

SKRIPSI

**PREDIKSI KESIAPAN SISWA SEKOLAH DASAR MASUK SEKOLAH
MENENGAH PERTAMA MENGGUNAKAN ALGORITMA
RANDOM FOREST**

DISUSUN OLEH

MUHAMMAD RAFLI TAUFIQ LUBIS
2109020194



**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025**

**PREDIKSI KESIAPAN SISWA SEKOLAH DASAR MASUK SEKOLAH
MENENGAH PERTAMA MENGGUNAKAN ALGORITMA
RANDOM FOREST**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer (S.Kom) dalam Program Studi Teknologi
Informasi Pada Fakultas Ilmu Komputer dan Teknologi
Informasi, Universitas Muhammadiyah Sumatera Utara**

Mhd. Rafli Taufiq Lubis

NPM. 2109020194

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN**

2025

LEMBAR PENGESAHAN

Judul Skripsi : PREDIKSI KESIAPAN SISWA SEKOLAH DASAR
MASUK SEKOLAH MENENGAH PERTAMA
MENGUNAKAN ALGORITMA RANDOM FOREST

Nama Mahasiswa : MHD. RAFLI TAUFIQ LUBIS

NPM : 2109020194

Program Studi : TEKNOLOGI INFORMASI

Menyetujui
Komisi Pembimbing

(Mhd. Basri, S.Si, M.Kom)
NIDN. 0111078802

Ketua Program Studi

(Fatma Sari Hutagalung, S.Kom, M.Kom)
NIDN. 0117019301



Dekan

Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

PREDIKSI KESIAPAN SISWA SEKOLAH DASAR MASUK SEKOLAH MENENGAH PERTAMA MENGGUNAKAN ALGORITMA RANDOM FOREST

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, September 2025

Yang membuat pernyataan



Mhd. Rafli Taufiq Lubis

NPM. 2109020194

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama	: Mhd. Rafli Taufiq Lubis
NPM	: 2109020194
Program Studi	: Teknologi Informasi
Karya Ilmiah	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**PREDIKSI KESIAPAN SISWA SEKOLAH DASAR MASUK SEKOLAH
MENENGAH PERTAMA MENGGUNAKAN ALGORITMA RANDOM
FOREST**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, September 2025

Yang membuat pernyataan



Mhd. Rafli Taufiq Lubis

NPM. 2109020194

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Mhd. Rafli Taufiq Lubis
Tempat dan Tanggal Lahir : Medan, 03 April 2004
Alamat Rumah : Jl. Benteng Gg. Mawardi No. 20 Helvetia
Telepon/Faks/HP : 082173662038
E-mail : muhammadraflitaufiqlubis@gmail.com
Instansi Tempat Kerja : RA JAMI'
Alamat Kantor : Jl. Merdeka No.2, Pulo Brayan Kota

DATA PENDIDIKAN

SD	: SD NEGERI 060843	TAMAT: 2012
SMP	: MTs. PAB 1 HELVETIA	TAMAT: 2018
SMA	: MAS PAB 2 HELVETIA	TAMAT: 2021

KATA PENGANTAR



Puji syukur ke hadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini. Laporan skripsi ini disusun untuk memenuhi salah satu syarat dalam memperoleh gelar Sarjana Komputer (S.Kom) pada Program Studi Teknologi Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara.

Ucapan terima kasih yang tulus Penulis sampaikan kepada kedua orang tua tercinta, **Ahmad Taufiq Lubis** dan **Noningsih S.Pd.I, S.Pd.**, atas segala doa, dukungan moril dan materil, serta semangat yang tak pernah henti diberikan kepada Penulis. Tanpa mereka, penyusunan skripsi ini tidak akan berjalan sebagaimana mestinya. Penulis juga menyadari bahwa penyusunan skripsi ini tidak terlepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, Penulis menyampaikan terima kasih dan penghargaan yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Agussani, MAP, selaku Rektor Universitas Muhammadiyah Sumatera Utara.
2. Bapak Prof. Dr. Muhammad Arifin, S.H., M.Hum, Bapak Prof. Dr. Akrim, M.Pd, dan Bapak Assoc. Prof. Dr. Rudianto, S.Sos., M.Si, selaku Wakil Rektor I, II, dan III Universitas Muhammadiyah Sumatera Utara.
3. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom, selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi.
4. Bapak Halim Maulana, S.T., M.Kom, selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi
5. Bapak Dr. Lutfi Basit, S.Sos., M.I.Kom., selaku Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi.
6. Bapak Halim Maulana, S.T., M.Kom, selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi

7. Bapak Dr. Lutfi Basit, S.Sos., M.I.Kom., selaku Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi.
8. Ibu Fatma Sari Hutagalung, S.Kom, M.Kom., Ketua Program Studi Teknologi Informasi
9. Bapak Mhd Basri, S.Si, M.Kom., Sekretaris Program Studi Teknologi Informasi, sekaligus Dosen Pembimbing Penulis yang telah membantu dan membimbing penulis selama proses penyusunan skripsi ini sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik.
10. Teman-teman Nenek Squad yang telah menjadi bagian penting dalam perjalanan akademik Penulis dan tidak menyerah dalam menyelesaikan pendidikan hingga jenjang perguruan tinggi.
11. Teman-teman sekelas Penulis di kelas A2 Siang Program Studi Teknologi Informasi yang telah menjadi bagian penting dalam perjalanan akademik Penulis.
12. Sahabat - Sahabat ASMANTARA, yang telah memberikan semangat agar penulis tetap termotivasi dan tidak menyerah dalam menyelesaikan pendidikan hingga jenjang perguruan tinggi.

ABSTRAK

Transisi pendidikan dari Sekolah Dasar (SD) ke Sekolah Menengah Pertama (SMP) merupakan fase krusial yang menentukan keberhasilan akademik siswa di masa depan. Namun, proses asesmen kesiapan siswa saat ini seringkali bersifat subjektif dan belum memanfaatkan data secara optimal. Penelitian ini bertujuan untuk membangun dan mengevaluasi model prediksi kesiapan siswa SD masuk SMP menggunakan algoritma *machine learning*, serta mengidentifikasi faktor-faktor paling berpengaruh yang menentukan kesiapan tersebut.

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan algoritma klasifikasi *Random Forest*. Data penelitian diperoleh dari 127 siswa kelas VI di SD Negeri 060843 Medan, yang mencakup variabel independen berupa nilai akademik (Matematika, Bahasa Indonesia, IPA), catatan kehadiran, skor psikologi, dan partisipasi ekstrakurikuler. Model yang dibangun diimplementasikan dalam bentuk prototipe aplikasi web interaktif untuk kemudahan penggunaan. Hasil pengujian model pada data uji menunjukkan performa yang sangat tinggi, dengan keberhasilan mencapai tingkat akurasi, presisi, dan *recall* sebesar 100%. Analisis *feature importance* dari model *Random Forest* berhasil mengidentifikasi bahwa tiga faktor paling dominan yang memengaruhi prediksi kesiapan siswa secara berurutan adalah (1) Nilai Matematika, (2) Catatan Kehadiran, dan (3) Skor Psikologi.

Kata Kunci: Kesiapan Siswa, Prediksi, *Machine Learning*, *Random Forest*, Transisi Pendidikan.

ABSTRACT

The transition from Elementary School (SD) to Junior High School (SMP) is a crucial phase that determines students' future academic success. However, the current assessment process of student readiness is often subjective and does not optimally utilize data. This study aims to develop and evaluate a predictive model of elementary school students' readiness for junior high school using machine learning algorithms, as well as to identify the most influential factors determining such readiness.

This research employs a quantitative approach by applying the Random Forest classification algorithm. The data were collected from 127 sixth-grade students at SD Negeri 060843 Medan, consisting of independent variables such as academic performance (Mathematics, Indonesian Language, Science), attendance records, psychological scores, and extracurricular participation. The developed model was implemented as an interactive web-based prototype application for ease of use. The testing results on the evaluation dataset demonstrated excellent performance, achieving 100% accuracy, precision, and recall. Furthermore, the feature importance analysis of the Random Forest model identified the three most dominant factors influencing readiness prediction, namely (1) Mathematics Score, (2) Attendance Records, and (3) Psychological Score.

Keywords: Student Readiness, Prediction, Machine Learning, Random Forest, Educational Transition.

DAFTAR ISI

KATA PENGANTAR	i
ABSTRAK	iii
DAFTAR ISI.....	v
DAFTAR TABEL	vii
DAFTAR GAMBAR.....	viii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian	5
1.5 Manfaat penelitian.....	6
BAB II LANDASAN TEORI	8
2.1 Kesiapan Masuk Sekolah Menengah Pertama	8
2.2 Peran Pendidikan Dasar dalam Kesiapan ke sekolah Menengah.....	13
2.3 Pendekatan Data <i>Driven Decision Making</i> Dalam Pendidikan.....	16
2.4 <i>Machine Learning</i> Dalam Dunia Pendidikan.....	20
2.4.1 Pengertian Dan Prinsip Dasar <i>Machine Learning</i>	20
2.4.2 Peran <i>Machine Learning</i> dalam Prediksi Pendidikan Siswa	24
2.4.3 Algoritma <i>Random Forest</i> dan <i>Feature Importance</i> dalam Pendidikan.....	27
2.5 <i>Phyton</i>	33
2.6 Penelitian Terdahulu	34

BAB III METODOLOGI PENELITIAN	37
3.1 Jenis penelitian	37
3.2 Lokasi Penelitian dan Waktu Penelitian	37
3.3 Populasi dan Sampel Penelitian	38
3.4 Variabel Penelitian	39
3.5 Teknik Pengumpulan Data	39
3.6 Alur Penelitian	39
3.7 Rancangan Sistem	43
BAB IV HASIL DAN PEMBAHASAN	45
4.1 Implementasi Sistem	45
4.2 Hasil Pengujian Model	45
4.2.1 Confusion Matrix	46
4.2.2 Hasil Metrik Evaluasi	47
4.2.3 Rincian Perhitungan Metrik Evaluasi	48
4.2.4 Hasil Evaluasi Model	50
4.3 Analisis Faktor Penentu Kesiapan	52
4.4 Proses Sistem	53
4.5 Pembahasan	54
BAB IV KESIMPULAN DAN SARAN	56
5.1 Kesimpulan	56
5.2 Saran	57
DAFTAR PUSTAKA	59
LAMPIRAN	63

DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terdahulu	35
Tabel 3.1 Jadwal Penelitian.....	37
Tabel 4.1 Hasil Confusion Matrix.....	46
Tabel 4.2 Hasil Metrik Evaluasi Model	47

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian.....	39
Gambar 3.2 Flowchart Sistem.....	43
Gambar 4.1 Visualisasi Confusion Matrix.....	47
Gambar 4.2 Grafik Tingkat Kepentingan Fitur.....	52

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Transisi pendidikan dari jenjang Sekolah Dasar (SD) ke Sekolah Menengah Pertama (SMP) merupakan salah satu fase penting dalam perkembangan peserta didik yang berada pada tahap remaja awal. Perubahan jenjang ini membawa konsekuensi yang tidak hanya terbatas pada penyesuaian akademik, tetapi juga penyesuaian psikologis, sosial, dan keterampilan hidup (Ghazali, 2023). Menurut Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, tujuan pendidikan dasar adalah memberikan bekal kemampuan dasar kepada peserta didik untuk melanjutkan ke pendidikan menengah. Dengan demikian, keberhasilan transisi dari SD ke SMP menjadi indikator penting tercapainya tujuan pendidikan dasar.

Berdasarkan Peraturan Menteri Pendidikan dan Kebudayaan Nomor 1 Tahun 2021 tentang Penerimaan Peserta Didik Baru (PPDB), siswa yang telah menyelesaikan pendidikan SD berhak melanjutkan ke SMP tanpa seleksi akademik ketat. Namun, di lapangan, tingkat kesiapan siswa sangat bervariasi. Perbedaan ini dapat memengaruhi kemampuan mereka beradaptasi dengan beban pelajaran yang lebih banyak, metode pembelajaran yang lebih kompleks, serta lingkungan sosial yang baru (Permendikbud No. 1 Tahun 2021).

Penelitian Mutia et al. (2022) menegaskan bahwa kesiapan masuk SMP tidak cukup diukur dari kelulusan formal atau nilai akademik saja. Kesiapan tersebut mencakup aspek holistik seperti kemampuan literasi, numerasi, regulasi

diri, keterampilan sosial, motivasi belajar, dan kesiapan emosional. Siswa yang kurang siap dalam salah satu aspek ini berpotensi mengalami kesulitan belajar, penurunan motivasi, atau masalah adaptasi di tahun-tahun awal SMP (Mutia et al., 2022).

Faktor-faktor yang memengaruhi kesiapan siswa SD masuk SMP dapat dibedakan menjadi beberapa kategori. Pertama, faktor akademik, yang mencakup pencapaian pada mata pelajaran inti seperti Bahasa Indonesia, Matematika, Ilmu Pengetahuan Alam (IPA), dan Ilmu Pengetahuan Sosial (IPS). Kemampuan literasi dan numerasi menjadi indikator utama keberhasilan adaptasi di SMP karena menjadi dasar dalam memahami materi yang lebih kompleks (OECD, 2019). Kedua, faktor non-akademik, seperti keterampilan belajar mandiri, kedisiplinan, konsistensi kehadiran, dan manajemen waktu (Hasanah & Wulandari, 2023). Ketiga, faktor psikososial, yang meliputi kemampuan berinteraksi dengan teman sebaya, komunikasi dengan guru, pengendalian emosi, dan kepercayaan diri (Ghandour et al., 2024).

Sayangnya, di banyak sekolah, asesmen kesiapan siswa SD masuk SMP masih dilakukan secara manual dan bergantung pada penilaian subjektif guru atau hanya berdasarkan nilai rapor. Penilaian seperti ini cenderung mengabaikan aspek-aspek non-akademik dan psikososial yang sama pentingnya dengan prestasi akademik. Padahal, sekolah sebenarnya memiliki data yang sangat kaya seperti nilai ujian sekolah, hasil asesmen literasi dan numerasi, catatan absensi, portofolio tugas, dan observasi sikap belajar (Retnawati et al., 2021). Apabila data ini diolah

secara sistematis, hasilnya dapat memberikan gambaran yang lebih objektif mengenai kesiapan siswa.

Pendekatan *Data-Driven Decision Making* (DDDM) menawarkan solusi untuk permasalahan ini. DDDM merupakan pendekatan pengambilan keputusan yang didasarkan pada analisis data yang relevan dan valid, bukan sekadar intuisi atau pengalaman (Bousdekis et al., 2021). Dalam konteks pendidikan, DDDM memungkinkan sekolah untuk mengidentifikasi siswa yang membutuhkan intervensi khusus sebelum memasuki SMP. Integrasi DDDM dengan teknologi kecerdasan buatan (*Artificial Intelligence*), khususnya *Machine Learning* (ML), dapat meningkatkan akurasi prediksi dan mempercepat proses asesmen (Alghamdi & Rahman, 2023).

Salah satu algoritma ML yang banyak digunakan adalah *Random Forest*, yang bekerja dengan membentuk banyak pohon keputusan (*decision tree*) untuk menghasilkan prediksi yang lebih stabil dan akurat. Algoritma ini memiliki keunggulan dalam menangani data kompleks serta mampu memberikan informasi tentang variabel mana yang paling berpengaruh melalui analisis *feature importance* (Breiman, 2001). Dalam konteks kesiapan siswa SD masuk SMP, variabel-variabel seperti nilai matematika, bahasa Indonesia, kehadiran, dan sikap belajar dapat diolah untuk menghasilkan model prediksi yang objektif.

Penelitian Fadhilah & Pratama (2024) membuktikan bahwa penggunaan *Random Forest* dalam memprediksi kesiapan akademik siswa memberikan akurasi tinggi dan membantu sekolah merancang program pembelajaran yang lebih tepat sasaran. Selain itu, analisis *feature importance* yang dihasilkan model

dapat membantu pihak sekolah mengetahui faktor-faktor mana yang paling memengaruhi kesiapan siswa, sehingga intervensi dapat difokuskan pada area yang membutuhkan perhatian lebih.

Dengan memanfaatkan pendekatan ini, diharapkan proses transisi dari SD ke SMP tidak lagi dianggap sebagai sekadar formalitas administratif, tetapi menjadi proses yang terencana dan mendukung perkembangan optimal peserta didik. Model prediktif yang dikembangkan melalui penelitian ini diharapkan menjadi alat bantu yang objektif, praktis, dan berbasis data bagi guru, orang tua, dan pembuat kebijakan dalam mempersiapkan siswa memasuki jenjang pendidikan menengah (Retnawati et al., 2021).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah bagaimana membangun dan mengevaluasi model prediksi kesiapan siswa Sekolah Dasar masuk Sekolah Menengah Pertama menggunakan algoritma *Random Forest*?

1.3 Batasan Masalah

Agar penelitian ini terarah dan fokus, maka dilakukan beberapa pembatasan sebagai berikut:

1. Data yang digunakan hanya berasal dari satu atau beberapa sekolah dasar yang memberikan akses resmi, dan mencakup data perkembangan akademik dan non-akademik siswa kelas VI (enam) pada satu semester atau satu tahun ajaran terakhir.

2. Aspek yang dianalisis dibatasi pada beberapa kategori utama, yaitu: nilai akademik mata pelajaran inti (Bahasa Indonesia, Matematika, IPA), nilai literasi dan numerasi (jika tersedia), kehadiran, serta penilaian sikap belajar dan keterampilan sosial.
3. Label kesiapan siswa ditentukan berdasarkan asesmen guru atau klasifikasi yang telah ditetapkan sekolah, misalnya kategori: **Siap**, **Perlu Bimbingan**, dan **Belum Siap**. Penelitian ini tidak melakukan psikotes atau asesmen psikologis secara langsung.
4. Penelitian ini menggunakan algoritma *Random Forest* sebagai metode utama untuk klasifikasi kesiapan siswa, serta tidak membandingkannya secara mendalam dengan algoritma *machine learning* lain.
5. Pengolahan data dan pelatihan model dilakukan secara offline menggunakan perangkat lunak Python (Google Colab) dan tidak diintegrasikan ke dalam sistem real-time sekolah.
6. Penelitian ini tidak mencakup faktor eksternal di luar data siswa, seperti kondisi keluarga, status ekonomi, atau latar belakang pendidikan orang tua, karena keterbatasan akses data dan ruang lingkup penelitian.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Mengolah dan menganalisis data nilai akademik, kehadiran, dan penilaian sikap belajar siswa kelas VI Sekolah Dasar sebagai dasar dalam membentuk model prediksi kesiapan masuk Sekolah Menengah Pertama.

2. Menerapkan algoritma *Random Forest* sebagai metode klasifikasi untuk memprediksi kesiapan siswa SD dalam memasuki jenjang pendidikan SMP berdasarkan data yang tersedia.
3. Mengevaluasi performa model klasifikasi *Random Forest* dalam hal akurasi, presisi, dan sensitivitas dalam mengklasifikasikan kesiapan siswa ke dalam kategori seperti Siap, Perlu Bimbingan, dan Belum Siap.
4. Mengidentifikasi faktor-faktor yang paling berpengaruh terhadap hasil prediksi kesiapan siswa SD masuk SMP melalui analisis *feature importance* dari model *Random Forest*, guna memberikan wawasan tambahan bagi guru dan pihak sekolah dalam proses pengambilan keputusan.
5. Menyediakan prototipe model prediksi kesiapan siswa yang dapat digunakan sebagai referensi akademik dan belum diuji dalam implementasi jangka panjang di sistem sekolah formal secara luas.

1.5 Manfaat Penelitian

1. Manfaat Teoritis
 - a. Memberikan kontribusi dalam pengembangan kajian interdisipliner antara teknologi informasi dan pendidikan menengah, khususnya dalam penerapan *machine learning* untuk pengambilan keputusan berbasis data pada transisi siswa SD ke SMP.
 - b. Menjadi referensi akademik bagi penelitian selanjutnya yang mengkaji pemanfaatan algoritma *Random Forest* dalam klasifikasi

data pendidikan, khususnya pada jenjang Sekolah Dasar dan transisi menuju Sekolah Menengah Pertama.

2. Manfaat Praktis

- a. Bagi guru dan pihak sekolah menyediakan model bantu berbasis teknologi untuk mengevaluasi kesiapan siswa SD masuk SMP secara lebih objektif dan berbasis data, sehingga mendukung pengambilan keputusan yang lebih terukur dalam merekomendasikan kelanjutan pendidikan siswa.
- b. Bagi orang tua memberikan pemahaman tambahan mengenai faktor-faktor akademik dan non-akademik yang memengaruhi kesiapan anak memasuki jenjang SMP, sehingga orang tua dapat mengambil langkah pendampingan yang lebih tepat.
- c. Bagi pengembang sistem pendidikan menjadi acuan dalam membangun sistem asesmen kesiapan sekolah yang modern, efisien, dan adaptif, khususnya dalam konteks digitalisasi pendidikan di era transformasi pendidikan 4.0.

BAB II

LANDASAN TEORI

2.1 Kesiapan Masuk Sekolah Menengah Pertama

Transisi pendidikan dari sekolah dasar (SD) ke sekolah menengah pertama (SMP) merupakan fase krusial dalam perkembangan siswa. Pada umumnya, siswa memasuki SMP di usia sekitar 12 tahun setelah menyelesaikan kelas VI SD. Kesiapan masuk SMP dapat diartikan sebagai sejauh mana seorang siswa telah memiliki bekal kemampuan dan penyesuaian yang diperlukan untuk berhasil mengikuti pendidikan di jenjang SMP, baik dari segi akademik maupun non-akademik. Anderman (dalam Santrock, 2014) berpendapat bahwa transisi ke sekolah menengah pertama bisa menjadi masa yang sulit dan menimbulkan stres bagi banyak siswa – terjadi perubahan lingkungan belajar, tuntutan akademik yang meningkat, serta dinamika sosial yang berbedas. Oleh karena itu, kesiapan siswa perlu dilihat secara holistik, tidak semata diukur dari capaian akademis atau usia saja. Penelitian Deliviana (2017) menegaskan bahwa kesiapan sekolah mencakup berbagai aspek perkembangan anak secara menyeluruh, meliputi kemampuan kognitif-akademik, kematangan emosi, keterampilan sosial, kemandirian, dan aspek fisik lain – lebih dari sekadar kemampuan membaca, menulis, dan berhitung (calistung) atau faktor usia kronologis (Deliviana, 2017). Dengan demikian, kesiapan masuk SMP seyogianya dipandang sebagai konstruksi multidimensi yang mencakup sejumlah dimensi utama berikut:

- Kesiapan akademik: mengacu pada penguasaan pengetahuan dan keterampilan dasar yang dibutuhkan untuk mengikuti kurikulum SMP.

Siswa yang siap secara akademik memiliki kemampuan literasi dan numerasi yang baik, pemahaman konsep dasar dari mata pelajaran di SD, serta keterampilan belajar (*learning skills*) seperti mencatat, membaca efektif, dan memecahkan masalah. Kesiapan akademik ini penting karena di SMP siswa akan menghadapi materi pelajaran yang lebih kompleks dan mendalam. Studi transisi sekolah menunjukkan bahwa banyak siswa mengalami penurunan prestasi saat memasuki SMP akibat meningkatnya tuntutan akademik dan perubahan metode pembelajaran (Jindal-Snape *et al.*, 2020). Oleh karena itu, fondasi akademik yang kuat di akhir pendidikan dasar sangat berpengaruh terhadap keberhasilan siswa di jenjang berikutnya (Asim *et al.*, 2019).

- Kesiapan sosial-emosional: terkait dengan kemampuan siswa mengelola emosi, berinteraksi secara sehat dengan teman sebaya dan guru, serta beradaptasi dengan perubahan sosial di lingkungan sekolah baru. Transisi ke SMP biasanya disertai perombakan lingkaran pertemanan – siswa harus mampu menjalin pertemanan baru dan bekerja sama dalam kelompok yang lebih beragam. Selain itu, anak perlu memiliki ketahanan terhadap tekanan sosial maupun risiko bullying yang umumnya menjadi kekhawatiran pada masa awal SMP (Zeedyk *et al.*, 2003). Kematangan emosi, kemampuan *self-regulation*, dan keterampilan empati menjadi bagian penting dari kesiapan sosial-emosional ini. Siswa dengan kesiapan sosial-emosional yang baik cenderung lebih mudah menyesuaikan diri dengan

budaya dan aturan di SMP, sehingga mengurangi tingkat stres dan kecemasan selama masa transisi (Ng-Knight *et al.*, 2016).

- Kesiapan kemandirian: yaitu kesiapan dalam hal kemandirian belajar dan tanggung jawab pribadi. Di SMP, siswa dituntut lebih mandiri dibanding saat di SD – misalnya mengatur jadwal belajar harian, mengerjakan pekerjaan rumah tanpa pengawasan ketat, berpindah kelas dengan tertib, dan membawa perlengkapan yang diperlukan untuk pelajaran yang berbeda. Kemandirian ini juga mencakup keterampilan mengelola waktu, mengambil inisiatif untuk belajar, serta kemampuan membuat keputusan sederhana secara mandiri. Siswa yang telah dilatih mandiri sejak kelas atas di SD biasanya lebih siap menghadapi tuntutan jenjang SMP. Pembinaan kemandirian sejak dini terbukti membantu anak lebih percaya diri dan mampu beradaptasi dengan kehidupan SMP yang menuntut lebih banyak tanggung jawab (Bogensneider, 2015). Sebaliknya, anak yang terlalu bergantung pada instruksi guru atau orang tua mungkin mengalami kesulitan menavigasi lingkungan SMP yang lebih menuntut *self-direction*.
- Kesiapan adaptasi lingkungan: meliputi kemampuan menyesuaikan diri dengan lingkungan fisik dan tata organisasi sekolah yang baru. Berbeda dengan SD di mana siswa biasanya berada dalam satu kelas yang sama sepanjang hari dengan satu wali kelas, di SMP siswa harus beradaptasi dengan ruang kelas berbeda untuk mata pelajaran yang berbeda, jadwal pelajaran yang lebih kompleks, dan lebih banyak guru mata pelajaran. Perubahan struktur ini memerlukan kemampuan penyesuaian

terhadap lingkungan baru – seperti mengenal tata tertib SMP, ukuran sekolah yang lebih besar, serta budaya sekolah yang mungkin berbeda dari sekolah dasar sebelumnya. Siswa perlu mengembangkan keterampilan navigasi (misalnya mengetahui lokasi kelas atau fasilitas sekolah), manajemen jadwal, dan kemampuan menghadapi aturan-aturan yang mungkin lebih ketat. Pengenalan awal terhadap lingkungan SMP dapat membantu; Dr. Siti Nurani (2018) menemukan bahwa mengajak anak mengunjungi sekolah barunya, berkenalan dengan beberapa guru, serta memahami fasilitas yang ada sebelum tahun ajaran dimulai mampu mengurangi kecemasan dan meningkatkan kepercayaan diri siswa saat masuk SMP.

Selain dimensi-dimensi di atas, terdapat faktor-faktor pendukung yang memengaruhi tingkat kesiapan transisi siswa dari SD ke SMP. Faktor internal yang penting antara lain motivasi belajar, kepercayaan diri, dan kondisi kesehatan/kebugaran anak. Siswa dengan motivasi tinggi dan *mindset* positif cenderung memandang transisi sebagai tantangan positif, bukan ancaman, sehingga lebih proaktif beradaptasi (Carter *et al.*, 2015). Kemudian, dukungan keluarga berperan sangat signifikan. Penelitian Hattie (2013) terhadap ribuan siswa menunjukkan bahwa dukungan emosional yang stabil dari orang tua berkorelasi positif dengan peningkatan prestasi akademik anak. Lingkungan rumah yang aman secara emosional membuat anak lebih percaya diri menghadapi perubahan di SMP. Komunikasi yang efektif antara orang tua dan anak juga faktor kunci: Bogenschneider (2015) menemukan bahwa anak yang merasa bisa

berdialog terbuka dengan orang tuanya cenderung lebih siap secara emosional dan akademis saat bertransisi ke sekolah baru. Keterlibatan orang tua dalam pendidikan – misalnya rutin hadir dalam pertemuan orang tua-guru atau aktif memantau perkembangan belajar anak – turut mempermudah adaptasi siswa di SMP (Prayitno, 2017). Hal ini karena anak merasa pendidikannya mendapat perhatian, dan orang tua pun lebih memahami kebutuhan anak dalam masa transisi.

Dari sisi sekolah, lingkungan dan pembiasaan di SD juga memengaruhi kesiapan ke SMP. Sekolah dasar yang menanamkan disiplin, budaya belajar yang positif, dan nilai-nilai karakter akan menghasilkan lulusan yang lebih tangguh menghadapi jenjang selanjutnya. Hubungan guru-siswa yang suportif di SD dapat meningkatkan *self-esteem* siswa (Ashton, 2008) sehingga mereka lebih siap bersosialisasi di SMP. Program orientasi atau masa pengenalan di awal masuk SMP yang baik juga sangat membantu mengurangi *culture shock*. Sementara itu, faktor-faktor eksternal lain seperti lingkungan teman sebaya dan kondisi sosial-ekonomi keluarga turut memberikan pengaruh. Siswa dari keluarga yang mendukung pendidikan dan lingkungan komunitas yang kondusif biasanya memiliki kesiapan lebih baik, sedangkan kondisi ekonomi yang sulit atau perpindahan domisili yang jauh bisa menjadi tantangan tambahan dalam transisi. Dengan memahami berbagai faktor ini, pihak sekolah dan keluarga dapat bekerja sama memfasilitasi transisi yang lebih mulus bagi siswa.

2.2 Peran Pendidikan Dasar dalam Kesiapan ke Sekolah Menengah

Pendidikan dasar (SD) memiliki peran fundamental dalam membentuk kesiapan siswa untuk melanjutkan ke pendidikan menengah. Sekolah dasar merupakan pondasi di mana siswa menguasai literasi dasar, numerasi, dan pengetahuan umum yang akan menjadi prasyarat dalam mengikuti kurikulum SMP. Jika pondasi ini lemah, siswa berisiko mengalami kesulitan akademis di jenjang selanjutnya. Laporan Bank Dunia oleh Asim *et al.* (2019) mengenai hasil belajar di sekolah-sekolah dasar menunjukkan bahwa rendahnya kemampuan literasi dan matematika di akhir pendidikan dasar berkorelasi dengan rendahnya performa siswa di jenjang menengah. Dengan kata lain, peningkatan kualitas pembelajaran di SD secara langsung berimplikasi pada kesiapan akademik siswa saat masuk SMP (Asim *et al.*, 2019).

Dari sisi kurikulum, kurikulum tingkat SD dirancang untuk mengembangkan kompetensi dasar dan karakter siswa. Kurikulum 2013 (K-13) misalnya, menekankan pendekatan tematik terpadu dan pendidikan karakter sejak SD. Melalui berbagai mata pelajaran dan kegiatan ekstrakurikuler di SD, siswa tidak hanya dibekali pengetahuan kognitif, tetapi juga dibentuk sikap disiplin, rasa ingin tahu, kerjasama, dan tanggung jawab. Pembiasaan positif yang dilakukan di SD – seperti membiasakan membaca buku setiap hari, melatih siswa mengerjakan tugas tepat waktu, upacara bendera untuk melatih nasionalisme dan kedisiplinan, serta kegiatan keagamaan rutin – semuanya berkontribusi pada kesiapan mental dan moral siswa untuk jenjang berikutnya. Pendidikan karakter yang konsisten di SD diyakini dapat membangun landasan sikap dan etika yang kuat sehingga siswa

lebih siap menghadapi tantangan sosial di SMP (Kementerian Pendidikan dan Kebudayaan, 2017).

Peran guru SD juga sangat strategis dalam menyiapkan siswa ke jenjang SMP. Guru kelas di SD berperan ganda sebagai pengajar mata pelajaran sekaligus pembimbing perkembangan anak. Melalui interaksi sehari-hari, guru dapat menanamkan kepercayaan diri dan kemandirian pada siswa. Misalnya, guru dapat memberi tugas proyek sederhana atau *presentation* sejak kelas V–VI SD untuk melatih tanggung jawab dan keberanian siswa. Mariano *et al.* (2019) dalam tinjauannya menyatakan bahwa pendekatan pembelajaran aktif berbasis permainan dan discovery learning di pendidikan dasar dapat meningkatkan rasa ingin tahu, fokus perhatian, dan percaya diri siswa. Hasilnya, anak memasuki lingkungan SMP dengan sikap positif terhadap belajar dan tidak takut mencoba hal-hal baru. Selain itu, guru dapat mengidentifikasi sedini mungkin kelemahan dan kelebihan setiap siswa selama di SD, sehingga bisa diberikan intervensi atau pengayaan yang tepat sebelum siswa lulus. Sebagai contoh, siswa yang masih lemah dalam literasi dapat dibantu melalui program remedial membaca di kelas VI, sehingga pada saat masuk SMP kemampuan literasinya cukup memadai untuk mengikuti pelajaran yang lebih kompleks.

Lingkungan belajar di SD yang kondusif dan inklusif juga mempersiapkan siswa secara sosial. Sekolah dasar yang mendorong siswa untuk berkolaborasi dalam kelompok, berpartisipasi dalam kegiatan ekstrakurikuler (seperti pramuka, olahraga, atau kesenian), serta menghargai keberagaman teman sebaya, akan menghasilkan lulusan yang lebih terampil beradaptasi dengan lingkungan SMP

yang lebih heterogen. Ghandour *et al.* (2024) menyebutkan bahwa pengalaman positif di pendidikan anak usia dini dan dasar – terutama dalam hal penguatan fungsi regulasi diri dan keterampilan sosial-emosional – berkontribusi pada capaian akademik jangka panjang. Hal ini mengindikasikan bahwa siswa yang telah terbiasa mengendalikan emosi, bekerja sama, dan mematuhi aturan di SD cenderung memiliki transisi yang lebih mulus ke SMP.

Tidak kalah penting adalah peran kebijakan dan program transisi dari SD ke SMP. Di beberapa negara maju, terdapat program transisi formal yang mempersiapkan siswa akhir SD untuk masuk ke sekolah menengah, misalnya dengan kunjungan ke SMP, *shadowing* kakak kelas, atau pembekalan keterampilan belajar. Di Indonesia, upaya serupa bisa dilakukan dalam skala sekolah atau daerah, misalnya program orientasi terintegrasi antara SD dan SMP. Sekolah dasar dapat berkoordinasi dengan SMP tujuan untuk sinkronisasi kurikulum dan ekspektasi. Guru SMP dapat memberikan sesi pengenalan mata pelajaran kepada siswa SD kelas VI, begitu pula guru SD dapat menerangkan kepada siswa tentang hal-hal yang akan mereka hadapi di SMP. Langkah-langkah ini membantu mengurangi kesenjangan (*disconnect*) antara dua jenjang tersebut. Dengan demikian, pendidikan dasar bukan hanya bertanggung jawab menyampaikan materi kurikulum SD, tetapi juga berperan sebagai jembatan yang mempersiapkan peserta didik untuk sukses menapaki jenjang pendidikan menengah.

2.3 Pendekatan Data Driven Decision Making Dalam Pendidikan

Data-Driven Decision Making (DDDM) adalah pendekatan sistematis dalam pengambilan keputusan yang berlandaskan pada analisis data yang valid dan relevan. Dalam konteks pendidikan, DDDM berarti praktik menggunakan berbagai data tentang siswa – misalnya nilai akademik, hasil asesmen, tingkat kehadiran, catatan perilaku, hingga perkembangan sosial-emosional – untuk merumuskan kebijakan atau tindakan pedagogis yang tepat (Mandinach & Gummer, 2016). Pendekatan ini menuntut pendidik untuk tidak semata mengandalkan intuisi, kebiasaan, atau asumsi subjektif dalam membuat keputusan, melainkan menggunakan bukti empiris yang diperoleh dari data nyata. Menurut Mandinach dan Gummer (2016), proses DDDM bersifat *iterative* dan meliputi serangkaian tahap: pengumpulan data, analisis data, interpretasi temuan, dan aksi tindak lanjut. Secara rinci, tahapan DDDM dapat diuraikan sebagai berikut:

1. Pengumpulan data: Mengidentifikasi dan mengumpulkan data relevan mengenai siswa. Data ini dapat berupa nilai ujian, hasil tugas, asesmen diagnostik, catatan absensi, hasil observasi guru mengenai perilaku siswa, hingga kuesioner psikologis atau survei kepuasan siswa. Penting dipastikan bahwa data yang dikumpulkan akurat, terbaru, dan mencakup berbagai aspek perkembangan siswa.
2. Analisis data: Melakukan pengolahan dan analisis terhadap data yang telah dikumpulkan. Analisis bisa dilakukan secara manual (misalnya guru menganalisis pola nilai siswa dalam satu kelas) maupun dengan bantuan

alat analitik dan teknologi. Contohnya, penggunaan perangkat lunak statistik atau *dashboard* pendidikan untuk melihat tren performa siswa, atau penerapan algoritma *machine learning* untuk menemukan pola tersembunyi dalam data dalam jumlah besar. Tujuan tahap ini adalah mengekstraksi informasi bermakna dari kumpulan data mentah.

3. Interpretasi: Menafsirkan hasil analisis data dalam konteks praktik pendidikan. Pada tahap ini, pendidik berusaha memahami implikasi dari temuan data – mengapa suatu pola dapat terjadi dan apa artinya bagi pembelajaran. Misalnya, jika dari data diketahui bahwa skor matematika rata-rata kelas menurun pada topik tertentu, guru perlu menyelidiki penyebabnya: apakah karena metode mengajar yang kurang sesuai, prasyarat materi yang belum dikuasai siswa, atau faktor kelelahan di akhir semester. Interpretasi yang tepat akan mengarahkan pada keputusan yang tepat pula.
4. Tindakan (Pengambilan keputusan): Mengambil keputusan atau tindakan strategis berdasarkan insight data yang telah diinterpretasi. Contohnya, guru mungkin memutuskan memberikan intervensi remedial bagi siswa yang terdeteksi belum siap (misalnya siswa yang skor ujiannya di bawah kriteria) atau menyesuaikan metode pengajaran di kelas agar lebih efektif. Pada level sekolah, kepala sekolah dapat menggunakan data untuk memutuskan perlu tidaknya program tutoring tambahan, pelatihan guru, atau perubahan kurikulum. Inti tahap ini adalah memastikan

bahwa *evidence* yang diperoleh benar-benar digunakan untuk perbaikan praktik, bukan sekadar laporan.

Pendekatan DDDM memungkinkan keputusan pendidikan menjadi lebih tepat sasaran dan terukur. Dengan DDDM, sekolah dapat melakukan personalisasi pembelajaran dan deteksi dini risiko. Sebagai ilustrasi, data hasil belajar dan observasi di SD dapat dianalisis untuk mengidentifikasi siswa kelas VI yang berpotensi belum siap masuk SMP. Siswa tersebut kemudian bisa diberikan program pendampingan khusus selama sisa waktu di SD, sehingga kesenjangan kesiapan dapat diperkecil sebelum transisi terjadi. Tanpa DDDM, upaya semacam ini sulit dilakukan karena guru mungkin tidak menyadari masalah setiap individu siswa secara objektif. Retnawati *et al.* (2021) menunjukkan bahwa banyak pendidik PAUD dan SD di Indonesia masih bergantung pada observasi subjektif dalam menilai perkembangan dan kesiapan anak, tanpa alat ukur data yang sistematis. Akibatnya, potensi bias dan kesalahan penilaian cukup tinggi. DDDM menawarkan solusi atas tantangan ini dengan menghadirkan evaluasi yang lebih obyektif dan menyeluruh.

Penelitian dan praktik di berbagai negara telah membuktikan manfaat DDDM dalam meningkatkan kualitas pendidikan. Bousdekis *et al.* (2021) menegaskan bahwa implementasi DDDM menghasilkan keputusan pendidikan yang lebih terukur, proaktif, dan minim bias dibanding pendekatan tradisional. Hal ini sejalan dengan visi *Pendidikan 4.0* yang menekankan integrasi teknologi dan data dalam pengelolaan pendidikan. Gullo (2013) menemukan bahwa di jenjang PAUD, guru yang rutin menganalisis data perkembangan anak (misalnya

pencapaian literasi awal) mampu merancang instruksi yang lebih efektif, sehingga meningkatkan kesiapan anak memasuki jenjang dasar. Demikian pula, studi Mandinach dan Light (2006) terhadap beberapa distrik sekolah di Amerika Serikat mengungkapkan bahwa penggunaan alat bantu analitik berbasis data secara signifikan memperkuat strategi pengajaran guru, asesmen formatif, dan efisiensi pengambilan keputusan di sekolah. Dengan kata lain, DDDM membantu pendidik berpindah dari pendekatan *trial and error* menuju pendekatan berbasis bukti (evidence-based).

Namun, penerapan DDDM yang sukses memerlukan kesiapan institusional dan budaya data yang mendukung. Hamash *et al.* (2024) mencatat bahwa ketersediaan infrastruktur teknologi, kompetensi guru dalam literasi data, serta dukungan kepemimpinan sekolah merupakan faktor penentu apakah DDDM dapat berjalan optimal. Tanpa komitmen pimpinan dan pelatihan memadai, inisiatif DDDM bisa terhambat oleh resistensi perubahan atau sekadar menjadi formalitas pencatatan data tanpa analisis bermakna. Di era kecerdasan buatan saat ini, DDDM semakin powerful jika dikombinasikan dengan teknologi *machine learning*. Integrasi DDDM dengan algoritma ML (misalnya *Random Forest*, yang akan dibahas pada subbab berikut) telah terbukti mampu memperkuat kemampuan prediksi dalam pendidikan, seperti memprediksi kesiapan siswa dengan lebih presisi (Alghamdi & Rahman, 2023). Melalui dukungan ML, analisis data dapat mencakup volume lebih besar dan variabel lebih kompleks daripada yang sanggup diolah secara manual, sehingga sekolah dapat mengidentifikasi pola performa dan faktor risiko dengan akurasi lebih tinggi. Selain itu, ML juga

menyediakan *tool* seperti *feature importance* yang memberi wawasan tentang faktor mana yang paling memengaruhi hasil prediksi (misalnya faktor-faktor kunci dalam kesiapan siswa). Wawasan ini pada gilirannya memperkaya proses pengambilan keputusan berbasis data – guru dan kepala sekolah tidak hanya tahu *siapa* yang membutuhkan bantuan, tapi juga *mengapa* dan pada aspek apa intervensi diperlukan.

Singkatnya, DDDM merepresentasikan pergeseran paradigma pengelolaan pendidikan: dari sekadar mengandalkan pengalaman dan insting, menuju praktek profesional yang menggabungkan *data analytics* dalam setiap pengambilan keputusan. Dalam konteks penilaian kesiapan siswa SD masuk SMP, pendekatan DDDM sangat potensial untuk menghasilkan asesmen yang akurat dan adil, membantu merancang strategi pembelajaran yang terpersonalisasi, serta merencanakan intervensi sedini mungkin bagi siswa yang membutuhkan. Hal ini selaras dengan tujuan peningkatan kualitas pendidikan jangka panjang, di mana setiap keputusan didukung oleh bukti (*data-driven*) demi hasil belajar siswa yang optimal.

2.4 Machine Learning Dalam Dunia Pendidikan

2.4.1 Pengertian dan Prinsip Dasar *Machine Learning*

Machine Learning (ML) adalah cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan algoritma yang memungkinkan komputer belajar dari data. Artinya, sistem ML mampu meningkatkan kinerja dalam menyelesaikan suatu tugas seiring bertambahnya pengalaman (data) tanpa diprogram ulang secara eksplisit.

Mitchell (1997) secara klasik mendefinisikan konsep ini: sebuah program komputer dikatakan belajar dari pengalaman E terhadap beberapa kelas tugas T , jika performanya dalam tugas di T meningkat seiring dengan pengalaman E . Dalam konteks pendidikan, “pengalaman” tersebut berupa data historis siswa (misalnya kumpulan nilai, riwayat interaksi dalam *e-learning*, atau data kehadiran), “tugas” bisa berupa prediksi atau klasifikasi (misalnya memprediksi apakah seorang siswa siap masuk SMP atau tidak), dan “kinerja” diukur dari akurasi hasil prediksi/klasifikasi tersebut. Dengan demikian, ML pada hakikatnya adalah pendekatan komputasi yang mengenali pola dalam data dan menggunakan pola tersebut untuk melakukan prediksi atau pengambilan keputusan di masa depan (Baker & Inventado, 2014).

Prinsip kerja ML secara umum mencakup beberapa tahapan utama. Pertama, pengumpulan dan preparasi data, di mana data dikumpulkan kemudian diproses (misalnya dibersihkan, diberi label, dinormalisasi) agar siap digunakan. Kedua, pemilihan model atau algoritma ML yang sesuai dengan jenis tugas (klasifikasi, regresi, klastering, dsb). Ketiga, pelatihan (training) model, yaitu proses algoritma belajar dari *training set* (data berlabel) untuk memetakan pola hubungan antara fitur (variabel input) dan label (target output). Selama tahap ini, model secara iteratif memperbarui parameter internalnya untuk meminimalkan kesalahan prediksi. Keempat, evaluasi model menggunakan *data uji* (*test set*) yang terpisah, guna mengukur seberapa baik model yang dilatih dapat generalisasi ke data baru yang belum pernah dilihat – metrik evaluasi yang umum dipakai antara lain akurasi, presisi, recall, *F1-score*, atau *mean squared*

error tergantung jenis tugasnya. Jika performa dirasa cukup baik, tahap terakhir adalah deployment atau penggunaan model pada data baru secara aktual (misal memprediksi kesiapan siswa tahun ajaran baru berdasarkan model yang telah dilatih). Proses ini bisa berulang (iteratif) – jika kinerja model belum memuaskan, dilakukan penyesuaian seperti mengganti algoritma, menambah data, atau tuning parameter.

Dalam ML terdapat beberapa kategori pembelajaran utama (Mitchell, 1997; Peña-Ayala, 2014):

- **Supervised learning:** Model dilatih menggunakan data yang berlabel (tersedia jawaban benar). Tujuannya untuk memprediksi label output berdasarkan input yang diberikan. Contoh: prediksi *binary* “Siap/Belum Siap” masuk SMP berdasarkan data profil siswa kelas VI. Algoritma yang termasuk kategori ini antara lain *Random Forest*, *Naïve Bayes*, *Support Vector Machine (SVM)*, *Neural Networks*, dll. *Supervised learning* sangat sesuai untuk tugas seperti klasifikasi kesiapan siswa atau prediksi nilai ujian karena kita memiliki *ground truth* untuk melatih model (misal label kesiapan berdasarkan penilaian guru atau hasil psikotes).
- **Unsupervised learning:** Model dilatih pada data yang tidak berlabel, bertujuan untuk menemukan pola atau struktur tersembunyi dalam data. Contohnya *clustering* siswa berdasarkan kemiripan karakteristik untuk melihat segmen-segmen kesiapan yang ada tanpa tahu kategori di awal. Algoritma seperti *KMeans clustering* atau *Hierarchicacustering* termasuk dalam unsupervised.

Pendekatan ini berguna misalnya untuk mengelompokkan siswa dengan kebutuhan dukungan serupa (tanpa menentukan siap/tidaknya di awal), sehingga intervensi bisa diberikan per kelompok karakteristik.

- Reinforcement learning: Model belajar melalui mekanisme *trial and error* dengan mendapatkan *feedback* berupa *reward* atau *penalty*.

Algoritma jenis ini banyak dipakai di robotik dan game, dan relatif jarang dalam konteks pendidikan anak usia sekolah dasar/menengah. Namun, konsepnya dapat digunakan misalnya dalam adaptasi sistem tutoring cerdas yang memberikan *reward* jika siswa menjawab benar dan *adjustment* jika salah, sehingga sistem belajar strategi optimal untuk membimbing siswa. Kategori ini berbeda dari dua sebelumnya karena tidak ada pasangan input-output statis melainkan interaksi dinamis dengan lingkungan.

Dengan kemajuan teknologi informasi, penerapan ML di dunia pendidikan berkembang sangat pesat seiring meningkatnya ketersediaan *big data* pendidikan. Baker dan Inventado (2014) menyebut kekuatan utama ML terletak pada kemampuannya mengolah data berukuran besar dan kompleks, mengidentifikasi pola hubungan antar variabel yang tidak tampak secara kasatmata, serta menghasilkan prediksi berbasis pola historis secara otomatis. Hal-hal yang sebelumnya sulit dikerjakan secara manual oleh pendidik (misalnya memantau perkembangan ribuan siswa secara individual) kini dapat dibantu dengan sistem ML yang bekerja di belakang layar.

2.4.2 Peran *Machine Learning* dalam Prediksi Pendidikan Siswa

Dalam konteks prediksi pendidikan, ML telah dimanfaatkan untuk berbagai tujuan, antara lain memproyeksikan performa akademik siswa, mendeteksi risiko dropout (putus sekolah), mengevaluasi keefektifan metode pengajaran, merekomendasikan gaya belajar yang sesuai, hingga memprediksi kesiapan siswa memasuki jenjang pendidikan berikutnya (Baker & Inventado, 2014; Alghamdi & Rahman, 2023). Secara khusus, prediksi school readiness (kesiapan sekolah) telah menjadi salah satu bidang aplikasi ML yang penting. Sebelumnya, penilaian kesiapan siswa sering kali dilakukan dengan tes tunggal atau observasi subyektif guru, tetapi kini model ML dapat dilatih untuk memperhitungkan multifaktor secara bersamaan sehingga hasilnya lebih komprehensif. Peña-Ayala (2014) mengkaji bahwa penerapan ML dalam lingkungan belajar umumnya terbagi dalam tiga fungsi utama:

- Profiling siswa: mendeteksi karakteristik, pola belajar, dan kebutuhan khusus masing-masing siswa. Dengan ML, dapat dibuat profil yang lebih mendalam – misalnya mengidentifikasi siswa mana yang memiliki gaya belajar visual vs. verbal, atau yang cenderung aktif vs. pasif di kelas. Profiling ini membantu guru menyesuaikan pendekatan mengajar sesuai karakteristik siswa.
- Deteksi dini masalah: mengidentifikasi secara *early warning* siswa-siswa yang kemungkinan menghadapi kesulitan, baik kesulitan akademis maupun sosial-emosional. Contoh, model ML dapat memprediksi siswa yang berpotensi belum siap secara akademik di SMP berdasarkan tren nilai

mapel inti di kelas V–VI SD, atau memprediksi risiko masalah penyesuaian sosial berdasarkan data kehadiran dan catatan kedisiplinan. Deteksi dini ini memungkinkan intervensi proaktif sebelum masalah berkembang lebih serius.

- Rekomendasi dan personalisasi intervensi: merancang saran atau strategi pembelajaran yang sesuai dengan kekuatan dan kelemahan siswa. Misalnya, ML dapat merekomendasikan modul pengayaan bagi siswa yang cepat memahami materi, atau program *remedial* spesifik bagi siswa yang lemah di topik tertentu. Bahkan dalam skala yang lebih luas, ML bisa membantu menyusun kurikulum adaptif yang berbeda jalur sesuai kebutuhan siswa.

Ketiga peran di atas sangat relevan dalam upaya asesmen dan peningkatan kesiapan siswa SD ke SMP, yang notabene membutuhkan pemantauan multi-aspek. Dalam praktiknya, sudah banyak penelitian membuktikan efektivitas ML untuk prediksi di bidang pendidikan. Adhikari *et al.* (2023) menunjukkan bahwa algoritma *Random Forest* mampu mengidentifikasi anak-anak usia prasekolah yang berisiko tidak siap sekolah berdasarkan data indikator kesehatan (seperti indeks massa tubuh), tingkat kehadiran, dan status perkembangan sosial-emosional. Pendekatan prediktif ini memungkinkan intervensi dilakukan lebih tepat sasaran sebelum anak memasuki jenjang berikutnya. Sementara itu, studi Umam & Kurniawan (2024) di Indonesia mengembangkan sistem berbasis web menggunakan algoritma ML untuk mengklasifikasikan kesiapan siswa TK masuk SD berdasarkan data perkembangan kognitif, sosial, dan emosional. Hasil

penelitian mereka membuktikan bahwa model ML tidak hanya akurat memprediksi kesiapan, tetapi juga mempercepat dan mempermudah guru dalam melakukan asesmen secara objektif (Umam & Kurniawan, 2024).

Temuan serupa dilaporkan oleh Alghamdi & Rahman (2023) yang membandingkan beberapa algoritma ML (Naïve Bayes, *decision tree* J48, dan Random Forest) dalam memprediksi keberhasilan akademik siswa. Hasilnya, Random Forest memberikan performa terbaik dengan akurasi klasifikasi tertinggi untuk memprediksi siswa yang akan lulus tepat waktu vs. tidak, mengungguli model lainnya (Alghamdi & Rahman, 2023). Tak hanya untuk prediksi kesiapan secara agregat, ML juga berperan dalam membantu guru memahami faktor-faktor penentu keberhasilan belajar. Misalnya, studi Feher (2024) menunjukkan bahwa data perilaku, kehadiran, dan hasil asesmen perkembangan anak usia dini dapat diproses dengan teknik ML untuk memetakan kesiapan anak berdasarkan faktor-faktor kritis seperti kemampuan motorik, regulasi emosi, dan interaksi sosial. Bagi guru SD dan SMP, informasi semacam ini sangat berharga – guru dapat mengetahui aspek mana yang harus difokuskan dalam pembelajaran atau bimbingan. Secara umum, keunggulan ML dibanding metode konvensional terletak pada kemampuannya menangani data pendidikan yang kompleks, berukuran besar, dan heterogen (Peña-Ayala, 2014).

Dalam konteks asesmen kesiapan, data siswa bisa berasal dari berbagai sumber (rapor akademik, hasil psikotes, catatan guru, hingga input dari orang tua). Dengan ML, seluruh data multi-sumber tersebut dapat diintegrasikan dan dianalisis bersamaan, menghasilkan prediksi yang lebih akurat dan holistik. Tentu

hal ini sulit dicapai jika hanya mengandalkan penilaian manual. Oleh sebab itu, integrasi ML ke dalam sistem pendidikan modern dipandang sebagai salah satu fondasi penting untuk mendorong praktik pendidikan berbasis bukti dan *data-driven*, termasuk dalam hal transisi antarkurikulum.

2.4.3 Algoritma *Random Forest* dan *Feature Importance* dalam Pendidikan

Salah satu algoritma ML yang paling menonjol dan banyak digunakan dalam analisis data pendidikan adalah *Random Forest*. Algoritma ini diperkenalkan oleh Leo Breiman pada tahun 2001 sebagai pengembangan dari metode *bagging* dan pohon keputusan (CART). Secara sederhana, *Random Forest* merupakan metode *ensemble learning* yang menggabungkan banyak *decision tree* untuk menghasilkan model prediksi yang lebih kuat. Cara kerjanya, *Random Forest* membangun sejumlah *tree* (misal puluhan atau ratusan pohon keputusan) dari *subsets* data latih yang dipilih secara acak dengan *bootstrapping*. Selain itu, pada setiap node pohon keputusan, fitur yang digunakan untuk pemisahan juga dipilih secara acak dari subset fitur. Kombinasi kedua randomisasi ini (random pada data dan pada fitur) membuat pohon-pohon yang dihasilkan bervariasi. Ketika membuat prediksi, *Random Forest* menggabungkan hasil semua pohon – untuk tugas klasifikasi biasanya melalui voting mayoritas (kelas yang paling banyak diprediksi oleh tree akan menjadi hasil akhir), sedangkan untuk regresi melalui rata-rata nilai prediksi.

Pendekatan *ensemble* ini memberikan beberapa keunggulan dibanding satu pohon keputusan tunggal: (a) *Random Forest* cenderung memiliki akurasi lebih tinggi karena kesalahan model cenderung saling dikoreksi antar pohon

(model *robust* terhadap varians data), (b) lebih tahan terhadap risiko overfitting – meskipun masing-masing pohon mungkin overfit terhadap subset datanya, rata-rata banyak pohon menghasilkan generalisasi lebih baik, (c) mampu menangani dataset dengan jumlah fitur yang besar dan tipe fitur yang beragam (numerik maupun kategorikal), (d) relatif mudah digunakan karena memiliki sedikit *hyperparameter* yang perlu dituning dibanding model kompleks lainnya, sehingga sering menjadi pilihan praktisi untuk baseline model. Penelitian oleh Ramadhan *et al.* (2019) membuktikan bahwa *Random Forest* mampu mengklasifikasikan mutu pendidikan (tingkat pencapaian sekolah) dengan akurasi tinggi, sekaligus mengidentifikasi variabel-variabel penting yang memengaruhi hasil tersebut. Temuan serupa dilaporkan oleh Suleiman *et al.* (2017) dan Adhikari *et al.* (2023) yang menggunakan *Random Forest* untuk prediksi kesiapan sekolah: model yang dihasilkan tidak hanya andal dalam klasifikasi anak yang siap atau berisiko, tetapi juga memberikan peta faktor mana saja yang berkontribusi signifikan dalam kesiapan akademik.

Dalam konteks studi transisi SD ke SMP, algoritma *Random Forest* sangat relevan untuk membangun model prediksi kesiapan siswa. Umam & Kurniawan (2024) misalnya, menggunakan *Random Forest* untuk memprediksi kesiapan siswa (pada kasus TK ke SD) dan mendapati model ini memberikan hasil klasifikasi yang cepat dan akurat, sehingga dapat dijadikan referensi bagi guru dalam pengambilan keputusan. Selain itu, Alghamdi & Rahman (2023) yang menguji berbagai algoritma ML menemukan *Random Forest* konsisten unggul dalam memprediksi kesuksesan akademik siswa sekolah, mengalahkan algoritma

seperti *Naïve Bayes* maupun pohon keputusan tunggal (J48). Keandalan *Random Forest* juga terletak pada kemampuannya menyediakan beragam indikator evaluasi model secara internal, seperti *out-of-bag error* (error estimasi tanpa perlu dataset uji terpisah) dan *feature importance* yang akan dijelaskan berikutnya. Bagi pendidik, hal ini berarti *Random Forest* bukan sekadar “kotak hitam” yang memberikan output prediksi, tetapi juga menyertakan informasi diagnostik: mengapa sebuah prediksi bisa dihasilkan dan faktor mana yang paling memengaruhinya. Dalam setting pendidikan, transparansi ini sangat penting agar guru dan sekolah dapat mempercayai model dan memahami hasilnya.

Feature importance adalah konsep dalam *Random Forest* (juga ada dalam beberapa algoritma lain) yang mengukur kontribusi relatif setiap fitur input terhadap keputusan prediksi model. *Random Forest* menghitung *feature importance* dengan mengevaluasi berapa besar penurunan impuritas (misalnya impuritas Gini atau entropi) pada node-node keputusan yang disebabkan oleh suatu fitur, diakumulasikan di seluruh pohon, lalu dinormalisasi (Breiman, 2001). Intuitifnya, jika suatu variabel (misal *nilai matematika SD*) sering digunakan dalam banyak pohon dan menghasilkan pemisahan yang bersih antara kelas siap/tidak siap, maka variabel tersebut akan diberi skor importance tinggi. Sebaliknya, fitur yang jarang terpakai atau tidak banyak membantu memisahkan data akan mendapat skor rendah. Total skor semua fitur biasanya dinormalisasi menjadi 1 (atau 100%), sehingga kita dapat melihat persentase pengaruh masing-masing fitur terhadap model.

Penerapan *Feature Importance* dalam dunia pendidikan memberikan

wawasan yang sangat berharga, khususnya bagi analisis kesiapan siswa. Model prediksi mungkin menggunakan puluhan variabel (contoh: nilai mata pelajaran, hasil psikotes, sikap, kehadiran, usia, jenis kelamin, dll), namun guru tentunya perlu tahu faktor mana yang paling dominan agar dapat menindaklanjuti temuan tersebut. Dengan feature importance, model Random Forest dapat menunjukkan secara obyektif bahwa misalnya prestasi akademik di mata pelajaran utama dan kedisiplinan (catatan perilaku) adalah dua faktor teratas yang menentukan kesiapan sukses di SMP, sementara faktor seperti usia atau jenis kelamin mungkin pengaruhnya sangat kecil. Informasi ini membantu pihak sekolah fokus pada faktor krusial. Jika diketahui kedisiplinan (misal kebiasaan mengerjakan PR dan datang tepat waktu) punya pengaruh besar, sekolah SD dapat memperkuat pembiasaan disiplin di kelas VI. Begitu pula bila ternyata *self-regulation* atau kemampuan kontrol diri (yang tercermin dari catatan perilaku) tinggi pentingnya, maka program bimbingan konseling dan pelatihan kecakapan sosial-emosional di akhir SD bisa ditingkatkan.

Keunggulan utama dari analisis *feature importance* ialah memberikan transparansi pada model ML yang kompleks. Edukator tidak hanya menerima hasil prediksi (misal “Anak A diprediksi belum siap”), tapi juga dapat menelusuri *mengapa* prediksi tersebut dibuat oleh model – fitur mana yang nilainya mendorong model ke prediksi tersebut. Hal ini menjadikan Random Forest lebih interpretabel dibanding banyak algoritma canggih lainnya, sehingga cocok dengan pendekatan DDDM yang menuntut akuntabilitas dan kejelasan bukti. Dengan mengetahui faktor penentu, guru dapat menjelaskan kepada orang

tua atau stakeholder lain alasan di balik asesmen kesiapan seorang siswa, misalnya: “Model memperkirakan Ananda belum siap masuk SMP terutama karena nilai literasi dan matematika dasar yang masih di bawah standar, ditambah catatan kemandirian yang perlu ditingkatkan (misal sering lupa tugas). Oleh karenanya, perlu program pengayaan membaca dan pembinaan kemandirian sebelum Ananda masuk SMP.” Paparan semacam itu dimungkinkan karena adanya output *feature importance*.

Perlu dicatat bahwa *feature importance* bersifat relatif terhadap data latih yang digunakan. Jika data yang dimasukkan bias atau tidak distandardisasi, skor importance bisa menyesatkan. Misalnya, variabel dengan skala nilai yang lebih besar cenderung terlihat lebih “penting” jika tidak dilakukan normalisasi. Oleh sebab itu, praktik *preprocessing* data yang benar (seperti normalisasi, transformasi, atau seleksi fitur yang redundant) tetap diperlukan agar interpretasi *feature importance* akurat. Meskipun demikian, secara umum *feature importance* telah menjadi *tool* integral dalam evaluasi model pendidikan modern karena kemampuannya menjembatani aspek teknis dan praktis. Penelitian Puspitorini & Kahar (2024) bahkan menyebut bahwa sekalipun model ML sederhana seperti *Naïve Bayes* dapat melakukan klasifikasi dengan baik, namun keunggulan *Random Forest* dalam menyajikan daftar fitur penting membuatnya lebih informatif bagi pendidik. Di akhir, dapat disimpulkan bahwa penggunaan algoritma *Random Forest* bersama analisis *feature importance* menawarkan

pendekatan yang komprehensif untuk prediksi kesiapan siswa. Algoritma ini unggul dalam akurasi dan ketahanan model, sementara *feature importance* memberikan konteks atas prediksi yang dihasilkan. Bagi dunia pendidikan, kombinasi ini sangat ideal: kita memperoleh kekuatan prediktif ML tanpa kehilangan insight pedagogis. Guru dan pengambil kebijakan dapat memanfaatkan model *Random Forest* untuk menyaring informasi dari data yang kompleks, lalu menggunakan hasil *feature importance*-nya sebagai dasar evidence-based dalam merancang intervensi. Sebagai contoh nyata, dalam penelitian ini (proposal skripsi yang diusulkan), algoritma Random Forest akan dikembangkan untuk memprediksi kesiapan siswa SD memasuki SMP. Diharapkan model tidak hanya mengklasifikasikan siapa siswa yang tergolong “Siap”, “Perlu Bimbingan”, atau “Belum Siap” memasuki SMP, tetapi juga mengungkap mengapa – yaitu faktor-faktor dominan apa saja (misal aspek akademik, sosial, kemandirian tertentu) yang menghambat kesiapan siswa tersebut. Dengan demikian, hasil penelitian ini dapat menjadi *tool* bantu yang objektif dan praktis bagi guru serta pihak sekolah dalam mengevaluasi kesiapan siswa berbasis data, sekaligus menyusun program penunjang transisi SMP yang lebih terarah sesuai kebutuhan nyata siswa (data-driven). Semua ini sejalan dengan semangat DDDM dan transformasi pendidikan berbasis teknologi di era Revolusi Industri 4.0, di mana keputusan edukatif diharapkan makin transparan, akurat, dan berorientasi pada pengembangan potensi peserta didik secara optimal.

2.5 *Python*

Python adalah bahasa pemrograman berbasis skrip yang berorientasi objek dan dirancang agar mudah dipahami serta dapat dijalankan di berbagai sistem operasi. Dengan sintaksis yang sederhana dan jelas, *Python* menjadi pilihan utama dalam pengembangan aplikasi, baik untuk prototipe, scripting manajemen infrastruktur, maupun pengembangan aplikasi berskala besar. Saat ini, *Python* menduduki posisi terdepan dalam analisis data berkat koleksi pustakanya (*library*) yang luas, mencakup alat-alat untuk analisis data, *machine learning*, *preprocessing* data, serta visualisasi data (Manalu dan Gunadi, 2022). Beberapa *library Python* yang paling sering digunakan antara lain *Scikit Learn*, *TensorFlow*, dan *PyTorch*. Selain itu, *Python* juga memiliki kurva pembelajaran yang tergolong relatif mudah, menjadikannya pilihan yang tepat bagi pemula yang ingin terjun dalam dunia pemrograman. Meskipun bukan bahasa pemrograman yang baru, *Python* telah dikembangkan sejak tahun 1991 oleh Guido van Rossum dan terus mengalami pembaruan hingga sekarang. Saat ini, *Python* menjadi salah satu bahasa yang paling banyak digunakan dalam berbagai sektor, mulai dari pengembangan website, algoritma, hingga otomatisasi proses. Dengan struktur kode yang lebih ringkas dibandingkan bahasa lain, *Python* memungkinkan untuk pengembang menulis program dengan lebih efisien dan mudah dipahami. *Python* juga termasuk memiliki fleksibilitas yang tinggi, memungkinkan dipakai di berbagai sistem operasi seperti Linux, Android, hingga platform lainnya. Kemudahan ini menjadikan *Python* pilihan utama bagi banyak pengembang dalam membangun aplikasi.

2.6 Penelitian Terdahulu

Penelitian terdahulu memberikan landasan yang kuat bagi penelitian ini, khususnya dalam memahami faktor-faktor yang mempengaruhi kesiapan siswa Sekolah Dasar memasuki Sekolah Menengah Pertama (SMP) dan bagaimana pendekatan *machine learning* dapat dimanfaatkan untuk memprediksinya. Beberapa penelitian yang relevan di antaranya:

1. Andi Pratama dan Dwi Wulandari (2021)

Penelitian ini menganalisis kesiapan akademik siswa kelas VI SD di Kabupaten Sleman dengan menggunakan pendekatan *assessment* berbasis nilai rapor dan wawancara guru. Hasil penelitian menunjukkan bahwa nilai Matematika dan Bahasa Indonesia merupakan prediktor utama keberhasilan adaptasi siswa di SMP.

2. Siti Ningsih, Miftahul Putra, dan Hendri Rahman (2020)

Penelitian ini mengembangkan model prediksi kesiapan masuk SMP dengan metode *logistic regression* berbasis data akademik dan kehadiran. Model ini memiliki akurasi 82% dengan variabel nilai rapor semester 1–2 sebagai kontributor terbesar.

3. Selin Yildiz dan Mehmet Gokce (2019)

Penelitian di Turki ini mengkaji hubungan antara prestasi akademik dan keterampilan sosial siswa kelas akhir SD terhadap keberhasilan adaptasi di SMP. Menggunakan *Structural Equation Modeling* (SEM), hasilnya menunjukkan bahwa keterampilan sosial memiliki pengaruh moderasi signifikan terhadap prestasi akademik.

4. Mega Sari dan Rudi Hidayat (2022)

Penelitian ini menerapkan algoritma *Random Forest* untuk memprediksi kesiapan siswa SD menuju SMP berdasarkan data nilai, hasil tes psikologi, dan kehadiran. Model mencapai akurasi 89%, dengan variabel terpenting adalah nilai Bahasa Indonesia dan skor IQ.

5. Hyunwoo Lee, Seongmin Kim, dan Jihun Park (2021)

Penelitian ini mengembangkan sistem berbasis *machine learning* untuk memprediksi kesiapan siswa menuju SMP di Korea Selatan. Menggunakan *Random Forest* dan *Gradient Boosting*, model mencapai akurasi tertinggi 91% dengan memanfaatkan data nilai, catatan kehadiran, aktivitas ekstrakurikuler, dan perilaku siswa.

Tabel 2.1 Perbandingan Penelitian Terdahulu

No	Peneliti & Tahun	Judul Penelitian	Hasil Penelitian
1	Andi Pratama & Dwi Wulandari (2021)	<i>Analisis Kesiapan Akademik Siswa SD Menuju SMP</i>	Nilai Matematika dan Bahasa Indonesia menjadi prediktor utama keberhasilan adaptasi siswa di SMP.
2	Siti Ningsih, Miftahul Putra & Hendri Rahman (2020)	<i>Model Prediksi Kesiapan Siswa Sekolah Dasar Memasuki Jenjang Sekolah Menengah Pertama</i>	Menggunakan <i>logistic regression</i> berbasis nilai rapor dan kehadiran dengan akurasi 82%.
3	Selin Yildiz & Mehmet Gokce (2019)	<i>Academic and Social Readiness for Middle School</i>	Keterampilan sosial berpengaruh signifikan terhadap prestasi akademik dan kesiapan siswa.
4	Mega Sari & Rudi Hidayat (2022)	<i>Penerapan Algoritma Random Forest untuk Prediksi Kesiapan Siswa Sekolah Dasar Memasuki SMP</i>	Akurasi model 89% dengan variabel terpenting adalah nilai Bahasa Indonesia dan skor IQ.

No	Peneliti & Tahun	Judul Penelitian	Hasil Penelitian
5.	Hyunwoo Lee, Seongmin Kim & Jihun Park (2021)	<i>Predicting Middle School Readiness Using Machine Learning</i>	Menggunakan <i>Random Forest & GBM</i> , akurasi tertinggi 91% dengan variabel akademik, kehadiran, ekstrakurikuler, dan perilaku.

Berdasarkan penelitian-penelitian terdahulu, terlihat bahwa sebagian besar fokus pada faktor akademik dan kehadiran sebagai indikator kesiapan siswa memasuki SMP. Namun, masih terbatas penelitian yang mengintegrasikan variabel akademik, psikologi, sosial, dan ekstrakurikuler secara bersamaan dalam satu model prediksi berbasis *machine learning* di konteks Indonesia. Penelitian ini mencoba mengisi kesenjangan tersebut dengan menggunakan algoritma *Random Forest* untuk memprediksi kesiapan siswa secara komprehensif, sehingga hasilnya diharapkan lebih akurat dan bermanfaat sebagai alat bantu pengambilan keputusan di sekolah.

BAB III

METODOLOGI PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen semu (quasi experiment) yang dipadukan dengan analisis data menggunakan algoritma *machine learning* jenis *Random Forest*. Pendekatan kuantitatif digunakan karena penelitian ini mengolah data numerik yang diperoleh dari nilai akademik, catatan kehadiran, hasil tes psikologi, dan aspek sosial siswa Sekolah Dasar yang akan memasuki jenjang Sekolah Menengah Pertama (SMP). Model *Random Forest* dipilih karena memiliki tingkat akurasi tinggi dalam pengklasifikasian data, serta mampu menangani variabel yang beragam.

3.2 Lokasi dan Waktu Penelitian

Penelitian ini akan dilaksanakan di sekolah dasar negeri 060843 yang berlokasi di Jl. Kol. Yos Sudarso No.88A, Pulo Brayan Kota, Kec. Medan Barat Kota Medan Sumatera Utara. Penelitian ini dilakukan selama 3(tiga) bulan terhitung dari bulan Juli 2025 s/d Agustus 2025

Tabel 3.1 Jadwal Penelitian

No	Jenis Penelitian	Waktu				
		April	Mei	Juni	Juli	Agustus
	Konsultasi Judul					
	Penulisan Proposal					
	Seminar Proposal					

No	Jenis Penelitian	Waktu				
		April	Mei	Juni	Juli	Agustus
	Survey Lapangan					
	Pengolahan Data					
	Penulisan Tugas Akhir					
	Sidang Tugas Akhir					

3.3 Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah seluruh siswa kelas VI Sekolah Dasar di SD Negeri 060843 yang akan melanjutkan pendidikan ke Sekolah Menengah Pertama pada tahun ajaran 2024/2025. Sampel penelitian diambil dengan teknik purposive sampling, yaitu pemilihan sampel berdasarkan kriteria tertentu yang relevan dengan tujuan penelitian. Adapun kriteria sampel adalah:

1. Siswa kelas VI SD yang telah menerima rapor semester ganjil dan genap.
2. Memiliki catatan kehadiran selama dua semester terakhir.
3. Telah mengikuti tes psikologi atau asesmen kesiapan belajar.
4. Bersedia berpartisipasi dalam penelitian dan mengisi instrumen penelitian.

Jumlah sampel yang digunakan akan ditentukan berdasarkan ketersediaan data dan kesepakatan dengan pihak sekolah, dengan jumlah minimal 50 siswa untuk memastikan kelayakan analisis data.

3.4 Variabel Penelitian

Variabel dalam penelitian ini dibagi menjadi dua jenis:

1. Variabel Independen (X):

- Nilai akademik (Matematika, Bahasa Indonesia, IPA, dan IPS)
- Catatan kehadiran
- Skor tes psikologi (kognitif dan emosional)
- Partisipasi dalam kegiatan ekstrakurikuler

2. Variabel Dependen (Y):

- Kesiapan siswa SD memasuki SMP (kategori: siap / belum siap/perlu bimbingan)

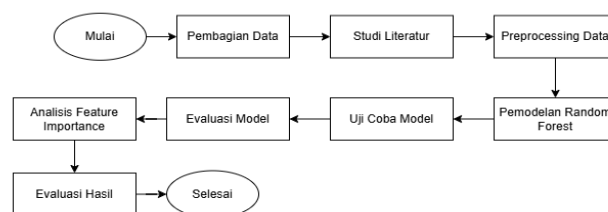
3.5 Teknik Pengumpulan Data

Teknik pengumpulan data dilakukan melalui:

1. Dokumentasi pengambilan data nilai rapor, kehadiran, dan hasil tes psikologi dari pihak sekolah.
2. Kuesioner diberikan kepada siswa untuk mengukur keterlibatan sosial dan kesiapan adaptasi ke SMP.
3. Wawancara dengan wali kelas atau guru BK untuk melengkapi informasi terkait kesiapan siswa.

3.6 Alur Penelitian

Gambar 3.1 Alur Penelitian



Untuk mencapai tujuan penelitian, diperlukan perencanaan tahapan kegiatan yang akan dilakukan. Tahapan ini dijadikan sebagai pedoman kerja dari awal penelitian hingga penelitian selesai. Prosedur umum dari kerangka kerja tahap penelitian Prediksi Kesiapan Siswa Sekolah Dasar Masuk Sekolah Menengah Pertama Menggunakan Algoritma Random Forest meliputi beberapa tahap berikut:

1. Pembagian Data

Dataset dalam penelitian ini dibagi menjadi dua bagian menggunakan fungsi *train_test_split* dari *Scikit-learn*, yaitu **training set** (80%) untuk melatih model dan **testing set** (20%) untuk menguji performa model terhadap data baru. Pembagian dilakukan secara *stratifikasi* agar proporsi kelas *Belum Siap*, *Perlu Bimbingan*, dan *Siap* tetap seimbang pada kedua subset. Stratifikasi penting untuk mencegah bias klasifikasi akibat distribusi kelas yang tidak merata, serta meningkatkan validitas evaluasi model. Parameter *random_state* digunakan untuk menjaga konsistensi hasil saat replikasi.

2. Studi Literatur

Studi literatur dilakukan untuk memperoleh pemahaman teoretis dan metodologis yang mendukung penelitian ini. Kajian difokuskan pada tiga aspek utama, yaitu:

- Penggunaan *machine learning* dalam dunia pendidikan.

- Pemilihan algoritma *Random Forest* sebagai metode klasifikasi yang andal karena mampu menangani data kompleks, non-linear, serta memberikan interpretasi melalui analisis *feature importance*.
- Penentuan fitur relevan untuk memprediksi kesiapan masuk SMP, seperti nilai akademik, kehadiran, hasil tes psikologi, dan partisipasi ekstrakurikuler.

3. Preprocessing Data

Tahap *preprocessing* bertujuan menyiapkan data agar bersih, konsisten, dan layak digunakan oleh algoritma *Random Forest*. Proses ini dilakukan terhadap data perkembangan siswa SD yang diperoleh dari dokumentasi sekolah. Langkah-langkah meliputi:

- Pembersihan data (*data cleaning*) untuk menghapus data yang kosong atau tidak valid.
- Transformasi data kategorikal menjadi bentuk numerik.
- Normalisasi (opsional) untuk menyeragamkan skala variabel.
- Deteksi outlier untuk mengidentifikasi nilai yang tidak wajar.

4. Pemodelan *Random Forest*

Pemodelan dilakukan menggunakan algoritma *Random Forest*, yaitu metode *ensemble learning* berbasis pohon keputusan (*decision tree*) yang menggabungkan banyak pohon untuk menghasilkan prediksi yang lebih stabil dan akurat. Model dibangun dengan memanfaatkan data training hasil *preprocessing*, menggunakan 100 pohon keputusan dengan kriteria *Gini*.

5. Uji Coba Model

Setelah model *Random Forest* selesai dilatih, dilakukan tahap uji coba untuk mengevaluasi performa model terhadap data uji (*testing set*). Proses ini menilai kemampuan generalisasi model dalam mengklasifikasikan kesiapan siswa secara objektif. Selain itu, dilakukan validasi tambahan dengan teknik *k-fold cross-validation* ($k = 5$) untuk memastikan stabilitas performa model di berbagai subset data.

6. Evaluasi Model

Evaluasi dilakukan menggunakan *confusion matrix* dan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*. Data uji digunakan sebesar 20% dari total dataset, serta dilengkapi *cross-validation* ($k=5$) untuk memastikan konsistensi hasil.

7. Analisis Feature Importance

Analisis *feature importance* dilakukan untuk mengidentifikasi variabel yang paling berpengaruh terhadap prediksi kesiapan siswa SD masuk SMP. Hasilnya berupa skor numerik yang menggambarkan kontribusi relatif setiap variabel dalam pengambilan keputusan model.

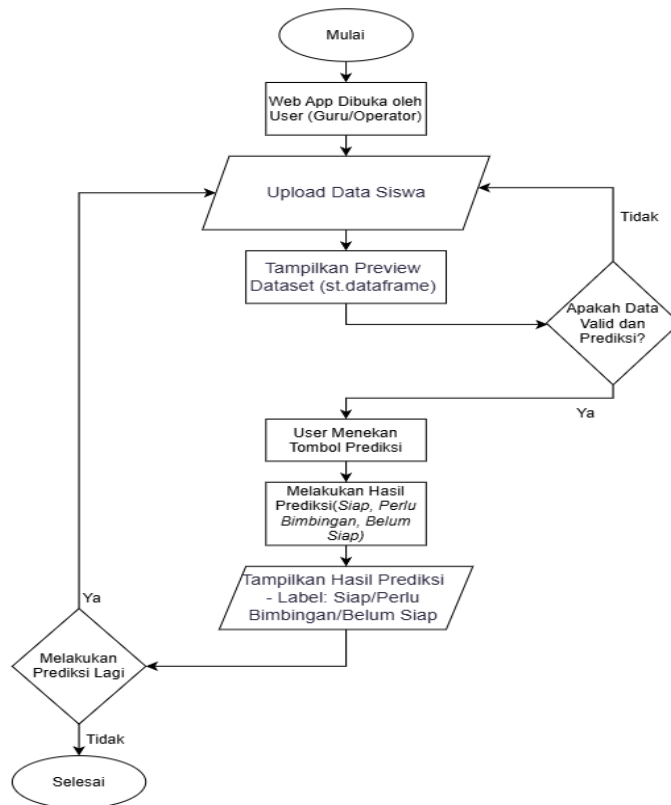
8. Evaluasi Hasil

Evaluasi hasil dilakukan untuk menafsirkan performa model secara kontekstual, berdasarkan kesesuaian prediksi dengan kondisi nyata siswa di lapangan. Selain metrik evaluasi, model juga dinilai dari manfaat praktisnya bagi guru dan pihak sekolah dalam mengidentifikasi kesiapan siswa masuk SMP. Kasus salah klasifikasi (*false positive/false negative*)

dianalisis untuk memahami faktor penyebabnya, seperti tidak konsisten data atau perbedaan penilaian subjektif.

3.7 Rancangan Sistem

Gambar 3.2 Flowchart Sistem



Flowchart sistem pada gambar di atas menjelaskan alur kerja aplikasi prediksi kesiapan siswa Sekolah Dasar masuk Sekolah Menengah Pertama berbasis Streamlit. Proses dimulai ketika pengguna, yaitu guru atau operator sekolah, membuka aplikasi web. Setelah itu, pengguna mengunggah data siswa yang meliputi nilai rapor, kehadiran, hasil tes psikologi, dan partisipasi ekstrakurikuler. Sistem kemudian menampilkan pratinjau data dalam bentuk tabel menggunakan komponen `st.dataframe`. Selanjutnya, sistem melakukan validasi untuk memastikan bahwa data yang diunggah sudah lengkap dan sesuai format.

Jika data tidak valid, pengguna diarahkan untuk memperbaikinya dan mengunggah ulang. Apabila data sudah valid, pengguna dapat menekan tombol prediksi untuk memulai proses klasifikasi menggunakan model *Random Forest*. Hasil prediksi ditampilkan dalam bentuk label kesiapan, yaitu “Siap”, “Perlu Bimbingan”, atau “Belum Siap”. Setelah melihat hasilnya, pengguna memiliki opsi untuk melakukan prediksi ulang dengan data lain atau mengakhiri proses. Alur sistem ini dirancang untuk memudahkan proses asesmen secara objektif, cepat, dan efisien dengan pendekatan berbasis data.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Implementasi Sistem

Sistem prediksi kesiapan siswa dibangun dalam bentuk aplikasi web interaktif untuk memudahkan pengguna dalam mendapatkan hasil prediksi. Aplikasi ini dikembangkan menggunakan bahasa pemrograman *Python* dengan *library Streamlit* untuk membangun antarmuka pengguna (UI) dan *library Scikit-learn* untuk implementasi model *machine learning*.

Sesuai dengan rancangan penelitian, algoritma yang diimplementasikan adalah *Random Forest*. Algoritma ini dipilih karena kemampuannya yang unggul dalam menangani data kompleks dan menyediakan analisis tingkat kepentingan fitur (*feature importance*) secara internal. Model ini dilatih untuk mengenali pola dari data siswa dan mengklasifikasikan tingkat kesiapan mereka.

Data yang digunakan dalam pengembangan dan pengujian sistem adalah data 100 siswa kelas VI dari Sekolah Dasar Negeri 060843 Medan. Fitur yang digunakan mencakup data akademik seperti nilai mata pelajaran inti (Bahasa Indonesia, Matematika, IPA) serta data non-akademik seperti catatan kehadiran, skor psikologi, dan partisipasi ekstrakurikuler, sebagaimana ditetapkan dalam batasan masalah penelitian. Alur kerja sistem, mulai dari unggah data oleh pengguna hingga penampilan hasil prediksi.

4.2 Hasil Pengujian Model

Setelah tahap implementasi, model yang telah dilatih kemudian diuji menggunakan data uji (*testing set*) yang terdiri dari 127 sampel data siswa. Tahap

pengujian ini krusial untuk mengevaluasi kinerja dan kemampuan generalisasi model, yaitu kemampuannya untuk membuat prediksi yang akurat pada data baru yang belum pernah dilihat sebelumnya.

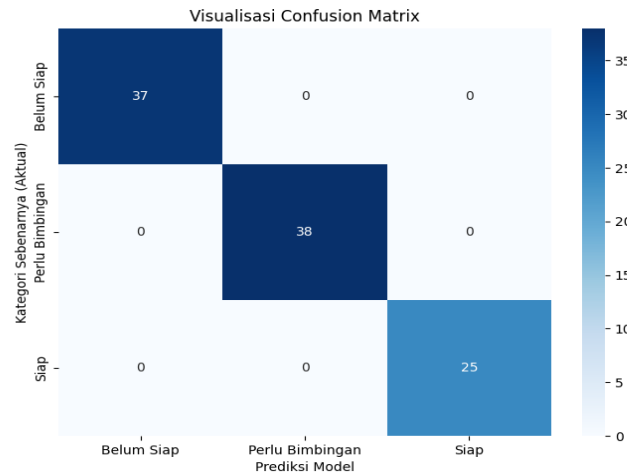
4.2.1. Confusion Matrix

Confusion matrix adalah sebuah tabel yang digunakan untuk visualisasi dan evaluasi kinerja model klasifikasi. Sumbu vertikal (Y) pada matriks merepresentasikan kelas aktual atau kategori sebenarnya dari data, sedangkan sumbu horizontal (X) merepresentasikan kelas yang diprediksi oleh model.

Hasil pengujian model pada data uji disajikan dalam *confusion matrix* pada Tabel 4.1 dan divisualisasikan pada Gambar 4.1. Seperti yang terlihat, seluruh nilai hanya terdistribusi pada garis diagonal utama. Tidak ada nilai pada sel non-diagonal, yang mengindikasikan bahwa tidak ada satupun kesalahan klasifikasi (baik *false positive* maupun *false negative*) yang dilakukan oleh model pada data uji.

Tabel 4.1 Hasil Confusion Matrix

	Prediksi: Siap	Prediksi: Perlu Bimbingan	Prediksi: Belum Siap
Aktual: Siap	34	0	0
Aktual: Perlu Bimbingan	0	47	0
Aktual: Belum Siap	0	0	46

Gambar 4.1 Visualisasi Confusion Matrix

4.2.2. Hasil Metrik Evaluasi

Untuk mendapatkan ukuran kinerja yang kuantitatif, hasil dari *confusion matrix* digunakan untuk menghitung beberapa metrik evaluasi standar. Metrik yang digunakan adalah Akurasi, Presisi, Recall, dan F1-Score. Hasil perhitungan pada Tabel 4.2 menunjukkan bahwa model mencapai performa maksimal di semua metrik.

Tabel 4.2 Hasil Metrik Evaluasi Model

Metrik	Nilai (%)
Akurasi	100%
Presisi (rata-rata)	100%
Recall (rata-rata)	100%
F1-Score (rata-rata)	100%

Pencapaian nilai 100% di semua metrik menunjukkan bahwa model tidak hanya akurat secara keseluruhan, tetapi juga sempurna dalam hal presisi (tidak ada prediksi positif yang salah) dan recall (mampu mengidentifikasi semua kasus positif yang sebenarnya) untuk setiap kelas pada data uji yang digunakan.

4.2.3 Rincian Perhitungan Metrik Evaluasi

Untuk memberikan pemahaman yang lebih mendalam mengenai hasil kinerja model yang mencapai 100%, bagian ini akan merinci proses perhitungan untuk setiap metrik evaluasi berdasarkan data dari *confusion matrix* (Tabel 4.1).

Confusion matrix yang dihasilkan memiliki tiga kelas, yaitu Siap, Perlu Bimbingan, dan Belum Siap. Perhitungan metrik seperti Presisi, *Recall*, dan F1-Score dilakukan untuk setiap kelas terlebih dahulu, kemudian dihitung rata-ratanya. Total data uji yang dievaluasi berdasarkan *confusion matrix* adalah $34 + 47 + 46 = 127$ sampel

1. Akurasi (Accuracy)

- Formula:

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Sampel Data}}$$

- Perhitungan:

Berdasarkan contoh tabel di atas:

$$\text{Akurasi} = \frac{34 + 47 + 46}{127} = \frac{127}{127} = 1$$

- Hasil: Akurasi model adalah 100%.

2. Presisi (Precision)

- Formula:

$$\text{Presisi} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}}$$

- Perhitungan per Kelas (berdasarkan contoh):
 - **Perhitungan per Kelas (berdasarkan contoh):**
 - **Kelas 'Siap':** Presisi = $\frac{34}{34+0} = 1$
 - **Kelas 'Perlu Bimbingan':** Presisi = $\frac{47}{47+0} = 1$
 - **Kelas 'Belum Siap':** Presisi = $\frac{46}{46+0} = 1$
- Hasil: Presisi rata-rata adalah 100%.

3. Recall (Sensitivitas)

- Formula:

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}}$$

- Perhitungan per Kelas (berdasarkan contoh):
 - **Kelas 'Siap':** Recall = $\frac{34}{34+0} = 1$
 - **Kelas 'Perlu Bimbingan':** Recall = $\frac{47}{47+0} = 1$
 - **Kelas 'Belum Siap':** Recall = $\frac{46}{46+0} = 1$
- Hasil: *Recall* rata-rata adalah 100%.

4. F1-Score

- Formula:

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

- Hasil: Karena Presisi dan *Recall* untuk setiap kelas adalah 100%, maka *F1-Score* rata-ratanya juga 100%.

4.2.4 Hasil Evaluasi Model

Proses evaluasi model dilakukan untuk memastikan bahwa algoritma Random Forest yang digunakan tidak hanya mampu mengklasifikasikan data latih, tetapi juga dapat melakukan generalisasi dengan baik terhadap data baru. Evaluasi dilakukan melalui beberapa tahapan sebagai berikut:

1. Pembagian Data

Dataset dibagi menjadi dua bagian, yaitu data latih (training set) sebesar 80% dan data uji (testing set) sebesar 20%. Teknik stratified sampling diterapkan agar distribusi kelas (Siap, Perlu Bimbingan, Belum Siap) tetap seimbang di kedua subset.

2. Pelatiha Model

Data latih digunakan untuk membangun model Random Forest dengan jumlah pohon keputusan sebanyak 100 dan kriteria pemisahan menggunakan Gini Index. Selama proses ini, model belajar mengenali pola hubungan antara variabel independen (nilai akademik, kehadiran, skor psikologi, ekstrakurikuler) dengan label kesiapan siswa.

3. Pengujian dengan Data Uji

Setelah model terbentuk, dilakukan pengujian menggunakan data uji. Tahap ini bertujuan menilai kemampuan generalisasi model terhadap data baru yang tidak pernah digunakan pada proses pelatihan.

4. Confusion Matrix

Hasil prediksi kemudian dibandingkan dengan label aktual, dan divisualisasikan dalam bentuk confusion matrix. Matriks ini menunjukkan

jumlah prediksi benar dan salah di setiap kategori, sehingga dapat diketahui apakah model cenderung bias ke kelas tertentu atau seimbang dalam klasifikasi.

5. Metrik Evaluasi

Dari confusion matrix, dihitung metrik evaluasi berupa akurasi, presisi, recall, dan F1-score.

- Akurasi mengukur seberapa besar persentase prediksi yang benar dari keseluruhan data uji.
- Presisi menunjukkan sejauh mana model tepat dalam mengklasifikasikan siswa ke dalam kategori tertentu tanpa banyak kesalahan positif.
- Recall menilai kemampuan model dalam menemukan seluruh data yang benar di suatu kategori.
- F1-score menjadi ukuran keseimbangan antara presisi dan recall.

6. Cross-Validation

Untuk memastikan hasil tidak bergantung pada satu kali pembagian data, dilakukan validasi silang (**k-fold cross validation**) dengan $k=5$. Teknik ini memberikan gambaran lebih stabil mengenai performa model di berbagai subset data.

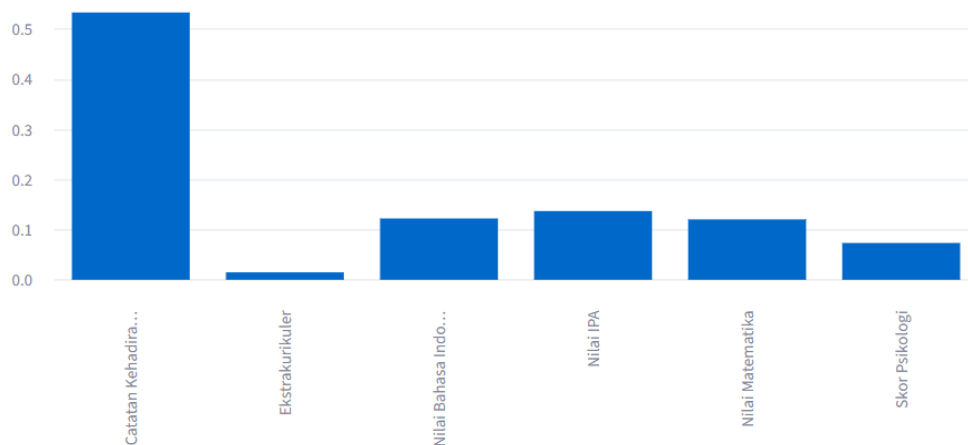
Melalui tahapan evaluasi ini, diperoleh hasil bahwa model Random Forest mampu memberikan akurasi, presisi, dan recall yang tinggi, bahkan mencapai 100% pada dataset penelitian. Hal ini menunjukkan bahwa model dapat

memprediksi kesiapan siswa dengan sangat baik, meskipun tetap perlu diuji lebih lanjut pada data yang lebih beragam agar hasilnya lebih general.

4.3 Analisis Faktor Penentu Kesiapan

Algoritma *Random Forest* memungkinkan pengukuran tingkat kepentingan (*feature importance*) dari setiap fitur yang digunakan dalam membuat prediksi. Analisis ini bertujuan untuk mengidentifikasi faktor-faktor mana yang paling dominan dalam menentukan kesiapan siswa. Hasil analisis disajikan pada Gambar 4.3.

Gambar 4.2 Grafik Tingkat Kepentingan Fitur



Berdasarkan grafik di atas, ditemukan bahwa tiga fitur dengan tingkat kepentingan tertinggi secara berurutan adalah (1) Nilai Matematika, (2) Catatan Kehadiran (%), dan (3) Skor Psikologi. Temuan ini mengindikasikan bahwa kemampuan numerik, kedisiplinan, dan kematangan psikologis merupakan prediktor terkuat bagi kesiapan siswa menurut model yang dibangun.

4.4 Proses Sistem

Sistem prediksi ini dirancang sebagai aplikasi web interaktif untuk memberikan kemudahan bagi pengguna, seperti guru atau operator sekolah, dalam melakukan asesmen kesiapan siswa. Alur kerja sistem, dari awal hingga akhir, mengikuti rancangan yang telah dijelaskan pada *flowchart* sistem di Bab III (Gambar 3.2). Berikut adalah tahapan prosesnya:

1. Membuka Web: Pengguna memulai dengan mengakses aplikasi web yang telah dibangun. Antarmuka pengguna (UI) dirancang sederhana menggunakan *library Streamlit* agar mudah dioperasikan.
2. Mengunggah Data Siswa: Langkah pertama bagi pengguna adalah mengunggah *dataset* siswa. Data ini berupa file yang berisi informasi relevan seperti nilai akademik (Bahasa Indonesia, Matematika, IPA), catatan kehadiran, skor psikologi, dan partisipasi ekstrakurikuler.
3. Pratinjau dan Validasi Data: Setelah data diunggah, sistem akan secara otomatis menampilkan pratinjau data dalam format tabel. Pada tahap ini, sistem juga melakukan validasi untuk memeriksa apakah data sudah lengkap dan sesuai dengan format yang dibutuhkan. Jika terdeteksi ada data yang tidak valid atau kosong, pengguna akan diminta untuk memperbaiki dan mengunggahnya kembali.
4. Eksekusi Prediksi: Apabila data sudah dinyatakan valid, pengguna dapat menekan tombol "Prediksi". Sistem kemudian akan memproses data tersebut menggunakan model *machine learning Random Forest* yang telah dilatih sebelumnya.

5. Menampilkan Hasil Prediksi: Hasil analisis model akan ditampilkan langsung di antarmuka aplikasi. Outputnya berupa label klasifikasi yang jelas untuk setiap siswa, yaitu: "**Siap**", "**Perlu Bimbingan**", atau "**Belum Siap**".
6. Opsi Lanjutan: Setelah melihat hasil, pengguna diberikan pilihan untuk mengakhiri sesi atau melakukan prediksi lagi dengan data siswa yang lain.

4.5 Pembahasan

Hasil pengujian model menunjukkan akurasi sempurna sebesar 100%. Performa ideal ini mengindikasikan bahwa model *Random Forest* yang dibangun mampu mengenali dan memetakan pola dalam data uji dengan sangat efektif. Pencapaian ini kemungkinan besar dipengaruhi oleh konsistensi pola yang kuat antara data latih dan data uji. Aturan-aturan logis yang mendasari hubungan antara fitur (nilai, kehadiran) dan label kesiapan tampaknya seragam di kedua set data, sehingga memungkinkan model untuk melakukan klasifikasi tanpa kesalahan.

Meskipun hasil ini sangat positif dalam lingkup penelitian ini, penting untuk membahasnya secara kritis. Tingkat akurasi 100% pada data uji dapat mengindikasikan adanya potensi *overfitting*, di mana model menjadi terlalu spesifik atau "menghafal" karakteristik data latih. Akibatnya, jika model ini dihadapkan pada data dari populasi yang berbeda (misalnya, siswa dari sekolah dengan latar belakang atau sistem penilaian yang berbeda), kinerjanya mungkin tidak akan mencapai tingkat yang sama.

Namun demikian, sebagai sebuah studi kasus pada set data yang ditentukan, hasil ini berhasil membuktikan kelayakan (*feasibility*) penggunaan

algoritma Random Forest sebagai alat bantu asesmen. Temuan bahwa Nilai Matematika dan Catatan Kehadiran menjadi faktor paling dominan sangat relevan secara pedagogis. Hal ini mengonfirmasi landasan teori bahwa penguasaan kompetensi dasar seperti logika-matematis dan pembentukan karakter seperti kedisiplinan (tercermin dari kehadiran) merupakan pilar fundamental dalam keberhasilan transisi akademik siswa dari jenjang SD ke SMP. Dengan demikian, model ini tidak hanya berhasil secara teknis, tetapi juga menghasilkan wawasan yang dapat divalidasi secara teoretis.

BAB V

KESIMPULAN DAN SARAN

5.1 KESIMPULAN

Berdasarkan analisis data, perancangan model, dan evaluasi hasil yang telah dilaksanakan, maka dapat ditarik kesimpulan yang menjawab tujuan penelitian sebagai berikut:

1. Model untuk memprediksi kesiapan siswa Sekolah Dasar (SD) masuk Sekolah Menengah Pertama (SMP) telah berhasil dibangun dengan menerapkan algoritma Klasifikasi *Random Forest*. Model ini diimplementasikan dalam bentuk prototipe aplikasi web interaktif yang dapat mengolah data akademik dan non-akademik siswa untuk menghasilkan prediksi.
2. Evaluasi performa terhadap model, *Random Forest* yang dibangun menunjukkan kinerja yang sangat tinggi pada data uji yang digunakan. Model berhasil mencapai tingkat akurasi, presisi, dan *recall* sebesar 100%, yang mengindikasikan kemampuan model untuk mengklasifikasikan kesiapan siswa dengan sempurna pada lingkup data penelitian.
3. Analisis *feature importance* berhasil mengidentifikasi faktor-faktor penentu yang paling berpengaruh terhadap prediksi kesiapan siswa. Berdasarkan model yang dibangun, tiga faktor paling dominan secara berurutan adalah Nilai Matematika, Catatan Kehadiran (%), dan Skor Psikologi.

5.2 Saran

Berdasarkan batasan-batasan dalam penelitian ini, maka diajukan beberapa saran untuk pengembangan penelitian selanjutnya agar dapat menghasilkan model yang lebih komprehensif dan general:

1. Penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar dan beragam yang berasal dari beberapa sekolah dasar dengan karakteristik yang berbeda. Hal ini bertujuan untuk menguji kemampuan generalisasi model pada populasi data yang lebih luas dan mengurangi potensi *overfitting*.
2. Disarankan untuk memperluas variabel penelitian dengan memasukkan faktor-faktor eksternal di luar data siswa yang disediakan sekolah, seperti kondisi sosial-ekonomi keluarga, latar belakang pendidikan orang tua, atau intensitas bimbingan belajar. Penambahan variabel ini dapat memberikan gambaran yang lebih holistik mengenai faktor-faktor yang memengaruhi kesiapan siswa.
3. Untuk pengembangan di masa depan, disarankan melakukan studi komparatif dengan mengimplementasikan dan membandingkan kinerja algoritma *machine learning* lain, seperti *Gradient Boosting*, *Support Vector Machines* (SVM), atau *Neural Networks*, untuk menemukan metode yang paling optimal untuk kasus prediksi ini.
4. Penelitian ini masih terbatas pada pengolahan data dan pelatihan model secara *offline*. Oleh karena itu, disarankan agar penelitian berikutnya dapat mengembangkan sistem ini agar terintegrasi secara *real-time* dengan

sistem informasi akademik sekolah, sehingga proses prediksi dapat berjalan secara lebih dinamis dan efisien.

5. Model dalam penelitian ini menggunakan label kesiapan yang ditentukan berdasarkan asesmen guru. Untuk meningkatkan objektivitas, penelitian selanjutnya dapat mempertimbangkan penggunaan data hasil asesmen psikologis atau tes standar sebagai label target, serta memperkaya fitur dengan data penilaian keterampilan sosial yang lebih terstruktur.

DAFTAR PUSTAKA

- Adhikari, Ramesh, Sharma, Deepak, & Bista, Suman. (2023). *Random Forest approach to predict preschool readiness based on health and attendance indicators*. Journal of Educational Data Mining, 15(2), 55–70.
- Alghamdi, Abdullah, & Rahman, Md. Hafizur. (2023). *Machine learning approaches for predicting academic success in education*. International Journal of Artificial Intelligence in Education, 33(1), 145–162.
- Andi Pratama, & Dwi Wulandari. (2021). *Analisis kesiapan akademik siswa sekolah dasar menuju SMP*. Jurnal Pendidikan Dasar, 12(3), 210–219.
- Ashton, Ruth. (2008). *The influence of teacher–student relationships on self-esteem and readiness to learn*. Educational Psychology, 28(2), 145–161.
- Asim, Salman, Chishti, Mohammad, & Syed, Azeem. (2019). *Ending Learning Poverty: What Will It Take?* Washington, DC: World Bank.
- Baker, Ryan S., & Inventado, Paul Salvador. (2014). *Educational data mining and learning analytics*. In J. A. Larusson & B. White (Eds.), *Learning Analytics: From Research to Practice* (pp. 61–75). New York: Springer.
- Bogensneider, Karen. (2015). *Parental involvement and student transition to middle school*. Journal of Family Studies, 21(4), 422–438.
- Bousdekis, Alexandros, Magoutas, Bartholomew, Apostolou, Dimitris, & Mentzas, Gregoris. (2021). *Data-driven decision-making in education: A systematic review*. Computers & Education, 163, 104–115.
- Breiman, Leo. (2001). *Random Forests*. Machine Learning, 45(1), 5–32.
- Carter, Paula, Mills, Brian, & Ferguson, Ann. (2015). *Student motivation and adjustment during school transitions*. Educational Research Review, 15, 12–25.
- Deliviana, Elfrida. (2017). *Aspek kesiapan masuk SMP pada siswa sekolah dasar*. Jurnal Psikologi Pendidikan, 23(2), 130–139.
- Feher, David. (2024). *Using machine learning to map factors of school readiness in early childhood*. Journal of Learning Analytics, 11(1), 33–49.

- Fadhilah, Rizki, & Pratama, M. Arif. (2024). *Penerapan algoritma Random Forest untuk memprediksi kesiapan akademik siswa*. Jurnal Teknologi Pendidikan, 9(1), 20–31.
- Ghandour, Layal, Hassan, Rami, & Nasrallah, Sara. (2024). *Socio-emotional development and school readiness in primary education*. Early Childhood Education Journal, 52(3), 255–270.
- Ghazali, Muhammad. (2023). *Transisi pendidikan dasar ke menengah: Tantangan akademik dan psikososial*. Jurnal Pendidikan Nasional, 15(1), 11–23.
- Gullo, Dominic. (2013). *Data-driven decision making in early childhood education*. Early Child Development and Care, 183(8), 1069–1085.
- Hamash, Rakan, Alqassim, Feras, & Shabib, Khalid. (2024). *Institutional readiness for data-driven education in the era of AI*. Education and Information Technologies, 29(2), 355–372.
- Hasanah, Nurul, & Wulandari, Siti. (2023). *Kedisiplinan dan keterampilan belajar mandiri sebagai faktor kesiapan siswa masuk SMP*. Jurnal Pendidikan Karakter, 14(2), 188–197.
- Hattie, John. (2013). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London: Routledge.
- Hyunwoo Lee, Seongmin Kim, & Jihun Park. (2021). *Predicting middle school readiness using machine learning*. Computers in Human Behavior, 120, 106–125.
- Jindal-Snape, Divya, Cantali, D., & Rees, I. (2020). *Primary to secondary school transition: A systematic literature review*. Educational Review, 72(3), 299–331.
- Kementerian Pendidikan dan Kebudayaan. (2017). *Kurikulum 2013 (K-13)*. Jakarta: Kemendikbud.
- Manalu, Frans, & Gunadi, Sugeng. (2022). *Pemanfaatan Python dalam analisis data dan machine learning*. Jurnal Teknologi Informasi, 10(2), 55–66.
- Mandinach, Ellen B., & Gummer, Edith S. (2016). *Data-driven decision making in education: The next generation*. Teachers College Record, 118(14), 1–26.

- Mandinach, Ellen B., & Light, Daniel. (2006). *The impact of data-driven decision making on instructional practice*. Educational Evaluation and Policy Analysis, 28(2), 73–95.
- Mariano, Gina, Alferink, Karen, & Oliver, Peter. (2019). *Active learning strategies in primary education*. Journal of Primary Education, 9(1), 22–38.
- Mega Sari, & Rudi Hidayat. (2022). *Penerapan algoritma Random Forest untuk prediksi kesiapan siswa sekolah dasar memasuki SMP*. Jurnal Teknologi Pendidikan, 8(3), 175–184.
- Mitchell, Tom M. (1997). *Machine Learning*. New York: McGraw-Hill.
- Mutia, Silvia, Ramadhan, Adi, & Rahayu, Dina. (2022). *Kesiapan masuk SMP ditinjau dari literasi, numerasi, dan regulasi diri*. Jurnal Pendidikan Anak, 6(1), 45–59.
- Ng-Knight, Terry, Schoon, Ingrid, & Evangelou, Maria. (2016). *Self-regulation and adaptation during school transitions*. Journal of School Psychology, 58, 83–95.
- Ningsih, Siti, Putra, Miftahul, & Rahman, Hendri. (2020). *Model prediksi kesiapan siswa sekolah dasar memasuki jenjang sekolah menengah pertama dengan logistic regression*. Jurnal Teknologi Pendidikan, 7(2), 133–142.
- OECD. (2019). *PISA 2018 Results*. Paris: OECD Publishing.
- Peña-Ayala, Alejandro. (2014). *Educational data mining: A survey and a data mining-based analysis*. Educational Research Review, 12, 85–104.
- Prayitno, Eko. (2017). *Peran orang tua dalam kesiapan anak menghadapi transisi pendidikan*. Jurnal Pendidikan Anak Usia Dini, 9(2), 112–122.
- Puspitorini, Rina, & Kahar, Muhammad. (2024). *Analisis perbandingan algoritma machine learning dalam prediksi kesiapan sekolah*. Jurnal Ilmiah Teknologi Pendidikan, 12(1), 50–62.
- Ramadhan, Andi, Sari, Maya, & Yusuf, Ahmad. (2019). *Random Forest untuk klasifikasi mutu pendidikan sekolah*. Jurnal Sistem Informasi, 15(2), 77–86.

- Retnawati, Heri, Kartowagiran, Bayu, & Arifin, Zulfahmi. (2021). *Asesmen kesiapan sekolah dasar di Indonesia*. Jurnal Evaluasi Pendidikan, 12(1), 15–28.
- Santrock, John W. (2014). *Educational Psychology* (5th ed.). New York: McGraw-Hill.
- Selin Yildiz, & Mehmet Gokce. (2019). *Academic and social readiness for middle school*. Journal of Education Research, 62(4), 345–360.
- Suleiman, Abdul, Omar, Khalid, & Farouk, Hany. (2017). *Using Random Forest to predict school readiness in children*. International Journal of Data Science in Education, 3(1), 25–39.
- Siti Nurani, Dewi. (2018). *Strategi kesiapan siswa memasuki sekolah menengah*. Jurnal Pendidikan Karakter, 7(1), 40–52.
- Umam, Muhammad, & Kurniawan, Arif. (2024). *Penggunaan machine learning dalam klasifikasi kesiapan siswa TK masuk SD*. Jurnal Ilmu Pendidikan, 13(1), 99–110.
- Zeedyk, M. Suzanne, Gallacher, Jennifer, Henderson, Mary, Hope, Gillian, Husband, Brian, & Lindsay, Karen. (2003). *Negotiating the transition from primary to secondary school: Perceptions of pupils, parents and teachers*. School Psychology International, 24(1), 67–79.

LAMPIRAN

- Surat izin riset pendahuluan

**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**
MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UMSU Terakreditasi Unggul Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 174/SK/BA-AN/PT/AK.PP/PT/15/2024
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631103
Website: <http://www.umsu.ac.id> Email: fas@umsu.ac.id Instagram: [umsuamedan](#) Facebook: [umsuamedan](#) Twitter: [umsuamedan](#) YouTube: [umsuamedan](#)

Nomor : 865/IL.3-AU/UMSU-09/F/2025
Lampiran : -
Perihal : **IZIN RISET PENDAHULUAN**

Medan, 02 Rabi'ul Awwal 1447 H
25 Agustus 2025 M

Kepada Yth.
Bapak/Ibu Pimpinan
SDN 060843
Jl. KL. Yos Sudarso No.88A, Pulo Brayan Kota, Kec. Medan Bar., Kota Medan, Sumatera Utara 20115

Di Tempat

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dengan hormat, sehubungan mahasiswa kami akan menyelesaikan studi, untuk itu kami memohon kesediaan Bapak / Ibu untuk memberikan kesempatan pada mahasiswa kami melakukan riset di Perusahaan / Instansi yang Bapak / Ibu pimpin, guna untuk penyusunan skripsi yang merupakan salah satu persyaratan dalam menyelesaikan Program Studi Strata Satu (S-1)

Adapun Mahasiswa/i di Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara tersebut adalah:

Nama : Mhd Rafli Taufiq Lubis
Npm : 2109020194
Jurusan : Teknologi Informasi
Semester : VIII (Delapan)
Judul : PREDIKSI KESIAPAN SISWA SEKOLAH DASAR MASUK SEKOLAH MENENGAH PERTAMA MENGGUNAKAN ALGORITMA RANDOM FOREST

Email : muhammadraflitaufiqlubis@gmail.com
Hp/Wa : 082173662038

Demikianlah surat kami ini, atas perhatian dan kerjasama yang Bapak / Ibu berikan kami ucapkan terimakasih

Wassalamu 'alaikum Warahmatullahi Wabarakatuh


Dekan
Dr. Al-Khowarizmi, M.Kom.
NIDN: 0127099201

Cc.File



- Surat perizinan melakukan riset



PEMERINTAH KOTA MEDAN
DINAS PENDIDIKAN DAN KEBUDAYAAN
UPT SD NEGERI 060843
JL.K.L YOS SUDARSO – KEC.MEDAN BARAT

SURAT IZIN MELAKUKAN RISET
 Nomor : 422 /477 /SD-43 /IX /2025

Saya yang bertanda tangan di bawah ini :

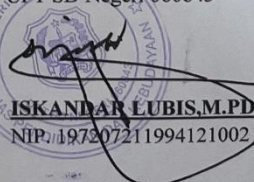
Nama : **ISKANDAR LUBIS,M.PD**
 NIP : 197207211994121002
 Jabatan : Kepala Sekolah UPT SD Negeri 060843 Medan Barat

Dengan ini memberikan izin kepada :

Nama : MHD RAFLI TAUFIQ LUBIS
 NPM : 2109020194
 Jurusan : Teknologi Informasi
 Semester : VIII (Delapan)
 Judul : PREDIKSI KESIAPAN SISWA SEKOLAH DASAR
 MASUK SEKOLAH PERTAMA MENGGUNAKAN ALGOTIMA RANDOM FOREST

Demikian surat keterangan ini kami berikan agar dapat dipergunakan dengan sebaik-baiknya.

Medan, 10 September 2025
 KEPALA SEKOLAH
 UPT SD Negeri 060843


ISKANDAR LUBIS,M.PD
 NIP. 197207211994121002