

Implementasi *Naive Bayes* pada Analisis Sentimen Opini Masyarakat di Twitter Terhadap Kondisi *New Normal* di Indonesia

E A Lisangan^{*1}, A Gormantara², R Y Carolus^{3,4}

^{1,2,3} Teknik Informatika, Fakultas Teknologi Informasi, Universitas Atma Jaya Makassar

⁴ Magister Informatika, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta

E-mail: erick_lisangan@lecturer.uajm.ac.id¹, alfredo_gormantara@lecturer.uajm.ac.id², 215311546@students.uajy.ac.id⁴

Abstrak. Pandemi yang melanda dunia saat ini menyebabkan masyarakat perlu beradaptasi dalam melaksanakan aktivitas sehari-hari atau yang dikenal dengan istilah kondisi “*New Normal*”. Analisis sentimen dibutuhkan agar pemerintah dapat mengetahui respon masyarakat terhadap kebijakan yang dikeluarkan dalam menanggulangi penularan Covid-19. Pada penelitian ini akan diimplementasikan algoritma *Naive Bayes* dan dipadukan dengan *metode stemming Sastrawi*. Terdapat beberapa proses yang dilalui dimulai dari data *crawling* untuk mengumpulkan *dataset*. Setelah *dataset* terkumpul kemudian melalui proses data *preprocessing*, *feature extraction*, dan klasifikasi menggunakan *Naive Bayes*. Pengujian algoritma menggunakan *confusion matrix* dengan memperhatikan nilai akurasi, *precision*, dan *recall*. Hasil pengujian terhadap *ratio* data *training* dan *testing* diperoleh *ratio* terbaik 70% dan 30% dengan nilai *accuracy*, *precision*, dan *recall* secara berturut-turut sebesar 94.55%, 93.55%, dan 93.55%. Model analisis sentimen diharapkan dapat membantu pemerintah dalam memperoleh *feedback* terhadap kebijakan terkait Covid-19 yang dikeluarkan,

Kata kunci: sentimen analisis, *Naive Bayes classifier*, *Sastrawi stemming*

Abstract. The pandemic that is engulfing the world today causes people to need to adapt in carrying out daily activities or what is known as the “*New Normal*” condition. Sentiment analysis is needed so that the government can find out the public's response to the policies issued in tackling the transmission of Covid-19. In this research, the *Naive Bayes* algorithm will be implemented and combined with the literary *stemming* method. There are several processes that are passed starting from data *crawling* to collecting *dataset*. After the *dataset* is collected, it then goes through the process of data *pre-processing*, *feature extraction*, and classification using *Naive Bayes*. Testing the algorithm using a *confusion matrix* by paying attention to the values of *accuracy*, *precision*, and *recall*. The results of the test on the *ratio* of training and testing data obtained the best *ratio* of 70% and 30% with the values of *accuracy*, *precision*, and *recall* respectively 94.55%, 93.55%, and 93.55%. The sentiment analysis model is expected to assist the government in obtaining *feedback* on policies related to Covid-19 issued.

Keywords: sentiment analysis, *Naive Bayes classifier*, *Sastrawi stemming*

1. Pendahuluan

Pandemi yang melanda dunia saat ini menyebabkan masyarakat perlu beradaptasi dalam melaksanakan aktivitas sehari-hari atau yang dikenal dengan istilah kondisi “*New Normal*”. Pemerintah Indonesia telah mengeluarkan berbagai kebijakan terkait situasi pandemi untuk mencegah penularan Covid-19 di tengah masyarakat, seperti Pembatasan Sosial Berskala Besar (PSBB) hingga Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM). Setiap kebijakan tentunya mendapatkan respon yang beragam dari masyarakat. Respon masyarakat inilah yang kemudian dijadikan masukan bagi pemerintah dalam menyempurnakan kebijakan yang telah dikeluarkan. Salah satu cara untuk menganalisis respon masyarakat dengan cepat adalah dengan memanfaatkan analisis sentimen.

Analisis sentimen telah banyak diimplementasikan pada berbagai isu pada beberapa bidang mulai dari sosial [1], ekonomi [2], politik [3], hingga kesehatan [4]. Analisis sentimen merupakan suatu bidang ilmu yang dapat mendapat membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks menggunakan *Natural Language Processing (NLP)*. Analisis sentimen juga dikenal sebagai penambangan opini yang mempelajari sentimen orang terhadap hal-hal tertentu [5].

Analisis sentimen terhadap isu pandemi Covid-19 telah diteliti sebelumnya seperti pada [6-10]. Pada [6] mengimplementasikan algoritma *Support Vector Machine (SVM)* untuk analisis sentimen kondisi *New Normal* di Indonesia. Hasil penelitian dengan memanfaatkan algoritma *stemming*. Nazief & Adriani menunjukkan nilai *accuracy* 76.5 %, *recall* 90.91 %, dan *precision* 70.80%. Pada [7] mengkolaborasi algoritma klasifikasi *lexicon-based* dan *naïve bayes classifier* untuk menganalisis opini publik terhadap Covid-19 secara umum dan diperoleh hasil pengujian dengan tingkat akurasi sebesar 99,48%, persentase *recall* sebesar 98,59% dan persentase *precision* sebesar 100%.

Pada [8] mengimplementasikan *Naive Bayes* untuk menganalisis sentimen penanganan Covid-19 di Indonesia selama periode April-Juni 2020 dan hasil pengujian memperoleh nilai akurasi sebesar 89.13%. Pada [9] melakukan analisis terhadap opini terkait vaksinasi Covid dengan membandingkan metode *Naive Bayes* dan *SVM* dimana diperoleh hasil klasifikasi *Naive Bayes* lebih baik dari segi akurasi dibandingkan *SVM* dengan nilai 85.59% berbanding 84.41%. Pada [10] membandingkan implementasi algoritma *Naive Bayes* dan *k-Nearest Neighbour (kNN)* untuk analisis sentimen terkait Covid-19 secara umum. Hasil pengujian menunjukkan nilai akurasi sebesar 85% untuk algoritma *Naive Bayes* dan 82% untuk algoritma *K-Nearest Neighbor*.

Berdasarkan penelitian sebelumnya dapat diperoleh bahwa algoritma *Naive Bayes* mampu menghasilkan nilai akurasi yang lebih dibandingkan *SVM* dan *K-NN*. Sumber data yang digunakan adalah menggunakan media sosial *Twitter*. Hal ini disebabkan karena terdapat 175,4 juta pengguna internet di Indonesia. Menurut statistik yang dikeluarkan oleh *We Are Social (2020)*, terdapat 126 juta orang pengguna aktif setiap hari yang menggunakan *Twitter* [11].

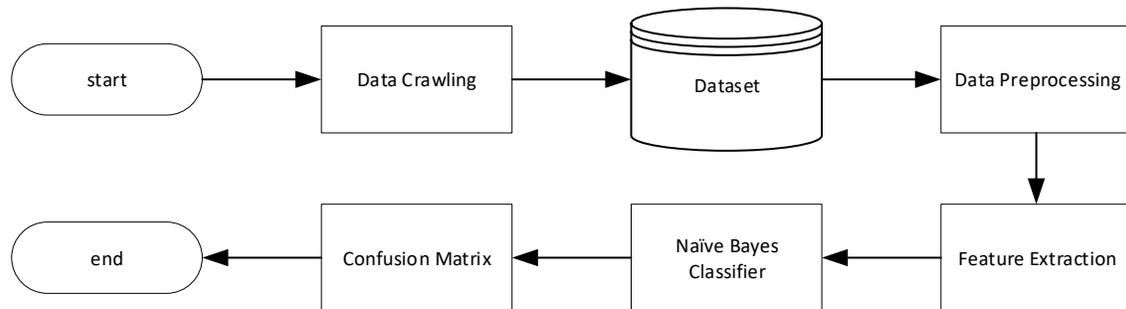
Pada [12] diimplementasikan *library* Sastrawi sebagai *stemming* untuk melakukan *text preprocessing* pada data komplain mahasiswa. Berdasarkan hasil penelitian dapat diperoleh bahwa *library* Sastrawi menunjukkan hasil *stemming* mencapai 92% pada kategori *Exact Match* dibandingkan dengan *Porter Stemming*.

Pada penelitian ini akan diimplementasikan algoritma *Naive Bayes* dan dipadukan dengan metode *stemming* Sastrawi dalam mengetahui opini publik terhadap kondisi *New Normal* di Indonesia. Implementasi algoritma akan menggunakan *Python* sebagai bahasa pemrograman dan *tools RapidMiner* untuk proses *crawling* data *tweet* pada *Twitter*. Model analisis sentimen yang dirancang diharapkan dapat membantu pemerintah dalam memperoleh *feedback* dari opini masyarakat terhadap kebijakan terkait Covid-19 yang dikeluarkan.

2. Metode

Alur kerja analisis sentimen yang digunakan pada penelitian ini dapat dilihat pada Gambar 1. Terdapat beberapa proses yang dilalui dimulai dari data *crawling* untuk mengumpulkan *dataset*. Setelah *dataset*

terkumpul kemudian melalui proses *data preprocessing*, *feature extraction*, dan klasifikasi menggunakan *Naive Bayes*. Pengujian algoritma menggunakan *confusion matrix* dengan memperhatikan nilai akurasi, *precision*, dan *recall*.



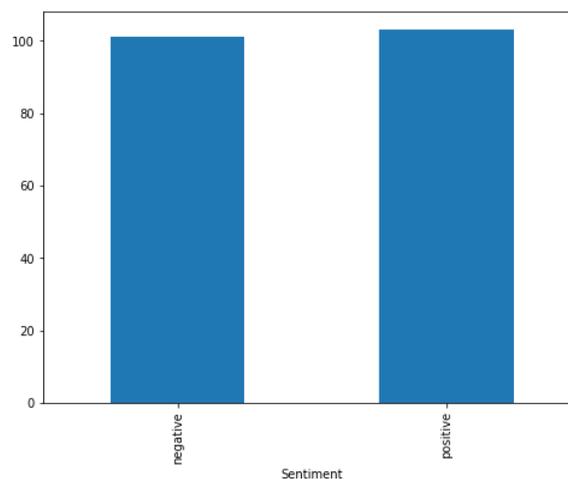
Gambar 1. Proses Analisis Sentimen

2.1. Data Crawling

Pada proses ini dilakukan pengumpulan data *tweet* menggunakan tools *RapidMiner*. Beberapa contoh *hashtag* yang digunakan adalah *#newnormal*, *#psbb*, hingga *#lockdown*. *Dataset* yang terkumpul sebanyak 204 *tweet* pada periode April-Mei 2020. *Tweet* tersebut kemudian diklasifikasikan menjadi 2 (dua) kelas, yaitu positif dan negatif, dengan jumlah *tweet* masing-masing 101 dan 100 *tweet* (Gambar 3).

	Sentiment	Text
0	negative	AirinAirinNZ Bagaimana pun kita sebagai masya...
1	negative	sincos07 new propethic normal kenormalan baru...
2	negative	muhammadiyah TVMuhammadiyah pres conference te...
3	negative	PTTransjakarta Besok new normal jam operasiona...
4	negative	just a few more days facing new normal Alias g...

Gambar 2. Sampel Data *Tweet* yang Diperoleh



Gambar 3. Distribusi Sentimen *Dataset*

2.2. Data Preprocessing.

Setelah diperoleh *dataset* kemudian dilakukan beberapa tahapan *preprocessing*, yaitu *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Pada *cleaning* dilakukan pembersihan terhadap beberapa simbol dan angka, seperti tanda pagar (#), tanda penghubung, hingga tanda baca. Pada proses *case folding* dilakukan perubahan struktur kata dari huruf besar menjadi huruf kecil.

Pada proses *tokenizing* dilakukan pemisahan kata dalam kalimat dengan menggunakan spasi sebagai indikator pemecah setiap kata. Selanjutnya dilakukan penghapusan beberapa kata yang kurang memiliki makna pada *stopword removal*. Kata-kata yang dihapus seperti “yang”, “dan”, “di”, dan “akan”. Tahapan terakhir *preprocessing* adalah *stemming*. Pada proses ini dilakukan penghilangan *sufiks*, *prefix* dan *konfiks* sehingga kata yang diperoleh berbentuk kata dasar. Pada penelitian ini menggunakan metode *stemming* Sastrawi [13]. Contoh hasil *data pre-processing* dapat dilihat pada Tabel 1.

Tabel 1. Hasil *Text Preprocessing*

Data Text	Sesudah <i>Preprocessing</i>
Biaya test PCR bisa lebih mahal dari harga tiketnya #COVID19indonesia https://t.co/DVlymxdXnN	biaya <i>test pcr</i> lebih mahal harga tiket covid19indonesia
BREAKING NEWS: Gugus Tugas Percepatan Penanganan Covid-19 di DKI Jakarta memutuskan untuk menetapkan status PSBB DKI Jakarta diperpanjang dan menetapkan bulan Juni sebagai masa transisi. #CariBeritaditvOne #diRumahNontontvOne #BreakingNews #COVID19Indonesia #PSBB #MasaTransisi https://t.co/oBXocQNkhR .	breaking news: gugus tugas cepat tangan covid-19 dki jakarta putus tetap status psbb dki jakarta panjang tetap bulan juni masa transisi. cariberitaditvone dirumahnontontvone breakingnews covid19indonesia psbb masatransisi.

2.3. Feature Extraction.

Pada tahapan ini dilakukan proses pembobotan terhadap setiap kata diperoleh dari *text preprocessing* menggunakan TF-IDF (*Term Frequency Inverse document frequency*) merupakan suatu pembobotan berasal dari kandungan kata tertentu pada sebuah dokumen. Tahapan pembobotan dengan TF-IDF adalah sebagai berikut [14], [15]:

1. Hitung term frequency $tf_{t,d}$ dengan membandingkan jumlah kemunculan sebuah kata dengan jumlah keseluruhan kata dalam sebuah dokumen.
2. Hitung *weighting term frequency* atau bobot *frekuensi term* ($Wtf_{t,d}$) menggunakan Persamaan (1)

$$Wtf_{t,d} = 1 + \log tf_{t,d} \quad (1)$$

3. Hitung *document frequency* (df) atau jumlah frekuensi dokumen yang mengandung term.

4. Hitung bobot *inverse document frequency* (*idf*) menggunakan Persamaan (2), dimana N adalah jumlah keseluruhan dokumen.

$$idf_t = \log \frac{N}{df_t} \quad (2)$$

5. Hitung nilai bobot TF-IDF menggunakan Persamaan (3) dimana $W_{t,d}$ adalah bobot TF-IDF.

$$W_{t,d} = W_{tf_{t,d}} + idf_t \quad (3)$$

2.4. Naive Bayes Classifier.

Naive Bayes Classifier merupakan salah satu metode *machine learning* yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [16]. Dasar dari *Naive Bayes* yang dipakai dalam pemrograman adalah rumus Bayes pada persamaan (4). Peluang kejadian A sebagai B atau $P(A|B)$ ditentukan dari peluang B saat A atau $P(B|A)$, peluang A atau $P(A)$, dan peluang B atau $P(B)$.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad (4)$$

2.5. Confusion Matrix.

Untuk pengukuran hasil klasifikasi metode *Naive Bayes Classifier*, digunakan *Confusion Matrix*. *Confusion matrix* adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data *mining*. Evaluasi dengan menghasilkan nilai *accuracy*, *precision* dan *recall* [17].

fij		Kelas Hasil Prediksi (j)	
		Kelas (+)	Kelas (-)
Kelas Asli (i)	Kelas (+)	TP (<i>True Positive</i>)	FN (<i>False Negative</i>)
	Kelas (-)	FP (<i>False Positive</i>)	TN (<i>True Negative</i>)

Gambar 4. Confusion matrix

Accuracy dalam klasifikasi adalah persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi.

Precision adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya.

Recall adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar.

Persamaan :

$$\text{Recall} = TP / (TP + FN)$$

$$\text{Precision} = TP / (TP + FP)$$

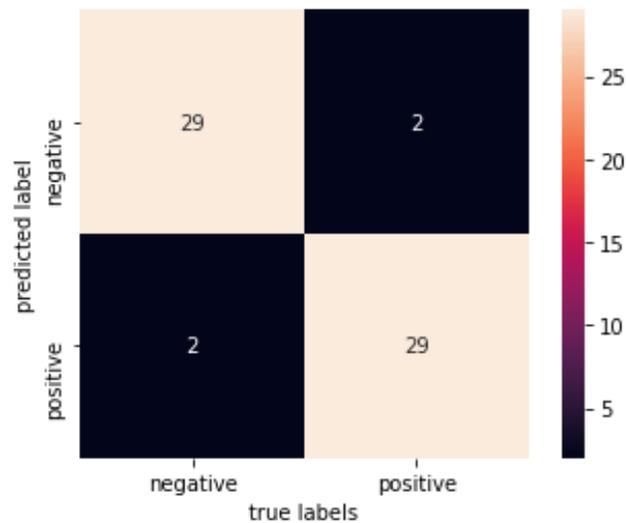
$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN)$$

3. Hasil dan Pembahasan

Berdasarkan *dataset* yang diperoleh kemudian dilakukan pembagian data *training* dan *testing* dimana digunakan beberapa perbandingan pada penelitian ini, seperti 70%:30%, 75% dan 25%, 80% berbanding

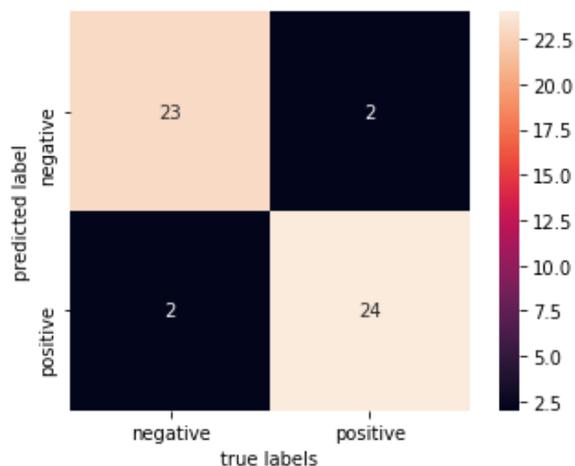
20%, serta 90%:10%. Algoritma untuk analisis sentimen dirancang menggunakan bahasa pemrograman *Python*. Hasil analisis pengujian terhadap perbandingan *data training* dan *data testing* yang dapat dilihat pada Tabel 2.

Berdasarkan pengujian terhadap penggunaan *ratio* 70%: 30% (Gambar 5) dapat dilihat bahwa nilai TP dan TN adalah sebanyak 29, serta FP dan FN sebesar 2. Hasil perhitungan nilai *recall* diperoleh sebesar $29 / (29+2) = 0.9355$, nilai *precision* sebesar $29 / (29+2) = 0.9355$. Sedangkan untuk nilai akurasi diperoleh nilai sebesar $(29+29) / (29+2+2+29) = 0.9455$ atau 94.55%. Berdasarkan hasil pengujian diperoleh nilai akurasi yang cukup baik sebesar 94.55%.



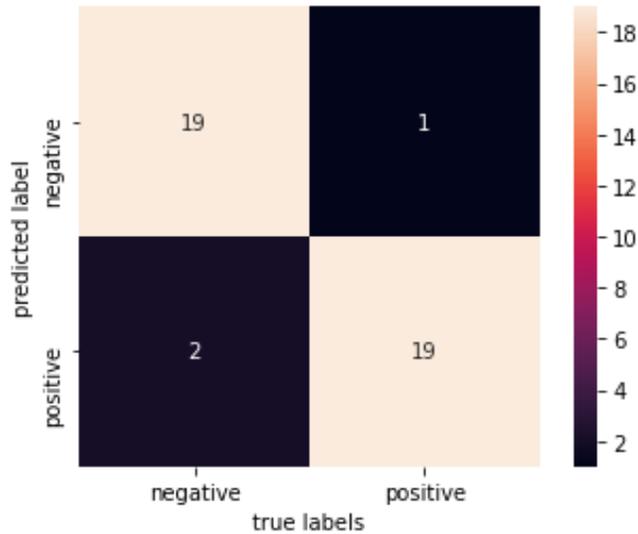
Gambar 5. *Confusion Matrix* untuk *Ratio* Perbandingan 70%:30%

Berdasarkan pengujian terhadap penggunaan *ratio* 75%: 25% (Gambar 6) dapat dilihat bahwa nilai TP adalah sebanyak 24, TN sebesar 23, serta FP dan FN sebesar 2. Hasil perhitungan nilai *recall* diperoleh sebesar $19 / (19+1) = 0.9231$, nilai *precision* sebesar $24 / (24+2) = 0.9231$. Sedangkan untuk nilai akurasi diperoleh nilai sebesar $(24+23) / (24+2+2+23) = 0.9216$ atau 92.16%. Berdasarkan hasil pengujian diperoleh nilai akurasi yang cukup baik sebesar 92.16%.



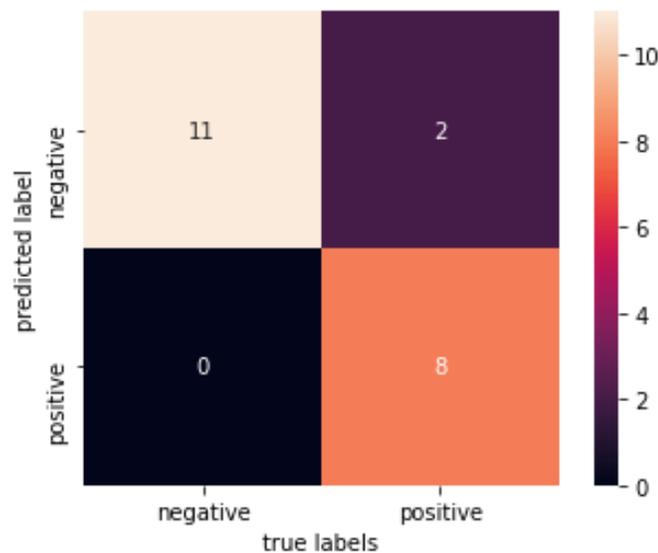
Gambar 6. *Confusion Matrix* untuk *Ratio* Perbandingan 75%:25%

Berdasarkan pengujian terhadap penggunaan ratio 80%:20% (Gambar 7), dapat dilihat bahwa nilai TP dan TN adalah sebanyak 19, nilai FP sebesar 1, dan FN sebesar 2. Hasil perhitungan nilai *recall* diperoleh sebesar $19 / (19+2) = 0.9047$, nilai *precision* sebesar $19 / (19+1) = 0.9231$. Sedangkan untuk nilai akurasi diperoleh nilai sebesar $(19+19) / (19+2+1+19) = 0.9267$ atau 92.67%. Berdasarkan hasil pengujian diperoleh nilai akurasi yang cukup baik sebesar 92.67%.



Gambar 7. Confusion Matrix untuk Ratio Perbandingan 80%:20%.

Berdasarkan pengujian terhadap penggunaan *ratio* 90%:10% (Gambar 8) dapat dilihat bahwa nilai TP sebesar 8 dan TN adalah sebanyak 11, nilai FP sebesar 2, dan FN sebesar 0. Hasil perhitungan nilai *recall* diperoleh sebesar $8 / (8+0) = 1.000$, nilai *precision* sebesar $8 / (8+2) = 0.800$. Sedangkan untuk nilai akurasi diperoleh nilai sebesar $(8+11) / (8+2+0+11) = 0.9047$ atau 90.47%. Berdasarkan hasil pengujian diperoleh nilai akurasi yang cukup baik sebesar 90.47%.



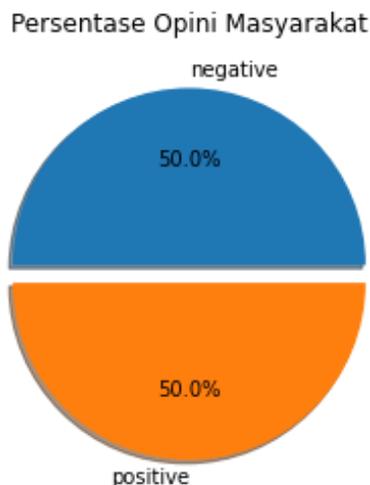
Gambar 8. Confusion Matrix untuk Ratio Perbandingan 90%:10%.

Pada Tabel 2 dapat dilihat bahwa nilai *accuracy*, *precision*, dan *testing* memiliki nilai yang saling mendekati untuk masing-masing *ratio* perbandingan. Hal ini dapat disebabkan karena distribusi kelas pada *dataset* hampir memiliki jumlah yang seimbang, yaitu 101 untuk positif dan 100 untuk negatif. Nilai akurasi terbaik diperoleh untuk *ratio* perbandingan 70% dan 30% sebesar 94.55%.

Tabel 2. Hasil pengujian terhadap ratio data training dan testing.

Perbandingan	70%:30%	75%:25%	80%:20%	90%:10%
Accuracy	0.9455	0.9216	0.9267	0.9047
Precision	0.9355	0.9231	0.9500	0.8000
Recall	0.9355	0.9231	0.9047	1.000

Pada Gambar 9 dapat dilihat visualisasi terhadap hasil analisis sentimen opini masyarakat untuk *ratio* perbandingan 70%:30%. Dapat dilihat bahwa opini masyarakat seimbang terhadap positif dan negatif terhadap kebijakan *New Normal* di Indonesia. Model analisis sentimen yang dirancang diharapkan dapat membantu pemerintah dalam memperoleh *feedback* dari opini masyarakat terhadap kebijakan terkait Covid-19 yang dikeluarkan. Berdasarkan *feedback* tersebut dapat dijadikan masukan bagi pemerintah dalam membuat kebijakan yang lebih baik lagi di masa depan.



Gambar 9. Tampilan Persentase Opini Masyarakat.

Pada penelitian ini juga dilakukan perbandingan terhadap penelitian sebelumnya [6] dimana diperoleh nilai akurasi terbaik sebesar nilai akurasi untuk setiap perbandingan *ratio data training* dan *testing*. Hasil perbandingan dapat dilihat pada Tabel 3. Hasil perbandingan menunjukkan bahwa nilai *accuracy*, *precision*, dan *recall* yang dihasilkan oleh *Naive Bayes* lebih baik jika dibandingkan dengan *SVM*. Hal ini dapat disebabkan karena pada *Naive Bayes*, setiap kata memiliki probabilitas masing-masing untuk diklasifikasikan pada kelas tertentu. Sedangkan pada *SVM*, setiap kalimat diwakilkan dalam model vektor 2 dimensi atau dengan kata lain setiap kata berkaitan dengan kalimatnya.

Tabel 3. Hasil perbandingan antara *Naive Bayes* dan *SVM*.

Perbandingan	70%:30%		80%:20%		90%:10%	
	NB	SVM	NB	SVM	NB	SVM
Accuracy	0.9455	0.7650	0.9267	0.7500	0.9047	0.7607
Precision	0.9355	0.8710	0.9500	0.8676	0.8000	0.8163
Recall	0.9355	0.6207	0.9047	0.5900	1.000	0.6897

4. Kesimpulan

Implementasi *Naive Bayes* mampu melakukan analisis sentimen opini masyarakat terhadap isu *New Normal* di Indonesia. Hasil pengujian terhadap *ratio* data training dan testing diperoleh *ratio* terbaik 70% dan 30% dengan nilai *accuracy*, *precision*, dan *recall* secara berturut-turut sebesar 94.55%, 93.55%, dan 93.55%. Model analisis sentimen yang dirancang diharapkan dapat membantu pemerintah dalam memperoleh *feedback* dari opini masyarakat terhadap kebijakan terkait Covid-19 yang dikeluarkan. Hasil perbandingan dengan *SVM* menunjukkan bahwa *Naive Bayes* menghasilkan nilai pengujian yang lebih baik untuk setiap perbandingan *ratio*. Hal ini disebabkan karena setiap kata merupakan variabel independen yang memiliki nilai probabilitas terhadap kelas tertentu.

Referensi

- [1] M. Karamibekr and A. A. Ghorbani, "Sentiment Analysis of Social Issues," *2012 International Conference on Social Informatics*, 2012, pp. 215-221.
- [2] V. Kumar and Sri Khetwat Saritha. 2021, "Sentiment Time Series Analysis on US Economic News." *New Opportunities for Sentiment Analysis and Information Processing IGI Global*, pp. 253-268.
- [3] F. A. Wenando, R. Hayami, Bakaruddin and A. Y. Novermahakim, 2020, "Tweet Sentiment Analysis for 2019 Indonesia Presidential Election Results using Various Classification Algorithms," *2020 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering (ICITAMEE)*, pp. 279-282.
- [4] A. Zunic, P. Corcoran, and I. Spasic, 2020, "Sentiment Analysis in Health and Well-Being: Systematic Review," *JMIR Medical Informatics*, vol. 8, no. 1, pp. 1-22.
- [5] X. Fang and J. Zhan, 2015, "Sentiment analysis using product review data," *Journal of Big Data*, vol. 2, no. 1, pp. 1-14.
- [6] A. Gormantara, 2020, "Analisis Sentimen Terhadap New Normal Era di Indonesia pada Twitter Menggunakan Metode Support Vector Machine," *Konferensi Nasional Ilmu Komputer (KONIK) 2020*, pp. 1-5.
- [7] E. Sutanty, and M. B. Siregar, 2020, "Analisis Sentimen Opini Publik Mengenai Covid-19 Melalui Platform Twitter Dengan Metode Naive Bayes Classifier," Repository Universitas Gunadarma, (<https://library.gunadarma.ac.id/repository/analisis-sentimen-opini-publik-mengenai-covid-19-melalui-platform-twitter-dengan-metode-nave-bayes-classifier-penelitian>, diakses 10 Januari 2022).
- [8] T. Yulianita, T. W. Utami, and M. A. Haris, 2020, "Analisis Sentimen Dalam Penanganan Covid-19 di Indonesia Menggunakan Naive Bayes Classifier," *Seminar Nasional Variansi*, pp. 235-243.
- [9] B. Laurensz, and E. Sedyono, 2021, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 10, no. 2, pp. 118-123.

- [10] A. L. Fairuz, R. D. Ramadhani , N. A. F. Tanjung, 2021, “Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter.” *Indonesian Journal of Data Science, IOT, Machine Learning and Artificial Intelligence*, vol. 1, no. 1, pp. 1-10.
- [11] A. Tri, R. Dirilis, I. Harga, and R. Note, “Non-Relieving Regulators Riset : Ada 175 , 2 Juta Pengguna Internet di Indonesia,” 2020. (<https://inet.detik.com/cyberlife/d-4907674/riset-ada-1752-juta-pengguna-internet-di-indonesia>, diakses 05 Januari 2022).
- [12] M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, dan H. A. Gozali, 2019, “Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi,” *2019 International Conference on Engineering, Technologies, and Applied Sciences*, pp. 1-6.
- [13] A. Librian, 2017, *High quality stemmer library for Indonesian Language (Bahasa)*, (<https://github.com/sastrawi>, diakses 20 Januari 2022).
- [14] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, 2018, “Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Univ. Brawijaya*, vol. 2, no. 11, pp. 4704–4713.
- [15] N. K. Widyasanti, I K. G. D. Putra, N. K. D. Rusjyanthi, 2018, “Seleksi Fitur Bobot Kata dengan Metode TFIDF untuk Ringkasan Bahasa Indonesia,” *Jurnal MERPATI*, vol. 6, no. 2, pp. 119-126.
- [16] S. L. B. Ginting, and R. P. Trinanda, 2013, “Teknik Data Mining Menggunakan Metode Bayes Classifier untuk Optimalisasi Pencarian pada Aplikasi Perpustakaan (Studi Kasus: Perpustakaan Universitas Pasundan Bandung),” *Jurnal Teknologi dan Informasi*, vol. 3, no. 2, pp. 37-50.
- [17] S. L. Ting, W. H. Ip, and A. H. C. Tsang, 2011, “Is Naive Bayes a Good Classifier for Document Classification?,” *International Journal of Software Engineering and Its Applications*, vol. 5, no. 3, pp. 37-46.