

BAB 2

TINJAUAN PUSTAKA

Rendahnya optimalisasi pemanfaatan data *Tracer Study* dalam mendukung pengelolaan lulusan Sekolah Menengah Kejuruan (SMK), khususnya prediksi lulusan pada jalur Bekerja, Melanjutkan Pendidikan, dan Berwirausaha (BMW). Meskipun sistem pendataan lulusan sudah tersedia melalui Bursa Kerja Khusus (BKK), data yang dihasilkan belum dimanfaatkan secara optimal oleh BKK dalam layanan bimbingan karir lulusan.

2.1 LANDASAN TEORI

2.1.1 Sekolah Menengah Kejuruan (SMK)

Pendidikan vokasi memiliki orientasi yang lebih luas daripada sekadar menyiapkan peserta didik untuk bekerja. Konsep BMW (Bekerja, Melanjutkan, dan Wirausaha) menggambarkan arah utama keberhasilan lulusan SMK dalam memulai karier sesuai minat, kemampuan, dan peluang yang tersedia. Dengan demikian, kualitas pendidikan kejuruan dapat diukur dari seberapa baik sekolah mampu memfasilitasi lulusan untuk terserap didunia kerja, melanjutkan studi ke jenjang yang lebih tinggi, atau mengembangkan usaha mandiri secara berkelanjutan.

Karakteristik Sosial-Ekonomi merupakan kombinasi dari faktor-faktor sosial (seperti pekerjaan orang tua, gaya hidup, serta lingkungan sosial) dan faktor ekonomi (seperti pendapatan, kepemilikan aset, dan kondisi kesejahteraan keluarga). Faktor-faktor ini memiliki pengaruh yang signifikan terhadap prestasi belajar siswa serta kesiapan kerja lulusan SMK.

Salah satu faktor penting lulusan merupakan karakteristik sosial–ekonomi yang mempengaruhi arah keterserapan lulusan SMK dalam konteks Bekerja, Melanjutkan Pendidikan, dan Wirausaha (BMW). Faktor-faktor sosial, seperti tingkat pendidikan dan jenis pekerjaan orang tua, pola dukungan keluarga, serta lingkungan sosial, berperan dalam membentuk motivasi, aspirasi, dan kesiapan lulusan dalam menentukan pilihan karier. Sementara itu, faktor ekonomi, termasuk kepemilikan aset, akses terhadap sarana penunjang belajar, serta kondisi kesejahteraan keluarga, turut mempengaruhi peluang lulusan untuk bekerja,

melanjutkan pendidikan, atau berwirausaha. Oleh karena itu, analisis keterkaitan antara karakteristik sosial–ekonomi dan arah keterserapan lulusan menjadi penting untuk memperoleh pemahaman yang lebih komprehensif dan berbasis data mengenai faktor-faktor dominan yang memengaruhi keberhasilan lulusan SMK.

2.1.2 Sosial -Ekonomi

Karakteristik Sosial -Ekonomi merupakan kombinasi dari faktor-faktor sosial (seperti Pekerjaan ayah, Pekerjaan ibu, Status dalam keluarga) dan faktor ekonomi (seperti pendapatan, kepemilikan aset, dan kondisi kesejahteraan keluarga). Faktor-faktor ini memiliki pengaruh yang signifikan terhadap prestasi belajar siswa serta kesiapan kerja lulusan SMK.

Menurut Kuntoro (2012), variabel-variabel sosial-ekonomi dapat digunakan sebagai prediktor performa akademik dan keberhasilan karier siswa, karena siswa dari latar belakang ekonomi yang stabil cenderung memiliki akses lebih besar terhadap sumber belajar, pelatihan, dan dukungan emosional. Sebaliknya, kondisi ekonomi yang rendah dapat menjadi hambatan bagi siswa untuk mencapai hasil optimal, baik di sekolah maupun di dunia kerja.

2.1.3 Rough Set

Rough Set Theory merupakan metode matematis yang digunakan untuk menangani ketidakpastian, ketidaklengkapan, dan ketidakkonsistenan informasi dalam suatu dataset. Metode ini digunakan untuk menganalisis hubungan antarobjek berdasarkan kesamaan nilai atribut tanpa memerlukan informasi tambahan, seperti probabilitas awal. Dalam proses analisis data, Rough Set dapat digunakan untuk mengidentifikasi atribut-atribut yang memiliki pengaruh terhadap atribut keputusan serta menghilangkan atribut yang redundan tanpa mengurangi informasi penting yang terkandung dalam dataset (Yue-hua & Shan, 2011).

Konsep utama dalam Rough Set adalah indiscernibility relation, yaitu hubungan ketidakterbedaan antara dua objek yang memiliki nilai sama pada sekumpulan atribut tertentu. Objek-objek yang tidak dapat dibedakan berdasarkan atribut kondisi akan dikelompokkan ke dalam kelas ekuivalensi. Melalui kelas ekuivalensi tersebut, Rough Set membentuk dua pendekatan, yaitu lower

approximation dan upper approximation. Lower approximation merupakan kumpulan objek yang secara pasti termasuk ke dalam suatu kelas keputusan, sedangkan upper approximation merupakan kumpulan objek yang mungkin termasuk ke dalam kelas keputusan tersebut. Selisih antara upper approximation dan lower approximation disebut boundary region, yang menggambarkan tingkat ketidakpastian dalam proses klasifikasi.

Rough Set memiliki konsep reduct dan core. Reduct merupakan sekumpulan atribut minimum yang mampu mempertahankan kemampuan klasifikasi yang sama dengan keseluruhan atribut. Reduct digunakan untuk menghilangkan atribut yang tidak relevan atau redundan tanpa kehilangan informasi penting mengenai keputusan. Dan core merupakan atribut yang selalu muncul pada setiap reduct dan dianggap sebagai atribut yang paling penting dan tidak dapat dihilangkan. Proses reduksi atribut tersebut dapat menyederhanakan dataset, mengurangi kompleksitas komputasi, dan meningkatkan interpretabilitas model. Selain melakukan reduksi atribut, Rough Set menghasilkan aturan keputusan dalam bentuk hubungan kondisi dan keputusan. Aturan tersebut dinyatakan dalam bentuk if-then. Apabila atribut sosial-ekonomi tertentu memenuhi suatu kondisi, maka lulusan dapat diklasifikasikan ke dalam kategori Bekerja, Melanjutkan Pendidikan, Wirausaha, atau Lainnya. Aturan keputusan yang dihasilkan dapat dievaluasi berdasarkan nilai support, confidence, dan coverage untuk mengetahui kekuatan dan tingkat keandalan aturan dalam merepresentasikan pola pada dataset.

Pada penelitian ini, Rough Set digunakan sebagai tahap awal pemodelan untuk menyeleksi atribut yang paling informatif dari data sosial-ekonomi dan data Tracer Study lulusan. Proses dilakukan dengan mengidentifikasi atribut dominan dan mengeliminasi atribut yang redundan, sehingga model klasifikasi menjadi lebih sederhana tanpa kehilangan informasi esensial. Atribut hasil seleksi Rough Set selanjutnya digunakan dalam proses pembentukan model Decision Tree. Dengan demikian, Rough Set berperan dalam meningkatkan efisiensi pemodelan sekaligus membantu menghasilkan pola klasifikasi yang lebih mudah dipahami (Sahid et al., 2020).

Penerapan Rough Set dalam bidang pendidikan telah dilakukan pada berbagai penelitian. Prajana (2016) menerapkan Rough Set untuk memprediksi tingkat kelulusan siswa dalam Ujian Nasional di SMA Negeri 5 Kota Banda Aceh. Penelitian tersebut menunjukkan bahwa Rough Set dapat digunakan untuk menganalisis data pendidikan, menentukan atribut yang berkaitan dengan hasil kelulusan, serta menghasilkan aturan keputusan sebagai dasar prediksi. Hal ini mendukung penggunaan Rough Set dalam penelitian keterserapan lulusan karena data pendidikan umumnya memiliki banyak atribut kategorikal, kemungkinan redundansi, dan pola hubungan yang perlu dijelaskan secara interpretatif.

Berdasarkan karakteristik tersebut, Rough Set dipilih karena memiliki beberapa kelebihan, yaitu tidak memerlukan asumsi distribusi statistik, mampu mengolah data kategorikal, dapat melakukan reduksi atribut, dan menghasilkan aturan keputusan yang mudah diinterpretasikan.

2.1.4 Decision Tree

Decision Tree merupakan salah satu algoritma machine learning yang digunakan untuk melakukan klasifikasi dan prediksi berdasarkan serangkaian aturan keputusan yang direpresentasikan dalam bentuk struktur pohon. Klasifikasi merupakan proses dalam data mining yang bertujuan untuk memetakan suatu objek atau data ke dalam kelas tertentu berdasarkan karakteristik atribut dan pola yang dipelajari dari data latih. Model klasifikasi yang telah terbentuk selanjutnya digunakan untuk memprediksi kelas dari data baru secara objektif (Adli & Sahid, 2021).

Struktur Decision Tree terdiri atas root node, internal node, branch, dan leaf node. Root node merupakan simpul awal yang memuat seluruh data dan atribut yang akan dianalisis. Internal node merupakan simpul yang menunjukkan proses pengujian terhadap suatu atribut, sedangkan branch merupakan cabang yang merepresentasikan hasil dari pengujian tersebut. Leaf node merupakan simpul akhir yang menunjukkan kelas keputusan atau hasil prediksi. Struktur ini menghasilkan aturan keputusan berbentuk if-then, sehingga proses klasifikasi dapat ditelusuri dan dipahami dengan mudah. Pembentukan Decision Tree dilakukan dengan membagi

data secara berulang berdasarkan atribut yang memberikan pemisahan terbaik terhadap kelas target. Pemilihan atribut pada setiap simpul dapat dilakukan menggunakan ukuran tertentu, seperti Gini Index, Entropy, atau Information Gain.

Proses pembentukan Decision Tree dimulai dengan menempatkan seluruh data latih pada root node. Algoritma memilih atribut terbaik untuk membagi data menjadi beberapa kelompok yang lebih homogen. Proses dilakukan secara berulang pada setiap cabang hingga memenuhi kriteria penghentian tertentu, seperti kedalaman maksimum pohon, jumlah minimum data pada setiap simpul, atau kondisi ketika data pada suatu simpul telah berada dalam kelas yang sama. Hasil akhirnya berupa pohon keputusan yang dapat digunakan untuk mengklasifikasikan data baru berdasarkan jalur keputusan dari root node hingga leaf node.

Dalam penelitian ini, Decision Tree digunakan untuk mengklasifikasikan arah keterserapan lulusan berdasarkan data sosial-ekonomi, data akademik, dan data Tracer Study. Kelas keputusan yang digunakan terdiri atas Bekerja, Melanjutkan Pendidikan, Wirausaha, dan Lainnya. Proses klasifikasi dilakukan dengan mempelajari pola hubungan antara atribut kondisi, seperti asal sekolah, alamat, pekerjaan orang tua, status dalam keluarga, kepemilikan sarana transportasi, akses teknologi dan internet, nilai akademik, serta karakteristik Tracer Study, dengan status keterserapan lulusan. Penggunaan Decision Tree dalam klasifikasi arah karier lulusan didasarkan pada kemampuannya menangani permasalahan klasifikasi multikelas dan menghasilkan pola keputusan yang dapat divisualisasikan. Suryaningrat et al. (2025) menggunakan algoritma klasifikasi untuk mengidentifikasi arah karier lulusan berdasarkan karakteristik data yang tersedia.

Decision Tree dipilih karena memiliki beberapa kelebihan. Algoritma ini mampu mengolah data numerik maupun kategorikal, tidak memerlukan asumsi distribusi statistik tertentu, menghasilkan model yang transparan, serta menyediakan aturan keputusan yang mudah ditelusuri. Visualisasi dalam bentuk pohon memungkinkan pihak BKK SMKN 1 Bangkinang mengetahui atribut yang digunakan dalam setiap proses pengambilan keputusan.

2.1.5 Hybrid Rough Set – Decision Tree

Metode hybrid merupakan pendekatan yang menggabungkan dua atau lebih algoritma untuk memanfaatkan keunggulan masing-masing metode serta mengurangi keterbatasan yang dimiliki apabila algoritma digunakan secara terpisah. Pada penelitian ini, metode hybrid dibentuk melalui penggabungan Rough Set dan Decision Tree. Rough Set berperan pada tahap seleksi dan reduksi atribut, sedangkan Decision Tree berperan sebagai algoritma klasifikasi akhir. Penggabungan tersebut dilakukan secara berurutan atau sequential hybrid, yaitu keluaran dari proses Rough Set digunakan sebagai masukan dalam pembentukan model Decision Tree.

Dasar penggunaan Rough Set dalam penelitian ini didasarkan pada karakteristik data akademik, sosial-ekonomi, dan Tracer Study yang terdiri atas banyak atribut kategorikal, memiliki kemungkinan keterkaitan antaratribut, serta mengandung atribut yang kurang relevan atau redundan. Penggunaan seluruh atribut secara langsung dapat meningkatkan kompleksitas model dan menyebabkan pola klasifikasi sulit diinterpretasikan. Rough Set dipilih karena mampu mengevaluasi ketergantungan atribut terhadap kelas keputusan serta menghasilkan reduct tanpa memerlukan informasi awal berupa probabilitas. Atribut yang digunakan dalam proses klasifikasi merupakan atribut yang masih memiliki kemampuan untuk membedakan status keterserapan lulusan.

Decision Tree dipilih sebagai algoritma klasifikasi karena mampu menangani data kategorikal dan numerik setelah melalui tahap preprocessing, serta menghasilkan model dalam bentuk struktur pohon dan aturan Keputusan. Karakteristik tersebut sesuai dengan kebutuhan penelitian karena hasil klasifikasi tidak hanya dituntut memiliki performa yang baik, tetapi juga harus mudah dipahami dan dimanfaatkan oleh pihak BKK. Penggabungan Rough Set dan Decision Tree dilakukan karena penggunaan Decision Tree terhadap seluruh atribut berpotensi menghasilkan pohon yang kompleks dan mengalami overfitting. Oleh karena itu, Rough Set digunakan untuk menyederhanakan ruang atribut, sedangkan Decision Tree digunakan untuk membentuk pola klasifikasi berdasarkan atribut hasil reduct.

Rough Set menganalisis ketergantungan antaratribut, mengidentifikasi atribut yang relevan, serta menghilangkan atribut redundan tanpa mengurangi kemampuan dataset dalam membedakan kelas keputusan. Proses tersebut menghasilkan reduct. Atribut-atribut yang terdapat dalam reduct dianggap telah mampu mempertahankan informasi klasifikasi yang dimiliki oleh keseluruhan atribut awal. Rough Set digunakan untuk menyederhanakan ruang fitur sebelum proses klasifikasi dilakukan.

Decision Tree menggunakan atribut hasil reduksi Rough Set untuk membentuk model klasifikasi. Algoritma ini membagi data berdasarkan atribut yang memberikan pemisahan terbaik terhadap kelas target. Proses pembagian dilakukan secara bertahap dari root node hingga leaf node dengan menggunakan ukuran seperti Gini Index atau Entropy. Hasilnya berupa struktur pohon dan aturan keputusan berbentuk if-then yang dapat digunakan untuk memprediksi kelas suatu data baru.

Tahapan metode Hybrid Rough Set-Decision Tree, data akademik, sosial-ekonomi, dan Tracer Study dikumpulkan serta diintegrasikan berdasarkan identitas lulusan. Kemudian dilakukan tahap preprocessing, seperti penanganan data kosong, transformasi data kategorikal, diskretisasi atribut numerik, dan pembentukan kelas target. Rough Set digunakan untuk membentuk decision system yang terdiri dari atribut kondisi dan atribut Keputusan. Selanjutnya dilakukan analisis ketergantungan atribut dan pembentukan reduct untuk memperoleh atribut yang paling relevan. Atribut hasil reduct digunakan sebagai input dalam pembentukan model Decision Tree. Model Decision Tree dilatih untuk mengklasifikasikan arah keterserapan lulusan. Terakhir, performa model hybrid dievaluasi menggunakan accuracy, precision, recall, F1-score, confusion matrix, dan ROC-AUC.

Pada penelitian ini, atribut kondisi terdiri atas karakteristik sosial-ekonomi dari data akademik dan data Tracer Study lulusan, sedangkan atribut keputusan berupa status keterserapan lulusan. Status tersebut diklasifikasikan ke dalam empat kelas, yaitu Bekerja, Melanjutkan Pendidikan, Wirausaha, dan Lainnya. Rough Set digunakan untuk menentukan atribut yang memiliki kemampuan paling besar dalam membedakan keempat kelas tersebut, sedangkan Decision Tree digunakan untuk membentuk pola dan aturan klasifikasi berdasarkan atribut yang telah dipilih.

Penggunaan Rough Set sebelum Decision Tree bertujuan untuk mengurangi jumlah atribut yang kurang relevan atau redundan.

Penggabungan Rough Set dan Decision Tree juga memberikan manfaat dari sisi interpretabilitas. Rough Set dapat menunjukkan atribut yang termasuk dalam reduct dan menghasilkan aturan keputusan, sedangkan Decision Tree dapat memperlihatkan urutan atribut yang digunakan dalam proses klasifikasi melalui struktur pohon. Kombinasi ini diharapkan dapat menghasilkan model klasifikasi arah keterserapan lulusan SMKN 1 Bangkinang yang lebih sederhana, dan mudah dimanfaatkan oleh pihak BKK.

2.2 PENELITIAN TERDAHULU

Penelitian mengenai prediksi klasifikasi lulusan siswa SMK telah banyak dilakukan dengan pendekatan dan metode yang beragam. Beberapa studi terdahulu telah memanfaatkan algoritma machine learning maupun Rough Set, Decision Tree, variable-variabel karakteristik Sosial Ekonomi untuk tujuan tersebut. Namun, sebagian besar penelitian tersebut masih memiliki keterbatasan dari segi integrasi metode, optimasi, cakupan variabel, serta interpretasi hasil yang belum aplikatif. Berikut beberapa peneliti terdahulu :

Penelitian pertama memiliki latar belakang untuk membuka peluang bagi dunia Pendidikan hasil belajar dan kelulusan siswa. Sekolah menengah atas (SMA) maupun kejuruan (SMK) menghadapi tantangan untuk meningkatkan persentasi kelulusan siswa pada Ujian Nasional, yang sering kali dipengaruhi oleh banyak faktor seperti nilai pra-ujian, prestasi akademik, dan tingkat kedisiplinan. Penelitian ini bertujuan untuk menerapkan teori Rough Set dalam memprediksi tingkat kelulusan siswa pada Ujian Nasional di SMA Negeri 5 Kota Banda Aceh (Prajana dkk., 2016). Penelitian ini menggunakan pendekatan teori Rough Set, yang merupakan salah satu teknik data mining untuk menemukan aturan keputusan (decision rules) berdasarkan data siswa. Data yang digunakan mencakup nilai rata-rata siswa, nilai pra-UN, serta nilai UN pada SMA Negeri 5 Banda Aceh. Dari hasil analisis pada jurnal ini jika nilai pra-UN dan rata-rata nilai UN berada di atas ambang batas tertentu, maka siswa diprediksi “lulus”, sedangkan jika nilai

keduanya rendah, siswa diprediksi “tidak lulus”. Model yang dibangun menggunakan teori Rough Set menghasilkan akurasi tinggi dan tingkat kesalahan rendah, serta mampu menjelaskan hubungan antar variabel secara logis tanpa memerlukan asumsi probabilistik. Dengan demikian, metode Rough Set dinilai cocok untuk membantu sekolah dalam memantau, mengevaluasi, dan memprediksi kelulusan siswa secara objektif dan transparan. Penelitian ini memiliki keunggulan pada penggunaan teori Rough Set yang mampu menghasilkan aturan keputusan yang transparan dan mudah dipahami. Seleksi atribut yang dilakukan membantu mengurangi kompleksitas data dan meningkatkan akurasi prediksi. Seleksi atribut yang diterapkan dalam penelitian tersebut mampu menekan kompleksitas data dan memberikan dampak positif terhadap akurasi model yang dihasilkan. Namun demikian, kajian ini masih berfokus pada aspek kelulusan akademik dan belum mengakomodasi pengaruh faktor sosial-ekonomi maupun kondisi lulusan setelah meninggalkan bangku sekolah. Di samping itu, hasil penelitian belum dimanfaatkan secara langsung sebagai dasar pengambilan keputusan strategis bagi lembaga pendidikan, terutama dalam pengelolaan data lulusan dan layanan bimbingan karier. Oleh sebab itu, diperlukan pengembangan penelitian yang mengintegrasikan variabel sosial-ekonomi serta mengarahkan analisis pada pemetaan keterserapan lulusan ke dalam jalur bekerja, melanjutkan pendidikan, dan berwirausaha.

Penelitian selanjutnya menggunakan Decision Tree yang bertujuan meningkatkan tingkat kelulusan siswa agar dapat bersaing di dunia kerja. Penelitian ini untuk membandingkan kinerja algoritma Decision Tree C4.5 dan Naïve Bayes dalam menganalisis dan memprediksi tingkat kelulusan siswa di SMK Swadhipa 2 Natar Kabupaten Lampung Selatan. Melalui pendekatan algoritma Decision Tree C4.5 dan Naïve Bayes, penelitian ini berfokus untuk menemukan model klasifikasi yang paling efektif dalam memprediksi kelulusan siswa. Hasil analisis Decision Tree C4.5 mencapai akurasi 80%, presisi 82%, recall 75%, dan F1-score 78%, sedangkan Naïve Bayes memiliki akurasi 75%, presisi 78%, recall 80%, dan F1-score 76%. Penelitian ini membantu pihak sekolah dapat melakukan intervensi dini terhadap siswa yang berisiko tidak lulus (Susanti dkk., 2023). Focus penelitian ini

masih pada prediksi kelulusan dan belum mengkaji kondisi lulusan setelah menyelesaikan Pendidikan. Penelitian ini juga belum menerapkan tahapan seleksi atribut untuk menyederhanakan data dan meningkatkan efisiensi model.

Penelitian berikutnya adalah memprediksi kemampuan belajar dan kemungkinan keberhasilan kelulusan siswa berdasarkan faktor-faktor sosial-ekonomi dan kedisiplinan. Hasil menunjukkan bahwa faktor ekonomi orang tua merupakan variabel paling berpengaruh terhadap keberhasilan akademik siswa (Kuntoro & Sudarwanto, 2017). Tujuan utama penelitian yang berjudul *Prediction of Student Performance Using Decision Tree C4.5 Algorithm* adalah memprediksi kemampuan belajar dan kemungkinan keberhasilan kelulusan siswa berdasarkan faktor-faktor sosial-ekonomi dan kedisiplinan. Algoritma yang digunakan adalah Decision Tree C4.5. Hasil menunjukkan bahwa faktor ekonomi orang tua merupakan variabel paling berpengaruh terhadap keberhasilan akademik siswa. Penelitian ini memiliki fokus analisis yang masih diarahkan pada performa akademik dan keberhasilan belajar siswa, tanpa memperluas kajian pada kondisi lulusan setelah menyelesaikan pendidikan. Selain itu, penelitian ini belum menekankan penggunaan metode seleksi atribut untuk menyaring variabel yang paling relevan, sehingga potensi penyederhanaan model dan peningkatan efisiensi analisis masih terbuka.

Penelitian berikutnya adalah prediksi kelulusan siswa dengan penerapan metode Knowledge Discovery in Database (KDD) dan algoritma Decision Tree (Al-giffary, 2024). Tujuan penelitian ini adalah membangun model klasifikasi untuk menentukan status lulus atau tidak lulus berdasar dataset 120 siswa dengan atribut nama, nilai KKM, jurusan, dan status kelulusan. Proses analisis mengikuti tahapan Knowledge Discovery in Database (KDD) dan evaluasi dilakukan menggunakan confusion matrix, accuracy, precision, recall, dan ROC. Decision Tree digunakan karena mudah dipahami dan mampu memetakan atribut. Hasil penelitian menunjukkan akurasi sebesar 91,43%, recall 75%, precision 60%, dan AUC 0,500. Akan tetapi penelitian ini memiliki jumlah atribut yang terbatas dan distribusi data yang seimbang, yang menyebabkan nilai precision lebih rendah. Dan

pada penelitian ini berhasil menunjukkan bahwa Decision Tree dapat menjadi alat analisis yang efektif dalam memprediksi kelulusan.

Penelitian berikutnya adalah model prediksi performa siswa dengan kolaborasi Rough Set dan Backpropagation Neural Network (BPNN) untuk meningkatkan akurasi prediksi dan mengurangi atribut yang tidak relevan (Madeira, 2021). Tujuan penelitian ini untuk menilai efektivitas kombinasi dari kedua metode tersebut dalam memprediksi nilai akhir siswa secara lebih efisien dan kompleksitas data lebih rendah. Rough Set yang digunakan untuk feature selection, yang menyisakan 5 atribut terpenting (assignments, reading, quizzes, speaking, activities), sebelum dilakukan training BPNN. Hasil penelitian mampu memprediksi performa siswa. Penelitian ini kuat dalam metodologi reduksi fitur dan validasi model, namun ukuran dataset kecil (53 siswa) yang membatasi generalisasi hasil, dan model hanya berfokus pada satu Lembaga dengan struktur penelitian tertentu sehingga kurang representative.

Penelitian berikutnya ini merupakan implementasi algoritma Decision Tree menggunakan metode CRISP-DM untuk mengklasifikasikan peserta didik secara lebih efektif berdasarkan atribut nilai rata-rata pengerjaan kuis dan jumlah sesi pengerjaan kuis (Sutoyo et al., 2018). Tujuan penelitian ini menghasilkan model klasifikasi yang mampu mengelompokkan siswa ke kategori kelas A, B, C, atau D. algoritma yang digunakan adalah Decision Tree C4.5 dan Random Forest, dengan validasi menggunakan 10-Fold Cross Validation. Hasil penelitian menunjukkan Decision Tree C4.5 memiliki performa paling tinggi dengan akurasi 96,73%, mengungguli Random Forest yang mencapai 95,13%. Penelitian ini menggunakan Decision Tree kuat dari sisi metodologi dan evaluasi model, namun masih terbatas menggunakan dua atribut sehingga tidak menangkap factor lain yang juga mempengaruhi performa siswa.

Tabel 2. 1 Penelitian Terkait

No	Judul	Algoritma	Dataset yang digunakan
1	Penerapan Teory Rough Set Untuk Memprediksi Tingkat Kelulusan Siswa Dalam Ujian Nasional Pada Sma Negeri 5 Kota Banda Aceh. (Prajana & Andika, 2016)	Rough Set	Rata-rata siswa, nilai pra-UN, serta nilai UN pada SMA Negeri 5 Banda Aceh.
2	Perbandingan Penerapan Algoritma Decision Tree C.45 dan Naïve Bayes Dalam Analisa Kelulusan Siswa pada SMK Swadhipa 2 Natar Kabupaten Lampung Selatan. (Susanti et al., 2023)	Decision Tree & Naïve Bayes	Transkrip nilai akhir Bahasa Indonesia, Bahasa Inggris, matematika Siswa pada SMK Swadhipa 2 Natar Kabupaten Lampung Selatan.
3	Prediction Of Student Performance Using Decision Tree C 4 . 5 Algorithm. (Kuntoro & Sudarwanto, 2017)	Decision Tree C4.5	Status sosial-ekonomi orang tua, tingkat kedisiplinan siswa, dan prestasi akademik 390 data siswa Sekolah Dasar Negeri 4 Trimulyo, Kabupaten Lampung Timur.
4	Klasifikasi Kelulusan Siswa Tahun 2024 Menggunakan Metode Decision Tree (Studi Kasus SMA Islam Al-Azhar 5 Cirebon). (Al-Giffary, F. R., & Martanto, 2024)	Decision Tree	Memiliki 4000 atribut utama yaitu Nama siswa, Nilai, KKm Jurusan (IPA 1, IPA 2, IPA 3, dan IPS) dari 119 siswa kelas XII SMA Islam AL-Azhar 5 Cirebon.
5	Prediction of Student Performance Using Rought Set Theory And Backpagation Neural Networks. (Kuntoro, R. K., Sudarwanto, R., & Sriyanto, 2017).	Rough Set dan Backpropagation Neural Network (BPNN)	53 siswa, 10 atribut perilaku dan keterampilan Bahasa, 1 atribut keputusan.
6	Implementasi Algoritma Decision Tree untuk Klasifikasi Data Peserta Didik. (Kuntoro, R. K., Sudarwanto, R., & Sriyanto, 2017)	Decision Tree metode CRISP-DM	Total 123 siswa dengan atribut ID siswa, nama, jumlah sesi pengerjaan kuis, nilai rata-rata.