

BAB I

PENDAHULUAN

1.1. Latar Belakang

Saat ini, internet telah mengubah cara orang mengekspresikan pandangan dan pendapat melalui posting blog, forum online, dan media sosial. Jutaan orang menggunakan jaringan sosial seperti Facebook, Twitter, Instagram, dan lain-lain untuk mengekspresikan emosi, pendapat, dan berbagi pandangan tentang kehidupan sehari-hari. Media sosial menghasilkan sejumlah besar data sentimen dalam bentuk tweet, pembaruan status, *posting blog*, komentar, ulasan, dan lain-lain (Kharde, 2016).

Jutaan pengguna berbagi pendapat dan mengekspresikannya menggunakan *micro-blogging* setiap hari. Twitter adalah situs *micro-blogging* yang sangat populer dimana pengguna dibatasi penggunaan karakter menjadi 140 karakter; pembatasan semacam ini membuat penggunaannya menjadi ringkas sekaligus ekspresif (Huq, 2017). Seiring berjalannya waktu, tidak dipungkiri bahwa Twitter sebagai *platform* komunikasi semakin menyusup ke dalam kehidupan sehari-hari. Setiap harinya, sekitar 554,7 juta orang di dunia menggunakan layanan tersebut dan mem-*posting* 58 juta “tweet” setiap harinya (Statistic Brain, 2013).

Seperti yang dijelaskan sebelumnya, Twitter adalah salah satu media sosial penyebaran informasi tercepat. Banyak hal yang bisa disampaikan melalui situs tersebut, seperti keluhan pengguna Twitter terhadap dunia asli, dan informasi tentang masalah-masalah yang terjadi pada saat itu. Tidak jarang banyak berita yang ramai dibicarakan di Twitter namun tidak diangkat di televisi maupun surat kabar. Salah satu contohnya yaitu informasi tentang penyebaran virus 2019 Novel Coronavirus yang dapat mengakibatkan penyakit Covid-19 di seluruh dunia, terutama pada Indonesia.

Pertama kali penyebaran 2019 Novel Coronavirus itu terdeteksi muncul pertama kali di Kota Wuhan, Tiongkok pada akhir 2019 lalu. Virus ini pertama kali

muncul di pasar hewan dan makanan laut di Kota Wuhan dan yang jatuh sakit akibat virus ini adalah seorang pedagang dari pasar tersebut (Febrian, 2020). Pada kasus pertama, Covid-19 yang diamati di Cina cukup parah untuk mendapatkan perhatian medis dan menghasilkan pengujian, tetapi jumlah total orang yang terinfeksi telah sulit dipahami. Rasio fatalitas kasus yang diperkirakan di antara pasien yang dirawat secara medis sejauh ini adalah sekitar 2%, tetapi rasio sebenarnya belum diketahui untuk beberapa waktu kedepan (Marc Lipstich, 2020)

Pada Indonesia, kasus ini pertama kali ditemukan pada dua warga Depok, Jawa Barat awal maret tahun 2020. Dari awal pandemi sampai saat ini, kurva dari penyebaran corona ini masih saja naik, terlihat dari jumlah warga yang dinyatakan positif terkena virus tersebut sudah mencapai 965.283. Dari jumlah kasus yang sudah terjadi, diantaranya ada 781.147 yang sembuh dan 27.453 yang telah meninggal dari kasus ini (Covid-19, 2021). Cepat terjadinya penyebaran virus ini terjadi karena masih banyak warga yang melakukan aktifitas di luar rumah dan masih kurang dalam menaati protokol kesehatan.

Pandemi global semacam ini menimbulkan berbagai reaksi dari orang-orang yang bergantung pada orientasi budaya. Komunikasi media mungkin tampak akurat dan efektif untuk menginformasikan kepada publik. Namun, itu juga dapat memberikan informasi yang salah dan berkontribusi pada kepanikan publik yang tidak perlu dan mengakibatkan tanggapan yang tidak diinginkan. Kecemasan sosial masing-masing berkaitan dengan masalah dasar sedangkan kepanikan moral, bersama dengan kesadaran risiko industri, sebagian besar dimasukkan dalam wacana keselamatan. Salah satu penyebab kepanikan ini adalah karena informasi dari para ahli tidak cukup disebarluaskan untuk menjangkau masyarakat pada waktu yang tepat. Demikian pula, kurangnya informasi dan informasi yang salah, yang sering kali dipicu oleh berita utama dan fokus media populer yang berlebihan, telah terbukti memperkuat ketakutan dan fobia terkait kesehatan (Nicomedes, 2020)

Di twitter banyak sekali opini yang disampaikan terkait dampak dari pandemi Covid-19 pada saat hari-hari besar dan *event* tertentu terutama di kota-kota besar

seperti Pilkada 2020, Hari Natal 2020 dan Tahun Baru 2021, Vaksinasi Pertama di Indonesia dan juga pada saat Hari raya Idul Fitri. Beberapa dampak umum pada pandemi diantaranya adalah frustasi, kebosanan, kecemasan dan ketakutan yang tidak terkontrol terkait Covid-19 ini (Serafini, et al., 2020). Opini yang disampaikan nanti akan dianalisa berupa analisis sentimen. Analisis sentimen dapat didefinisikan sebagai proses yang mengotomatiskan penambangan sikap, pendapat, pandangan dan emosi dari teks, ucapan, tweet, dan sumber basis data melalui Natural Language Processing (NLP). Analisis sentimen melibatkan pengelompokan pendapat dalam teks ke dalam kategori seperti "positif", "negatif" atau "netral". (Kharde, 2016).

Dalam mengelompokkan pendapat ke dalam kategori "positif", "negatif" atau "netral" dibutuhkan bantuan dari sistem *Indonesia Sentiment Lexicon* (InSet). Selain itu diperlukan juga algoritma analisis sentiment untuk menganalisis sentimen tersebut yaitu *Sentiment Classification Algorithm* berdasarkan *K-Nearest Neighbor* dan *Support vector Machine*. Kedua algoritma ini bekerja dengan dataset dan fitur yang sama. Pada penelitian Aulianita (2016), peneliti mengkomparasikan analisis sentimen dari review kamera. Uji coba dari pengujian data training pada penelitian tersebut menghasilkan akurasi dari K-NN sebesar 79% dan untuk akurasi dari SVM yaitu 72%. Selain itu pada penelitian Lidya dkk (2015), mereka mengkomparasikan teks analisis sentimen pada teks bahasa Indonesia menggunakan metode yang sama. Dalam uji coba tersebut, mereka menggunakan beberapa keyword, salah satunya yaitu pemilu 2014. Akurasi yang didapatkan pada K-NN menghasilkan 60,3% dan pada SVM menghasilkan 66.9%. Pada penelitian Baita dkk (2021), mereka mengkomparasikan teks analisis sentimen mengenai Vaksin Sinovac menggunakan metode yang sama. Akurasi yang didapatkan pada K-NN menghasilkan 56% dan pada SVM menghasilkan 70%. Algoritma SVM yang digunakan adalah kernel linear. Kernel ini memiliki keunggulan dalam analisis teks tetapi sulit pada jumlah data berskala besar. Sedangkan pada K-NN perlu menentukan nilai dari parameter k. Karena hasil dari dua penelitian itu berbeda, maka saya ingin meneliti lagi dari kedua algoritma tersebut.

Tahun 2020 merupakan tahun penyebaran virus mematikan seperti Covid-19 terjadi pertama kali di dunia termasuk Indonesia. Jika pada suatu saat terjadi penyebaran penyakit yang mengarah ke kematian, masyarakat dapat mengantisipasi kejadian ini dan sebagai bahan evaluasi untuk pemerintah setempat untuk menghadapi kejadian pandemi ini. Untuk memudahkan pengerjaan, dibutuhkan analisa yang dapat melakukan pengklasifikasian opini berupa teks sentimen pada Twitter. Penelitian ini melakukan Analisis Sentimen terhadap Pandemi Covid-19 di Indonesia menggunakan *K-Nearest Neighbor* dan *Support Vector Machine*.

1.2. Rumusan Masalah

Dari latar belakang diatas, maka rumusan masalah yang diangkat sebagai berikut:

- a. Bagaimana mengklasifikasikan dan menganalisa pola sentimen masyarakat Indonesia terhadap pandemi Covid-19 pada saat *event* tertentu menggunakan *Text Mining*?
- b. Bagaimana hasil perbandingan algoritma untuk mengklasifikasikan sentimen menggunakan *K-Nearest Neighbor* dan *Support Vector Machine*?

1.3. Tujuan Penelitian

Dari rumusan masalah diatas, tujuan dari penelitian ini sebagai berikut:

- a. Mengklasifikasikan dan menganalisa pola sentimen masyarakat Indonesia terhadap pandemi Covid-19 pada saat *event* tertentu.
- b. Mendapatkan hasil perbandingan algoritma untuk mengklasifikasikan sentimen menggunakan *K-Nearest Neighbor* dan *Support Vector Machine*.

1.4. Manfaat Penelitian

Adapun manfaat penelitian ini sebagai berikut:

- a. Sebagai bahan evaluasi pada masyarakat Indonesia dan pemerintahan untuk menangani pandemi Covid-19.

- b. Mengetahui respon dan pola sentimen masyarakat Indonesia pada dampak pandemi pada saat *event* tertentu.
- c. Mengetahui algoritma yang terbaik dalam pengklasifikasian respon dampak pandemi Covid-19.

1.5. Batasan Masalah

Dalam penelitian ini, batasan masalah yang digunakan sebagai berikut:

- a. Data yang digunakan dalam penelitian adalah data twitter yang berbahasa Indonesia.
- b. *Event* yang telah ditentukan adalah Pilkada 2020, Hari Natal 2020 dan Tahun Baru 2021, Vaksinasi Pertama di Indonesia dan Hari raya Idul Fitri.
- c. Data yang diambil dari Bulan November 2020 – Juni 2021 sesuai *event* yang telah ditentukan.
- d. Data yang diambil berupa kata tentang dampak pandemi pada hari-hari besar atau *event* di Indonesia misalkan “dampak pandemi vaksinasi”, “dampak pandemi libur tahun baru / natal”, “dampak pandemi pilkada”, dan “dampak pandemi lebaran”
- e. Pengklasifikasian ini dibagi menjadi 3 kategori yaitu “positif”, “negatif” dan “netral”