

BAB I

PENDAHULUAN

1. 1. Latar Belakang

Pandemi Corona Virus Disease 2019 (Covid-19) yang mulai berkembang pada awal tahun 2020 berdampak sangat buruk bagi masyarakat dunia. Wabah ini disebabkan oleh virus yang mulai berkembang dari kota Wuhan, China (Burhanuddin & Abdi, 2020). Perkembangan virus ini cukup cepat dan sangat sulit untuk diantisipasi (Sherman dkk., 2020). Seluruh dunia merasakan dampak yang signifikan termasuk di Indonesia. Indonesia bahkan termasuk negara dengan penularan paling tinggi di Asia setelah India (Worldmeter, 2020). Banyak upaya yang telah dilakukan pemerintah untuk mencegah meluasnya penyebaran virus Covid-19 ini. Upaya tersebut dimulai dengan pemberlakuan protokol kesehatan yang sangat ketat dengan disertai penelitian untuk menemukan vaksin yang dapat melawan virus ini.

Harapan pemerintah tersebut tidak serta merta disambut baik oleh masyarakat Indonesia. Keraguan masyarakat terhadap vaksin ini bisa juga karena masih kurangnya informasi dan sosialisasi yang dilakukan pemerintah. Keraguan tersebutlah yang selanjutnya menimbulkan pro dan kontra di tengah masyarakat (Mufarida, 2020). Opini masyarakat terkait vaksin ini banyak dikemukakan pada layanan media sosial yang digunakan masyarakat, salah satunya pada Twitter (Irawan dkk., 2020).

Twitter adalah salah satu dari 5 media sosial dengan pengguna terbanyak di Indonesia selain *Youtube*, *Whatsapp*, *Instagram* dan *Facebook*(Datareportal, 2021). *Youtube* adalah media sosial dengan konten video(Balakrishnan & Griffiths, 2017), sehingga tidak dapat dijadikan sebagai sumber data dalam penelitian text mining. *Whatsapp* adalah media komunikasi untuk saling bertukar informasi antar user yang tergabung dalam group *Whatsapp*(Nabilla & Kartika, 2020), tidak terbuka untuk umum. *Instagram* adalah media sosial untuk bertukar gambar dan video(Mittal dkk., 2017), sehingga dalam hal ini tidak dapat dijadikan sebagai sumber data dalam penelitian. *Facebook* sejak tahun 2014 telah mengeluarkan kebijakan menghentikan ijin instan akses data pos ke user lain(Dance dkk., 2018), yang artinya kita tidak bisa mengambil data yang ditulis orang lain dalam aplikasi facebook dengan fasilitas API atau pengambilan data secara otomatis dengan programming. *Twitter* merupakan media sosial terbesar dengan akses data secara gratis dan terbuka untuk penelitian(Sharma & Ghose, 2020).

Pada era internet sekarang ini penggunaan media sosial menyebar luas dan sangat pesat di semua lapisan masyarakat(Nakov dkk., 2017). *Twitter* adalah salah satu media sosial yang dipakai generasi sekarang untuk intens berkomunikasi dan berinteraksi pada semua kalangan(Zukhrufillah, 2018). Media twitter merupakan media digital yang penyebarannya tidak dapat dibatasi(Zukhrufillah, 2018).

Vaksin adalah harapan baru bagi penanganan virus Covid-19 ini. Penduduk yang telah mendapatkan vaksin diharapkan memiliki imunitas untuk melawan virus Covid-19 (Dror dkk., 2020). Presiden Republik Indonesia mengeluarkan Perpres

nomor 99 Tahun 2020 yang isinya tentang pedoman pengadaan dan pelaksanaan vaksin Covid-19 (Presiden Republik Indonesia, 2020). Pemerintah melalui Kementerian Kesehatan juga mengeluarkan Keputusan Menteri Kesehatan nomor 9860 Tahun 2020 yang isinya menetapkan lima jenis vaksin Covid-19 yang boleh dipakai di Indonesia (Kementerian Kesehatan Republik Indonesia, 2020). Terbitnya dua peraturan tersebut merupakan langkah serius pemerintah agar vaksinasi pencegahan Covid-19 segera dilaksanakan.

Oleh sebab itu, penting bagi pemerintah untuk mengetahui bagaimana sentimen masyarakat terkait kebijakan vaksin Covid-19 ini. Sentimen masyarakat terhadap kebijakan vaksin berdasarkan opini yang disampaikan dapat diketahui dengan menggunakan metode *text mining*. *Text mining* adalah suatu proses pengambilan informasi dari suatu dokumen. Hasil dari *text mining* akan diolah untuk mengkategorikan setiap kalimat opini tersebut bersifat positif, negatif atau netral (Kaur & Sharma, 2020).

Penelitian sebelumnya tentang analisis sentimen pada twitter tentang opini performa klub sepak bola menggunakan *SVM* dan perbaikan kata menggunakan algoritma *Levenshtein* (Pakpahan dkk., 2019). Penelitian ini pada proses teks preprocessing menggunakan algoritma perbaikan kata *Levenshtein*. Kata-kata yang salah diperbaiki dengan mencari nilai jarak antara kata yang salah dengan kata yang ada pada kamus KBBI. Kata yang dianggap diperbaiki adalah kata yang memiliki nilai jarak *Levenshtein* terdekat atau terkecil, dan kata tersebut dipakai sebagai kata baru untuk perbaikan. Pada perbaikan kata belum membandingkan

dengan kata yang paling sering muncul dalam dokumen sebagai pertimbangan kata yang lebih cocok. Akurasi yang didapat pada penggunaan metode *SVM* dan tanpa menggunakan *Levenshtein Distance* sebesar 82,75% dan dengan penggunaan algoritma *Levenshtein Distance* adalah sebesar 83,25%. Pada penelitian ini penulis menggunakan algoritma *Damerau Levenshtein Distance* dengan harapan dapat menghasilkan peningkatan akurasi yang lebih baik.

Penelitian lainnya tentang analisis sentimen pada perusahaan penerbangan dengan data *Twitter*, dilakukan perbandingan algoritma *SVM* dan *Naïve Bayes* (Rahat dkk., 2019). Penelitian ini mengambil data dari *Twitter* dengan *Twitter API*. *Preprocessing* data dilakukan dengan proses stemming, remote url, remove @tag dan hastag, menghilangkan stop word. Data hasil *preprocessing* diolah dengan algoritma *Naïve Bayes* dan algoritma *Support Vector Machine (SVM)*. Pada perbandingan tersebut didapat bahwa *SVM* menghasilkan akurasi sebesar 82,48 % dan *Naïve Bayes* dengan akurasi 76,56%. Penelitian analisis sentiment ini menyimpulkan metode *SVM* lebih baik dari *Naïve Bayes*. Penelitian lainnya yang membahas analisis sentimen data pada *Twitter*, juga membandingkan metode *Naïve Bayes* dan *Support Vector Machine* (Ogotu dkk., 2019). Pada jumlah data dibawah 1000 data, metode *SVM* lebih unggul dibandingkan *Naïve Bayes*, namun pada jumlah data 1000 data sampai dengan 2000 data metode *Naïve Bayes* menghasilkan akurasi yang lebih baik. Percobaan dengan jumlah data 3000 data kembali metode *SVM* memiliki akurasi yang lebih baik. Hasil tersebut dapat dikatakan metode *SVM* rata-rata lebih unggul, namun perlu penelitian lebih lanjut. Berdasarkan dua

penelitian diatas, penulis mengambil pertimbangan untuk menggunakan metode *Support Vector Machine*, disamping metode ini sudah sangat terkenal dan banyak digunakan dalam penelitian klasifikasi (Hadna dkk., 2016).

Preprocessing data pada penelitian analisis sentimen diperlukan untuk meningkatkan akurasi dari klasifikasi. Pemrosesan data tersebut biasanya terdiri dari menghapus duplikat data, menghapus URL, tag “#” dan tag “@”, menghapus tanda baca, menghapus kata hubung, pemeriksaan ejaan, dan perbaikan kata gaul (Singh & Kumari, 2016). Pada tahap pemeriksaan ejaan, salah satu penelitian terdahulu adalah penelitian tentang pengoreksian ejaan teks berbahasa Indonesia dengan algoritma *Damerau Lanvenshtein Distance* dikombinasikan dengan *Recurrent Neural Network* (Augusfian dkk., 2019). Penelitian ini menyatakan bahwa algoritma *Damerau Lanvenshtein Distance* dan RNN menghasilkan akurasi yang cukup baik dengan rata-rata akurasi 74%.

Penelitian lainnya tentang pemeriksaan ejaan adalah penelitian tentang pencarian dan keroksi kata dengan algoritma *Damerau Lanvenshtein Distance* (Oktaviyani dkk., 2019). Pada penelitian ini dibangun sebuah sistem pencarian artikel dengan *flatfrom* website. Pada saat pencarian, user sering salah dalam mengetikkan kata kunci. Kesalahan pengetikan kata kunci tersebut diperbaiki dengan algoritma *Damerau Lanvenshtein Distance*. Penelitian ini menyimpulkan perbaikan kata menghasilkan akurasi sampai 91,24% dengan presisi 89,58% dan algoritma ini sangat bagus untuk memperbaiki kesalahan kata.

Berdasarkan paparan diatas, penulis mengusulkan penelitian tentang analisis sentimen pada data Twitter berbahasa Indonesia menggunakan metode *Support Vector Machine*. Pada tahap *preprocessing* perbaikan ejaan, penulis mengusulkan algoritma *Damerau Lanvenshtein Distance*. Kombinasi kedua algoritma ini diharapkan dapat memperbaiki akurasi analisis sentimen terkait vaksin Covid-19.

1. 2. Identifikasi Masalah

Identifikasi masalah yang dapat diambil dari penjelasan latar belakang yang dapat dijadikan bahan penelitian antara lain:

1. Opini terkait vaksin Covid-19 berkembang pesat di masyarakat melalui media sosial seiring dengan kebijakan pemerintah untuk pengadaan vaksin Covid-19.
2. Identifikasi sentimen sebuah opini mengenai vaksin Covid-19 sulit dilakukan mengingat jumlah opini yang sangat banyak.

1. 3. Pembatasan Masalah

Penelitian ini dibatasi oleh beberapa hal sebagai pertimbangan ruang lingkup dan untuk memfokuskan hasil penelitian antara lain sebagai berikut.

1. Data yang digunakan dalam penelitian ini adalah data komentar pada twitter dengan bahasa Indonesia.
2. Data yang akan diklasifikasikan adalah komentar yang berkaitan dengan vaksin Covid-19.

1. 4. Rumusan Masalah

Rumusan masalah yang dapat disusun berdasarkan identifikasi masalah dan beberapa batasan masalah antara lain.

1. Bagaimana mengimplementasikan algoritma *Damerau Levenshtein Distance* pada tahapan teks *preprocessing*?
2. Bagaimana perbandingan hasil, penggunaan dan tanpa penggunaan algoritma *Damerau Levenshtein Distance* pada tahap teks *preprocessing* analisis sentimen data *twitter* menggunakan *Support Vector Machine*?

1. 5. Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut.

1. Menerapkan algoritma *Damerau Levenshtein Distance* pada tahapan teks *preprocessing*.
2. Membandingkan hasil klasifikasi *Support Vector Machine* yang didapat, menggunakan algoritma *Damerau Levenshtein Distance* dan tanpa algoritma tersebut pada tahap teks *preprocessing*.

1. 6. Manfaat Penelitian

Penelitian ini diharapkan dapat memiliki manfaat baik secara teoritis di bidang penelitian dan aplikasinya pada masyarakat luas.

1. Manfaat Teoritis.

Secara teoritis hasil penelitian ini diharapkan mampu menambah wawasan peneliti serta memberikan sumbangan pemikiran bagi pengembangan

penelitian analisis sentimen pada media sosial twitter dengan algoritma *Damerau Levenshtein Distance* dan *Support Vector Machine*.

2. Manfaat Praktis

Penelitian ini akan menghasilkan gambaran opini masyarakat terhadap vaksin Covid-19, sehingga dengan hasil penelitian dapat dijadikan pertimbangan tentang perbaikan kebijakan publik terkait vaksin Covid-19 kedepannya.

