

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bantuan Sosial COVID-19 merupakan bentuk tanggung jawab negara kepada masyarakat terdampak untuk setidaknya mengantisipasi permasalahan sosial ekonomi yang berkaitan dengan COVID-19. Hanya saja, penyaluran bantuan sosial memiliki terlalu banyak regulasi dan peraturan yang tidak harmonis dari pemerintah. Bantuan sosial yang diberikan menjadi terkesan membingungkan, waktu penerimaan yang cukup lama, dan diterima oleh masyarakat kurang tepat sasaran [1]. Permasalahan tersebut kemudian terungkap dan sempat ramai diperbincangkan di sosial media semenjak Desember 2020 yang membahas tentang kasus korupsi Bansos COVID-19 [2]. Dari permasalahan yang dialami, pemerintah terus berupaya untuk melakukan perbaikan. Seperti percepatan penyaluran, perbaikan data, dan cara pengalokasian. Salah satu hal yang diupayakan pada penyaluran bansos adalah percepatan pendistribusian sebelum hari raya Idul Fitri 1432 H [3].

Selama rentang lima bulan dari Desember 2020 pemerintah terus berupaya untuk menyalurkan bantuan sosial kepada mereka yang terdampak. Terdapat berbagai macam tanggapan dari masyarakat tentang program bantuan sosial, misalnya trending #BangkitnyaKKN pada tanggal 25 Januari 2021 yang mengungkit kembali masalah korupsi bantuan [4]. Dari tanggapan itulah bisa didapatkan berbagai macam informasi serta penilaian mengenai program bantuan sosial.

Salah satu cara masyarakat mengemukakan pendapat atau opini adalah melalui *twitter*. Laporan terbaru dari *We Are Social* dan *Hootsuite* pada Januari 2021 memperlihatkan potensi didapatkannya informasi dari data *twitter*. Dilaporkan bahwa *twitter* menjadi salah satu sosial media di Indonesia yang memiliki pengguna aktif terbanyak (63,6%) setiap bulannya dengan rentang usia 16-64 tahun. Usia tersebut merupakan rentang dimana seseorang dianggap dewasa dan memiliki kemampuan untuk mengamati dan menanggapi isu yang sedang terjadi [5]. Luasnya data dari *twitter* juga dapat menunjang akurasi dari algoritma yang akan digunakan.

Ekstraksi opini dari *twitter* membutuhkan ketelitian untuk menghasilkan ketepatan. Hal tersebut dikarenakan hasil penambangan data *twitter* memiliki struktur yang tidak rapi, dan seringkali ditulis dengan bahasa yang tidak sesuai dengan standar. Metode yang dapat digunakan untuk dapat mengklasifikasikan opini adalah analisis sentimen [6]. Pada penelitian sebelumnya, analisis sentimen mengenai wacana pemindahan Ibu Kota Indonesia berhasil memberikan persentase hasil pro dan kontra dari tanggapan masyarakat. Jumlah pro yang menjadi mayoritas pada penelitian ini bisa dipertimbangkan karena data langsung dari komentar masyarakat ketika isu tersebut sempat menjadi *trending topic*. Dengan kata lain tanggapan dari masyarakat bagi suatu kebijakan penting untuk pemerintah mengambil keputusan. Hasil analisis sentimen ini mengambil jenis kata dengan mayoritas kata kerja dan kata sifat [7]. Begitu pula dengan penelitian analisis sentimen yang dilakukan, mengangkat isu bantuan sosial dengan hubungan kata kerja dari upaya pemerintah menanggulangi masalah, dan kata sifat yang menjadi

penghubung dalam menentukan sentimen. Selain itu penelitian ini mengambil momen ketika isu bantuan sosial hangat diperbincangkan.

Dibutuhkan algoritma untuk mengklasifikasikan penilaian mengenai topik yang dibahas. Klasifikasi sentimen akan dilakukan dengan menggunakan Algoritma Random Forest. Penelitian terdahulu menunjukkan bahwa Random Forest merupakan algoritma yang baik untuk klasifikasi multi kelas. Selain sentimen multi kelas, sentimen dengan pernyataan kondisional dapat diidentifikasi dengan menambahkan beberapa aspek pada data. Diperlukan data yang lebih luas dalam implementasi Algoritma Random Forest [8]. Hal tersebut dikarenakan, *splitting* yang dilakukan Random Forest ketika membuat pohon tidak mengidentifikasi semua variabel, hanya memilih sebagian variabel untuk mendapatkan hasil yang terbaik dari banyaknya data. Ketika pengecekan variabel lebih sedikit maka korelasi dari setiap pohon akan cenderung lebih kecil. Korelasi yang kecil itu akan meningkatkan hasil performa klasifikasi sentimen [9].

Performa yang tinggi mengindikasikan bahwa klasifikasi dan penyimpulan data menjadi informasi dapat dilakukan dengan baik. Pada penelitian analisis sentimen sebelumnya, evaluasi performa klasifikasi model berhasil dilakukan dengan mengukur *accuracy*, *precision*, *recall* dan *f1-score*. Dengan performa tersebut didapatkan analisis tipe error dari perbandingan antara hasil klasifikasi algoritma dan klasifikasi yang sebenarnya [10].

Setiap data yang diolah membutuhkan metode pengelolaan yang baik. CRISP-DM merupakan salah satu metode untuk mengekstrak data menjadi sebuah pengetahuan. Berdasarkan banyak survey dan polling bahwa standar untuk

pengembangan data mining adalah metode CRISP-DM [11]. Metode CRISP-DM dalam implementasinya pada domain ilmu sosial dapat meningkatkan produktivitas dari segi waktu dan kualitas penelitian. Ekstraksi pengetahuan yang kompleks menghasilkan karakteristik yang unik [12]. Domain tersebut berkaitan dengan domain penelitian ini yaitu ilmu sosial dan ekonomi.

Hasil analisis tanggapan masyarakat mengenai bantuan sosial akan berupa penilaian tendensi netral, positif, dan negatif. Penilaian dari opini tersebut ditentukan dengan melihat nilai persentase sentimen tertinggi. Dari masing-masing nilai tendensi, akan didapatkan topik-topik yang bisa dievaluasi atau dipertimbangkan untuk dikaji dalam hal bantuan sosial. Dianalisis pula setiap faktor yang mempengaruhi hasil yang disajikan. Termasuk evaluasi algoritma Random Forest pada analisis sentimen. Evaluasi tersebut dapat menjadi referensi implementasi atau pengembangan dari metode yang digunakan. Berdasarkan pemaparan tersebut maka dilakukanlah penelitian dengan judul “ANALISIS SENTIMEN SEBAGAI PENILAIAN OPINI TERHADAP POLEMIK BANTUAN SOSIAL COVID-19 MENGGUNAKAN ALGORITMA RANDOM FOREST”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan maka perumusan masalah dari penelitian ini dapat diuraikan sebagai berikut:

1. Bagaimana hasil analisis sentimen sebagai penilaian terhadap bantuan sosial COVID-19 dengan menggunakan algoritma Random Forest melalui akumulasi persentase sentimen netral, negatif, dan positif ?

2. Apa saja topik yang dibahas pada sentimen netral, negatif, dan positif dari hasil analisis sentimen bantuan sosial COVID-19?
3. Bagaimana performa *accuracy*, *precision*, *recall* dan *f1-score* algoritma Random Forest untuk analisis sentimen terhadap bantuan sosial COVID-19?

1.3 Tujuan

Tujuan dari penelitian sentimen analisis menggunakan algoritma Random Forest mengenai analisis sentimen terhadap Bantuan Sosial COVID-19 adalah sebagai berikut:

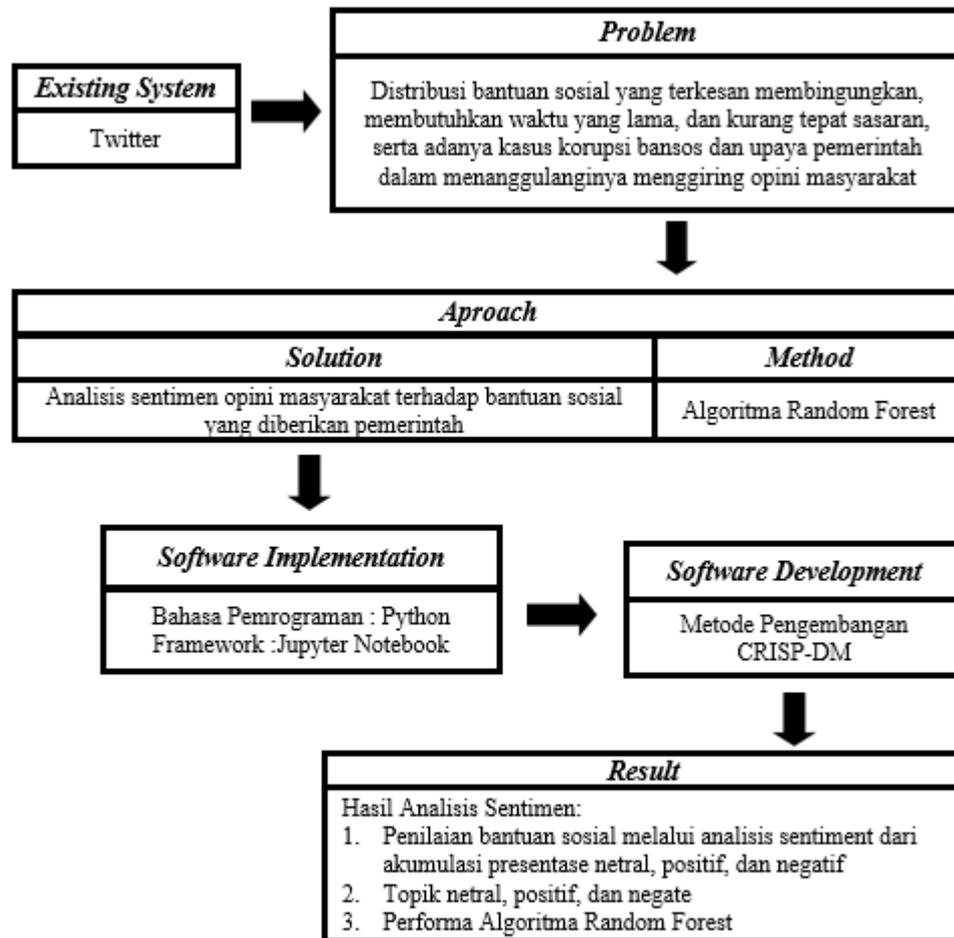
1. Mengetahui persentase netral, positif, dan negatif analisis sentimen terhadap bantuan sosial COVID-19 sebagai kriteria penilaian dan menganalisis faktor dari hasil tersebut.
2. Membahas topik netral, negatif, dan positif analisis sentimen bantuan sosial COVID-19 mulai dari proses penambangan data sampai hasil pengklasifikasian.
3. Mengetahui performa *accuracy*, *precision*, *recall* dan *f1-score* algoritma Random Forest dalam mengklasifikasikan sentimen.

1.4 Batasan Masalah

Ditetapkan beberapa batasan masalah pada penelitian yang dilakukan agar hasil sesuai dengan perencanaan. Batasan masalah tersebut meliputi:

1. Metode yang digunakan untuk melakukan klasifikasi adalah Random Forest dengan parameter dari *sklearn*.
2. Labelling yang dilakukan hanya berdasarkan penilaian otomatis dari mesin tanpa labelling manual.
3. Bahasa pemrograman yang digunakan adalah Python dengan Framework Jupyter Notebook.
4. Periode pengambilan data dimulai dari tanggal 5 Mei 2021 sampai 20 Mei 2021.
5. *Keyword* yang digunakan adalah 'bansos'.
6. Data yang akan diambil setiap harinya minimal berjumlah 500 dengan limit 10,000.
7. Data yang dianalisis berbahasa Indonesia.
8. Analisis Sentimen ini mengklasifikasi teks menjadi netral, negatif, dan positif.
9. Hasil analisis sentimen tidak diimplementasikan dalam bentuk aplikasi.

1.5. Kerangka Pemikiran



Gambar 1. 1 Kerangka Pemikiran

Kerangka pemikiran pada Gambar 1.1 diawali dengan isu bantuan sosial hangat diperbincangkan di *twitter* pada bulan Desember 2020 lalu karena adanya kasus korupsi bantuan sosial COVID-19. Pemerintah pun berusaha menanggulangi permasalahan tersebut. Topik yang didapatkan dari sosial media menghasilkan teks yang tidak terstruktur, oleh karena itu dibutuhkan proses pengelolaan data untuk menghasilkan analisis sentimen. Dibutuhkan pula algoritma untuk memodelkan data tersebut algoritma yang digunakan pada penelitian ini adalah Random Forest, dan untuk setiap langkah dari pemrosesan data digunakan metode pengembangan

CRISP-DM. Hasil dari analisis sentimen ini berupa analisis sentimen netral, positif, dan negatif dengan mengetahui persentase dari keduanya serta topik dari sentimen tersebut yang bisa menjadi acuan untuk mengetahui hal-hal yang perlu diperhatikan pada pembahasan bantuan sosial. Pembahasan performa dari algoritma yang digunakan untuk pertimbangan dan perbaikan pada penelitian selanjutnya dengan algoritma yang sama.

1.6. Sistematika Penulisan

Penulisan dari informasi yang didapat disusun dalam bentuk laporan tugas akhir dengan uraian sebagai berikut:

BAB I PENDAHULUAN

Bahasan dari Bab ini terdiri dari Latar Belakang, Rumusan Masalah, Tujuan dari penelitian, Batasan Masalah, Kerangka Pemikiran dan Sistematika Penulisan.

BAB II KAJIAN PUSTAKA

Kajian pustaka ini membahas penelitian terdahulu, konsep-konsep, teori-teori, model, rumus utama dan turunannya yang berhubungan dengan penelitian yang dilakukan. Hal-hal tersebut akan membantu dalam pemecahan masalah serta acuan agar penelitian lebih terarah.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan langkah-langkah dan teknik yang dilakukan dalam penelitian. Metodologi ini dibahas sesuai dengan kebutuhan pada penelitian.

BAB IV HASIL DAN PEMBAHASAN

Temuan dari tahapan penelitian dipaparkan pada bagian hasil. Setiap hasil yang dipaparkan kemudian akan dibahas berurutan sesuai dengan rumusan masalah.

BAB V KESIMPULAN DAN SARAN

Kesimpulan dari bab ini membahas jawaban dari pertanyaan pada rumusan masalah. Bab ini pun memuat saran dari hal-hal utama sebagai penunjang peningkatan kualitas untuk penelitian selanjutnya.

DAFTAR PUSTAKA

Bagian ini berisi sumber-sumber yang digunakan untuk penelitian.

LAMPIRAN

Memuat dokumen-dokumen tambahan yang terdapat pada proses penyusunan dan perancangan dimuat pada lampiran.

