

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sosial media di Indonesia berkembang sangat pesat, menurut Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada tahun 2019 masyarakat Indonesia sebanyak kurang lebih 196 juta jiwa terhubung ke internet dan juga 95% aktivitas adalah membuka sosial media [1]. Dengan pengguna aktif tertinggi pada Facebook lalu disusul dengan Instagram dan yang setelahnya diduduki oleh Twitter.

Tetapi tidak seluruh pengguna sosial media, menggunakan teknologi ini dengan bijak. Tak sedikit pengguna melakukan tindakan negatif seperti penipuan, penyebaran berita bohong, menulis hal - hal yang cenderung mengandung ujaran kebencian, hingga perundungan secara daring (cyberbullying). Hal ini tentu menjadi akibat negatif dengan adanya sosial media pada masyarakat.

Dengan banyaknya pengguna sosial aktif maka akan ada banyak dampak yang ditimbulkan dan salah satunya ialah cyberbullying, pada data (APJII) sebanyak 51,2% pernah mengalami bullying yang mana angka tersebut cukup besar mengingat jumlah pengguna sosial media yang banyak [1]. Perundungan di media sosial dilakukan dengan menulis komentar ataupun dengan mengkritik melalui status yang menyinggung dengan menggunakan kata-kata hinaan unsur ras, agama dan juga suku.

Pada penelitian serupa sebelumnya hanya tertuju kepada model yang didesain khusus hanya untuk bagaimana pendeteksian ujaran kebencian bullying dengan satu metode saja. Pada penelitian kali ini berfokus kepada bagaimana pendeteksian ujaran kebencian berdasarkan akurasi dari dua metode yang berbeda yang akan dibandingkan. Dengan ini maka akan diperoleh hasil dari metode yang digunakan dan akan dilakukan analisa lebih lanjut untuk mengetahui akurasi dan kelayakan metode sebagai acuan penelitian selanjutnya.

Penelitian akan berfokus kepada metode Support Vector Machine (SVM) dan Random Forest (RF). Sesuai pada penelitian - penelitian serupa sebelumnya, belum ditemukan contoh yang didesain spesifik untuk mendeteksi atau mengklasifikasi cuitan yang bermakna cyberbullying dengan metode tersebut maka penelitian ini

bertujuan buat mewujudkan hal tersebut.

Pada penelitian sebelumnya yang dilakukan oleh Ahmad Fathan Hidayatullah yaitu mengidentifikasi cuitan kasar pada Twitter dengan metode atau algoritma Support Vector Machine (SVM) didapati keakuratan hingga 0.9928 yang menunjukkan angka tersebut sudah cukup tinggi dalam mengklasifikasi dan mendeteksi cuitan Twitter [2]. Penelitian serupa namun dengan metode yang berbeda dilakukan oleh Nassharah Abdulloh dengan penelitian yang berjudul deteksi cyberbullying pada sosial media Twitter menggunakan metode Multinomial Naïve Bayes 0.9616, Logistic Regression 0.9949, Support Vector Machine with linear kernel 0.9975, dan K-Nearest Neighbor (KNN) 0.9188, menurutnya metode tersebut memiliki keakuratan yang cukup tinggi pula dalam mengklasifikasikan cuitan Twitter [3].

Begitu juga dengan penelitian yang berkaitan klasifikasi teks dengan menggunakan metode Random Forest (RF) yang dipadukan dengan beberapa metode pendukung lainnya yaitu dengan KNN classifier dan F-Score dengan menggunakan dataset laman berita BBC News mendapatkan akurasi dari 90% sampai 96% akurat dengan klasifikasi beritanya [4].

Walaupun sudah banyak yang melakukan penelitian terhadap analisis sentimen namun masih sedikit yang meneliti tentang perbandingan antar metode yang sebenarnya cukup penting dalam penelitian lebih lanjut. Dengan demikian penelitian ini diharapkan dapat menjadi salah satu penelitian yang berkontribusi terhadap pengembangan ilmu pengetahuan tentang analisis sentimen.

1.2 Perumusan masalah

Berdasarkan penjabaran latar belakang, didapatkan rumusan masalah sebagai berikut:

1. Bagaimana mendapatkan hasil dari dua metode yaitu *Support Vector Machine* (SVM) dan *Random Forest* (RF).
2. Perolehan hasil performa perbandingan terhadap deteksi *cyberbullying* terhadap sosial media.
3. Evaluasi hasil dari perbandingan metode *Support Vector Machine* (SVM) dan *Random Forest* (RF).

1.3 Tujuan Penelitian

Maksud dan tujuan dari penelitian adalah sebagai berikut:

1. Mencari akurasi terbaik dari metode *Support Vector Machine* (SVM) dan *Random Forest* (RF).
2. Menganalisis parameter akurasi terbaik dari setiap metode.

1.4. Batasan Masalah

Adapun pembatasan dalam penelitian agar penelitian lebih terarah adalah sebagai berikut:

1. Dataset menggunakan data Indonesian Abusive and Hate Speech Twitter Text
2. Data yang digunakan adalah data tweet dalam bahasa Indonesia dengan jumlah 10.000 tweet.
3. Uji validitas dilakukan dengan membandingkan metode *Support Vector Machine* (SVM) dan *Random Forest* (RF).

1.5 Manfaat Penelitian

1. Menambah wawasan ilmu pengetahuan tentang analisis sentimen.
2. Mempermudah pencarian validasi akurasi metode.
3. Menambah pemahaman mengenai metode analisis sentimen.