

1. Pendahuluan

COVID-19 adalah penyakit menular melalui transmisi udara (droplet, batuk, bersin) dan permukaan yang terkontaminasi. Berdasarkan Statistik Global Menurut WHO pada bulan Oktober 2020, Indonesia termasuk peringkat pertama tertinggi di Regional Asia Tenggara[1]. Pasien covid 19 dapat mengalami dampak Kesehatan Panjang yang dapat menyebabkan kerusakan paru-paru secara permanen. [2]. Berbagai upaya Kesehatan telah dilakukan untuk mencegah penyebaran COVID-19 salah satunya melalui pemberian Vaksin. Saat ini terdapat lebih dari 100 kandidat vaksin COVID-19 yang sedang dikembangkan, dan beberapa di antaranya masih dalam tahap uji coba pada manusia[3]. Kementerian Kesehatan Republik Indonesia (Kemenkes RI) menuturkan bahwa tidak semua golongan usia akan menerima vaksinasi COVID-19. Ada kelompok usia yang dikecualikan yakni kelompok usia 0-18 tahun, 60 tahun keatas serta orang dengan penyakit penyerta (komorbid) berat [4].

Rencana vaksinasi masal tersebut menuai banyak tanggapan beragam di masyarakat salah satunya di media social twitter. Opini yang beragam tersebut, bisa digunakan oleh pemerintah sebagai bahan masukan dalam menyusun kebijakan mengenai rencana pemberian vaksin masal covid-19. Berdasarkan Hasil riset Wearesocial Hootsuite yang dirilis Januari 2019 pengguna media sosial di Indonesia mencapai 150 juta atau sebesar 56% dari total populasi[5]. Dalam rilis Statista disebutkan pengguna aktif twitter Indonesia per Mei 2016 saja mencapai 24,34 juta [6]. Berdasarkan data statista pada tahun 2019 pengguna twitter banyak didominasi oleh umur 18-34 tahun [7]. Dimana rentang umur tersebut masuk kedalam kelompok usia yang akan menerima vaksinasi.

Untuk mengetahui opini masyarakat di twitter diperlukan sentiment analisis. Sentimen analisis disebut sebagai penambangan opini, bertujuan untuk menentukan sentimen orang-orang tentang suatu topik dengan menganalisis postingan mereka dan berbagai tindakan di media sosial [8]. Sentiment analisis bisa digunakan untuk menambang berbagai data di berbagai bidang ke ilmuwan. Misalnya, dalam bidang politik sentiment analisis pernah digunakan dalam memprediksi kemenangan pemilu seperti pada pemilihan presiden di Indonesia pada tahun 2019 bersumber dari *tweet* dan *tag* terkait[9]. Sentiment analisis dibagi menjadi dua pendekatan yaitu pendekatan berbasis lexicon dan juga pendekatan berbasis machine learning[8]. metode pembelajaran mesin (machine learning) lebih banyak digunakan dan populer karena dianggap lebih efektif dalam menganalisis berbasis teks[10].

Banyak metode klasifikasi yang dapat digunakan untuk melakukan analisis sentiment berbasis teks. Mengenai rencana kebijakan vaksin covid-19 ini belum banyak penelitian sentiment yang membahas hal tersebut terutama di Indonesia saat ini. Seperti pada penelitian Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter menggunakan metode LDA (*Latent Dirichlet Allocation*) menunjukkan penggunaan metode analisis yang lebih tervalidasi diharapkan dapat dilakukan karena penelitian ini terbatas pada penggunaan kamus positif-negatif yang digunakan[15].

Perbandingan mengenai keunggulan metode klasifikasi dapat dilihat Dalam penelitian sentimen analisis pada pemindahan ibukota indonesia menggunakan empat metode klasifikasi yaitu Naïve Bayes classifier(NBC), logistic regression, support vector machine(SVM), dan K-nearest neighbor (KNN) hasilnya menunjukkan bahwa support vector machine (SVM) terbukti lebih unggul dengan nilai *F-measure* sebesar 97.57% [11]. Namun kelas klasifikasi sentimennya hanya menggunakan dua kelas sentimen yaitu positif dan negatif.

Keakuratan SVM juga bisa dilihat Dalam penelitian mengenai prediksi kemenangan Obama dalam pemilu, [12] hasil prediksi SVM berkorelasi positif dengan IEM (Iowa Electronics Markets), keduanya memprediksi Obama memenangkan pemilu dan menyiratkan bahwa Twitter dapat dianggap sebagai sumber yang valid dalam memprediksi hasil. Kemudian, pemilihan fitur dan representasi yang tepat dapat meningkatkan akurasi analisis sentimen yang menggunakan SVM salah satunya penerapan teknik pre-processing dan TF-IDF pada Indonesian-General-Sentiment-Analysis-Dataset[14].

Berdasarkan penelitian analisis dan implemtnasi SVM dengan string kernel dalam klasifikasi berita berbahasa Indonesia didapatkan bahwa kernel RBF mampu menambahkan hasil akurasi menjadi 47,43% meski tidak terlalu signifikan. Hal tersebut dapat disebabkan pada proses preprocessing hanya menerapkan *tokenization* dan *stopword removal* saja. Dan tidak melakukan proses preprocessing pada data uji. Kernel string adalah fungsi kernel seperti rbf, polynomial, linear dan sigmoid yang beroperasi pada string. Merupakan salah satu metode yang dapat mengatasi permasalahan klasifikasi non linear dan merubah data kedalam dimensi tinggi supaya dapat menemukan model klasifikasi yang baik. Terdapat beberapa jenis kernel namun, pada penelitian hanya melakukan pengujian terhadap dua kernel SVM yaitu RBF dan linear[20].

Untuk dapat mengetahui gambaran dari kebijakan rencana vaksinasi masal covid-19 di Indonesia ada beberapa hal yang dilakukan pada penelitian ini. Pertama mencari model klasifikasi menggunakan algoritma SVM dengan penerapan string sehingga dapat memberikan solusi untuk pengklasifikasi sebaran data yang tidak dapat

dipisahkan secara linear[23]. Hal tersebut dapat membarikan peluang menemukan model yang terbaik. Klasifikasi akan dibagi menjadi tiga aakelas yaitu positif,negatif dan netral. Pembobotan TF-IDF dan proses *preprocessing* dilakukan untuk peningkatan model. Kedua prediksi gambaran opini masyarakat dilakukan dengan menggunakan salah satu string kernel yang memiliki hasil uji yang terbaik. Ketiga dilakukan perhitungan data - data penting dan juga topik-topik yang sering di bahas masyarakat saat berbicara mengenai vaksinasi COVID-19.

Kerangka penulisan penelitian ini adalah sebagai berikut: Studi terkait,Sistem penelitian yang dibangun , Evaluasi hasil penelitian dan juga menyimpulkan penulisan penelitian ini secara keseluruhan.