

PROPOSAL TUGAS AKHIR



UNIVERSITAS TRISAKTI

**Analisis Sentimen Terhadap Vaksin di Twitter Menggunakan Naïve Bayes dan
K-Nearest Neighbors**

Diajukan Sebagai Syarat Dalam Memperoleh Gelar Sarjana Strata Satu (S1)
Program Studi Sistem Informasi

Disusun Oleh:

Nama: Tasya Aulia

NIM: 065001800022

FAKULTAS TEKNOLOGI INDUSTRI

PRODI SISTEM INFORMASI

UNIVERSITAS TRISAKTI

JAKARTA BARAT

HALAMAN PENGESAHAN

Analisis Sentimen Terhadap Vaksin di Twitter Menggunakan Naïve Bayes dan K-Nearest Neighbors

PROPOSAL TUGAS AKHIR

Diajukan Sebagai Syarat Dalam Memperoleh Gelar Sarjana Strata Satu (S1) Program
Studi Sistem Informasi

Universitas Trisakti

Disusun Oleh:

Nama: Tasya Aulia

NIM: 065001800022



UNIVERSITAS TRISAKTI

Jakarta, 28 September 2021

Pembimbing Utama

Dr. Dedy Sugiarto, S.Si, MM

Pembimbing Pendamping

Ir. Teddy Siswanto, M.Si

ABSTRAK

Nama : Tasya Aulia
Program Studi : Sistem Informasi
Judul : Analisis Sentimen Terhadap Vaksin di Twitter Menggunakan Naïve Bayes dan K-Nearest Neighbors

Twitter merupakan salah satu media sosial yang digunakan di Indonesia. Twitter sering digunakan oleh penggunanya untuk menyampaikan pendapat terkait sebuah produk, lembaga maupun peristiwa. Kata Vaksin terutama Vaksin Sinovac menjadi kata kunci yang sedang menjadi topik yang sedang dibicarakan karena keputusan pemerintah terkait wajib vaksin di Indonesia, untuk mengetahui nilai sentimen yang terdapat pada opini masyarakat, analisis sentimen menjadi salah satu metode yang dapat digunakan untuk menilai opini publik tersebut dengan membaginya menjadi beberapa nilai. Nilai-nilai tersebut adalah nilai positif, nilai negatif dan nilai netral. Untuk dapat mengetahui nilai tersebut, dapat digunakan metode *Lexicon-Based* yaitu salah satu metode pelabelan dengan mencocokkan kata-kata yang terdapat dalam dokumen dengan kata-kata yang terdapat dalam kamus. Setelah dilakukan pelabelan, data tersebut diuji dengan metode klasifikasi *Naive Bayes* dan *K-Nearest Neighbors* untuk mengetahui tingkat keakuratannya.

Kata Kunci: Twitter, Analisis Sentimen, Vaksin Sinovac, *Lexicon-Based*, *Naive Bayes*, *K-Nearest Neighbors*

1. Latar Belakang

Akhir-akhir ini, vaksin menjadi perbincangan hangat masyarakat Indonesia di media sosial terutama di Twitter. Sudah beberapa kali kata kunci Vaksin menjadi topik yang sedang tren yang banyak dibicarakan oleh pengguna Twitter. Saat ini, Vaksin kali ini kaitannya sangat erat dengan Virus Covid-19. Hal tersebut diawali oleh dua kasus pasien positif pertama virus Covid-19 di Indonesia di awal bulan Maret 2020 yang membuat kasus virus Covid-19 tersebut terus menanjak di Indonesia. Sampai saat ini, kasus positif virus Covid-19 belum benar-benar hilang. Pemerintah Indonesia menerapkan berbagai macam cara dan perintah guna menekan angka kasus positif yang ada karena Indonesia pernah sampai mengalami kekurangan stok oksigen di setiap rumah sakit di pertengahan tahun 2021. Pada akhirnya pemerintah menerapkan kebijakan vaksinasi guna memperlambat laju virus covid-9 di Indonesia yang tertuang dalam Peraturan Presiden Republik Indonesia Nomor 14 Tahun 2021 tentang Perubahan atas peraturan Presiden Nomor 99 Tahun 2020 tentang Pengadaan Vaksin dan Pelaksanaan Vaksinasi dalam Rangka Penanggulangan Pandemi *Coronavirus Disease* 2019 (COVID 19) Pasal 13 A yang berbunyi: (1) Kementerian Kesehatan melakukan pendataan dan menetapkan sasaran penerima Vaksin COVID-19. (2) Setiap Orang yang telah ditetapkan sebagai sasaran penerima Vaksin COVID-19 berdasarkan pendataan sebagaimana dimaksud pada ayat (1) wajib mengikuti Vaksinasi COVID-19. (3) Dikecualikan dari kewajiban sebagaimana dimaksud pada ayat (2) bagi sasaran penerima Vaksin COVID-19 yang tidak memenuhi kriteria penerima Vaksin COVID-19 sesuai dengan indikasi Vaksin COVID-19 yang tersedia. (4) Setiap orang yang telah ditetapkan sebagai sasaran penerima Vaksin COVID-19 yang tidak mengikuti Vaksinasi COVID-19 sebagaimana dimaksud pada ayat (2) dapat dikenakan sanksi administratif berupa: a. Penundaan atau penghentian pemberian jaminan sosial

atau bantuan sosial; b. Penundaan atau penghentian layanan administrasi pemerintahan; dan/ atau c. denda. (5) Pengenaan sanksi administratif sebagaimana dimaksud pada ayat (4) dilakukan oleh kementerian, lembaga, pemerintah daerah, atau badan sesuai dengan kewenangannya[1].

Kebijakan tersebutlah yang membuat pengguna Twitter terutama masyarakat Indonesia ramai membicarakan tentang vaksin, terkait efek samping yang dapat timbul, kehalalan dari vaksin tersebut, ada juga yang mendukung sikap pemerintah atas perintah wajib vaksin untuk menekan angka kasus COVID-19 sampai ada yang membantu untuk menjelaskan secara detail terkait proses vaksinasi, proses pembentukan antibodi oleh vaksin itu sendiri dan masih banyak pro dan kontra yang tercipta akibat dari vaksin. Selain masyarakat yang pro dan kontra terhadap vaksin, masih banyak juga masyarakat yang mencari tahu informasi lebih detail tentang vaksin di media sosial twitter untuk meyakinkan dirinya apakah mereka mau ikut vaksin atau tidak.

Dengan banyaknya pro dan kontra, vaksinasi pertama di Indonesia akhirnya dilakukan oleh Presiden Indonesia Bapak Joko Widodo pada tanggal 13 Januari 2021 untuk dosis 1 Vaksin Sinovac[2]. Seiring berjalannya waktu dan pemerintah juga semakin memperketat vaksinasi di Indonesia, semakin banyak juga masyarakat yang sudah melakukan vaksinasi. Menurut data dari Kementerian Kesehatan dan *website* resmi pemerintah untuk menanggulangi Covid-19 yaitu Covid19.go.id mencatat sudah 54.441.979 orang yang sudah melakukan vaksinasi dosis ke-2 dan sebanyak 975.390 orang yang mendapatkan vaksinasi dosis ke-3 per tanggal 6 Oktober 2021[3].

2. Rumusan Masalah

Berdasarkan latar belakang di atas, berikut merupakan rumusan masalah pada penelitian skripsi ini:

1. Kata apa saja yang sering dituliskan oleh pengguna Twitter terkait Jenis Vaksin dari tanggal 30 September sampai 6 Oktober 2021.
2. Bagaimana respon sentimen masyarakat terhadap kata Kunci Vaksin Sinovac di Twitter dari tanggal 30 September 2021 sampai 6 Oktober 2021.
3. Bagaimana hasil akurasi dari metode Naive Bayes dan K-Nearest Neighbors terhadap kata kunci Vaksin Sinovac di media sosial Twitter.

3. Batasan Masalah

Untuk pembahasan penelitian skripsi ini, akan dibatasi sebagai berikut:

1. Menggunakan bahasa R
2. Menggunakan data yang berasal dari Twitter dengan kata kunci “Jenis Vaksin“ untuk mengetahui kata apa saja yang sering dibicarakan
3. Menggunakan data yang berasal dari Twitter dengan kata Kunci “Vaksin Sinovac” sebagai data untuk analisis sentimen.
4. Data yang dikumpulkan menggunakan Bahasa Indonesia
5. Data yang digunakan dari tanggal 30 September 2021 sampai 6 Oktober 2021 sebanyak 1624 data untuk kata kunci Jenis Vaksin dan 2427 data untuk kata kunci Vaksin Sinovac dari Twitter

6. Menggunakan metode Naive Bayes dan K-Nearest Neighbors sebagai metode pengklasifikasian terhadap analisis sentimen.

4. Tujuan Penelitian

Untuk mengetahui kata apa saja yang sering muncul terhadap kata kunci Jenis Vaksin, mencari tahu terkait nilai sentimennya dari kata yang sering muncul tersebut dan menguji keakuratannya menggunakan metode Naive Bayes dan K-Nearest Neighbors.

5. Manfaat

Hasil dari penelitian kali ini, diharapkan dapat memberikan manfaat sebagai berikut ini:

1. Memberikan informasi terkait pembahasan yang berhubungan dengan kata kunci tersebut memiliki nilai sentimen positif, negatif atau netral.
2. Untuk pemerintah atau instansi terkait dapat mengetahui respon masyarakat terkait vaksin dan dapat membuat strategi kedepannya untuk vaksinasi di Indonesia
3. Mendapatkan nilai akurasi dari hasil analisis menggunakan metode Naive Bayes dan K-Nearest Neighbors.

6. Kajian Pustaka

6.1. Tinjauan Pustaka

Pada umumnya, sebelum melakukan penelitian membutuhkan banyak referensi hasil penelitian yang telah dilakukan sebelumnya untuk mendukung hasil penelitian yang sedang dikerjakan.

Berdasarkan penelitian sebelumnya, ada beberapa kesulitan yang dialami oleh peneliti dalam melakukan analisis sentimen yaitu pelabelan. Banyak peneliti yang merasa kesulitan saat melakukan pelabelan data secara manual dengan banyaknya data yang tersedia. Pelabelan yang dilakukan oleh peneliti secara manual juga bisa menjadi subjektif[4]. Hal tersebut dikarenakan tidak ada campur tangan dari seorang ahli bahasa yang mempunyai kredibilitas dalam tata bahasa.

Selain pelabelan, melakukan analisis sentimen dalam bahasa Indonesia dari data yang diambil di media sosial memiliki suatu tantangan tersendiri dikarenakan bahasa yang digunakan dalam media sosial biasanya menggunakan bahasa yang tidak baku. Hal ini yang membuat peneliti biasanya melakukan preprocessing data terlebih dahulu dengan mengubah bahasa yang tidak baku menjadi bahasa baku.

Pada penelitian sebelumnya yang dilakukan oleh Nabilah Putri Aprilia, Dian Pratiwi dan Anung Barlianto yaitu Sentiment Visualization of Covid-19 Vaccine Based on Naive Bayes. Penelitian ini menggunakan data sebanyak 8837 data dengan membaginya menjadi 80% sebagai data latih dan 20% data uji yang menghasilkan akurasi sebesar 73,1% dengan menggunakan klasifikasi Naive Bayes. Berdasarkan Confusion Matrix, data aktual negatif sebanyak 120 data yang benar masuk ke kelas negatif, 444 data negatif masuk ke kelas positif dan data aktual Positif sebanyak 31 data positif

masuk ke kelas negatif, 1173 data positif yang masuk ke dalam kelas positif[5].

Penelitian selanjutnya yang dilakukan oleh Nur Hadi Saputra yaitu Analisis Sentimen pada Pemilihan Gubernur Jawa Barat pada tahun 2018 dari data Twitter dengan menggunakan metode Naive Bayes. Pada penelitian ini, akurasi yang dihasilkan oleh metode naive bayes yaitu sebesar 57,14% dengan sensitivitas kelas positif sebesar 85,21%, kelas netral sebesar 48,05% dan 26,31% berada di kelas Negatif. Untuk Spesifitasnya, sebesar 40,86% pada kelas positif, pada kelas netral sebesar 59,81% dan 87.77% berada pada kelas negatif[4].

Pada penelitian lain yang dilakukan oleh Siti Saidah dan Joanna Mayary yaitu dengan Analisis Sentimen Pengguna Twitter Terhadap Dompot Elektronik Dengan Metode Lexicon Based dan K-Nearest Neighbor. dengan menggunakan data OVO sebanyak 357 tweets, data GOPAY sebanyak 337 tweets dan data LinkAja sebanyak 255 tweets. Data tersebut dilakukan pelabelan dengan menggunakan metode Lexicon-Based yang hasil akhirnya diklasifikasikan menggunakan metode K-Nearest Neighbors. Pada penelitian ini menghasilkan nilai cenderung positif dengan akurasi sebesar 86,91% pada data OVO. Sedangkan untuk data GOPAY menghasilkan akurasi sebesar 94,05% untuk nilai positif dan hasil dari data yang terakhir yaitu data tweets LinkAja menghasilkan akurasi sebesar 76,31% untuk nilai positif[6].

Selain itu, ada juga penelitian yang membandingkan metode kombinasi dengan menggunakan metode machine learning dan lexicon based. Penelitian ini dilakukan oleh Alaa Thamer Mahmood, Siti Sakira Kamaruddin, Raed Naser dan Maslinda Mohd Nadzir dengan melakukan sentimen analisis terkait komentar pada postingan Universitas Utara Malaysia (UUM) pada Facebook. Metode yang digunakan pada penelitian ini adalah Naive Bayes, SVM, Lexicon-Based dan kombinasi antara Lexicon-Based

dengan Naive Bayes dan SVM. Pada Metode Machine learning yang digunakan yaitu Naive Bayes dan SVM, metode ini menggunakan cara pelabelan secara manual. Sedangkan untuk Lexicon-Based menggunakan pelabelan secara otomatis berdasarkan kamus yang disediakan. Untuk Metode kombinasi, metode ini menggunakan hasil dari Lexicon-Based yang nantinya diklasifikasikan menggunakan metode Naive Bayes dan SVM. Hasilnya, pada metode Naive Bayes dengan pelabelan manual menghasilkan akurasi sebesar 73%, untuk metode SVM dengan pelabelan manual menghasilkan akurasi sebesar 80% dan untuk Lexicon-Based menghasilkan 85%. Pada metode kombinasi menghasilkan akurasi sebesar 86% untuk metode Naive Bayes dan 90% untuk metode SVM[7].

6.2. Landasan Teori

6.2.1. Twitter

Twitter merupakan salah satu media sosial berbasis *online* yang digunakan pengguna untuk berbagi pembaruan status dengan menulis atau melihat pesan singkat berbasis teks sampai 140 karakter[8].

Selain itu, twitter juga memiliki fitur *trending topics* (topik hangat). Fitur ini disediakan twitter untuk mengetahui topik apa saja yang sedang ramai dibicarakan berdasarkan kata kunci atau tagar yang sering digunakan oleh pengguna di berbagai negara.

6.2.2. Vaksin Covid-19

Covid-19 atau Coronavirus Disease 2019 adalah penyakit dari Coronavirus jenis baru yaitu Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2)[9]. Virus jenis ini merupakan sekumpulan virus dari subfamili Orthocoronavirinae dalam keluarga Coronaviridae dan ordo Nidovirales[10]. Virus ini dapat memiliki

gejala yang berbeda-beda setiap orang, biasanya orang yang terinfeksi Covid-19 akan mengalami gangguan saluran pernapasan. Perkembangan dari virus ini juga sangat tinggi yang membuat orang yang terinfeksi harus segera melakukan isolasi mandiri untuk orang yang memiliki gejala ringan, sedangkan untuk orang yang memiliki gejala berat harus segera dirujuk ke rumah sakit.

Vaksin merupakan antigen (mikroorganisme) yang sudah dilemahkan. Vaksin yang diberikan kepada orang sehat akan menyebabkan pembentukan antibodi yang spesifik dari mikroorganisme yang sudah dilemahkan. Akhirnya, tubuh mengenali mikroorganisme tersebut dan dilatih untuk melawannya ketika virus tersebut masuk ke dalam tubuh. Untuk menemukan sebuah vaksin untuk sebuah virus juga membutuhkan waktu yang sangat lama[9]. Beruntungnya, sekarang Indonesia telah memiliki beberapa vaksin yang dapat digunakan antara lain Astrazeneca, Bio Farma, Johnson & Johnson, Moderna, Pfizer, Sinovac.

6.2.3. Analisis Sentimen

Analisis Sentimen (Opinion Mining) merupakan sebuah bidang untuk menganalisis berbagai aspek terhadap sebuah entitas. Aspek-aspek tersebut dapat berupa sebuah pendapat atau penilaian terhadap sebuah entitas seseorang, sebuah lembaga bahkan untuk sebuah produk dan topik peristiwa yang sedang terjadi. Analisis ini bertujuan untuk mencari nilai emosi yang terkandung dalam aspek-aspek tersebut. Nilai tersebut dapat berupa nilai positif atau negatif[11].

Dalam kehidupan sehari-hari, analisis sentimen ini dapat berupa pencarian testimoni dari pelanggan lain sebelum membeli

sebuah produk atau untuk sebuah perusahaan dapat mencari testimoni dari pelanggan mereka untuk dapat meningkatkan kualitas dan layanan dari produk yang mereka berikan. Hasil dari analisis ini akan dijadikan pengambilan keputusan dan perencanaan masa depan terkait strategi bisnis.

6.2.4. Pengumpulan Data

Pengumpulan data merupakan aktivitas yang dilakukan saat pertama kali melakukan penelitian. Pengumpulan data ini bertujuan untuk mendukung kegiatan penelitian yang sedang dilakukan. Untuk mendapatkan sebuah data yang baik, maka diperlukan metode pengumpulan data yang tepat juga[12].

6.2.5. Preprocessing

Preprocessing merupakan teknik mengolah data dari data yang sebelumnya sudah dikumpulkan. Hal tersebut dilakukan karena data mentah biasanya berupa data yang tidak berstruktur[13]. Maka dari itu, preprocessing berfungsi mengubah data yang tidak terstruktur tersebut menjadi data terstruktur dengan melakukan Case Folding, Cleaning Data, Stopword Removal dan Normalization.

6.2.6. Stemming

Stemming merupakan sebuah proses yang dilakukan untuk memberikan pemetaan pada sebuah kata-kata dalam sebuah dokumen yang akan di ubah menjadi kata dasar. Dalam bahasa Indonesia, untuk membuat kata menjadi kata dasar harus dihilangkan Prefiks, Sufiks dan Konfiks[14]. Prefiks merupakan imbuhan yang berada di awal kata, Sufiks adalah imbuhan yang

berada di akhiran kata, sedangkan konfiks merupakan gabungan antara Prefiks dan Sufiks.

Contoh Prefiks sebagai berikut:

Ber + Main = Bermain

Ber + Lari = Berlari

Ter + Baca = Terbaca

Contoh Sufiks sebagai berikut:

Jalan + kan = Jalankan

Makan + An = Makanan

Cemar + I = Cemari

Contoh dari Konfiks sebagai berikut:

Per + Guna + Kan = Penggunaan

Ber + Datang + An = Berdatangan

Meng + Hias + I = Menghiasi

6.2.7.Lexicon Based

Lexicon based merupakan salah satu teknik pelabelan yang dilakukan dalam analisis sentimen. Teknik ini merupakan teknik untuk memberikan label pada sebuah data text secara otomatis berdasarkan kamus[15]. Kamus ini berupa kumpulan kata kata yang dibagi menjadi

kata positif dan kata negatif. Pelabelan ini dapat memberikan pengetahuan terkait polaritas dari data berdasarkan positif, netral ataupun negatif.

6.2.8.TF-IDF

TF-IDF merupakan sebuah algoritma yang dapat memberikan informasi terkait seberapa sering sebuah kata dalam dokumen akan muncul. Untuk menentukan bobotnya, TF-IDF memiliki beberapa komponen yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF)[16].

$$TF - IDF_{t,d} = TF_{t,d} \times IDF_t$$

Dengan keterangan:

$TF_{t,d}$ = Term Frequency pada kata t di dokumen d

IDF_t = Inverse Document Frequency pada kata t

$$IDF_t = \log \left(\frac{N}{df_t} \right)$$

N = jumlah dokumen

df_t = jumlah dokumen yang mengandung kata t

TF atau Term Frequency adalah bentuk seberapa sering sebuah kata muncul dalam dokumen. IDF atau Inverse Document Frequency merupakan bentuk pengurangan bobot dalam setiap kata yang muncul dalam dokumen yang berfungsi untuk dapat mempermudah pencarian kata yang jarang ditemukan.

6.2.9. Naive Bayes

Naive Bayes merupakan salah satu metode pengklasifikasian statistik yang menggunakan probabilitas sederhana secara independen atau tidak bergantung. Metode ini didasari oleh Bayes Rule yaitu memprediksikan sebuah peluang masa depan dari pembelajaran di masa sebelumnya[17].

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

Dengan keterangan:

x = Data yang class nya belum diketahui

c = Hipotesis data X merupakan sebuah kelas spesifik

$P(C|X)$ = Probabilitas hipotesis C berdasarkan kondisi X
(Posterior Probability)

$P(C)$ Probabilitas hipotesis C (Prior Probability)

6.2.10. K-Nearest Neighbors

KNN atau K-Nearest Neighbors merupakan sebuah teknik klasifikasi dengan cara mengukur tingkat kemiripan pada tetangganya atau mengukur jarak euclidean berdasarkan data latih dengan data ujinya[18].

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Dengan Keterangan:

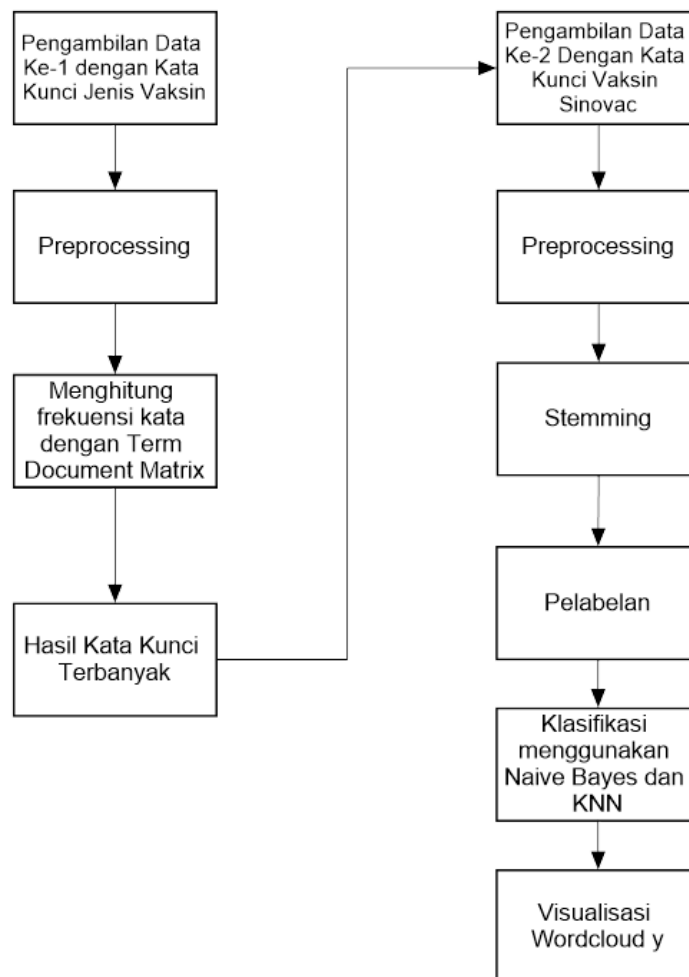
d = Jarak Euclidean

x_i = Atribut dari data latih

y_i = Atribut dari data uji

Metode ini akan membuat sebuah kategori dari data uji berdasarkan kemiripan tetangga terdekatnya dari data latih. Ketika tingkat kemiripannya mendekati atau hampir sama, maka data uji tersebut dapat dimasukkan ke dalam kategori yang sama dengan data latihnya.

7. Metode Penelitian



Gambar 1 *Flowchart* Penelitian

7.1. Pengumpulan data

Pengumpulan data yang dilakukan dalam penelitian ini yaitu dengan menggunakan data dari Twitter dengan kata kunci “Jenis Vaksin” untuk menentukan jenis vaksin apa saja yang banyak dibicarakan oleh pengguna twitter dalam waktu 30 September 2021 sampai 6 Oktober 2021. Setelah menemukan jenis vaksin yang banyak dibicarakan, selanjutnya melakukan pengumpulan data dengan kata kunci “Vaksin Sinovac” dari Twitter dalam waktu 30 September sampai 6 Oktober. Kata kunci “Vaksin Sinovac” inilah yang akan diolah analisis sentimennya. Dari dua kata kunci tersebut, keduanya di ambil di media sosial twitter dengan menggunakan akses API yang diberikan oleh Twitter Developer.

7.2. Preprocessing

Dari data yang sebelumnya yang telah dikumpulkan, selanjutnya melakukan preprocessing data untuk memperbaiki data mentah tersebut sebelum dilakukan pelabelan. Preprocessing yang dilakukan berupa Case Folding, Cleaning, Tokenizing, Stopword Removal dan Normalization.

7.2.1. Case Folding

Case Folding yang merupakan pengubahan teks menjadi lowercase atau huruf kecil.

7.2.2. Cleaning Data

Cleaning yaitu melakukan pembersihan data berupa penghapusan URL, tagar, mention, tanda baca, angka, dan beberapa kata yang tidak digunakan.

7.2.3. Tokenizing

Tokenizing merupakan proses pemecahan sebuah kalimat menjadi sebuah token-token tertentu.

7.2.4. Stopword Removal

Stopword Removal adalah proses pembuangan kata-kata yang tidak diperlukan berdasarkan kamus.

7.2.5. Normalization

Normalization yaitu melakukan normalisasi kata yang tidak baku menjadi kata baku.

7.3. Stemming

Stemming merupakan kegiatan untuk mengubah kata yang memiliki imbuhan di ubah menjadi kata dasar. Pengubahan kata ini dilakukan dengan menghapus imbuhan Prefiks, Sufiks dan Konfiks. Stemming pada penelitian ini menggunakan library khusus yang di *install* dari github.com.

7.4. Pelabelan

Pelabelan merupakan kegiatan memberikan label terhadap teks yang terdapat dalam data yang sudah di preprocessing. Pelabelan ini dilakukan secara otomatis dengan memberikan bobot pada teks. Bobot tersebut diberikan berdasarkan persamaan kata yang terdapat dalam teks dengan kata-kata yang terdapat dalam kamus kata positif dan kamus kata negatif. Kamus kata positif dan kamus kata negatif tersebut bersumber dari github.com. Kata yang sama berdasarkan kamus tersebut mendapatkan nilai 1. Hasilnya, jumlah kata positif - jumlah kata negatif yang menghasilkan bobot pada teks.

Setelah bobot sudah diberikan maka dilanjutkan dengan pelabelan berdasarkan bobot yang ada. Jika bobot < 0 maka hasilnya negatif. Jika

bobot = 0 maka labelnya netral dan jika bobot >0 maka labelnya bernilai positif.

7.5. Pengklasifikasian Data

Metode Naive Bayes merupakan metode pengklasifikasian data yang digunakan pada penelitian setelah data yang akan digunakan telah selesai diberi label. Naive Bayes menggunakan probabilitas sederhana yang memprediksikan sebuah peluang masa depan dari pembelajaran di masa sebelumnya.

Metode selanjutnya yaitu menggunakan metode K-Nearest Neighbors. Dimana untuk dapat mengklasifikasikannya menggunakan jarak euclidean atau jarak tetangga terdekat.

7.6. Visualisasi

Setelah melakukan klasifikasi, selanjutnya yang dilakukan adalah membuat visualisasi dari data yang sudah di beri label. Tujuan dari visualisasi ini yaitu untuk mempermudah membaca makna dari hasil analisis. Visualisasi ini dapat menggunakan Word Cloud yaitu membuat sekumpulan kata yang terdapat dalam data berkumpul membentuk kumpulan kata yang memiliki perbedaan ukuran dan warna berdasarkan jumlah frekuensi kemunculannya. Selain Word Cloud, visualisasi data juga dapat menggunakan histogram untuk mengetahui jumlah frekuensi terhadap klasifikasi.

7.6.1. WordCloud

Sebelum melakukan visualisasi menggunakan word cloud, kata-kata yang terdapat dalam data harus dihitung terlebih dahulu jumlah frekuensi kemunculannya. Selanjutnya dibuat Word Cloud.

Daftar Pustaka

- [1] Peraturan Presiden, “Peraturan Presiden Republik Indonesia Nomor 14 tahun 2021 tentang Perubahan atas Peraturan Presiden Nomor 99 Tahun 2020 tentang Pengadaan Vaksin dan Pelaksanaan Vaksinasi dalam Rangka Penanggulangan Pandemi Coronavirus Disease 2019 (Covid-19),” vol. 2019, no. 039471, p. 13 pages, 2021.
- [2] “Program Vaksinasi COVID-19 Mulai Dilakukan, Presiden Orang Pertama Penerima Suntikan Vaksin COVID-19 – P2P Kemenkes RI.” [Online]. Available:
<http://p2p.kemkes.go.id/program-vaksinasi-covid-19-mulai-dilakukan-presiden-orang-pertama-penerima-suntikan-vaksin-covid-19/>. [Accessed: 6-Oct-2021].
- [3] “Berita Terkini | Satgas Penanganan Covid-19.” [Online]. Available:
<https://covid19.go.id/berita/data-vaksinasi-covid-19-update-6-oktober-2021>.
[Accessed: 7-Oct-2021].
- [4] N. H. Saputra, “Analisis Sentimen Pemilihan Gubernur Jawa Barat 2018 Pada Data Twitter Menggunakan Naïve Bayes” 2020.
- [5] N. P. Aprilia, D. Pratiwi, and A. Barlianto, “View of Sentiment Visualization of Covid-19 Vaccine Based On Naive Bayes Analysis,” *J. Inf. Technol. Comput. Sci.*, vol. 6, no. 2, pp. 195–208, 2021.
- [6] S. S. Salim and J. Mayary, “Analisis Sentimen Pengguna Twitter Terhadap Dompot Elektronik Dengan Metode Lexicon Based Dan K – Nearest

Neighbor,” *J. Ilm. Inform. Komput.*, vol. 25, no. 1, pp. 1–17, 2020, doi: 10.35760/ik.2020.v25i1.2411.

- [7] A. T. Mahmood, S. S. Kamaruddin, R. K. Naser, and M. M. Nadzir, “A combination of lexicon and machine learning approaches for sentiment analysis on facebook,” *J. Syst. Manag. Sci.*, vol. 10, no. 3, pp. 140–150, 2020, doi: 10.33168/JSMS.2020.0310.
- [8] K. Makice, “Twitter API : up and running,” 2009.
- [9] Aditama, Prof. Dr. Tjandra Yoga. Covid-19 Dalam Tulisan Prof. Tjandra. Lembaga Penerbit Badan Penelitian dan Pengembangan Kesehatan (LPB), 2020.
- [10] Yunus, N. R., & Rezki, A. (2020). Kebijakan Pemberlakuan Lockdown Sebagai Antisipasi Penyebaran Coronavirus Covid-19. *SALAM: Jurnal Sosial Dan Budaya Syar-I*, 7(3), 227- 238
- [11] “Sentiment Analysis and Opinion Mining - Bing Liu - Google Buku.” [Online]. Available:
https://books.google.co.id/books?id=AZBfAQAAQBAJ&printsec=frontcover&hl=id&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.
[Accessed: 19-Oct-2021].
- [12] A. Zaki and Dkk., “Metode Pengumpulan dan Teknik Pengumpulan Data,” p. xxii + 362 hlm, 2018.
- [13] H, Aris Tri Jaka, “Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining,” *Inform. UPGRIS*, vol. 1, pp. 1–9, 2015.

- [14] F. Z. Tala, “A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia.”
- [15] S. Kiritchenko, X. Zhu, and S. M. Mohammad, “Sentiment analysis of short informal texts,” *J. Artif. Intell. Res.*, vol. 50, no. October, pp. 723–762, 2014, doi: 10.1613/jair.4272.
- [16] J. Patterson and A. Gibson, *Deep learning : a practitioner’s approach*. O’Reilly Media, 2017.
- [17] C. Sammut and G. Webb, *Encyclopedia of Machine Learning - Google Books*. Springer, 2011.
- [18] E. V Rahcmanani, S. F. Pane, and N. H. Harani, “Algoritma C4.5 dan K-Nearest Neighbors (KNN),” 2019.