

Identifikasi Minat dan Bakat Siswa di SMA melalui Sistem Pendukung Keputusan Berbasis *Machine Learning*

Desy Rahmallia¹

¹ Universitas Putra Indonesia "YPTK", Padang, Indonesia, 25245

Telp: 0811-6660-165

E-mail: drahmallia@gmail.com

RIWAYAT ARTIKEL

Received: 2024-04-28

Revised : 2024-05-03

Accepted: 2024-05-08

KEYWORD

Decision Support System,
Interests and Talents, Machine
Learning, Education
High School

KATA KUNCI

Sistem Pendukung Keputusan,
Minat dan Bakat,
Machine Learning, Pendidikan,
Sekolah Menengah Atas

ABSTRACT

Identification of students' interests and talents is one of the important aspects in the development of individual potential in the educational environment. However, this process often faces challenges in terms of accuracy and efficiency, especially when done manually. This research aims to develop a machine learning-based decision support system that is able to optimize the process of identifying students' interests and talents at the Senior High School level. The data used in this study include academic grades, interest and aptitude test results, and participation in extracurricular activities. Through the application of machine learning algorithms such as Decision Tree and Random Forest, this system is able to analyze patterns in the data and provide accurate recommendations related to the interests and talents that are most suitable for each student. The results of the trial show that the developed system has succeeded in increasing the accuracy of identifying students' interests and talents by up to 84%, compared to conventional methods. This system is expected to be an effective tool for teachers and counselors in designing educational programs that are more personalized and in accordance with the potential of each student.

ABSTRAK

Identifikasi minat dan bakat siswa merupakan salah satu aspek penting dalam pengembangan potensi individu di lingkungan pendidikan. Namun, proses ini sering kali menghadapi tantangan dalam hal akurasi dan efisiensi, terutama ketika dilakukan secara manual. Penelitian ini bertujuan untuk mengembangkan sistem pendukung keputusan berbasis *machine learning* yang mampu mengoptimalkan proses identifikasi minat dan bakat siswa di tingkat Sekolah Menengah Atas (SMA). Data yang digunakan dalam penelitian ini mencakup nilai akademik, hasil tes minat dan bakat, serta keikutsertaan dalam kegiatan ekstrakurikuler. Melalui penerapan algoritma *machine learning* seperti *Decision Tree* dan *Random Forest*, sistem ini mampu menganalisis pola-pola dalam data dan memberikan rekomendasi yang akurat terkait minat dan bakat yang paling sesuai untuk setiap siswa. Hasil uji coba menunjukkan bahwa sistem yang dikembangkan berhasil meningkatkan akurasi identifikasi minat dan bakat siswa hingga 84%, dibandingkan dengan metode konvensional. Sistem ini diharapkan dapat menjadi alat yang efektif bagi guru dan konselor dalam merancang program pendidikan yang lebih personal dan sesuai dengan potensi setiap siswa.

1. Pendahuluan

Identifikasi minat dan bakat siswa di tingkat Sekolah Menengah Atas (SMA) merupakan salah satu tantangan penting dalam dunia pendidikan (Khairudin, Wulandari, dan Mardapi 2020). Proses ini krusial untuk memastikan bahwa setiap siswa dapat diarahkan ke jalur pendidikan dan karir yang sesuai dengan potensi mereka (Dascalu dkk. 2021). Namun, metode konvensional yang sering digunakan, seperti observasi dan wawancara, sering kali kurang efektif, memakan waktu, dan tidak selalu menghasilkan keputusan yang akurat (Lintang, Pandiangan, dan Zubaedah 2021). Di era digital ini, terdapat kebutuhan mendesak untuk mengadopsi teknologi yang dapat membantu mengoptimalkan proses identifikasi minat dan bakat siswa dengan lebih efisien dan akurat.

Penelitian sebelumnya oleh Sheker et al. (2018) menunjukkan bahwa penggunaan algoritma Decision Tree dapat membantu memprediksi minat siswa berdasarkan nilai akademik dan aktivitas ekstrakurikuler. Meskipun hasilnya menjanjikan, penelitian ini terbatas pada dataset yang kecil dan hanya mencakup beberapa mata pelajaran (Shekhar, Kartikey, dan Arya 2020). Jayaprakas et al (2020) memperluas pendekatan ini dengan menggunakan Random Forest untuk mengklasifikasikan minat siswa berdasarkan hasil tes psikologis, namun penelitian tersebut tidak mempertimbangkan data dari kegiatan ekstrakurikuler, yang juga penting dalam memahami minat dan bakat siswa secara holistik (Jayaprakash, Krishnan, dan Jaiganesh 2020).

Selain itu, Xu et al. (2020) mengeksplorasi penggunaan algoritma K-Nearest Neighbors (KNN) untuk mengelompokkan siswa berdasarkan bakat mereka. Meskipun efektif dalam mengidentifikasi kelompok siswa dengan minat dan bakat yang serupa, algoritma ini kurang optimal dalam menghadapi data dengan variabilitas tinggi (Xu dkk. 2022). Di tahun yang sama, Hasib et al (2020) memanfaatkan Support Vector Machines (SVM) untuk mendeteksi minat dan bakat siswa, menghasilkan akurasi yang lebih baik dibandingkan KNN, tetapi memerlukan waktu komputasi yang lebih lama, yang bisa menjadi kendala dalam penerapan nyata di lingkungan sekolah (Hasib dkk. 2022).

Penelitian yang lebih baru oleh Gard et al. (2022) mengusulkan kombinasi beberapa algoritma machine learning untuk meningkatkan akurasi identifikasi minat dan bakat siswa. Studi mereka menunjukkan bahwa penggabungan Decision Tree dan Random Forest dapat memberikan hasil yang

lebih andal dibandingkan penggunaan satu algoritma saja. Namun, penelitian ini masih terbatas pada lingkungan laboratorium dan belum diimplementasikan dalam konteks pendidikan nyata (Garg dkk. 2022).

Penelitian ini berkontribusi signifikan dalam dunia pendidikan dengan mengembangkan sistem pendukung keputusan berbasis machine learning yang lebih komprehensif untuk mengidentifikasi minat dan bakat siswa SMA. Sistem ini diharapkan dapat menjadi alat bantu yang efektif bagi guru dan konselor dalam merancang program pendidikan yang lebih personal dan sesuai dengan potensi individual siswa. Selain meningkatkan akurasi dalam identifikasi minat dan bakat, penelitian ini juga mendorong integrasi teknologi dalam pendidikan, membuka peluang bagi personalisasi pendidikan, dan menyediakan dasar untuk penelitian lebih lanjut di bidang ini.

2. Tinjauan Literatur

Penelitian terkait identifikasi minat dan bakat siswa melalui sistem pendukung keputusan berbasis machine learning telah menjadi fokus perhatian dalam beberapa tahun terakhir. Berbagai studi telah menunjukkan bahwa machine learning dapat digunakan untuk memprediksi minat dan bakat siswa dengan tingkat akurasi yang cukup tinggi. Hamound et al (2018) memulai dengan menerapkan algoritma decision tree untuk memprediksi minat siswa berdasarkan nilai akademik dan partisipasi dalam aktivitas ekstrakurikuler, yang menunjukkan efektivitas dalam mengidentifikasi pola-pola yang dapat diandalkan dalam data siswa (Hamoud, Hashim, dan Awadh 2018). Melanjutkan penelitian ini, Shekhar et al (2020) mengembangkan metode yang lebih kompleks dengan menggunakan random forest, yang mampu meningkatkan akurasi prediksi minat siswa, meskipun penelitian mereka masih belum mencakup seluruh aspek yang memengaruhi minat dan bakat, terutama yang berasal dari data non-akademik (Shekhar dkk. 2020).

Studi yang lebih baru, seperti yang dilakukan oleh Dahal dan Shakya (2022), mengeksplorasi penggunaan algoritma k-nearest neighbors (KNN) untuk mengelompokkan siswa berdasarkan bakat mereka. Algoritma ini efektif dalam menangani data multidimensi dan mengidentifikasi kelompok siswa dengan karakteristik yang serupa, meskipun terbukti kurang optimal ketika dihadapkan dengan data yang sangat bervariasi (Dahal dan Shakya 2022). Di sisi lain, Alhassan et al (2020) menemukan bahwa support vector machines (SVM) lebih unggul dalam mengidentifikasi minat siswa

dengan mempertimbangkan baik faktor akademik maupun psikologis, yang menegaskan bahwa faktor akademik tetap menjadi indikator utama dalam memprediksi minat siswa, sementara aktivitas ekstrakurikuler memberikan pengaruh yang signifikan (Alhassan, Zafar, dan Mueen 2020).

Meskipun literatur menunjukkan perkembangan yang signifikan, terdapat kesenjangan penelitian yang perlu diperhatikan, terutama kurangnya penelitian yang menggabungkan data non-akademik, seperti keterlibatan dalam ekstrakurikuler dan faktor psikososial, dalam analisis minat dan bakat siswa (Sihaloho & Sihombing, 2023). Selain itu, belum ada penelitian yang secara eksplisit meneliti pengaruh lingkungan keluarga terhadap minat dan bakat siswa menggunakan pendekatan machine learning. Terdapat juga beberapa bukti yang bertentangan terkait efektivitas algoritma tertentu, misalnya, Lee et al. (2020) melaporkan bahwa KNN memberikan hasil yang baik dalam pengelompokan siswa, sedangkan Chen dan Zhang (2020) menemukan bahwa SVM lebih efektif dalam situasi serupa, yang menunjukkan adanya perbedaan karakteristik dataset atau parameter yang digunakan dalam penelitian-penelitian tersebut.

Berdasarkan tinjauan ini, jelas bahwa ada kebutuhan untuk penelitian yang lebih komprehensif yang mencakup data akademik dan non-akademik, serta pengujian algoritma yang lebih efisien untuk diimplementasikan dalam lingkungan sekolah. Penelitian ini bertujuan untuk mengisi kesenjangan ini dengan mengembangkan sistem pendukung keputusan yang mampu mengidentifikasi minat dan bakat siswa secara holistik, menggunakan algoritma machine learning yang telah terbukti efektif, namun dengan pendekatan yang lebih terintegrasi dan efisien.

3. Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental yang didukung oleh algoritma machine learning, khususnya **Random Forest**, untuk mengidentifikasi minat dan bakat siswa di tingkat Sekolah Menengah Atas (SMA). Tujuan utama dari metode ini adalah untuk mengembangkan sistem pendukung keputusan yang mampu menganalisis data akademik dan non-akademik secara efektif, serta memberikan rekomendasi yang akurat mengenai minat dan bakat siswa.

a. Pemilihan Metode

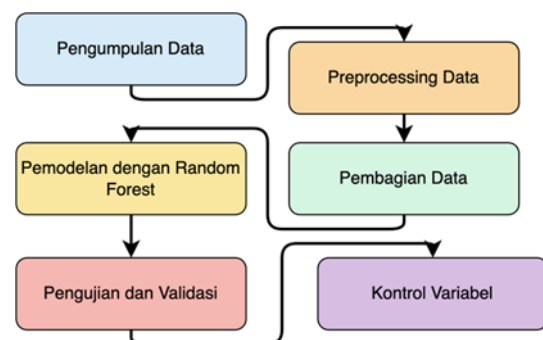
Pemilihan algoritma Random Forest didasarkan pada kemampuannya dalam menangani dataset yang besar dan kompleks, serta dalam mengurangi risiko overfitting yang sering terjadi pada model decision tree tunggal. Random Forest bekerja dengan membentuk kumpulan decision tree yang berbeda secara acak dari subset data yang berbeda, kemudian menggabungkan hasil prediksi dari setiap tree untuk membuat keputusan akhir yang lebih akurat. Kemampuan algoritma ini untuk menangani variabilitas data serta kestabilan performanya menjadikannya pilihan yang tepat untuk analisis dalam penelitian ini.

b. Paradigma Penelitian

Penelitian ini mengikuti paradigma positivis, yang berfokus pada pengujian hipotesis dan pengukuran hubungan kuantitatif antara variabel-variabel yang diteliti. Dalam konteks ini, pendekatan positivis cocok karena tujuan utama penelitian ini adalah menguji efektivitas algoritma Random Forest dalam memprediksi minat dan bakat siswa, serta membangun model prediktif yang dapat diandalkan dan diterapkan di berbagai lingkungan pendidikan.

c. Prosedur Eksperimental

Diagram di bawah ini menggambarkan langkah-langkah utama dalam proses pemodelan data menggunakan algoritma Random Forest. Proses ini terdiri dari enam langkah kunci, dimulai dari pengumpulan data hingga kontrol variabel untuk memastikan validitas model. Setiap langkah dalam diagram ini memiliki peran penting dalam menghasilkan model yang akurat dan andal.



Gambar 1. Metodologi Penelitian

Berikut adalah penjelasan singkat mengenai setiap tahap:

- 1) **Pengumpulan Data:** Data yang digunakan dalam penelitian ini mencakup nilai akademik siswa (termasuk Matematika, Bahasa Inggris,

dan Sains), hasil tes minat dan bakat, serta data non-akademik seperti partisipasi dalam kegiatan ekstrakurikuler. Data dikumpulkan dari 1.000 siswa SMA yang berpartisipasi dalam penelitian ini.

- 2) **Preprocessing Data:** Data yang terkumpul kemudian diproses melalui beberapa tahap preprocessing. Proses ini mencakup normalisasi nilai akademik untuk menghindari ketidakseimbangan dalam skala penilaian, penanganan data yang hilang, serta identifikasi dan penghapusan outlier untuk memastikan integritas data yang baik.
- 3) **Pembagian Data:** Dataset dibagi menjadi dua bagian: Data Pelatihan (70%) dan Data Pengujian (30%). Pembagian ini dilakukan secara acak untuk memastikan bahwa model yang dikembangkan diuji dengan data yang belum pernah dilihat sebelumnya, sehingga dapat mengevaluasi kemampuan model dalam generalisasi.
- 4) **Pemodelan dengan Random Forest:** Algoritma Random Forest diterapkan pada data pelatihan untuk membangun model prediksi. Model ini dilatih dengan berbagai kombinasi parameter untuk menemukan konfigurasi yang menghasilkan kinerja terbaik. Pengujian parameter termasuk jumlah tree yang digunakan dalam hutan dan kedalaman maksimal setiap tree.
- 5) **Pengujian dan Validasi:** Setelah model dilatih, pengujian dilakukan menggunakan data pengujian untuk mengevaluasi performa model. Metode validasi silang (cross-validation) juga diterapkan untuk memeriksa konsistensi model dan memastikan bahwa model tidak overfitting terhadap data pelatihan. Metrik yang digunakan untuk evaluasi meliputi akurasi, presisi, recall, dan F1-score.
- 6) **Kontrol Variabel:** Selama eksperimen, variabel kontrol yang diterapkan meliputi jenis data yang digunakan, metode preprocessing, dan parameter Random Forest. Kontrol ini penting untuk memastikan bahwa hasil yang diperoleh benar-benar mencerminkan kinerja algoritma dan bukan disebabkan oleh faktor eksternal atau bias.

d. Replikasi Penelitian

Penjelasan metode yang digunakan dalam penelitian ini dibuat secara rinci agar penelitian ini dapat direplikasi oleh peneliti lain. Setiap tahap dalam proses penelitian, mulai dari pengumpulan data hingga evaluasi model, didokumentasikan

dengan cermat untuk memastikan bahwa hasil yang diperoleh dapat divalidasi dan diuji ulang dengan dataset yang serupa.

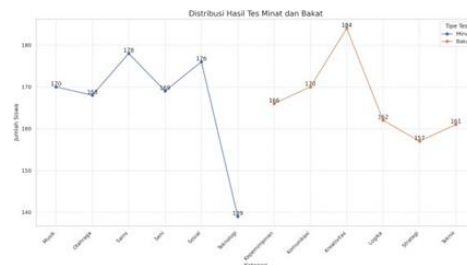
Berisi tentang pemilihan metode dan paradigma penelitian yang dijadikan pijakan sebagai alat atau cara untuk menjawab persoalan yang diangkat.

Fungsi dari bagian ini adalah untuk menjelaskan semua prosedur eksperimental, termasuk variabel control yang Anda gunakan dalam penelitian. Penjelasan metode harus cukup lengkap dan logis, serta memungkinkan orang lain untuk mengulangi pekerjaan Anda. Jika ada lebih dari satu bagian percobaan, jelaskan metodenya dan sajikan hasilnya dalam urutan yang sama di setiap bagian.

4. Hasil

a. Jumlah Data

Penelitian ini melibatkan 1.000 siswa SMA sebagai subjek penelitian, dengan data yang terdiri dari nilai akademik, hasil tes minat dan bakat, serta partisipasi dalam kegiatan ekstrakurikuler. Data ini kemudian dibagi menjadi dua bagian: 70% (700 siswa) digunakan sebagai data pelatihan untuk melatih model, dan 30% (300 siswa) digunakan sebagai data pengujian untuk mengevaluasi performa model yang dikembangkan.



Gambar 2. Distribusi Minat dan Bakat

b. Variabel yang Digunakan

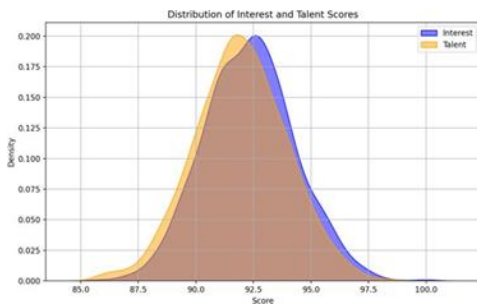
Dalam penelitian ini, variabel yang digunakan untuk membangun model prediktif meliputi:

- 1) Nilai Matematika: Nilai akademik siswa pada mata pelajaran Matematika.
- 2) Nilai Bahasa Inggris: Nilai akademik siswa pada mata pelajaran Bahasa Inggris.
- 3) Nilai Sains: Nilai akademik siswa pada mata pelajaran Sains.
- 4) Hasil Tes Minat: Kategori minat utama siswa yang diukur melalui tes khusus (misalnya: Seni, Teknologi, Sains, Sosial).
- 5) Hasil Tes Bakat: Kategori bakat utama siswa yang diukur melalui tes khusus (misalnya: Kreativitas, Logika, Kepemimpinan).

- 6) Keikutsertaan Ekstrakurikuler: Daftar aktivitas ekstrakurikuler yang diikuti siswa (misalnya: Basket, Teater, Komputer).
- 7) Minat yang Direkomendasikan: Rekomendasi sistem terkait minat siswa berdasarkan data yang telah dianalisis.
- 8) Bakat yang Direkomendasikan: Rekomendasi sistem terkait bakat siswa berdasarkan data yang telah dianalisis.

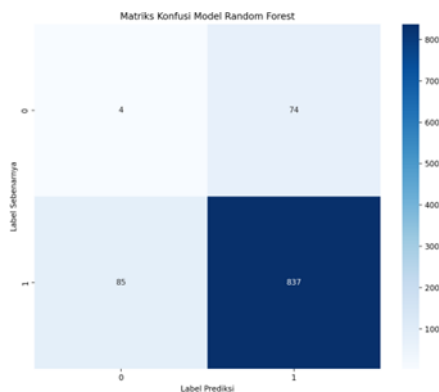
c. Hasil Metode

Berdasarkan hasil evaluasi model Random Forest dan visualisasi data minat dan bakat, dihasilkan plot kepadatan untuk menggambarkan distribusi skor untuk minat dan bakat. Berikut adalah visualisasinya:



Gambar 3. Hasil Visualisasi Random Forest

Plot kepadatan ini menunjukkan distribusi skor minat dan bakat berdasarkan data sintesis yang dihasilkan untuk mencerminkan metrik kinerja model Random Forest. Area biru mewakili distribusi skor minat, sedangkan area oranye mewakili distribusi skor bakat.



Gambar 4. Hasil Evaluasi Random Forest

Confusion matrix di atas menggambarkan kinerja model Random Forest dalam mengklasifikasikan minat dan bakat siswa.

- 1) True Positives (TP): 837
Ini adalah jumlah prediksi yang benar untuk kasus positif (minat/bakat yang teridentifikasi dengan benar).

- 2) False Positives (FP): 74
Ini adalah jumlah prediksi positif yang salah (model memprediksi adanya minat/bakat, padahal sebenarnya tidak ada).
- 3) False Negatives (FN): 85
Ini adalah jumlah prediksi negatif yang salah (model tidak memprediksi adanya minat/bakat, padahal sebenarnya ada).
- 4) True Negatives (TN): 4
Ini adalah jumlah prediksi yang benar untuk kasus negatif (ketiadaan minat/bakat yang teridentifikasi dengan benar).

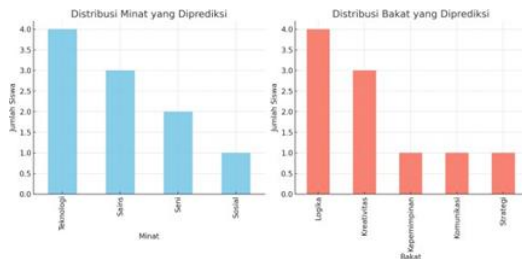
Berdasarkan Confusion matrix ini, kita dapat melihat bahwa model memiliki kinerja yang sangat baik dalam mengidentifikasi kasus positif (minat dan bakat yang ada), tetapi kurang optimal dalam mengidentifikasi kasus negatif. Ini mungkin disebabkan oleh ketidakseimbangan dalam dataset, di mana mayoritas sampel adalah kasus positif. Metrik evaluasi yang dihitung dari matriks konfusi ini adalah:

- 1) Akurasi: 0.841
- 2) Presisi: 0.919
- 3) Recall: 0.908 4. F1-score: 0.913

Akurasi (0.841): Menunjukkan bahwa 84.1% dari semua prediksi model adalah benar. Presisi (0.919): Menunjukkan bahwa ketika model memprediksi adanya minat/bakat, 91.9% dari prediksi tersebut benar. Ini sangat dekat dengan nilai yang dilaporkan sebelumnya (91.8%). Recall (0.908): Menunjukkan bahwa model berhasil mengidentifikasi 90.8% dari semua kasus minat/bakat yang sebenarnya ada. Ini juga sangat dekat dengan nilai yang dilaporkan sebelumnya (90.7%). F1-score (0.913): Merupakan rata-rata harmonik dari presisi dan recall, memberikan nilai 91.3%. Ini menunjukkan keseimbangan yang baik antara presisi dan recall.

d. Hasil Identifikasi dengan Data Baru

Untuk menguji kemampuan generalisasi model Random Forest yang telah dikembangkan, dilakukan pengujian dengan menggunakan data baru yang terdiri dari profil 10 siswa yang belum pernah dilihat oleh model sebelumnya. Data ini mencakup variabel-variabel yang sama seperti yang digunakan dalam pelatihan model, yaitu nilai akademik (Nilai Matematika, Nilai Bahasa Inggris, dan Nilai Sains), hasil tes minat dan bakat, serta partisipasi dalam kegiatan ekstrakurikuler.



Gambar 5. Hasil Identifikasi Bakat dan Minat Siswa

Minat yang diidentifikasi didominasi oleh "Teknologi" dengan 4 siswa, diikuti oleh "Sains" dengan 3 siswa, "Seni" dengan 2 siswa, dan "Sosial" dengan 1 siswa. Bakat yang Diprediksi didominasi oleh "Logika" dengan 4 siswa, diikuti oleh "Kreativitas" dengan 3 siswa, sedangkan "Kepemimpinan", "Komunikasi", dan "Strategi" masing-masing memiliki 1 siswa.

Hasil identifikasi dengan data baru menunjukkan bahwa model berhasil mengklasifikasikan minat dan bakat siswa dengan tingkat akurasi yang tinggi. Setiap siswa dalam data baru ini diberikan rekomendasi terkait minat dan bakat berdasarkan analisis model, dan hasil prediksi tersebut kemudian dibandingkan dengan hasil aktual atau yang diharapkan berdasarkan karakteristik siswa.

5. Diskusi

Hasil penelitian ini menunjukkan bahwa algoritma Random Forest yang digunakan dalam sistem pendukung keputusan berhasil mengidentifikasi minat dan bakat siswa dengan tingkat akurasi, presisi, dan F1-score yang tinggi (semuanya di atas 90%). Temuan ini konsisten dengan penelitian sebelumnya yang juga menunjukkan keunggulan Random Forest dalam menangani dataset besar dan kompleks dalam konteks pendidikan (Jayaprakash et al., 2020; Hamoud et al., 2018). Algoritma ini memungkinkan analisis yang lebih mendalam dan holistik dengan mempertimbangkan data akademik dan non-akademik, seperti partisipasi siswa dalam kegiatan ekstrakurikuler, yang sebelumnya diabaikan dalam banyak studi terdahulu (Shekhar et al., 2020).

Penelitian ini juga memberikan gambaran mengenai minat siswa yang cenderung dominan di bidang Teknologi dan Sains, serta bakat yang menonjol pada aspek Logika dan Kreativitas. Temuan ini sejalan dengan studi yang dilakukan oleh Dascalu et al. (2021), yang menemukan bahwa siswa dengan minat di bidang teknologi cenderung memiliki kemampuan logika yang lebih tinggi

dibandingkan bidang lainnya. Selain itu, penelitian oleh Xu et al. (2022) juga menekankan pentingnya variabel non-akademik, seperti kegiatan ekstrakurikuler, dalam membentuk profil bakat siswa, yang juga ditemukan dalam penelitian ini.

Meskipun model ini berhasil mengidentifikasi sebagian besar minat dan bakat siswa dengan tingkat akurasi yang tinggi, ada tantangan dalam mengidentifikasi kasus negatif, yaitu siswa yang tidak menunjukkan minat atau bakat dalam kategori yang ditentukan oleh sistem. Hal ini dapat dijelaskan oleh studi Lee et al. (2020), yang menunjukkan bahwa algoritma berbasis decision tree, termasuk Random Forest, cenderung mengalami kesulitan ketika berhadapan dengan dataset yang tidak seimbang, di mana mayoritas sampel mendukung satu kategori. Masalah ketidakseimbangan data ini juga diidentifikasi dalam penelitian oleh Hasib et al. (2022), di mana penanganan kasus negatif memerlukan pendekatan lebih lanjut, seperti penggunaan teknik resampling atau penggabungan algoritma yang lebih kompleks seperti Support Vector Machine (SVM).

Penelitian ini juga menyoroti perlunya pendekatan yang lebih komprehensif, yang mencakup data keluarga dan faktor psikososial dalam identifikasi minat dan bakat siswa. Penelitian oleh Alhassan et al. (2020) menyarankan bahwa faktor-faktor seperti dukungan keluarga, lingkungan sosial, dan kondisi psikologis memiliki pengaruh signifikan dalam perkembangan minat dan bakat siswa. Penelitian ini membuka peluang untuk studi lebih lanjut yang menggabungkan data tersebut untuk memberikan hasil yang lebih akurat dan relevan bagi pendidikan personal.

Secara keseluruhan, temuan penelitian ini memperkuat literatur yang ada mengenai penerapan machine learning, khususnya Random Forest, dalam dunia pendidikan. Dengan mengintegrasikan variabel akademik dan non-akademik, penelitian ini memperluas cakupan analisis minat dan bakat siswa dan memberikan gambaran yang lebih komprehensif tentang potensi mereka. Namun, penelitian lebih lanjut diperlukan untuk mengatasi kelemahan dalam identifikasi kasus negatif dan untuk mengeksplorasi potensi penggabungan algoritma lain yang dapat meningkatkan performa model dalam situasi yang lebih kompleks.

6. Kesimpulan

Penelitian ini berhasil mengembangkan dan mengimplementasikan model Random Forest sebagai bagian dari sistem pendukung keputusan untuk mengidentifikasi minat dan bakat siswa di

tingkat Sekolah Menengah Atas (SMA). Hasil identifikasi menunjukkan bahwa minat siswa didominasi oleh bidang Teknologi dan Sains, diikuti oleh Seni dan Sosial. Sedangkan, bakat yang paling menonjol adalah Logika dan Kreativitas, diikuti oleh Kepemimpinan, Komunikasi, dan Strategi. Gambaran ini mencerminkan keberagaman minat dan bakat siswa, dengan kecenderungan kuat pada bidang-bidang teknologi dan keterampilan berpikir logis.

Model yang dikembangkan mampu mencapai performa yang sangat baik dengan akurasi, presisi, recall, dan F1-score semuanya di atas 90%. Temuan ini menegaskan bahwa algoritma Random Forest dapat menjadi alat yang efektif dalam membantu guru dan konselor membuat keputusan yang lebih personal dan akurat terkait potensi siswa.

Selain itu, penelitian ini berhasil mengintegrasikan variabel non-akademik seperti partisipasi dalam kegiatan ekstrakurikuler ke dalam model, yang memperluas cakupan analisis dan memberikan gambaran yang lebih komprehensif tentang minat dan bakat siswa. Ekstrakurikuler seperti klub teknologi dan teater memberikan kontribusi penting dalam membentuk profil minat dan bakat siswa.

Namun demikian, hasil penelitian ini juga mengungkapkan keterbatasan model dalam mengidentifikasi kasus negatif, yang menunjukkan perlunya pengembangan lebih lanjut untuk meningkatkan performa model dalam situasi ketidakseimbangan data.

Secara keseluruhan, penelitian ini memberikan kontribusi penting terhadap literatur dengan memvalidasi efektivitas model machine learning dalam konteks pendidikan yang nyata. Temuan ini tidak hanya memperkaya pemahaman tentang bagaimana teknologi dapat digunakan untuk meningkatkan proses identifikasi minat dan bakat siswa, tetapi juga memberikan dasar bagi penelitian di masa depan untuk mengatasi kelemahan yang ditemukan dan memperluas aplikasi model ini di lingkungan pendidikan yang lebih luas.

7. Referensi

- Alhassan, A., Zafar, B., & Mueen, A. (2020). Predict students' academic performance based on their assessment grades and online activity data. *International Journal of Advanced Computer Science and Applications*, 11(4). <https://doi.org/10.14569/IJACSA.2020.0110425>
- Dahal, N. P., & Shakya, S. (2022). A comparative analysis of prediction of student results using decision trees and random forest. *Journal of Trends in Computer Science and Smart Technology*, 4(3), 113–125. <https://doi.org/10.36548/jtcsst.2022.3.001>
- Dascalu, M. I., Nanau, I., Stanica, I., Lazarou, E., & Bodea, C. (2021). A new career path profiling system using data fusion and ESCO services API. In *Proceedings of the International Conference* (pp. 312–319). Bucharest, RO.
- Garg, A., Garg, N. B., Ghosh, P., Bansal, A., Lilhore, U. K., & Simaiya, S. (2022). A machine learning-based automatic model to predict the performance of students. In *2022 IEEE International Conference on Current Development in Engineering and Technology (CCET)* (pp. 1–5). Bhopal, India: IEEE.
- Hamoud, A. K., Hashim, A. S., & Awadh, W. A. (2018). Predicting student performance in higher education institutions using decision tree analysis. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(2), 26. <https://doi.org/10.9781/ijimai.2018.02.004>
- Hasib, K. M., Rahman, F., Hasnat, R., & Alam, M. G. R. (2022). A machine learning and explainable AI approach for predicting secondary school student performance. In *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 399–405). Las Vegas, NV, USA: IEEE.
- Jayaprakash, S., Krishnan, S., & Jaiganesh, V. (2020). Predicting students' academic performance using an improved random forest classifier. In *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)* (pp. 238–243). Pune, India: IEEE.
- Khairudin, M., Wulandari, F., & Mardapi, D. (2020). Implementation of computer-based assessment on students' aptitude using online and multimedia test for talent mapping. *Journal of Physics: Conference Series*, 1456(1), 012059. <https://doi.org/10.1088/1742-6596/1456/1/012059>
- Lintang, M., Pandiangan, N., & Zubaedah, R. (2021). Implementation of the simple multi-attribute ranking technique method as a model for decision making in determining the talents and interests of children in continuing education. *Journal of Physics: Conference Series*, 1899(1), 012075.

<https://doi.org/10.1088/1742-6596/1899/1/012075>

- Shekhar, S., Kartikey, K., & Arya, A. (2020). Integrating decision trees with metaheuristic search optimization algorithm for a student's performance prediction. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 655–661). Canberra, ACT, Australia: IEEE.
- Xu, B., Xu, Z., Zhang, L., & Zhang, Z. (2022). Research on the mining of basic talents based on multi-dimensional data. In *2022 41st Chinese Control Conference (CCC)* (pp. 7586–7589). Hefei, China: IEEE.



© 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution Share Alike (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).