

Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier

Winda Yulita^{1)*}, Eko Dwi Nugroho²⁾, Muhammad Habib Algifari³⁾

^{1,2,3}Program Studi Teknik Informatika, Institut Teknologi Sumatera
^{1,2,3}Jl. Terusan Ryacudu, Way Huwi, Jati Agung, Lampung Selatan, Lampung
Email:¹winda.yulita@if.itera.ac.id*

Abstract

The coronavirus outbreak has brought about unprecedented measures, forcing the authorities to make decisions regarding the imposition of lockdowns in some areas hit by the pandemic. Social media has become an important support for people when going through this difficult time. On November 9, 2020, when the first vaccine with an effective rate of over 90% was announced, social media had reacted and people all over the world began to express their feelings regarding vaccination. This study aims to analyze opinions about COVID-19 vaccination in Indonesia. Analysis was carried out on 3780 tweets related to vaccination using the Naïve Bayes Classifier algorithm. Based on the analysis, it can be observed that most tweets have a positive attitude (60.3%), while the number of neutral tweets (34.4%) exceeds the number of opposing tweets (5.4%). The resulting accuracy value is 0.93 (93%).

Keyword: COVID-19, vaccine, Naïve Bayes Classifier

Abstrak

Wabah virus corona telah membawa langkah-langkah yang belum pernah terjadi sebelumnya, yang memaksa pihak berwenang untuk membuat keputusan terkait dengan penerapan *lockdown* di beberapa daerah yang dilanda pandemi. Media sosial telah menjadi pendukung penting bagi orang-orang saat melewati masa sulit ini. Pada tanggal 9 November 2020, ketika vaksin pertama dengan tingkat efektif lebih dari 90% telah diumumkan, media sosial telah bereaksi dan orang-orang di seluruh dunia mulai mengekspresikan perasaan mereka terkait dengan vaksinasi. Penelitian ini bertujuan untuk menganalisis pendapat tentang vaksinasi COVID-19 di Indonesia. Analisis dilakukan terhadap data 3780 tweet yang berkaitan vaksinasi dengan menggunakan algoritma Naïve Bayes Classifier. Berdasarkan analisis, dapat diamati bahwa sebagian besar tweet memiliki sikap positif (60,3 %), sementara jumlah tweet yang netral (34,4 %) melebihi jumlah tweet yang menentang (5,4 %). Nilai akurasi yang dihasilkan sebesar 0,93 (93 %).

Kata Kunci: COVID-19, vaksin, Naïve Bayes Classifier

1. PENDAHULUAN

Wabah virus corona yang disebabkan oleh coronavirus SARS-CoV-2 telah membawa serangkaian perubahan dalam banyak aspek kehidupan ekonomi dan sosial masyarakat. Sejak kemunculannya, pandemi virus corona terus berkembang di berbagai belahan dunia, mencapai 220 negara dan wilayah pada 9 Desember 2020. Pemerintah telah mencoba mengatasi wabah dengan mempertimbangkan serangkaian tindakan, tidak semuanya sesuai dengan opini publik. Selama ini pesatnya pertumbuhan jumlah kasus secara global telah menimbulkan kepanikan, ketakutan, dan kecemasan di kalangan masyarakat [2].

Situasi saat ini yang disebabkan oleh *lockdown* di beberapa bagian dunia dan adanya penerapan *social distancing*, penggunaan media sosial secara global telah meningkat [2]. Hal ini disebabkan keberhasilan menghubungkan orang-orang dari tempat yang berbeda secara geografis dan memungkinkan mereka untuk bertukar ide. Terlebih lagi, banyak orang tampaknya mengandalkan media social untuk memperoleh informasi lebih lanjut. Akibatnya, platform media sosial telah menjadi saluran mediator antara setiap individu dan seluruh dunia, bahkan mejadi aplikasi sosial yang cepat berkembang. [3], [4]. Melalui media sosial ini, orang menunjukkan pandangan, pendapat, dan emosi yang berbeda terhadap berbagai peristiwa yang terjadi akibat pandemi coronavirus [3].

Di antara beberapa platform media social yang terkenal, Twitter telah mendapatkan perhatian khusus karena pengguna dapat dengan mudah menyiarkan informasi tentang pendapat mereka tentang topik tertentu melalui pesan publik, yang disebut tweet [5]. Selain informasi yang ditawarkan secara sukarela oleh pengguna, tweet juga dapat menyimpan informasi yang terkait dengan lokasi pengguna dan memungkinkan berisi *links*, *emoticons*, dan *hashtags* yang dapat membantu pengguna mengekspresikan perasaannya dengan lebih baik dan menjadikannya sumber informasi yang berharga. [5], [6]. Terlebih lagi, Twitter telah digunakan oleh pejabat pemerintah dan tokoh politik untuk menginformasikan kepada masyarakat umum baik mengenai aktivitas mereka atau jika terjadi peristiwa besar [7].

Seiring waktu, informasi yang diambil dari Twitter telah digunakan dalam berbagai penelitian, seperti: menganalisis opini publik terkait krisis pengungsi [8], bencana alam dan gerakan sosial [9], mengevaluasi layanan perusahaan [10] dan reputasi [11], vaksinasi [5], opini film [1] dan lain-lain.

Topik vaksinasi merupakan salah satu tema yang menimbulkan serangkaian pertanyaan di media sosial, sebagian besar terkait dengan keamanan seluruh proses. Akibatnya, serangkaian penelitian telah menganalisis dampak dari kampanye media sosial yang berbeda pada keraguan vaksinasi [12] sentimen masyarakat umum berkaitan dengan proses vaksinasi [13]. Selain itu, vaksinasi COVID-19 menimbulkan pertanyaan baru terkait dengan waktu yang relatif singkat yang dibutuhkan untuk pengembangan vaksin. Seperti diketahui, proses pengembangan vaksin biasanya memakan waktu satu dekade. Pengembangan vaksin tercepat sebelumnya adalah empat tahun dalam kasus vaksin gondok. Bahkan hampir empat puluh tahun setelah penemuan HIV, belum ada vaksin yang efektif yang dikembangkan. Namun, waktu pembentukan vaksin untuk COVID-19 sangat singkat karena keadaan darurat [14]. Pada tanggal 18 Desember 2020, situs web COVID-19 Vaccine Tracker, yang diadakan oleh Milken Institute, menunjukkan 236 vaksin sedang dalam pengembangan, 38 sekarang dalam uji klinis dan 7 telah mencapai keputusan regulasi. Namun demikian, pada 8 Desember 2020 vaksin pertama telah diberikan di Inggris.

Dalam konteks ini, penelitian berikut ini akan menganalisis opini publik terkait proses vaksinasi dalam kasus COVID-19, dengan mempertimbangkan pesan yang diposting di Twitter.

Pengumpulan data pada penelitian ini dilakukan dengan melakukan pengambilan data set dari Kaggle. Data set tersebut merupakan tweet dr masyarakat dan tweet itu akan dikelompokkan dengan label positif, negatif dan netral menggunakan metode naïve bayes classifier. Pada akhirnya, tujuan dari analisis ini adalah untuk menentukan apakah vaksin akan diterima secara luas oleh masyarakat atau tidak.

2. METODE PENELITIAN

Dalam menganalisis sentiment, ada beberapa tahapan yang perlu dilakukan untuk mendapatkan hasil pengujian yang terbaik. Langkah-langkahnya terdiri dari *collection and labeling data*, *preprocessing* dan analisis sentimen.

2.1. Collection and Labeling Data

Pengumpulan data dilakukan dengan *crawling* data tweet menggunakan fasilitas Application Program Interface (API) yang disediakan oleh Twitter. Twitter Search API memiliki kelemahan yaitu hanya dapat mencari tweet yang dipublikasikan dalam tujuh hari terakhir. Dalam penelitian ini mengumpulkan data tweet sebanyak 3780 tweet yang tersimpan di file.csv. Kemudian tweet yang digunakan sebagai *training data* diberi label positif, negatif dan netral. Dalam pelabelan setiap tweet dilakukan secara manual yang kemudian diverifikasi oleh Balai Bahasa, sedangkan *testing data* diproses tanpa label.

2.2. Preprocessing

Setelah memberi label pada data, langkah selanjutnya adalah *preprocessing*. Tahap ini merupakan tahap dimana data disiapkan menjadi data yang siap untuk dianalisis. Data yang diperoleh ditampilkan dalam keadaan apa adanya, sehingga mungkin berupa data yang tidak diinginkan karena merupakan data yang tidak relevan. Data yang tidak relevan tersebut terkadang merupakan data yang sulit ditangani. Oleh karena itu, penghapusan bagian data yang tidak relevan yang sangat disarankan dilakukan dan merupakan tahap yang harus dilewati, sehingga efisiensi sistem dapat terwujud. Berikut struktur yang terdapat pada data tweet yaitu User ID (disimbolkan dengan @), URL, teks, tanggal dan waktu, lokasi, file multimedia (gambar, video dll), emoticon, hashtag (disimbolkan dengan #). Masing-masing bagian struktur tersebut memiliki signifikansi sendiri selama analisis sentimen tetapi ada beberapa data yang tidak memiliki efek signifikan sehingga disarankan untuk menghilangkan data ini saat analisis sentiment.

Nama pengguna dalam tweet selalu diawali dengan simbol @, ini untuk memberitahu siapa yang menulis tweet tersebut. Pada proses analisis sentimen tidak ada pengaruh yang signifikan dari nama pengguna sehingga dengan menerapkan filter, nama pengguna dikeluarkan dari *training data* dan *testing data*. Ada batasan karakter dalam tweet sehingga pengguna menyertakan beberapa tautan URL untuk menjelaskannya dengan lebih baik. URL ini (umumnya dimulai dengan http://) tidak diperlukan selama *training data* dan *testing data* karena tidak mengandung informasi berguna di dalamnya yang dapat digunakan selama analisis sentimen.

Teks adalah isi tweet sebenarnya yang dapat lebih panjang maksimal 140 karakter dan berisi semua yang ingin di-tweet oleh pengguna mana pun. Teks ini terutama digunakan untuk analisis sentimen tetapi harus bebas dari data yang tidak relevan. Tanggal dan waktu diselipkan pada setiap tweet yang memberi tahu tentang kapan tweet tertentu di-tweet. Fitur ini sangat berguna untuk mengetahui seberapa sering pengguna men-tweet dan menemukan kata-kata twitter yang paling iral untuk interval waktu tertentu.

Lokasi juga dapat dilacak di tweet tetapi ini adalah fitur opsional pengguna apakah dia ingin berbagi lokasi atau tidak. Fitur ini membantu untuk menemukan tren di area demografis tertentu. Dalam penelitian ini, fitur lokasi tidak digunakan dan harus dihilangkan. Emoticon adalah simbol yang sangat berguna yang ada dalam teks tweet, mereka menekankan sentimen tweet dan juga membantu menemukan sentimen sebenarnya dari tweet.

Ada beberapa tahapan dalam preprocessing ini, antara lain cleansing, convert negation, convert emoticon, case folding, tokenization, filtering stopword dan stemming dalam bahasa Indonesia.

2.2.1. Cleansing

Cleansing adalah tahapan dimana karakter dan tanda baca yang tidak diperlukan dihilangkan dari teks [28]. Berfungsi untuk mengurangi *noise* pada *dataset*. Contoh karakter yang dihilangkan seperti URL, tag (#), tanda baca seperti titik (.), Koma (,) dan tanda baca lainnya. Contoh kalimat yang melalui proses *cleansing* yaitu “Vaksin menimbulkan pro dan kontra di masyarakat <http://dlvr.it/Qb83n8> pic.twitter.com/8MucIMhUMO. Keluaran atau *output*-nya adalah “Vaksin menimbulkan pro dan kontra di masyarakat”.

2.2.2. Convert Negation

Dalam bahasa Indonesia terdapat kata “tidak” yang disebut dengan kata negasi yang merupakan kata yang dapat membalikkan arti kata yang sebenarnya. Berikut ini adalah contoh kalimat *convert negation*, input “Saya tidak mau mengulur2 waktu nih”, output “Saya tidak_mau mengulur2 waktu nih”.

2.2.3. Convert Emoticon

Emoticon adalah ekspresi wajah yang diwakili oleh kombinasi huruf, tanda baca, dan angka. Pengguna biasanya menggunakan *emoticon* untuk mengekspresikan *mood* yang mereka rasakan. *Convert emoticon* adalah salah satu cara untuk mengekspresikan ekspresi perasaan secara tekstual. Ini adalah contoh kalimat yang melalui proses *convert emoticon* dengan *input* “wah, senang. sudah banyak yg makin peduli :)”, maka *output*-nya ”wah, senang. sudah banyak yg peduli senang”. Berikut adalah salah satu contoh *convert emoticon* kata:

Tabel 1. Contoh *convert emoticon*

Sebelum	Sesudah
:(:'(:[;(:/): x(--" :# :-@ :c :f ;(:v :x :s)" : * _ *	Sedih
:) :] (^_^) ^^v <3 ^^ ^_^ 0:) <:} :* (^.^) =) :3	senang

2.2.4. Case Folding

Dalam penulisan sebuah tweet pasti ada perbedaan bentuk huruf. Tahap ini merupakan proses penyeragaman huruf baik huruf kecil maupun huruf besar. Ini adalah contoh kalimat *case folding*, *input* “Pemerintah mewajibkan yang melakukan bepergian harus memiliki sertifikat vaksin COVID”, dan *output* “pemerintah mewajibkan yang melakukan bepergian harus memiliki sertifikat vaksin covid”.

2.2.5. Tokenization

Suatu proses yang dilakukan untuk memotong atau memecah kalimat menjadi bagian-bagian atau kata-kata. Hasil kata dari proses ini disebut token. Dalam beberapa kasus, proses *tokenization* juga dilakukan dengan menghilangkan tanda baca yang tidak diperlukan. Ada beberapa model *tokenization* yang dapat digunakan yaitu unigram, bigram, trigram, dan ngram. Ini adalah contoh kalimat *tokenization*, input “pemerintah mewajibkan yang melakukan bepergian harus memiliki sertifikat vaksin covid”. *Output*-nya adalah “pemerintah”, “mewajibkan”, “yang”, “melakukan”, “bepergian”, “harus”, “memiliki”, “sertifikat”, “vaksin”, “covid”.

2.2.6. Filtering

Filtering adalah tahapan menghilangkan kata-kata yang muncul dalam jumlah yang banyak tetapi dianggap tidak memiliki arti (*stopwords*). Pada dasarnya, daftar *stopwords* adalah kumpulan kata yang banyak digunakan dalam berbagai bahasa. Alasan menghapus *stopword* di banyak program aplikasi yang berhubungan dengan *text mining* adalah penggunaannya yang terlalu umum, sehingga pengguna bisa fokus pada kata lain yang jauh lebih penting. Ini adalah contoh kalimat *stopwords*, input “pemerintah mewajibkan yang melakukan bepergian harus memiliki sertifikat vaksin covid”, output “pemerintah mewajibkan melakukan bepergian memiliki sertifikat vaksin covid”. Berikut adalah beberapa contoh kata dari *stopwords*:

Tabel 2. Contoh *stopword*

ada	di	kalau	pada	yaitu
aku	dia	kami	saja	bila
bapak	ini	lalu	tentu	hari
berbagai	itu	lewat	untuk	masa
cara	jadi	meski	yang	tapi
cuma	juga	oleh	wah	hal

2.2.7. Stemming

Stemming adalah tahapan mengubah kata menjadi kata dasar menurut kaidah bahasa Indonesia yang benar. Berikut adalah contoh *stemming* kalimat, *input* “pemerintah mewajibkan bepergian memiliki sertifikat vaksin covid”, dan *output* “pemerintah wajib pergi milik sertifikat vaksin covid”. Berikut ini adalah beberapa contoh kata yang melakukan proses *stemming*.

Tabel 3. Contoh kata yang melakukan proses *stemming*

Sebelum	Setelah
akhiri	akhir
mewajibkan	wajib
diberikan	beri
secukupnya	cukup
dipergunakan	guna
adanya	ada

2.2.8. Weighting Word

Weighting word atau pembobotan kata merupakan mekanisme untuk memberikan skor pada frekuensi kemunculan suatu kata dalam dokumen teks. Salah satu metode yang populer untuk pembobotan kata adalah TF-IDF (*Term Frequency-Inverse Document Frequency*). *Term Frequency-Inverse Document Frequency* merupakan metode pembobotan yang menggabungkan dua konsep yaitu *Term Frequency* dan *Document Frequency*. *Term Frequency* adalah konsep pembobotan dengan mencari seberapa sering (frekuensi) kemunculan suatu *term* dalam satu dokumen. Karena setiap dokumen memiliki panjang yang berbeda-beda, maka bisa saja sebuah kata muncul lebih banyak pada dokumen yang panjang dibandingkan dengan dokumen yang pendek. Dengan demikian, *term frequency* sering dibagi dengan panjang dokumen (jumlah kata dalam dokumen).

Sedangkan *Document Frequency* adalah jumlah dokumen dimana suatu *term* muncul. Semakin kecil frekuensi kemunculannya, semakin kecil nilai bobotnya. Saat menghitung *term frequency*, semua kata di dalamnya dianggap penting. Namun, ada kata-kata yang sebenarnya kurang penting dan tidak perlu diperhitungkan seperti “di-”, “ke-”, “dan”, dll. Oleh karena itu, kata-kata yang kurang penting ini perlu dikurangi bobotnya dan menambah bobotnya pada kata-kata penting lainnya. Ini adalah ide dasar mengapa *stopword* diperlukan. Oleh karena itu diperlukan perhitungan TFIDF, agar skor dapat diperoleh dengan menggunakan persamaan (1):

$$tf - idf_{t,d} = tf_{t,d} * idf_t \quad (1)$$

2.3. Classification of Sentiment analysis Using Naive Bayes Classifier

Pengklasifikasi Naive Bayes adalah pengklasifikasi paling sederhana dan paling umum digunakan. Model klasifikasi Naive Bayes menghitung probabilitas posterior suatu kelas berdasarkan distribusi kata dalam dokumen. Hal itu bergantung pada representasi dokumen yang sangat sederhana sebagai *Bag of words*. Model ini bekerja dengan mengekstraksi fitur *bag of words* yang mengabaikan posisi kata dalam dokumen. Ini menggunakan Teorema Bayes untuk memprediksi probabilitas bahwa set fitur yang diberikan milik label tertentu. Untuk analisis sentimen twitter, bigram dari data twitter digunakan sebagai fitur di Naive Bayes. Proses ini

mengklasifikasikan tweet menjadi label positif, negatif dan netral [15].

Naïve Bayes merupakan machine learning yang menggunakan perhitungan probabilitas yang menggunakan konsep pendekatan Bayes. Penggunaan teorema Bayes pada algoritma Naïve Bayes adalah dengan menggabungkan *prior probability* dan *conditional probability* dalam suatu rumus yang dapat digunakan untuk menghitung probabilitas dari setiap kemungkinan klasifikasi.

$$P(H|X) = \frac{P(H)P(X|H)}{P(X)} \quad (2)$$

3. HASIL DAN PEMBAHASAN

Setelah proses klasifikasi analisis sentimen selesai dilakukan, diperlukan satu langkah lagi untuk menentukan kualitas proses yang telah dilakukan yaitu mengevaluasi hasil. Pada tahap ini kinerja dari perhitungan yang telah dilakukan akan diuji dengan parameter akurasi. Akurasi (A) adalah jumlah dokumen yang diklasifikasikan dengan benar, baik *True Positive* maupun *True Negative*. Menghitung nilai akurasi dapat menggunakan persamaan (3):

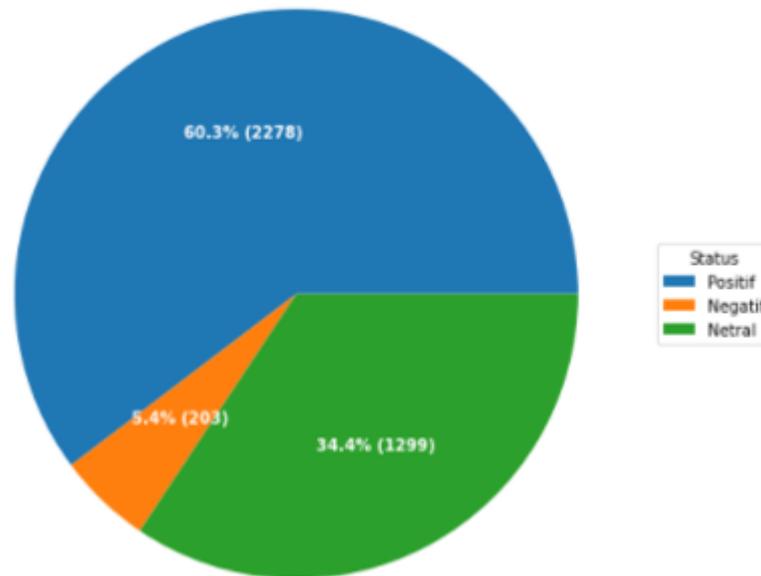
$$A = \frac{(TP + TN)}{(TP + FP + TN + FN)} \times 100\% \quad (3)$$

Variabel seperti TP, TN, FP, dan FN berasal dari *confusion matrix*. TN adalah singkatan dari True Negative, data negatif tergolong negatif. TP adalah singkatan dari True Positive, data positif tergolong positif. FN adalah singkatan dari False Negative, data positif tergolong negatif. FP adalah singkatan dari False Positive, data negatif tergolong positif. Tabel 4 merupakan penjelasan dari *confusion matrix*.

Tabel 4. *Confusion matrix*

	Prediction Yes	Prediction No
True Yes	TP	FN
True No	FP	TN

Setelah melakukan proses pelatihan dan pengujian, didapatkan hasil sebagai berikut:



Gambar 2. Hasil Penelitian

Hasil yang didapatkan dari penelitian ini adalah masyarakat Indonesia rata - rata memberikan respon yang positif terkait kebijakan vaksinasi Covid-19 dengan persentase data sebesar 60.3% dan jumlah data sebanyak 2278 data. Respon negatif terhadap kebijakan lebih kecil dibandingkan respon netralnya, sehingga membuktikan hanya sedikit orang yang menentang kebijakana vaksinasi ini. Hal ini ditunjukkan dengan nilai respon negatif adalah 5.4% (203 data), dan respon netral adalah 34.4% (1299 data). Nilai akurasi yang dihasilkan sebesar 0,93 (93 %).

4. KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini adalah pendapat masyarakat Indonesia khususnya pengguna media sosial Twitter, rata - rata memberikan respon positif terkait adanya kebijakan vaksinasi Covid - 19 di Indonesia dibuktikan dengan persentase respon positif sebanyak 60.3% (2278 respon). Penggunaan algoritma Naive Bayes Classifier untuk melakukan analisis sentimen ini sudah sangat baik ditunjukkan dengan hasil akurasi yang tinggi sebesar 93%.

REFERENCES

- [1] S. Styawati and K. Mustofa, "A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 3, p. 219, 2019, doi: 10.22146/ijccs.41302.
- [2]. K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag and A. E. Hassanien, "Sentiment analysis of COVID-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media", *Appl. Soft Comput.*, vol. 97, Dec. 2020.
- [3]. A. H. Alamoodi, B. B. Zaidan, A. A. Zaidan, O. S. Albahri, K. I. Mohammed, R. Q. Malik, et al., "Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review", *Expert Syst. Appl.*, vol. 167, Apr. 2021.



- [4]. G. Appel, L. Grewal, R. Hadi and A. T. Stephen, "The future of social media in marketing", *J. Acad. Marketing Sci.*, vol. 48, no. 1, pp. 79-95, Jan. 2020.
- [5]. E. D'Andrea, P. Ducange, A. Bechini, A. Renda and F. Marcelloni, "Monitoring the public opinion about the vaccination topic from tweets analysis", *Expert Syst. Appl.*, vol. 116, pp. 209-226, Feb. 2019.
- [6]. A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods", *ACM Comput. Surv.*, vol. 49, no. 2, Nov. 2016.
- [7]. J. Golbeck, J. M. Grimes and A. Rogers, "Twitter use by the U.S. Congress", *J. Amer. Soc. Inf. Sci. Technol.*, vol. 61, no. 8, pp. 1612-1621, May 2010.
- [8]. N. Öztürk and S. Ayvaz, "Sentiment analysis on Twitter: A text mining approach to the syrian refugee crisis", *Telematics Informat.*, vol. 35, no. 1, pp. 136-147, Apr. 2018.
- [9]. G. A. Ruz, P. A. Henríquez and A. Mascareño, "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers", *Future Gener. Comput. Syst.*, vol. 106, pp. 92-104, May 2020.
- [10]. P. Tiwari, P. Yadav, S. Kumar, B. K. Mishra, G. N. Nguyen, S. P. Gochhayat, et al., "Sentiment analysis for airlines services based on Twitter dataset" in *Social Network Analytics*, New York, NY, USA:Academic, pp. 149-162, 2019.
- [11]. N. A. Vidya, M. I. Fanany and I. Budi, "Twitter sentiment to analyze net brand reputation of mobile phone providers", *Procedia Comput. Sci.*, vol. 72, pp. 519-526, Jan. 2015.
- [12]. E. A. Pedersen, L. H. Loft, S. U. Jacobsen, B. Søbørg and J. Bigaard, "Strategic health communication on social media: Insights from a Danish social media campaign to address HPV vaccination hesitancy", *Vaccine*, vol. 38, no. 31, pp. 4909-4915, Jun. 2020.
- [13]. S. Martin, E. Kilich, S. Dada, P. E. Kummervold, C. Denny, P. Paterson, et al., "Vaccines for pregnant women...?! Absurd'—Mapping maternal vaccination discourse and stance on social media over six months", *Vaccine*, vol. 38, no. 42, pp. 6627-6637, Sep. 2020.
- [14]. T. T. Le, Z. Andreadakis, A. Kumar, R. G. Román, S. Tollefsen, M. Saville, et al., "The COVID-19 vaccine development landscape", *Nature Rev. Drug Discovery*, vol. 19, no. 5, pp. 305-306, Apr. 2020.
- [15]. R. Jose and V. S. Chooralil, "Prediction of election result by enhanced sentiment analysis on twitter data using classifier ensemble Approach," *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, 2016, pp. 64-67, doi: 10.1109/SAPIENCE.2016.7684133.