



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Pada akhir bulan Desember 2019 lalu, 41 kasus pneumonia tanpa penyebab yang jelas ditemukan di Wuhan, China [1]. Setelah diteliti oleh Dr. Jianguo Xu beserta tim ilmuwan dari Chinese Academy of Engineering, ditemukan tipe coronavirus baru yaitu SARS-CoV-2 [2]. Penyakit yang disebabkan oleh virus tersebut kemudian diperkenalkan oleh WHO sebagai COVID-19 [3]. COVID-19 kemudian mengguncang dunia, hingga pada akhir Januari 2020, WHO mendeklarasikan situasi darurat terhadap kesehatan dunia [4]. Dilaporkan dari portal berita online Detik, COVID-19 diketahui telah memasuki Indonesia untuk pertama kalinya pada tanggal 2 Maret 2020 [5].

Pemberian vaksin pada masyarakat Indonesia telah diterapkan pemerintah sejak tanggal 13 Januari 2021. Menurut data yang diperoleh dari situs resmi Kementerian Kesehatan Indonesia yang diakses pada 20 Oktober 2021 [6], pemberian vaksin dosis pertama telah mencapai 109 juta penduduk dengan persentase 52,62%. Sedangkan pemberian vaksin dosis kedua baru mencapai 64 juta penduduk. Dengan jumlah demikian, pemberian vaksin kedua di Indonesia baru mencapai persentase 30,23%, sedangkan target pencapaian pemberian vaksin terdapat pada angka 208 juta penduduk.

Dari data tersebut, dapat diketahui bahwa masih terdapat lebih dari 50% masyarakat Indonesia yang masih belum menerima vaksin. Terdapat beragam alasan lambatnya dilaksanakan proses vaksinasi, namun untuk mencari tahu beragam alasan dan pandangan publik tersebut secara manual akan membutuhkan waktu yang lama serta tenaga yang lumayan besar. Untuk mengatasinya, berbagai teknik dan metode penelitian telah dirancang dan dikembangkan.

Sampai saat ini, pandangan dan bahasan oleh publik terhadap suatu pandemik telah menerima banyak perhatian dari para peneliti, seperti penelitian pada pandangan publik terhadap pandemik Ebola [7], penyakit yang disebabkan oleh virus Dengue di Brazil [8], dan pandemik COVID-19 [9][10][11]. Algoritma Latent Dirichlet Allocation (LDA) sendiri telah diimplementasikan untuk mengklasifikasikan pandangan publik pada penelitian oleh Mutanga dkk [12]. Pada penelitian tersebut, LDA digunakan untuk mencari tahu klasifikasi atas pandangan masyarakat

Afrika Selatan terhadap pandemik COVID-19 pada platform Twitter. Hasil klasifikasi oleh penelitian tersebut berhasil memperoleh *coherence score* sebesar 56%, dan pada data yang sama, teknik *sentiment analysis* juga diimplementasikan.

Terkait metode yang digunakan dalam klasifikasi pandangan publik, beberapa penelitian lebih berfokus pada perbandingan performa metode-metode tersebut. Seperti pada penelitian yang dilakukan Hassan dkk [13], dilakukan perbandingan terhadap 2 teknik *topic modelling*, Latent Dirichlet Allocation (LDA) dan Latent Semantic Analysis (LSA), dalam mengklasifikasi corpora berupa kumpulan *e-book*. Pada penelitian tersebut, LDA menunjukkan performa yang lebih baik dengan *coherence score* berupa 59%, dengan selisih 2% di atas LSA. Seperti penelitian sebelumnya, penelitian oleh Alash, H dkk [14] juga membandingkan algoritma LDA dan LSA pada set data berupa kumpulan *tweet*. Namun tidak seperti penelitian terhadap *tweet* lainnya, fitur *hashtag* pada tweets tidak dihilangkan, melainkan dianggap sebagai fitur yang penting dan perlu dipertimbangkan dalam proses klasifikasi. Hasilnya pun menunjukkan bahwa LDA unggul dalam *coherence score* dari LSA. Hal itu disebabkan karena algoritma LDA mencari topik sebanyak-banyaknya, sedangkan algoritma LSA berusaha mengelompokkan dokumen ke dalam jumlah topik yang sesedikit mungkin. Hal tersebut juga merupakan alasan mengapa LDA lebih efektif jika diaplikasikan untuk jumlah pengelompokan yang banyak.

Pendapat dan pandangan publik juga sudah mudah untuk diperoleh di jaman digital ini. Hal tersebut dapat diperoleh melalui pengisian survei, *interview*, maupun melalui media sosial. Media sosial seperti Twitter telah sering digunakan untuk berdiskusi mengenai suatu kejadian yang besar, termasuk pandemi COVID-19. Pada peninjauan 132 publikasi terkait media sosial dan jejaring sosial, diperoleh data bahwa hampir semua publikasi yang mengulas kegiatan berbagi informasi mengenai bencana alam dan *critical events* berfokus pada data Twitter [15].

Penelitian terkait vaksin COVID-19 juga sudah banyak diterapkan. Salah satunya merupakan penelitian yang dilakukan oleh Akarsu dkk [16] berupa survei mengenai pandangan pengguna media sosial terhadap vaksin COVID-19. Dari 759 penjawab, jawaban didominasi oleh rasa takut dan tidak aman terhadap vaksin. Penelitian ini juga menyimpulkan bahwa agar kampanye vaksin COVID-19 dapat berhasil ke depannya, pemerintah perlu mengikuti perkembangan vaksin, transparan akan informasi mengenai vaksin, dan ikut mempertimbangkan keprihatinan publik.

Indonesia dengan jutaan penduduk, di lain sisi, akan menimbulkan kesulitan bagi pemerintah dalam memperhitungkan masing-masing pendapat masyarakat. Karena itu, penelitian ini ditujukan untuk membantu dalam mengetahui pendapat

masyarakat, termasuk apa bagaimana sikap masyarakat terhadap kegiatan vaksinasi. Penelitian akan dilakukan menggunakan data yang diperoleh dari pendapat dan keluhan masyarakat terkait vaksin COVID-19, yang disebarakan dalam bentuk tweets pada media sosial Twitter.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang telah dijabarkan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara mengimplementasikan algoritma Latent Dirichlet Allocation untuk menganalisa aspek apa saja yang menjadi pembahasan masyarakat Indonesia terkait vaksin COVID-19?
2. Bagaimana cara mengimplementasikan teknik Lexicon-based Sentiment Analysis untuk mencari bobot persentase *tweet* yang setuju, netral, dan tidak setuju terhadap masing-masing topik bahasan hasil klasifikasi Latent Dirichlet Allocation terkait vaksin COVID-19?
3. Berapa tingkat akurasi dari teknik analisis sentimen yang digunakan?

1.3 Batasan Permasalahan

Berdasarkan perumusan masalah yang telah dijabarkan, maka yang menjadi batasan masalah adalah sebagai berikut:

1. Set data yang digunakan pada penelitian ini diambil dari tweets berbahasa Indonesia pada media sosial Twitter.
2. Set data berupa tweets yang digunakan berkaitan dengan kata kunci yang sedang tren di Twitter, berupa vaksin, astrazeneca, sinovac, vaksin covid, dan vaksin corona.
3. Pengambilan set data berupa *tweet* dilakukan 3 kali. Pengambilan pertama dilakukan dari rentang waktu antara 12 September hingga 17 September 2021, dan pengambilan kedua dilakukan dari rentang waktu antara 4 Oktober hingga 9 Oktober 2021. Sedangkan untuk pengambilan ketiga dilakukan dari rentang waktu antara 1 November hingga 6 November 2021.

1.4 Tujuan Penelitian

Berdasarkan perumusan masalah yang telah dijabarkan, tujuan dari penelitian ini mencakup sebagai berikut:

1. Mengimplementasikan algoritma Latent Dirichlet Allocation untuk menganalisa berbagai aspek yang menjadi pembahasan masyarakat Indonesia terkait vaksin COVID-19.
2. Mengimplementasikan teknik Lexicon-based Sentiment Analysis untuk mencari bobot persentase *tweet* yang setuju, netral, dan tidak setuju terhadap masing-masing topik bahasan hasil klasifikasi Latent Dirichlet Allocation terkait vaksin COVID-19.
3. Menghitung tingkat akurasi dari teknik analisis sentimen yang digunakan.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Dapat memperoleh gambaran sikap masyarakat terhadap vaksin COVID-19 dan informasi tersebut dapat digunakan oleh pemerintah beserta tenaga kesehatan dalam mengantisipasi cara baru untuk kelancaran vaksinasi COVID-19 pada masyarakat.
2. Dapat memperoleh gambaran cara mengimplementasikan Latent Dirichlet Allocation dan analisis sentimen pada suatu set data.

1.6 Sistematika Penulisan

Sistematika penulisan laporan berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian yang terdiri atas 5 bab, yaitu:

- Bab 1 PENDAHULUAN
Bab pertama berisi penguraian terhadap latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan

- Bab 2 LANDASAN TEORI

Bab kedua berisi kumpulan teori sebagai landasan dari penelitian yang dilakukan, mencakup vaksin COVID-19, media sosial Twitter, *preprocessing*, *topic modelling*, algoritma Latent Dirichlet Allocation, *topic coherence*, analisis sentimen, serta *Lexicon-based Sentiment Analysis*.

- Bab 3 METODOLOGI PENELITIAN

Bab ketiga menguraikan tahap metode penelitian yang digunakan, dengan menggunakan visualisasi *flowchart* dan *mockup* aplikasi.

- Bab 4 HASIL DAN DISKUSI

Bab keempat menjelaskan implementasi kode, pengujian dan evaluasi berupa validasi terhadap sistem yang telah dibangun.

- Bab 5 SIMPULAN DAN SARAN

Bab kelima berisi penjelasan terkait simpulan terhadap penelitian yang telah dilakukan, dan saran untuk pengembangan selanjutnya.

