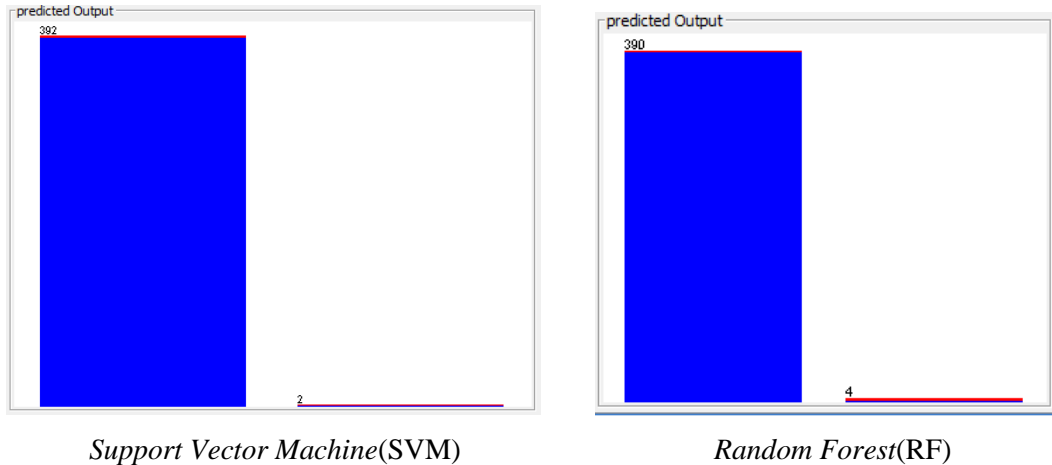
Berdasarkan nilai Ujian Akhir Madrasah, Ujian praktik dan sikap dari tahun kelulusan 2020 sampai dengan 2022 dihasilkan grafik seperti dibawah ini :



Gambar 6. Grafik Pola Kelulusan Siswa

2. Prediksi Kelulusan Siswa

Grafik diatas dapat membuktikan tingkat kelulusan siswa dari tahun ke tahun yang selalu naik secara signifikan. Namun berdasarkan metode *support vector machine*(SVM) dan *Random forest*(RF) diprediksi untuk kelulusan siswa di tahun yang akan datang dapat dilihat pada grafik dibawah ini :



Gambar 7. Predicted Output dengan kedua metode

Berdasarkan grafik diatas maka hasil prediksi tertinggi menggunakan *tools* Weka mendapatkan prosentase kelulusan siswa Madrasah Aliyah Swasta di tahun depan sebesar 99,49%.

KESIMPULAN

Penelitian ini bertujuan dua(2) hal yaitu (1) untuk memprediksi kelulusan siswa Madrasah Aliyah Swasta(MAS) dengan Algoritma *Support Vector Machine*(SVM) dan *Random Forest*(RF) dan (2) melakukan komparasi performa algoritma SVM dan RF untuk prediksi kelulusan siswa MAS. Berdasarkan hasil penelitian yang telah dilakukan maka didapatkan kesimpulan bahwa untuk tujuan pertama algoritma SVM dan RF sangat baik dalam memprediksi kelulusan siswa MAS terlihat dari akurasi yang sangat tinggi yaitu SVM(98,98%) dan RF(99,49%). Untuk tujuan kedua didapatkan kesimpulan bahwa algoritma RF sedikit lebih baik daripada algoritma SVM dalam hasil pengukuran parameter *Accuracy*(99,49% vs 98,98%), *Precision*(99,74% vs 99,23%), *Recall*(99,74% vs 99,74%), *F-Measure*(99,74% vs 99,48%), dan *Classification error*(0,005 vs 0,010), akan tetapi untuk parameter waktu konsumsi SVM(0,04 detik) lebih baik daripada RF(0,26 detik). Dengan demikian kesimpulannya algoritma RF sedikit lebih baik daripada SVM meskipun RF memerlukan waktu penyelesaiannya yang lebih lama dari SVM.

UCAPAN TERIMA KASIH

Terimakasih kepada Prodi Sistem Informasi dan Fakultas Teknik Universitas Islam Madura dan Civitas Akademika yaitu Kepala Sekolah dan Staf TU Madrasah Aliyah Swasta yang telah membantu sehingga penelitian ini terselenggara dengan baik di tahun 2023.

REFERENSI

- Cahaya, L., Hiryanto, L., & Handhayani, T. (2017). Student graduation time prediction using intelligent K-Medoids Algorithm. *2017 3rd International Conference on Science in Information Technology (ICSITech)*, 263–266. <https://doi.org/10.1109/ICSITech.2017.8257122>
- Effendi, M. M., & Setiawan, A. (2020). MENENTUKAN PREDIKSI KELULUSAN SISWA DENGAN MEMBANDINGKAN ALGORITMA C4.5 DAN NAIVE BAYES STUDI KASUS SMKN. 1 CIKARANG SELATAN. *SIGMA: Jurnal Teknologi Pelita Bangsa*, 11(3).
- Galla, B. M., Shulman, E. P., Plummer, B. D., Gardner, M., Hutt, S. J., Goyer, J. P., D’Mello, S. K., Finn, A. S., & Duckworth, A. L. (2019). Why High School Grades Are Better Predictors of On-Time College Graduation Than Are Admissions Test Scores: The Roles of Self-Regulation and Cognitive Ability. *American Educational Research Journal*, 56(6), 2077–2115. <https://doi.org/10.3102/0002831219843292>
- Hartatik, Kusriani, K., & Budi Prasetyo, A. (2020). Prediction of Student Graduation with Naive Bayes Algorithm. *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 1–5. <https://doi.org/10.1109/ICIC50835.2020.9288625>
- Heckhausen, J., Chang, E. S., Greenberger, E., & Chen, C. (2013). Striving for Educational and Career Goals During the Transition After High School: What is Beneficial? *Journal of Youth and Adolescence*, 42(9), 1385–1398. <https://doi.org/10.1007/s10964-012-9812-5>
- Kodratillah, E. Y., & Naya, C. (2021). PENERAPAN DATA MINING UNTUK PREDIKSI KELULUSAN SISWA MENGGUNAKAN ALGORITMA NAIVE BAYES PADA SMK GARUDA. *SIGMA: Jurnal Teknologi Pelita Bangsa*, 12(4).
- Kurniawati, G., & Maulidevi, N. U. (2022). Multivariate Sequential Modelling for Student Performance and Graduation Prediction. *2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 293–298. <https://doi.org/10.1109/ICITACEE55701.2022.9923971>
- Lagman, A. C., Alfonso, L. P., Goh, M. L. I., Lalata, J. P., Magcuyao, J. P. H., & Vicente, H. N. (2020). Classification Algorithm Accuracy Improvement for Student Graduation Prediction Using Ensemble Model. *International Journal of Information and Education Technology*, 10(10), 723–727. <https://doi.org/10.18178/ijiet.2020.10.10.1449>
- Larson, L. M., Pesch, K. M., Surapaneni, S., Bonitz, V. S., Wu, T.-F., & Werbel, J. D. (2015). Predicting Graduation: The Role of Mathematics/Science Self-Efficacy. *Journal of Career Assessment*, 23(3), 399–409. <https://doi.org/10.1177/1069072714547322>
- Margita, S. (2022). MODEL PREDIKSI UNTUK MENENTUKAN PREDIKAT KELULUSAN SISWA MENGGUNAKAN ALGORITMA NAIVE BAYES DAN MLP: STUDI KASUS SMK BUDDHI TANGERANG. *JURNAL ALGOR*, 3(2).
- Marhamah, M. (2020). *IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM) DALAM MEMPREDIKSI KELULUSAN JALUR MASUK PERGURUAN TINGGI BANDA ACEH (STUDI KASUS MAHASISWA BARU TAHUN AJARAN 2019)*. <https://library.ar-raniry.ac.id/>
- Nurhuda, A., & Rosita, D. (2017). Prediction Student Graduation on Time Using Artificial Neural Network on Data Mining Students STMIK Widya Cipta Dharma Samarinda. *Proceedings of the 2017 International Conference on E-Commerce, E-Business and E-Government*, 86–89. <https://doi.org/10.1145/3108421.3108431>

- Ojha, T., Heileman, G. L., Martinez-Ramon, M., & Slim, A. (2017). Prediction of graduation delay based on student performance. *2017 International Joint Conference on Neural Networks (IJCNN)*, 3454–3460. <https://doi.org/10.1109/IJCNN.2017.7966290>
- Oztekin, A. (2016). A hybrid data analytic approach to predict college graduation status and its determinative factors. *Industrial Management & Data Systems*, 116(8), 1678–1699. <https://doi.org/10.1108/IMDS-09-2015-0363>
- Purnamasari, E., Rini, D. P., & Sukemi. (2019). Prediction of the Student Graduation's Level using C4.5 Decision Tree Algorithm. *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)*, 192–195. <https://doi.org/10.1109/ICECOS47637.2019.8984493>
- Ramis, M., Chang, A., & Nissen, L. (2019). Factors Influencing Undergraduate Students' Intention to Use Evidence-Based Practice After Graduation: Development and Validation of a Theory-Based Prediction Model. *Worldviews on Evidence-Based Nursing*, 16(5), 397–407. <https://doi.org/10.1111/wvn.12384>
- Salam, A., Zeniarja, J., & Anthareza, D. M. (2022). Student Graduation Prediction Model using Deep Learning Convolutional Neural Network (CNN). *2022 International Seminar on Application for Technology of Information and Communication (ISemantic)*, 362–366. <https://doi.org/10.1109/iSemantic55962.2022.9920449>
- Sanjaya, A., & Wahyana, T. (2022). PENERAPAN METODE K-NEAREST NEIGHBOUR UNTUK SISTEM PREDIKSI KELULUSAN SISWA MTS NURUL MUSLIMIN BERBASIS. *Journal Transformation of Mandalika*, 3(1).
- Tampakas, V., Livieris, I. E., Pintelas, E., Karacapilidis, N., & Pintelas, P. (2019). Prediction of Students' Graduation Time Using a Two-Level Classification Algorithm. In M. Tsitouridou, J. A. Diniz, & T. A. Mikropoulos (Eds.), *Technology and Innovation in Learning, Teaching and Education* (Vol. 993, pp. 553–565). Springer International Publishing. https://doi.org/10.1007/978-3-030-20954-4_42
- Vegas, E., Murnane, R. J., & Willett, J. B. (2001). From High School to Teaching: Many Steps, who Makes It? *Teachers College Record: The Voice of Scholarship in Education*, 103(3), 427–449. <https://doi.org/10.1111/0161-4681.00121>
- Vulperhorst, J., Lutz, C., De Kleijn, R., & Van Tartwijk, J. (2018). Disentangling the predictive validity of high school grades for academic success in university. *Assessment & Evaluation in Higher Education*, 43(3), 399–414. <https://doi.org/10.1080/02602938.2017.1353586>
- Wirawan, C., Khudzaeva, E., Hasibuan, T. H., Karjono, & Lubis, Y. H. K. (2019). Application of Data mining to Prediction of Timeliness Graduation of Students (A Case Study). *2019 7th International Conference on Cyber and IT Service Management (CITSM)*, 1–4. <https://doi.org/10.1109/CITSM47753.2019.8965425>
- Worrell, F. C., & Hale, R. L. (2001). The relationship of hope in the future and perceived school climate to school completion. *School Psychology Quarterly*, 16(4), 370–388. <https://doi.org/10.1521/scpq.16.4.370.19896>
- Yaqin, A., Laksito, A. D., & Fatonah, S. (2021). Evaluation of Backpropagation Neural Network Models for Early Prediction of Student's Graduation in XYZ University. *International Journal on Advanced Science, Engineering and Information Technology*, 11(2), 610. <https://doi.org/10.18517/ijaseit.11.2.11152>
- Yaqin, A., Rahardi, M., & Abdulloh, F. F. (2022). Accuracy Enhancement of Prediction Method using SMOTE for Early Prediction Student's Graduation in XYZ University. *International Journal of Advanced Computer Science and Applications*, 13(6). <https://doi.org/10.14569/IJACSA.2022.0130652>