

BAB III

METODELOGI PENELITIAN

3.1 Metode Pengumpulan Data

Adapun metode pengumpulan data yang digunakan dalam penelitian ini sebagai berikut.

a. Kuesioner

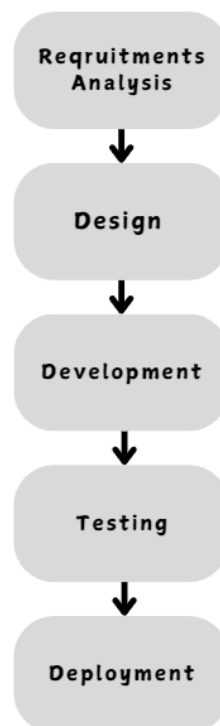
Pengumpulan data melalui kuesioner atau survei anonim dilakukan dengan membuat kuesioner online yang berisi pertanyaan terkait bullying dan pelecehan seksual, baik bagi siswa yang menjadi korban maupun saksi. Survei ini mencakup pertanyaan tentang jenis insiden yang dialami (fisik, verbal, online, atau seksual), lokasi kejadian (kelas, koridor, online, dll.), pelaku (siswa, guru, staf, dll.), waktu kejadian, dan dampak insiden terhadap korban. Data yang dikumpulkan melalui survei ini akan diolah menjadi dataset yang akan digunakan dalam program untuk memahami pola insiden serta melatih algoritma decision tree C4.5 dalam menganalisis dan mendeteksi perilaku *bullying dan pelecehan seksual*.

b. Wawancara

Selain pengumpulan data melalui kuesioner, peneliti juga akan melaksanakan wawancara dengan guru Bimbingan dan Konseling (BK). Wawancara ini berfokus pada pengumpulan informasi terkait jenis-jenis bullying dan pelecehan seksual yang sering terjadi, serta bagaimana sistem klasifikasi dapat diterapkan untuk mengidentifikasi setiap jenis bullying dan pelecehan seksual tersebut. Tujuan dari wawancara ini adalah untuk memahami karakteristik masing-masing jenis insiden, pola kejadian, dan faktor-faktor yang memengaruhinya. Informasi yang diperoleh akan digunakan untuk merancang sistem klasifikasi yang akurat dan relevan, sehingga sistem dapat membantu dalam mendeteksi, menganalisis, serta menangani kasus bullying dan pelecehan seksual dengan lebih terstruktur dan efektif.

3.2 Metode pengembangan perangkat lunak

Sistem pelaporan bullying dan pelecehan seksual berbasis website menggunakan algoritma decision tree C4.5 sebagai rekomendasi tindakan dikembangkan menggunakan metode Waterfall. Dalam penerapan metode ini, setiap langkah pengembangan dilakukan secara berurutan dan sistematis, dimulai dari tahap analisis kebutuhan, desain, implementasi, pengujian, hingga pemeliharaan. Metode ini memungkinkan setiap tahapan untuk diselesaikan sepenuhnya sebelum melanjutkan ke tahap berikutnya. Kelebihan menggunakan metode Waterfall adalah prosesnya yang terstruktur dan terdokumentasi dengan baik, sehingga memudahkan pengelolaan proyek serta memberikan panduan yang jelas dalam setiap langkah pengembangan. Alur metode Waterfall dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Alur Metode Waaterfall

3.2.1 *Reqrutments Analysis*

Pada tahapan ini penulis akan melakukan analisis kebutuhan untuk menunjang proses berjalannya penelitian ini. Adapun analisis kebutuhan yang dibutuhkan adalah sebagai berikut.

1. Analisis Kebutuhan Perangkat Lunak

Perangkat lunak atau *software* yang digunakan untuk membuat serta menjalankan program pada penelitian ini adalah sebagai berikut.

- a) Windows 11 64-bit
- b) HTML & CSS
- c) PHP
- d) Laravel
- e) MySQL
- f) Visual Studio Code
- g) Rapid Miner

2. Analisis Kebutuhan Perangkat Keras

Perangkat keras atau *hardware* yang digunakan untuk membuat serta menjalankan program pada penelitian ini adalah sebagai berikut.

- a) Ryzen 5 3550H 3.7Ghz
- b) Nvidia Gtx 1650
- c) 16 GB RAM DDR4
- d) 500 GB SSD
- e) 500 GB HDD
- f) Mouse

3.2.2 *Design*

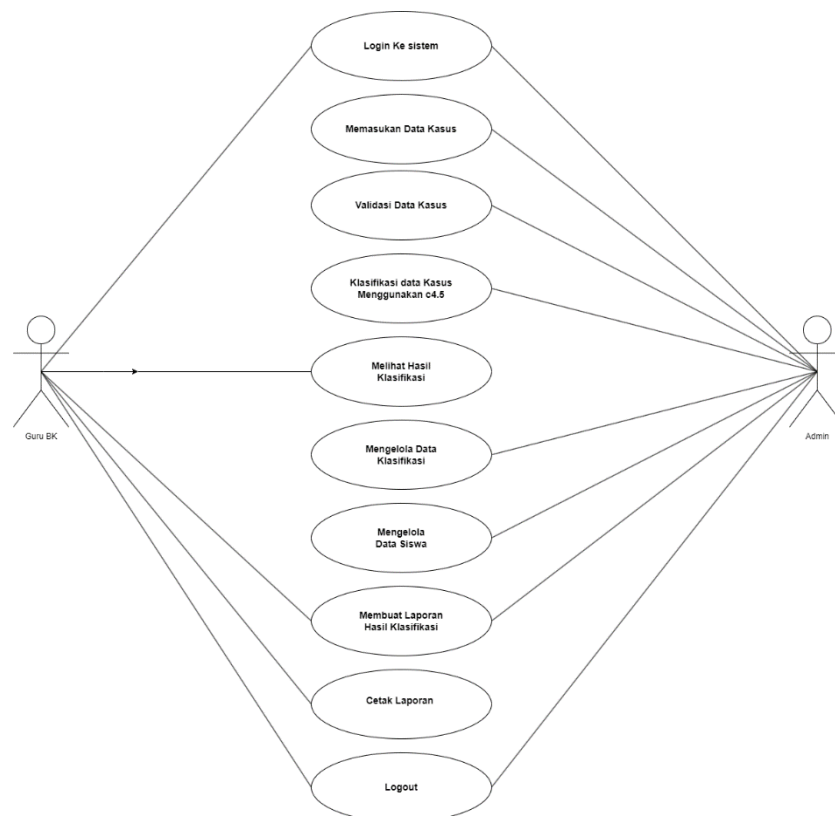
Pada tahap perancangan (*design*), dilakukan proses pemodelan sistem berdasarkan hasil analisis kebutuhan yang telah didapatkan. Selanjutnya, dikembangkan pemodelan berbasis data untuk menjelaskan keterkaitan antarelemen data. Dalam

pemodelan sistem ini, digunakan *Unified Modeling Language* (UML) yang terdiri atas beberapa jenis diagram, seperti *Diagram* :

1. *Use-Case Diagram*
2. Activity Diagram
3. Class Diagram

3.2.2.1 Use Case Diagram

Use case diagram menggambarkan fungsi-fungsi yang diharapkan dari suatu sistem, menjelaskan secara garis besar cara kerja sistem dengan mempresentasikan interaksi antara aktor yang terlibat, serta memberikan gambaran mengenai fungsi-fungsi pada sistem tersebut. Gambar 3.2 merupakan diagram use case dari sistem yang dimaksud.



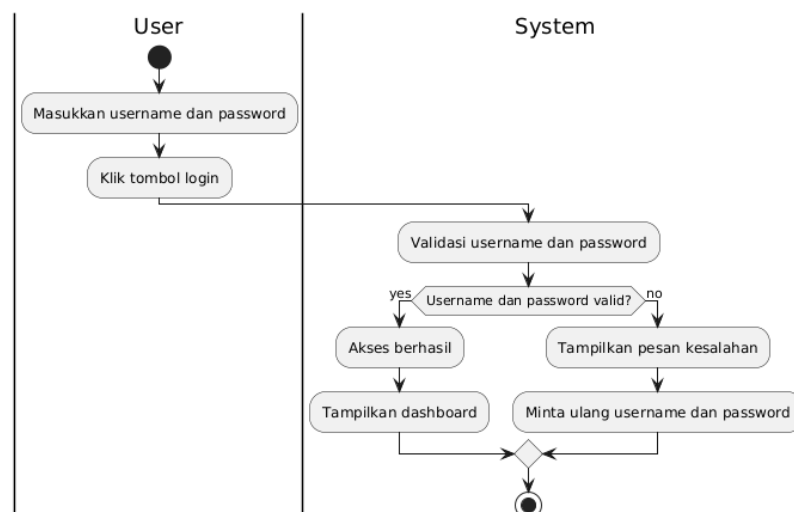
Gambar 3. 2 Use Case Diagram

3.2.2.2 Activity Diagram

Activity diagram menggambarkan alur kerja atau proses yang terjadi dalam suatu sistem. Diagram ini menjelaskan urutan aktivitas, pengambilan keputusan, serta bagaimana proses-proses tersebut berinteraksi satu sama lain. Dengan menampilkan langkah-langkah yang harus dilalui dan kondisi yang mempengaruhi jalannya aktivitas, activity diagram memberikan gambaran yang jelas tentang alur kerja sistem. Gambar dibawah merupakan diagram activity dari sistem yang dimaksud, yang menggambarkan langkah-langkah atau proses yang terjadi dalam sistem tersebut.

a) Activity Diagram login

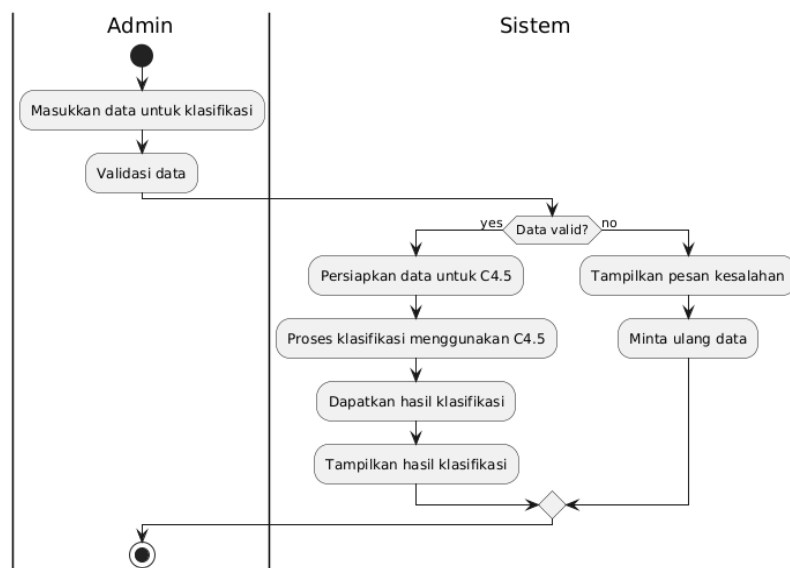
Activity diagram login menggambarkan alur proses yang terjadi saat seorang pengguna mencoba untuk masuk ke dalam sistem. Pengguna mulai dengan memasukkan username dan password, kemudian menekan tombol login. Sistem memvalidasi data yang dimasukkan, dan jika username serta password yang diberikan valid, sistem memberikan akses dan menampilkan dashboard pengguna. Namun, jika validasi gagal, sistem akan menampilkan pesan kesalahan dan meminta pengguna untuk mengulang proses dengan memasukkan ulang data yang benar. Diagram ini menggambarkan keputusan yang terjadi berdasarkan hasil validasi, baik berhasil maupun gagal.



Gambar 3. 3 Activity Diagram Login

b) Activity Diagram Klasifikasi algoritma C4.5

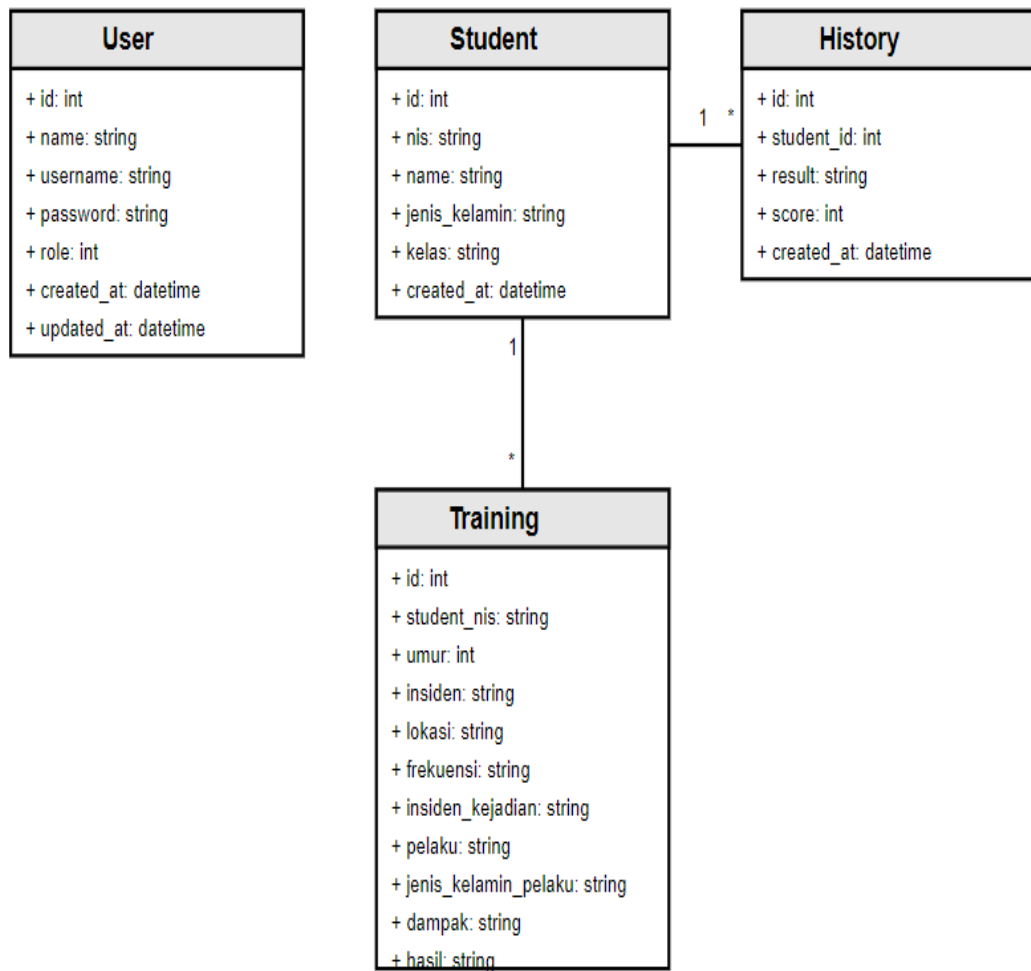
Diagram ini mengilustrasikan alur kerja seorang admin dalam sistem, dimulai dengan proses login yang divalidasi oleh sistem, kemudian admin dapat mengelola data kasus dan memantau hasil klasifikasi, sementara sistem memproses data untuk laporan yang kemudian dapat dihasilkan oleh admin, dan proses berakhir ketika admin melakukan logout dari sistem.



Gambar 3. 4 Diagram Activity klasifikasi C4.5

3.2.2.3 Class Diagram

Class diagram menggambarkan struktur dan hubungan antar kelas dalam suatu sistem. Diagram ini menjelaskan bagaimana kelas-kelas saling berinteraksi, termasuk atribut, metode, serta relasi seperti asosiasi, pewarisan, dan agregasi. Dengan menampilkan elemen-elemen yang membentuk struktur sistem, class diagram memberikan gambaran yang jelas tentang bagaimana data dan fungsi diorganisir dalam sistem. Gambar di bawah ini merupakan class diagram dari sistem yang dimaksud, yang menunjukkan keterkaitan antar kelas serta perannya dalam pengembangan sistem.



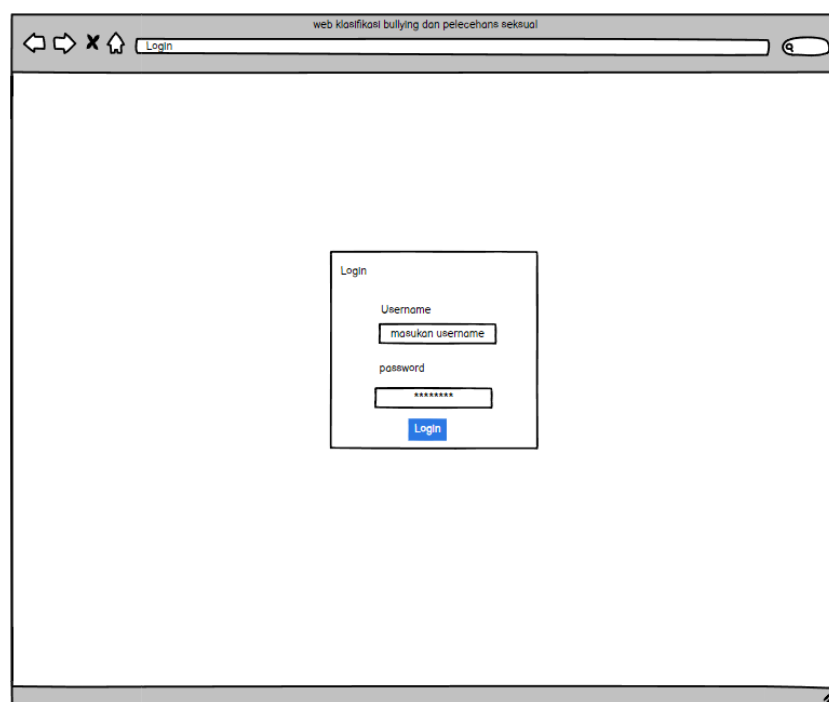
Gambar 3. 5 Class Diagram

3.2.2.4 Design Interface

Tahap ini merupakan proses implementasi dalam menyelesaikan permasalahan pada penelitian ini, yang dikembangkan berdasarkan tahapan-tahapan sebelumnya. Hasil dari analisis kebutuhan serta perancangan yang telah dilakukan kemudian diterapkan dalam bentuk program. Berikut ini adalah tampilan rancangan antarmuka dari sistem yang akan dibangun.

a) Halaman Login

yang berfungsi sebagai pintu masuk bagi Admin atau Guru Bk yang sudah memiliki akun untuk mengakses sistem. Halaman ini memiliki dua field input utama: username untuk memasukkan nama pengguna dan password untuk memasukkan kata sandi yang telah didaftarkan sebelumnya. Setelah kedua field tersebut diisi dengan benar, pengguna dapat menekan tombol "Login" untuk melakukan proses autentikasi dan masuk ke dalam sistem.

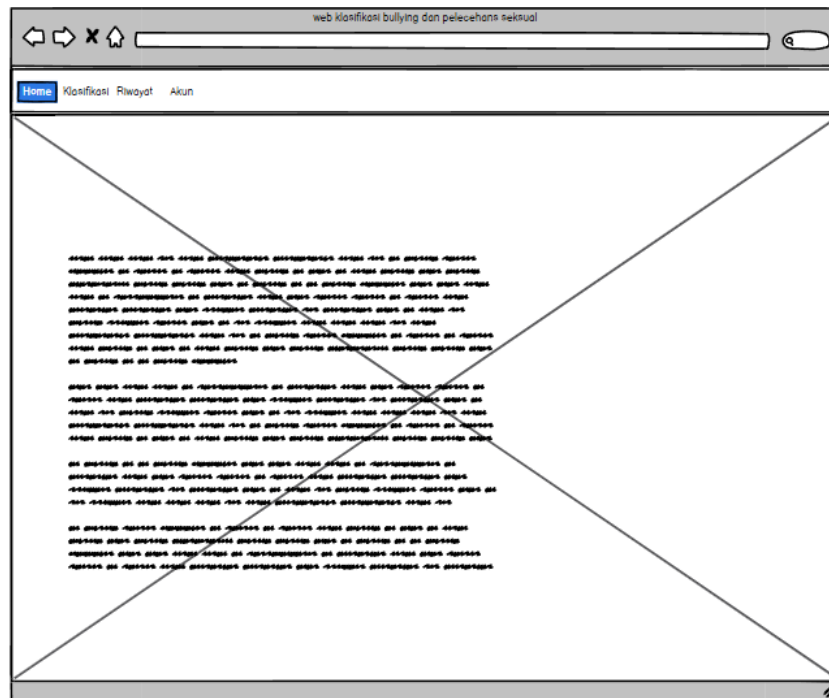


The image shows a web browser window with the title "web klasifikasi bullying dan pelecehan seksual". The address bar contains the word "Login". The main content area displays a login form with the following elements:

- A title "Login" above the input fields.
- A label "Username" followed by a text input field containing the placeholder text "masukan username".
- A label "password" followed by a password input field containing a series of asterisks "*****".
- A blue button labeled "Login" positioned below the password field.

Gambar 3. 6 Halaman login

- b) halaman utama atau dashboard setelah pengguna berhasil login. Di bagian atas terdapat menu navigasi dengan beberapa pilihan: "Beranda", "Klasifikasi", dan "Akun".



Gambar 3. 7 halaman utama

3.3 Algoritma C4.5

C4.5 adalah algoritma yang digunakan untuk membangun pohon keputusan, yang berguna dalam tugas klasifikasi. Proses utama dalam algoritma ini melibatkan pemilihan atribut terbaik berdasarkan perhitungan Information Gain (Keuntungan Informasi) yang diperoleh dari perhitungan entropi. Berikut adalah penjelasan tentang cara algoritma ini bekerja dalam memilih atribut terbaik dan membangun pohon keputusan.

Perhitungan Entropi

Entropi digunakan untuk mengukur sejauh mana ketidakpastian atau keraguan dalam dataset. Semakin tinggi nilai entropi, semakin tidak pasti data tersebut. Berikut adalah langkah-langkah dalam perhitungan entropi:

1. Menghitung Entropi Dataset Awal: Entropi dataset, $H(D)$, dihitung dengan rumus:

$$H(D) = - \sum_{i=1}^m p_i \log_2 p_i$$

Di mana:

- p_i adalah proporsi kelas ke- i dalam dataset D .
 - m adalah jumlah kelas dalam dataset.
2. Menghitung Entropi Berdasarkan Atribut: Setiap atribut A akan dihitung entropinya berdasarkan pembagian dataset menurut nilai-nilai atribut tersebut. Entropi setelah pembagian untuk atribut A , disebut Conditional Entropy, dihitung dengan rumus:

$$H(D | A) = \sum_{v \in \text{Values}(A)} \frac{|D_v|}{|D|} H(D_v)$$

Di mana:

- (A) adalah nilai-nilai yang dimiliki oleh atribut A .
 - D_v adalah subset data D yang memiliki nilai v untuk atribut A .
3. Menghitung Information Gain: Setelah menghitung entropi awal dan entropi berdasarkan pembagian untuk setiap atribut, Information Gain dihitung dengan rumus:

$$\text{Gain}(A) = H(D)$$

Atribut dengan Information Gain terbesar akan dipilih sebagai atribut pembagi di node pohon keputusan.

Setelah memilih atribut terbaik berdasarkan Information Gain, pohon keputusan akan dibangun dengan langkah-langkah berikut:

1. Memilih Atribut Terbaik: Pilih atribut dengan Information Gain terbesar untuk membagi dataset.
2. Membagi Dataset Berdasarkan Nilai Atribut: Dataset dibagi menjadi beberapa subset berdasarkan nilai dari atribut yang terpilih.
3. Penerapan Rekursif pada Setiap Subset: Langkah-langkah yang sama diterapkan secara rekursif pada setiap subset data, hingga tercapai kondisi di mana semua data dalam subset memiliki kelas yang sama.

3.3.1 Pengujian Algoritma Decision Tree C4.5

Setelah dataset yang berisi informasi terkait jenis kelamin, kelas, insiden yang dialami, pelaku, lokasi kejadian, frekuensi insiden, dampak psikologis, dan hasil klasifikasi terkumpul, tahap selanjutnya adalah menguji dan menganalisis data dengan menggunakan algoritma Decision Tree C4.5. Pada tahap ini, algoritma ini digunakan untuk membangun pohon keputusan dengan menghitung entropy dan information gain dari setiap atribut yang ada, untuk memilih atribut terbaik yang akan digunakan untuk membagi data pada setiap node pohon keputusan.

Perhitungan Entropi dan Information Gain

1. Entropi Dataset Total $H(D)$:

Langkah pertama adalah menghitung entropi total untuk seluruh dataset yang nantinya akan digunakan sebagai dasar dalam menghitung information gain untuk atribut-atribut lainnya. Dalam dataset ini, terdapat dua kategori klasifikasi yang telah dikelompokkan, yaitu Bullying dan Pelecehan Seksual. Berdasarkan data, terdapat 120 kasus, dengan 142 kasus berlabel Bullying dan 29 kasus berlabel Pelecehan Seksual. Oleh karena itu, perhitungan entropi untuk dataset total adalah sebagai berikut:

$$H(D) = - \left(\frac{93}{120} \log_2 \frac{93}{120} + \frac{27}{120} \log_2 \frac{27}{120} + \frac{13}{120} \log_2 \frac{13}{120} + \frac{12}{120} \log_2 \frac{12}{120} + \frac{8}{120} \log_2 \frac{8}{120} + \frac{7}{120} \log_2 \frac{7}{120} + \frac{6}{120} \log_2 \frac{6}{120} + \frac{5}{120} \log_2 \frac{5}{120} \right) = 2.44989$$

Jadi, entropi total untuk dataset adalah 2.44989.

2. Menghitung Entropi Berdasarkan Atribut

Pada bagian ini, entropi dihitung untuk setiap atribut dalam dataset untuk menilai tingkat ketidakpastian dalam distribusi kelas di tiap kategori insiden. Entropi digunakan untuk menggambarkan sejauh mana distribusi kelas dalam kategori tertentu terdistribusi merata. Nilai entropi yang lebih tinggi menunjukkan distribusi kelas yang lebih merata, sementara nilai yang lebih rendah menunjukkan distribusi yang lebih terpusat pada satu kelas.

Sebagai contoh, untuk kategori "Diejek, dihina, diberi julukan negatif", kita menghitung entropi dengan cara sebagai berikut:

$$H(D_{diejek}) = - \left(\frac{30}{40} \log_2 \frac{30}{40} + \frac{10}{40} \log_2 \frac{10}{40} \right) = 0.513$$

Hasil perhitungan entropi untuk kategori ini adalah 0.513, yang menunjukkan adanya ketidakpastian yang cukup besar dalam distribusi kelas "Bullying" dan "Pelecehan Seksual".

Selanjutnya, entropi dihitung untuk kategori insiden lainnya, dengan hasil sebagai berikut:

Tabel 3. 1 Tabel Perhitungan Entropy Kategori Insiden

Insiden	Entropi
Diejek, dihina, diberi julukan negatif	0.513
Dikomentari seksual, dicatcalling, dilecehkan secara verbal	0.811
Dikucilkan, dirumorkan, memengaruhi orang lain	0.000
Dipukul, ditendang, didorong, dirampas barang anda	0.000
Gestur tidak pantas, menatap dengan maksud tertentu	0.000
Mengirim pesan teks, gambar, atau video seksual melalui media sosial	0.000
Menyentuh, memegang, atau memeluk tanpa izin	0.000

Penjelasan:

1. Kategori "Diejek, dihina, diberi julukan negatif" memiliki entropi 0.513, yang menunjukkan ketidakpastian yang cukup besar dalam distribusi kelasnya.
2. Kategori "Dikomentari seksual, dicatcalling, dilecehkan secara verbal" memiliki entropi 0.811, yang lebih tinggi, mengindikasikan distribusi kelas yang lebih merata.
3. Beberapa kategori lainnya, seperti "Dipukul, ditendang, didorong, dirampas barang anda" dan "Gestur tidak pantas", memiliki entropi 0.000, yang berarti distribusi kelasnya sangat terfokus pada satu kategori saja.

Perhitungan entropi ini memberikan gambaran yang jelas mengenai distribusi kelas pada setiap insiden. Semakin tinggi entropi, semakin beragam distribusi kelas di dalam kategori tersebut, sedangkan nilai entropi yang lebih rendah menandakan distribusi kelas yang lebih terpusat.

3. Weighted Entropy untuk Atribut "Insiden"

Pada bagian ini, dilakukan perhitungan Weighted Entropy untuk atribut "Insiden apa yang pernah dialami?". Weighted Entropy memberikan bobot pada entropi masing-masing kategori berdasarkan jumlah kejadian dalam kategori tersebut. Dengan kata lain, perhitungan ini mempertimbangkan frekuensi kemunculan setiap insiden, sehingga ketidakpastian yang diukur akan mencerminkan pengaruh dari banyaknya kejadian pada masing-masing kategori.

Terdapat beberapa kategori insiden dengan entropi yang sudah dihitung sebagai berikut :

Insiden	Jumlah Kejadian	Entropi
Diejek, dihina, diberi julukan negatif	105	0.513
Dikomentari seksual, dicatcalling, dilecehkan secara verbal	8	0.811
Dikucilkan, dirumorkan, memengaruhi orang lain	27	0.000
Dipukul, ditendang, didorong, dirampas barang anda	13	0.000

Mengirim pesan teks, gambar, atau video seksual melalui media sosial	6	0.000
Menyentuh, memegang, atau memeluk tanpa izin	5	0.000
Gestur tidak pantas, menatap dengan maksud tertentu	5	0.000
Gestur tidak pantas, menatap dengan maksud tertentu	2	0.000

$$H(D \mid \text{Insiden})$$

$$\begin{aligned}
 &= \frac{105}{120} \times 0.513 + \frac{8}{120} \times 0.811 + \frac{27}{120} \times 0.000 \\
 &+ \frac{13}{120} \times 0.000 + \frac{6}{120} \times 0.000 + \frac{5}{120} \times 0.000 \\
 &+ \frac{5}{120} \times 0.000 + \frac{2}{120} \times 0.000 = 0.3528
 \end{aligned}$$

Weighted Entropy untuk atribut "Insiden" adalah 0.561. Nilai ini menunjukkan ketidakpastian keseluruhan dari distribusi kelas di seluruh kategori insiden, dengan mempertimbangkan frekuensi kemunculan masing-masing insiden dalam dataset.

4. Information Gain untuk Atribut "Insiden"

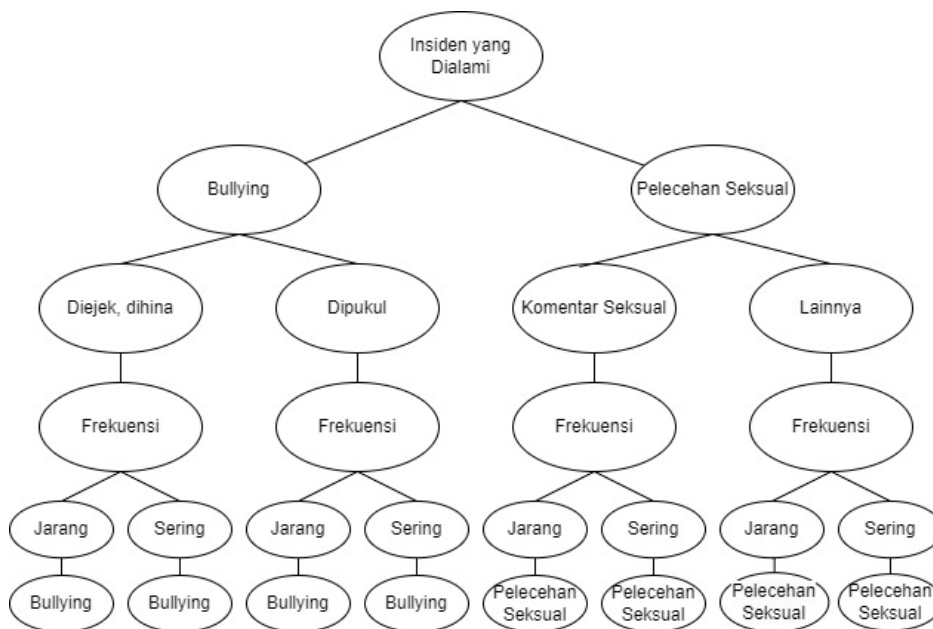
Kemudian, information gain dihitung dengan mengurangi entropi total dengan entropi berbobot yang telah dihitung sebelumnya:

$$\text{Gain}(\text{Insiden}) = H(D) - H(D \mid \text{Insiden}) = 2.164 - 0.353 = 1.811$$

5. Mengulang Proses untuk Atribut Lainnya

Langkah yang sama diterapkan untuk atribut lainnya, seperti "Lokasi Kejadian", "Pelaku", dan seterusnya. Atribut yang memiliki information gain terbesar akan dipilih untuk membagi dataset pada setiap node pohon keputusan.

Setelah proses perhitungan entropi dan information gain selesai, langkah berikutnya adalah membangun pohon keputusan dan menganalisis hasil yang didapat. Berdasarkan dataset yang ada, pohon keputusan dibentuk untuk mengklasifikasikan jenis insiden (seperti Bullying atau Pelecehan Seksual) berdasarkan beberapa atribut penting, seperti jenis insiden, frekuensi kejadian, dan dampak psikologis yang dirasakan oleh korban.



Gambar 3. 8 Pohon Keputusan

Pohon keputusan pada Gambar 3.9 yang terbentuk berdasarkan atribut-atribut yang dipilih menggambarkan struktur keputusan yang terorganisir dengan baik. Setiap node berfungsi sebagai titik keputusan, di mana dataset dibagi berdasarkan nilai atribut tertentu. Proses ini diulang secara rekursif pada setiap subset data, menghasilkan struktur yang jelas dan mudah dipahami.

- a) Pada root node, atribut "Insiden yang Dialami" dipilih karena atribut ini memiliki information gain yang lebih besar dibandingkan dengan atribut lainnya. Hal ini menunjukkan bahwa jenis insiden (seperti Diejek, dihina atau Komentar Seksual) memberikan pembeda yang signifikan dalam klasifikasi data.
- b) Setelah dibagi berdasarkan jenis insiden, pohon keputusan selanjutnya menggunakan frekuensi kejadian sebagai atribut pembagi. Ini menunjukkan bahwa seringnya insiden terjadi dapat memberikan informasi tambahan yang berguna dalam klasifikasi. Insiden yang sering terjadi lebih cenderung masuk ke kategori Bullying atau Pelecehan Seksual, sedangkan insiden yang jarang terjadi lebih cenderung berada di kategori yang sama, tetapi dengan intensitas yang lebih rendah.

Tabel 3. 2 tabel Perhitungan Entrophy dan Gain

Node	Atribut	Nilai Atribut	Jumlah Kasus	Bullying	Pelecehan Seksual	Entropy	Information Gain
Root Node	Insiden yang Dialami	Diejek, dihina, diberi julukan negatif	40	30	10	0.811	0.052
		Dipukul, ditendang, dirampas barang	20	10	10	0.811	
		Komentar seksual, pelecehan verbal	20	5	5	0.811	

Level 2 (Cabang dari "Diejek")	Frekuensi Insiden	Jarang (1- 3 kali)	20	15	5	0.811	0.032
		Sering (> 7 kali)	10	5	5	0.811	
Level 2 (Cabang dari "Dipukul")	Frekuensi Insiden	Jarang (1- 3 kali)	10	7	3	0.918	0.021
		Sering (> 7 kali)	10	3	7	0.918	
Level 2 (Cabang dari "Komentar Seksual")	Frekuensi Insiden	Jarang (1- 3 kali)	10	2	8	0.722	0.089
		Sering (> 7 kali)	10	3	7	0.722	

Pada tabel 3.1 yang disajikan sebelumnya menggambarkan proses pembagian data yang dilakukan pada setiap node dalam pohon keputusan. Tabel ini menunjukkan bagaimana dataset dibagi berdasarkan atribut yang dipilih dan bagaimana hasil klasifikasi Bullying dan Pelecehan Seksual ditentukan berdasarkan nilai information gain yang dihitung untuk setiap atribut.

1. **Root Node:** Pada node pertama pohon keputusan, atribut "Insiden yang Dialami" dipilih sebagai pembagi utama. Dataset dibagi menjadi dua kategori besar: Bullying dan Pelecehan Seksual.
 - a) Berdasarkan data, entropi untuk dataset total dihitung sebesar 0.811.
 - b) Kemudian, untuk masing-masing sub-kategori dalam "Insiden yang Dialami", dihitung entropi yang lebih kecil dan information gain-nya. Hasilnya menunjukkan bahwa atribut "Diejek, dihina" memiliki information gain terbesar untuk kategori Bullying, sementara "Komentar Seksual" memiliki information gain terbesar untuk Pelecehan Seksual.

2. **Level Kedua (Cabang dari Bullying dan Pelecehan Seksual):** Setelah pembagian pertama berdasarkan jenis insiden, dataset selanjutnya dibagi berdasarkan frekuensi kejadian insiden, yaitu Jarang (1-3 kali) atau Sering (>7 kali). Tabel menunjukkan bagaimana frekuensi kejadian mempengaruhi klasifikasi, dengan nilai information gain yang dihitung di setiap cabang.
- a) Misalnya, pada cabang "Diejek, dihina" dengan frekuensi "Jarang", hasil klasifikasi menunjukkan bahwa insiden ini lebih cenderung menjadi Bullying. Jika terjadi "Sering", hasil klasifikasinya tetap Bullying, namun dengan intensitas insiden yang lebih tinggi.
 - b) Untuk Pelecehan Seksual, "Komentar Seksual" yang sering terjadi berhubungan dengan Pelecehan Seksual. Begitu juga dengan "Lainnya", yang menunjukkan hubungan serupa.