

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengguna Internet di Indonesia mengalami perkembangan yang cukup pesat hal tersebut dibuktikan dari data Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pengguna internet pada tahun 1998 berjumlah 512.000 dan pada tahun 2002 mengalami peningkatan menjadi 4.5 juta pengguna dan pada akhir 2007 pengguna internet Indonesia sejumlah 25 juta dan ada 2016 pengguna Internet Indonesia tercatat sebanyak 79 juta (Mujahiddin, 2017; Setyaningrum & Dwilestari, 2017).

Di antara penggunaan internet di Indonesia penggunaan media sosial yang paling dirasakan pertumbuhannya hal tersebut berdasarkan data Kementerian Komunikasi dan Informatika RI (KEMKOMINFO) pada tahun 2019 pengguna internet di Indonesia sebanyak 150 juta jiwa dan penetrasi penggunaan media sosial sebanyak 56% (Rizkinaswara, 2019). Salah satu Media sosial yang banyak digunakan di Indonesia adalah Twitter. Menurut data KEMKOMINFO Indonesia menempati urutan kelima pengguna media sosial Twitter (KEMKOMINFO, 2017).

Twitter merupakan media sosial bertipe *micro-blogging* (blog berukuran kecil) yang didirikan oleh Jack Dorsey pada Maret 2006 dan diluncurkan pada Juli 2006, Twitter memiliki pengguna aktif sebanyak 300 juta akun setiap bulannya (Alfina et al., 2017).

Keunikan dari Twitter adalah mempunyai kicauan dengan ukuran maksimum 280 karakter (Kartini, 2017). Kicauan menurut Kamus Besar Bahasa Indonesia (KBBI) berarti bunyi burung, berbicara sembarang saja atau berceloteh dan pesan yang dikirim melalui layanan jejaring sosial Twitter (KBBI, 2020).

Dengan kebebasan yang ada pada media sosial dan bertambahnya pengguna Twitter menjadikan Twitter memiliki berbagai macam kicauan di antaranya adalah kicauan yang mengandung konten asusila, pencemaran nama baik, penghinaan, berita bohong, ujaran kebencian dll. Permasalahan ini disikapi Pemerintah Indonesia dengan mengesahkan undang-undang no. 11 tahun 2008 mengenai ITE

yang mengatur informasi dan transaksi elektronik (Setyaningrum & Dwilestari, 2017) dan juga dengan terbitnya surat edaran Kapolri SE/6/X/2015 pada tahun 2015 mengenai penanganan ujaran kebencian (Anam & Hafiz, 2015; Polri, 2015).

Di antara hal-hal yang diatur oleh undang-undang no. 11 tahun 2008 tentang ITE dan surat edaran Kapolri SE/6/X/2015 adalah informasi yang mengandung ujaran kebencian. Ujaran kebencian merupakan ungkapan atau bahasa yang digunakan untuk mengekspresikan kebencian terhadap suatu individu atau golongan dengan tujuan untuk menghinakan atau mempermalukan pihak yang ditujukan ujaran kebencian tersebut (Munir, Fauzi, & Perdana, 2018). Menurut KOMNAS HAM ujaran kebencian adalah segala tindakan dan usaha baik langsung maupun tidak langsung yang didasarkan pada kebencian atas dasar suku, agama, aliran keagamaan, keyakinan/kepercayaan, ras, antar golongan, warna kulit, gender, kaum difabel, dan orientasi seksual yang merupakan hasutan terhadap individu atau kelompok agar terjadi diskriminasi, kekerasan, penghilangan nyawa dan/ atau konflik sosial yang dilakukan dengan berbagai sarana (KOMNAS HAM, 2015).

Ujaran kebencian sangat berbahaya karna merendahkan orang lain, menimbulkan kerugian material dan korban manusia, dapat berdampak konflik dan dapat berdampak pada pemusnahan kelompok (KOMNAS HAM, 2015). Ujaran kebencian juga dapat menyebabkan pelabelan, stigma, pengucilan, diskriminasi, kekerasan, kebencian terhadap kelompok, kekerasan terhadap kelompok dan pembantaian terhadap etnis/kelompok agama/bangsa (KOMNAS HAM, 2015).

Pada tahun 2014 istilah *haters* mulai populer dikarenakan pemilihan umum, hal tersebut menjadikan warga Indonesia saling berkubu dan saling melakukan argumen-argumen politik di sosial media, tidak sedikit dari argumen tersebut yang merupakan ujaran kebencian (Juliswara, 2017). Laman Kompas merilis pada tahun 2015 terdapat 671 kasus ujaran kebencian yang sudah dilaporkan meliputi pencemaran nama baik, pelecehan, fitnah, provokasi, dan ancaman (Movanita, 2016).

Pada 2017 Kasus ujaran kebencian kembali marak dikarenakan adanya pemilihan kepala daerah (Pilkada) serentak. Pada 2017 kicauan ujaran kebencian mengenai Pilkada DKI Jakarta merupakan ujaran kebencian yang paling banyak,

Umar Syahid Aulia Rahman, 2020

IMPLEMENTASI METODE MULTINOMIAL NAIVE BAYES UNTUK MENDETEKSI KICAUAN YANG MENGANDUNG UJARAN KEBENCIAN PADA DATA TWITTER BAHASA INDONESIA

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.epu.edu

hal tersebut dibahas pada penelitian yang dilakukan oleh (Alfina et al., 2017) pada tahun 2017 yang bertujuan mengidentifikasi ujaran kebencian dalam rentang pelaksanaan Pilkada DKI dengan cara dikumpulkan 1100 kicauan yang diberikan label oleh relawan yang berbeda agama dan mencapai kesepakatan 100% bahwa dari 1100 kicauan tersebut 260 kicauan mengandung ujaran kebencian .

Dengan banyaknya data kicauan yang ada pada media sosial Twitter akan sulit untuk membedakan kicauan yang mengandung ujaran kebencian dan yang tidak mengandung ujaran kebencian oleh karna itu pada penelitian kali ini penulis mengajukan solusi yaitu dengan pengklasifikasian kicauan yang mengandung ujaran kebencian dengan implementasi metode *Multinomial Naive Bayes* (MNB) yang dalam hipotesis kami algoritma *Multinomial Naive Bayes* memiliki performa yang cukup baik atau memiliki hasil akurasi atau *f-measure* yang lebih tinggi dari 70% dalam kasus pengklasifikasian teks, hipotesis ini berdasarkan dari hasil penelitian sebelumnya mengenai *sentiment analysis* pada data Twitter dengan menggunakan berbagai algoritma di antaranya adalah MNB, *Support Vector Machine* (SVM) dan *Logistic Regression* (LR) yang menghasilkan akurasi sebesar 73.7% pada algoritma MNB dengan ekstraksi fitur menggunakan TF-IDF dan 73.9% menggunakan ekstraksi fitur *count vectorizer* (Devi & Saharia, 2020).

Penelitian mengenai performa MNB selanjutnya adalah penelitian yang dilakukan oleh (Daga, Gupta, Vardhan, & Mukherjee, 2020) mengenai prediksi *retweet* dan *like* pada Twitter dengan menggunakan SVM, LR, *Random Forest* (RF), *Neural Network* (NN) dan MNB dengan fitur ekstraksi menggunakan TF-IDF dan *Doc2vec* menghasilkan akurasi untuk prediksi *retweet* MNB dengan TF-IDF pada *data training* sebesar 75.9% merupakan yang tertinggi dibanding algoritma lain yang dipakai pada penelitian tersebut, sedangkan akurasi untuk prediksi *like* dengan TF-IDF menghasilkan akurasi 77.7% dan merupakan hasil tertinggi dibanding algoritma lain pada penelitian tersebut.

Penelitian selanjutnya mengenai performa MNB pada klasifikasi teks terbukti pada penelitian (Wongso, Luwinda, Trisnajaya, Rusli, & Rudy, 2017) mengenai klasifikasi berita pada situs berita CNN Indonesia menggunakan MNB, *Multivariate Bernouli* dan SVM menghasilkan hasil terbaik ada pada klasifikasi

menggunakan metode MNB dengan seleksi fitur menggunakan TF-IDF dengan *precision* 0.984 dan *recall* 0.984.

Pada saat penulis mengajukan penelitian ini pada 2018 penulis tidak dapat menemukan dataset yang sesuai dengan tema penelitian ini yaitu klasifikasi kicauan bahasa Indonesia apakah mengandung ujaran kebencian atau tidak, oleh karna itu penulis mengajukan untuk membangun dataset kicauan bahasa Indonesia dengan label mengandung ujaran kebencian dan tidak mengandung ujaran kebencian. Pada prosesnya penulis menemukan dataset *abusive and hate Twitter* dalam bahasa Indonesia pada situs Kaggle.com yang diunggah pada tanggal 17 Mei 2020 oleh karna itu penulis menjadikan dataset tersebut sebagai referensi dalam pembangunan dataset dalam penelitian ini (Ibrohim & Budi, 2019).

Pada penelitian sebelumnya oleh (Alfina et al., 2017; Antariksa, Purnomo WP, & Ernawati, 2019; Davidson, Warmesley, Macy, & Weber, 2017) *dataset* yang didapatkan merupakan *dataset* yang tidak seimbang antara kicauan yang mengandung ujaran kebencian dan yang tidak mengandung ujaran kebencian sehingga ukuran performa model klasifikasi menggunakan akurasi tidak dapat dilakukan, oleh karna itu ukuran performa pada model klasifikasi ujaran kebencian dilakukan menggunakan *f-measure*. Hal ini dilakukan agar dapat melihat performa algoritma yang dipakai terhadap topik bahasan yang dicari dimana pada penelitian ini fokus klasifikasi adalah kicauan yang mengandung ujaran kebencian.

1.2 Rumusan Masalah

1. Bagaimana membangun *dataset* ujaran kebencian pada kicauan (*tweet*) bahasa Indonesia?
2. Bagaimana membedakan kicauan bahasa Indonesia yang mengandung ujaran kebencian dan yang tidak mengandung ujaran kebencian secara otomatis?
3. Bagaimana tingkat *f-measure* klasifikasi dengan metode *Multinomial Naive Bayes* pada kasus kicauan bahasa Indonesia yang mengandung ujaran kebencian?

1.3 Tujuan Penelitian

Berikut adalah tujuan yang ingin dicapai dari penelitian ini:

1. Membangun *dataset* kicauan (*tweet*) ujaran kebencian dalam bahasa Indonesia.
2. Membedakan kicauan dalam bahasa Indonesia dengan memberikan label (klasifikasi) yang mengandung ujaran kebencian dan yang tidak mengandung ujaran kebencian secara otomatis dengan metode *Multinomial Naive Bayes* .
3. Menganalisis tingkat *f-measure* klasifikasi dengan metode *Multinomial Naive Bayes* pada kasus kicauan bahasa Indonesia yang mengandung ujaran kebencian.

1.4 Batasan Masalah

Batasan masalah penelitian ini yaitu:

1. *Dataset* merupakan data kicauan yang diambil di www.twitter.com
2. *Dataset* merupakan kicauan berbahasa Indonesia
3. *Dataset* kicauan merupakan data kicauan dengan kata kunci tertentu
4. *Dataset* pada penelitian ini dilabeli sendiri oleh penulis dan tidak melalui proses validasi oleh ahli.
5. Penelitian menggunakan metode *Multinomial Naive Bayes* dengan menggunakan bahasa pemrograman python

1.5 Sistematika Penulisan

Berikut ini adalah sistematika penulisan yang dilakukan dalam menyusun skripsi :

BAB 1 PENDAHULUAN

Bab ini berisikan masalah yang diangkat dalam penelitian meliputi latar belakang mengapa penelitian ini dilakukan. Dimana pada saat ini pengguna media sosial Twitter sangat banyak sehingga sulit untuk membedakan suatu kicauan apakah mengandung ujaran kebencian atau tidak, oleh karna itu dibutuhkan cara yang lebih

Umar Syahid Aulia Rahman, 2020

IMPLEMENTASI METODE MULTINOMIAL NAIVE BAYES UNTUK MENDETEKSI KICAUAN YANG MENGANDUNG UJARAN KEBENCIAN PADA DATA TWITTER BAHASA INDONESIA

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.epu.edu

praktis dan mudah untuk membedakan kicauan pada media sosial Twitter dimana pada penelitian ini kami menggunakan metode *Multinomial Naive Bayes* untuk mengklasifikasi ujaran kebencian pada kicauan bahasa Indonesia karna kami melihat algoritma *Multinomial Naive Bayes* terbukti memiliki performa yang cukup baik dalam pengklasifikasian teks. Pada bab ini juga menjelaskan mengapa kami membangun dataset pada penelitian ini dimana pada awal penelitian tidak ditemukan dataset yang sesuai. Bab ini terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian yang akan dilakukan, manfaat penelitian dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Bab ini berisikan mengenai kajian teori yang mendukung penelitian ini yang bertujuan untuk memperdalam teori yang berkaitan pada penelitian ini sehingga dengan ilmu dan teori yang didapatkan dari pengkajian tersebut dapat membantu dalam proses penelitian ini. Pada Bab ini berisi kajian teori meliputi teori tentang klasifikasi teks, *dataset*, *Multinomial Naive Bayes*, Twitter, ujaran kebencian dan *confusion matrix*.

BAB III METODOLOGI PENELITIAN

Pada bab ini berisikan mengenai langkah-langkah dalam melakukan penelitian ini, dimulai dari melakukan perumusan masalah hingga dokumentasi hasil penelitian, serta mendeskripsikan alat dan bahan yang digunakan dalam penelitian ini.

BAB IV PEMBAHASAN

BAB IV menjabarkan mengenai pembahasan penelitian dimulai dengan membangun *dataset*, praproses dataset ekstraksi fitur, learning model dan perhitungan *confusion matrix* untuk menghitung tingkat *f-measure* dari penerapan algoritma *Multinomial Naive Bayes* pada kasus ujaran kebencian pada kicauan bahasa Indonesia dan diakhiri dengan analisis hasil konfigurasi penelitian yang menjelaskan mengenai mengapa penelitian tersebut menghasilkan hasil seperti itu dan membandingkan dengan hasil penelitian sebelumnya.

BAB V KESIMPULAN DAN SARAN

Bab V berisikan kesimpulan penelitian ini berdasarkan dari masalah dan tujuan yang sudah dideskripsikan pada Bab I diikuti dengan kelebihan dan kekurangan penelitian ini dan saran untuk menghindari melakukan kesalahan yang sama pada penelitian selanjutnya.