

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sudah satu tahun lebih sejak secara resmi diumumkannya kasus pertama di Indonesia oleh Presiden Joko Widodo pada tanggal 2 Maret 2020 dengan 2 orang terkonfirmasi positif COVID-19 (World Health Organization, 2020). Berbagai macam upaya dan kebijakan telah dilakukan oleh pemerintah untuk menanggulangi pandemi COVID-19 ini, mulai dari pembatasan sosial hingga *lockdown*. Pada 3 September 2020, Presiden Indonesia membentuk Tentang Tim Nasional Percepatan Pengembangan Vaksin *Corona Virus Disease* (COVID-19) sebagai upaya lanjutan dalam menanggulangi pandemi COVID-19 yang tertuang dalam Keputusan Presiden Nomor 18 Tahun 2020 Tentang Tim Nasional Percepatan Pengembangan Vaksin *Corona Virus Disease* (COVID-19) (Presiden Republik Indonesia, 2020). Dalam Keputusan Menteri Kesehatan Nomor HK.01.07/Menkes/9860/2020 tentang Penetapan Jenis Vaksin untuk Pelaksanaan Vaksin *Corona Virus Disease* 2019 (Covid-19), jenis vaksin yang digunakan di Indonesia antara lain yang diproduksi oleh PT Bio Farma (Persero), AstraZeneca, *China National Pharmaceutical Group Corporation* (Sinopharm), Moderna, Pfizer Inc. And BioNTech, dan Sinovac Biotech Ltd. (Kementrian Kesehatan Republik Indonesia, 2020). Program vaksinasi COVID-19 di Indonesia dimulai pada 13 Januari 2021 di Istana Negara dengan Presiden Joko Widodo sebagai orang pertama yang menerima vaksin (CNN Indonesia, 2021). Namun, program vaksinasi tersebut menuai banyak

pro dan kontra dari masyarakat. Kementerian Komunikasi dan Informatika menyebutkan bahwa 30% masyarakat Indonesia meragukan keamanan dan kehalalan vaksin COVID-19 (Media Indonesia, 2021).

Respons dan opini masyarakat dituangkan dalam berbagai media. Media sosial menjadi yang paling populer bagi masyarakat untuk menuangkan pendapatnya mengenai berbagai hal. Hal ini terjadi karena media sosial merupakan suatu hal yang wajib dimiliki oleh seluruh masyarakat (Rachman dan Pramana, 2020). Berdasarkan data Digital 2021: Indonesia oleh Simon Kemp di DATAREPORTAL, pada Januari 2021 pengguna media sosial di Indonesia mencapai 170 juta pengguna. Salah satu media sosial yang paling banyak digunakan di Indonesia adalah Twitter yang mencapai 63,6% dari pengguna internet dengan umur 16 hingga 64 tahun (Kemp, 2020). Hal tersebut membuka peluang sumber data yang sangat besar yang dapat dimanfaatkan untuk menghasilkan suatu pengetahuan. Pemanfaatan data pada media sosial menjadi sebuah terobosan baru dalam melakukan survei. Data dari media sosial dapat memberikan informasi mengenai respons dan opini masyarakat mengenai berbagai hal. Seperti respons dan opini masyarakat terkait dengan program vaksinasi Covid-19 di Indonesia. Hal ini menjadi sangat penting bagi pemerintah untuk menjadi bahan pertimbangan dalam menanggapi sikap publik (Hikmawan, Pardamean dan Khasanah, 2020).

Sentiment Analysis atau yang biasa disebut dengan *opinion mining* merupakan salah satu cabang penelitian dari *text mining* yang bertujuan untuk menentukan persepsi atau subjektivitas publik (khalayak) terhadap suatu topik

pembahasan, kejadian, ataupun permasalahan (Pramana *dkk.*, 2018). *Sentiment Analysis* dapat diterapkan untuk mengetahui opini publik terkait vaksin Covid-19 di Indonesia. Terdapat beberapa penelitian yang menganalisis opini masyarakat terkait Covid-19. Seperti penelitian yang dilakukan oleh (Adrian *dkk.*, 2021) yang melakukan klasifikasi sentimen Pembatasan Sosial Berskala Besar atau PSBB pada media sosial *Twitter* dengan membandingkan algoritma *Support Vector Machine* dengan *Random Forest*. Data yang digunakan pada penelitian tersebut merupakan data *tweet* tentang PSBB yang berjumlah 466 data *tweet*. Hasil penelitian tersebut menunjukkan nilai akurasi 58% untuk algoritma *Random Forest* dan 56% untuk algoritma *Support Vector Machine*. Hasil akurasi kedua algoritma tersebut sangat kecil. Hal ini dikarenakan penggunaan jumlah data yang sedikit, yaitu hanya berjumlah 466 data *tweet*. Sedangkan pada penelitian (Santoso *dkk.*, 2021) melakukan klasifikasi persepsi pengguna media sosial *Twitter* terhadap kasus covid-19 menggunakan algoritma *Logistic Regression* dengan *hyperparameter* L2 dan *None*. Data yang digunakan pada penelitian tersebut diambil dari situs *Kaggle* dengan nama *Coronavirus tweets NLP - Text Classification* dengan jumlah data latih sebanyak 41.157 data *tweet* serta data tes sebanyak 3.798 data *tweet*. Hasil dari penelitian tersebut menunjukkan algoritma *Logistic Regression* dengan parameter L2 menghasilkan nilai akurasi sebesar 77% dan nilai *f1 - score* sebesar 74%, sedangkan parameter *None* menghasilkan nilai akurasi sebesar 74% dengan nilai *f1 - score* sebesar 70%. Penggunaan data yang besar pada penelitian tersebut dapat memengaruhi nilai akurasi pada algoritma *Logistic Regression*. Hal ini dikarenakan beberapa algoritma seperti *Support Vector Machine* (SVM) tidak efisien saat

melatih data dengan kapasitas besar karena kompleksitas waktu meningkat secara eksponensial (Meel *dkk.*, 2020).

Stacking Ensemble merupakan sebuah teknik mengombinasikan beberapa algoritma *machine learning* berdasarkan pembelajarannya dapat meningkatkan akurasi yang dihasilkan. *Stacking* yang dimaksud merupakan prosedur umum dimana pelajar (*learner*) dilatih untuk menggabungkan pelajar individu (*individual learner*) (Zhou, 2012). *Stacking ensemble* mampu mengklasifikasikan sentimen pada tweet Bengali dengan akurasi yang lebih baik dari algoritma *Naive Bayes* (Sarkar, 2020). Pada penelitian (Meel *dkk.*, 2020) juga menggunakan metode *stacking ensemble classifier* dengan mengombinasikan algoritma *Logistic Regression*, *Support Vector Machine*, dan *Random Forest* yang menghasilkan akurasi mencapai 94.32% untuk klasifikasi *binary class*.

Pada penelitian ini, menerapkan *stacking ensemble* untuk klasifikasi sentimen masyarakat terkait dengan vaksin Covid-19 di Indonesia pada media sosial *Twitter* dengan klasifikasi *multi-class*. Hal ini dilakukan karena metode *Random Forest* dan *SVM* pada penelitian (Adrian *dkk.*, 2021) serta algoritma *Logistic Regression* pada penelitian (Santoso *dkk.*, 2021) untuk klasifikasi *multi-class* yang mendapatkan nilai akurasi dan *f1-score* kurang dari 80%. Pada penelitian ini mencoba menerapkan *stacking ensemble classifier* dengan mengombinasikan algoritma *Logistic Regression*, *Random Forest*, dan *Support Vector Machine*, serta *XGBoost* untuk klasifikasi sentimen *multi-class* yang diharapkan dapat menghasilkan performa yang lebih baik.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, pada penelitian ini dapat dirumuskan masalah sebagai berikut:

1. Bagaimana mengimplementasikan metode *stacking ensemble* untuk klasifikasi multi-kelas sentimen masyarakat terhadap vaksin Covid-19 di Indonesia pada media sosial *Twitter* ?
2. Seberapa tinggi nilai akurasi yang didapatkan oleh model *stacking ensemble classifier* untuk klasifikasi multi-kelas sentimen masyarakat terhadap vaksin Covid-19 di Indonesia pada media sosial *Twitter* ?

1.3 Batasan Masalah

Agar penelitian ini dapat dilaksanakan secara spesifik, terdapat beberapa batasan masalah pada penelitian ini. Batasan masalah tersebut adalah sebagai berikut:

1. Data yang digunakan diambil dari *Twitter* selama bulan Mei sampai dengan Juli 2021.
2. Membahas analisis sentimen serta performa metode *stacking ensemble classifier* untuk klasifikasi multi kelas sentimen masyarakat terhadap vaksin Covid-19 di Indonesia pada media sosial *Twitter*.
3. Menggunakan metode *stacking ensemble classifier* untuk klasifikasi multi-kelas sentimen masyarakat terhadap vaksin Covid-19 di Indonesia pada media sosial *Twitter*.
4. Implementasi model *stacking ensemble classifier* menggunakan sintak bahasa pemrograman *python* dengan beberapa *library* pada *python*.

5. Evaluasi model *stacking ensemble classifier* dalam klasifikasi multi-kelas sentimen masyarakat terhadap vaksin Covid-19 di Indonesia pada media sosial *Twitter* menggunakan *Confusion Matrix*.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini adalah:

1. Mengimplementasikan metode *stacking ensemble* untuk klasifikasi sentimen masyarakat terhadap vaksin Covid-19 di Indonesia pada media sosial *Twitter*.
3. Mengidentifikasi nilai akurasi yang didapatkan dari metode *stacking ensemble* untuk klasifikasi sentimen masyarakat terhadap vaksin Covid-19 di Indonesia pada media sosial *Twitter*.

1.5 Manfaat Penelitian

Penelitian ini dilakukan dengan harapan dapat bermanfaat bagi seluruh pihak yang terkait, diantaranya :

1. Secara Aplikatif
 - a. Diharapkan dari hasil analisis klasifikasi sentimen masyarakat terhadap vaksin Covid-19 di Indonesia pada media sosial *Twitter* ini dapat membantu pemerintah dalam menanggapi sikap publik terhadap program vaksinasi di Indonesia.
2. Secara Akademis
 - a. Mengetahui cara implementasi model *machine learning* dengan menggunakan metode *stacking ensemble classifier* untuk klasifikasi multi-kelas sentimen masyarakat terhadap vaksin Covid-19 di Indonesia pada media sosial *Twitter*.

- b. Memberikan wawasan berapa tinggi nilai akurasi model *stacking ensemble classifier* dalam klasifikasi multi-kelas sentimen masyarakat terhadap vaksin Covid-19 di Indonesia pada media sosial Twitter.

1.6 Metodologi Penelitian

Metodologi penelitian merupakan langkah – langkah yang digunakan dalam melakukan sebuah penelitian dan melakukan analisis kritikal dari metode penelitian (Siyoto dan Sodik, 2015). Langkah – langkah penelitian dapat dituangkan ke dalam sebuah peta jalan (*roadmap*) penelitian. Pada *roadmap* penelitian ini terdapat tiga tahap penelitian, yaitu tahap inialisasi dengan waktu jangka pendek, tahap pengembangan dengan waktu jangka menengah, dan tahap lanjut dengan waktu jangka panjang.

Tahap inialisasi merupakan tahap awal dalam penelitian. Tahap ini terdiri dari studi literatur tentang objek penelitian yaitu topik vaksin Covid-19 di Indonesia pada media sosial *twitter* serta metode yang akan digunakan yaitu *stacking ensemble* dan algoritma *machine learning* yang akan digunakan. Pada tahapan ini merangkum *paper* penelitian – penelitian sebelumnya yang berhubungan dengan penelitian ini.

Tahap pengembangan merupakan tahap mengembangkan model yang diusulkan. Pada tahapan ini dilakukan *data crawling* dari *Twitter* menggunakan salah satu *library* dengan topik “vaksin covid-19”. Pada tahapan ini juga dilakukan persiapan pada dataset yang didapatkan dengan memanfaatkan metode *text-preprocessing* yang terdiri dari *lowercasing*, *stemming*, *tokenization*, *stopword removal*, dan *normalization*. Tahap terakhir pada tahap pengembangan yaitu

mengembangkan model, pengujian model, serta evaluasi model. Model yang dikembangkan menggunakan metode *stacking ensemble classifier* dengan mengkombinasikan algoritma *Logistic Regression*, *Support Vector Machine*, *Random Forest*, serta *XGBoost* untuk klasifikasi multi-kelas sentimen masyarakat terhadap vaksin Covid-19 di Indonesia pada media sosial *Twitter*. Pengujian model diperlukan untuk menguji apakah model sudah mencapai akurasi yang diinginkan atau belum. Evaluasi model dilakukan apabila nilai akurasi yang didapatkan tidak sesuai dengan keinginan.

Tahap lanjut merupakan tahapan dengan jangka waktu yang panjang. Pada tahapan ini menganalisis kembali kekurangan dari model yang telah dibuat serta mempelajari metode – metode lain yang dapat digunakan dalam *sentiment analysis* untuk memperbaiki kekurangan pada model yang telah dibuat.

1.7 Sistematika Penulisan

Adapun aturan dan sistematika penulisan yang digunakan dalam penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab pendahuluan ini berisi latar belakang umum yaitu gambaran secara garis besar tentang isi laporan, rumusan masalah, batasan permasalahan pada penelitian, tujuan penelitian, manfaat penelitian yang diperoleh, metode penelitian serta sistematika penulisan laporan tugas akhir.

BAB II LANDASAN TEORI

Bab ini berisi pembahasan teori - teori yang saling berhubungan dengan penelitian seperti konsep serta metode dan algoritma yang terkait dengan penelitian

ini. Pada bab ini juga berisi penjelasan dari penelitian sebelumnya yang relevan dengan penelitian ini.

BAB III METODOLOGI

Dalam bab ini berisi uraian metode yang digunakan dalam melakukan penelitian, mulai dari objek penelitian, matriks penelitian serta tahapan-tahapan yang dilakukan pada penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi pemaparan hasil serta pembahasan terhadap perancangan pada bab sebelumnya, yaitu bagaimana alur sebuah data *tweet* yang diolah dengan memanfaatkan *text-preprocessing*, bagaimana mengombinasikan algoritma *machine learning* yaitu *Logistic Regression*, *Random Forest*, dan *Support Vector Machine* (SVM) pada metode *stacking ensemble classifier* dengan bahasa pemrograman *python*, serta dilakukan evaluasi terhadap model yang telah dibuat.

BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bab terakhir berisi kesimpulan dan saran dari hasil penelitian, serta merupakan garis besar dari metode penelitian yang telah dilakukan. Kesimpulan adalah hasil akhir dari penelitian yang dilakukan, sedangkan Saran berisi tentang rekomendasi sesuai dengan keterbatasan yang ada pada sistem.