

BAB I

PENDAHULUAN

1.1. Latar Belakang

COVID-19 adalah penyakit yang disebabkan oleh turunan *coronavirus* baru, 'CO' diambil dari *corona*, 'VI' virus, dan 'D' *disease* (penyakit). Sebelumnya, penyakit ini disebut '2019 novel *coronavirus*' atau '2019-nCoV.' Virus *COVID-19* adalah virus baru yang terkait dengan keluarga virus yang sama dengan *Severe Acute Respiratory Syndrome (SARS)* dan beberapa jenis virus flu biasa (WHO, 2020). *Coronavirus* 2019 (*Covid-19*) adalah penyakit menular yang disebabkan oleh sindrom pernapasan akut *coronavirus* 2 (*Sars-CoV-2*). Berbagai upaya pencegahan penyebaran virus *Covid-19* pun dilakukan oleh pemerintah di negara-negara di dunia guna memutus rantai penyebaran virus *Covid-19* ini, yang disebut dengan istilah *lockdown* dan *social distancing* (Supriatna, 2020). Angka penyebaran yang begitu besar ini membuat beberapa masyarakat menjadi sedikit parno. Hal ini dikarenakan mereka merasa bingung, cemas, stress, depresi dan frustrasi akan pandemi *COVID-19* yang tak kunjung selesai ini (Kompas, 2020). Mereka merasa khawatir diri mereka sendiri akan tertular virus mematikan tersebut. Tentunya bagi sebagian orang, kondisi tersebut dapat mengganggu kesehatan mental. Sehingga dibutuhkan suatu cara agar mereka tidak merasa khawatir di tengah masa pandemi *COVID-19* ini (Hanafi & Ferdiansyah, 2020). Gesekan opini sosial di masa *Covid-19* terjadi akibat dampak wabah *Covid-19* yang juga melanda Indonesia, menyerpa semua segi hidup bermasyarakat, perilaku hidup sehat, komunikasi sosial, perekonomian keluarga, hingga pandangan politik terpengaruh wabah virus korona baru di kalangan masyarakat (Kompas, 2020). Media sosial merupakan wadah opini masyarakat dalam mengemukakan opini-opininya. Dalam penanggulangan pandemi *Covid-19* ini, pemerintah juga melaksanakan vaksinasi terhadap warga Indonesia namun masih banyak yang belum percaya terhadap vaksinasi ini. Seperti yang dilansir pada laman harian Kompas yaitu Survei Litbang Kompas antara tanggal 27 Desember 2020 hingga 9 Januari 2021, mengungkap tingkat kepercayaan publik atas keamanan dan perlindungan program vaksinasi *corona*. Survei terhadap sebanyak 2.000 responden di 34 provinsi dengan tingkat kepercayaan 95% dan margin *error* 2,83% menyatakan mayoritas responden, yakni

sebanyak 46% percaya dengan vaksin, 31% tidak percaya dan 23% lainnya tidak tahu (Kompas, 2021). Ketidakpercayaan masyarakat terhadap vaksinasi ini juga tertulis pada laman beranda Twitter yang ditulis oleh akun @wingkyp yaitu “silahkan teruskan promo vaksin kalau mau tanggung dosa jariah” (Twitter, 2021).

Twitter merupakan salah satu media sosial yang banyak digunakan oleh masyarakat saat ini. Twitter merupakan salah satu media sosial yang cukup ramai di Indonesia. Pada laporan finansial Twitter kuartal ke-3 tahun 2019, pengguna aktif harian di platform Twitter global dicatat meningkat 17 persen, ke angka 145 juta pengguna (Kompas, 2019). Berbagai macam manfaat dapat diperoleh dari *tweet* dimulai dari *event detection* (deteksi kejadian, salah satunya bencana alam), prediksi pergerakan pasar saham, prediksi pemilu hingga penyebaran penyakit di suatu wilayah. Sebagai contoh, untuk prediksi pergerakan pasar saham, analisa dilakukan dengan cara menganalisa *tweet* yang berisi *mood* positif dan negatif (sentimen analisis) yang berkaitan dengan pasar saham seperti Dow Jones, S&P 500, NASDAQ (Dong et al., 2010). Analisis sentimen merupakan salah satu topik penelitian yang banyak dilakukan.

Data yang dihasilkan dari sosial media khususnya Twitter memiliki potensi dijadikan sebagai sumber data untuk dilakukan pengekstrakan opini atau yang disebut Analisis Sentimen. Analisis Sentimen yang merupakan bagian dari *opinion mining* adalah proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi (Nur, 2011). Analisis sentimen pada dasarnya merupakan upaya mengklasifikasikan sekumpulan opini apakah masuk ke dalam kategori opini positif atau opini negatif. Opini yang telah diklasifikasikan dapat menjadi informasi yang bernilai untuk sebuah perusahaan. Dalam beberapa penelitian analisis sentimen, untuk mendapatkan hasil yang diinginkan dilakukan dengan pendekatan menggunakan metode *data mining*. *Data mining* adalah suatu konsep yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang tersimpan di dalam *database* besar. (Turban et al, 2005).

Beberapa penelitian analisis sentimen dengan menggunakan teknik *data mining* di dalamnya yang menjadi rujukan dalam penelitian ini antara lain, Penelitian lain menyebutkan dalam judul penelitiannya yaitu Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan *Naïve Bayes* dan Pembobotan *Emoji*, di mana dalam penelitian ini menggunakan data Twitter sebagai data penelitiannya dan metode *Naive Bayes* sebagai klasifikasinya (Lestari et al., 2017). Juga pada penelitian dengan judul Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode *Naive Bayes* dan Seleksi Fitur *Query Expansion Ranking*, dimana pada penelitian tersebut ditulis dalam pada kesimpulan yaitu Pada penelitian ini metode yang digunakan adalah *Multinomial Naive Bayes* dengan seleksi fitur *Query Expansion Ranking* untuk mengurangi jumlah fitur pada proses klasifikasi. Proses dari analisis sentimen terdiri dari *preprocessing*, seleksi fitur dengan metode *Query Expansion Ranking*, dan klasifikasi dengan *Naive Bayes*. Pengujian pada penelitian ini adalah uji akurasi dengan menggunakan variasi rasio seleksi fitur, hasilnya seleksi fitur 75% memiliki akurasi terbaik sebesar 86.6%. Metode *Naive Bayes* dan *Multinomial Naive Bayes* merupakan metode serumpun yang memiliki tingkat performa tinggi dalam mengklasifikasi data teks. Seperti hasil penelitian dengan judul “Perbandingan Metode Klasifikasi Berita Hoaks Berbahasa Indonesia Berbasis Pembelajaran Mesin” yang dikutip, dimana dalam penelitian ini dilakukan perbandingan antara metode *Naive Bayes* dan *Multinomial Naive Bayes* dalam mengklasifikasi berita hoax yang bersumber dari portal berita *online*. Hasil dari penelitian ini disebutkan bahwa nilai akurasi yang dicapai metode *Multinomial Naive Bayes* sebesar 92% dan tingkat akurasi yang dicapai metode *Naive Bayes* sebesar 93%. Penelitian ini juga menjadi dasar atau bahan pertimbangan dalam pemilihan metode untuk penelitian yang akan dilakukan oleh penulis. Dari fakta-fakta di atas penulis selaku peneliti akan mengajukan usulan penelitian dengan judul Prediksi Pendapat Masyarakat Terhadap Penggunaan Vaksin Dalam Mengantisipasi *Covid-19* Menggunakan Teknik Sentimen Analisis Berbasis *Naive Bayes*.

1.2. Rumusan Masalah

Berdasar pada latar belakang di atas, maka dapat dirumuskan masalah dalam tesis ini sebagai berikut:

1. Berapakah nilai akurasi, presisi dan *recall* yang diperoleh metode *Naive Bayes* dan *Multinomial Naive Bayes* dalam mengklasifikasi sentimen data *Tweet*?
2. Metode mana yang lebih baik antara *Naive Bayes* dan *Multinomial Naive Bayes* dalam mengklasifikasi sentimen data *Tweet* berdasarkan nilai akurasi yang diperoleh?

1.3. Tujuan

Beranjak dari latar belakang permasalahan, tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui tingkat akurasi, presisi dan *recall* yang diperoleh metode *Naive Bayes* dan *Multinomial Naive Bayes* dalam melakukan klasifikasi sentimen data *Tweet*.
2. Mengetahui metode terbaik antara *Naive Bayes* dan *Multinomial Naive Bayes* dalam melakukan klasifikasi sentimen data *tweet* berdasarkan nilai akurasi yang diperoleh.

1.4. Manfaat

Keluaran dari penelitian ini diharapkan bermanfaat bagi semua kalangan, di antaranya :

1. Penelitian ini secara teoritis diharapkan akan bermanfaat dan menambah khasanah dan variasi penelitian dalam penerapan sentimen analisis dan teknik-teknik *data mining* pada media sosial.
2. Hasil dari penelitian ini diharapkan dapat dijadikan sebagai perbandingan bagi peneliti lain yang tertarik dalam penerapan sentimen analisis dan teknik-teknik *data mining* pada media sosial.

3. Secara praktis hasil penelitian ini juga dapat bermanfaat bagi pihak pemerintahan sebagai landasan atau dasar untuk mengambil keputusan strategis terhadap pandemi *Covid-19* ini.

1.5. Batasan Masalah

Mengingat luasnya ruang lingkup penelitian dalam implementasi teknik - teknik *data mining* pada sentimen analisis, khususnya pada media sosial Twitter, maka penelitian ini dibatasi pada:

1. Data *tweet* yang akan dianalisis berbentuk bahasa Indonesia.
2. Kata kunci yang digunakan adalah “vaksin” dan “*Covid-19*”.
3. Skenario uji menggunakan *K Fold Cross Validation*.
4. *Output* pada klasifikasi adalah perbandingan kecenderungan opini positif dan opini negatif. Kondisi netral tidak akan menjadi pokok bahasan dalam penelitian ini.
5. Pengukuran kinerja klasifikasi adalah akurasi, presisi dan *recall*.
6. Jumlah data yang digunakan sebanyak 1000 *tweet*.
7. Teknik seleksi fitur yang digunakan adalah *chi-square*.
8. *Tools* yang digunakan dalam penelitian ini adalah *Rstudio* untuk *crawling* data *Tweet*, Excel untuk penyimpanan dan *Jupyter Notebook Python* untuk proses analisis.