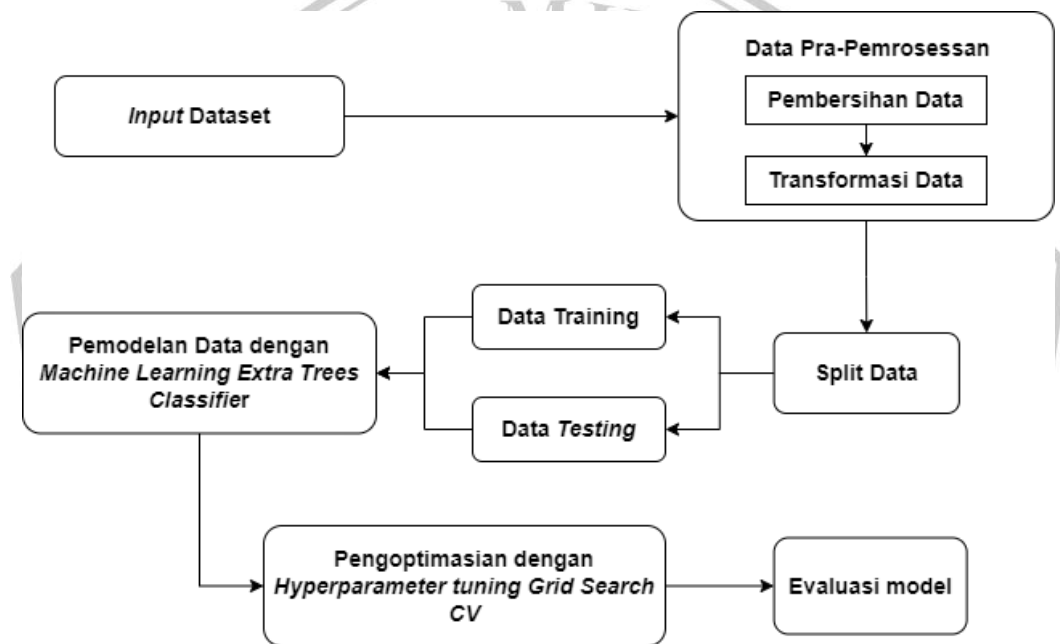


## BAB III

### METODE PENELITIAN

#### 3.1 Rancangan Penelitian

Penelitian ini mengimplementasikan metode *Extra Trees Classifier* dengan optimasi *Hyperparameter Tuning Grid Search CV* untuk mengklasifikasikan tingkat kemampuan adaptasi siswa dalam pendidikan online. Berikut ini adalah langkah-langkah dalam metode penelitian ini :



**Gambar 2.** Alur Penelitian Model

Tahapan penelitian pada Gambar 2 dimulai dari penginputan dataset hingga evaluasi model. Pertama-tama, dataset tersebut akan melalui pra-pemrosesan seperti pembersihan data dan transformasi data. Setelah melalui pra-pemrosesan, dataset tersebut akan di split menjadi dua yaitu data training dan data testing. Dataset tersebut akan dimodelkan dengan *Extra Trees Classifier*, dari model tersebut menghasilkan nilai akurasi. Nilai akurasi tersebut akan dioptimasi lagi dengan *Hyperparameter Tuning Grid Search CV* yang diharapkan dapat meningkatkan nilai akurasi yang lebih baik. Langkah terakhir yaitu evaluasi model yang digunakan pada penelitian.

### 3.2 Gathering Data

Penelitian ini menggunakan data tingkat adaptasi siswa terhadap pendidikan online. Dataset diperoleh dari Kaggle dan berasal dari survei yang dilakukan secara offline dan online pada salah satu institusi sekolah di Bangladesh, India(13). Dataset memiliki 1205 data yang terdiri dari 14 atribut, yaitu jenis kelamin, usia, tingkat pendidikan, jenis Institusi, siswa bidang IT, lokasi tinggal di kota, tingkat beban, kondisi keuangan, tipe internet, tipe konektivitas jaringan internet, durasi kelas, status institusi memiliki LMS, perangkat yang digunakan dan tingkat adaptasi. Tabel 1 menjelaskan rincian dari seluruh 14 atribut yang digunakan dan nilai kelas yang. Pemilihan dataset ini didasarkan pada kesesuaian atribut-atribut tersebut untuk digunakan sebagai input dalam proses klasifikasi pada penelitian ini. Dataset ini juga mencakup atribut-atribut yang berkaitan erat dengan tingkat adaptasi mahasiswa dalam pendidikan online.

**Tabel 1.** Atribut dan Nilai Dataset yang Digunakan

No	Atribut	Nilai
1.	Jenis kelamin	<i>Boy, Girl</i>
2.	Usia	<i>1-5, 6-10, 11-15, 16-20, 21-25, 26-30, 30+</i>
3.	Tingkat pendidikan	<i>School, University, College</i>
4.	Jenis institusi	<i>Non Government, Government</i>
5.	Siswa bidang IT	<i>Yes, No</i>
6.	Lokasi tinggal di kota	<i>Yes, No</i>
7.	Tingkat beban	<i>Low, High</i>
8.	Kondisi Keuangan	<i>Poor, Mid, Rich</i>
9.	Tipe jaringan internet	<i>2G, 3G, 4G</i>
10.	Tipe konektivitas jaringan internet	<i>Tab, Mobile, Computer</i>
11.	Perangkat yang digunakan saat di kelas	<i>Mobile Data, Wifi</i>
12.	Durasi kelas sehari- hari	<i>0, 1-3 Hours, 3-6 Hours</i>
13.	Status institusi memiliki LMS	<i>Yes, No</i>
14.	Tingkat adaptasi	<i>Low, Moderator, High</i>

### 3.3 Data Pra-Pemrosesan

Sebelum dataset diproses lebih lanjut, dilakukan langkah pra-pemrosesan data untuk memudahkan proses klasifikasi. Tahapan pra-pemrosesan data menjadi aspek penting dalam konteks pengembangan model *Machine Learning*, yang bertujuan untuk membersihkan data asli dan secara hasilnya meningkatkan akurasi serta efisiensi model secara keseluruhan [4]. Dalam penelitian ini, tahap pra-pemrosesan data mencakup dua aspek utama, yaitu pembersihan data dan transformasi data.

Nilai-nilai yang kurang penting atau tidak tersedia (null) disingkirkan selama proses pembersihan data [14]. Dalam dataset yang digunakan, langkah ini dilakukan terlebih dahulu dengan mengecek nilai-nilai yang termasuk dalam kategori *missing value* pada setiap atribut. Hasil dari langkah tersebut menunjukkan bahwa setiap atribut menunjukkan hasil *False*, yang mengindikasikan bahwa tidak ada nilai yang hilang. Hal ini diartikan sebagai tanda bahwa dataset memiliki kualitas yang baik.

Transformasi data bertujuan untuk mengalihkan skala pengukuran dari nilai awal ke bentuk nilai yang berbeda [10]. Dalam konteks dataset yang digunakan, semua atribut memiliki nilai kategorikal, sehingga langkah transformasi data diterapkan dengan menggunakan metode *Standard Scaler*. Pada awalnya, nilai dari data kategorikal diubah menjadi bentuk data numerik yang dinormalisasi. Informasi hasil dari proses transformasi ini dapat ditemukan dalam Tabel 2, yang menampilkan data lima teratas dari dataset setelah melalui tahap transformasi. Hal ini dilakukan untuk memberikan gambaran visual tentang perubahan dan penyesuaian nilai-nilai dalam dataset setelah proses transformasi, sesuai dengan metode yang diterapkan.

**Tabel 2.** Dataset yang telah ditransformasi

Jenis Kelamin	Usia	Tingkat Pendidikan	Jenis Institusi	Siswa Bidang IT	Lokasi Tinggal di Kota	Tingkat Beban	Kondisi Keuangan	Tipe Internet	Tipe Konektivitas	Durasi Kelas	Status LMS	Perangkat	Tingkat Adaptasi
0	3	2	1	0	1	1	0	1	2	2	0	2	2
1	3	2	1	0	1	0	0	0	2	1	1	1	2
1	2	0	0	0	1	1	0	1	2	1	0	1	2
1	1	1	1	0	1	1	0	0	2	1	0	1	2
1	2	1	1	0	1	1	1	0	1	0	0	1	1

Setelah melalui semua tahapan pengecekan, dataset disusun menjadi dua bagian dengan proporsi 8:2 untuk data *training* dan data *testing* [12]. Detail

pembagian ini tercatat secara terinci dalam Tabel 3. Langkah selanjutnya melibatkan identifikasi objek data, di mana masing-masing bagian terdiri dari data fitur (x) dan data target (y). Data fitur (x) mencakup seluruh atribut kecuali atribut tingkat adaptasi, sementara data target (y) hanya memuat atribut tingkat adaptasi. Proses ini bertujuan untuk memastikan bahwa dataset terstruktur secara optimal untuk digunakan dalam proses pelatihan dan pengujian model *Machine Learning*.

**Tabel 3.** Hasil Pembagian Data

Keterangan	Presentase (%)	Jumlah Data
Data <i>Training</i>	80	964
Data <i>Testing</i>	20	241
Total	100	1205

Data *training* merupakan subset dari dataset yang digunakan untuk melatih dan mengembangkan model-model *Machine Learning*. Selain itu, pada data training diterapkan cross-validation untuk meningkatkan validitas model [15]. Proses ini dilakukan untuk memastikan bahwa model *Machine Learning* yang dikembangkan mampu generalisasi dengan baik pada data yang tidak terlihat sebelumnya, sehingga hasilnya dapat diandalkan dan relevan ketika diterapkan pada situasi dunia nyata.

### 3.4 Modelling

Metode ini menggunakan *Decision Trees* yang tidak terkorelasi dalam sebuah hutan untuk menghasilkan hasil pemisahan data. Pembangunan *Decision Trees* dilakukan dalam Hutan Pohon Keputusan Ekstra, dihasilkan dari pelatihan awal dengan pemilihan fitur unggulan. Pemilihan fitur ini terjadi setelah setiap pengujian node, diikuti oleh pemilihan acak k-fitur untuk setiap pohon. Hal ini membantu formulasi *Decision Trees* baru yang tidak terkorelasi [16].

Struktur hutan yang sama digunakan untuk memastikan pemilihan fitur ketika hutan dibangun, di mana setiap fitur yang dipilih mengakibatkan penurunan keseluruhan dalam kriteria matematika yang digunakan dalam keputusan pemisahan fitur. Istilah "Kebermaknaan Gini Unggulan" digunakan untuk menggambarkan hal ini. Pengguna memilih k-fitur dengan peringkat tinggi

berdasarkan preferensinya untuk pemilihan fitur; semua fitur diatur dalam urutan menurun berdasarkan nilai GINI dari setiap fitur.

Klasifikasi *Extra Trees Classifier* menggunakan Gini impurity secara default, dan menggunakan Entropi sebagai alternatif untuk klasifikasi. Sementara itu, untuk regresi, *Mean Square Error* dan *Mean Absolute Error* digunakan:

$$\text{Gini Impurity} = \sum_{j=1}^O f_j (1 - f_j)$$

$$\text{Entropy} = \sum_{j=1}^O -f_j \log (f_j)$$

Dimana  $f_j$  adalah frekuensi label  $j$  pada suatu node dan  $O$  adalah jumlah label unik [17].

### 3.5 Model Optimization

Proses penyetelan hyperparameter dilakukan pada metode *Extra Trees Classifier*, dan informasi mengenai parameter yang dapat diubah variabelnya melalui teknik *Hyperparameter Tuning* menggunakan *Grid Search CV* dapat ditemukan dalam Tabel 4. Langkah ini diambil untuk memastikan bahwa seluruh proses optimasi dapat berjalan secara maksimal. Pelatihan atau pemasangan model ini memerlukan 10 kali estimasi waktu, menandakan bahwa penyesuaian parameter sedang dilakukan secara cermat untuk mencapai konfigurasi yang optimal. Proses ini menjadi penting agar model dapat mencapai performa terbaiknya dengan memodifikasi parameter-parameter yang krusial dalam algoritma *Extra Trees Classifier*.

**Tabel 4.** Variasi Parameter yang digunakan

<i>Hyperparameter</i>	Nilai
<i>n_estimators</i>	300
<i>Criterion</i>	<i>entropy</i>
<i>min_samples_split</i>	2,
	4,

	15,
<i>min_weight_fraction_leaf</i>	0.0,
	0.15,
	0.35,
	0.5,
<i>max_depth</i>	54

### 3.6 Model Evaluation

Setelah proses penelitian diselesaikan, model menjalani proses evaluasi untuk menentukan apakah melakukan pelatihan ulang dengan mengoptimalkan *Grid Search CV* pada penyesuaian *hyperparameter* dapat membawa peningkatan kinerja. Evaluasi model melibatkan prediksi hasil menggunakan kumpulan data uji atau validasi, dan hasil prediksi ini dibandingkan dengan hasil aktual yang sebenarnya. Untuk menilai kinerja model, berbagai metrik seperti akurasi, presisi, recall, perolehan, dan skor F1 dihitung. Evaluasi ini menjadi sangat penting dalam pemahaman sejauh mana model mampu melakukan generalisasi terhadap data baru yang belum terlihat sebelumnya. Hasil evaluasi ini memberikan panduan yang berharga untuk potensi penyempurnaan lebih lanjut dalam mengoptimalkan model *Machine Learning* yang dikembangkan.