

**SKRIPSI**

**RANCANG BANGUN APLIKASI ANALISIS SENTIMEN  
PERMENDIKBUD RISTEK “PENCEGAHAN DAN  
PENANGANAN KEKERASAN SEKSUAL DI LINGKUNGAN  
PERGURUAN TINGGI”**

**Disusun dan diajukan oleh:**

**IRMA JUFRI  
D121 17 1008**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS HASANUDDIN  
GOWA  
2024**

## LEMBAR PENGESAHAN SKRIPSI

### RANCANG BANGUN APLIKASI ANALISIS SENTIMEN PERMENDIKBUD RISTEK “PENCEGAHAN DAN PENANGANAN KEKERASAN SEKSUAL DI LINGKUNGAN PERGURUAN TINGGI”

Disusun dan diajukan oleh

**IRMA JUFRI**  
**D121 17 1008**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian  
Studi Program Sarjana Program Studi Teknik Informatika  
Fakultas Teknik Universitas Hasanuddin  
Pada tanggal 15 Januari 2024  
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,

  
Dr. Ir. Zahir Zainuddin, M.Sc.  
NIP 196404271989101002

  
Elly Warni, S.T., M.T.  
NIP 198202162008122001

Ketua Program Studi,

  
Prof. Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng.  
NIP 197507162002121004



## PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;

Nama : IRMA JUFRI  
NIM : D121171008  
Program Studi : Teknik Informatika  
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Rancang Bangun Aplikasi Analisis Sentimen Permendikbud Ristek “Pencegahan dan Penanganan Kekerasan Seksual di Lingkungan Perguruan Tinggi”

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gowa, 15 Januari 2024

Yang Menyatakan

The image shows a 2000 Rupiah postage stamp from Indonesia. The stamp features the Garuda Pancasila emblem and the text 'REPUBLIK INDONESIA', '2000', and 'METERAN TEMPEL'. A handwritten signature is written over the stamp. Below the stamp, the name 'IRMA JUFRI' is printed in a bold, black, sans-serif font.

**IRMA JUFRI**

## ABSTRAK

**IRMA JUFRI.** *Rancang Bangun Aplikasi Analisis Sentimen Permendikbud Ristek “Pencegahan dan Penanganan Kekerasan Seksual di Lingkungan Perguruan Tinggi”* (dibimbing oleh Zahir Zainuddin dan Elly Warni)

Kekerasan seksual di lingkungan perguruan tinggi telah menjadi isu yang mendapat perhatian serius dari Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbud-Ristek) Republik Indonesia. Sebagai respon terhadap isu tersebut, Kemendikbud-Ristek menerbitkan Permendikbud Ristek Nomor 30 Tahun 2021. Peraturan ini dirancang untuk mencegah dan menangani kasus-kasus kekerasan seksual di kampus. Banyak dari tanggapan masyarakat disuarakan melalui media sosial, khususnya Twitter. Namun, peraturan ini mendapat berbagai tanggapan dari masyarakat, baik yang mendukung maupun yang kontra. Penelitian ini dilakukan mengingat pentingnya memahami persepsi masyarakat terhadap peraturan ini.

Tujuan utama dari penelitian ini adalah untuk membangun model klasifikasi sentimen yang dapat mengidentifikasi dan mengklasifikasikan opini masyarakat terhadap peraturan tersebut yang dapat diterapkan pada aplikasi.

Sistem ini dirancang untuk mengklasifikasikan sentimen *tweet* dengan mengumpulkan data menggunakan metode *scraping* data *tweet*, kemudian dilakukan pelabelan data secara manual dan dilakukan *text preprocessing* untuk mengolah data menjadi lebih terstruktur agar lebih mudah diolah. Selanjutnya, dengan pembobotan kata menggunakan TF-IDF untuk mengubah teks menjadi numerik yang digunakan sebagai *input* pembelajaran *Decision Tree*, yang digunakan untuk membangun model klasifikasi analisis sentimen.

Melalui analisis sentimen, penelitian ini berhasil mengembangkan model dengan akurasi sebesar 94%. Model ini diintegrasikan ke dalam aplikasi berbasis web yang dikembangkan dengan *framework* Flask. Aplikasi ini dirancang sedemikian rupa sehingga memungkinkan pengguna untuk mengunggah data *tweet* dan mendapatkan hasil analisis sentimen secara langsung.

Kata Kunci: Analisis Sentimen, *Decision Tree*, Permendikbud-Ristek, Flask

## ABSTRACT

**IRMA JUFRI.** *Design and Development of Sentiment Analysis Application Permendikbud Ristek "Prevention and Handling of Sexual Violence in the Higher Education Institutions"* (supervised by Zahir Zainuddin and Elly Warni)

Sexual violence in higher learning environments has become an issue that has received serious attention from the Ministry of Indonesia's Ministry of Education, Culture, Research and Technology (Kemendikbud-Ristek). In response to this issue, Kemendikbud-Ristek issued Permendikbud Ristek Number 30 of 2021. This regulation is designed to protect against and handle cases of sexual violence on campus. Many of the public responses were voiced through social media, particularly Twitter. However, this regulation received various responses from the public, both in favor and against. This research was conducted given the importance of understanding people's perceptions of this regulation.

The ultimate purpose of this research is to build a sentiment classification model that can identify and classify public opinion on these regulations that can be applied to applications.

This system is designed to classify tweet sentiment by collecting data using the tweet data scraping method, then manually labeling the data and text preprocessing is used to process the data to be more structured so that it is easier to process. Then using TF-IDF to convert the text into numeric words which are used as input for Decision Tree learning, which is used to build a sentiment analysis classification model.

Through sentiment analysis, this research successfully developed a model with an accuracy of 94%. This model was integrated into a web-based application developed with the Flask framework. The application is designed in such a way that it allows users to upload tweet data and get sentiment analysis results directly.

**Keywords:** Sentiment Analysis, Decision Tree, Permendikbud-Ristek, Flask

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	ii
PERNYATAAN KEASLIAN.....	iii
ABSTRAK .....	iv
ABSTRACT.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR .....	viii
DAFTAR TABEL.....	ix
DAFTAR SINGKATAN DAN ARTI SIMBOL .....	x
DAFTAR LAMPIRAN.....	xi
KATA PENGANTAR .....	xii
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian .....	2
1.4 Manfaat Penelitian .....	2
1.5 Ruang Lingkup .....	3
<b>BAB 2 TINJAUAN PUSTAKA .....</b>	<b>4</b>
2.1 Permendikbud-Ristek Nomor 30 .....	4
2.2 Twitter.....	4
2.3 Analisis Sentimen .....	5
2.4 <i>Natural Language Processing (NLP)</i> .....	6
2.5 <i>Text Preprocessing</i> .....	6
2.6 <i>Web Scraping</i> .....	7
2.7 TF-IDF.....	8
2.8 <i>Decision Tree</i> .....	9
2.9 <i>Confusion Matrix</i> .....	10
2.10 Flask .....	13
2.11 <i>Use Case Diagram</i> .....	13
2.12 <i>Activity Diagram</i> .....	14
2.13 <i>Black Box Testing</i> .....	14

BAB 3 METODE PENELITIAN.....	15
3.1 Tahapan Penelitian.....	15
3.2 Waktu dan Lokasi Penelitian .....	16
3.3 Instrumen Penelitian .....	16
3.4 Pengambilan Data .....	17
3.5 Rancangan Sistem.....	17
3.5.1 Pelabelan.....	19
3.5.2 <i>Preprocessing</i> .....	19
3.5.3 TF-IDF.....	21
3.5.4 Klasifikasi <i>Decision Tree</i> .....	26
3.6 Perancangan Web .....	29
3.6.1 <i>Use Case Diagram</i> .....	30
3.6.1 <i>Activity Diagram</i> .....	30
3.6.2 Rancangan Antarmuka ( <i>Interface</i> ) .....	32
3.7 Skenario Pengujian .....	35
3.7.1 Pengujian Performa Model .....	35
3.7.2 Pengujian Kinerja Aplikasi.....	37
BAB 4 HASIL DAN PEMBAHASAN.....	39
4.1 Pengumpulan Data dan <i>Labelling</i> .....	39
4.2 <i>Preprocessing</i> .....	40
4.3 Vektorisasi TF-IDF .....	41
4.4 Analisis Sentimen dengan <i>Decision Tree</i> .....	43
4.4.1 Evaluasi Model <i>Decision Tree</i> .....	44
4.4.2 Interpretasi Model.....	45
4.5 Implementasi Sistem .....	46
4.5.1 Implementasi Model .....	46
4.5.2 Implementasi Web .....	47
4.6 Pengujian Fungsional Sistem .....	52
4.6.1 Pengujian Halaman Dataset.....	52
4.6.2 Pengujian Halaman Hasil Analisis .....	53
4.6.3 Pengujian Hasil Prediksi.....	54
BAB 5 KESIMPULAN DAN SARAN .....	55
5.1 Kesimpulan .....	55
5.2 Saran .....	55
DAFTAR PUSTAKA .....	56
LAMPIRAN.....	60

## DAFTAR GAMBAR

Gambar 1 Contoh <i>Decision Tree Classifier</i> .....	9
Gambar 2 Diagram alur penelitian .....	15
Gambar 3 Diagram alur perancangan sistem .....	18
Gambar 4 <i>Flowchart</i> pembobotan kata.....	22
Gambar 5 Diagram alir pemodelan <i>Decision Tree</i> .....	27
Gambar 6 Contoh pohon keputusan dari TF-IDF .....	28
Gambar 7 Sampel Visualisasi <i>Decision Tree</i> .....	28
Gambar 8 <i>Use case diagram</i> .....	30
Gambar 9 <i>Activity diagram</i> web sentimen .....	31
Gambar 10 Rancangan tampilan halaman utama.....	32
Gambar 11 Rancangan tampilan halaman dataset .....	33
Gambar 12 Rancangan tampilan halaman hasil analisis .....	33
Gambar 13 Rancangan tampilan halaman Interpretasi .....	34
Gambar 14 Rancangan tampilan halaman setelah diprediksi .....	35
Gambar 15 Sampel data dan label.....	39
Gambar 16 Rincian jumlah dataset .....	39
Gambar 17 Visualisasi <i>Term Frequency</i> .....	41
Gambar 18 Visualisasi TF-IDF.....	42
Gambar 19 <i>Confusion Matrix Decision Tree</i> .....	43
Gambar 20 Hasil prediksi model .....	45
Gambar 21 Fungsi Flask prediksi .....	46
Gambar 22 Halaman utama.....	47
Gambar 23 Halaman dataset .....	48
Gambar 24 Hasil analisis <i>tweet</i> positif.....	48
Gambar 25 Hasil analisis <i>tweet</i> negatif.....	49
Gambar 26 Hasil analisis <i>tweet</i> netral.....	49
Gambar 27 Halaman Interpretasi .....	50
Gambar 28 Hasil prediksi positif .....	51
Gambar 29 Hasil prediksi netral .....	51
Gambar 30 Hasil prediksi negatif .....	52

## DAFTAR TABEL

Tabel 1. <i>Confusion Matrix</i> .....	10
Tabel 2. <i>Multiclass Confusion Matrix</i> .....	11
Tabel 3. Contoh <i>cleansing</i> .....	20
Tabel 4. Contoh <i>case folding</i> .....	20
Tabel 5. Contoh tokenisasi .....	20
Tabel 6. Contoh <i>stopword removal</i> .....	21
Tabel 7. Contoh <i>stemming</i> .....	21
Tabel 8. Sampel dataset.....	23
Tabel 9. Contoh hasil perhitungan TF .....	24
Tabel 10. Contoh hasil perhitungan IDF.....	25
Tabel 11. Contoh hasil perhitungan TF-IDF.....	26
Tabel 12. Skenario <i>Black Box Testing</i> .....	38
Tabel 13. Sampel hasil <i>preprocessing</i> .....	40
Tabel 14. Nilai <i>Inverse Document Frequency</i> .....	42
Tabel 15. Evaluasi model <i>Decision Tree</i> .....	44
Tabel 16. Perhitungan <i>Classification Report</i> .....	45
Tabel 17. <i>Black Box Testing</i> halaman dataset.....	52
Tabel 18. <i>Black Box Testing</i> halaman hasil analisis .....	53
Tabel 19. <i>Black Box Testing</i> hasil prediksi .....	54

## DAFTAR SINGKATAN DAN ARTI SIMBOL

<b>Lambang/Singkatan</b>	<b>Arti dan Keterangan</b>
CSV	<i>comma separated value</i>
IDF	<i>Inverse Document Frequency</i>
NLP	<i>Natural Language Processing</i>
TF-IDF	<i>Term Frequency - Inverse Document Frequency</i>
TF	<i>Term Frequency</i>
KBBI	Kamus Besar Bahasa Indonesia
URL	<i>Uniform Resource Locator</i>
Permendikbud-Ristek	Peraturan Menteri Pendidikan Kebudayaan Riset dan Teknologi
TPos	<i>True Negative</i>
FPosNeg	<i>False Positive Negative</i>
FPosNet	<i>False Positive Neutral</i>
TNeg	<i>True Negative</i>
FNegPos	<i>False Negative Positive</i>
FNegNet	<i>False Negative Neutral</i>
TNet	<i>True Neutral</i>
FNetPos	<i>False Neutral Positive</i>
FNetNeg	<i>False Neutral Negative</i>
RT	<i>Repost Tweet</i>

**DAFTAR LAMPIRAN**

Lampiran 1 Surat penugasan pembimbing.....	60
Lampiran 2 Lembar perbaikan skripsi .....	61

## KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT, Tuhan yang maha Esa, yang memberikan rahmatnya sehingga penulis dapat menyelesaikan tugas akhir dengan judul **Rancang Bangun Aplikasi Analisis Sentimen Permendikbud Ristek “Pencegahan dan Penanganan Kekerasan Seksual di Lingkungan Perguruan Tinggi”** sebagai salah satu persyaratan dalam menyelesaikan program Strata-1 pada Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin.

Penulis menyadari bahwa dalam penyusunan tugas akhir ini tidak lepas dari bantuan, dukungan dan bimbingan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Kedua orang tua penulis, Bapak Jufri dan Ibu Idia, atas cinta, doa, dan dukungan tanpa henti yang selalu mengiringi setiap langkah penulis dalam perjalanan pendidikan ini.
2. Dr. Ir. Zahir Zainuddin, M.Sc., selaku dosen pembimbing I dan Elly Warni, S.T., M.T., selaku dosen pembimbing II, atas bimbingan, arahan, dan ilmu yang telah diberikan dalam membimbing penulis selama proses penyelesaian tugas akhir ini.
3. Bapak Prof. Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng. selaku Ketua Departemen Teknik Informatika Universitas Hasanuddin.
4. Segenap Dosen dan Staf Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah membantu penulis selama perkuliahan dan penyelesaian tugas akhir.
5. Adik-adik penulis, Irna, Iswan dan Ilham, atas semangat dan dukungan mereka yang turut mendorong penulis untuk tetap berjuang selama ini.
6. Devy, Jume, Lina dan Yudi telah menemani dalam suka maupun duka serta membantu dan mendukung penulis.
7. Teman-teman RECOGNIZER atas dukungan, semangat, bantuan serta pengalaman yang telah diberikan kepada penulis.

8. Serta berbagai pihak yang turut memberikan dukungan, bantuan, dan kontribusi dalam berbagai bentuk, yang tidak dapat penulis sebutkan satu persatu.

Penulis berharap semoga Allah SWT membalas dengan sebaik-baiknya segala kebaikan yang telah penulis terima dari berbagai pihak yang telah membantu memudahkan dalam menyelesaikan tugas akhir ini. Penulis menyadari banyak kekurangan dan masih jauh dari kata sempurna dalam penulisan tugas akhir ini. Oleh karena itu, penulis menerima segala bentuk masukan, kritikan dan saran guna menyempurnakan tugas akhir ini. Semoga tugas akhir ini dapat memberikan manfaat bagi semua pihak yang membacanya.

Gowa, 15 Januari 2024

Penulis  
IRMA JUFRI

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Kementerian Pendidikan, Kebudayaan, Riset, Dan Teknologi (Kemendikbud-Ristek) Republik Indonesia ini telah menerbitkan peraturan terkait pencegahan dan penanganan kekerasan seksual di lingkungan kampus. Aturan ini dimuat dalam Peraturan Mendikbud Ristek (Permendikbud Ristek) Nomor 30 Tahun 2021 Tentang Pencegahan dan Penanganan kekerasan Seksual di Lingkungan Perguruan Tinggi yang ditandatangani oleh Menteri Pendidikan Kebudayaan Riset dan Teknologi pada tanggal 31 Agustus 2021. Hal ini banyak menuai pro maupun kontra dari masyarakat yang menyuarakan pendapatnya melalui Twitter. Nadiem Makarim selaku Mendikbud Ristek mengatakan bahwa situasi perguruan tinggi saat ini dapat dibilang gawat darurat karena banyaknya kasus kekerasan seksual. Data dari komnas perempuan dari 2015-2020 sebanyak 27% dari aduan mengenai kekerasan dan pelecehan seksual terjadi di jenjang perguruan tinggi. Dari survei eksternal (2019) terhadap 127 responden dari 79 kampus 29 kota, 89% yang berisi perempuan pernah mengalami kekerasan atau pelecehan seksual, dan 4% lainnya berisi laki-laki juga mengaku pernah mengalami hal serupa. Data terakhir yaitu hasil survei Ditjen Diktiristek terhadap dosen perguruan tinggi (2020) sebanyak 77% dosen mengaku bahwa kasus kekerasan seksual pernah terjadi di kampus mereka, dan 63% dari kasus tersebut tidak dilaporkan. Hal tersebut yang menjadi pertimbangan dibuatnya peraturan spesifik dan khusus, untuk melindungi korban di dalam lingkungan kampus (Muhammad Tri Ajie, 2021).

Analisis sentimen adalah sebuah proses untuk menentukan sentimen atau opini dari seseorang yang diwujudkan dalam bentuk teks dan biasa dikategorikan sebagai sentimen positif, negatif dan netral. Pengguna internet banyak menuliskan pengalaman, opini, dan segala hal yang menjadi perhatian mereka. Tulisan tentang apa yang mereka rasakan ini merupakan perasaan positif, negatif dan netral yang biasa diungkapkan dengan cara yang cukup kompleks (Lestari et al., 2020).

Melihat permasalahan yang ada, pada penelitian ini dilakukan analisis sentimen untuk mengetahui opini masyarakat Indonesia terhadap Peraturan Menteri Pendidikan, Kebudayaan, Riset dan Teknologi perihal “Pencegahan dan Penanganan Kekerasan Seksual di Lingkungan Perguruan Tinggi”. Meningkatnya pengguna media sosial Twitter, yang bekerja secara *real-time* untuk masyarakat dapat menyampaikan keluh kesah maupun membenarkan terhadap pemberlakuan dengan demikian perlu dibuat sebuah sistem yang dapat melakukan klasifikasi suatu *tweet* yang berisikan opini termasuk ke dalam kelas sentimen. Adapun Tujuan penelitian ini adalah untuk membangun model yang dapat mengklasifikasikan komentar masyarakat ke dalam kategori sentimen positif, negatif dan netral menggunakan metode *Decision Tree*. *Decision Tree* dapat diterapkan ke semua tipe data seperti nominal, ordinal, numerik, dll (Rathi et al., 2018).

## **1.2 Rumusan Masalah**

Bagaimana cara membangun model klasifikasi sentimen terhadap Peraturan Menteri Pendidikan Kebudayaan Riset dan Teknologi Nomor 30 tentang “Pencegahan dan Penanganan Kekerasan Seksual di Lingkungan Perguruan Tinggi” yang dapat diimplementasikan ke dalam aplikasi?

## **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah membangun model klasifikasi sentimen yang dapat diterapkan pada aplikasi untuk menganalisis sentimen Permendikbud Ristek Nomor 30 perihal “Pencegahan dan Penanganan Kekerasan Seksual di Lingkungan Perguruan Tinggi”.

## **1.4 Manfaat Penelitian**

Adapun manfaat yang diharapkan dengan dilakukannya penelitian ini adalah:

1. Masyarakat, penelitian ini diharapkan dapat memberikan informasi dan pemahaman kepada masyarakat terkait sentimen publik terhadap Permendikbud-Ristek Nomor 30.

2. Pembuat Kebijakan, penelitian ini diharapkan dapat membantu dalam mengevaluasi kebijakan Permendikbud-Ristek Nomor 30 berdasarkan sentimen masyarakat.
3. Peneliti, penelitian ini diharapkan dapat menjadi referensi bagi peneliti mendatang terkait analisis sentimen.

### **1.5 Ruang Lingkup**

Adapun batasan masalah pada penelitian ini sebagai berikut:

1. Kumpulan data yang digunakan menggunakan bahasa Indonesia yang berasal dari Twitter.
2. Penelitian ini menggunakan data dengan topik Permendikbud-Ristek Nomor 30 tahun 2021.
3. Metode klasifikasi yang digunakan, yaitu *Decision Tree* dengan menerapkan teknik TF-IDF sebagai representasi dari fitur teks.
4. Menggunakan *framework* Flask dalam membangun aplikasi berbasis web.

## **BAB 2 TINJAUAN PUSTAKA**

### **2.1 Permendikbud-Ristek Nomor 30**

Permendikbud Ristek merupakan singkatan dari Peraturan Menteri Pendidikan, Budaya, Riset dan Teknologi yang mengatur kebijakan pendidikan khususnya di tingkatan perguruan tinggi. Sebagai Mendikbud Ristek yang memangku kebijakan berhak menentukan peraturan yang akan diterapkan. Setiap kebijakan yang dibuat dan diputuskan oleh pemangku kebijakan tentu memiliki tujuan yang jelas. Sebagaimana definisi kebijakan menurut beberapa ahli dapat berarti satu tindakan yang dilakukan oleh badan pemerintah dengan maksud dan tujuan tertentu. Definisi lain tentang kebijakan adalah satu pilihan yang harus dilakukan atau tidak dilakukan oleh pemerintah untuk mengatasi suatu persoalan atau masalah yang sedang terjadi seperti persoalan yang terjadi di dunia pendidikan (Adawiyah, 2022).

Peraturan Menteri Pendidikan, Kebudayaan, Riset dan Teknologi Nomor 30 Tahun 2021 mendefinisikan dengan sangat jelas tentang kekerasan seksual sebagai *focus* daripada pembahasannya. Dalam Permendikbud perlindungan dan hak korban dijadikan prioritas utama. Nadiem Makarim menegaskan Permendikbud ini menjadi wadah perlindungan korban serta mencegah terjadinya keberlanjutan kasus kekerasan yang dialaminya. Diuraikan lebih jelas bahwa “untuk mencegah dan menangani kekerasan seksual, Permendikbud menggunakan prinsip kepentingan terbaik bagi korban, keadilan serta kesetaraan gender, kesetaraan hak dan aksesibilitas bagi penyandang disabilitas, akuntabilitas, independen, kehati-hatian, konsisten, dan jaminan ketidakberulangan”. Seluruh elemen yang ada di Lingkungan Perguruan Tinggi baik pelajar maupun pendidik dijadikan target sesuai dengan yang termuat dalam Permendikbud (Di & Muhammadiyah, 2022).

### **2.2 Twitter**

Twitter merupakan media sosial yang memungkinkan pengguna untuk mengekspresikan opini dan perasaan mereka mengenai banyak isu atau

permasalahan. Berbeda dengan media sosial yang lain yang harus menjadi teman terlebih dahulu baru dapat berinteraksi, Twitter memungkinkan antar pengguna tetap terhubung walaupun mereka tidak saling berteman (Arsya Monica Pravina, Imam Cholissodin, 2019).

Twitter dapat digunakan untuk analisis sentimen. Data *tweet* di Twitter sering kali digunakan untuk melacak dan menganalisis pendapat dan perasaan pengguna terhadap topik, merek, atau peristiwa tertentu (Desai & Mehta, 2017). Twitter merupakan media komunikasi yang dapat digunakan masyarakat untuk berekspresi dan lebih diminati oleh masyarakat dunia. Pada twitter terdapat istilah *tweet* yaitu pengguna twitter dapat memberikan kabar terbaru, berekspresi, beraspirasi, dan beropini yang ditulis oleh pengguna twitter lainnya terutama topik atau hal yang sedang menjadi perbincangan utama (Darwis et al., 2021).

### **2.3 Analisis Sentimen**

Analisis sentimen adalah bagian dari teks mining yang mengelompokkan polaritas dari teks, pengelompokan tersebut dilakukan untuk melihat bagaimana polaritas dari suatu *tweet* apakah opini yang diberikan bersifat positif, negatif atau netral (Kurniawan & Apriliani, 2020).

Analisis sentimen terdiri atau juga biasa disebut dengan *opinion mining* adalah salah satu bagian dari *Text Mining*. Tugas dasar dalam analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral. *Sentiment analysis* juga dapat menyatakan perasaan emosional sedih, gembira, atau marah. Analisis sentimen adalah proses pengumpulan, pemrosesan, pemahaman terhadap opini, perasaan, dan sikap yang diungkapkan dalam teks oleh individu atau kelompok. Tujuannya adalah untuk mengklasifikasikan sentimen tersebut menjadi positif, negatif, atau netral (Sukhavasi, 2021).

## **2.4 Natural Language Processing (NLP)**

*Natural Language Processing* (NLP) adalah salah satu bidang ilmu komputer yang merupakan cabang dari kecerdasan buatan, dan bahasa (*linguistik*) yang berkaitan dengan interaksi antara komputer dan bahasa alami manusia, seperti bahasa Indonesia atau bahasa Inggris. Tujuan utama dari studi NLP adalah membuat mesin yang mampu mengerti dan memahami makna bahasa manusia lalu memberikan respon yang sesuai (Yunefri et al., 2021).

NLP merupakan proses pembuatan model komputasi dari bahasa sehingga memungkinkan terjadinya interaksi manusia dengan komputer dengan perantara alami yang dipakai oleh manusia. NLP memodelkan pengetahuan terhadap bahasa dari segi kata, bagaimana kata-kata bergabung menjadi satu kalimat dan konteks kata dalam kalimat. Pada proses pemodelan komputasi dibutuhkan *text preprocessing* yang terdiri dari beberapa tahapan yaitu, *case folding*, *tokenizing*, *filtering* dan *stemming* (Agustina Purwitasari et al., 2022).

## **2.5 Text Preprocessing**

*Text Preprocessing* adalah proses perubahan bentuk data yang belum terstruktur menjadi data yang terstruktur sesuai dengan kebutuhan untuk proses *data mining*. Tujuan *preprocessing* dalam *data mining* adalah mentransformasi data ke suatu format yang prosesnya lebih mudah dan efektif sehingga diperoleh hasil kualitas *mining* yang baik (Khairunnisa et al., 2021). *Preprocessing* yang biasa dilakukan terhadap dokumen adalah sebagai berikut:

### **1. Case Folding dan Cleansing**

Pada tahapan ini dilakukan penyeragaman seluruh teks menjadi huruf kecil (*lowercase*) dan pembersihan atau penghapusan pada semua dokumen yang berisi angka, url (<http://>), *username* (@), tanda pagar (#), *delimiter* seperti koma (,) dan titik (.) dan juga tanda baca lainnya (Septian et al., 2019).

## **2. Stopwords Removal**

*Stopwords Removal* adalah suatu proses menghilangkan kata-kata umum yang tidak memiliki makna. Penghilangan beberapa kata kerja, kata sifat, dan kata keterangan lainnya dapat dimasukkan ke dalam daftar *stopwords*. Karakteristik utama dalam pemilihan *stopword* biasanya adalah kata yang mempunyai frekuensi kemunculan yang tinggi misalnya kata penghubung seperti “dan”, “atau”, “tapi”, “sehingga” dan lainnya. Tujuan utama dari *stopwords removal* adalah menghilangkan kata-kata yang tidak memiliki makna, sehingga dapat meningkatkan kecepatan dan performa pemrosesan. *Stopword* termasuk penentu, konjungsi, preposisi dan sejenisnya (Christianto et al., 2020).

## **3. Stemming**

Proses untuk melakukan pemotongan awalan atau akhiran kata dan diubah ke bentuk akarnya yang bertujuan untuk menghilangkan imbuhan (Agustina Purwitasari et al., 2022). Tujuan dari proses *stemming* adalah menghilangkan imbuhan-imbuhan. Jika imbuhan tersebut tidak dihilangkan maka setiap satu kata dasar akan disimpan dengan berbagai macam bentuk yang berbeda sesuai dengan imbuhan yang melekatinya sehingga hal tersebut akan menambah beban *database*.

## **4. Tokenisasi**

Pada tahapan ini dilakukan proses pemotongan pada sebuah dokumen ataupun kalimat menjadi kata atau biasa disebut token. Setiap kata pada proses *tokenizing* akan diidentifikasi berdasarkan spasi (Kurniawan & Apriliani, 2020).

## **2.6 Web Scraping**

*Web scraping* adalah proses pengambilan data dari halaman web secara otomatis menggunakan perangkat lunak atau skrip yang dirancang khusus. Teknik ini memungkinkan pengguna untuk mengambil informasi yang terdapat pada halaman web dan mengubahnya menjadi data terstruktur yang dapat diolah secara komputasional. *Web Scraping* merupakan suatu teknik untuk mengutip dan mengekstraksi data atau informasi dari suatu *website* dengan menggunakan *low-level HTML*. *Web Scraping* ini mengambil data kotor secara *real-time* dari

*website* twitter, yang selanjutnya akan dipilih menjadi data *tweet* bersih (Hamdana, 2021).

## 2.7 TF-IDF

TF-IDF (*Term Frequency – Inverse Document Frequency*) adalah ukuran statistik numerik yang digunakan untuk mengetahui seberapa penting suatu suku kata dalam sebuah dokumen atau korpus. TF-IDF dapat digunakan untuk pembobotan kata sebagai fitur untuk klasifikasi sentimen kalimat. TF (*Term Frequency*) adalah frekuensi kemunculan kata pada tiap dokumen yang diberikan, menunjukkan seberapa penting kata itu di dalam tiap dokumen tersebut. DF (*Document Frequency*) adalah frekuensi dokumen yang mengandung kata menunjukkan seberapa umum kata tersebut. IDF adalah *inverse* dari nilai DF. Hasil dari pembobotan kata menggunakan TF-IDF ini adalah hasil perkalian dari TF dikalikan dengan IDF. Bobot kata semakin besar jika sering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam banyak dokumen (Septian et al., 2019).

Untuk menghitung nilai DF digunakan persamaan 1.

$$Idf_t = \log \left( \frac{N}{df_t} \right) \quad (1)$$

Untuk menghitung pembobotan TF-IDF, menggunakan persamaan 2

$$Tf.idf_{t,d} = tf_{t,d} \times Idf_t \quad (2)$$

Dimana:

$Idf_t$  = inversi frekuensi dokumen dari kata t

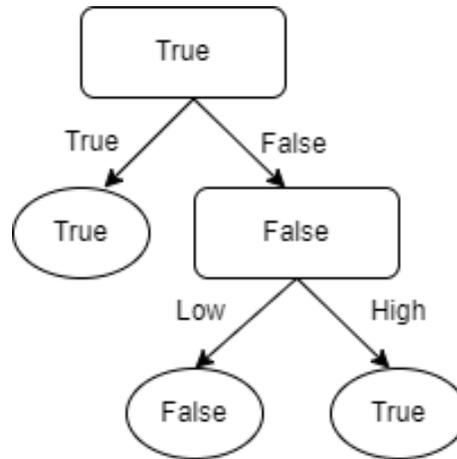
N = banyaknya dokumen

$df_t$  = banyaknya dokumen yang mengandung kata t

$Tf.idf_{t,d}$  = nilai bobot kata t pada dokumen d

$tf_{t,d}$  = frekuensi kemunculan term t pada dokumen d

## 2.8 Decision Tree



Gambar 1 Contoh *Decision Tree Classifier*

Berdasarkan Gambar 1, *Decision tree* merupakan salah satu metode untuk mengklasifikasikan data. Model *Decision Tree* merupakan sebuah *tree* yang terdiri dari *root node*, *internal node* dan *terminal node*. Sementara *root node* dan *internal node* merupakan variabel/fitur, *terminal node* adalah label kelas. Dalam melakukan klasifikasi, sebuah data *query* akan menelusuri *root node* dan *internal node* sampai mencapai *terminal node*. Pelabelan kelas data *query* berdasarkan label di *internal node*. Pada *Decision Tree* tradisional, data yang digunakan adalah data dengan nilai fitur yang sudah pasti. Pada penelitian ini bertujuan untuk membangun suatu *Decision Tree* yang dapat menangani data dengan nilai fitur yang tidak pasti yaitu data yang memiliki *range* nilai. Selama ini penanganan data dengan fitur yang tidak pasti menggunakan metode *averaging*. Akan tetapi ada penelitian lain yang menyebutkan bahwa ada metode lain yang lebih handal dalam menangani data dengan nilai fitur yang tidak pasti. Metode tersebut adalah metode berbasis distribusi (Tri Romadloni et al., 2019).

Algoritma CART (*Classification and Regression Trees*) adalah algoritma berbasis *Decision Tree* yang dapat digunakan untuk masalah klasifikasi dan regresi dalam pembelajaran mesin. Algoritma ini bekerja dengan mempartisi data pelatihan secara rekursif menjadi *subset* yang lebih kecil menggunakan pemisahan biner. CART adalah algoritma yang kuat dan populer karena kemampuannya untuk menangani fitur kategorikal dan kontinu, kemampuan interpretasinya, dan kemampuannya untuk menangkap hubungan non-linear antara fitur dan variabel

target. Algoritma *Classification and Regression Trees* (CART) selalu membuat pohon biner, yang yang berarti bahwa setiap *node* non-terminal memiliki dua *node* anak. Hal ini berbeda dengan metode metode berbasis pohon lainnya, yang memungkinkan beberapa *node* anak. Setelah pohon dibangun, pohon tersebut dapat digunakan untuk membuat prediksi dengan menelusuri pohon dari *node* awal ke *node* daun. Untuk masalah regresi, prediksi adalah rata-rata dari nilai target pada *node* daun. Untuk masalah klasifikasi, prediksinya adalah kelas mayoritas di *node* daun ( Ayush et al., 2023).

## 2.9 Confusion Matrix

*Confusion Matrix* adalah tabel silang (*cross table*) yang menunjukkan jumlah kemunculan antara dua nilai kelas yaitu klasifikasi yang sebenarnya atau bernilai aktual dan klasifikasi prediksi, seperti yang ditunjukkan pada Tabel 1, kolom-kolom menunjukkan prediksi model sedangkan baris-baris menunjukkan klasifikasi yang sebenarnya (Grandini et al., 2020).

Tabel 1. *Confusion Matrix*

		Kelas Prediksi	
		Positif	Negatif
Kelas Aktual	Positif	TP	FN
	Negatif	FP	TN

Dimana:

TP (*True Positive*): Jumlah data positif yang diprediksi dengan benar menjadi positif oleh model.

TN (*True Negative*): Jumlah data negatif yang diprediksi dengan benar menjadi negatif oleh model.

FP (*False Positive*): Jumlah data negatif yang salah diprediksi sebagai positif oleh model

FN (*False Negative*): Jumlah data positif yang salah diprediksi sebagai negatif oleh model .

*Confusion matrix* membantu dalam memahami sejauh mana model dapat membedakan antara kelas positif dan negatif serta seberapa sering model menghasilkan prediksi yang salah. Informasi dari *confusion matrix* juga dapat digunakan untuk menghitung metrik evaluasi seperti *akurasi*, *presisi*, *recall*, dan *F1-score*. Dalam penelitian ini, digunakanlah *Multiclass Confusion matrix* karena terdapat tiga nilai sentimen yang berbeda, yaitu netral, positif, dan negatif. Tabel *Multiclass Confusion Matrix* dapat dilihat pada Tabel 2.

Tabel 2. *Multiclass Confusion Matrix*

		Kelas Prediksi		
		Negatif	Netral	Positif
Kelas Aktual	Negatif	TNeg	FNetNeg	FPosNeg
	Netral	FNegNet	TNet	FPosNet
	Positif	FNegPos	FNetPos	TPos

Dimana:

TNeg (*True Negative*) : Jumlah data yang sebenarnya kelas negatif dan model dengan benar memprediksi bahwa data tersebut adalah negatif.

FNetNeg (*False Neutral Negative*) : Jumlah data yang seharusnya kelas negatif, tetapi model memprediksi sebagai kelas netral.

FPosNeg (*False Positive Negatif*) : Jumlah data yang seharusnya negatif, tetapi model memprediksi sebagai kelas positif.

TNet (*True Neutral*) : Jumlah data yang sebenarnya kelas netral dan model memprediksi bahwa data tersebut adalah kelas netral.

FNegNet (*False Negative Neutral*) : Jumlah data yang sebenarnya kelas netral, tetapi model memprediksi sebagai kelas negatif.

FPosNet (*False Positive Neutral*) : Jumlah data yang seharusnya netral, tetapi model memprediksi sebagai kelas positif.

TPos (*True Positive*) : Jumlah data yang sebenarnya kelas positif dan model dengan benar memprediksi bahwa data tersebut adalah kelas positif.

FNegPos (*False Negative Positive*) : Jumlah data yang seharusnya kelas positif, tetapi model memprediksi sebagai bukan kelas negatif.

FNetPos (*False Neutral Positive*) : Jumlah data yang seharusnya positif, tetapi model memprediksi sebagai kelas netral.

Untuk menentukan performa matriks pada sebuah algoritma diperlukan *precision*, *recall*, *f1-score* dan akurasi yang dihitung berdasarkan nilai dari *confusion matrix* (Singh et al., 2021).

- a. *Precision* (P) didefinisikan sebagai rasio *item* relevan yang dipilih terhadap semua *item* yang terpilih. *Precision* merupakan probabilitas bahwa sebuah *item* yang dipilih adalah relevan. Dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan itu (Siddik et al., 2019). *Precision* dapat dihitung menggunakan persamaan 3.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

- b. *Recall* (R) didefinisikan sebagai rasio dari *item* relevan yang dipilih terhadap total jumlah *item* relevan yang tersedia. *Recall* merupakan probabilitas bahwa suatu *item* yang relevan akan dipilih. *Recall* dapat dihitung dengan jumlah rekomendasi yang relevan yang dipilih oleh *user* dibagi dengan jumlah semua rekomendasi yang relevan baik dipilih maupun rekomendasi yang tidak terpilih (Siddik et al., 2019). *Recall* dapat dihitung menggunakan persamaan 4.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

- c. *f1-score* adalah metrik evaluasi yang mengukur keseimbangan antara *precision* dan *recall* dalam suatu model klasifikasi. *F1-Score* sangat berguna ketika terdapat ketidakseimbangan antara kelas positif dan negatif dalam dataset, di mana akurasi saja mungkin tidak memberikan gambaran yang akurat tentang kinerja model (Singh et al., 2021). Sehingga dapat dihitung menggunakan persamaan 5.

$$f1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

- d. Akurasi (A) adalah jumlah dokumen yang diklasifikasikan dengan benar, baik itu *true positive* maupun *true negative*. Akurasi yang dihasilkan dari sebuah klasifikasi yang memberikan hasil latih dengan bentuk persentase dari kelompok data latih yang telah diklasifikasikan dengan benar (Siddik et al., 2019). *Accuracy* dapat diperoleh dengan menggunakan persamaan 6.

$$Akurasi = \frac{TP+TN}{(TP+TN+FP+FN)} 100\% \quad (6)$$

## 2.10 Flask

Flask adalah sebuah kerangka kerja (*framework*) web yang ditulis dalam bahasa pemrograman Python. Kerangka kerja web pada dasarnya adalah sebuah kerangka kerja atau struktur yang menyediakan alat-alat dan komponen yang diperlukan yang untuk membangun aplikasi web. Flask berfungsi sebagai kerangka aplikasi dan tampilan dari suatu web. Dengan menggunakan Flask dan bahasa Python, pengembang dapat membuat sebuah web yang berstruktur dan dapat mengatur *behavior* suatu web dengan mudah (Ningtyas & Setiyawati, 2021).

*Web framework* Flask ditulis menggunakan bahasa Python, sehingga sebelum Flask dapat digunakan, maka *developer* harus menginstall Python pada perangkat yang akan digunakan. Oleh sebab itu, *web developer* yang akan menggunakan Flask sebagai *web framework* untuk *web development* harus setidaknya mempelajari bahasa pemrograman Python terlebih dahulu, sebelum dapat menggunakan Flask seutuhnya (Somya, 2018).

## 2.11 Use Case Diagram

*Use case diagram* merupakan komponen penting dalam fase spesifikasi kebutuhan dalam pengembangan perangkat lunak. Fase ini penting untuk mendefinisikan fungsionalitas dan interaksi sistem dari perspektif pengguna. Use case diagram membantu dalam memastikan kejelasan, konsistensi, dan kelengkapan dalam proses pengembangan perangkat lunak (Bogumila & Mateusz, 2021).

### **2.12 Activity Diagram**

*Activity diagram* memainkan peran penting dalam fase kebutuhan dalam pengembangan perangkat lunak. Fase ini selalu ada terlepas dari metodologi yang digunakan. *Activity diagram* sebagai alat untuk menggambarkan proses dan alur kerja dalam pengembangan perangkat lunak, dengan fokus khusus pada transformasi dari spesifikasi teks ke representasi grafis. Ini membantu dalam mengklarifikasi dan memformalkan kebutuhan dalam bentuk yang lebih terstruktur dan mudah dipahami (Bogumila & Mateusz, 2021).

### **2.13 Black Box Testing**

Pengujian fungsionalitas dalam suatu *website* sangat penting dilakukan untuk mempermudah pengguna dalam implementasinya. Pengujian tersebut dilakukan dengan menggunakan metode *Black Box Testing*. Pengujian *Black Box* merupakan pengujian yang didasarkan pada detail aplikasi seperti tampilan *website*, fungsi-fungsi yang ada pada *website*, dan kesesuaian alur fungsi dengan bisnis proses yang diinginkan oleh pengguna (Febriyanti et al., 2021).

Metode *Black Box* merupakan metode yang digunakan atau dipakai untuk menguji sebuah aplikasi tanpa harus melihat lebih detail aplikasi tersebut. Proses *Black Box Testing* dilakukan dengan cara mencoba program yang telah dibuat dengan mencoba memasukkan data pada setiap *form*. Pengujian ini diperlukan untuk mengetahui program sudah berjalan dengan yang dibutuhkan oleh *user* (Yulistyanti et al., 2022).