

BAB I PENDAHULUAN

1.1 Latar Belakang

Konektivitas antara manusia di era modern merupakan sesuatu yang tersedia secara luas dengan adanya perkembangan teknologi berupa internet dan media sosial [1]. Diperkirakan bahwa dari seluruh populasi manusia sebanyak 7,81 miliar, 4,66 miliar memiliki akses internet dan 4,14 miliar lainnya merupakan pengguna media sosial yang aktif. Khususnya di Indonesia, [2] mencatat bahwa sebanyak 196,71 juta jiwa penduduk di Indonesia memiliki akses internet, dengan penetrasi total sebesar 73,7%. Ketersambungan manusia dalam skala sedemikian besar yang difasilitasi kedua teknologi tersebut, khususnya media sosial, dapat menjadi peluang untuk memajukan aktivitas bisnis dan komersial.

Persebaran paparan suatu merek dagang di media sosial berkorelasi dengan jumlah penjualan merek dagang tersebut, dan media sosial juga mempermudah interaksi antara konsumen dan perusahaan [3], [4]. Media sosial juga memberikan kesempatan bagi konsumen yang memiliki pengalaman positif dengan suatu merek dagang untuk menyebarkan promosi merek tersebut berupa *e-word of mouth* (e-WOM) secara organik [5]. Penelitian terdahulu [6], [7] mencatat bahwa sedang terjadi peralihan tren di bidang pemasaran, yang beralih dari pendekatan tradisional menuju pendekatan pribadi di ruang media sosial. Oleh karena itu, penting bagi suatu perusahaan komersial untuk bisa menetapkan eksistensi media sosial yang signifikan. Keberadaan media sosial suatu yang demikian tidak hanya berkaitan dengan pemasaran, tetapi juga penanganan umpan balik dari konsumen terkait produk yang dipasarkan oleh perusahaan.

Salah satu *platform* media sosial yang paling banyak digunakan khususnya di Indonesia adalah Twitter [1]. Menemukan bahwa dari seluruh pengguna internet di Indonesia, jumlah pengguna yang juga aktif di media

sosial mencapai 170 juta jiwa, dan sebanyak 63,6% diantaranya aktif menggunakan Twitter. Sebagai situs media sosial berbasis *microblogging*, Twitter banyak dimanfaatkan oleh pelaku sektor bisnis untuk sarana komunikasi dan layanan pelanggan [8] dan dapat digunakan untuk menyebarkan eksistensi *brand* produk serta menggali respon konsumen secara langsung.

Literatur ilmiah terdahulu telah banyak meneliti pendekatan penambangan data di Twitter untuk berbagai topik, seperti politik [9][10], Kesehatan [11] dan pendidikan [12], [13]. Ditinjau dari teknik yang digunakan, beberapa penelitian tersebut [9], [13] melakukan *sentiment analysis* sebagai bagian dari penelitian mereka. *Sentiment analysis*, sebagaimana didefinisikan oleh [14], adalah tugas (*task*) dalam penambangan data yang bertujuan untuk mengelompokkan data dalam bentuk kalimat berbahasa tertentu ke dalam kelas positif, negatif, atau netral. Beberapa teknik yang dapat digunakan untuk melakukan *sentiment analysis* meliputi random forest (RF) dan support vector machine; kedua teknik ini telah dibandingkan dalam penelitian terdahulu ketika diterapkan pada konteks deteksi nada suatu kalimat tertulis [15]. Beberapa penelitian terdahulu lainnya [16]-[17]-[18] membahas penambangan data di Twitter dari segi yang berbeda, yaitu pelaksanaan topic modelling, yang didefinisikan oleh [18] sebagai tugas (*task*) dalam penambangan data untuk mengelompokkan data leksikal ke dalam golongan-golongan tertentu yang disesuaikan dengan suatu kategori besar atau topik.

Khususnya Twitter, telah menjadi sarana utama bagi interaksi antara perusahaan komersial dan konsumen. Pertumbuhan pesat penggunaan media sosial telah membuka peluang baru bagi bisnis dalam pemasaran dan manajemen merek. Namun, di tengah dinamika ini, memahami sentimen konsumen secara efektif menjadi tantangan yang signifikan. Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi secara mendalam respons konsumen terhadap merek perusahaan publik di media sosial, khususnya di

platform Twitter. Tujuan utamanya adalah membandingkan kinerja berbagai algoritma klasifikasi untuk analisis sentimen dan memberikan wawasan tentang bagaimana perusahaan dapat meningkatkan kehadiran mereka di media sosial serta merespons dengan cepat dan efektif terhadap umpan balik konsumen. Penelitian sebelumnya telah memberikan berbagai wawasan tentang analisis sentimen di Twitter menggunakan berbagai pendekatan. Telah teliti studi oleh [19] membandingkan metode Naïve Bayes dan Support Vector Machine (SVM) dalam analisis sentimen terhadap program acara di SCTV, mendapatkan labeling terbaik dengan labeling otomatis dan untuk akurasi terbaik yang didapatkan dengan menggunakan metode SVM seluruh program acara akurasi 88,57%, akurasi berita 79,81% Entertainments 89,80% Sinetron 73,68% dan FTV 87,74%. Di sisi lain, penelitian yang dilakukan oleh [20] menggunakan SVM dan Naïve Bayes berbasis Particle Swarm Optimization untuk menganalisis sentimen terkait topik yang berbeda dan didapatkan hasil pengklasifikasikan sentimen positif dan negatif dengan baik, dengan akurasi 94,81 dan AUC 0,974. Seiring dengan konteks yang beragam ini, penelitian ini akan mengeksplorasi kinerja algoritma Random Forest dan SVM dalam menganalisis sentimen terhadap merek perusahaan, dengan fokus pada ulasan dan umpan balik konsumen di platform Twitter. Dengan demikian, diharapkan penelitian ini akan memberikan kontribusi yang berharga tidak hanya bagi praktisi bisnis dalam meningkatkan strategi pemasaran dan manajemen merek mereka di era media sosial, tetapi juga bagi penelitian akademis yang berkelanjutan dalam pemahaman sentimen konsumen dalam konteks yang terus berkembang ini.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, dapat disimpulkan permasalahan yang akan diselesaikan dalam penelitian ini adalah berapa nilai performa (*Accuracy*, *Precision*, *Recall*, dan *F1 Score*) yang dihasilkan oleh model klasifikasi pada perbandingan model, *Random Forest* dan *Support Vector Machine*?

1.3 Batasan Masalah

Batasan masalah dari penelitian ini adalah,

1. *Data tweet* aduan yang digunakan dalam penelitian ini akan dikumpulkan melalui API Twitter dengan menggunakan kata kunci terkait PT Garuda Indonesia. Hanya *tweet* aduan yang diakses secara publik yang akan digunakan dalam analisis.
2. Penelitian ini akan menggunakan *sentimen analisis* dengan dua metode klasifikasi, yaitu *Random Forest* dan *Support Vector Machine*.
3. Penelitian ini akan membatasi pengumpulan data dan analisis pada *tweet* aduan terhadap PT Garuda Indonesia yang di posting dalam periode satu tahun terakhir.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah,

1. Membandingkan metode klasifikasi *Random Forest* dan *Support Vector Machine* dalam analisis sentimen untuk menemukan metode dengan tingkat akurasi terbaik dalam mendeteksi *tweet* aduan terhadap PT Garuda Indonesia.
2. Mengidentifikasi topik yang paling penting dalam data media sosial PT Garuda Indonesia, seperti permasalahan pelayanan, kepuasan pelanggan, atau promosi produk.

1.5 Manfaat Penelitian

Ditinjau dari sisi keilmuan, hasil penelitian ini diharapkan dapat menjadi referensi untuk penelitian di masa depan yang berfokus pada penerapan teknik penambangan data dalam konteks analitik media sosial.

1.6 Sistematika Penulisan

1. Bab 1 menjelaskan latar belakang penelitian, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian.

2. Bab 2 berisi studi literatur yang dilakukan untuk menunjukkan dasar akademis penelitian, baik berupa teori-teori dasar yang menjadi landasan maupun kajian pustaka dari penelitian-penelitian relevan yang sebelumnya, dan menjelaskan hipotesis/pertanyaan penelitian.
3. Bab 3 menjelaskan objek penelitian, alur penelitian serta alat dan bahan.
4. Bab 4 berisi laporan implementasi penelitian.
5. Bab 5 berisi hasil penelitian dan saran.

