

BAB I

PENDAHULUAN

I.1. Latar Belakang

Perkembangan bidang teknologi dan informasi saat ini sedang mengalami pertumbuhan yang pesat dan tidak terbendung. Hal ini tidak lepas dari kebutuhan masyarakat yang beranekaragam dan ketatnya persaingan penyedia teknologi informasi sehingga menuntut teknologi pembaruan setiap waktu. Teknologi informasi yang menjadi prioritas pengembangan adalah teknologi informasi yang dapat memudahkan kegiatan sehari-hari masyarakat. Tidak hanya untuk memudahkan kegiatan manusia, tetapi *trend* perkembangan teknologi informasi saat ini telah menjadi ajang *pamer* bagi sebagian masyarakat. Pengembangan teknologi informasi dengan tidak diimbangi dengan pengembangan karakter manusia akan menimbulkan permasalahan baru di kemudian hari.

Salah satu bidang teknologi informasi yang mengalami perkembangan pesat beberapa dekade terakhir yaitu teknologi informasi media sosial. Munculnya istilah kaum milenial merupakan istilah untuk generasi masyarakat yang fasih dalam menggunakan media sosial dan teknologi terkini dalam kegiatannya sehari-hari. Bahkan dalam setiap aspek pekerjaan yang dilakukan tidak pernah terlepas dari penggunaan media sosial dan teknologi tersebut. Namun, permasalahan belakangan ini yang terjadi yaitu kebebasan penggunaan media sosial tidak diimbangi dengan

pemantauan yang ketat, telah menimbulkan hal-hal yang mengarah kepada perilaku kejahatan di media sosial. Salah satu perilaku kejahatan dalam media sosial tersebut adalah tindakan *Cyberbullying*.

Yang dimaksud dengan tindakan *Cyberbullying* yaitu berupa tindakan menyakiti atau merendahkan orang lain menggunakan sarana media sosial. *Cyberbullying* dapat menimbulkan dampak depresi kepada orang yang mendapat perlakuan ini yang bahkan beberapa diantaranya telah sampai kepada dampak yang lebih ekstrim, yaitu bunuh diri (Kartono, 1992). Dalam sumber berita detiknews.com, tahun 2017 pemerintah menyatakan 84% remaja berusia 12 sampai 17 tahun di Indonesia menjadi korban tindakan perundungan (*bullying*) dan kebanyakan kasus *bullying* yang ditemukan merupakan *cyberbullying* (Laksana, 2017).

Cyberbullying didefinisikan sebagai tindakan penyalahgunaan sosial media yang diatur oleh Undang-Undang Informasi dan Transaksi Elektronik (UU ITE) No. 11 Tahun 2008 pasal 27 ayat 3 yang berisi tentang pencemaran nama baik ataupun penghinaan”. Untuk dapat mengklasifikasikan suatu informasi bersifat *bullying* atau tidak, tidak dilakukan secara manual seperti dulu. Metode yang lebih cepat digunakan adalah menggunakan bantuan teknologi. Untuk mewujudkan teknik klasifikasi menggunakan media teknologi (dalam hal ini komputer), dibutuhkan penelitian dan pembuatan korpus dalam bahasa manusia karena pada dasarnya bahasa yang terprogram pada komputer jauh berbeda dengan yang dimengerti mesin komputer.

Topik dalam penelitian ini adalah bagaimana mengidentifikasi teks atau komentar-komentar pada jejaring sosial. Teknik mengidentifikasi teks pada sebuah kumpulan data disebut *Text Mining*. *Text mining* adalah sub bagian dari proses pengolahan data yang lebih besar dan kompleks yaitu *data mining*. *Text mining* bertujuan untuk mendapatkan data pengetahuan dan informasi yang berarti dari sekumpulan data yang besar dan kompleks. Tantangan pada *text mining* adalah kompleksitas struktur teks yang tidak lengkap, arti kurang jelas dan tidak baku, menggunakan bahasa yang berbeda serta penerjemahan yang kurang presisi (Sussolaikah & Alwi, 2016).

Pada semua teknik *data mining*, dibutuhkan metode dan algoritma yang sesuai untuk mengolah data masukan menjadi hasil keluaran yang sesuai. Sehingga, saat ini ada banyak algoritma *data mining* pada berbagai kasus. Contoh, pada teknik klasifikasi: *Rule-Based Algorithms*, *Distance-Based Algorithms*, *Statistical-Based Algorithms*, *Neural Network-Based Algorithms*, dan *Decision Tree-Based Algorithms* (Norwawi, 2020).

Dalam Wu, et al. (2009), telah dilakukan penelitian pada bidang *data mining* dan dirumuskan sepuluh algoritma terbaik dan populer di kalangan peneliti pembelajaran mesin. Hasilnya, diketahui bahwa kebanyakan membahas masalah algoritma klasifikasi. Beberapa algoritma klasifikasi populer dan terbaik yaitu *k-Nearest Neighbor (k-NN)*, *pohon keputusan C4.5*, *Classification and Regression Trees (CART)* dan *Naive Baiyes* (Wu & Kumar, 2009).

Syahfitri Kartika Lidya, et al. (2015) melakukan penelitian dengan melakukan perbandingan kinerja algoritma *Support Vector Machine* (SVM) dan *k-Nearest Neighbor* (*k-NN*) dalam analisis sentimen teks Bahasa Indonesia. Acuan kinerja didasarkan pada banyaknya data dan banyaknya atribut. Kesimpulannya, diketahui bahwa *running time* pada SVM relatif lebih lambat daripada *k-NN*, karena dalam proses validasi menggunakan *K-Fold Cross Validation*, matriks vektor di algoritma SVM cukup besar sehingga mengakibatkan iterasi validasi membutuhkan waktu tambahan (Lidya et al., 2015).

Penelitian sentimen analisis juga dilakukan oleh Baghat (2020), di mana dalam penelitian tersebut dibandingkan kinerja dari algoritma *Naïve Bayes* dan *k-Nearest Neighbor* (*k-NN*). Data yang digunakan adalah kumpulan komentar dalam jejaring sosial Twitter kemudian diidentifikasi kalimat yang mengandung unsur perundungan (*bullying*). Hasilnya, metode *k-Nearest Neighbor* (*k-NN*) memberikan tingkat akurasi yang lebih baik. Menguji algoritma adalah hal yang perlu untuk memberikan gambaran algoritma yang baik untuk kasus yang berbeda (Sitepu, 2018).

Berdasarkan latar belakang tersebut, penelitian ini mencoba melakukan analisis kinerja berdasarkan akurasi tiga jenis algoritma yang merupakan bagian dari sepuluh algoritma populer, yaitu algoritma *Naïve Bayes*, *Support Vector Machine* (SVM) dan *k-Nearest Neighbor* (*k-NN*) pada *text mining* berupa klasifikasi komentar *cyber bullying*. Adapun sumber data sets adalah mengumpulkan data set dari komentar-komentar yang ada pada jejaring sosial twitter terutama saat pemilihan umum tahun 2019 yang terdapat banyak komentar *bullying* dari *haters*

masing-masing kandidat. Uji hasil klasifikasi menggunakan teknik *confusion matrix* berdasarkan pada penilaian *precision*, *recall*, dan *F-measure*. Atas dasar inilah penelitian ini mengambil judul: “Perbandingan Analisis Kinerja Naïve Bayes, K-Nearest Neighbor dan Support Vector Machine dalam Mengidentifikasi Komentar Cyber Bullying pada Jejaring Sosial”.

I.2. Rumusan Masalah

Pada penelitian ini, rumusan masalah penelitian meliputi:

1. Mengklasifikasikan *text mining* berupa data komentar *cyber bullying* pada jejaring sosial menggunakan algoritma *Naïve Bayes*, k-NN dan SVM.
2. Menganalisa *performance* dari algoritma *Naïve Bayes*, k-NN dan SVM dari sudut pandang berdasarkan *confusion matrix* pada permasalahan *text mining*.

I.3. Tujuan

Adapun tujuan penelitian ini yaitu:

1. Memperoleh hasil klasifikasi dari data penelitian, yaitu berupa data komentar *cyber bullying* pada jejaring sosial, menggunakan algoritma *Naïve Bayes*, k-NN dan SVM.
2. Memperoleh algoritma terbaik pada penelitian *text mining* dari segi *performance* algoritma klasifikasi *Naïve Bayes*, k-NN dan SVM berdasarkan *confusion matrix*.

I.4. Manfaat

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Dapat mengetahui lebih dalam tentang tahapan klasifikasi pada masalah *text mining* berbahasa Indonesia.
2. Dapat mengetahui penilaian akurasi algoritma *Naïve Bayes*, k-NN dan SVM berdasarkan *confusion matrix*, meliputi *precision*, *recall*, *F-measure*, dan *success rate*.
3. Hasil penelitian dapat menjadi rujukan para pembaca dalam memahami algoritma *Naïve Bayes*, k-NN dan SVM, terutama aspek akurasi.

I.5. Batasan Masalah

Dalam menganalisis komentar *cyber bullying* menggunakan algoritma *Naïve Bayes*, k-NN dan SVM dilakukan beberapa batasan sebagai berikut:

1. Dataset yang digunakan adalah komentar berbahasa Indonesia yang dikumpulkan melalui twitter.
2. Menggunakan bahasa pemrograman *python*.
3. Dataset yang digunakan berupa teks komentar-komentar pada jejaring sosial dalam menganalisis *cyber bullying* menggunakan algoritma *Naïve Bayes*, k-NN dan SVM.

I.6. Sistematika Penulisan

Sistematika tesis yang dalam penelitian ini terdiri dari 6 bab, yaitu:

BAB I PENDAHULUAN

Bab pendahuluan akan membahas tentang latar belakang dari masalah penelitian yang di angkat, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan yang diuraikan dalam laporan penelitian ini.

BAB II TINJAUAN PUSTAKA

Bab tinjauan pustaka membahas tentang teori-teori yang dibutuhkan dalam menyusun penelitian berdasarkan penelitian yanf telah dilakukan dahulu. Rumus-rumusan penting dan pengertian-pengertian istilah juga disusun dalam bab ini.

BAB III METODOLOGI PENELITIAN

Bab metodologi penelitian membahas tentang gambaran umum tahapan dalam proses penelitian yang terdiri dari pengumpulan data, input data, *preprocessing data*, analisa menggunakan algoritma machine learning beserta teknik uji performa dari algoritma.

BAB IV HASIL

Bab ini melaporkan hasil pelaksanaan penelitian. Data diolah melalui tahap *text processing*, pembobotan TF-IDF kemudian dikelompokkan

dengan menggunakan metode KNN, Naive Bayes dan SVM. Bab ini juga mendeskripsikan cara penyelesaian masalah secara matematis dan secara manual, sehingga diperoleh gambaran bagaimana sebenarnya cara kerja sistem yang dibangun.

BAB V PEMBAHASAN

Bab ini membahas makna dari hasil yang diperoleh untuk menjawab rumusan masalah penelitian. Selain itu, bab ini menjelaskan bagaimana cara kerja sistem menyelesaikan masalah penelitian. Hasil akhir proses ini berisi hasil perbandingan ketiga metode menggunakan *confussion matrix*.

BAB VI KESIMPULAN DAN SARAN

Bab ini memuat kesimpulan dan saran berdasarkan penjelasan dan hasil pengujian dari beberapa bab sebelumnya. Penulis berharap penelitian ini dapat bermanfaat untuk penelitian selanjutnya.