

BAB I

PENDAHULUAN

1.1. Latar Belakang

Media sosial merupakan platform media yang memfokuskan pada eksistensi pengguna serta memfasilitasi mereka dalam beraktivitas maupun berkolaborasi. Karena itu media sosial dapat dilihat sebagai medium (fasilitator) online yang menguatkan hubungan antar pengguna sekaligus sebuah ikatan sosial. Melalui media sosial yang semakin banyak berkembang memungkinkan informasi menyebar dengan mudah di masyarakat. Informasi dalam bentuk apa pun dapat disebarluaskan dengan mudah dan cepat sehingga mempengaruhi cara pandang, gaya hidup, serta budaya suatu bangsa. (Manning et al., 2008)

Salah satu media sosial yang sangat populer saat ini yaitu Twitter. Twitter merupakan media sosial yang berbentuk microblogging atau ngeblog secara singkat dalam satu paragraf dengan maksimal 280 huruf, karena jumlah huruf dalam satu kali tweet terbatas/dibatasi. Pengguna Twitter terdiri dari berbagai kalangan, seperti pejabat, selebritis, artis, hingga masyarakat biasa pada umumnya. Akan tetapi, kemudahan yang diberikan untuk berbagi informasi melalui media sosial tak luput dari penyalahgunaan yang dilakukan oleh penggunanya. Salah satu bentuk penyalahgunaan tersebut yaitu cyberbullying.

Cyberbullying merupakan bentuk tindak kekerasan yang dilakukan oleh seseorang terhadap korbannya di dunia maya, dimana korban dihina, diejek, dipermalukan dan diintimidasi oleh pelaku. Dampak dari cyberbullying yaitu mempengaruhi mental korban, bahkan banyak dari korban bullying berakhir dengan bunuh diri karena tidak tahan dengan banyak tekanan. Ada hubungan positif dan signifikan antara perilaku pelaku cyberbullying dan perilaku korban cyberbullying yang mana semakin reaktif perilaku pelaku cyberbullying maka semakin reaktif pula perilaku korban cyberbullying. Hal ini membuktikan betapa besarnya pengaruh cyberbullying terhadap kehidupan sosial. (Khaira et al., 2020)

Oleh karena itu, sampai Saat ini, tidak ada data statistik yang tersedia informasi konkret tentang anak-anak yang terkena dampak kasus cyberbullying di Indonesia. Interaksi di internet tersebut membuat anak-anak dan remaja khususnya remaja penggemar K-Pop, rentan menjadi korban bahkan pelaku cyberbullying di media sosial, mengingat aktivitas yang mereka lakukan banyak dihabiskan di internet dan media sosial. Kekhawatiran tersebut juga tergambar dalam Survei Penetrasi Internet dan Perilaku Pengguna Internet di Indonesia 2018 yang dirilis oleh Asosiasi Penyelenggara Jasa Internet di Indonesia (APJII) yang menunjukkan, 49% pengguna internet pernah dirisak (di-bully) dalam bentuk diejek atau dilecehkan di media sosial. hal tersebut menjadikan baik orang dewasa, remaja maupun anak-anak rentan untuk mendapatkan perlakuan negatif tersebut bahkan menjadi pelaku.”

K-Pop juga dapat dikenali dengan munculnya fans club K-Pop di media sosial yang menyediakan semua informasi tentang artis K-Pop kepada para penggemar K-Pop. Di Indonesia, khususnya di kalangan remaja. Oleh karena itu, pendekatan pembelajaran mesin dapat digunakan untuk mendeteksi yang mengandung unsur intimidasi. Artinya, menganalisa sentimen. Klasifikasi membutuhkan pendekatan pembelajaran mesin yang dapat membedakan antara kata-kata yang mengandung cyberbullying dan yang tidak.(Ortega et al., 2008)

Analisis sentimen merupakan sub bagian dari Natural Language Processing (NLP) yang fokus pada menentukan perasaan yang terkandung pada sebuah teks. Analisis sentimen dikenal dengan opinion mining yang merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat. Ide dasar dari analisis setimen adalah untuk mendeteksi polaritas teks pada dokumen, kalimat, dan tweet. Polaritas sentimen terbagi tiga yaitu positif, negatif, dan netral.

Berbagai penelitian terkait analisis sentimen telah banyak dilakukan. Terdapat dua pendekatan untuk melakukan analisis sentimen, pendekatan yang pertama adalah berbasis machine learning yaitu dengan melatih classifier pada dataset yang telah dilabelkan secara manual. Pendekatan yang kedua adalah berbasis leksikal yang tidak memerlukan pelatihan dataset, mengukur polaritas

suatu kalimat atau dokumen berdasarkan pada sentimen kata-kata dan frasa-frasa sambil menerapkan aturan-aturan tertentu yang diambil dari linguistik. Penelitian ini akan menganalisis sentimen terkait cyberbullying dari komentar masyarakat pada media sosial Twitter, untuk itu diperlukan metode yang dapat mengklasifikasikan komentar ke dalam kelas positif, negatif, dan netral. Kelas negatif berarti komentar yang mengandung elemen cyberbullying, kelas positif berarti komentar mengandung unsur motivasi atau dukungan, dan kelas netral adalah komentar yang tidak mengandung elemen cyberbullying.

Dalam penelitian sebelumnya, SentiStrength dievaluasi pada beberapa dataset yang terdiri dari teks bahasa Inggris. Dalam penelitian ini, kami mereplikasi studi sebelumnya untuk teks bahasa Indonesia. Penelitian ini menggunakan metode SentiStrength berbasis kamus/leksikon. Kamus/leksikon SentiStrength berisi terms serta bobot kekuatan sentimennya. Metode SentiStrength menggunakan daftar idiom dan emotikon serta beberapa aturan lain seperti peningkatan kekuatan sentimen ketika huruf kapital digunakan. (Nurzahputra & Muslim, 2016)

Penelitian ini menggunakan metode *Prototype* dan *naive bayes*. Mengingat sistem yang akan di bangun ini baru sehingga diperlukan metode yang teratur, perancangan suatu sistem untuk memudahkan proses pembuatan.

Harapan dengan adanya sistem yang di usulkan ini dapat digunakan oleh pengguna media sosial bisa melihat penyebaran cyberbullying dan mencari kata-kata yang terkandung bullying. Berdasarkan latar belakang diatas maka diangkatlah penelitian yang berjudul **“Analisis Sentimen Cyberbullying Penggemar Kpop Di Media Sosial Twitter Menggunakan Metode Naive Bayes”**.

1.2. Rumusan Masalah

Adapun rumusan masalah pada penelitian ini untuk menentukan apakah suatu pesan yang dikirim melalui media sosial khususnya *Twitter* mengandung unsur *bullying* atau tidak.

1.3. Tujuan Penelitian

1. Untuk mengklasifikasikan sentimen pada komentar tweet berdasarkan kelas positif, negatif, dan netral
2. Untuk mengetahui trend yang ada di kalangan remaja dalam tindakan terhadap bullying di sosial media
3. Mengimplementasikan metode Naïve Bayes untuk melakukan klasifikasi sentimen cyberbullying pada media sosial twitter.

1.4. Batasan Masalah

Untuk melakukan penelitian secara spesifik dan jelas, diperlukan batasan yang diterapkan pada penelitian ini. Batasan yang dilakukan pada penelitian, adalah:

1. Komentar tweet yang dianalisis adalah komentar berbahasa Indonesia
2. Sentimen analysis dilakukan dengan metode klasifikasi
3. Algoritma yang diterapkan adalah Naïve bayes.
4. Komentar tweet akan dikelompokkan menjadi komentar positif, komentar negatif, dan netral.
5. Media yang digunakan untuk mengambil data *tweet* pada *twitter* ialah *twitter API*.

1.5. Manfaat Penelitian

1. Membantu mengetahui kata positif, negatif, dan netral.
2. Mengetahui kata yang mengandung konten cyberbullying.
3. Dapat membantu pengguna sosial media khususnya pada remaja dan orang dewasa agar lebih bijak lagi dalam menggunakan media sosial dan tidak melakukan tindakan cyberbully agar tidak merugikan orang lain.

1.6. Metodologi Penelitian

1.6.1. Alat dan Bahan

Adapun alat dan bahan yang digunakan untuk mendukung penelitian ini baik itu perangkat keras (*Hardware*) maupun perangkat lunak (*Software*) yang digunakan :

1. Perangkat Keras (*Hardware*)

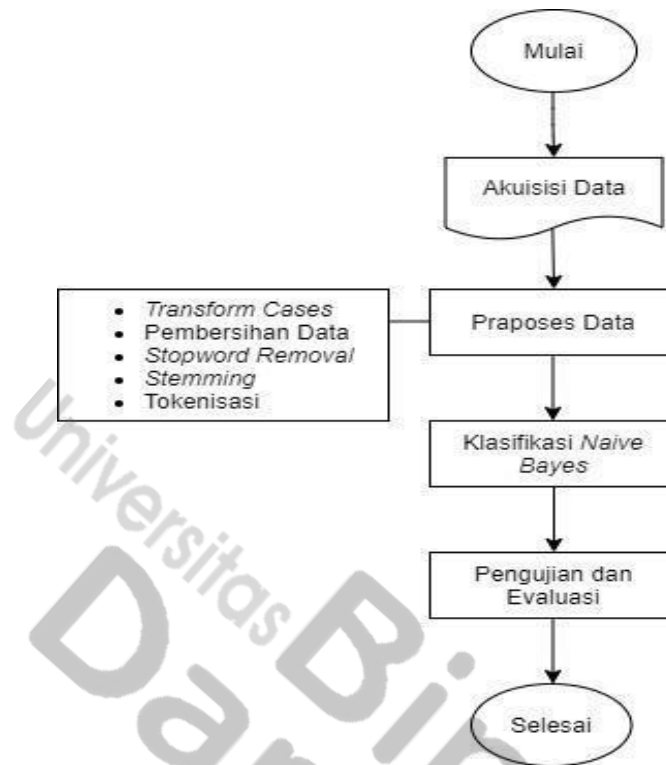
1. Asus VivoBook processor Intel Core i5
2. RAM 8GB
3. VGA Nvidia GeForce MX252Gb
4. SSD 512Gb
5. Hard disk 1 Tera Byte

2. Perangkat Lunak (*Software*)

1. Sistem operasi Windows 11
2. Google Chrome
3. Microsoft word 2016 (sebagai pembuatan laporan)
4. Pemograman menggunakan Bahasa Python
5. Twitter API (untuk menarik data)
6. Draw.io
7. Mysql
8. Jupyter Notebook
9. Xampp

1.7. Metode Analisis Data

Metode Analisis data yang akan digunakan adalah sebagai berikut:



Gambar 1. Gambar di atas berisi setiap proses yang telah dilakukan dalam mengerjakan penelitian ini.

1.7.1. Akuisisi Data

Tahap ini merupakan tahap untuk mendapatkan data yang dibutuhkan, data yang dimaksud dalam penelitian ini adalah tweet pengguna twitter yang mention akun publik figur politik di Indonesia yang akan dipakai untuk data latih dan data uji. Data tweet didapatkan dengan cara menggunakan API (Application Programming Interface) yang sudah disediakan oleh twitter sebanyak 540 data tweet yang dibagi 440 data untuk data latih yang diberikan label positif dan negatif serta 100 tweet untuk data uji.

1.7.2. Praproses Data

Pada tahap ini sangat penting untuk mengurangi attribute pada data yang kurang berpengaruh pada proses klasifikasi dalam penelitian ini. Data yang digunakan pada langkah ini merupakan data data mentah yang memiliki noise, sehingga hasil dari tahap ini ialah data yang sudah siap untuk memudahkan proses klasifikasi. Pada praproses data terdapat beberapa tahap yaitu case folding, tokenisasi, stopwords removal, stemming, dan padding

sehingga data yang masih terdapat noise akan bersih dan dapat digunakan untuk tahap selanjutnya.

1.7.3. Klasifikasi Naive Bayes

Tahap selanjutnya dari penelitian ini yaitu pengklasifikasian data tweet yang telah melewati tahap pra proses data metode yang digunakan yaitu klasifikasi dan penelitian ini menggunakan Naïve Bayes. Algoritma tersebut akan mengklasifikasikan tweet yang mengandung unsur cyberbullying. Pengerjaan menggunakan *Naïve Bayes* terdapat dua proses, yaitu data latih dan data uji. Langkah pertama ialah melakukan pelatihan sistem dengan data latih, langkah selanjutnya yaitu proses uji sistem dengan mengacu probabilitas data latih. Klasifikasi *Naïve Bayes* menggunakan library dari NLTK (Natural Language Toolkit) merupakan library python yang digunakan dalam proses machine learning teks.

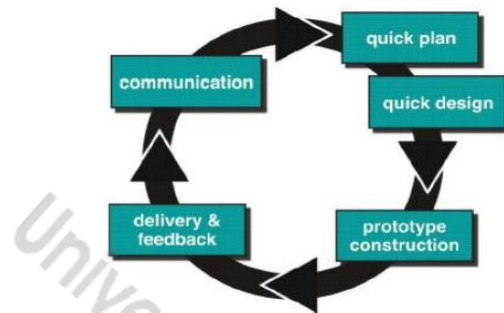
1.7.4. Pengujian dan Evaluasi

Merupakan tahap untuk menguji hasil klasifikasi dengan menggunakan metode confusion matrix dengan sejumlah data yang diuji. Pada tahap ini menghitung nilai akurasi, precision, recall, dan specificity.

1.8. Metode Pengumpulan Data

Pada tahapan pengumpulan data di ambil secara langsung dari akun twitter @bts_bighit dengan memanfaatkan search *API Key Twitter (API)* yang telah disediakan oleh *Twitter dan rapid miner*. Kemudian setelah mendapatkan akses API key maka dapat dicari data yang diinginkan melalui fasilitas kolom pencarian di twitter yang berhubungan dengan penelitian ini. Tahap pertama dalam melakukan proses analisis sentimen adalah pengumpulan data. Data dari Twitter dengan 2000 Cari "cyberbulliyng, kpop, dan bullying" menggunakan aplikasi Rapidminer. Twitter Search digunakan untuk mengambil data dari media sosial Twitter. Data ini disimpan dan disimpan dalam format Excel.csv dengan membuat Excel.csv. , untuk membantu mengidentifikasi duplikat atau konten duplikat yang harus dihapus. Pemilihan atribut, penggantian nilai yang hilang, subprocessing, dan analisis sentimen diikuti, dan hasil akhirnya adalah data yang digunakan sebagai data pelatihan dalam algoritma naive Bayes.

1.9. Metode Pengembangan Sistem



Gambar 2 Pengembangan Sistem

Metode yang akan digunakan dalam pengembangan sistem adalah metodologi *prototyping*. *Prototyping* merupakan suatu metode dalam pengembangan sistem yang menggunakan pendekatan untuk membuat suatu program dengan cepat dan bertahap sehingga segera dapat dievaluasi oleh pemakai. *Prototyping* mewakili model produk yang akan dibangun atau mensimulasikan struktur, fungsionalitas dan operasi sistem. Di perancangan *prototyping* ini akan ada dashboard berbasis web untuk visualisasi sebaran cyberbully dan juga ada fitur untuk prediksi kata-kata yang mengandung cyberbully. (Manalu, 2019)

Kelebihan dan Kekurangan Metode Prototype

A. Kelebihan Metode Prototype

1. Menghemat waktu dalam pengembangan sistem.
2. Penentuan kebutuhan lebih mudah diwujudkan.
3. Pelanggan / klien berpartisipasi aktif dalam pengembangan sistem, sehingga hasil perangkat lunak mudah disesuaikan dengan kebutuhan dan keinginan pelanggan.
4. Komunikasi yang baik antaral pelanggan dan pengembang.
5. Pengembang dapat lebih mudah dalam menentukan kebutuhan pelanggan.

B. Kekurangan Metode Prototype

1. Proses perancangan dan analisi terlalu singkat.
2. Biasanya Kurang fleksibel dalam menghadapi perubahan.
3. Pengembang kadang-kadang membuat kompromi implementasi dengan menggunakan sistem operasi yang tidak relevan dan algoritma yang tidak efisien.

1.10. Sistematika Penulisan

Sistematika penulisan skripsi yang digunakan, dapat diuraikan sebagai berikut:

BAB I PENDAHULUAN

Dalam BAB ini, berisi penjelasan mengenai latar belakang, perumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penelitian, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Berisi penjelasan terkait *cyberbullying*, serta uraian mengenai teori-teori yang digunakan sebagai landasan atau acuan dalam melaksanakan penelitian.

BAB III ANALISIS DAN PERANCANGAN

Bab ini menjelaskan tentang analisa kebutuhan dalam pembangunan sistem serta rancangan sistem yang akan dibuat.

BAB IV HASIL DAN PEMBAHASAN

Berisi tentang penjelasan hasil pengembangan Sistem Informasi tracer study menggunakan PHP pada SMK Muhammadiyah 1 Palembang beserta penjelasan terkait penggunaan sistem dan fitur-fitur di dalamnya.

BAB V KESIMPULAN DAN SARAN

Yaitu bagian terakhir yang memuat kesimpulan dan juga saran dari peneliti untuk pengembangan selanjutnya