

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sejak kasus pertama tanggal 2 Maret 2020, pandemi COVID-19 telah melanda Indonesia menurut Almuttaqi (2020). Setelah lebih dari setahun, muncul varian-varian dari virus COVID-19. Dikutip dari situs Kementerian Kesehatan (promkes.kemkes.go.id), pada bulan Mei 2021, Kementerian Kesehatan memastikan tiga varian dari COVID-19 yaitu varian *alpha*, *beta*, dan *delta* telah masuk ke Indonesia. Pemerintah mulai menerapkan lagi pembatasan kegiatan masyarakat yang disebut dengan PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat) Darurat pada Juli 2021. Hal ini juga bersamaan dengan dimulainya program vaksinasi untuk masyarakat umum.

PPKM Darurat yang pada awalnya diberlakukan hanya selama 2 minggu kembali diperpanjang dan pada Agustus 2021 PPKM darurat diubah menjadi PPKM Level dimana *level* disini disesuaikan dengan tingkat keparahan kasus COVID-19 per kabupaten / kota. Pada 23 Agustus 2021, pemerintah mulai memperbolehkan kegiatan di luar rumah dengan syarat penggunaan aplikasi PeduliLindungi sebagai sarana skrining dan *tracing*. Hasil dari skrining dan *tracing* pedulilindungi akan digunakan sebagai bahan pertimbangan penentuan level PPKM di tiap daerah. Masyarakat yang ingin melakukan kegiatan di luar rumah seperti pergi ke pusat perbelanjaan, makan di restoran, dan bekerja di gedung perkantoran wajib melakukan proses *check-in* menggunakan aplikasi PeduliLindungi dengan cara memindai kode batang di pintu masuk.

Dikutip dari website resmi Pedulilindungi, PeduliLindungi adalah aplikasi yang dikembangkan oleh KOMINFO (Kementerian Komunikasi dan Informatika) yang bekerja sama dengan Kemenkes (Kementerian Kesehatan), KPCPEN (Komite Penanganan Covid-19 dan Pemulihan Ekonomi Nasional), dan Kementerian BUMN dimana aplikasi ini merupakan aplikasi untuk tracing Covid-19 yang diterapkan di Indonesia untuk melakukan pelacakan kontak secara digital. PeduliLindungi sebelumnya bernama TraceTogether. Namun karena dinilai mirip dengan nama aplikasi pelacakan kontak digital dari Singapura, namanya diubah menjadi PeduliLindungi. PeduliLindungi sebagai aplikasi pelacakan kontak digital menggunakan perangkat GPS ponsel dalam melakukan pelacakan keberadaan penggunanya. Perangkat pengguna akan saling bertukar identitas anonim jika berada dalam jangkauan GPS yang berdekatan dan data ini akan disimpan dalam waktu 14 hari. Pelacakan ini berfungsi untuk melacak apabila seseorang memiliki histori berada dekat dengan kontak erat, kasus suspek, dan konfirmasi. PeduliLindungi juga dapat memberi informasi kepada pengguna mengenai status zona dimana pengguna tersebut berada lengkap dengan data kasus Covid-19. Pada PeduliLindungi, masyarakat dapat melakukan pendaftaran program vaksinasi dan setelahnya sertifikat vaksin tersimpan pada data akun PeduliLindungi.

Meskipun aplikasi PeduliLindungi memiliki banyak manfaat baik itu untuk pemerintah maupun masyarakat, banyak kontroversi timbul di masyarakat. Twitter merupakan salah satu platform yang digunakan masyarakat untuk mengutarakan pendapatnya mengenai PeduliLindungi. Twitter adalah aplikasi media sosial dan mikroblog daring dimana penggunanya dapat mengirim dan membaca pesan yang disebut tweet. Tweet adalah pesan berbasis teks pada media sosial Twitter. Pada Maret 2006, Jack Dorsey mendirikan Twitter bersama dengan beberapa media sosial lainnya pada bulan Juli. Semenjak diluncurkan, Twitter masuk dalam jajaran sepuluh

situs yang paling banyak dikunjungi pengguna di Internet, dan mendapat julukan "pesan singkat dari Internet". Di media sosial Twitter, pengguna yang belum terdaftar hanya bisa melihat dan membaca tweet, sedangkan pengguna yang telah terdaftar dapat menulis tweet melalui situs web atau aplikasi dari Twitter. Beragam opini dalam bentuk tweet diutarakan masyarakat, baik itu opini positif dan negatif. Opini tersebut membahas hal-hal terkait Pedulilindungi seperti aplikasi dan regulasinya. Sentimen dalam opini ini dapat bermanfaat untuk pengembangan aplikasi dan regulasi Pedulilindungi oleh pihak terkait sehingga perlu dilakukan analisis sentimen. Dalam penelitian ini akan dilakukan klasifikasi terhadap tweet tersebut ke dalam dua kelas sentimen yaitu positif dan negatif.

Tweet yang akan dianalisis adalah *tweet* pada kurung waktu September 2021 hingga Oktober 2021. Karena pada kurung waktu tersebut topik mengenai pedulilindungi sedang hangat diperbincangkan sehingga sering masuk di kolom *trending topic* di Twitter. Hal ini mungkin terjadi karena pada awal September pemerintah mulai mewajibkan penggunaan pedulilindungi sehingga masyarakat mulai banyak yang menggunakan dan membahas aplikasi tersebut. Melalui *tweet* yang diklasifikasikan ke dalam sentimen negatif, dapat diketahui faktor apa saja yang menyebabkan PeduliLindungi kurang disukai oleh masyarakat yang dapat dilihat melalui visualisasi *wordcloud*.

Untuk melakukan analisis sentimen, digunakan pendekatan *machine learning* menggunakan *supervised learning* dengan metode *Support Vector Machine* yang termasuk dalam kategori *linear classifier*. Pada tahap *preprocess* juga akan menggunakan metode pemilihan fitur dua tahap. Kedua metode ini sudah pernah diterapkan pada penelitian sebelumnya yaitu seperti *Emotion Mining of Indonesia Presidential Political Campaign 2019 using Twitter Data* oleh Kurniawan (2020) dan *Analisa Sentimen Pelanggan WIFI.ID Pada Twitter Dengan Support Vector Machine*

oleh Apriandi (2016) yang menggunakan metode *Support Vector Machine* sebagai algoritma dalam melakukan klasifikasi. Sedangkan pemilihan fitur dua tahap menggunakan *Information Gain* dan *Genetic Algorithm* pernah dilakukan dalam penelitian *A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm* oleh Uguz (2011).

Support Vector Machine (SVM) merupakan metode dalam klasifikasi linier dengan mengidentifikasi hyperplane terbaik yang kemudian dapat digunakan sebagai pemisah antara dua kelas pada *input space*. SVM menerapkan prinsip dasar klasifikasi linear, yang kemudian dikembangkan dengan menambahkan konsep kernel trik pada ruang dengan dimensi tinggi menjadi pengklasifikasi non-linear. Pada umumnya, masalah nyata memiliki bentuk *non-linearly separable*, sehingga *hyperplane* tidak dapat mengklasifikasi kedua class secara sempurna. Oleh sebab itu, diterapkan prinsip kernel pada SVM. Konsep SVM non-linear yaitu dengan menerapkan fungsi $\Phi(x)$ untuk memetakan mengubah data x ke ruang vektor dengan dimensi lebih tinggi (Saifinnuha, 2015).

Untuk meningkatkan efektivitas dan efisiensi model, akan dilakukan pemilihan fitur secara dua tahap dengan *Information Gain* (IG) di tahap pertama dan *Genetic Algorithm* (GA) seperti dibuktikan oleh Uguz (2011). *Information Gain* adalah salah satu metode dalam pemilihan fitur yang menggunakan teknik scoring untuk nominal ataupun pembobotan atribut kontinu yang didiskretkan menggunakan maksimal entropy dimana *entropy* tersebut akan digunakan untuk menentukan nilai *Information Gain*. *Entropy* berguna untuk memberi gambaran tentang banyaknya informasi yang dibutuhkan dalam proses pengkodean suatu kelas. Nilai IG dari suatu istilah diukur dengan menghitung banyaknya bit informasi yang diambil dari prediksi pengkategorian dengan ada atau tidaknya istilah dalam sebuah dokumen. Fitur terpilih

merupakan fitur dengan nilai IG yang tidak sama dengan nol dan atau melebihi suatu nilai threshold tertentu. Ide dalam *Information Gain* untuk memilih fitur adalah menyeleksi fitur dengan informasi yang paling memiliki dampak yang signifikan terhadap kategori. *Genetic Algorithm (GA)* adalah salah satu *Evolutionary Algorithm* dimana algoritma ini mencontoh proses evolusi dimana konsep utama yang diterapkan dalam algoritma ini adalah setiap individu yang lemah akan punah dan setiap individu yang unggul akan bertahan hidup. Keunggulan setiap individu akan diuji menggunakan fungsi yang dikenal dengan *fitness function*. *Fitness* dalam *Genetic Algorithm* merupakan gambaran kelayakan sebuah solusi terhadap sebuah permasalahan. Fungsi *Fitness* akan menghasilkan *output* berupa *fitness value*.

Pemilihan fitur secara dua tahap lebih unggul dibandingkan dengan pemilihan fitur secara tradisional dengan satu tahap. Hal ini dibuktikan oleh Uguz (2011) dimana dengan konfigurasi pemilihan fitur menggunakan IG dan metode KNN menghasilkan *F-measure* 93.74% hingga 96.34%. Sedangkan konfigurasi menggunakan IG dan GA dalam pemilihan fitur serta modeling menggunakan KNN menghasilkan *F-measure* 95.03% hingga 97.84%. Dengan tambahan perbandingan yaitu pemilihan fitur dua tahap menggunakan IG dan PCA menghasilkan *F-measure* 93.86% hingga 97.4%, sedikit lebih rendah dibandingkan IG dan GA.

Dari hasil studi literatur, masing-masing jurnal referensi memiliki metode-metode yang diterapkan untuk tujuan efisiensi dan efektivitas model, serta ada pula metode untuk menambah informasi atau pandangan mengenai data yang digunakan pada penelitian. Namun dari jurnal-jurnal tersebut tidak ada yang membahas atau menggunakan data sentimen mengenai PeduliLindungi dimana topik ini memiliki urgensi untuk diteliti terutama di masa pandemi seperti saat penelitian ini dilakukan dimana PeduliLindungi dapat membantu pemerintah dan masyarakat dalam melakukan tracing atau pendataan mengenai COVID-19. Selain itu metode-metode

tersebut pada penelitian terdahulu belum pernah diterapkan secara bersamaan pada sebuah penelitian, oleh karena itu akan dilakukan penelitian dengan judul “Analisis Sentimen Tweet Mengenai Pedulilindungi Menggunakan Metode Support Vectir Machine”. Metode-metode tersebut sebelumnya pernah digunakan pada beberapa peneltiain terdahulu. Selain pemilihan fitur dua tahap dan metode SVM, akan diterapkan pula metode emotikon handling untuk mentrasnformasi emoji dan emotikon menjadi kata yang dapat berpotensi menjadi fitur yang relevan pada suatu kelas. Diterapkan juga penggunaan botometer dengan asumsi terdapat tweet yang dibuat oleh akun bot, oleh karena itu digunakan botometer untuk mengidentifikasi kemungkinan profil Twitter merupakan bot.

Setelah model selesai dibuat, dilakukan visualiasi hasil klasifikasi dalam bentuk *wordcloud*. Dari *wordcloud* ini akan dilakukan analisis deskripsi data untuk mengetahui faktor apa saja yang mempengaruhi masing-masing kelas sentimen. Hasil analisis deskripsi tersebut dapat dimanfaatkan oleh pihak pengembang aplikasi dan regulasi Pedulilindungi dalam mengembangkan aplikasi dan peraturan terkait Pedulilindungi. Model yang telah dibuat juga akan diimplementasikan pada aplikasi berbasis web. Aplikasi ini akan berfungsi untuk melakukan prediksi kelas sentimen pada sebuah kalimat terutama pada topik Pedulilindungi, serta menampilkan visualisasi hasil klasifikasi.

1.2 Rumusan Masalah

Dari uraian di atas, ditemukan permasalahan yang akan dibahas dalam penelitian ini, yaitu :

- Bagaimana membuat model klasifikasi sentimen terhadap tweet mengenai pedulilindungi?
- Apa saja hal yang membuat pedulilindungi mendapat sentimen negatif dari masyarakat pengguna Twitter?

1.3 Batasan Masalah

Diperlukan batasan masalah agar pembahasan dalam penelitian ini tidak keluar dari topik penelitian. Berikut adalah batasan masalah dalam penelitian ini :

1. Data yang akan dilakukan analisis sentimen merupakan tweet dari Twitter dalam rentang waktu September-Oktober 2021 menggunakan kata kunci “pedulilindungi” dalam bahasa Indonesia
2. Klasifikasi sentimen terbagi atas sentimen positif dan negatif

1.4 Tujuan Penelitian

Tujuan dalam penelitian ini adalah :

- Membangun aplikasi prediksi sentimen dengan menggunakan model yang menggunakan algoritma *Support Vector Machine* dan menerapkan pemilihan fitur dua tahap dengan algoritma *Information Gain* dan *Genetic Algorithm*.
- Mengetahui apa saja hal yang membuat pedulilindungi mendapat sentimen negatif.

1.5 Sistematika Penulisan

Sistematika penulisan skripsi ini akan membantu mengarahkan penyusunan laporan supaya tidak menyimpang dan sebagai acuan dalam mencapai tujuan penulisan laporan skripsi sesuai dengan apa yang diharapkan. Laporan skripsi ini terbagi menjadi 5 bab, yaitu :

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang terkait permasalahan yang diangkat, rumusan masalah, batasan masalah, tujuan dan sistematika penulisan yang digunakan dalam penelitian ini.

BAB II TINJAUAN PUSTAKA

Bab ini berisi tentang dasar teori yang berhubungan dengan masalah yang akan dibahas, penelitian terdahulu, dan tools yang akan digunakan dalam penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab ini berisi tentang metodologi penelitian yang digunakan dalam melakukan penelitian ini seperti pembuatan model, evaluasi model, visualisasi data hasil klasifikasi, analisis deskripsi data, dan pembuatan aplikasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil dan pembahasan yang telah dilakukan selama penelitian diantaranya yaitu hasil pembuatan model, evaluasi model, visualisasi data hasil klasifikasi, analisis deskripsi data, dan pembuatan aplikasi.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan yang dapat diambil dari keseluruhan isi dari laporan skripsi serta saran yang disampaikan penulis untuk penelitian selanjutnya agar didapat hasil yang lebih baik.

DAFTAR PUSTAKA

Berisi tentang literatur yang digunakan sebagai pedoman yang membantu pengerjaan skripsi.

LAMPIRAN

Pada bagian ini berisi tentang data atau pelengkap yang menunjang dalam pembuatan skripsi.