KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 5, No. 4, Oktober (2024), pp. 1741-1747

Application of Data Mining Techniques in Assessing the Performance of Vocational High School Students in Computer Engineering at SMK Negeri 1 Braja Selebah Using Support Vector Machines (SVM), Naive Bayes, and k-Nearest Neighbors (k-NN) Algorithms

Yoga Adi Putra[1], Handoyo WD Nugroho[2], Chairani[3], Joko Trilokai[4]*
[1,2,3,4]Fakultas Ilmu Komputer, IIB Darmajaya Lampung, Indonesia
E-mail: yogaadiputra22@gmail.com[1], handoyowd@darmajaya.ac.id[2], chairani@darmajaya.ac.id[3], jokotriloka@darmajaya.ac.id[4]

Abstract

This study discusses the performance comparison of three classification algorithms: K-Nearest Neighbors (k-NN), Naive Bayes, and Support Vector Machines (SVM), in assessing student performance at a Vocational High School specializing in Computer Engineering. The objective of this research is to identify the most effective algorithm for classification based on various evaluation metrics such as accuracy, precision, recall, and F1-Score. The experimental results show that the SVM algorithm has the best performance with an accuracy of 93.2%, precision of 93.4%, recall of 93.2%, and F1-Score of 93.1%. Naive Bayes ranks second with an accuracy of 86.2%, precision of 86.8%, recall of 86.2%, and F1-Score of 86.4%. The k-NN algorithm is in the last position with an accuracy of 81.0%, precision of 81.0%, recall of 82.0%, and F1-Score of 80.0%. Therefore, the SVM algorithm is recommended as the best model for classification in the context of this research.

Keywords: *Data Mining, Support Vector Machines (SVM), Naive Bayes, k-Nearest Neighbors (k-NN), Student Performance Evaluation, Vocational Education, SMK Negeri 1 Braja Selebah*

1. Pendahuluan

Dalam era digital yang terus berkembang, pendidikan di Indonesia, terutama di tingkat Sekolah Menengah Kejuruan (SMK), dituntut untuk menghasilkan lulusan yang siap memasuki dunia kerja dengan kompetensi yang memadai. Salah satu cara untuk memastikan kualitas pendidikan adalah dengan melakukan evaluasi kinerja siswa secara akurat dan objektif. Teknologi informasi, khususnya teknik data mining, telah terbukti sebagai alat yang efektif dalam menganalisis data besar dan memberikan wawasan yang mendalam terhadap kinerja siswa[1][2].

Data mining adalah proses mengekstraksi informasi yang bermanfaat dari kumpulan data yang besar dan kompleks. Dalam konteks pendidikan, data mining dapat digunakan untuk menganalisis data siswa, seperti nilai akademik, kehadiran, dan aktivitas ekstrakurikuler, untuk menilai kinerja mereka secara menyeluruh. Dengan penerapan algoritma machine learning, seperti Support Vector Machines (SVM), Naive Bayes, dan k-Nearest Neighbors (k-NN), evaluasi kinerja siswa dapat dilakukan secara lebih efisien dan akurat[3][4]

SMK Negeri 1 Braja Selebah merupakan salah satu institusi pendidikan yang berfokus pada bidang kejuruan, khususnya Teknik Komputer. Dalam upaya meningkatkan kualitas pendidikan, sekolah ini memerlukan metode yang tepat untuk mengevaluasi kinerja siswa. Penilaian kinerja siswa secara manual tidak hanya memakan waktu, tetapi juga rentan terhadap subjektivitas dan kesalahan manusia. Oleh karena itu, diperlukan sebuah sistem yang mampu mengevaluasi kinerja siswa secara otomatis dan objektif[5][6].

ISSN: 2720-992X
Copyright © KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)

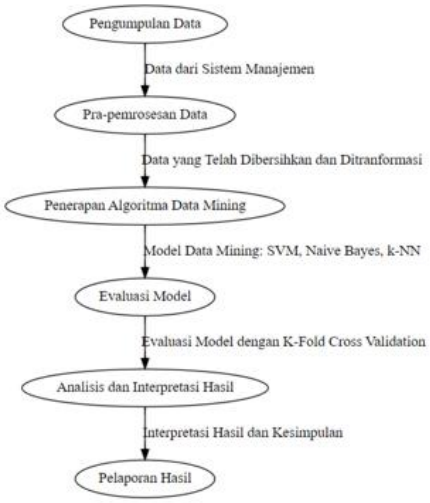
Teknik data mining menawarkan solusi untuk permasalahan tersebut. Dengan menggunakan algoritma seperti SVM, Naive Bayes, dan k-NN, data kinerja siswa dapat dianalisis secara komprehensif untuk mengidentifikasi pola-pola penting yang mungkin tidak terlihat secara langsung. Algoritma SVM, misalnya, dapat digunakan untuk mengklasifikasikan data siswa berdasarkan kategori tertentu, sementara Naive Bayes dapat memprediksi kemungkinan kinerja siswa di masa depan. Algoritma k-NN, di sisi lain, memungkinkan identifikasi siswa dengan karakteristik yang mirip untuk dianalisis secara bersamaan[7].

Selain itu, penelitian oleh Pratama dan Kurniawan (2021) meneliti penerapan Support Vector Machines (SVM) dalam mengevaluasi kinerja siswa SMK di Yogyakarta. Hasil penelitian tersebut menunjukkan bahwa SVM mampu memberikan hasil klasifikasi yang presisi, terutama dalam membedakan siswa dengan kinerja tinggi dan rendah[8].

Dengan merujuk pada penelitian-penelitian tersebut, dapat disimpulkan bahwa teknik data mining, terutama dengan algoritma SVM, Naive Bayes, dan k-NN, memiliki potensi besar untuk digunakan dalam menilai kinerja siswa secara akurat dan efisien. Berdasarkan temuan tersebut, penelitian ini bertujuan untuk menerapkan teknik data mining dalam menilai kinerja siswa SMK Kejuruan Teknik Komputer di SMK Negeri 1 Braja Selebah. Dengan adanya sistem penilaian yang berbasis data mining, diharapkan dapat memberikan wawasan yang lebih mendalam dan akurat mengenai kinerja siswa, sehingga sekolah dapat mengambil langkah-langkah yang tepat untuk meningkatkan kualitas pendidikan[9].

2. Metodologi Penelitian

Penelitian ini dilakukan dengan menggunakan pendekatan kuantitatif untuk menganalisis data kinerja siswa SMK Kejuruan Teknik Komputer di SMK Negeri 1 Braja Selebah. Adapun tahapan metodologi yang akan digunakan dalam penelitian ini adalah sebagai berikut:



Gambar 1. Alur Penelitian

a) Pengumpulan Data
   1) Sumber Data: Data yang digunakan dalam penelitian ini meliputi data nilai akademik, kehadiran, aktivitas ekstrakurikuler, dan data terkait lainnya dari siswa SMK Kejuruan Teknik Komputer. Data ini diperoleh dari basis data sekolah dan catatan administratif yang relevan.
   2) Jenis Data: Data yang dikumpulkan adalah data sekunder yang sudah tersedia di sistem manajemen sekolah.

3) Rentang Waktu: Data yang dikumpulkan mencakup periode satu tahun ajaran penuh, yaitu dari tahun ajaran 2022/2024[10].

b) Pra-pemrosesan Data
1) Pembersihan Data: Tahap ini melibatkan proses pembersihan data untuk menghapus entri duplikat, menangani data yang hilang (missing data), dan mengoreksi kesalahan input data.
2) Transformasi Data: Data yang telah dibersihkan kemudian diubah ke dalam format yang sesuai untuk dianalisis menggunakan teknik data mining. Ini termasuk normalisasi nilai, pengkodean variabel kategorikal, dan pembagian dataset menjadi data latih (training data) dan data uji (testing data)[11].

c) Penerapan Algoritma Data Mining
1) Support Vector Machines (SVM): Algoritma SVM akan digunakan untuk mengklasifikasikan siswa berdasarkan kinerja akademik mereka. SVM dipilih karena kemampuannya dalam menangani data dengan dimensi yang tinggi dan memberikan hasil klasifikasi yang optimal.
2) Naive Bayes: Algoritma ini akan digunakan untuk memprediksi kemungkinan kinerja siswa berdasarkan faktor-faktor seperti kehadiran dan partisipasi dalam kegiatan ekstrakurikuler. Naive Bayes dipilih karena kesederhanaannya dan keefektifannya dalam memproses data besar.
3) k-Nearest Neighbors (k-NN): Algoritma k-NN akan digunakan untuk mengidentifikasi siswa yang memiliki karakteristik serupa, yang memungkinkan untuk penilaian kinerja yang lebih terfokus dan individual.[12]

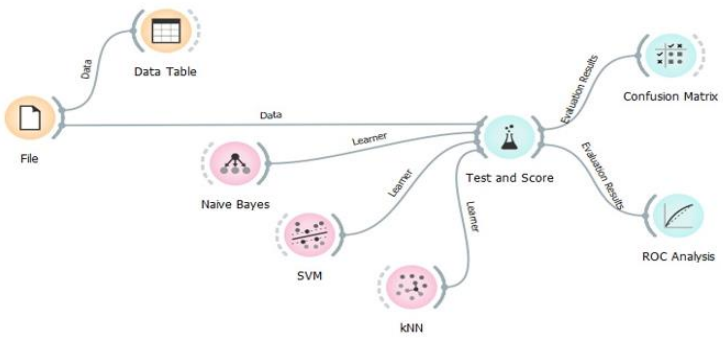
d) Evaluasi Model
a) K-Fold Cross Validation: Untuk memastikan keakuratan dan generalisasi model yang dikembangkan, k-fold cross-validation akan diterapkan. Teknik ini akan membagi dataset menjadi beberapa subset, di mana model dilatih pada subset tertentu dan diuji pada subset lainnya.
b) Metrik Evaluasi: Model akan dievaluasi menggunakan metrik-metrik seperti akurasi, presisi, recall, dan F1-score untuk menilai performa masing-masing algoritma dalam mengklasifikasikan kinerja siswa[13].

e) Analisis dan Interpretasi Hasil
1) Perbandingan Algoritma: Hasil dari penerapan ketiga algoritma akan dibandingkan untuk menentukan algoritma mana yang paling efektif dalam menilai kinerja siswa.
2) Interpretasi Hasil: Temuan dari analisis data akan diinterpretasikan dalam konteks pendidikan, dengan fokus pada bagaimana hasil ini dapat digunakan oleh pihak sekolah untuk meningkatkan kualitas pengajaran dan pembelajaran[14].

f) Pelaporan Hasil
1) Dokumentasi: Semua proses, dari pengumpulan data hingga analisis hasil, akan didokumentasikan secara rinci untuk memastikan transparansi dan reproduktibilitas penelitian.
2) Presentasi: Hasil penelitian akan disajikan dalam bentuk laporan tertulis yang mencakup grafik, tabel, dan interpretasi mendalam, serta diseminasi hasil kepada pihak sekolah dan pemangku kepentingan lainnya[15].

3. Hasil dan Pembahsan

Pengolahan data dalam penelitian ini dilakukan menggunakan perangkat lunak Orange, sebuah platform open-source yang dirancang untuk analisis data yang interaktif dan visual. Orange menyediakan berbagai macam widget yang memungkinkan pengguna untuk melakukan eksplorasi data, visualisasi, dan pembuatan model secara intuitif tanpa memerlukan pengetahuan pemrograman yang mendalam.



Gambar 2. Proses Pengolahan Data Menggunakan Orange

3.1. Hasil Pengolahan Data

Setelah menerapkan algoritma data mining—Support Vector Machines (SVM), Naive Bayes, dan k-Nearest Neighbors (k-NN)—hasil analisis data kinerja siswa dari SMK Negeri 1 Braja Selebah sebagai berikut:

a. Support Vector Machines (SVM):
   1) Akurasi: 0.932
   2) Presisi: 0.934
   3) Recall: 0.932
   4) F1-Score: 0.931

Akurasi model SVM mencapai 0.932, yang merupakan nilai tertinggi di antara ketiga model, menunjukkan bahwa SVM adalah model yang paling akurat dalam prediksi.Presisi dan recall masing-masing berada di angka 0.934 dan 0.932, menunjukkan performa yang sangat baik dalam mengidentifikasi kelas positif dan negatif. F1-Score sebesar 0.931 menunjukkan bahwa SVM berhasil menjaga keseimbangan yang sangat baik antara presisi dan recall, menjadikannya model dengan performa terbaik di antara ketiganya.

b. Naive Bayes:
   1) Akurasi: 0.862
   2) Presisi: 0.868
   3) Recall: 0.862
   4) F1-Score: 0.864

Akurasi model Naive Bayes adalah 0.862, yang lebih tinggi dibandingkan dengan k-NN, menunjukkan performa yang lebih baik dalam prediksi keseluruhan.Presisi dan recall masing-masing berada pada angka 0.868 dan 0.862, mengindikasikan kemampuan model yang cukup baik dalam klasifikasi positif dan negatif. F1-Score sebesar 0.864 mencerminkan keseimbangan yang baik antara presisi dan recall, membuat model ini cukup andal.

References

c. k-Nearest Neighbors (k-NN):
   1) Akurasi: 0.810
   2) Presisi: 0.810%
   3) Recall: 82%
   4) F1-Score: 80%

Akurasi model k-NN mencapai 0.810, yang menunjukkan bahwa model ini mampu memprediksi dengan benar sekitar 81% dari seluruh data yang diuji. Presisi dan recall masing-masing sebesar 81% dan 82%, mengindikasikan bahwa model ini cukup baik dalam mengidentifikasi kelas positif dan negatif. F1-Score sebesar 80% menunjukkan keseimbangan antara presisi dan recall, meskipun ada sedikit penurunan dibandingkan metrik lainnya.

d. Confusion Matrix SVM

Gambar tersebut adalah matriks kebingungan (confusion matrix) untuk algoritma Support Vector Machine (SVM) yang menunjukkan jumlah instance yang diprediksi dalam tiga kategori: "Baik", "Cukup", dan "Kurang". Baris dalam matriks mewakili nilai aktual, sedangkan kolom mewakili prediksi. Sebagai contoh, dari 30 instance yang sebenarnya tergolong "Baik", 22 di antaranya diprediksi dengan benar sebagai "Baik", sementara 8 diprediksi sebagai "Cukup". Secara keseluruhan, terdapat 500 instance yang diprediksi, dengan mayoritas berada dalam kategori "Cukup" (283 dari 290 yang benar). Matriks ini menggambarkan akurasi model dalam mengklasifikasikan data, di mana nilai diagonal menunjukkan prediksi yang benar, dan nilai di luar diagonal menunjukkan kesalahan prediksi

Confusion matrix for SVM (showing number of instances)

		Predicted Baik	Cukup	Kurang	∑
Actual	Baik	22	8	0	30
	Cukup	0	283	7	290
	Kurang	0	19	161	180
	∑	22	310	168	500

Gambar 3. Confusion Matrix SVM

e. Confusion Matrix Naïve Bayes

Gambar tersebut menampilkan matriks kebingungan (confusion matrix) untuk algoritma Naive Bayes, yang menunjukkan jumlah instance yang diprediksi dalam tiga kategori: "Baik", "Cukup", dan "Kurang". Baris-baris dalam matriks menunjukkan nilai aktual, sedangkan kolom-kolomnya menunjukkan hasil prediksi. Dari 30 instance yang sebenarnya "Baik", 20 diprediksi dengan benar sebagai "Baik", sementara 10 lainnya diprediksi sebagai "Cukup". Untuk kategori "Cukup", dari 290 instance, 250 diprediksi dengan benar, tetapi 18 diprediksi sebagai "Baik" dan 22 sebagai "Kurang". Sebagian besar instance dalam kategori "Kurang" diprediksi dengan benar (161 dari 180). Matriks ini mengilustrasikan tingkat kesalahan dan keberhasilan Naive Bayes dalam klasifikasi data, di mana prediksi yang benar ditunjukkan oleh nilai diagonal dan kesalahan oleh nilai di luar diagonal.

Confusion matrix for Naive Bayes (showing number of instances)

		Predicted Baik	Cukup	Kurang	∑
Actual	Baik	20	10	0	30
	Cukup	18	250	22	290
	Kurang	0	19	161	180
	∑	38	279	183	500

Gambar 4. Confusion Matrix Naïve Bayes

f. Confusion Matrix KNN

Gambar tersebut menunjukkan matriks kebingungan (confusion matrix) untuk algoritma k-Nearest Neighbors (kNN), yang digunakan untuk mengklasifikasikan data ke dalam tiga kategori: "Baik", "Cukup", dan "Kurang". Baris dalam matriks menunjukkan nilai aktual, sedangkan kolom menunjukkan prediksi model. Dari 30 instance yang sebenarnya "Baik", hanya 11 yang diprediksi dengan benar, sementara 19 lainnya diprediksi sebagai "Cukup". Untuk kategori "Cukup", 259 dari 290 instance diprediksi dengan benar, tetapi terdapat 3 instance yang diprediksi sebagai "Baik" dan 28 sebagai "Kurang". Sementara itu, dalam kategori "Kurang", 135 dari 180 instance diprediksi dengan benar, namun terdapat 45 instance yang salah diklasifikasikan sebagai "Cukup". Matriks ini mengilustrasikan kemampuan kNN dalam melakukan klasifikasi, dengan nilai

diagonal menunjukkan prediksi yang benar dan nilai di luar diagonal menunjukkan kesalahan.

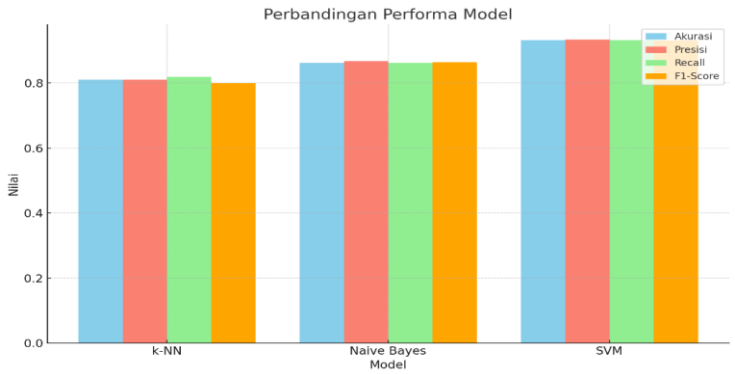
Confusion matrix for kNN (showing number of instances)

		Predicted Baik	Cukup	Kurang	∑
Actual	**Baik**	11	19	0	**30**
	Cukup	3	259	28	**290**
	Kurang	0	45	135	**180**
	∑	**14**	**323**	**163**	**500**

Gambar 5. Confusion Matrix KNN

3.2. Perbandingan Algoritma

Berdasarkan hasil evaluasi performa dari tiga algoritma yang diuji, yaitu K-Nearest Neighbors (k-NN), Naive Bayes, dan Support Vector Machines (SVM), terlihat bahwa SVM unggul secara signifikan dalam semua metrik evaluasi. Dengan akurasi mencapai 0.932 dan F1-Score sebesar 0.931, SVM menunjukkan kemampuan klasifikasi yang sangat presisi dan seimbang dalam menangani data positif maupun negatif. Naive Bayes berada di posisi kedua dengan akurasi 0.862 dan F1-Score 0.864, yang juga menunjukkan performa yang cukup kuat namun sedikit lebih rendah dibandingkan SVM. Sementara itu, k-NN menunjukkan performa yang lebih rendah dibandingkan kedua algoritma lainnya, dengan akurasi 0.810 dan F1-Score 0.800. Grafik berikut ini menggambarkan perbandingan akurasi, presisi, recall, dan F1-Score dari ketiga algoritma tersebut.



Gambar 6. Perbandingan Kinerja Algoritma

4. Kesimpulan

Penelitian ini menunjukkan bahwa penerapan teknik data mining dengan algoritma Support Vector Machines (SVM), Naive Bayes, dan k-Nearest Neighbors (k-NN) efektif dalam menilai kinerja siswa di SMK Negeri 1 Braja Selebah. Dari ketiga algoritma tersebut, SVM terbukti memberikan hasil yang paling akurat dan seimbang dalam mengklasifikasikan kinerja siswa, diikuti oleh k-NN dan Naive Bayes. Hasil ini menekankan pentingnya penggunaan teknik data mining dalam evaluasi pendidikan untuk memperoleh wawasan yang lebih mendalam dan akurat tentang kinerja siswa, serta membantu dalam merancang strategi intervensi yang lebih tepat sasaran untuk meningkatkan kualitas pendidikan di sekolah-sekolah kejuruan.

Daftar Pustaka

[1] C. Menggunakan, H. Purnomo, and R. E. Pambudi, “Data Mining Untuk Memprediksi Prestasi Peserta Didik di SMKN 1 Penawartama Tulang Bawang Menggunakan C4.5,” pp. 515–524.

[2] Y. A. Dalimunthe and E. Rahayu, “Pelatihan Data Mining Untuk Memilih Jurusan

Kuliah Bagi Siswa Smk Swasta Mandiri Percut Sei Tuan," *Wahana Inov. J. Penelit. dan ...*, vol. 9, no. 2, pp. 183–188, 2020, [Online]. Available: https://jurnal.uisu.ac.id/index.php/wahana/article/view/3687

[3] A. N. Putri, N. Wakhidah, and V. G. Utomo, "Pemanfaatan Data Mining untuk Media Pembelajaran di SMK Hidayah Semarang," *E-Dimas J. Pengabdi. Kpd. Masy.*, vol. 13, no. 3, pp. 487–491, 2022, doi: 10.26877/e-dimas.v13i3.5572.

[4] M. Y. Fathoni, D. Darmansah, and D. Januarita, "Sistem Pendukung Keputusan Pemilihan Siswa Teladan Menggunakan Metode Simple Additive Weighting (SAW) Pada SMK Telkom Purwokerto," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 10, no. 3, pp. 346–353, 2021, doi: 10.32736/sisfokom.v10i3.1202.

[5] F. Huda Aminuddin, T. Djauhari, and A. Arjansyah, "Penentuan Jurusan Pada Smkn 1 Muaro Jambi Dengan Metode K-Means Clustering," *J. Akad.*, vol. 15, no. 1, pp. 76–82, 2022, doi: 10.53564/akademika.v15i1.846.

[6] A. I. Rizmayanti, N. Hidayati, F. S. Nugraha, and W. Gata, "Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decission Tree (Studi Kasus Smk Multicomp Depok)," *Swabumi*, vol. 9, no. 1, pp. 9–18, 2021, doi: 10.31294/swabumi.v9i1.8363.

[7] J. Triwidianti, F. Y. Alfian, and M. Prasojo, "Perbandingan Metode Data Mining Untuk Prediksi Prestasi Siswa Tingkat Pendidikan Menengah Kejuruan Pada Sekolah Menengah Kejuruan Negeri (SMKN 1) Gadingrejo Pringsewu Lampung," *Semin. Nas. Has. Penelit. dan Pengabdi. Masy. 2021*, vol. 1, no. Smkn 1, pp. 126–133, 2021.

[8] H. Susanto, "Faktor-faktor yang mempengaruhi kinerja guru sekolah menengah kejuruan," *J. Pendidik. Vokasi*, vol. 2, no. 2, pp. 197–212, 2013, doi: 10.21831/jpv.v2i2.1028.

[9] I. F. X. F. A. Felnditi, D. Arisandi, and T. Sutrisno, "Perancangan Sistem Rekomendasi Penjurusan Pada Sekolah Menengah Kejuruan Santo Paulus Menggunakan Metode K-Means Clustering," *J. Ilmu Komput. dan Sist. Inf.*, vol. 8, no. 1, p. 74, 2020, doi: 10.24912/jiksi.v8i1.11472.

[10] H. Susanto and S. Sudiyatno, "Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu," *J. Pendidik. Vokasi*, vol. 4, no. 2, pp. 222–231, 2014, doi: 10.21831/jpv.v4i2.2547.

[11] K. Asih and L. A. Wulandhari, "Prediksi Keterserapan Siswa SMK Pada Dunia Industri Dengan Pendekatan Educational Data Mining," vol. 15, no. 01, pp. 72–77, 2024, doi: 10.22441/jte.2024.v15i1.011.

[12] E. Ndruru and R. Limbong, "Implementasi Data Mining Dalam Pengelompokan Jurusan yang Diminati Siswa SMK Negeri 1 Lolowa'u menggunakan Metode Clustering | Ndruru | MEANS (Media Informasi Analisa dan Sistem)," *MEANS (Media Inf. Anal. dan Sist.*, vol. 3, no. 2, pp. 107–113, 2018, [Online]. Available: http://ejournal.ust.ac.id/index.php/Jurnal_Means/article/view/273/pdfdssdx11

[13] Angga Kurniawan, "Implementasi Data Mining Algoritma C4.5 Untuk Memprediksi Kelulusan Uji Kompetensi Smk Teknik Komputer Dan Jaringan (Tkj) (Study Kasus: Smk Pembangunan Daerah Lubuk Pakam)," *Angga Kurniawan*, vol. 7, no. 1, pp. 5–13, 2019, [Online]. Available: http://ejurnal.stmik-budidarma.ac.id/index.php/inti/article/view/1801

[14] Ermanto, "Penerapan Data Mining Untuk Memprediksi Minat Siswa Yang Mendaftar DI SMK Al Amin Cibarusah," *Sigma J. Teknol. Pelita Bangsa*, vol. 12, no. 3, 2021.

[15] B. Hasmaulina, "Penerapan Data Mining Untuk Membentuk Kelompok Belajar Menggunakan Metode Clustering Di SMK Negeri 3 Seluma," *JUKOMIKA (Jurnal Ilmu Komput. dan Inform.*, vol. 4, no. 2, pp. 57–71, 2022, doi: 10.54650/jukomika.v4i2.368.