1. Pendahuluan

Dengan adanya situs jejaring sosial dan situs mikroblog, komunikasi antara orang-orang dari berbagai latar belakang budaya dan psikologis menjadi lebih langsung, mengakibatkan semakin banyak konflik “cyber” diantara para pengguna internet. Akibatnya, ungkapan kebencian semakin banyak, dan sampai ketitik dimana telah menjadi masalah serius untuk diperhatikan [1]. SARA (Suku, Agama, Ras dan Antar golongan) adalah setiap tindakan yang melibatkan kekerasan, diskriminasi, dan pelecehan yang didasarkan atas identitas diri. SARA merupakan cara yang sering digunakan dalam dunia politik yaitu cara untuk menjatuhkan saingan politik dari salah satu kubu seperti yang terjadi pada pemilu tahun 2019 kemarin, menurut [3] dia menyatakan bahwa sentimen isu SARA masih mengancam keberlangsungan Pemilihan Presiden 2019 disebabkan oleh persaingan ketat para kontestan partai politik untuk mendapatkan kursi kekuasaan.

Penelitian terkait ujaran kebencian yang cukup serupa dengan SARA telah dilakukan sebelumnya, penelitian terkait penyaringan ujaran kebencian pertama kali diusulkan oleh [2] pada tahun 2010, hanya saja dengan ruang lingkup yang sangat terbatas dan metode supervised yang sangat sederhana, dalam jurnal itu penulis mengusulkan suatu klasifikasi dengan metode semantic filtering atau menyaring suatu kata negatif dalam kalimat, Hanya saja jurnal tersebut tidak begitu relevan dengan apa yang akan di bahas dalam jurnal ini. Jurnal ini akan mengklasifikasikan suatu ujaran/komentar agar dapat dinyatakan sebagai ujaran SARA atau tidak.

Pada penelitian kali ini yang akan dilakukan adalah pengklasifikasian *tweet* SARA yang bertebaran di internet dengan menggunakan proses analisis sentimen. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau fitur/tingkat aspek dan menentukan apakah pendapat yang dikemukakan dalam dokumen, kalimat atau fitur entitas/aspek bersifat positif, negatif atau netral[4].

Pada jurnal ini akan dilakukan perbandingan antara dua metode klasifikasi yaitu *SVM (Support Vector Machine)* dan *Naïve Bayes* sehingga akan diketahui metode manakah yang dapat mengklasifikasikan ujaran SARA dengan lebih baik. Perbandingan dengan kasus yang berbeda pernah dilakukan sebelumnya sebagai contohnya adalah penelitian [4] dalam jurnal ini dilakukan perbandingan hasil dari 3 algoritma yaitu *Support Vector Machine (SVM)*, *Naïve Bayes* dan *Decision Tree*. Dalam percobaannya tersebut *SVM* terbukti memiliki performansi yang lebih baik dibanding 2 Algoritma lainnya dikarenakan memiliki nilai *Precision*, *Recall*, dan *F-measure* yang lebih tinggi. Penelitian lainnya dilakukan oleh A. Sopharak dkk pada tahun 2014 penelitian yang dilakukan adalah perbandingan terkait pendeteksian microaneurysm secara otomatis melalui foto pupil seseorang, pada penelitian ini metode yang digunakan adalah *SVM* dan *Naïve Bayes*, dan hasil yang didapat menunjukkan bahwa *SVM* memiliki tingkat akurasi yang lebih tinggi dibanding *Naïve Bayes* dalam pendeteksian microaneurysm yang mana merupakan tanda awal terjadinya retinopati diabetik yang dapat digunakan oleh dokter spesialis[19]. Selanjutnya penelitian yang dilakukan oleh [20] jurnal ini membandingkan metode *Naïve Bayes* dan *Support Vector Machine* untuk klasifikasi emosi pada teks bahasa Indonesia. Hasilnya *SVM* memperoleh hasil yang lebih baik dibandingkan *Naïve Bayes*. dan dari beberapa penelitian ini memunculkan pertanyaan terkait Perbandingan *Support Vector Machine* dan *Naïve Bayes* dalam Pengklasifikasian ujaran SARA di *Twitter*.

Seperti yang disebutkan diatas *Sentimen analysis* adalah proses yang dapat mengklasifikasikan ekspresi manusia kedalam beberapa bentuk seperti positif, negatif dan netral, banyak sumber yang bisa digunakan untuk *sentiment analysis* seperti foto, text, video, audio juga merupakan sumber atau topik yang bisa dibawa untuk *sentiment analysis* [31]. Alasan mengapa penulis menggunakan algoritman *SVM* dan *Naïve Bayes* sebagai algoritma yang digunakan dalam penelitian ini salah satunya didasari oleh beberapa penelitian sebelumnya yang melakukan riset yang sama seperti. Pada penelitian [31] dalam kasus “perbandingan *SVM* dan *Naïve Bayes* dalam *Review data set*”. penulis memperoleh hasil dimana *SVM* memperoleh hasil yang lebih tinggi dibanding *Naïve Bayes* yaitu dengan 82.48% untuk *SVM* dan 76.56% *Naïve Bayes* jenis *SVM* yang digunakan dalam penelitian ini adalah linear *SVM* dan Multinomial *Naïve Bayes* dikarenakan metode serta jenis metode yang digunakan sama, penulis memilih menggunakan jurnal ini sebagai referensi patokan. Lalu juga dalam penelitian yang dilakukan [32]. Dalam kasus ini penulis juga melakukan perbandingan antar *SVM* dan *Naïve Bayes* dalam kasus “*sentiment analysis terhadap candidate gubernur jawa barat periode 2018-2023* berdasarkan opini publik di *twitter* dalam kasus ini penulis menggunakan fitur *Crawling* kemudian *tweet* akan dipilih secara manual untuk mendapatkan *tweet* yang paling relevan untuk penelitian itu. Dan untuk mencari tingkat akurasinya penelitian ini menggunakan algoritma *Confusion matrix*. Dimana pada penelitian ini *SVM* memperoleh akurasi yang jauh lebih rendah dibanding *Naïve Bayes* yaitu 94% untuk *Naïve Bayes* dan 75.50% untuk *SVM*. Pada kasus tersebut kedua algoritma menggunakan *data set* yang sama akan tetapi *SVM* mendapatkan akurasi yang jauh lebih rendah. Berdasarkan 2 penelitian diatas penulis memutuskan menggunakan algoritma *SVM* dan *Naïve Bayes* dikarenakan dalam beberapa kasus yang sudah di *review* sebelumnya memperlihatkan *Naïve Bayes* memiliki akurasi yang lebih dibandingkan *SVM* dalam kasus yang cukup mirip. Untuk melihat seberapa baik kedua algoritma dalam kasus perbandingan yang dilakukan dalam penelitian kali ini.

Topik yang diangkat dari penulisan tugas akhir ini adalah perbandingan *Support Vector Machine* dan *Naïve Bayes* dalam pengklasifikasian ujaran SARA di *twitter*. untuk mengetahui perbandingan tingkat performansi

serta akurasi dari metode tersebut. Sedangkan batasannya adalah penelitian hanya mencakup ujaran SARA yang menggunakan Bahasa Indonesia.

Tujuan dari penelitian tugas akhir ini adalah melakukan perbandingan antara *SVM* dengan *Naïve Bayes* dalam pengklasifikasian ujaran SARA di *twitter*. Untuk mengetahui metode mana yang memberikan tingkat akurasi yang lebih baik dari dua metode tersebut. Dan juga membangun sistem yang dapat mengklasifikasikan ujaran *SARA* di *twitter* secara akurat sehingga kita dapat mengetahui kebiasaan orang Indonesia dalam memberikan tanggapan di media sosial. Juga mengetahui hasil yang lebih baik diantara proses klasifikasi yang menggunakan data *Realtime* dan data *Offline*