

BAB I PENDAHULUAN

Pada bab ini dipaparkan apa yang akan dibahas mengenai permasalahan yang ditemukan.

1.1 Latar Belakang

Pada era sekarang penggunaan internet menjadi hal yang wajar, masyarakat sekarang lebih aktif bermain sosial media yang merupakan bagian dari internet. Aktivitas online telah menjadi sahabat karib yang tumbuh menjadi kekuatan besar dalam peluang berkelanjutan untuk bisnis, pengalihan perhatian, dan sebagainya. Media sosial salah satunya Twitter telah diubah menjadi mekanisme utama sebagai mitra online menerima tanggapan dari seluruh dunia [1]. Media sosial Twitter saat ini menjadi salah satu sarana pemasaran digital bagi organisasi atau perusahaan untuk mempromosikan produk dan layanan, serta dapat mendapatkan umpan balik dari pelanggan untuk mengembangkan bisnis mereka. Namun pada sebagian besar orang Indonesia aktif dan juga ekspresif melalui tweet mereka [2], yang kerap beropini mengandung sindiran pahit, celaan, dan bahkan dapat berupa kata-kata positif untuk mengungkapkan opini negatif terhadap suatu tujuan tertentu yaitu orang, organisasi, produk atau layanan.

Dalam bidang *Natural Language Processing* (NLP) mendeteksi sarkasme dalam sebuah kalimat atau pesan masih menjadi tantangan besar karena fitur leksikal (makna yang terdapat dalam kamus) yang didapat dari kalimat tersebut tidak memberikan informasi yang cukup untuk mendeteksi sarkasme, dimana salah satu jenis bahasa sarkasme yang digunakan pengguna untuk berbicara tentang sesuatu dapat berlawanan dengan maksud yang sebenarnya [2][3]. Misalnya: “Bekerja di hari libur adalah passion saya #sarkasme”. Dalam contoh ini, penulis mengatakan bahwa dia suka bekerja pada hari libur, tetapi karena hashtag (#) sarkasme, arti sebenarnya menunjukkan bahwa dia tidak suka bekerja pada hari libur. Biasanya tidak ada yang suka bekerja pada hari libur, sehingga sulit untuk mengidentifikasi sentimen semacam itu dengan menggunakan sistem analisis sentimen yang ada [4].

Beberapa penelitian telah dilakukan untuk menguji sarkasme dalam data tekstual. Penelitian Anukarsh G Prasad, dkk melakukan penelitian deteksi sarkasme

menggunakan beberapa metode klasifikasi seperti *Random Forest*, *Adaptive Boosting*, *Gaussian Naïve Bayes* dengan data tweet umum dengan label sarkastik dan non-sarkastik [5]. Hasilnya adaptive boosting menunjukkan akurasi yang baik dari pada metode klasifikasi yang lain. Penelitian Ashwitha A, Shruthi G, dkk mengusulkan kerangka kerja otomatis untuk mengidentifikasi komentar sarkastik dari Twitter menggunakan metode SVM dengan data terdiri dari Tweet dengan hashtag *#sarcasm* atau *#sarcastic* dan tweet yang memiliki arti positif atau negatif (misalnya *#happy*, *#sad*) [1], dan hasil yang diperoleh menunjukkan akurasi yang baik. Penelitian Dwi A. P. Rahayu dkk menggunakan 2 kombinasi berbeda untuk mendeteksi sarkasme, kombinasi pertama menggunakan metode *Bag of Words* dan *Naïve Bayes*, kombinasi kedua menggunakan *Interjection*, *punctuation*, TF-IDF dan KNN [2]. Data yang digunakan Tweet yang berisi trending topik dan secara manual diberikan label positif, sarkasme, netral, dan negatif. Dari dua kombinasi mendapatkan hasil yang berbeda, pada kombinasi pertama mendapatkan hasil yang kurang memuaskan dimana banyak kalimat sarkasme yang masih tergolong kalimat positif. Penelitian Mohd Suhairi Md Suhaimin, dkk mengusulkan deteksi sarkasme menggunakan SVM dengan data komentar di Facebook kemudian secara manual diberi label sarkasme atau tidak oleh tiga annotator [3], hasil penelitian ini menunjukkan akurasi yang baik.

Metode *Support Vector Machine* (SVM) mendapatkan hasil yang baik untuk deteksi sarkasme, dan SVM memiliki kinerja yang bagus ketika kumpulan data tertentu memiliki banyak atribut seperti klasifikasi teks, sesuai yang diuraikan oleh Ashwitha A, Shruthi G, dkk [1]. Oleh karena itu, pada penelitian ini akan menggunakan metode *Support Vector Machine* untuk klasifikasi sarkasme pada pesan tweet berbahasa Indonesia. Dengan ditambahkan metode *Adaptive Boosting* (*Adaboost*) sebagai metode *Ensemble* untuk mengoptimalkan kinerja klasifikasi, karena menurut Anukarsh G Prasad, dkk *Adaptive Boosting* bekerja dengan menetapkan bobot dalam pelatihan dan memperbaikinya, karena item data yang kuat dapat mempengaruhi klasifikasi menuju akurasi yang lebih baik [5].

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka dapat diperoleh perumusan masalah sebagai berikut:

1.2.1 Permasalahan

Di Indonesia sebagian besar orang menyampaikan sebuah opini mereka dalam tweet Twitter untuk menanggapi sebuah produk dan layanan yang sedang mereka gunakan, salah satunya adalah layanan internet yang telah menjadi kebutuhan sehari-hari. Orang Indonesia kerap menggunakan kalimat sarkasme dalam menyampaikan opini, sehingga dengan adanya sarkasme dapat membuat pembaca salah dalam memahami opini tersebut. Oleh karena itu, diperlukan sebuah pendekatan untuk menganalisis opini yang mengandung sarkasme atau bukan sarkasme.

1.2.2 Metode Usulan

Dari permasalahan yang telah dipaparkan maka akan diusulkan pada penelitian ini sebuah sistem klasifikasi untuk mengelompokkan kalimat sarkasme atau bukan sarkasme pada tweet. Klasifikasi sarkasme yang diusulkan pada penelitian ini akan menggunakan metode *Support Vector Machine Ensemble*.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan dan metode yang telah dipaparkan maka didapatkan pertanyaan penelitian sebagai berikut ini:

1. Berapa akurasi yang dihasilkan oleh metode *Support Vector Machine Ensemble* pada proses deteksi sarkasme?
2. Apakah dengan adanya penambahan metode *Ensemble*, terdapat peningkatan akurasi dalam proses deteksi sarkasme?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

1. Dapat mengklasifikasikan teks sebagai sarkasme atau bukan sarkasme dari tweet berbahasa Indonesia mengenai masalah jaringan Internet menggunakan metode *Support Vector Machine Ensemble*.
2. Mengetahui nilai akurasi deteksi sarkasme menggunakan metode *Support Vector Machine Ensemble*.

1.3.2 Manfaat

Dapat membantu pembaca untuk mengurangi kesalahan dalam pemahaman kalimat yang mengandung sarkasme dan bukan sarkasme.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Proses klasifikasi hanya menggunakan data teks berbahasa Indonesia.
2. Data tweet yang dikumpulkan secara manual, mengenai tanggapan para pengguna layanan provider di Indonesia di Twitter mengenai masalah jaringan internet.
3. Dataset berjumlah 1374 data teks dan pelabelan data dilakukan secara mandiri.
4. Menggunakan metode *Support Vector Machine Ensemble* untuk klasifikasi.

1.5 Sistematika Laporan

BAB 1 **PENDAHULUAN**

Pada bab ini dipaparkan apa yang akan dibahas mengenai permasalahan yang ditemukan. Latar belakang, perumusan masalah, tujuan dan manfaat, batasan masalah, dan sistematika penulisan.

BAB II **KAJIAN PUSTAKA**

Pada tahap ini dipaparkan mengenai tinjauan kembali teori yang mendukung dengan permasalahan yang ditemukan.

BAB III **METODE USULAN**

Pada bab ini dipaparkan metode penelitian yang dilakukan untuk menyelesaikan penelitian ini. Metodologi penelitian yang terbagi ke dalam beberapa tahapan yaitu membangun algoritma klasifikasi, rancangan sistem (Input, Proses, Ouput), data penelitian, serta rencana pengujian.

BAB IV **HASIL DAN PEMBAHASAN**

Pada bab ini dipaparkan berbagai ujicoba saat penelitian dilakukan dan pembahasan mengenai hasil pengujian.

BAB V PENUTUP

Pada bab ini dipaparkan kesimpulan dari penelitian yang telah dilakukan sesuai pertanyaan penelitian, dan saran untuk penelitian berikutnya.

DAFTAR Pada bab ini dipaparkan daftar jurnal yang dijadikan acuan
PUSTAKA dalam penelitian ini.