KLASIFIKASI *DATA MINING* UNTUK PENGKATEGORIAN *STATEMENT HARASSMENT BULLYING* MELALUI APLIKASI X DENGAN METODE *DECISION TREE*

SKRIPSI

Diajukan untuk Memenuhi Sebagian Syarat Guna Memperoleh
Gelar Sarjana Matematika dalam Ilmu Matematika

Oleh: Diah Ayu Anggraini

NIM: 1908046004

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI WALISONGO SEMARANG

2023

PERNYATAAN KEASLIAN NASKAH

Yang bertanda tangan di bawah ini:

Nama	: Diah Ayu Anggraini
NIM	: 1908046004
Jurusan/Program Studi	: Matematika

Menyatakan bahwa skripsi yang berjudul:

KLASIFIKASI *DATA MINING* UNTUK PENGKATEGORIAN *STATEMENT HARASSMENT BULLYING* MELALUI APLIKASI X DENGAN METODE *DECISION TREE*

Secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali bagian tertentu yang dirujuk sumbernya.

Semarang, 20 Desember 2023

Saya yang menyatakan,

Diah Ayu Anggraini

NIM. 1908046004

Halaman pengesahan

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI WALISONGO
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Prof. Dr. Hamka Ngaliyan Semarang
Telp. 024-7601295 Fax. 024-7615387

PENGESAHAN

Naskah skripsi berikut :

Judul	: **Klasifikai Data Mining Untuk pengkategorian Statement Harassmen Bullying melalui Aplikasi X dengan Metode Decision Tree**
Peneliti	: Diah Ayu Anggraini
NIM	: 1908046004
Program Studi	: Matematika

Telah diujikan dalam sidang *munaqosyah* oleh Dewan Penguji Fakultas Sains dan Teknologi UIN Walisongo dan dapat diterima sebagai salah satu syarat memperoleh gelar sarjana dalam Ilmu Matematika.

Semarang, 28 Desember 2023

DEWAN PENGUJI

Ketua Sidang,

Dr.Budi Cahyono , M. Si
NIP. 198012152009121003

Sekertaris Sidang,

Eva Khoirun Nisa, M.Si
NIP.198070102201903201O

Penguji Utama I

Any Muanalifah, M. Si, P.hD
NIP.198020113201101200 9

Penguji Utama II

Dinni Rahma Oktaviani, M.Si
NIP.199410092019032017

Pembimbing I

Ariska Kurnia Rachmawati, M. Sc
NIP. 198908112019032019

Pembimbing II

Eva Khoirun Nisa, M. Si
NIP. 198701022019032010

NOTA DINAS

Semarang, 20 Desember 2023

Yth. Ketua Program studi Matematika

Fakultas sains dan Teknologi

UIN Walisongo Semarang

Assalamu'alaikum warahmatullahi wabarakatuh

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan:

Judul	: Klasifikasi *Data Mining* Untuk Pengkategorian *Statement Harassment Bullying* Melalui aplikasi X Dengan Metode *Decision Tree*
Peneliti	: Diah Ayu Anggraini
NIM	: 1908046004
Program Studi	: Matematika

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada fakultas Sains dan Teknologi UIN Walisongo Semarang untuk di ajukan dalam Sidang Munaqosah.

Wasalamu'alaikum warahmatullahi wabarakuh

Pembimbing I

Ariska Kurnia R. M.Sc

NIP. 198908112019032019

NOTA DINAS

Semarang, 20 Desember 2023

Yth. Ketua Program studi Matematika

Fakultas sains dan Teknologi

UIN Walisongo Semarang

Assalamu'alaikum warahmatullahi wabarakatuh

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan:

Judul	: Klasifikasi *Data Mining* Untuk Pengkategorian *Statement Harassment Bullying* Melalui aplikasi X Dengan Metode *Decision Tree*
Peneliti	: Diah Ayu Anggraini
NIM	: 1908046004
Program Studi	: Matematika

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada fakultas Sains dan Teknologi UIN Walisongo Semarang untuk di ajukan dalam Sidang Munaqosah.

Wasalamu'alaikum warahmatullahi wabarakuh

Pembimbing II

Eva Khoirun Nisa. M.Si

NIP. 198701022019032010

ABSTRAK

Judul : **KLASIFIKASI *DATA MINING* UNTUK PENGKATEGORIAN *STATEMENT HARASSMENT BULLYING* MELALUI APLIKASI X DENGAN METODE *DECISION TREE***

Peneliti : Diah Ayu Anggraini

NIM : 1908046004

Bullying merupakan permasalahan yang perlu mendapat perhatian penting dari Masyarakat. *Bullying* melibatkan kebiasaan yang berdampak buruk, mulai dari kerusakan psikologis pada korban hingga kasus bunuh diri. Tujuan dari penelitian ini adalah untuk mengidentifikasi konten yang mengandung makna perundungan online *(cyberbullying)* di media sosial khususnya pada aplikasi X. Dalam hal ini sumber data penelitian ini diperoleh dari media sosial aplikasi X. Setidaknya 1532 data berhasil dikumpulkan. Data tersebut terdiri dari 812 dua jenis tweet yang masing-masing berisi statement negatif dan statement positif. Tujuan penelitian dicapai dengan melakukan beberapa langkah: pengumpulan data, pra-pemrosesan, klasifikasi, dan evaluasi. Dalam studi ini, algoritma yang digunakan dalam penelitian ini adalah *decision tree* dibantu dengan algoritma c4.5 dan *rapidminer*. Dapat disimpulkan dari hasil *accuracy* diperoleh 71,3%, *precision* didapatkan hasil 44,3% dan diperoleh nilai *recall* sebesar 96,2%

Kata kunci: *bullying, Decision Tree, Algoritma C4.5*

KATA PENGANTAR

Bismillahirahmanirrahim

Puji Syukur peneliti panjatkan kehadirat allah SWT yang sudah memberi Rahmat serta karunianya sehingga peneliti dapat menyelesaikan skripsi dengan baik dan tepat waktu. Sholawat serta salam peneliti sanjungkan kepada Nabi Muhammad SAW, Keluarga, sahabat beliau yang telah memberikan pencerahan dan pembelajaran bagi kita semua sehingga kita dapat merasakan nikmat iman dan islam atas kemuliaan ilmu dengan pengharapan semoga dapat memberi syafaat di hari kiamat kelak, Amiin ya Robbal 'alamiin.

Penulis skripsi ini yang berjudul "**KLASIFIKASI *DATA MINING* UNTUK PENGKATEGORIAN *STATEMENT HARASSMENT BULLYING* MELALUI APLIKASI X DENGAN METODE *DECISION TREE*** "bertujuan agar memenuhi persyaratan memperoleh gelar sarjana dalam menyelesaikan Pendidikan pada program studi S.1 Matematika Fakultas Sains dan Teknologi Universitas Islam negeri Walisongo Semarang.

Pada kesempatan kali ini perkenankanlah peneliti untuk mengucapkan rasa terimakasih terhadap semua pihak yang telah membantu dalam penyusunan skripsi ini karena dalam penulisan skripsi ini peneliti menemui halangan dan kesulitan, namun berkat bimbingan dan dukungan dari berbagai pihak, sehinnga penulisan skripsi ini terselesaikan dengan baik. Ucapan berterima kasih ini peneliti sampaikan dengan segala kerendahan hati dan rasa hormat kepada:

1. Prof. Dr. Nizar, M.Ag, selaku Rektor UIN Walisongo Semarang beserta Wakil Rektor I,II, dan III UIN Walisongo Semarang.
2. Dr. Ismail SM, M.Ag, selaku Dekan Fakultas Sains dan Teknologi UIN Walisongo Semarang, beserta Wakil Dekan I,II, dan III Fakultas Sains dan Teknologi UIN Walisongo Semarang.
3. Emy Siswanah, M.Sc, selaku Ketua Prodi Matematika Fakultas Sains dan Teknologi
4. Ahmad Aunur Rohman, M.Pd, selaku Sekertaris Program Studi Matematika Fakultas Sains dan Teknologi UIN Walisongo Semarang.
5. Ariska Kurnia Rachmawati, M.Sc, selaku Dosen Pembimbing I yang telah menyenggangkan waktu untuk memberikan bimbingan, dukungan, memotivasi serta memberi penjelasan sehingga skripsi ini dapat terselesaikan.
6. Eva Khoirun Nisa, M.Si, selaku Dosen Pembimbing II yang telah menyenggangkan waktu untuk memberikan bimbingan, semangat dan motivasi serta memberi penjelasan sehingga skripsi ini dapat terselesaikan.
7. Bapak/Ibu Dosen Serta Staf Fakultas Sains dan Teknologi UIN Walisongo Semarang atas Pelajaran ilmu dan moral serta pelayanan sewaktu masa kuliah hingga penyusunan skripsi ini selesai.
8. Tersepesial kedua orang tua peneliti tercinta bapak Saefudin dan ibu Mutlikah yang senantiasa mendidik dengan Ikhlas, memberi dukungan penuh terhadap

peneliti dan tak Lelah mendoakan pengharapan yang tulus dan baik untuk anaknya.

9. Adik peneliti Cindy Okhtavia yang tanpa henti memberikan semangat dan dukungan untuk meraih dan mewujudkan cita-cita. Serta kedua adik kecil peneliti yang sangat menggemaskan yang selalu membuat semangat dalam pengerjaan skripsi.
10. Keluarga besar peneliti yang selalu memberikan dukungan dan nasihat untuk menyelesaikan skripsi ini, semoga Allah memberikan balsan dalam setiap Langkah kita, Aamiin ya Robbalalamin.
11. Teman-teman Program studi Matematika, terkhusus angkatan 2019 yang berjuang bersama kurang lebih 4 tahun dan selalu memberikan semangat dan motivasi, terutama pada sahabat peneliti Rahyan Elena Mahatiara.
12. Terkhusus untuk calon masa depan saya selama pembuatan skripsi ini Arif Mahmudi yang menjadi support system selama penulisan dan menjadi *moodbooster* terbaik untuk dapat menyelesaikan skripsi ini.
13. Kepada para bujang saya *NCT all member* dan terkhusus untuk haechan, doyoung, mark, jaehyun, jeno, jaemin yang selama pembuatan skripsi selalu menghibur penulis.

Peneliti menyadari bahwa terdapat banyak kekurangan dalam skripsi ini, maka dari itu peneliti sangat membutuhkan kritik dan saran dari pembaca sebagai bahan memperbaikipenulisan di kemudian hari. Kritik dan saran dapat dikirim melalui email anggraini_1908046004@student.walisongo.ac.id.

Semarang, 20 Desember 2023

Diah Ayu Anggraini

NIM. 1908046004

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Bullying (perundungan) merupakan masalah serius yang terjadi di masyarakat, termasuk di Indonesia. Salah satu kasus bullying yang marak menjadi perbincangan belakangan ini yaitu kasus pengurus BEM FMIPA UNY yang berinisial MF. MF mendapatkan *hate comment* hingga *harassment bullying* disebabkan adanya tulisan di salah satu akun *menfes* yang mengatakan bahwa ada pengurus BEM FMIPA melakukan kekerasan seksual terhadap mahasiswa baru. Setelah ditelusuri ternyata bukti yang dikirim pada menfes merupakan *fake chat* dan berita tersebut *hoax*. *Menfes* tersebut dikirim oleh salah satu mahasiswa yang memiliki dendam terhadap MF dikarenakan tidak diterima menjadi pengurus BEM FMIPA UNY. Akibat dari berita *hoax* tersebut mental MF sempat tergoncang karena mendapatkan *hate comment* dari warga net(kumparan.com). Dengan demikian, *bullying* dapat memberikan dampak buruk bagi korbannya, baik secara fisik maupun mental. Pada era digital ini, pembullyan dapat terjadi secara online melalui media sosial, aplikasi *chatting*, dan game online.

Salah satu platform media sosial yang populer saat ini adalah *aplikasi X*, di mana pengguna dapat mengirimkan pesan dalam bentuk cuitan (*tweet*). Menurut laporan dari *Hootsuite* dan *We Are Social* per Januari 2023 didapatkan peningkatan yang masif sebesar 27,4% dari periode sebelumnya dimana terdapat 556 juta pengguna *aplikasi X* di seluruh dunia (Suparto & Habibullah, 2021). Jumlah pengguna *aplikasi X* telah meningkat lebih dari sepuluh kali lipat pada enam tahun terakhir. Publikasi statista menyebutkan bahwa per Januari 2023 didapatkan 24 juta pengguna *aplikasi X* di Indonesia sehingga dinobatkan menjadi negara dengan pengguna *aplikasi X* teraktif kelima di dunia. (Databoks, 2023).

Namun tidak semua pengguna media sosial memanfaatkan teknologi ini dengan baik. Tidak sedikit pengguna menggunakan *aplikasi X* untuk melakukan kegiatan negatif seperti berbuat curang, menyebarkan berita palsu, menulis postingan berisi ujaran kebencian, dan bahkan perundungan secara online (*cyberbullying*). Hal ini tentu merupakan dampak negatif media sosial terhadap masyarakat. Hal-hal negatif seperti ini sangat meresahkan. Dampak yang mungkin terjadi antara lain kerusuhan akibat penyebaran berita palsu, kerugian harta benda akibat penipuan, dan insiden akibat *cyberbullying*. Pemerintah menyatakan 84% remaja Indonesia berusia 12 hingga 17 tahun menjadi korban *bullying*, dan sebagian besar

insiden *bullying* yang ditemukan adalah pembullyian secara *cyber*. *Bullying* di media sosial terjadi dengan menulis *tweet* yang memuat kata-kata ofensif seperti to*o*, go*lok, kampung, bahkan kata-kata yang menghina ras, suku, dan agama(Abdulloh & Hidayatullah, 2021). Didalam agama islam sendiri telah melarang *bullying* (perundungan) dalam bentuk apapun dan di dalam Al-Quran telah menyebutkan larangan hal tersebut pada surat Al-Hujurat ayat 11

يٰٓاَيُّهَا الَّذِيْنَ اٰمَنُوْا لَا يَسْخَرْ قَوْمٌ مِّنْ قَوْمٍ عَسٰٓى اَنْ يَّكُوْنُوْا خَيْرًا مِّنْهُمْ وَلَا نِسَاۤءٌ مِّنْ نِّسَاۤءٍ عَسٰٓى اَنْ يَّكُنَّ خَيْرًا مِّنْهُنَّۚ وَلَا تَلْمِزُوْٓا اَنْفُسَكُمْ وَلَا تَنَابَزُوْا بِالْاَلْقَابِۗ بِئْسَ الِاسْمُ الْفُسُوْقُ بَعْدَ الْاِيْمَانِۚ وَمَنْ لَّمْ يَتُبْ فَاُولٰۤىِٕكَ هُمُ الظّٰلِمُوْنَ

Artinya:*" Wahai orang-orang yang beriman! Janganlah suatu kaum mengolok-olok kaum yang lain (karena) boleh jadi mereka (yang diperolok-olokkan) lebih baik dari mereka (yang mengolok-olok) dan jangan pula perempuan-perempuan (mengolok-olokkan) perempuan lain (karena) boleh jadi perempuan (yang diperolok-olokkan) lebih baik dari perempuan (yang mengolok-olok). Janganlah kamu saling mencela satu sama lain dan janganlah saling memanggil dengan gelar-gelar yang buruk. Seburuk-buruk panggilan adalah (panggilan) yang buruk (fasik) setelah beriman. Dan barang siapa tidak bertobat, maka mereka itulah orang-orang yang zalim."* (Qs. Al-Hujurat 49:11)

Pada surat Al-Hujurat ayat 11 dijelaskan adanya larangan mengkritik dan mengejek sesama umat Islam. Allah SWT mengingatkan umat Islam untuk tidak menebar kebencian dan hinaan terhadap kelompok dan individu umat Islam lainnya. Artinya, Allah SWT telah melarang melakukan tindakan bullying kepada sesama umat manusia.

Berdasarkan latar belakang sebelumnya maka penelitian ini penting dilakukan dalam mencegah dan mengurangi perilaku *bullying*. Salah satu upaya yang dapat dilakukan dalam penelitian ini adalah mendeteksi/mengklasifikasikan tweet yang mengandung makna bullying. Didapatkan data *tweet* pada *aplikasi X* yang sangat besar jumlahnya *(big data)* yang akan digunakan untuk pengklasifikasian tentu memerlukan pengelolaan data dengan baik dan benar. Sehingga, sangat penting untuk mewujudkan pemanfaatan *big data* sebagai sumber pengembangan informasi dan pengetahuan yang didukung dengan teknologi yang terlibat dalam proses pengambilan keputusan. Implementasi *big data* menjadi dasar dalam perumusan kebijakan melalui analisis data meliputi analisis data prediktif, historis, dan sosial.(Rahmanto et al., 2021) . Penggelolaan *big data* dapat dilakukan menggunakan *data mining*. *Data mining* merupakan suatu langkah yang berfungsi mengekstrak pengetahuan dari sekumpulan data yang berjumlah besar (*big data*). *Data mining* juga berfungsi untuk menganalisis

yang meneliti sekumpulan data dalam menemukan hubungan yang tidak terduga dan merangkum data dengan cara yang berguna, dimengeri, dan bervariatif bagi pemilik data.

Salah satu metode *data minin*g yang digunakan dalam penelitian ini adalah *decision tree*. *Decision Tree* adalah pendekatan yang sangat populer dan praktis dalam *machine learning* untuk menyelesaikan masalah klasifikasi. *Decision tree* juga memiliki beberapa algoritma salah satunya algoritma C4.5. Algoritma C4.5 ini merupakan program yang membuat *decision tree* berdasarkan kumpulan data input berlabel(Hozairi et al., 2021). Motode *decision tree* merupakan metode yang mudah diinterpretasikan dan diimplementasikan baik untuk nilai kontinu maupun diskrit. Kelebihan *decision tree* bisa dilihat pada penelitian yang dilakukan oleh (Sang et al., 2021) yaitu hasil penerapan data mining untuk mengklasifikasikan kualitas udara di DKI Jakarta, algoritma *decision tree* memiliki kinerja yang lebih baik dibandingkan dengan algoritma SVM (*support vector machine*) untuk mengklasifikasikan kualitas udara di DKI Jakarta. Penelitian yang dilakukan oleh (Sartika & Sensuse, 2017) dari hasil perbandingan algoritma klasifikasi *nearest neighbour*, *naïve bayes*, dan *decision tree* yang digunakan dalam studi kasus pengambilan keputusan pemilihan pola pakaian menunjukkan bahwa algoritma klasifikasi *decision tree* adalah algoritma klasifikasi yang paling akurat

dibandingkan algoritma klasifikasi *naïve bayes* dan *nearest neighbour*. Selanjutnya, pada penelitian (Alamsyah et al., 2023) metode *decision tree* memiliki tingkat akurasi yang tinggi dibandingkan algoritma K-NN (*k-nearest neighbour)* untuk klasifikasi kebakaran hutan diwilayah aljazair.

Berdasarkan latar belakang, penelitian ini mengklasifikasi data untuk mendeteksi *bullying* pada media sosial *aplikasi X* menjadi dua kelas yaitu negatif dan positif. Kelas negatif memuat *tweet* berupa kata-kata yang mengandung unsur *bullying*, dan kelas positif memuat *tweet* berupa kata-kata yang tidak mengandung unsur *bullying*. Pengembangan penelitian ini diharapkan dapat membantu orang tua, pemerintah, dan negara dalam melindungi generasi muda dari pembullyan secara online dan mengurangi jumlah pelaku bullying.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas didapatkan rumusan masalah sebagai berikut:

1. Bagaimana analisis pengkategorian *statement harassment bullying* melalui aplikasi X dengan metode *Decision tree*?
2. Bagaimana tingkat akurasi yang dimiliki metode *decision Tree* untuk menganalisis pengkategorian *statement harassment* bullying melalui *aplikasi X*?

1.3 Tujuan Penelitian

Dari rumusan masalah diatas, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mendeteksi atau mengidentifikasi pernyataan pernyataan apa yang masuk dalam kategori *statement harassment bullying*
2. Untuk mendapatkan hasil klasifikasi yang akurat dari metode *Decision Tree*

1.4 Manfaat Penelitian

Dari tujuan penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Tersedianya data tentang kata atau gabungan kata yang paling berpotensi digunakan untuk melakukan bullying di media social khususnya *aplikasi X*.
2. Data tersebut dapat dimanfaatkan oleh orang tua, pemerintah, dan negara dalam melindungi generasi muda dari pembullyan secara online dan mengurangi jumlah pelaku bullying.

1.5 Batasan Masalah

Penelitian ini memiliki Batasan masalah sebagai berikut:

1. Kata kunci yang digunakan dalam penelitian ini adalah anjing, bodoh dan jelek. Karena pada penelitian (Trihapsari et al., 2016) 3 kata kunci tersebut sudah dapat mewakili sebagai kata yang banyak digunakan dalam *bullying*
2. Data yang digunakan adalah data aplikasi *X* tahun 2019- 2023 di Indonesia
3. Metode yang digunakan adalah *Decision Tree*
4. *tool* yang digunakan adalah *Python* dan *Rapidminer*
5. Variabel yang digunakan yaitu pesan yang dikirim melalui *twitter* yang terdapat unsur *negatif dan positif.*

BAB II

KAJIAN PUSTAKA

2.1 *Data Mining*

Data mining merupakan bidang keilmuan yang relatif baru dengan beragam penerapan, menjadikannya salah satu dari sepuluh bidang keilmuan yang memiliki dampak terbesar terhadap perkembangan teknologi. *Data mining* juga merupakan teknik untuk mengekstraksi atau "menambang" pengetahuan dari sejumlah besar data. *Data mining* tidak hanya menggali pengetahuan saja, tetapi dapat menganalisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak terduga dan merangkum data dengan cara baru yang dapat dipahami dan berguna bagi pemilik data.(Widaningsih, 2019)

Data Mining merupakan proses mengekstraksi dan mengidentifikasi informasi bermanfaat dan pengetahuan yang relevan dari *database* besar menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning*. Secara umum, data mining dapat dikategorikan menjadi dua, yaitu:

a. *deskriptif,* berarti data mining yang mengacu pada penggunaan data untuk

mencari pola yang dapat dipahami manusia yang menggambarkan karakteristik data.

b. *Prediktif*, berarti menggunakan *data mining* untuk membentuk sebuah model pengetahuan yang akan digunakan untuk membuat prediksi.

Mengenai berbagai kegunaan data mining, tugas data mining dapat diklasifikasikan menjadi enam kelompok berdasarkan fungsinya, yaitu:

a. Klasifikasi *(classification).*Menggeneralisasi struktur yang diketahui dan di terapkan pada data baru. Misalnya, mengklasifikasikan penyakit ke dalam berbagai jenis atau mengklasifikasikan email sebagai spam atau bukan.
b. Klasterisasi *(clustering).* Mengelompokkan data dari kelas yang tidak diketahui ke dalam sejumlah kelompok tertentu menurut ukuran kemiripannya.
c. Regresi *(regression).* Menemukan suatu fungsi yang memodelkan data dengan galat (kesalahan prediksi) seminimal mungkin.
d. Deteksi anomial *(anomialy detection).* Mengidentifikasi data yang tidak umum, bisa berupa outlier, perubahan atau deviasi yang mungkin sangat penting dan perlu investigasi lebih lanjut.
e. Pembelajaran aturan asosiasi *(asosiasi rule learning).* Atau pemodelan kebergantungan

(dependency modeling) yaitu mencari relasi antar variabel.

f. Perangkuman (*summarization*).
Menyediakan representasi data yang lebih sederhana, meliputi visualisasi dan pembuatan laporan (Burch dan Grudnitski dalam (Fauzi, 2019)

Data mining merupakan salah satu dari rangkaian *Knowledge Discovery in Dattabase (KDD)*. KDD berhubungan dengan Teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola sejumlah data. langkah-langkah dalam serangkaian proses tersebut memiliki tahapan sebagai berikut:

a) Pembersihan data untuk membuang data yang tidak konsisten dan noise
b) Proses Penambangan *(process mining)* merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
c) Integrasi data adalah penggabungan data dari beberapa sumber
d) Transformasi Data (*Data Transformation)* data diubah menjadi bentuk yang sesuai untuk dimining
e) Aplikasi Teknik data mining, proses ekstraksi pola dari data yang ada

f) Evaluasi pola *(pattern evaluation).* Proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan(Muhidin et al., 2022)

2.2 Data

Data adalah kumpulan informasi atau pernyataan tentang sesuatu yang diperoleh dengan mempelajari atau mengamati sumber tertentu. Data yang diperoleh bisa menjadi fakta atau argumen karena belum diproses lebih lanjut.(Ibeng, 2020)

Algoritma umumnya memiliki dua data diantaranya adalah:

a) *Data training*

Training set digunakan oleh algoritma klassifikasi untuk membentuk sebuah model *classifier*. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada.

b) *Data Testing*

Testing set digunakan untuk mengukur sejauh mana classifier berhasil melakukan klasifikasi dengan benar. Karena itu, data yang ada pada *testing set* seharusnya tidak boleh ada pada *training set* sehingga dapat diketahui apakah *model classifier* sudah "pintar" dalam melakukan klasifikasi. Lain

lagi halnya dengan validation set. (Makasudede, 1953)

2.3 Klasifikasi

Klasifikasi adalah salah satu metode pembelajaran paling umum dalam data mining. Klasifikasi diartikan sebagai suatu bentuk analisis data untuk mengekstrak model yang digunakan untuk memprediksi label kelas. Dalam klasifikasi, kelas adalah atribut paling unik dalam suatu kumpulan data dan mewakili variabel bebas dalam statistik. Klasifikasi data terdiri dari dua proses yaitu tahap pembelajaran dan tahap klasifikasi. Tahap pembelajaran merupakan tahap terbentuknya model klasifikasi, dan tahap klasifikasi merupakan tahap dimana model klasifikasi digunakan untuk memprediksi label kelas dari data.(Sartika & Sensuse, 2017)

Klasifikasi merupakan metode *supervised learning* yang membutuhkan data training berlabel untuk menghasilkan sebuah aturan yang mengklasifikasikan data uji ke dalam kelompok atau kelas yang telah ditentukan. Beberapa teknik klasifikasi yang digunakan adalah *decision tree, rule-based classifier, neural network, support machine dan naïve Bayes classifier*. Setiap teknik menggunakan algoritme pembelajaran untuk mengidentifikasi model yang memberikan hubungan yang paling sesuai antara

himpunan atribut dan label kelas dari data input.(Setio et al., 2020)

2.4 *Decision Tree*

Decision tree atau dikenal dengan pohon keputusan adalah salah satu metode klasifikasi yang menggunkan representasi suatu struktur pohon yang yang berisi alternatif-alternatif untuk pemecahan suatu masalah. *Decision tree* ini juga menunjukkan faktor-faktor yang mempengaruhi hasil alternatif dari keputusan tersebut disertai dengan estimasi hasil akhir bila kita mengambil keputusan tersebut. Peranan *decision tree* ini adalah sebagai *Decision Support System* untuk membantu manusia dalam mengambil suatu keputusan. Manfaat dari *decision tree* adalah melakukan proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga orang yang mengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Konsep yang digunakan oleh *decision tree* adalah mengubah data menjadi suatu keputusan pohon dan aturan-aturan keputusan*(rule*). (Nilai et al., 2015)

Decision tree menggunakan struktur hierarki untuk pembelajaran *supervised*. Proses dari *decision tree* dimulai dari *root node* hingga *leaf node* yang dilakukan secara rekursif. Di mana setiap percabangan menyatakan suatu kondisi yang harus dipenuhi dan pada setiap ujung pohon menyatakan

kelas dari suatu data. Pada *decision tree* terdiri dari tiga bagian yaitu:

a. *Root node*
   Node ini merupakan node yang terletak paling atas dari suatu pohon atau akar pohon.
b. *Internal node*
   Node ini merupakan *node* percabangan, hanya terdapat satu input serta mempunyai minimal dua output.
c. *Leaf node*
   Node ini merupakan *node* akhir, hanya memiliki satu input, dan tidak memiliki output.

Kelebihan dari metode *Decesion tree* (pohon keputusan) adalah:

1) Daerah pengambilan keputusan yang sebelumnya kompleks dan sangat *Global* dapat diubah menjadi lebih simpel dan spesifik.
2) Eliminasi perhitungan-perhitungan yang tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka sampel diuji hanya berdasarkan kriteria atau kelas tertentu.
3) Fleksibel untuk memilih fitur dari *internal node* yang berbeda, fitur yang terpilih akan membedakan suatu kriteria dibandingkan kriteria yang lain dalam *node* yang sama.

kefleksibelan metode pohon keputusan ini meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode perhitungan satu tahap yang lebih konvensional.

4) Dalam analisis multivariat dengan kriteria dan kelas yang jumlahnya sangat banyak, seorang penguji biasanya perlu untuk mengestimasikan baik itu di deskripsi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut. metode pohon keputusan dapat menghindari munculnya permasalahan ini dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap motif internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan.

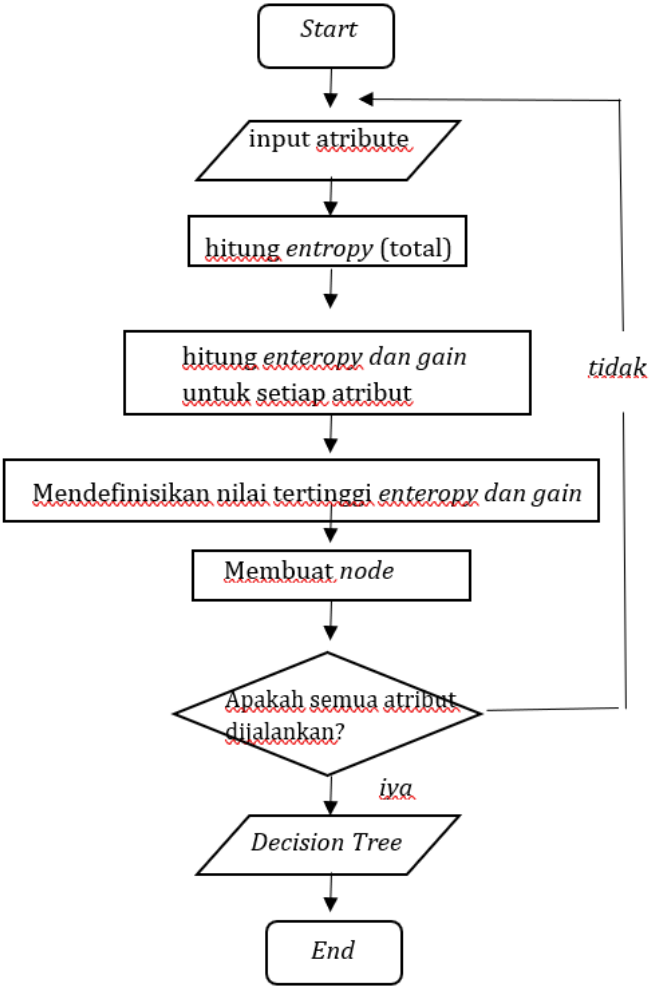
Model *Decision Tree* dibentuk menyerupai struktur *flowchart*, dimana setiap simpul yang bukan simpul daun merupakan atribut pengujian, setiap cabang mewakili output dari pengujian, dan setiap simpul daun (terminal *node*) menentukan *label class*. Simpul paling atas dari sebuah pohon adalah node akar. *Decision Tree* membedakan antara suatu kelas dengan kelas lainnya berdasarkan tingkat kemurnian kelas *(impurity)* pada suatu simpul. Alat ukur kemurnian kelas *(impurity)* yang umum digunakan adalah *GINI Index, Entropy, Misclassification measure, Chi-square measure, G-square measure*. Beberapa

algoritma yang termasuk kategori *Decission Tree* antara lain *ID3, C4.5, C5.0, CART, CHAID, dan lain-lain* (Gorunescu, 2011).

Pohon dibangun dengan memisahkan data secara rekursif sehingga setiap bagian terdiri dari kelas data yang sama. Format pemisahan yang digunakan untuk membagi data tergantung pada jenis atribut yang digunakan dalam pemisahan. Algoritma *Decision tree* dapat menangani data numerik *(kontinu)* dan diskrit. *Algoritma decision tree* menggunakan *rasio gain*.(Setio et al., 2020)

2.5 *Algoritma C4.5*

Pada klasifikasi perbandingan *algoritma C4.5* dan *algoritma Id3* untuk memprediksi kelulusan didapatkan hasil *algoritma C4.5* memprediksi lebih baik daripada *algoritma id3*.(Faizah & Jananto, 2021) Oleh karena itu, untuk pengklasifikasian ini menggunakan algoritma c4.5. *Algoritma c4.5* merupakan pengembangan dari *algoritma ID3*. *Algoritma C4.5* dan *ID3* diciptakan oleh seorang peneliti dibidang kecerdasan buatan bernama *j. Rose quinlan* pada akhir tahun 1970-an. *Algoritma C4.5* membuat pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun. Gambar 2.1 menunjukkan gambaran umum proses untuk membangun sebuah pohon keputusan.



Gambar 2. 1 Gambaran Umum Algoritma C4.5

Secara umum, *algoritma C4.5* untuk membangun sebuah pohon keputusan adalah sebagai berikut:

a) Menginput atribute.
b) Menghitung *entrophy* (total) pada semua atribute.
c) Menghitung setiap *entrophy* dan *gain* untuk semua atribute.
d) Mendefinisikan nilai tertinggi *entrophy* dan *gain*.
e) Membuat *node.* Jika ada anggota node yang memiliki nilai entrophy lebih besar dari nol, ulangi lagi proses dari awal dengan *node* sebagai syarat sampai semua anggota dari *node* bernilai nol.

Node adalah atribut yang mempunyai nilai *gain* tertinggi dari atribut-aribut yang ada. Rumus untuk menghitung *gain* adalah sebagai berikut:

$$Gain\ (S, A) = Entrophy(S) - \sum_{i=1}^{n} \frac{|S_1|}{|S|} * Entrophy(S_1) \quad (1)$$

Keterangan:

S : Himpunan Kasus

A : Atribut

n : Jumlah Partisi Atribut

$|S_1|$: Jumlah kasus pada partisi ke – i

$|S|$: Jumlah Kasus dalam S

Sementara itu, untuk menghitung nilai *entrophy* dapat ditulis sebagai berikut:

$$Entrophy\ (S) = \sum_{i=1}^{n}(-pi) * \log_2 pi \quad (2)$$

Keterangan:

S : Himpunan Kasus

n : Banyaknya Partisi

Pi : Probabolitas yang didapat dari kasus i di bagi total kasus

Entrophy merupakan jumlah bit yang dibutuhkan untuk menyatakan suatu kelas. Semakin kecil nilai *entrophy*, semakin baik digunakan dalam mengekstraksi suatu kelas. Setelah memeroleh nilai *entrophy*, Proses perhitungan *gain* dilakukan secara berulang-ulang hingga semua data masuk ke dalam kelas yang sama. Atribut yang telah dipilih tidak diinputkan lagi pada perhitungan nilai *gain.*

2.6 *Tool Data Mining*

Tool data mining adalah *software* yang digunakan untuk mempermudah seorang peneliti, akademis, dan pihak manapun dalam hal mengolah dan menambang sebuah data, Selain itu juga bisa dijadikan sebuah *library* sehingga bisa ditanam dalam sebuah program, sehingga program bisa melakukan apa yang bisa *tool* lakukan.

Python adalah bahasa pemrograman yang berguna pada banyak platform, dengan keuntungan tidak memerlukan *compiler* atau *interpreter*, sehingga menghemat banyak waktu dalam pengembangan program. *Interpreter* dapat digunakan secara interaktif, sehingga memudahkan pengembang aplikasi untuk mengevaluasi fungsionalitas program dari awal, bereksperimen dengan fitur bahasa, dan merancang program sekali pakai.(Handayani, 2023)

Rapidminer adalah salah satu *software* untuk pengolahan data mining. Pekerjaan yang dilakukan oleh *rapidminer text mining* adalah berkisar dengan analisis teks, mengekstrak pola-pola dari dataset yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database.(Faid et al., 2019)

2.7 *Confusion matrix*

Confusion Matrix merupakan pengukuran performa buat permasalahan klasifikasi *machine learning* dimana luaran bisa berbentuk 2 kelas ataupun lebih. *Confusion Matrix* merupakan tabel dengan 4 campuran berbeda dari nilai prediksi serta nilai aktual. Berdasarkan Tabel.1 dapat dijelaskan bahwa terdapat 4 sebutan representasi hasil proses klasifikasi pada *confusion matrix* ialah *True Positif (TP), True Negatif (TN), False Positif (FP), serta False Negatif (FN)*. *Confusion matrix* pula kerap disebut *error matrix*. Pada dasarnya confusion matrix membagikan data

perbandingan hasil klasifikasi yang dicoba oleh sistem dengan hasil klasifikasi sesungguhnya.(Hozairi et al., 2021)

Table 1. Pengujian Confusionan Matrix

Clasification		*Predicted class*	
		TRUE	*FALSE*
Actual: True		*True Positif (TP)*	*False Negatif (FN)*
Actual: False		*False Positif (FP)*	*True Negatif (TN)*

Tolak ukur acuan hasil perhitungan metode *confusion matrix* yaitu *Accuracy, Precision* dan *recall.*

1. *Accuracy* dapat didefinisikan selaku jenjang korelasi antara nilai prediksi dengan nilai aktual. Akurasi merupakan representasi keakuratan model dalam melakukan *clustering* dengan benar.

$$Accurary = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

2. *Precision* merupakan persentase keakuratan hasil evaluasi dengan metode yang digunakan. *Precision* adalah tingkat ketepatan antara data yang diharapkan oleh pengguna dengan jawaban yang diberikan oleh sistem.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

3. *Recall* merupakan gambaran kesesuaian suatu metode dalam pengambilan informasi. *Recall* adalah tingkat keberhasilan sistem dalam menciptakan kembali suatu data.(Amaliah & Dwi Nuryana, 2022)

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

2.8 *Bullying*

Kata "*bully*" dikenal sejak tahun 1530-an. Pada dasarnya *bullying* melibatkan dua orang, pelaku *bullying* dan korban. Pelaku mem-*bully* korban secara fisik, lisan, atau cara lain untuk mendapatkan rasa kekuasaan. Tindakan ini mungkin langsung (memukul, mencela, dan lain-lain) atau secara tidak langsung (gosip, rumor, fitnah, dan lain-lain). Bullying, yakni tindakan merendahkan kemanusiaan manusiayang tercermin dalam sikap dan perilaku. Perilaku bully kerap kali lahir karena hilangnya empati dan nurani. Karena melihat orang lain yang berbeda dengan si pembully, maka merasa yang berbeda dan minoritas layak di-*bully*, layak dihakimi. Sesungguhnya tindakan membully menunjukkan rendahnya sikap mental dan karakter dari pelakunya.(Trihapsari et al., 2016)

Bullying merupakan suatu tindakan agresif yang mengganggu kenyamanan dan menyakiti orang lain dengan adanya perbedaan kekuatan

maupun psikis dari korban dan pelaku yang dilakukan secara berulang. Berdasarkan medianya *bullying* dibedakan menjadi dua, yakni traditional *bullying* dan *cyber bullying*. Traditional bullying terjadi dengan kontak secara langsung antara korban dan pelaku. Sedangkan, cyber bullying terjadi melalui perantaraan media sosial dan korban dilecehkan atau dianiaya melalui media social. (Hidajat et al., n.d.)

Cyberbullying adalah segala bentuk kekerasan yang dialami anak atau remaja dan dilakukan teman seusia mereka melalui dunia cyber atau internet. *Cyberbullying* adalah kejadian manakala seorang anak atau remaja diejek, dihina, diintimidasi, atau dipermalukan oleh anak atau remaja lain melalui media internet, teknologi digital atau telepon seluler. Berbeda dengan bullying tradisional, karena *Cyberbullying* terjadi 24 jam/ hari, 7 hari/ minggu, dan mencapai korbannya dimanapun dia berada termasuk di rumah(Anwar, 2017). *Cyberbullying* memiliki banyak bentuk, antara lain:

a) Pelecehan/ provokasi emosi (*harassment/ trolling*), adalah mengirimkan pesan bersifat mengancam atau menyerang, berbagi foto atau video aib/vulgar, atau memposting pesan yang mengancam atau memancing amarah pada situs jejaring sosial.

b) Fitnah (*denigration*), adalah informasi palsu, salah,berupa gosip yang menyebar.
c) Penyulut kemarahan (*flaming*), menggunakan bahasa ekstrim untuk memancing perkelahian.
d) Mencuri identitas seseorang atau membajak situs seseorang (*hacking*).
e) Pengecualian (*exclusion*), meninggalkan seseorang secara sengaja.
f) Mengirimkan gambar atau memaksa seseorang untuk mengirim gambar seksual.

2.9 Penelitian Sebelumnya

Peneliti melakukan telaah terhadap beberapa penelitian, terdapat beberapa penelitian yang memiliki keterkaitan dengan penelitian yang akan peneliti lakukan. Baik keterkaitan pada metode maupun pada subyek penelitian. Berikut merupakan beberapa keterkaitan penelitian terdahulu.

Penelitian sebelumnya yang pertama**,** Penelitian mengenai klasifikasi *cyber bullying* pada media sosial *twitter* dengan menggunakan algoritma naïve bayes oleh (Trihapsari et al., 2016) Tujuan dari penelitian ini adalah menemukan kata atau gabungan kata yang paling berpotensi digunakan untuk melakukan cyberbullying pada media sosial khususnya Twitter, sehingga mempermudah pengembang untuk membuat sistem atau perangkat lunak yang dapat melakukan

deteksi dini terhadap terjadinya cyberbullying di media social. Penelitian sebelumnya yang kedua**,**Penelitian ini membahas Analisa sentimen *cyberbullying* di jejaring sosial twitter dengan algoritma *naïve bayes* oleh (Maulana et al., 2020) tujuan dari penelitian ini yaitu untuk mengklasifikasikan tweet positive dan negative. Pada penelitian sebelumnya tersebut memiliki kesamaan dengan peneliti yaitu sama membahas tentang klasifikasi sentiment cyberbullying. Hal yang berbeda dari penelitian tersebut dengan peneliti yaitu metode yang digunakan karena ada banyak metode yang dapat digunakan salah satunya decision tree.

Selanjutnya Ketiga**,** Penelitian tentang Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi Balita Di Kabupaten Simalungun oleh (Hafizan & Putri, 2020) tujuan dari penelitian ini adalah akan menguji coba model untuk klasifikasi status gizi balita berdasarkan baku rujukan WHO-2005 dengan memanfaatkan teknik data mining menggunakan metode *Decision Tree Algoritma C4.5.* penelitian keempat tentang Klasifikasi Opini Pengguna Media Sosial Twitter Terhadap JNT Di Indonesia dengan Algoritma Decision Tree oleh (Handoko et al., 2022) tujuan dari penelitian ini adalah Untuk dapat mengelompokkan opini yang banyak, dibutuhkan program machine learning yang dapat mempermudah proses pengelompokan opini tersebut. Terdapat banyak algoritma yang

dapat digunakan untuk mengelompokkan opini, salah satunya adalah Decision Tree. Kelima, Penelitian tentang Analisa tanggapan terhadap PSBB di indonesia dengan algoritma decision tree pada twitter oleh (Aditya Quantano Surbakti et al., 2021) tujuannya untuk mendapatkan keakurasian yang baik dari 3 metode yaitu *Decision Tree, Naïve Bayes Classifier dan K-Nearest Neighbors (K-NN).* Pada penelitian ketiga diatas yang membedakan penelitian ini dengan peneliti yaitu data dan juga metode yang digunakan. Peneliti hanya menggunakan satu metode untuk mengkelompokan opini dan mendapatkan keakurasian.

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif untuk mendapatkan pemahaman tentang bagaimana suatu konstruksi dapat dibangun dan bekerja. Penelitian kuantitatif merupakan penelitian yang menggunakan populasi atau sampel tertentu dengan mengumpulkan data dan mengolah data yang bersifat statistik.(Oktavia, 2023)

Penelitian ini bertujuan untuk mengetahui tingkat akurasi yang dihasilkan dalam metode algoritma *decision tree*. Sifat yang digunakan dalam penelitian ini adalah sifat deskriptif karena akan menjelaskan secara detail akibat dari hasil akurasi.

3.2 Sumber data Penelitian

Sumber data yang diambil dalam penelitian bersifat data sekunder karena dalam penelitian menggunakan sumber data yang berasal dari *sentiment* pada *aplikasi X* dengan beberapa kata kunci yang merupakan unsur *bullying*(Hijriana & Rasyidan, 2017). Untuk mendapatkan dataset peneliti memanfaatkan *tool Python* yang telah terhubung dengan *aplikasi X* untuk mengambil data dari tahun 2019 hingga tahun 2023.

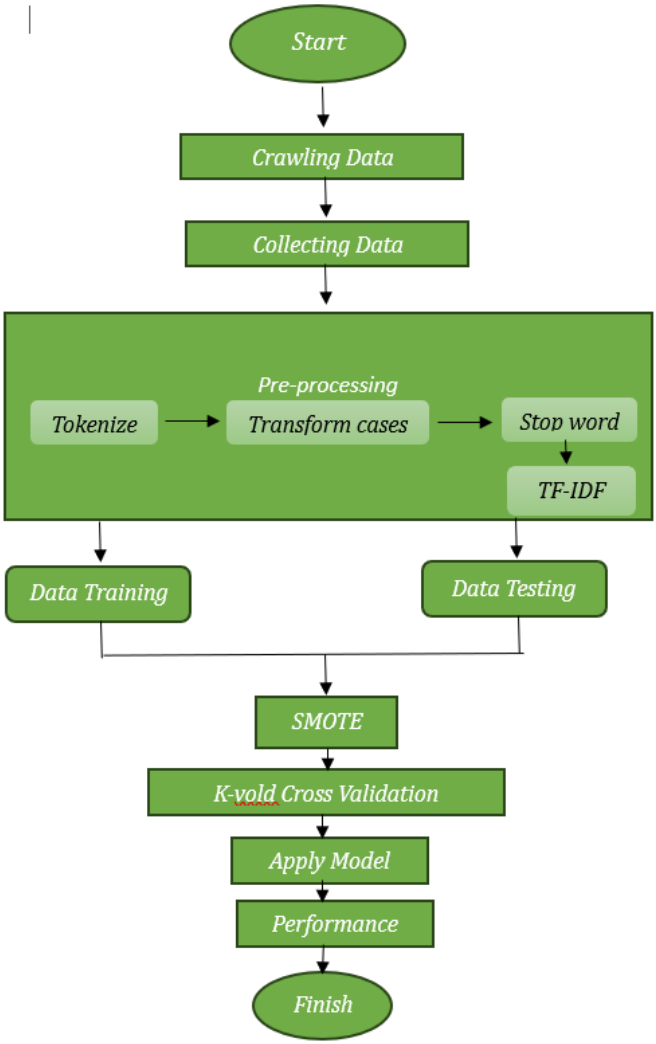
3.3 Variabel penelitian

Pendekatan *data mining* yang dilakukan dalam penelitian ini adalah teknik klasifikasi. Teknik klasifikasi adalah suatu teknik menggolongkan data ke dalam beberapa kategori atau label.(Alamsyah et al., 2023) Penggolongan tersebut dalam penelitian ini dibagi menjadi dua kelas yaitu:

a) *sentimen positif* yaitu *sentiment* yang memiliki sifat membangun dan tidak memiliki unsur kebencian
b) *sentimen negatif* yaitu *sentiment* yang memiliki unsur kebencian, mengkritik dan bersifat menjatuhkan

3.4 Teknik analisis data

Tahapan-tahapan dalam penelitian yang terdapat dalam metode klasifikasi algoritma *data mining* memiliki sifat interaktif, sifat interaktif merupakan sifat yang melibatkan langsung pemakai atau sama dengan sebagai perantara *knowledge base* (Gullo, 2015). Data yang dikumpulkan sebagai hasil dari penggolongan yang digunakan dalam penelitian. Penelitian akan menggunakan proses yang terstruktur dimulai dari input data hingga output data. Gambar 3.1 menunjukkan gambaran umum proses penelitian.



Gambar 3. 1 Alur Penelitian

Penjelasan mengenai teknis analisis diatas sebagai berikut:

1. Crawling data

Istilah "*crawling data*" mengacu pada proses pengumpulan data di aplikasi X. Tahap ini bekerja dengan menggunakan bantuan "*search X*" menggunakan *tool Python* untuk mengunduh data berbentuk tweet atau user dari *server aplikasi X*. Pada penelitian ini, kata kunci: Anjing,jelek dan bodoh digunakan untuk *crawling data*.

2. *Collecting* data

Penelitian ini menggunakan *data dari aplikasi X*, sebanyak 1532 tweet (data awal). Setelah mendapatkan data tersebut maka tahap selanjutnya *collecting* data. Tahapan ini bertujuan agar sistem komputer lebih mengenali bentuk *data set*. Pelabelan kelas sentimen atau opini pada data dikelompokkan menjadi dua kelas sentimen yaitu kelas positifve dan kelas negative untuk memudahkan dalam mengelola data dengan membaca maksud dari kalimat yang ada dalam data tersebut sehingga dapat diberi label sentimen bahwa tweet tersebut termasuk ke dalam sentimen positive atau negative. kelas negative menunjukkan komentar yang mengandung elemen bullying. Kelas positif menunjukkan komentar yang mengandung

motivasi atau dukungan yang tidak mengandung elemen bullying.

3. Pre-Prosesing data

Setelah proses pengumpulan data selanjutnya akan dilakukan *preprocessing* data. *Preprocessing* merupakan langkah untuk mengubah data mentah menjadi data atau format yang sesuai untuk tahap analisis berikutnya. *Preprocessing* meliputi penghapusan karakter khusus, *transform cases, stopword dan tokenisasi*. Tujuan dilakukan *preprocessing* data adalah untuk menyaring data penting dari kumpulan data.

1. *Tokenize*
   Tokenisasi adalah proses pemotongan seluruh urutan karakter menjadi satu potongan kata. Proses ini untuk membagi teks yang dapat berupa kalimat, paragraf atau dokumen, menjadi token-token.
2. *Transform cases*
   Transform Cases proses ini berfungsi menyelaraskan semua teks menjadi huruf kecil karena untuk menjauhi adanya masalah pada saat *stopword*.
3. *Stopword*
   Stopword adalah kata umum (*common words*) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna.
4. *Weighting* (Pembobotan)

TF-IDF adalah salah satu metode pembobotan sebuah kata didalam sistem pencarian informasi. TF-IDF metode yang menggambungkan *Term Frequency (TF)* dan *Inverse Document Frequency (IDF*). *Term Frequency* adalah kemunculuan frekuensi munculnya kata sama pada dokumen. *Inverse Document Frequency* banyaknya koleksi dokumen yang bersangkutan mengandung kata tertentu.(Anam et al., 2021)

4. SMOTE

Proses SMOTE merupakan tahap pengujian data dari hasil *preprocessing* dengan melakukan *balancing* terhadap label *sentiment* dengan menggunakan teknik *SMOTE (Synthetic Minority Over-sampling Technique).* Proses SMOTE berfungsi untuk mendapatkan data yang baik dengan proses mengimbangi kelas yang diakibatkan oleh *overfitting*.(Dessy Angelina et al., 2023)

5. K-Vold Cross Validation

K-vold cross validation adalah teknik untuk mengevaluasi/memvalidasi keakuratan model yang dibangun pada dataset tertentu. *Cross validation* juga menjalankan pengujian standar untuk memprediksi *eror rate. K-vold cross validation* merupakan metode validasi silang

yang digunakan untuk menghitung akurasi prediksi suatu sistem. Metode ini banyak digunakan oleh para peneliti karena dapat mempersingkat waktu perhitungan dan menjaga keakuratan perkiraan. Pada *k-vold cross validation*, data (D) dibagi menjadi k *subsets* D1, D2, ..., Dk yang mempunyai jumlah yang sama. Proses ini diulangi sebanyak k *subsets* dan hasil akurasi klasifikasi merupakan hasil rata-rata untuk setiap data *training* dan *testing*. K-vold yang umum digunakan adalah 3, 5, 10 dan 20. (Setio et al., 2020)

6. Apply Model

Untuk menerapkan model yang telah dilatih sebelumnya menggunakan data training pada *unlabeled* data (*data testing*). Tujuannya adalah untuk mendapatkan prediksi pada *data testing* yang belum memiliki label. Yang perlu diperhatikan adalah *data testing* harus memiliki urutan, jenis, maupun peran atribut yang sama dengan *data training*.

Model algoritma yang digunakan adalah *decision tree. Decision Tree* merupakan salah satu metode klasifikasi yang menggunkan representasi suatu struktur pohon yang yang berisi alternatif-alternatif untuk pemecahan suatu masalah.

7. Performance

Digunakan untuk mengevaluasi kinerja model yang memberikan daftar nilai kriteria kinerja secara otomatis sesuai dengan tugas yang diberikan. Misalkan untuk klasifikasi, kriteria yang diberikan adalah *accuracy, precision dan recall.*(Aditya Quantano Surbakti et al., 2021)

BAB IV

HASIL DAN PEMBAHASAN

4.1 Sumber Data

Sumber data dalam penelitian ini adalah data sekunder yang didapat melalui *sentiment-sentiment* masyarakat dengan kata-kata mengandung unsur *bullying* yang diambil dalam bentuk teks yang merupakan tweet dari pengguna aplikasi X. Data yang diambil dalam periode waktu 2019-2023, batasan dalam pencarian kata kata unsur *bullying* ini adalah dengan menggunakan bahasa Indonesia. Penelitian ini diperkuat dengan adanya jurnal terdahulu dan buku-buku pendukung.

Gambar 4. 1 Tampilan Pencarian Twitter dengan Kata Kunci Bodoh

4.2 *Crawling Data*

Tahapan setelah mempunyai sumber data adalah mengumpulkan data. *Crawling data* merupakan proses pencarian data. Proses *crawling data* dilakukan dengan bantuan *tool Python*. Data yang akan diambil dari aplikasi X. pada proses pengambilan data ini peneliti menggunakan kata bodoh, anjing dan jelek, *crawling* data dilakukan secara bertahap dan masing masing kata peneliti hanya mengambil 500 data.

Pertama, peneliti akan *crawling data* dengan kata "Bodoh". Berikut ini merupakan proses *crawling data* menggunakan *python*. *Script* yang digunakan untuk mengumpulkan data sebagai berikut:

Script python: (Helmi satria,2023)

#menginstalkan package pada pyhton

```
# Import required Python package
!pip install pandas

# Install Node.js (because tweet-harvest built
using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl
gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl                                    -fsSL
https://deb.nodesource.com/gpgkey/nodesource-
repo.gpg.key | sudo gpg --dearmor -o
/etc/apt/keyrings/nodesource.gpg

!NODE_MAJOR=20 && echo "deb [signed-
by=/etc/apt/keyrings/nodesource.gpg]
```

```
https://deb.nodesource.com/node_$NODE_MAJOR.x
nodistro main" | sudo tee
/etc/apt/sources.list.d/nodesource.list

!sudo apt-get update
!sudo apt-get install nodejs -y

!node -v
#proses mengumpulkan data dengan kata
kunci “bodoh”
# Crawl Data

filename = 'bodoh.csv'
search_keyword = 'bodoh until:2023 since:2019'
limit = 100

!npx --yes tweet-harvest@latest -o "{filename}"
-s "{search_keyword}" -l {limit} --token ""
import pandas as pd

#setelah didapatkan datanya maka data
akan disimpan dalam bentuk file csv

# Specify the path to your CSV file
file_path = f"tweets-data/{filename}"

# Read the CSV file into a pandas DataFrame
df = pd.read_csv(file_path, delimiter=";")

# Display the DataFrame
display(df)
```

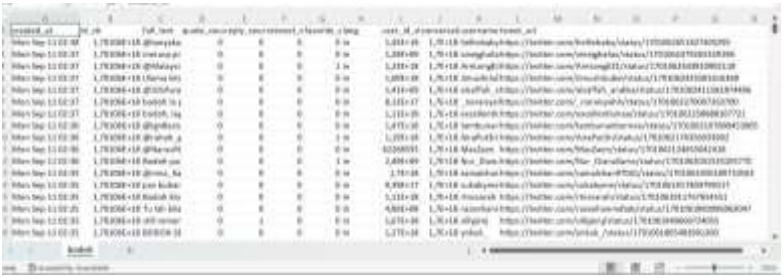
Proses pengumpulan data dengan kata kunci “anjing dan jelek” sama seperti di atas. Hanya saja pada tahap *crawling data* kata yang dimasukan sesuai

dengan kata kunci yang akan dicari. *Script* proses *crawling data* dengan kata kunci "anjing dan jelek" terdapat pada lampiran 5 dan 6.

Crawling data yang diambil dalam penelitian berbentuk teks yang merupakan tweet dari pengguna twitter, hasil yang didapat dari *crawling data* berjumlah 1532 tweet. Tahap *crawling data* menghasilkan 12 atribut yang berisi:

a. *Created_at* yaitu waktu yang dihasilkan saat pengguna mengunggah tweet.
b. *Id_str* yaitu merupakan nomor id twitter.
c. *Full_text* yaitu isi tweet yang diunggah
d. *Quote_count* yaitu jumlah kutipan pada tweet tersebut.
e. *Reply_count* yaitu jumlah reply pada tweet tersebut.
f. *Retweet_count* adalah jumlah retweet pada tweet tersebut.
g. *Favorite_count* adalah jumlah favorite pada tweet tersebut
h. *lang* adalah bahasa yang digunakan ketika membuat tweet.
i. *User_id_str* yaitu nomor id dari pengguna twitter.
j. *Conversation_id_str* merupakan nomor id yang terdapat dalam tweet.
k. *Username* adalah nama pengguna twitter.
l. *Tweet_url* yaitu link menuju ke tweet.

Seperti gambar dibawah ini:



Gambar 4. 2 Data Hasil Dari Crawling Data

Untuk mempermudah dalam proses menganalisis data, maka peneliti hanya menggunakan atribut text saja untuk proses selanjutnya,

Gambar 4. 3 Hasil Pemilihan Atribut

4.3 *Collecting* data

Tahap selanjutnya adalah *collecting data*. Tahapan ini bertujuan agar sistem komputer lebih mengenali bentuk *data set*. Selain itu, tahapan ini dapat mengubah data yang tidak tersusun menjadi data yang terstruktur. 1532 *raw data* yang dihasilkan dari proses

crawling di proses pada tahap *collecting* melalui 4 tahapan diantaranya yaitu *subproses, filter example,* dan *replace duplicate* dengan bantuan *tool Rapidmaner*. Proses *collecting* data menghasilkan 1488 data yang diasumsikan sudah terstruktur. Tahapan ini diawali dengan mengimportkan data berupa *file CSV*.

Gambar 4. 4 Import Data ke Rapidminer

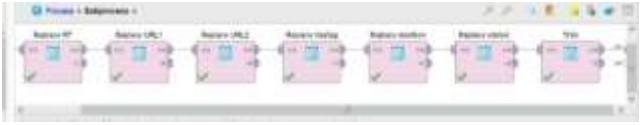
a. *Subproses*

Subproses merupakan proses penghapusan *RT, URL hastag, mention*, dan *symbol* yang ada pada data tweet.

contoh:

@navieses BOCAL BOCIL ANJING DASAR SEPUH, nih pacar aku ðŸ¥º https://t.co/jzQrKR8WMT

Proses:



Gambar 4. 5 Subproses

Output:

BOCAL BOCIL ANJING DASAR SEPUH, nih
pacar aku

b. *Filter example*

Filter example adalah proses yang digunakan untuk memfilter atau menghapus data tweet yang kosong.

Contoh: misal ada data yang isinya hanya titik maupun koma bahkan tidak ada isinya sama sekali

Gambar 4. 6 Filter Example

Output: maka data tersebut akan otomatis hilang

c. *Replace duplicate*

Replace duplicate digunakan untuk menghapus *tweet* yang sama pada data. Sehingga data yang akan digunakan tidak akan ada yang sama.

Contoh: jika terdapat tweet yang sama seperti ini

Apasih anjing
Apasih anjing

Gambar 4. 7 Replace Duplicate

Output: maka *ouput* yang akan muncul hanya satu *tweet* saja

Setelah dilakukan *collecting data*. Data hasil yang diperoleh yaitu sebanyak 1488 data, selanjutnya data tersebut akan dilabelli untuk diklasifikasikan dalam kelompok tweet yang mengandung bullying unsur positive atau negative, dengan format seperti ditunjukkan pada table 2. Saat dilalukan pelabelan hasil data yang diperoleh atau data yang digunakan oleh peneliti sebanyak 627 data. sedangkan 861 data yang lain tidak dapat digunakan dalam penelitian ini dikarenakan tidak memenuhi syarat.

Table 2. Contoh data tweet

No.	Kata kunci	Tweet	klasifikasi
1.	Anjing	Nah tambah gawat ini Anjing penjaga modal Kapitalis	negative
2.		fakultas gue fakultas tai bodo amat anjing	negative
3.		sumpah zize arhan lucubanget anjing	positive
4.		ganteng bgt anjing lu kenapa lucu ganteng bgt anjing gua kangen	positive
5.	Jelek	Biasanya yg model begini sih udh jelek gatau diri bnyk mau lgi	negative
6.		sayang banget tempatnya aesthetic di rusak sama tuh orang tulul mana fotonya jelek banget	negative
7.		Seolah tidak kata yang baik keluar dari mulut Kecuali ada kata2 yang jelek	positive
8.	Bodoh	Bodo amattt laki-laki yang takut sama perempuan berpendidikan tinggi itu artinya dia takut gabisa bodoh-bodohin ceweknya	negative
9.		Body shaming udah paling rendah dan bodoh Gak pny bahan tertawaan lain selain ngehina fisik orang	negative
10.		Begitulah strategi politiknya masyarakat dibikin bodoh agar tetap gampang dibodohi	negative

4.4 *Pre-Processing Data*

Tahapan *pre-prosesing* dilakukan sebelum mendapatkan *data training* dan *data testing*. dimana tahap *pre-prosesing* ini meliputi:

a. Tokenize

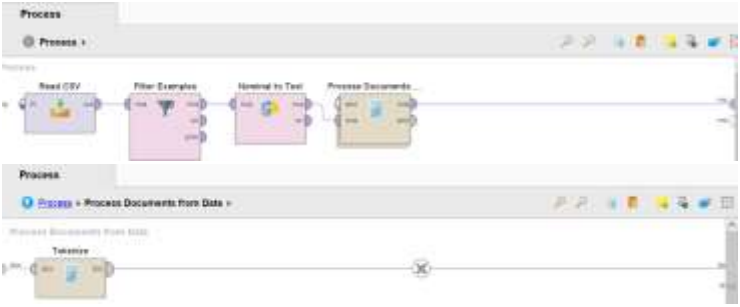
Proses tokenisasi adalah proses memecahkan kalimat menjadi potongan kata atau token untuk mengetahui asal munculnya kata dasar. Berikut proses *tokenize* menggunakan *tool Rapidminer.*

Text yang akan di input:

Table 3. Contoh Input Data

No.	Text
1.	Sumpah zize arhan lucu banget anjing
2.	Sayang banget tempatnya aesthetic dirusak sama tuh orang tulul mana fotonya jelek banget
3.	Body shaming udah paling rendah dan bodoh Gak pny bahan tertawaan lain selain ngehina fisik orang

Proses:



Gambar 4. 8 Proses Tokenize

Berikut ini merupakan hasil dari proses *tokenize*

Table 4. Hasil dari Proses Tokenize

No.	Text
1.	[sumpah, zize, arhan, lucu, banget, anjing]
2.	[sayang, banget, tempatnya, aesthetic, di, rusak, sama, tuh, orang, tulul, mana, fotonya, jelek, banget]
3.	[Body, shaming, udah, paling, rendah, dan, bodoh, Gak, pny, bahan, tertawaan, lain, selain, ngehina, fisik, orang]

b. Transform cases

Transform case merupakan proses pemerataan huruf dari huruf kapital menjadi huruf kecil atau sebaliknya. Pada penelitian ini data set diubah menjadi huruf kecil semua menggunakan fitur *lowercase.*

Input data:

Table 5. Input Data

No.	text
1.	WKWKWK ANJING NAJIS BGT ORG KEA GITU
2.	YA ALLAH RIFANA ANJ LU JELEK BANGEET KASIAN BAPAKNYA FAN AELAH
3.	EH SUMPAH COWO NGAPA MAKIN KESINI KELAKUANNYA MAKIN KEK APAAN SI BODOH

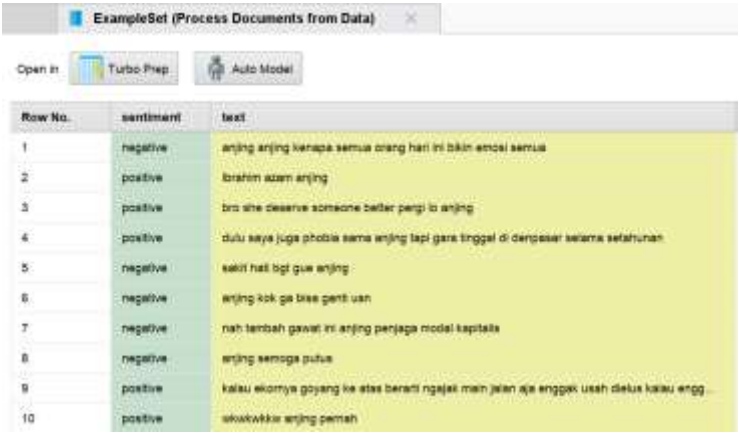
Proses data:

Gambar 4. 9 Proses Transform Cases

Hasil dari proses *transform cases*

Table 6. Hasil Transform Cases

No.	text
1.	wkwkwk anjing najis bgt org kea gitu
2.	ya allah rifana anj lu jelek bangeet kasian bapaknya fan aelah
3.	eh sumpah cowo ngapa makin kesini kelakuannya makin kek apaan si bodoh


ExampleSet (Process Documents from Data)

Open in Turbo Prep Auto Model

Row No.	sentiment	text
1	negative	anjing anjing kenapa semua orang hari ini bikin emosi semua
2	positive	ibrahim ajaib anjing
3	positive	bro she deserve someone better pergi lo anjing
4	positive	dulu saya juga phobia sama anjing tapi gara tinggal di denpasar selama setahunan
5	negative	sakit hati bgt gue anjing
6	negative	anjing kok ga bisa ganti ush
7	negative	nah tambah gawat ini anjing penjaga modal kapitalis
8	negative	anjing semoga putus
9	positive	kalau ekornya goyang ke atas berarti ngajak main jalan aja enggak usah dielus kalau engg...
10	positive	wkwkwkkk anjing pernah

Gambar 4. 10 Hasil Transform Cases

c. Stopword

Stopword merupakan proses menghilangkan kata umum (*common words*) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna.

Input data:

Table 7. Input Data Stop Word

No.	**text**
1	Iya nih
2	Aku sih tengah
3	susah

Proses data:

Gambar 4. 11 Proses Stopword

Output yang dihasilkan dari proses *stopword* sebagai berikut:

xampleSet (Process Documents from Data)

Turbo Prep | Auto Model

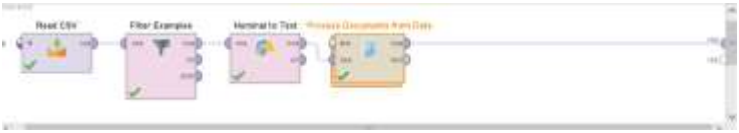
sentiment	text
positive	oo ini anak anjing kirain chiken cuyy
negative	orang power macam ni kau bagi gambar anjing tidur dekat tangga pun dia boleh buat bagi n...
negative	ini wajah yg nge body shaming in kakaknya weh maaf but iu jelek anjing
positive	aduh anjing lucu banget pacar siapa pacar gua
negative	cowok maupun cewek kalo punya hobi emang baiknya biarin dia me time sama hobinya nga...
positive	apa beda jenis binatang nya yg kucing yg anjing akhirnya kucingamp anjing bertarung dgn ...
positive	kemarin mimpiku seaneh ini dari tiba tiba yakin ex clientku anak jurusanku terus nyari dia iri ...
negative	emang anjing kalian ngaku lu pada ya apa yang diomongin
negative	pemilihan osis kenapa lapangan sih kan panas anjing pusing mual jadi satu
negative	minta double shot dikasih susu gula aren tuh gimana sih konsepnya anjing gue yang ngant...
negative	kenalan dr aplikasi dating emang rada rada sih w paling kesel sama yg ngechat duluan hi te...

Gambar 4. 12 Hasil Stopword

Contoh kata dalam data tersebut jika dilakukan proses *stop word* maka data tersebut akan menghilang dengan sendirinya. Yang artinya data yang tidak memiliki makna maka tidak akan digunakan dalam penelitian ini.

d. TF-IDF

TF-IDF merupakan proses penghapusan kemunculuan frekuensi munculnya kata sama pada dokumen dan banyaknya koleksi dokumen yang bersangkutan mengandung kata tertentu.



Gambar 4. 13 Proses TF_IDF

Output yang diperoleh dari proses diatas sebagai berikut:

Gambar 4. 14 Hasil TF_IDF

4.5 *SMOTE*

Peneliti melakukan proses sampling data dengan metode *SMOTE* untuk mendapatkan data yang baik dengan proses mengimbangi kelas

Gambar 4. 15 Proses SMOTE

Gambar 4. 16 Hasil SMOTE

Dapat dilihat dari gambar hasil dari *SMOTE* membuat kelas menjadi seimbang sehingga dapat membuat proses selanjutnya menjadi lebih mudah.

4.6 *K-VOLD Cros Validation*

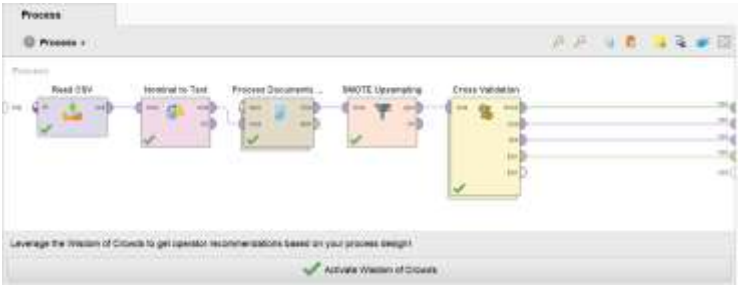
Penelitian ini menggunakan 10-*vold cross validation* untuk menguji *performa decision tree* dalam membentuk klasifikasi. Hasil akurasi klasifikasi merupakan hasil rata-rata untuk setiap data *training* dan *testing*.

Input:

Table 8. Input Data Cross Validation

NO.	Text	Sentiment
1.	anjing semoga putus	negative

2.	Trauma sampe skrg gara gara tetangga tolol punya anjing	negative
3.	Udah nangis kayaknya wkwkwk anjing	positive



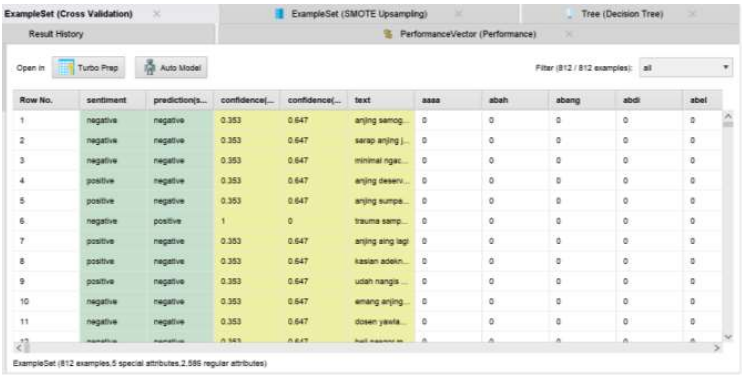
Gambar 4. 17 Proses Cross Validation

Output yang dihasilkan dari *cross validation* sebagai berikut:

Table 9. Hasil Cross Validation

NO.	**Text**	**Sentiment**	**Prediction(sentiment)**
1.	anjing semoga putus	negative	negative
2.	Trauma sampe skrg gara gara tetangga tolol	negative	positive

NO.	Text	Sentiment	Prediction(sentiment)
	punya anjing		
3.	Udah nangis kayaknya wkwkwk anjing	positive	negative



Gambar 4. 18 Hasil Cross Validation

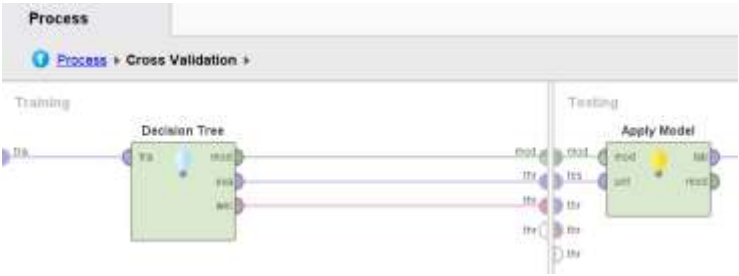
4.7 *Apply Model*

Klasifikasi dengan *Decision Tree*

Metode *Decision Tree* yang peneliti gunakan yaitu *Algoritma C4.5* yang mana merupakan algoritma dasar berdasarkan penerapan *Tree* dengan rumus 1 pada bab 2 seperti dibawah ini

$$Gain\ (S, A) = Entrophy(S) - \sum_{i=1}^{n} \frac{|s_1|}{|s|} * Entrophy(S_1)$$

Penerapan klasifiakasi ini menggunakan *subprosesing* yang akan memperoleh pohon keputusan. Proses *apply model* ini bisa dilihat pada gambar 2.1 pada bab dua.

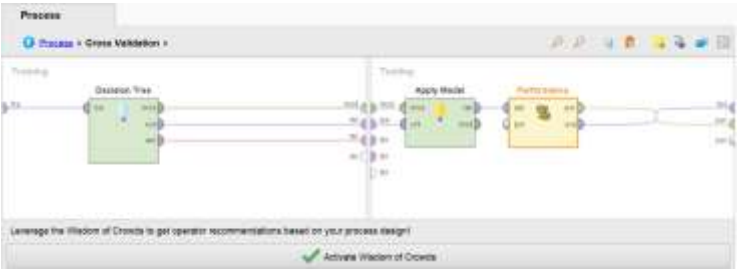


Gambar 4. 19 Proses Apply Model

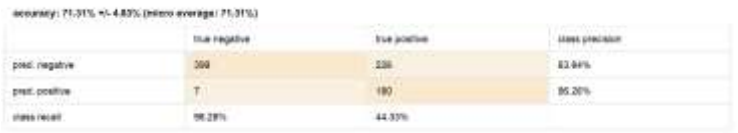
Dari proses diatas didapatkan pohon keputusan sebagai berikut:

Gambar 4. 20 Hasil Pohon Keputusan

Selanjutnya mencari probabilitas menggunakan metode *algoritma C4.5*. Berdasarkan hasil *eksperimen* yang telah peneliti lakukan ini menggunakan algoritma klasifikasi *Decision Tree* diketahui bahwa model juga telah berhasil membedakan data ke dalam kelas klasifikasi. Diketahui bahwa perhitungan akurasi dengan memanfaatkan *performance* yang mendapatkan *confusion matrix* sebagai berikut :



Gambar 4. 21 Proses Performance



Gambar 4. 22 Hasil Performance

Pada gambar 4.22 menunjukan bahwa dengan menggunakan algoritma *decision tree*, prediksi terhadap *statement harassment bullying* yang merupakan benar *statement harassment bullying* berjumlah 399 tweet, sementara 226 tweet yang di prediksi *bullying* ternyata tidak *bullying*. Sementara 7 tweet yang diprediksi tidak *bullying*

ternyata *bullying*, sementara 180 tweet yang di prediksi tidak *bullying* memang benar tidak *bullying*. Dari perhitungan diatas didapatkan nilai *accuracy* sebesar 71,31%.

4.8 *Performance*

Pada tahap pengujian kinerja metrik diuji dengan metode *confusion matrix*, dimana peneliti menghitung klasifikasi manual dengan hasil klasifikasi berdasarkan model *Decision Tree*.

Perhitungan akurasi pada data testing diatas dapat dikoreksi dengan perhitungan manual sehingga menghasilkan

$$Accurary = \frac{TP+TN}{TP+TN+FP+FN} X\ 100\%$$

$$= \frac{180+399}{180+399+226+7} X\ 100\% = 71{,}3\%$$

Diperoleh hasil perhitungan manual sebesar 71,3%. Untuk memperjelas hasil akurasi diperlukan perhitungan *Precision* dan *Recall.* Perhitungan manual sebagai berikut:

$$Precision = \frac{TP}{TP+FP} X\ 100\%$$

$$= \frac{180}{180+226} X\ 100\% = 44{,}3\%$$

Didapatkan nilai *precision* sebesar 44,3%.

$$Recall = \frac{TP}{TP+FN} X\ 100\%$$

$$= \frac{180}{180+7} X\ 100\% = 96{,}2\%$$

Hasil perhitungan manual *Recall* didapatkan nilai sebesar 96,2%.

Berdasarkan perhitungan algoritma yang telah dilakukan, Didapatkan hasil *Accuracy* sebesar 71,3%, *Precision* sebesar 44,3 %, dan *recall* sebesar 96,2%. Nilai tertinggi yang didapatkan yaitu *recall*. Untuk menentukan acuan performasi pada algoritma, pada artikel yang ditulis oleh (Reshika arthana, 2019) dikatakan dapat memilih algoritma yang memiliki nilai *recall* tinggi sebagai acuan performasi jika *False positive* lebih baik dari pada *false negative* artinya lebih baik algoritma memprediksi *statement negative* tetapi sebenarnya *positive* daripada algoritma salah memprediksi bahwa statement diprediksi *positive* yang sebenarnya *negative*. Dengan demikian, berdasarkan hasil pengujian yang dilakukan pada penelitian ini, model *decision tree* memiliki performa yang cukup baik pada penyelesaian klasifikasi statement harassment bullying.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan pembahasan hasil pengujian maka dapat diambil kesimpulan sebagai berikut:

1. Berdasarkan analisis dari klasifikasi pengkategorian statement *harassment bullying* melalui aplikasi X dengan metode *decision tree*, dengan kata kunci "anjing, jelek, dan bodoh" dengan data awal berjumlah 1538 tweet. Setelah itu dilakukan *collecting data* sehingga data hasil yang diperoleh yaitu 1488 data. Data digolongkan menjadi 2 label yaitu sentiment positif dan sentiment negatif. Data diolah menggunakan *tool Rapidminer*. Lalu dilakukan pembobotan *term frequency dan invers document frequency* (TF-IDF) untuk menghitung masing masing teks. Penelitian ini menggunakan 10-*vold cross validation* untuk menguji performa *decision tree* dalam membentuk klasifikasi. Kemudian dilakukan yang terakhir dilakukan klasifikasi menggunakan *Decision Tree*. Hasil klasifikasi menunjukkan 812 data dengan data ber *sentiment negatif* sebesar 625 dan data ber *sentiment positif* sebesar 187 data. Artinya 625 data *tweet* tersebut merupakan *statement*

harassment bullying. Dapat disimpulkan pelaku *bullying* di indonesia masih sangat tinggi.

2. Klasifikasi pengkategorian *statement harassment bullying* menggunakan metode *decision tree* baik digunakan karena nilai accuracy, precision dan recallnya cukup tinggi *accuracy* sebesar 71,3%, *precision* sebesar 44,3 % dan *Recall* sebesar 96,2%.

5.2 Saran

Ada beberapa hal yang peneliti sarankan untuk pengembangan penelitian selanjutnya:

1. Penelitian ini dapat diganti menggunakan topik yang berbeda dengan mencari topik masalah yang lebih spesifik.
2. Penelitian ini dapat di kembangkan dengan metode klasifikasi data mining lainnya.

DAFTAR PUSTAKA

Abdulloh, N., & Hidayatullah, A. F. (2021). Deteksi Cyberbullying pada Cuitan Media Sosial Twitter. *Automata*, *Vol 1*(1), 1–5.

Aditya Quantano Surbakti, Regiolina Hayami, & Januar Al Amien. (2021). Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter. *Jurnal CoSciTech (Computer Science and Information Technology)*, *2*(2), 91–97. https://doi.org/10.37859/coscitech.v2i2.2851

Alamsyah, M. F., Satriawan, T. P., Ramadanis, F. N., Mulyawan, R. A., Edmond, C., & Firmansyah, R. (2023). Analisa Komparasi Algoritma Naïve Bayes , Decision Tree Dan KKN Untuk Klasifikasi Kebakaran Hutan Pada Wilayah Aljazair. *Jurnal Sistem Informasi Dan Ilmu Komputer*, *1*(2), 72–86.

Amaliah, F., & Dwi Nuryana, I. K. (2022). Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter. *Journal of Informatics and Computer Science (JINACS)*, *3*(03), 384–393. https://doi.org/10.26740/jinacs.v3n03.p384-393

Anam, M. K., Pikir, B. N., & Firdaus, M. B. (2021). Penerapan Na¨ıve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemeritah. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(1), 139–150. https://doi.org/10.30812/matrik.v21i1.1092

Anwar, F. (2017). Perubahan dan Permasalahan Media Sosial. *Jurnal Muara Ilmu Sosial, Humaniora, Dan Seni*, *1*(1), 137. https://doi.org/10.24912/jmishumsen.v1i1.343

Burch dan Grudnitski dalam (Fauzi, 2017:19-21). (2019). Bab II Landasan Teori. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699.

Dessy Angelina, Hayati, U., & Dwilestari, G. (2023). Penerapan Metode Support Vector Machine Pada Sentimen Analisis Pengguna Twitter Terhadap Konser K-Pop. *Kopertip : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, *7*(1), 14–23. https://doi.org/10.32485/kopertip.v7i1.251

Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi. *Teknika*, *8*(1), 11–16. https://doi.org/10.34148/teknika.v8i1.95

Faizah, T., & Jananto, A. (2021). Perbandingan Algoritma C4 . 5 Dan Id3 Untuk Prediksi. *Jatisi*, *8*(2), 980–990.

Gullo, F. (2015). From patterns in data to knowledge discovery: What data mining can do. *Physics Procedia*, *62*, 18–22. https://doi.org/10.1016/j.phpro.2015.02.005

Hafizan, H. S. T. B., & Putri, A. N. S. T. B. (2020). Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi. *Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, *1*(2), 68–72.

Handayani, L. (2023). *Implementasi K-Nearest Neighbor Dalam Mengklasifikasikan Hate Tweet K-Popers Pada Twitter Teknologi, Dan*.

https://repository.uinjkt.ac.id/dspace/handle/123456789/73418

Handoko, W. T., Supriyanto, E., Purwadi, D. I., Budiarso, Z., & Listiyono, H. (2022). Klasifikasi Opini Pengguna Media Sosial Twitter Terhadap JNT Di Indonesia dengan Algoritma Decision Tree. *Jurnal Sains Komputer & Informatika (J-SAKTI*, *6*(2), 790–799.

Hidajat, M., Adam, A. R., & Danaparamita, M. (n.d.). *Dampak media sosial dalam*. *6*(1), 72–81.

Hijriana, N., & Rasyidan, M. (2017). Penerapan Metode Decision Tree Algoritma C4.5 Untuk Seleksi Calon Penerima Beasiswa Tingkat Universitas. *Al Ulum: Jurnal Sains Dan Teknologi*, *3*(1), 9. https://doi.org/10.31602/ajst.v3i1.983

Hozairi, H., Anwari, A., & Alim, S. (2021). Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes. *Network Engineering Research Operation*, *6*(2), 133. https://doi.org/10.21107/nero.v6i2.237

Ibeng, P. (2020). Pengertian Data, Fungsi Data, dan Macam Jenisnya. *13 Januari*, 1–10. https://pendidikan.co.id/pengertian-data/

Makasudede, Y. (1953). *Bab 2 tinjauan pustaka*. 8–45.

Maulana, F. A., Ernawati, I., Labu, P., & Selatan, J. (2020). Analisa sentimen cyberbullying di jejaring sosial twitter dengan algoritma naïve bayes. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA*, 529–538. https://conference.upnvj.ac.id/index.php/senamika/

article/view/619

Muhidin, N. F., Gustian, D., & Sihabudin, S. (2022). Analisis dan Penerapan Algoritma C4.5 Untuk Memprediksi Kualitas Penelitian dan Publikasi Ilmiah. *Jurnal Media Informatika Budidarma*, *6*(4), 2454. https://doi.org/10.30865/mib.v6i4.4463

Nilai, D., Yang, F., & Pasti, T. (2015). *IMPLEMENTASI METODE POHON KEPUTUSAN UNTUK KLASIFIKASI DATA*. *June*.

Oktavia, K. (2023). *Pendekatan multinomial Naive Bayes untuk klasifikasi jenis-jenis cyber harassment di media sosial Twitter*. http://etheses.uin-malang.ac.id/40836/%0Ahttp://etheses.uin-malang.ac.id/40836/1/19610079.pdf

Rahmanto, F., Pribadi, U., & Priyanto, A. (2021). Big Data: What are the Implications for Public Sector Policy in Society 5.0 Era? *IOP Conference Series: Earth and Environmental Science*, *717*(1), 0–7. https://doi.org/10.1088/1755-1315/717/1/012009

Sang, A. I., Sutoyo, E., & Darmawan, I. (2021). Analisis Data Mining untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree dan Support Vector Machine. *E-Proceeding of Engineering* , *8*(5), 8954–8963.

Sartika, D., & Sensuse, D. I. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *Jatisi*, *1*(2), 151–161.

Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020).

Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, *3*, 64–71.

Suparto, D., & Habibullah, A. (2021). Efektivitas Penggunaan Sosial Media Twitter dalam Penyebaran Informasi dalam Pelayanan Publik. *Indonesian Governance Journal : Kajian Politik-Pemerintahan*, *4*(2), 161–172. https://doi.org/10.24905/igj.v4i2.1927

Trihapsari, E., Pembimbing, D., Magister, P., Telematika-cio, B. K., Elektro, J. T., & Industri, F. T. (2016). *KLASIFIKASI CYBER BULLYING PADA MEDIA SOSIAL TWITTER DENGAN MENGGUNAKAN CYBER BULLYING CLASSIFICATION ON TWITTER SOCIAL MEDIA USING NAÏVE BAYES ALGORITHM*.

Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, *13*(1), 16–25. https://doi.org/10.36787/jti.v13i1.78

Abdulloh, N., & Hidayatullah, A. F. (2021). Deteksi Cyberbullying pada Cuitan Media Sosial Twitter. *Automata*, *Vol 1*(1), 1–5.

Aditya Quantano Surbakti, Regiolina Hayami, & Januar Al Amien. (2021). Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter. *Jurnal CoSciTech (Computer Science and Information Technology)*, *2*(2), 91–97. https://doi.org/10.37859/coscitech.v2i2.2851

Alamsyah, M. F., Satriawan, T. P., Ramadanis, F. N., Mulyawan, R. A., Edmond, C., & Firmansyah, R. (2023).

Analisa Komparasi Algoritma Naïve Bayes , Decision Tree Dan KKN Untuk Klasifikasi Kebakaran Hutan Pada Wilayah Aljazair. *Jurnal Sistem Informasi Dan Ilmu Komputer*, *1*(2), 72–86.

Amaliah, F., & Dwi Nuryana, I. K. (2022). Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter. *Journal of Informatics and Computer Science (JINACS)*, *3*(03), 384–393. https://doi.org/10.26740/jinacs.v3n03.p384-393

Anam, M. K., Pikir, B. N., & Firdaus, M. B. (2021). Penerapan Na ̈ıve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemeritah. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(1), 139–150. https://doi.org/10.30812/matrik.v21i1.1092

Anwar, F. (2017). Perubahan dan Permasalahan Media Sosial. *Jurnal Muara Ilmu Sosial, Humaniora, Dan Seni*, *1*(1), 137. https://doi.org/10.24912/jmishumsen.v1i1.343

Burch dan Grudnitski dalam (Fauzi, 2017:19-21). (2019). Bab II Landasan Teori. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699.

Dessy Angelina, Hayati, U., & Dwilestari, G. (2023). Penerapan Metode Support Vector Machine Pada Sentimen Analisis Pengguna Twitter Terhadap Konser K-Pop. *Kopertip : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, *7*(1), 14–23. https://doi.org/10.32485/kopertip.v7i1.251

Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi. *Teknika*, *8*(1), 11–16. https://doi.org/10.34148/teknika.v8i1.95

Faizah, T., & Jananto, A. (2021). Perbandingan Algoritma C4 . 5 Dan Id3 Untuk Prediksi. *Jatisi*, *8*(2), 980–990.

Gullo, F. (2015). From patterns in data to knowledge discovery: What data mining can do. *Physics Procedia*, *62*, 18–22. https://doi.org/10.1016/j.phpro.2015.02.005

Hafizan, H. S. T. B., & Putri, A. N. S. T. B. (2020). Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi. *Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, *1*(2), 68–72.

Handayani, L. (2023). *Implementasi K-Nearest Neighbor Dalam Mengklasifikasikan Hate Tweet K-Popers Pada Twitter Teknologi, Dan*. https://repository.uinjkt.ac.id/dspace/handle/123456789/73418

Handoko, W. T., Supriyanto, E., Purwadi, D. I., Budiarso, Z., & Listiyono, H. (2022). Klasifikasi Opini Pengguna Media Sosial Twitter Terhadap JNT Di Indonesia dengan Algoritma Decision Tree. *Jurnal Sains Komputer & Informatika (J-SAKTI*, *6*(2), 790–799.

Hidajat, M., Adam, A. R., & Danaparamita, M. (n.d.). *Dampak media sosial dalam*. *6*(1), 72–81.

Hijriana, N., & Rasyidan, M. (2017). Penerapan Metode Decision Tree Algoritma C4.5 Untuk Seleksi Calon Penerima Beasiswa Tingkat Universitas. *Al Ulum: Jurnal Sains Dan Teknologi*, *3*(1), 9.

https://doi.org/10.31602/ajst.v3i1.983

Hozairi, H., Anwari, A., & Alim, S. (2021). Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes. *Network Engineering Research Operation*, *6*(2), 133. https://doi.org/10.21107/nero.v6i2.237

Ibeng, P. (2020). Pengertian Data, Fungsi Data, dan Macam Jenisnya. *13 Januari*, 1–10. https://pendidikan.co.id/pengertian-data/

Makasudede, Y. (1953). *Bab 2 tinjauan pustaka*. 8–45.

Maulana, F. A., Ernawati, I., Labu, P., & Selatan, J. (2020). Analisa sentimen cyberbullying di jejaring sosial twitter dengan algoritma naïve bayes. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA*, 529–538. https://conference.upnvj.ac.id/index.php/senamika/article/view/619

Muhidin, N. F., Gustian, D., & Sihabudin, S. (2022). Analisis dan Penerapan Algoritma C4.5 Untuk Memprediksi Kualitas Penelitian dan Publikasi Ilmiah. *Jurnal Media Informatika Budidarma*, *6*(4), 2454. https://doi.org/10.30865/mib.v6i4.4463

Nilai, D., Yang, F., & Pasti, T. (2015). *IMPLEMENTASI METODE POHON KEPUTUSAN UNTUK KLASIFIKASI DATA*. *June*.

Oktavia, K. (2023). *Pendekatan multinomial Naive Bayes untuk klasifikasi jenis-jenis cyber harassment di media sosial Twitter*. http://etheses.uin-malang.ac.id/40836/%0Ahttp://etheses.uin-

malang.ac.id/40836/1/19610079.pdf

Rahmanto, F., Pribadi, U., & Priyanto, A. (2021). Big Data: What are the Implications for Public Sector Policy in Society 5.0 Era? *IOP Conference Series: Earth and Environmental Science*, *717*(1), 0–7. https://doi.org/10.1088/1755-1315/717/1/012009

Sang, A. I., Sutoyo, E., & Darmawan, I. (2021). Analisis Data Mining untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree dan Support Vector Machine. *E-Proceeding of Engineering* , *8*(5), 8954–8963.

Sartika, D., & Sensuse, D. I. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *Jatisi*, *1*(2), 151–161.

Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, *3*, 64–71.

Suparto, D., & Habibullah, A. (2021). Efektivitas Penggunaan Sosial Media Twitter dalam Penyebaran Informasi dalam Pelayanan Publik. *Indonesian Governance Journal : Kajian Politik-Pemerintahan*, *4*(2), 161–172. https://doi.org/10.24905/igj.v4i2.1927

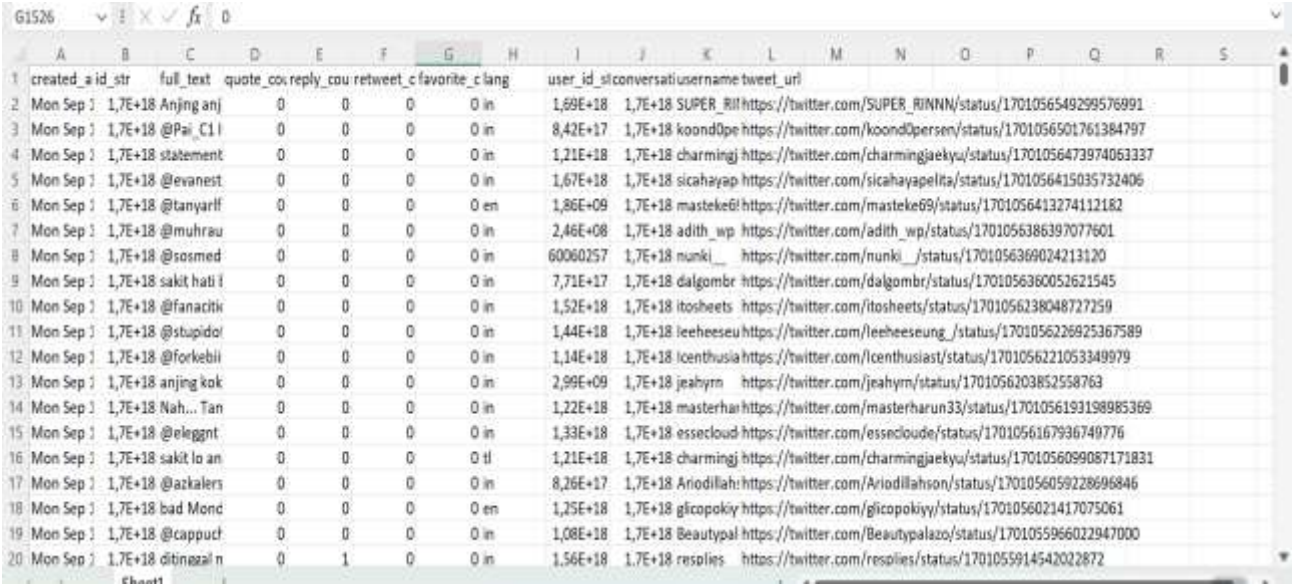
Trihapsari, E., Pembimbing, D., Magister, P., Telematika-cio, B. K., Elektro, J. T., & Industri, F. T. (2016). *KLASIFIKASI CYBER BULLYING PADA MEDIA SOSIAL TWITTER DENGAN MENGGUNAKAN CYBER BULLYING CLASSIFICATION ON TWITTER SOCIAL MEDIA USING*

NAÏVE BAYES ALGORITHM.

Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, *13*(1), 16–25. https://doi.org/10.36787/jti.v13i1.78

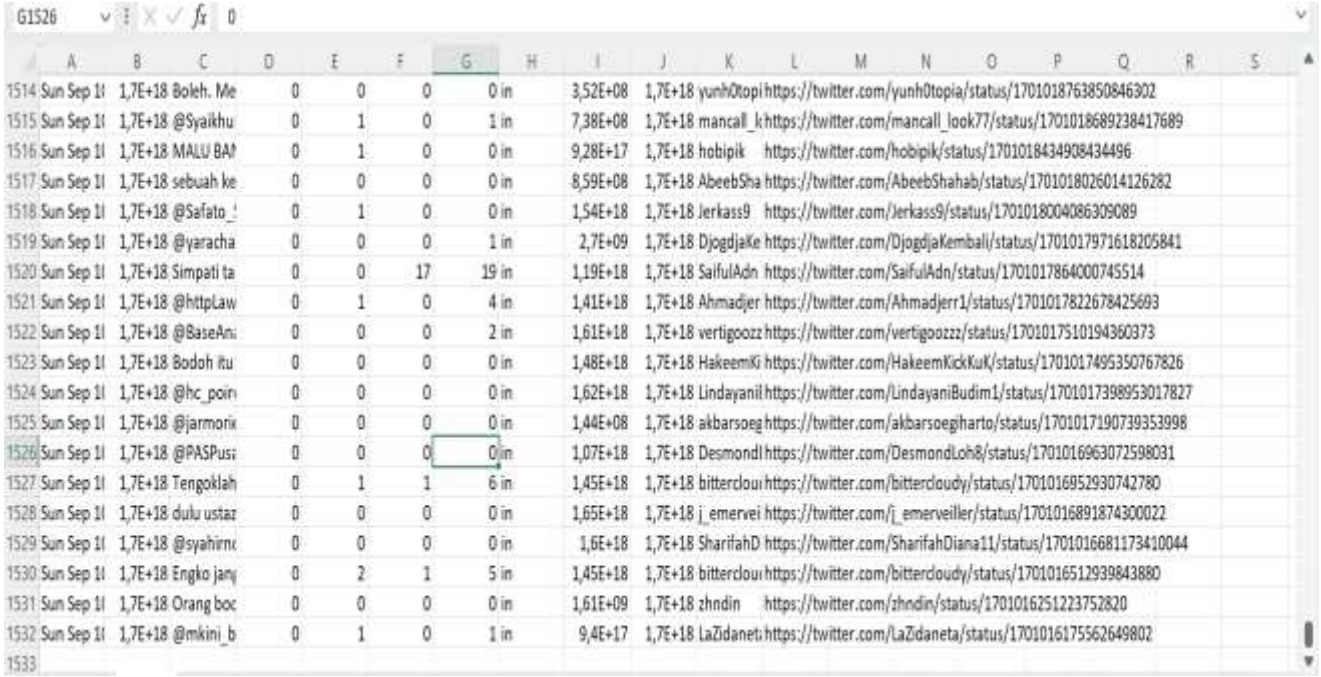
LAMPIRAN

Lampiran 1 Hasil Crawling Data



G1526 | 0

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	created_a	id_str	full_text	quote_cou	reply_cou	retweet_c	favorite_c	lang	user_id_st	conversati	username	tweet_url							
2	Mon Sep 1	1,7E+18	Anjing anj	0	0	0	0	in	1,69E+18	1,7E+18	SUPER_RI	https://twitter.com/SUPER_RINNN/status/1701056549299576991							
3	Mon Sep 1	1,7E+18	@Pai_C1 I	0	0	0	0	in	8,42E+17	1,7E+18	koond0pe	https://twitter.com/koond0persen/status/1701056501761384797							
4	Mon Sep 1	1,7E+18	statement	0	0	0	0	in	1,21E+18	1,7E+18	charmingj	https://twitter.com/charmingjaekyu/status/1701056473974063337							
5	Mon Sep 1	1,7E+18	@evanest	0	0	0	0	in	1,67E+18	1,7E+18	sicahayap	https://twitter.com/sicahayapelita/status/1701056415035732406							
6	Mon Sep 1	1,7E+18	@tanyarlf	0	0	0	0	en	1,86E+09	1,7E+18	masteke6!	https://twitter.com/masteke69/status/1701056413274112182							
7	Mon Sep 1	1,7E+18	@muhrau	0	0	0	0	in	2,46E+08	1,7E+18	adith_wp	https://twitter.com/adith_wp/status/1701056386397077601							
8	Mon Sep 1	1,7E+18	@sosmed	0	0	0	0	in	60060257	1,7E+18	nunki__	https://twitter.com/nunki__/status/1701056369024213120							
9	Mon Sep 1	1,7E+18	sakit hati l	0	0	0	0	in	7,71E+17	1,7E+18	dalgombr	https://twitter.com/dalgombr/status/1701056360052621545							
10	Mon Sep 1	1,7E+18	@fanacitic	0	0	0	0	in	1,52E+18	1,7E+18	itosheets	https://twitter.com/itosheets/status/1701056238048727259							
11	Mon Sep 1	1,7E+18	@stupido/	0	0	0	0	in	1,44E+18	1,7E+18	leeheeseu	https://twitter.com/leeheeseung_/status/1701056226925367589							
12	Mon Sep 1	1,7E+18	@forkebii	0	0	0	0	in	1,14E+18	1,7E+18	lcenthusia	https://twitter.com/lcenthusiast/status/1701056221053349979							
13	Mon Sep 1	1,7E+18	anjing kok	0	0	0	0	in	2,99E+09	1,7E+18	jeahyrn	https://twitter.com/jeahyrn/status/1701056203852558763							
14	Mon Sep 1	1,7E+18	Nah... Tan	0	0	0	0	in	1,22E+18	1,7E+18	masterhar	https://twitter.com/masterharun33/status/1701056193198985369							
15	Mon Sep 1	1,7E+18	@eleggnt	0	0	0	0	in	1,33E+18	1,7E+18	essecloud	https://twitter.com/essecloude/status/1701056167936749776							
16	Mon Sep 1	1,7E+18	sakit lo an	0	0	0	0	tl	1,21E+18	1,7E+18	charmingj	https://twitter.com/charmingjaekyu/status/1701056099087171831							
17	Mon Sep 1	1,7E+18	@azkalers	0	0	0	0	in	8,26E+17	1,7E+18	Ariodillah	https://twitter.com/Ariodillahson/status/1701056059228696846							
18	Mon Sep 1	1,7E+18	bad Monc	0	0	0	0	en	1,25E+18	1,7E+18	glicopokiy	https://twitter.com/glicopokiyy/status/1701056021417075061							
19	Mon Sep 1	1,7E+18	@cappuch	0	0	0	0	in	1,08E+18	1,7E+18	Beautypal	https://twitter.com/Beautypalazo/status/1701055966022947000							
20	Mon Sep 1	1,7E+18	ditinggal n	0	1	0	0	in	1,56E+18	1,7E+18	resplies	https://twitter.com/resplies/status/1701055914542022872							



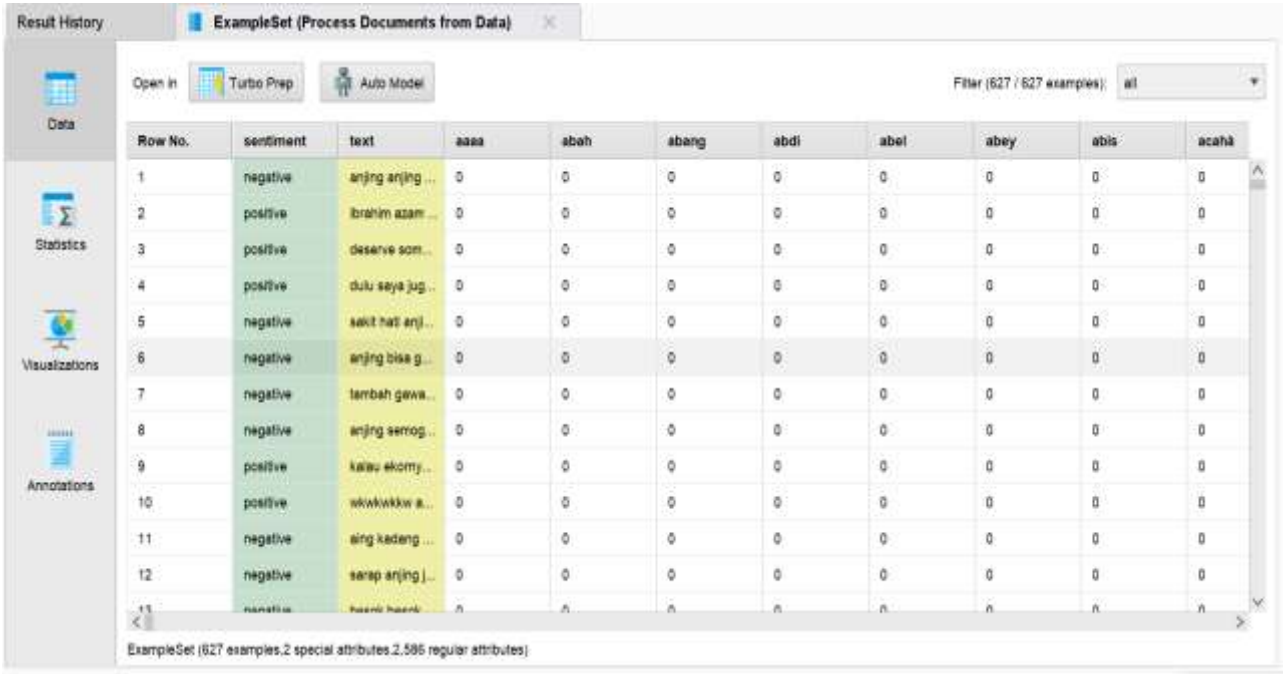
G1526 0

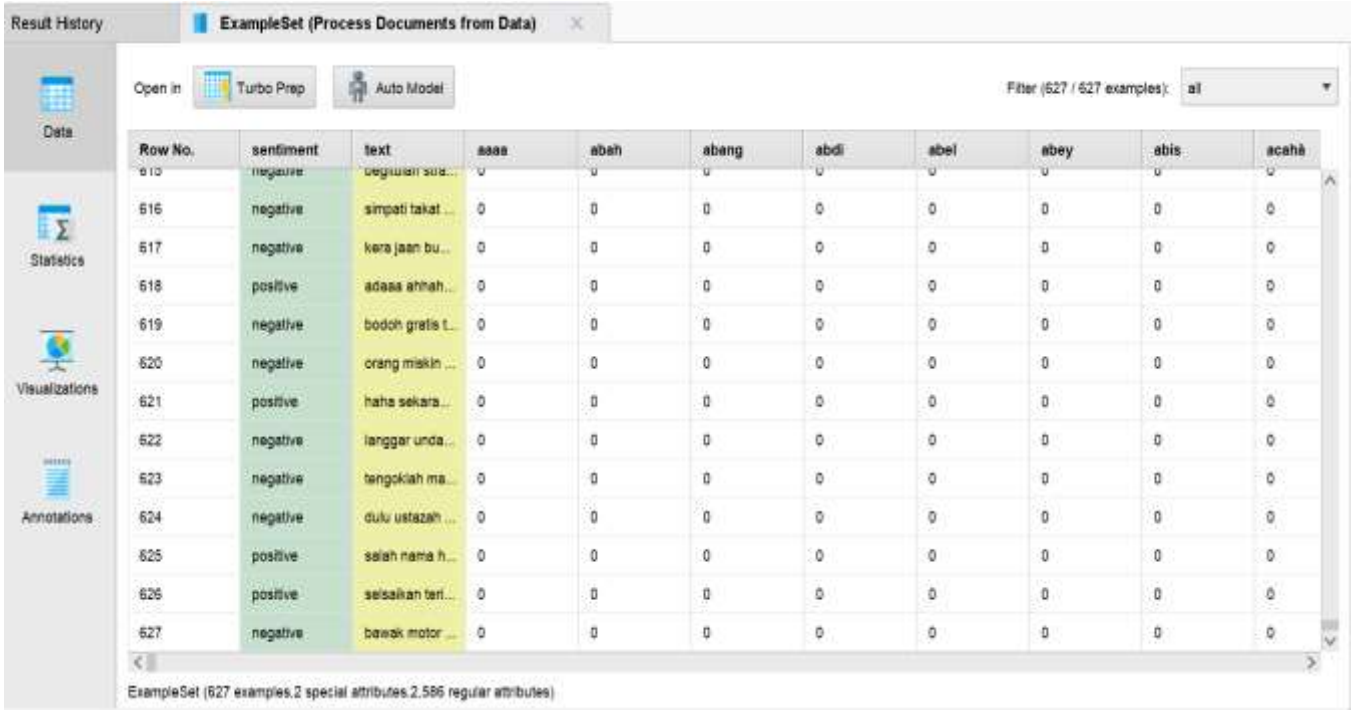
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1514	Sun Sep 1(	1,7E+18	Boleh. Me	0	0	0	0 in		3,52E+08	1,7E+18	yunh0topi	https://twitter.com/yunh0topia/status/1701018763850846302							
1515	Sun Sep 1(	1,7E+18	@Syaikhu	0	1	0	1 in		7,38E+08	1,7E+18	mancall_k	https://twitter.com/mancall_look77/status/1701018689238417689							
1516	Sun Sep 1(	1,7E+18	MALU BAI	0	1	0	0 in		9,28E+17	1,7E+18	hobipik	https://twitter.com/hobipik/status/1701018434908434496							
1517	Sun Sep 1(	1,7E+18	sebuah ke	0	0	0	0 in		8,59E+08	1,7E+18	AbeebSha	https://twitter.com/AbeebShahab/status/1701018026014126282							
1518	Sun Sep 1(	1,7E+18	@Safato_	0	1	0	0 in		1,54E+18	1,7E+18	Jerkass9	https://twitter.com/Jerkass9/status/1701018004086309089							
1519	Sun Sep 1(	1,7E+18	@yaracha	0	0	0	1 in		2,7E+09	1,7E+18	DjogdjaKe	https://twitter.com/DjogdjaKembali/status/1701017971618205841							
1520	Sun Sep 1(	1,7E+18	Simpati ta	0	0	17	19 in		1,19E+18	1,7E+18	SaifulAdn	https://twitter.com/SaifulAdn/status/1701017864000745514							
1521	Sun Sep 1(	1,7E+18	@httpLaw	0	1	0	4 in		1,41E+18	1,7E+18	Ahmadjer	https://twitter.com/Ahmadjerr1/status/1701017822678425693							
1522	Sun Sep 1(	1,7E+18	@BaseAn:	0	0	0	2 in		1,61E+18	1,7E+18	vertigoozz	https://twitter.com/vertigoozzz/status/1701017510194360373							
1523	Sun Sep 1(	1,7E+18	Bodoh itu	0	0	0	0 in		1,48E+18	1,7E+18	HakeemKi	https://twitter.com/HakeemKickKuK/status/1701017495350767826							
1524	Sun Sep 1(	1,7E+18	@hc_poin	0	0	0	0 in		1,62E+18	1,7E+18	Lindayanil	https://twitter.com/LindayaniBudim1/status/1701017398953017827							
1525	Sun Sep 1(	1,7E+18	@jarmorix	0	0	0	0 in		1,44E+08	1,7E+18	akbarsoeg	https://twitter.com/akbarsoegiharto/status/1701017190739353998							
1526	Sun Sep 1(	1,7E+18	@PASPus:	0	0	0	0 in		1,07E+18	1,7E+18	DesmondL	https://twitter.com/DesmondLoh8/status/1701016963072598031							
1527	Sun Sep 1(	1,7E+18	Tengoklah	0	1	1	6 in		1,45E+18	1,7E+18	bittercloui	https://twitter.com/bittercloudy/status/1701016952930742780							
1528	Sun Sep 1(	1,7E+18	dulu ustaz	0	0	0	0 in		1,65E+18	1,7E+18	j_emervei	https://twitter.com/j_emerveiller/status/1701016891874300022							
1529	Sun Sep 1(	1,7E+18	@syahirn	0	0	0	0 in		1,6E+18	1,7E+18	SharifahD	https://twitter.com/SharifahDiana11/status/1701016681173410044							
1530	Sun Sep 1(	1,7E+18	Engko jan	0	2	1	5 in		1,45E+18	1,7E+18	bittercloui	https://twitter.com/bittercloudy/status/1701016512939843880							
1531	Sun Sep 1(	1,7E+18	Orang boc	0	0	0	0 in		1,61E+09	1,7E+18	zhndin	https://twitter.com/zhndin/status/1701016251223752820							
1532	Sun Sep 1(	1,7E+18	@mkini_b	0	1	0	1 in		9,4E+17	1,7E+18	LaZidaneti	https://twitter.com/LaZidaneta/status/1701016175562649802							
1533																			

Lampiran 2 Hasil Data dari Collecting dan Pelabelan Data

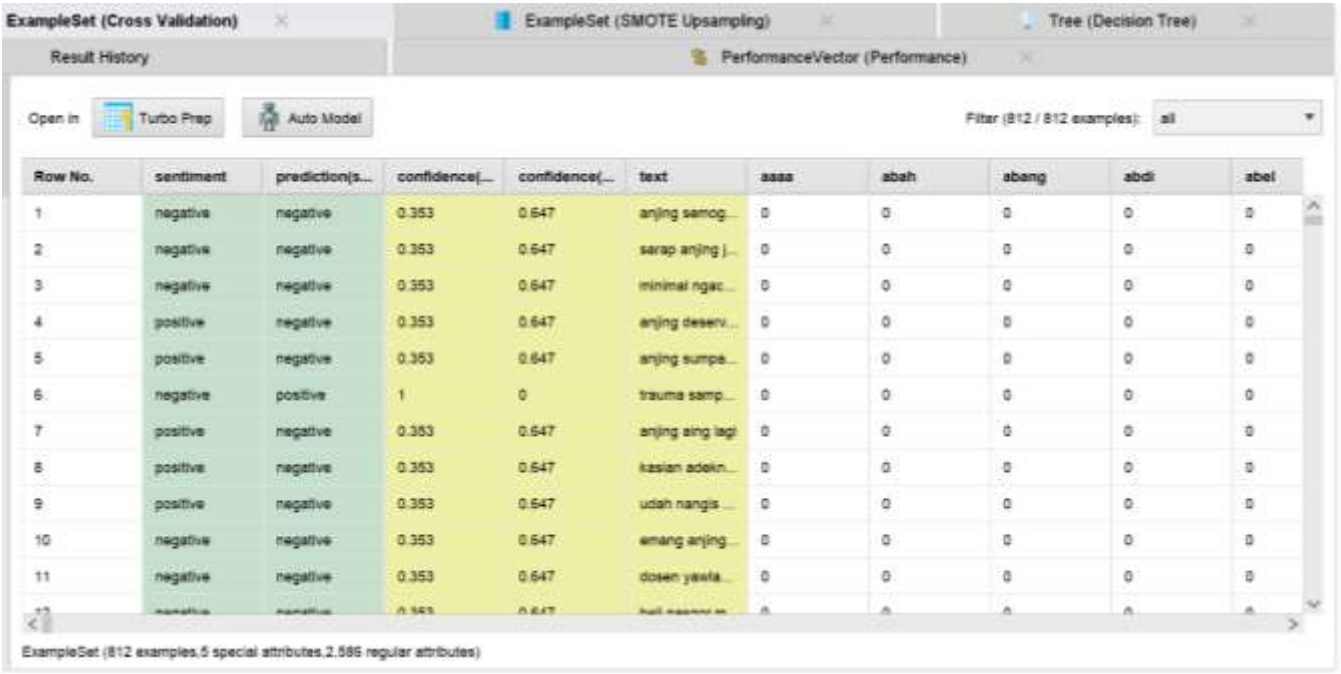
sentiment	**text**
negative	Otak geblek
negative	Makin lama makin bodoh puak PAS ni
positive	g pernah kenal dashcam bangpadahal internet kykny sdh masuk yabisa twitteran jgbukannya banyakin literasi malah anjing maju duluan
negative	gua gedeg di chat mulu diajakin photo box anjing
negative	ekspetasi lu tinggi banget anjing jauh jauh dari tu cewe dah ga cocok sama orang yang gabisa paham keinginan dan hak orang lain
positive	Nunut anjing
positive	kalo emang gak sreg yaudah tinggal mundur gausah malah ngomongin orangnya ditempat umum kaya gini gausah jadi ngomongin gak kuliah lah
positive	Anjing yg berjasa
positive	Anjing kalo gak di kasih makan emang gonggong terus sih KasihanðŸ
negative	Apasi anjing
negative	GUE BILANG JUGA APA DIEM ANJING
negative	oiya anjing
positive	emang elu siapa sih kok komenin orang sebegitunya yaudah ngga usah

Lampiran 3 Hasil Dari Pre-Precessing Data




Result History
ExampleSet (Process Documents from Data)
Data
Statistics
Visualizations
Annotations
Open in
Turbo Prep
Auto Model
Filter (627 / 627 examples):
all
Row No.
sentiment
text
aaaa
abah
abang
abdi
abel
abey
abis
acaha
616
negative
simpati takat ...
617
negative
kera jaan bu...
618
positive
adaaa ahhah...
619
negative
bodoh gratis t...
620
negative
orang miskin ...
621
positive
haha sekara...
622
negative
langgar unda...
623
negative
tengoklah ma...
624
negative
dulu ustazah ...
625
positive
salah nama h...
626
positive
selsaikan teri...
627
negative
bawak motor ...
ExampleSet (627 examples,2 special attributes,2,586 regular attributes)

Lampiran 4 Data Hasil Cross Validation



ExampleSet (Cross Validation) | ExampleSet (SMOTE Upsampling) | Tree (Decision Tree)

Result History | PerformanceVector (Performance)

Open in: Turbo Prep | Auto Model

Filter (812 / 812 examples): all

Row No.	sentiment	prediction(s...	confidence(...	confidence(...	text	aasa	abah	abang	abdi	abel
1	negative	negative	0.353	0.647	anjing samog...	0	0	0	0	0
2	negative	negative	0.353	0.647	serap anjing j...	0	0	0	0	0
3	negative	negative	0.353	0.647	minimal ngac...	0	0	0	0	0
4	positive	negative	0.353	0.647	anjing deserv...	0	0	0	0	0
5	positive	negative	0.353	0.647	anjing sumpe...	0	0	0	0	0
6	negative	positive	1	0	trauma samp...	0	0	0	0	0
7	positive	negative	0.353	0.647	anjing aing lagi	0	0	0	0	0
8	positive	negative	0.353	0.647	kasian adekn...	0	0	0	0	0
9	positive	negative	0.353	0.647	udah nangis ...	0	0	0	0	0
10	negative	negative	0.353	0.647	emang anjing...	0	0	0	0	0
11	negative	negative	0.353	0.647	dosen yawla...	0	0	0	0	0

ExampleSet (812 examples,5 special attributes,2.586 regular attributes)

Row No.	sentiment	prediction(s...	confidence(...	confidence(...	text	aaaa	abah	abang	abdi	abel
801	positive	positive	1	0	main sedjo pal...	0	0	0	0	0
802	positive	positive	1	0	anjing pening ...	0	0	0	0	0
803	positive	positive	1	0	oneus lagunya...	0	0	0	0	0
804	positive	positive	1	0	sayang banget...	0	0	0	0	0
805	positive	positive	1	0	selsaikan terim...	0	0	0	0	0
806	positive	positive	1	0	anjing ngajaki...	0	0	0	0	0
807	positive	positive	1	0	tahunan mung...	0	0	0	0	0
808	positive	positive	1	0	jiwa pembully ...	0	0	0	0	0
809	positive	positive	1	0	sorry kalo terla...	0	0	0	0	0
810	positive	positive	1	0	padahal abis b...	0	0	0	0	0
811	positive	positive	1	0	adaaa ahhaha...	0	0	0	0	0
812	positive	positive	1	0	masf pada anji...	0	0	0	0	0

Lampiran 5 *Script crawling data* kata kunci “anjing”

Script pyhton :

#menginstalkan package pada pyhton

```
# Import required Python package
!pip install pandas

# Install Node.js (because tweet-harvest built
using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl
gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl                                    -fsSL
https://deb.nodesource.com/gpgkey/nodesource-
repo.gpg.key   |   sudo   gpg   --dearmor   -o
/etc/apt/keyrings/nodesource.gpg

!NODE_MAJOR=20   &&   echo   "deb   [signed-
by=/etc/apt/keyrings/nodesource.gpg]
https://deb.nodesource.com/node_$NODE_MAJOR.x
nodistro      main"      |      sudo      tee
/etc/apt/sources.list.d/nodesource.list

!sudo apt-get update
!sudo apt-get install nodejs -y

!node -v
#proses mengumpulkan data dengan kata
kunci “anjing”
# Crawl Data

filename = 'anjing.csv'
search_keyword = 'anjing until:2023 since:2019'
limit = 100
```

```
!npx --yes tweet-harvest@latest -o "{filename}"
-s "{search_keyword}" -l {limit} --token ""
import pandas as pd

#setelah didapatkan datanya maka data
akan disimpan dalam bentuk file csv

# Specify the path to your CSV file
file_path = f"tweets-data/{filename}"

# Read the CSV file into a pandas DataFrame
df = pd.read_csv(file_path, delimiter=";")

# Display the DataFrame
display(df)
```

Lampiran 6 *Script crawling data* kata kunci "jelek"

Script pyhton :

#menginstalkan package pada pyhton

```
# Import required Python package
!pip install pandas

# Install Node.js (because tweet-harvest built
using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl
gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl                                    -fsSL
https://deb.nodesource.com/gpgkey/nodesource-
repo.gpg.key | sudo gpg --dearmor -o
/etc/apt/keyrings/nodesource.gpg
```

```
!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_$NODE_MAJOR.x nodistro main" | sudo tee /etc/apt/sources.list.d/nodesource.list

!sudo apt-get update
!sudo apt-get install nodejs -y

!node -v
#proses mengumpulkan data dengan kata kunci "jelek"
# Crawl Data

filename = 'jelek.csv'
search_keyword = 'jelek until:2023 since:2019'
limit = 100

!npx --yes tweet-harvest@latest -o "{filename}" -s "{search_keyword}" -l {limit} --token ""
import pandas as pd

#setelah didapatkan datanya maka data akan disimpan dalam bentuk file csv

# Specify the path to your CSV file
file_path = f"tweets-data/{filename}"

# Read the CSV file into a pandas DataFrame
df = pd.read_csv(file_path, delimiter=";")

# Display the DataFrame
display(df)
```

DAFTAR RIWAYAT HIDUP

1. **Identitas diri**

   Nama lengkap : Diah Ayu Anggraini
   Tempat, Tanggal Lahir : Demak, 25 Maret 2000
   Alamat : Ds. Purwareja Rt/Rw 15/02, Kec. Sematu Jaya, Kab. Lamandau, Kalimantan Tengah
   No Handphone : 081250452908
   Email : anggraini_1908046004@student.walisongo.ac.id

2. **Riwayat Pendidikan**

   Pendidikan Formal:
   SD : SDN Jatirejo 2
   SMP : SMP N 1 Mijen
   SMA : SMA N 1 Sematu Jaya
   Pendidikan non formal: -

Semarang, 20 Desember 2023

Diah Ayu Anggraini

NIM 1908046004