

# BAB I PENDAHULUAN

## I.1 Latar Belakang

*Corona Virus Disease 2019* (COVID-19) merupakan wabah dunia yang pertama kali dilaporkan di Wuhan pada Desember 2019. Kasus Covid-19 di Indonesia pertama kali dilaporkan pada 2 Maret 2020. Covid-19 disebabkan oleh infeksi virus SARS-Cov-2 (Sardjoko, 2021). Virus ini dapat menular melalui *droplet* dari seorang yang sedang terinfeksi kepada orang lain. Penularan yang cepat mengharuskan masyarakat untuk menerapkan protokol kesehatan seperti memakai masker, menjaga jarak dan mencuci tangan dengan sabun.

Salah satu tindakan yang dilakukan untuk menghambat penyebaran virus ini adalah dengan pelaksanaan vaksinasi. Dengan adanya vaksinasi, diharapkan imunitas tubuh akan meningkat dan terjadi *Herd Immunity*. Program vaksinasi di Indonesia sendiri baru mulai dilaksanakan pada Januari 2021 untuk Tenaga Kesehatan. Hingga Oktober 2021, terdapat beberapa jenis vaksin yang digunakan di Indonesia. Pada awalnya pelaksanaan vaksinasi banyak menuai pro dan kontra dari kalangan masyarakat mengenai efek samping yang ditimbulkan pasca vaksin. Kurangnya edukasi terhadap masyarakat mengenai vaksin menjadi salah satu penyebabnya. Selain itu, ada beberapa oknum yang menyebarkan *hoax* dan mempengaruhi orang lain untuk tidak mengikuti program vaksinasi dengan berbagai alasan. Berdasarkan data dari Kementerian Kesehatan Republik Indonesia hingga November 2021, total masyarakat yang sudah divaksinasi dosis pertama adalah 53,01% dan dosis 2 sebanyak 31,29 dari total sasaran sebanyak 208.265.720 (*Vaksinasi COVID-19 Nasional*, 2021). Sedangkan menurut penjelasan dari WHO (*World Health Organization*) untuk mencapai *herd immunity* sebanyak 160 juta hingga 187 juta masyarakat Indonesia harus mendapatkan dua dosis vaksinasi atau minimal 70% dari total sasaran. Jumlah tersebut masih di bawah angka 70% sehingga perlu dilakukan evaluasi situasi yang ada pada masyarakat terkait pelaksanaan vaksinasi itu sendiri. Oleh karena itu, analisis sentimen mengenai kebijakan yang dilakukan pemerintah terkait vaksin dirasa perlu untuk mengetahui kecenderungan sentimen terkait vaksin Covid-19 di Indonesia.

Melalui media sosial, orang menunjukkan pandangan, pendapat, dan emosi yang berbeda terhadap berbagai peristiwa yang terjadi akibat pandemi. Dengan penggunaan media *Twitter* pada saat ini, masyarakat dapat mengemukakan pendapatnya secara terbuka terhadap isu hangat yang ada di kalangan masyarakat. Saat ini media sosial banyak dimanfaatkan untuk mengetahui opini masyarakat umum. Umumnya, pada media sosial *Twitter*, isu yang sedang hangat dan banyak dibicarakan masyarakat akan menjadi *Trending Topic*. Isu vaksinasi merupakan salah satu rangkaian topik yang menimbulkan pertanyaan di media sosial, terutama terkait keamanan seluruh prosesnya. Dengan demikian, banyak penelitian telah menganalisis dampak dari berbagai kampanye media sosial pada kecurigaan vaksinasi dan opini publik tentang proses vaksinasi (Yulita et al., 2021). Selain itu, vaksinasi COVID-19 telah menimbulkan pertanyaan baru tentang waktu yang relatif singkat untuk pengembangan vaksin. Seperti kita tahu, proses pengembangan vaksin biasanya memakan waktu 10 tahun (Pedersen et al., 2020). Untuk mengetahui dan menyaring pendapat yang ada di media sosial tersebut apakah termasuk pendapat positif dan negatif dapat menggunakan Analisis Sentimen.

Analisis sentimen sangat membantu dalam proses pengambilan keputusan dan mengetahui opini terkini oleh masyarakat terhadap suatu isu tertentu. Beberapa artikel dan publikasi membuktikan bahwa metode *Support Vector Machine (SVM)* lebih unggul daripada teknik klasifikasi data yang lain (Rani & Bhatt, 2020). Sebagai bahan perbandingan penulis juga menerapkan algoritma *Random Forest* dimana algoritma tersebut dapat digunakan untuk klasifikasi data dalam jumlah yang besar. Selain itu *Random Forest* juga dapat menangani data kategoris lebih baik daripada SVM. Penelitian untuk *text classification* sudah banyak dilakukan sebelumnya, misalnya untuk klasifikasi teks berita BBC *Random Forest*, klasifikasi lagu berdasarkan lirik lagu dengan *Support Vector Machine* (Willy et al., 2021). Penelitian tersebut menghasilkan klasifikasi yang cukup baik dalam melakukan klasifikasi *high dimension* data pada kasus berita maupun lirik lagu. Menurut (Gupte et al., 2014) pada studi *comparative* terkait algoritma yang dapat digunakan untuk analisis sentiment menyatakan bahwa klasifikasi dengan algoritma *Random Forest* jelas memiliki keunggulan dengan akurasi dan kinerja

tinggi, kesederhanaan dalam pemahaman, dan peningkatan hasil selama periode waktu tertentu. Hal ini membuat algoritma tersebut paling cocok untuk situasi seperti analisis sentimen. Berdasarkan studi literatur pada penelitian tersebut, pada penelitian ini penulis melakukan perbandingan algoritma *Support Vector Machine* dan *Random Forest* sehingga dapat diketahui algoritma yang lebih sesuai untuk analisis sentimen pada opini masyarakat terhadap program vaksinasi covid-19 dan menghasilkan nilai akurasi yang lebih baik.

Untuk mengetahui sentimen pada opini masyarakat terhadap vaksinasi covid-19 pada media sosial *twitter*, data yang didapatkan perlu dilakukan pelabelan. Untuk memudahkan pelabelan data dalam jumlah besar, digunakan *tools* pelabelan otomatis. Salah satu *tools* untuk pelabelan otomatis adalah *Textblob*. Dengan menggunakan *Textblob*, analisis sentimen menjadi lebih sederhana. Meskipun penggunaan *Textblob* memudahkan untuk analisis sentimen tetapi *Textblob* hanya menghasilkan nilai *subjectivity* dan *polarity* sehingga pada opini yang bias, *Textblob* mungkin bukan *tools* yang tepat untuk analisis sentimen. Selain itu, penggunaan bahasa yang luas juga memungkinkan penggunaan *Textblob* untuk analisis sentimen menjadi lebih sulit untuk menganalisis emosi (Gujjar J & Kumar H R, 2021). Oleh karena itu pada penelitian ini juga menggunakan *tools* pelabelan lain untuk analisis sentimen, yaitu *Vader*. Penelitian menunjukkan bahwa *Vader* meningkatkan kegunaan lexicon sentimen tradisional seperti LIWC (*Linguistic Inquiry and WordCount*). Tidak seperti LIWC, *VADER* lebih sensitif terhadap ekspresi suasana hati dalam konteks media sosial dan lebih suka menggeneralisasi ke domain lain (Elbagir & Yang, 2019). Dengan menggunakan kedua *tools* pelabelan tersebut, penelitian ini dapat mengetahui *tools* yang lebih sesuai dan menghasilkan akurasi yang lebih baik untuk analisis sentimen pada kasus Vaksinasi Covid-19.

Dengan ini diharapkan dapat memberikan informasi mengenai pandangan ataupun opini masyarakat terhadap pelaksanaan Vaksinasi Covid-19 dengan valid. Berdasarkan penjelasan tersebut, penulis melakukan penelitian ini dengan tujuan untuk mengetahui nilai akurasi dari analisis sentimen masyarakat terhadap program Vaksinasi Covid-19 pada media sosial *Twitter* dengan menggunakan

algoritma *Support Vector Machine* dan *Random Forest* pada dataset yang dilakukan pelabelan dengan menggunakan *Vader* dan *Textblob*.

## **I.2 Rumusan Masalah**

Berdasarkan latar belakang yang ada, maka dibuatlah rumusan masalah sebagai berikut :

1. Bagaimana implementasi algoritma *Support Vector Machine* dan *Random Forest* untuk analisis sentimen masyarakat terhadap Vaksinasi Covid-19 pada media sosial *twitter*?
2. Bagaimana tingkat akurasi algoritma *Support Vector Machine* dan *Random Forest* pada analisis sentimen masyarakat terhadap Vaksinasi Covid-19 pada media sosial *twitter*?
3. Bagaimana perbandingan sentimen positif dan negatif masyarakat terhadap Vaksinasi Covid-19 pada media sosial *twitter* dengan menggunakan *Vader* dan *Textblob*?

## **I.3 Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah dibuat, maka tujuan dari dibuatnya penelitian ini adalah sebagai berikut :

1. Mengetahui implementasi algoritma *Support Vector Machine* dan *Random Forest* untuk analisis sentimen masyarakat terhadap Vaksinasi Covid-19 pada media sosial *twitter*
2. Mengetahui tingkat akurasi algoritma *Support Vector Machine* dan *Random Forest* pada analisis sentimen masyarakat terhadap Vaksinasi Covid-19 pada media sosial *twitter*.
3. Mengetahui perbandingan sentimen positif dan negatif masyarakat terhadap Vaksinasi Covid-19 pada media sosial *twitter*.

## **I.4 Batasan Penelitian**

Agar pembahasan dari penelitian ini tidak semakin meluas, maka dibuatlah Batasan masalah sebagai berikut :

1. Media sosial yang digunakan adalah *Twitter*.

2. *Tweet* yang diambil dan dianalisis hanya yang menggunakan Bahasa Indonesia.
3. *Tweet* yang akan dianalisis adalah *tweet* yang mengandung kata kunci vaksin covid-19.
4. Label yang digunakan untuk analisis sentimen adalah positif (*tweet* yang mendukung program vaksinasi) dan negatif (*tweet* yang kontra dengan program vaksinasi).
5. Kasus yang digunakan adalah Vaksinasi Covid-19 di Indonesia dan beberapa *keyword* yang berkaitan dengan kasus tersebut.

## **I.5 Manfaat Penelitian**

Manfaat dari penelitian ini antara lain adalah sebagai berikut :

1. Bagi Pemerintah, membantu menentukan strategi untuk pelaksanaan vaksinasi di Indonesia sehingga dapat terlaksana dengan lancar dan mengukur tingkat kesadaran masyarakat mengenai pentingnya vaksinasi covid-19 selama masa pandemi.
2. Bagi Masyarakat, memberikan informasi dan referensi yang dapat digunakan untuk mengambil keputusan.
3. Bagi Penulis, Dapat menerapkan ilmu yang didapatkan selama perkuliahan pada studi kasus yang nyata.

## **I.6 Sistematika Penulisan**

Penelitian ini diuraikan dengan sistematika penulisan sebagai berikut

### **Bab I           Pendahuluan**

Pada bab ini menjelaskan tentang uraian permasalahan, latar belakang, perumusan masalah yang diambil dari latar belakang permasalahan, tujuan penelitian, manfaat yang didapatkan dari penelitian, batasan masalah dari ruang lingkup penelitian.

### **Bab II          Tinjauan Pustaka**

Bab ini berisi mengenai literatur yang relevan dengan permasalahan dan penelitian yang dilakukan. Selain itu dibahas pula daftar referensi baik dari buku/penelitian lainnya yang akan digunakan untuk proses analisis serta evaluasi penelitian

yang meliputi konsep analisis sentimen, pengambilan data, algoritma yang digunakan hingga pengukuran dan evaluasi berupa *confusion matrix* dan grafik ROC.

**Bab III Metodologi Penelitian**

Bab ini menjelaskan metode/kerangka kerja yang dipilih untuk penyelesaian masalah mulai dari proses pengumpulan data, tahap implementasi hingga proses evaluasi hasil analisis yang telah dilakukan.

**Bab IV Analisis dan Perancangan**

Bab ini menjelaskan proses perancangan arsitektur *text mining* mulai dari proses pengambilan, pengolahan, dan analisis data.

**Bab V Hasil dan Evaluasi**

Pada bab ini berisi hasil dari implementasi algoritma dan sentimen analisis dengan data berupa teks yang berasal dari *twitter*. Hasil implementasi selanjutnya dievaluasi dengan menggunakan *confusion matrix* dan ROC.

**Bab VI Kesimpulan dan Saran**

Pada bab ini menjelaskan mengenai kesimpulan yang didapatkan dari keseluruhan proses implementasi hingga evaluasi yang menjawab rumusan masalah yang terdapat pada pendahuluan. Saran dari solusi sebagai bahan perbaikan untuk penelitian selanjutnya.