Building of Informatics, Technology and Science (BITS)
Volume 4, No 1, Juni 2022 Page: 106−109
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i1.1581

Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM

Rima Tamara Aldisa[1], Pandu Maulana[2]

Fakultas Teknologi Komunikasi dan Informatika, Informatika, Universitas Nasional, Jakarta, Indonesia
Fakultas Ilmu Komputer, Magister Teknologi Informasi, Universitas Indonesia, Depok, Indonesia
Email: [1]rimatamaraa@gmail.com, [2]pandu.maulana44@gmail.com
Email Penulis Korespondensi: rimatamaraa@gmail.com
Submitted: **11/05/2022**; Accepted: **20/05/2022**; Published**: 30/06/2022**

Abstrak−Dampak dari adanya pandemi COVID-19 sangatlah luas, terutama di Indonesia. Pada awal tahun 2022, Indonesia memasuki tahap awal pemulihan kondisi yang disebabkan oleh pandemi. Pemerintah melakukan opsi bagi masyarakat untuk melakukan vaksinasi dosis ketiga (booster). Namun, muncul sejumlah pro dan kontra di tengah masyarakat terkait vaksin booster tersebut. Penelitian ini bertujuan untuk melakukan analisis sentimen terkait opini masyarakat terhadap vaksinasi booster COVID-19 di Indonesia menggunakan model Naive Bayes. Sumber data yang digunakan berasal dari Twitter. Alur kerja penelitian ini meliputi crawling data, pemberian label, preprocessing, pembagian dataset, dan uji coba model serta perbandingannya dengan model lain, yakni Decision Tree dan SVM. Hasil pada penelitian ini menunjukkan skor AUC terbesar jatuh kepada model SVM (75.40%), namun untuk presisi yang lebih akurat jatuh kepada model Naive Bayes (83.81%). Selain itu, terdapat confusion matrix yang menunjukkan bahwa uji coba model Naive Bayes yang dilakukan berjalan dengan baik.

Kata Kunci: Pandemi; Booster; Analisis Sentimen; Naive Bayes; Twitter

Abstract−The impact of the COVID-19 pandemic is very broad, especially in Indonesia. In early 2022, Indonesia entered the early stages of recovering conditions caused by the pandemic. The government has an option for the community to carry out a third dose of vaccination (booster). However, there are a number of pros and cons in the community regarding the booster vaccine. This study aims to conduct sentiment analysis related to public opinion on the COVID-19 booster vaccination in Indonesia with Naive Bayes model. The data source used comes from Twitter. The workflow of this research includes data crawling, labeling, preprocessing, dataset sharing, and model testing and comparison with other models, namely Decision Tree and SVM. The results of this study indicate that the largest AUC score falls to the SVM model (75.40%), but for more accurate precision falls to the Naive Bayes model (83.81%). In addition, there is a confusion matrix which shows that the Naive Bayes model trial is running well.

Keywords: Pandemic; Booster; Sentiment Analysis; Naive Bayes; Twitter

I. PENDAHULUAN

Dampak pandemi COVID-19 saat ini memengaruhi berbagai sektor pada negara-negara di dunia, terutama di Indonesia. Dampak yang pada umumnya dirasakan yaitu pada sektor bisnis, pendidikan, politik, sosial, dan budaya [1]. Salah satu kebijakan pemerintah pada masa pandemi COVID-19 adalah vaksinasi terhadap masyarakat. Vaksinasi yang dilakukan pada setiap orang adalah dua kali, dengan jangka waktu yang telah ditentukan. Sifat pada kebijakan vaksinasi yang diberlakukan adalah wajib bagi setiap orang [2]. Selain untuk menekan angka penyebaran pandemi, vaksinasi berfungsi sebagai akses apabila masyarakat akan menggunakan fasilitas umum. Saat ini, jumlah seluruh masyarakat yang telah melakukan vaksinasi penuh adalah 162,103,305 orang [3]. Pada awal tahun 2022, Indonesia memasuki tahap awal pemulihan kondisi yang disebabkan oleh pandemi. Pemerintah melakukan opsi bagi masyarakat untuk melakukan vaksin dosis ketiga (*booster*) [4]. Hal tersebut dilakukan untuk meningkatkan imunitas tubuh. Namun, pada kebijakan tersebut muncul sejumlah pro dan kontra di tengah masyarakat [5]. Penelitian ini menggunakan teknik analisis sentimen, karena pada penelitian ini akan melakukan klasifikasi terhadap opini masyarakat. Salah satu media sosial yang paling banyak digunakan oleh masyarakat saat ini adalah Twitter [6]. Twitter memuat banyak sekali opini masyarakat terkait suatu topik pembahasan. Opini yang terdapat pada Twitter pun cukup variatif, hal tersebut dikarenakan pengguna Twitter memiliki latar belakang dan berasal dari golongan masyarakat yang berbeda [7]. Maka dari itu, penelitian ini menggunakan Twitter sebagai sumber data utama. Penelitian disini juga mengacu pada penelitian yang dilakukan sebelumnya oleh Sukma, dkk, 2020 Penelitian tersebut membahas tentang analisis sentimen mengenai salah satu kebijakan pemerintah Indonesia (Omnibus Law) dengan sumber data yang berasal dari Twitter. Penelitian dilakukan dengan menggunakan metode *Support Vector Machine* (*SVM*), yang kemudian dibandingkan dengan dua metode lainnya, *Decision Tree* dan *Naive Bayes*, Hasil penelitian ini menunjukkan bahwa topik yang memiliki sentimen negatif terbanyak adalah 'kerja', sedangkan yang paling sedikit 'inovasi' [8]. Penelitian yang dilakukan sebelumnya oleh S. Lestari dan S.Saepudin, 2021 penelitian membahas tentang Analisis sentimen vaksin sinovac pada twitter menggunakan algoritma naïve bayes, memiliki hasil penelitian bahwa tweets dengan sentimen positif sebanyak 86%, sedangkan tweets dengan sentimen negatif sebnayak 14% dengan nilai akurasi dari perhitungan menggunakan algoritma Naïve Bayes adalah 92,96%, [9]. Penelitian ini mengacu pada penelitian yang dilakukan oleh Z. Firmansyah, N.F. Puspitasari, 2021 penelitian membahas tentang Analisis sentimen masyarakat terhadap vaksinasi covid 19 berdasarkan opini twitter menggunakan algoritma naïve bayes, hasil penelitian membahas metode confussion matrix mendapatkan nilai akurasi sebesar 78% dan pengujian menggunakan metode k-fold cross validation dengan nilai k sebanyak 5 (lima) perulangan mendapatkan nilai akurasi sebesar 80%, [10]. Penelitian ini mengacu pada penelitian yang dilakukan oleh A.Harun, D.P Ananda, 2021 penelitian membahas tentang Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve bayes

Building of Informatics, Technology and Science (BITS)
Volume 4, No 1, Juni 2022 Page: 106−109
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i1.1581

dan Decission Tree, hasil penelitian membahas cenderung ke tanggapan negatif dengan nilai akurasi 100.00% menggunakan algoritma NBC dan 50.39% menggunakan algoritma Decision Tree. [11] Berdasarkan model yang digunakan pada penelitian ini, penulis akan melakukan hal yang sama dengan menggunakan ketiga model tersebut pada penelitian ini, yaitu dengan membandingkan model *Naive Bayes* dengan *Decision Tree* dan *SVM*.

2. METODOLOGI PENELITIAN

2.1 Analisis Sentimen

Analisis sentimen merupakan sebuah pengamatan yang dikelola berdasarkan pendapat, sentimen, evaluasi, sikap, dan perasaan orang lain, terutama berdasarkan apa yang mereka tulis. Pendapat lain menyebutkan bahwa analisis sentimen dilakukan untuk mengetahui apa yang dipikirkan orang lain berdasarkan informasi, seperti opini [12]. Dari kedua pendapat tersebut, dapat dikatakan bahwa analisis sentimen dilakukan atas pendapat tertulis. Orang sering kali mengungkapkan dan menuliskan pendapatnya di media sosial karena perkembangan era digital yang membuat kita tidak bisa lepas dari media sosial.

2.2 Naive Bayes

Naive Bayes merupakan model yang digunakan untuk pengelompokan berdasarkan probabilitas pada setiap kelas berdasarkan pembagian kata dalam suatu dokumen. Metode ini memiliki akurasi yang baik digunakan untuk analisis sentimen dan telah diuji dengan beberapa algoritma lain [13]. Persamaan pada *Naive Bayes* pada umumnya berbentuk seperti berikut:

$$P(A|B) = P(B|A) * P(A) / P(B) \quad (1)$$

P(B|A) merupakan probabilitas pada label sebelumnya , P(B) merupakan probabilitas sebelumnya yang terjadi.
P(A) merupakan probabilitas sebelumnya yang telah diklasifikasikan sebagai label.

2.3 Twitter

Twitter merupakan salah satu media sosial yang sangat populer dengan lebih dari 300 juta pengguna di seluruh dunia pada tahun 2018. Melakukan penelitian dan analisis dengan data yang bersumber dari Twitter telah menjadi salah satu topik yang sangat digemari akhir-akhir ini. Salah satu topik yang banyak dipelajari dalam menganalisis data dari Twitter merupakan analisis sentimen. Selain itu, pada Twitter terdapat banyak opini, sehingga dapat memudahkan penelitian yang menggunakan algoritma klasifikasi [14]. Maka dari itu, penulis memilih Twitter sebagai sumber data.

2.4 Pengumpulan Data dari Twitter

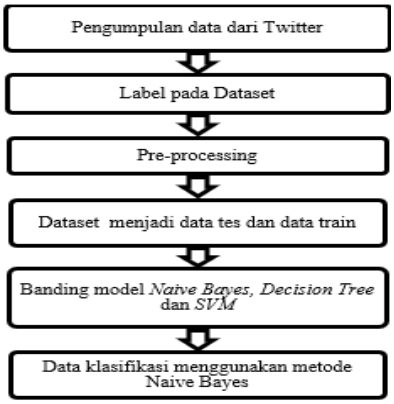
Tahap pengumpulan data dalam penelitian ini menggunakan teknik crawling. Teknik ini memanfaatkan API pada Twitter, sehingga akan didapatkan data tweet berdasarkan kata kunci tertentu. Data yang dikumpulkan adalah data tweet sepanjang tahun 2021. Pada dasarnya teknik ini memiliki alur yang sama dengan fitur pencarian pada website Twitter di *browser* [15]. Hasil yang diperoleh pada tahap ini adalah file tweet dengan 10,500 data opini.

2.5 Label pada Dataset

Pada tahap ini, tweet diberi label dalam dataset. Pelabelan dilakukan berdasarkan keadaan emosi pengguna yang terdapat pada setiap tweet [16]. Ada dua jenis klasifikasi yang ada pada dataset, yaitu positif (1) dan negatif (0). Karena proses ini melibatkan keadaan emosional, pelabelan dilakukan secara manual.

2.6 Alur Kerja Penelitian

Penelitian ini terdapat alur kerja dengan teknik analisis sentimen. Pada gambar 1 alur kerja terkait penelitian yang dilakukan



Gambar 1. Alur Kerja Penelitian

Building of Informatics, Technology and Science (BITS)
Volume 4, No 1, Juni 2022 Page: 106−109
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i1.1581

Alur kerja yang digunakan merupakan *framework* yang pada umumnya digunakan pada teknik analisis sentimen.

2.7 Preprocessing

Tahap *preprocessing* merupakan tahap yang cukup sulit untuk dilakukan. Tidak hanya pada penelitian ini, sebagian besar penelitian sejenis menggunakan banyak waktu untuk tahap ini. Ada beberapa tahapan yang dilakukan dalam *preprocessing*, antara lain *tokenizing*, *stop word removal*, dan *bag of words*. Pada tahap *tokenizing*, tanda baca dan URL dihapus. Selain itu, pemisahan kata dilakukan sehingga menjadi kata tersendiri (satu kata). Untuk *stop word removal* berfungsi untuk menghilangkan kata-kata yang tidak terlalu penting [17]. Dan untuk tahap *bag of words* dilakukan perubahan dokumen ke dalam bentuk vektor, yang memiliki nilai referensi dari 0 sampai 1. Dapat dikatakan bahwa pada tahap ini terdapat proses *mining* untuk setiap kata dalam setiap tweet [18].
Setelah dilakukan beberapa tahap *preprocessing*, data tersebut dibagi menjadi dua, dengan bobot 70% untuk *data train*, dan 30% untuk *data test*. Untuk tahap selanjutnya merupakan tahap klasifikasi dengan menggunakan *Naïve Bayes*. Yang dilanjutkan dengan melakukan perbandingan model, yaitu dengan model *SVM* dan *Decision Tree*.

3. HASIL DAN PEMBAHASAN

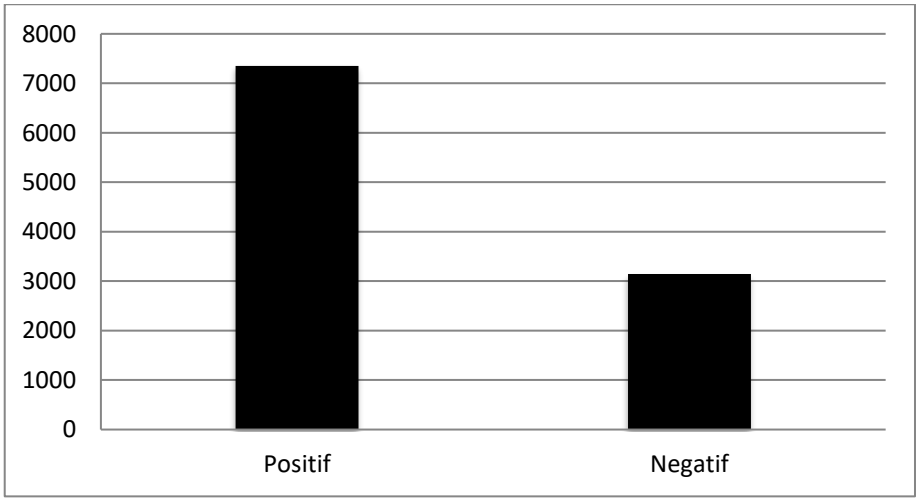
3.1 Evaluasi model Naive Bayes

Dilakukan evaluasi terhadap model *Naive Bayes* beserta perbandingannya, *Decision Tree* dan *SVM*. Evaluasi yang dilakukan menggunakan model *Hold Out*. Pada model *Hold Out*, terdapat skor akurasi, presisi, *recall*, F1, dan AUC. Pada Tabel 1 terdapat hasil yang telah didapatkan pada pengujian model.

Tabel 1. Perbandingan Evaluasi Model

Model	Accuracy	Precision	Recall	F1	AUC
Naive Bayes	70.00%	83.81%	67.78%	74.95%	71.06%
Decision Tree	79.00%	81.06%	89.09%	84.89%	74.15%
SVM	83.33%	79.97%	99.83%	88.80%	75.40%

Berdasarkan tabel perbandingan, model *SVM* memiliki skor AUC yang terbaik, yaitu sebesar 75.40%. Untuk urutan selanjutnya yaitu *Decision Tree*, dengan skor AUC sebesar 74.15%. Dan skor AUC yang terkecil merupakan *Naive Bayes*, yaitu sebesar 71.06%. Namun untuk skor presisi (precision) *Naive Bayes* memiliki skor yang terbesar. Hal tersebut berlaku karenakan pada *Naive Bayes* memiliki keunggulan yang baik untuk melakukan klasifikasi [16].
Pada Gambar 2 terdapat grafik pada jumlah tweet dengan sentimennya masing-masing, yaitu positive dan negative. Untuk positif berjumlah 7352, dan untuk negatif berjumlah 3148.



Gambar 2. Grafik Sentimen pada Dataset

Pada penelitian ini, disajikan hasil terakit topik yang paling banyak dibahas oleh publik di Twitter. Tentunya, topik tersebut merupakan peluang kerja di Indonesia selama masa pandemi COVID-19. Pada tahap uji coba model dengan *Naive Bayes*, terdapat hasil yang disajikan dalam bentuk *confusion matrix* pada Tabel 2.

Tabel 2. Hasil Uji Coba Model *Naive Bayes*

Model	Positive	Negative
Positive	1224 (TP)	114 (FP)
Negative	210 (FN)	252 (TN)

Building of Informatics, Technology and Science (BITS)
Volume 4, No 1, Juni 2022 Page: 106−109
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i1.1581

Untuk *true positive* (TP) memiliki jumlah terbanyak, yaitu sebesar 1224, serta *true negative* (TN) sebesar 252. Jumlah tersebut lebih banyak daripada *false poositive* (FP) dan *false negative* (FN), sehingga dapat dikatakan bahwa model *Naive Bayes* yang telah diuji pada penelitian ini berjalan dengan baik.

4. KESIMPULAN

Hasil pada penelitian ini menunjukkan bahwa model *SVM* memiliki kinerja rata-rata yang terbaik jika dibandingkan dengan dua model lainnya. Namun, untuk skor presisi model *Naive Bayes* menempati urutan yang terbaik, yaitu sebesar 83.81%. Selain itu, dilakukan perhitungan terhadap jumlah sentimen masing-masing yang ada pada dataset. Untuk positive berjumlah 2248, dan untuk negative berjumlah 751. Hasil yang terdapat pada confusion matrix menunjukkan bahwa uji coba model *Naive Bayes* berjalan dengan baik, dengan TP sebesar 1224, dan TN sebesar 252. Jumlah tersebut jauh lebih banyak jika dibandingkan dengan FP (114) dan FN (210).

REFERENCES

[1] D. F. Murad, R. Hassan, Y. Heryadi, B. D. Wijanarko, and Titan, “The Impact of the COVID-19 Pandemic in Indonesia (Face to face versus Online Learning),” Proceeding - 2020 3rd Int. Conf. Vocat. Educ. Electr. Eng. Strength. Framew. Soc. 5.0 through Innov. Educ. Electr. Eng. Informatics Eng. ICVEE 2020, pp. 4–7, 2020, doi: 10.1109/ICVEE50212.2020.9243202.

[2] N. A. Heartman, Y. Thahira, R. R. Ovelin and A. Helen, "Comparison of Adolescent Vaccination Data Accuracy by Urban Village in DKI Jakarta Province in July 2021 Using Several Data Mining Methods," 2021 International Conference on Artificial Intelligence and Big Data Analytics, 2021, pp. 82-87, doi: 10.1109/ICAIBDA53487.2021.9689724.

[3] Kementerian Kesehatan Republik Indonesia, “Vaksinasi COVID-19 Nasional,” Kemkes.go.id, 2022. [Online]. Available: https://vaksin.kemkes.go.id/#/vaccines [Accessed: 9-April-2021].

[4] A. Budiyanto, M. Anggraini and A. N. Hidayanto, "Predictive Analytics Comparison of Achieving Herd Immunity from COVID-19 in Indonesia and India Based on Fully Vaccinated People," 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE), 2022, pp. 289-294, doi: 10.1109/ISMODE53584.2022.9742823.

[5] R. R. K. Winahyu, M. Somantri and O. D. Nurhayati, "Predicting Creditworthiness of Smartphone Users in Indonesia during the COVID-19 pandemic using Machine Learning," 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE), 2022, pp. 223-227, doi: 10.1109/ISMODE53584.2022.9742831.

[6] W. Yue and L. Li, “Sentiment analysis using word2vec-cnn-bilstm classification,” 2020 7th Int. Conf. Soc. Netw. Anal. Manag. Secur. SNAMS 2020, pp. 3–7, 2020, doi: 10.1109/SNAMS52053.2020.9336549.

[7] Kusrini and M. Mashuri, “Sentiment analysis in twitter using lexicon based and polarity multiplication,” Proceeding - 2019 Int. Conf. Artif. Intell. Inf. Technol. ICAIIT 2019, pp. 365–368, 2019, doi: 10.1109/ICAIIT.2019.8834477.

[8] E. A. Sukma, A. N. Hidayanto, A. I. Pandesenda, A. N. Yahya, P. Widharto, and U. Rahardja, “Sentiment Analysis of the New Indonesian Government Policy (Omnibus Law) on Social Media Twitter”, Proc. - 2nd Int. Conf. Informatics, Multimedia, Cyber, Inf. Syst. ICIMCIS 2020, pp. 153–158, 2020, doi: 10.1109/ICIMCIS51567.2020.9354287

[9] S. Lestari dan S.Saepudin, “Analisis sentimen vaksin sinovac pada twitter menggunakan algoritma naïve bayes”, 2021, SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika) Universitas Nusa Putra, 7 Agustus 2021

[10] Z. Firmansyah, N.F. Puspitasari, “Analisis sentimen masyarakat terhadap vaksinasi covid 19 berdasarkan opini twitter menggunakan algoritma naïve bayes” 2021, Jurnal Teknik Informatika Vol. 14 No. 2, Oktober 2021 ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online) doi: https://doi.org/10.15408/jti.v14i2.24024

[11] A.Harun, D.P Ananda, “Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve bayes dan Decission Tree” 2021, Indonesian Journal of Machine Learning and Computer Science, Vol. 1 No. 1 (2021): MALCOM April 2021 .

[12] F. A. Wenando, R. Hayami, Bakaruddin, and A. Y. Novermahakim, “Tweet Sentiment Analysis for 2019 Indonesia Presidential Election Results using Various Classification Algorithms,” Proceeding - 1st Int. Conf. Inf. Technol. Adv. Mech. Electr. Eng. ICITAMEE 2020, pp. 279– 282, 2020, doi: 10.1109/ICITAMEE50454.2020.9398513.

[13] Z. Ferdous, I. Asad, and S. R. Deeba, “Analyzing the Factors Contributing to Graduate Unemployment,” 2019 IEEE Glob. Humanit. Technol. Conf. GHTC 2019, pp. 1–4, 2019, doi: 10.1109/GHTC46095.2019.9033029.

[14] R. Mulaudzi and R. Ajoodha, “An Exploration of Machine Learning Models to Forecast the Unemployment Rate of South Africa: A Univariate Approach,” 2020 2nd Int. Multidiscip. Inf. Technol. Eng. Conf. IMITEC 2020, 2020, doi: 10.1109/IMITEC50163.2020.9334090.

[15] S. H. Lee, Y. W. Cho, E. T. Im, and G. Y. Gim, “A Study on Customer Satisfaction Analysis of Public Institutions using Social Textmining,” Proc. - 20th IEEE/ACIS Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. SNPD 2019, pp. 385–394, 2019, doi: 10.1109/SNPD.2019.8935791.

[16] N. Tabassum and M. I. Khan, “Design an Empirical Framework for Sentiment Analysis from Bangla Text using Machine Learning,” 2nd Int. Conf. Electr. Comput. Commun. Eng. ECCE 2019, pp. 1–5, 2019, doi: 10.1109/ECACE.2019.8679347.

[17] H. Parveen and S. Pandey, “Sentiment analysis on Twitter Data-set using Naive Bayes algorithm,” Proc. 2016 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. iCATccT 2016, pp. 416–419, 2017, doi: 10.1109/ICATCCT.2016.7912034.

[18] G. Singh, B. Kumar, L. Gaur, and A. Tyagi, “Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification,” 2019 Int. Conf. Autom. Comput. Technol. Manag. ICACTM 2019, pp. 593– 596, 2019, doi: 10.1109/ICACTM.2019.8776800.