

Analisis Sentimen Media Sosial Twitter pada Kasus Pemberlakuan Pembatasan Kegiatan Masyarakat dengan menggunakan Metode Naïve Bayes Classifier

Muchammad Chandra Cahyo Utomo^{1, a)}, Mukhamad Tauhid^{2, b)}, dan Syamsul Mujahidin^{3, c)}

^{1,2,3}Prodi Informatika Institut Teknologi Kalimantan - Balikpapan, Kalimantan Timur.

^{a)}Penulis korespondensi: ccahyo@lecturer.itk.ac.id

^{b)}taukhid.muhammad31@gmail.com

^{c)}syamsul@lecturer.itk.ac.id

Abstrak. Media sosial merupakan media yang digunakan oleh pengguna untuk memperkenalkan diri, berinteraksi, berkolaborasi, dan berbagi informasi dengan pengguna lain menggunakan internet. Salah satu platform media sosial yang populer di Indonesia adalah Twitter. Twitter merupakan media sosial yang umumnya berfungsi sebagai pengirim pesan yang biasanya disebut sebagai kicauan atau *tweet*. Salah satu topik yang banyak dibahas adalah Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM), akibat dampak peningkatan kasus akibat munculnya COVID varian baru yaitu versi Omikron. salah satu tujuan penelitian ini yaitu untuk mengetahui hasil analisis sentimen mengenai opini masyarakat terhadap pemberlakuan pembatasan kegiatan masyarakat menggunakan metode *naive bayes*. Banyak teknik *machine learning* yang baik untuk analisis sentimen salah satunya adalah *Naive bayes classifier*, yang merupakan teknik pembelajaran mesin berdasarkan probabilitas. *Naive bayes classifier* adalah metode klasifikasi teks yang sederhana namun sangat akurat dan efektif yang pengklasifikasiannya sangat dipengaruhi oleh proses data latih. Data yang digunakan diambil melalui Twitter dengan 1594 *tweet*. *dataset* akan dibagi menjadi data pelatihan dan data pengujian. dengan membandingkan 90% pelatihan dan 10% pengujian. Sehingga rincian sebaran data yang digunakan dalam penelitian ini adalah 1434 *tweet* sebagai data latih dan 160 *tweet* sebagai data uji. Proses *Naive bayes classifier* yaitu melakukan *crawling* data, *pre-processing* data, pembagian data, pelabelan data, Klasifikasi model *naive bayes*, klasifikasi data latih. Hasil analisis sentimen opini masyarakat terhadap pemberlakuan pembatasan kegiatan masyarakat menggunakan metode *naive bayes* mendapatkan nilai sentimen sebanyak 53.5% negatif dan 46.5% positif, nilai akurasi 0,84, F1-Score 0,84, *precision* 0,85, dan nilai *recall* 0,84.

Kata Kunci : Analisis Sentimen, Naive Bayes, PPKM, Twitter

PENDAHULUAN

Penggunaan internet memungkinkan kita untuk berbagi informasi dengan cepat, hal ini dimanfaatkan oleh pengguna untuk saling berinteraksi berbagai kebutuhan atau hiburan. Perkembangan teknologi internet dan *smartphone*, juga mempengaruhi pesatnya perkembangan media sosial. kemudahan akses dalam mencari informasi membuat jejaring sosial menjadi fenomena yang sangat mempengaruhi arus informasi, hal ini terjadi hampir di seluruh dunia seperti Indonesia menjadi salah satu pengguna internet terbanyak. Media sosial yang saat ini termasuk sebagai media sosial dengan pengguna terbanyak di Indonesia adalah Twitter. Twitter umumnya berfungsi sebagai pengirim pesan yang biasanya disebut sebagai kicauan atau *tweet*. Indonesia menjadi salah satu pengguna Twitter terbanyak dengan 59% masyarakatnya menggunakan Twitter dan menjadi platform media sosial yang cukup

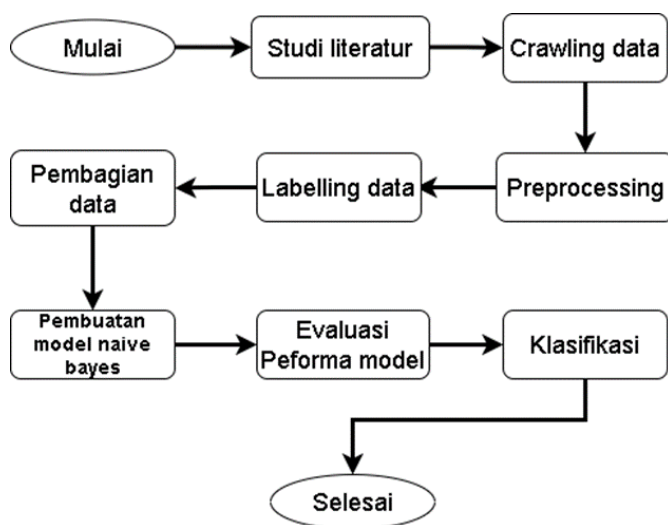
berpengaruh di Indonesia. Hal ini disebabkan tidak lain karna kemudahan untuk menyampaikan opini pengguna dalam membicarakan suatu topik masalah. [1]

Topik yang paling banyak diperbincangkan yaitu Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM), akibat dampak peningkatan infeksi varian baru COVID, khususnya COVID Omicron. Varian Omicron paling cepat menyebar dari pada varian sebelumnya, sehingga mendorong pemerintah untuk lebih memperluas kebijakan PPKM guna mencegah penyebaran virus ini. Hal ini menimbulkan banyak reaksi publik baik yang mendukung maupun menentang kebijakan tersebut. Untuk memahami sentimen atau opini publik terhadap kebijakan PPKM dapat dilakukan dengan analisis.[2]

Banyak teknik pembelajaran mesin yang bagus untuk klasifikasi analisis, salah satunya adalah pengklasifikasi *Naive Bayes Classifier* yaitu teknik pembelajaran mesin berbasis probabilitas. NBC merupakan metode klasifikasi teks yang sederhana namun sangat akurat dan efektif yang klasifikasinya sangat dipengaruhi oleh proses *training* data. [3]. Hal ini mendorong penulis melakukan penelitian analisis sentimen masyarakat mengenai PPKM dengan menggunakan metode *naive bayes classifier*.

METODOLOGI PENELITIAN

Diagram alir pada penelitian ini ditunjukkan seperti pada Gambar 1. Tahap awal penelitian dilakukan dengan melakukan *crawling* data pada Twitter menggunakan kata kunci “ppkm”. Kemudian *dataset* yang didapatkan dari hasil *crawling* data akan dilakukan tahap *pre-processing* dengan melakukan proses *cleaning text*, *casefolding*, *tokenization*, *normalization*, *stemming*, *filtering*, dan penghapusan kolom dan duplikat. Setelah *dataset* melalui tahap *pre-processing* tahap selanjutnya dilakukan pelabelan data dengan menggunakan kamus *lexicon* yang dibagi menjadi dua label, label positif dan negatif. Kemudian data hasil pelabelan akan dilakukan pembagian data menjadi dua data, data *training* dan data *testing*. [4] Pada data *training* digunakan untuk melatih model *naive bayes classifier* dan pada data *testing* digunakan untuk menguji model yang telah dibangun. Model yang telah dibangun dilakukan evaluasi performa model dengan melihat akurasi, presisi, dan *recall* dari hasil *testing*. kemudian model klasifikasi akan digunakan untuk analisis pada data pemberlakuan pembatasan kegiatan masyarakat. Setelah hasil analisis didapatkan maka penelitian selesai.



GAMBAR 1 Diagram Alir Penelitian

HASIL DAN PEMBAHASAN

Di sini kami memberikan beberapa saran dasar untuk memformat matematika Anda, tetapi kami tidak mencoba untuk mendefinisikan gaya atau spesifikasi rinci untuk pengaturan huruf matematika. Anda harus menggunakan gaya, simbol, dan konvensi standar untuk bidang/disiplin yang Anda tulis.

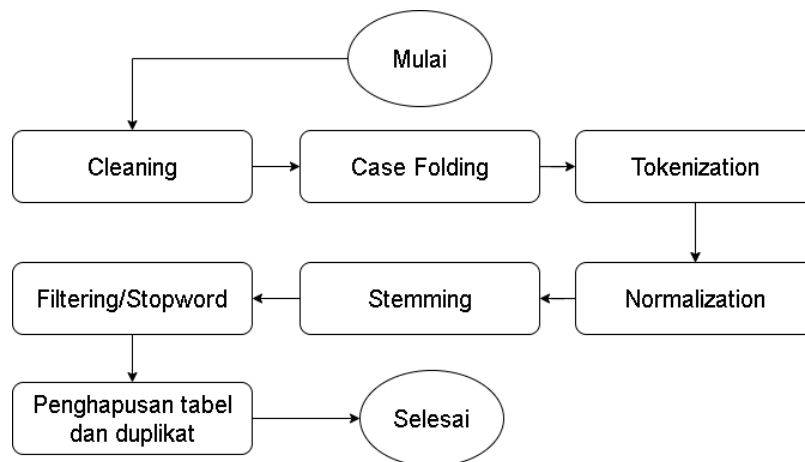
Crawling Data

Crawling data Twitter dilakukan untuk mengambil dataset tweet yang ada pada Twitter dengan menggunakan API Twitter. Crawling data dimulai dari tanggal 01 April 2022 hingga tanggal 03 Juni 2022, proses crawling menggunakan kata kunci “ppkm”, selama proses *crawling* didapatkan data *tweet* sebanyak 3504 data *tweet*. Hasil *crawling dataset* yang telah dilakukan akan disimpan ke dalam sebuah *file csv* dengan kolom waktu, *user*, dan *tweet*. Pada proses *crawling* data didapatkan data *tweet* yang juga mengambil kalimat yang tidak diperlukan seperti *link*, *username*, tanda baca, *emoticon*, dan juga *retweet* yang berarti *dataset* memiliki kalimat kembar atau *duplicate* yang mana hal ini dapat mempengaruhi hasil dari klasifikasi sehingga perlu dilakukan pembersihan data sebelum data dapat digunakan pada proses klasifikasi.

Pre-processing

Pada Tahap *Pre-Processing*, *dataset* yang telah didapat melalui proses *crawling* akan dibersihkan dengan cara membuang data yang tidak diperlukan agar sistem dapat dengan mudah memahami data. Ada beberapa tahapan dalam *pre-processing* yaitu *cleaning text*, *casefolding*, *tokenization*, *normalization*, *stemming*, *filtering*, penghapusan kolom dan tabel kosong.[5] *Pre-processing* yang dimaksud ditunjukkan pada Gambar 2.

. Tahap *pre-processing* dilakukan sesuai urutan agar setiap data Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) yang telah melalui tahap *pre-processing* dapat lebih efektif saat diolah oleh algoritma komputer. Hasil *pre-processing* digunakan sebagai pembuatan model analisis. Hasil *pre-processing* ditunjukkan pada Tabel 1.



GAMBAR 2 Diagram Alir Tahap *Pre-processing*

TABEL 1 Tahap *Pre-processing*

Proses	Hasil
Data	@prastow Dari sektor industri, sngat dirugikan dimulai bulan
<i>Crawling</i>	April, mulai dri efek PPKM, Solar naik. bahan baku naik. Kesulitan
<i>Cleaning</i>	Dari sektor industri sngat dirugikan dimulai bulan April mulai dri efek PPKM Solar naik bahan baku naik Kesulitan

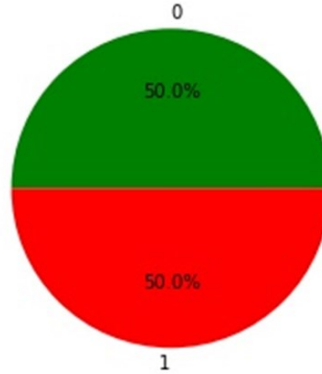
Proses	Hasil
<i>Case Folding</i>	dari sektor industri sngat dirugikan dimulai bulan april mulai dri efek ppkm solar naik bahan baku naik kesulitan
<i>Tokenization</i>	“dari”, “sektor”, “industri”, “sngat”, “dirugikan”, “dimulai”, “bulan”, “april”, “mulai”, “dri”, “efek”, “ppkm”, “solar”, “naik”, “bahan”, “baku”, “naik”, “kesulitan”
<i>Normalisasi</i>	“dari”, “sektor”, “industri”, “sangat”, “dirugikan”, “dimulai”, “bulan”, “april”, “mulai”, “dari”, “efek”, “ppkm”, “solar”, “naik”, “bahan”, “baku”, “naik”, “kesulitan”
<i>Stemming</i>	“dari”, “sektor”, “industri”, “sangat”, “rugi”, “mulai”, “bulan”, “april”, “mulai”, “dari”, “efek”, “ppkm”, “solar”, “naik”, “bahan”, “baku”, “naik”, “sulit”
<i>Filtering</i>	“sektor”, “industri”, “sangat”, “rugi”, “efek”, “ppkm”, “solar”, “baku”, “naik”, “sulit”

Labelling Data

Setelah menyelesaikan tahap *Pre-processing*, penulis melanjutkan penelitian hingga tahap pelabelan data. Untuk menentukan nilai sentimental suatu kata, penulis menggunakan kumpulan data *lexicon* positif dan negatif yang diperoleh. *Dataset* kosakata negatif memiliki nilai dengan skor -1 hingga -5, sedangkan *dataset* kosakata positif memiliki dengan skor 1 hingga 5.[6] Hasil proses pelabelan dari *dataset* yang digunakan dalam penelitian ini ditunjukkan pada Tabel 2. Data yang telah melalui proses *labeling* di mana terdapat 897 data negatif dan juga 697 data positif. Selanjutnya data yang telah memiliki label akan dilakukan perbaikan hasil *labelling* dan penyeimbangan jumlah *dataset* agar algoritma klasifikasi yang di buat mampu menghasilkan model yang dapat memprediksi hasil dengan akurat. Jumlah *dataset* setelah dilakukan penyeimbangan data dapat dilihat pada Gambar 3. Pada Gambar 3 warna hijau merupakan data sentimen negatif yang telah dilakukan penyeimbangan data. Warna merah merupakan data sentimen positif yang telah dilakukan penyeimbangan data.

TABEL 2 *Labeling*

<i>Filtering</i>	<i>Polarity score</i>	<i>Polarity</i>
['ppkm', 'level', 'kota', 'pekanbaru', 'zona', 'kuning', 'tular', 'COVID']	-9	0
['ppkm', 'longgar', '']	3	1
['kapasitas', 'ibadah', 'ppkm', 'level', 'lawan', 'pandemi']	-1	0



GAMBAR 3 Penyeimbangan Data

Ekstraksi Fitur

Setelah melakukan *labelling* pada *dataset*, maka akan dilakukan ekstraksi fitur. Penulis menggunakan ekstraksi *Ngram* dengan varian *Trigram*. *Ngram* yang dimaksud ditunjukkan pada Tabel 3. Sebagai contoh digunakan sampel *tweet* ‘ppkm level kota Pekanbaru zona kuning’. Selanjutnya untuk perhitungan manual pembobotan diambil contoh pada *term* kata “ppkm level” di mana kata tersebut muncul dua kali dari 4 dokumen yang ada sehingga perhitungan dapat dilihat pada Persamaan (1) dan Persamaan (2). Hasil dari perhitungan Persamaan (1) dan Persamaan (2) dapat dilihat bahwa *term* kata “ppkm level” memiliki nilai sebesar 1,5108.

TABEL 3 Tahap *Pre-processing*

Term	D1	D2	D3	D4	Bobot
ppkm	1	1	0	1	1,2231
ppkm level	1	0	1	0	1,5108
ppkm level kota	1	0	0	0	1,9162
level	1	0	0	1	1,5108
level kota	1	0	0	0	1,9162
level kota Pekanbaru	1	0	0	0	1,9162
kota	1	0	0	0	1,9162
kota pekanbaru	1	0	0	0	1,9162
kota pekanbaru zona	1	0	0	0	1,9162
pekanbaru	1	0	0	0	1,9162
pekanbaru zona kuning	1	0	0	0	1,9162
zona	1	0	1	0	1,5108
zona kuning	1	0	1	0	1,5108

$$W = \text{len} \left(\frac{n_{dn}}{1+n_{Term}} \right) + 1 \quad (1)$$

$$W = \text{len} \left(\frac{1+4}{1+2} \right) + 1 = 1,5108 \quad (2)$$

Pembagian Data

Pembagian data hasil pelabelan dibagi menjadi data *training* atau data latih dan data *testing* atau data uji. Data *training* atau data latih digunakan untuk melatih model sentimen *Naïve Bayes Classifier* sedangkan data uji atau data *testing* akan digunakan sebagai parameter performa model yang telah dibuat dengan menggunakan Matrix agar dapat menemukan nilai akurasi, presisi, *recall*, dan *f-1 score*. Pada penelitian ini penulis akan menggunakan tiga varian pembagian data untuk menentukan varian mana yang memiliki performa model paling baik yang kemudian akan digunakan untuk klasifikasi analisis. Pembagian *dataset* yang penulis gunakan dalam penelitian ini yaitu 7:3, 8:2, dan 9:1 untuk data *training* dan data *testing*.

Pembuatan Model *Naive Bayes Classifier*

Penelitian ini akan menggunakan *multinomial naive bayes* di mana metode ini sangat bagus digunakan pada klasifikasi yang memiliki dua kelas atau lebih, Sampel data *training* yang akan di gunakan ditunjukkan pada Tabel 4. Langkah pertama yang dilakukan yaitu melakukan Perhitungan *prior probability* dan Probabilitas kata atau *likelihood*. Sampel data *Testing* yang akan dihitung ditunjukkan pada Tabel 5. Berdasarkan Tabel 5 dapat diketahui probabilitas tiap kata yang ada pada dokumen satu untuk melihat nilai kelas negatif dan positif sebagai sampel data.

Dari data *Testing* yang telah dihitung nilai probabilitas mendapatkan nilai pada kelas positif 6.3539×10^{-7} dan nilai probabilitas pada kelas negatif 2.91507×10^{-7} sehingga dapat dikatakan bahwa dokumen satu pada data *Testing* memiliki kelas positif.

TABEL 4 Training Dataset

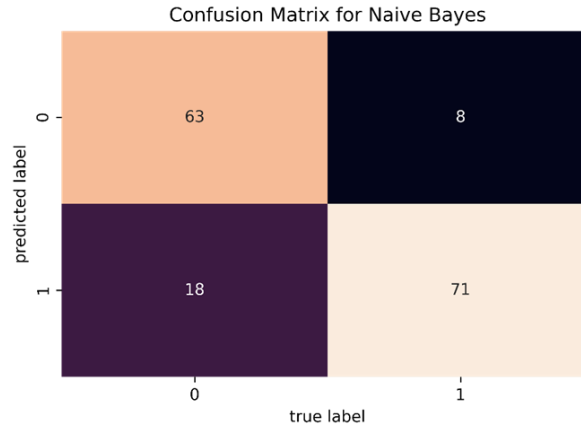
Dokumen	<i>Tweet</i>	Kelas
1	isu bohong bijak ppkm level resmi hapus via	Negatif
2	jabodetabek ppkm level giat ibadah ramadan maksimal	Positif
3	ppkm plonga plongo kaya monyet	Negatif
4	lamongan ppkm level laksana persen patuh protokol sehat	Positif

TABEL 5 Sampel Data *Testing*

Dokumen	<i>Tweet</i>	Kelas
1	patuh sehat level ppkm resmi hapus	?

Evaluasi Performa Model

Evaluasi data latih dilakukan untuk melihat performa model yang telah dibuat dengan menggunakan data *training*. Evaluasi data dilakukan dengan menggunakan Matrix. Matrix yang dimaksud ditunjukkan seperti pada Gambar 4. Pada penelitian ini dilakukan tiga variasi pembagian data untuk melihat performa dari model yang telah dibuat. Tiga variasi pembagian data yang dimaksud ditunjukkan pada Tabel 6. Pada Tabel 6 diketahui bahwa model yang telah di buat menggunakan Variasi data 9:1 mendapatkan nilai paling baik dengan *accuracy* sebesar 0,84, presisi 0,85, *recall* 0,84, dan *F1-Score* 0,84.



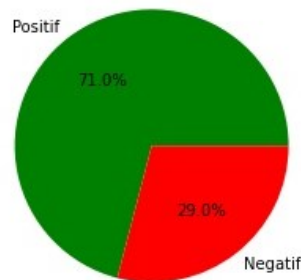
GAMBAR 4 Confusion matrix for Naive Bayes

TABEL 6 Hasil Evaluasi Performa Model

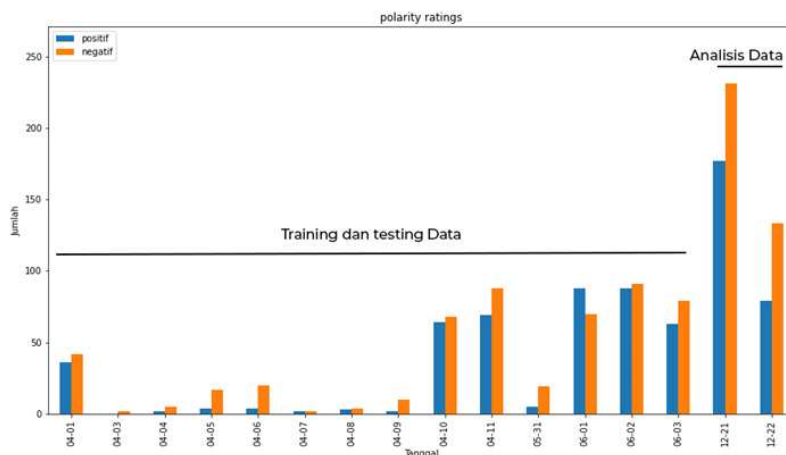
Pembagian <i>Dataset</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
9:1	0.84	0.85	0.84	0.84
8:2	0.71	0.72	0.71	0.71
7:3	0.75	0.76	0.75	0.75

Klasifikasi

Klasifikasi dilakukan untuk melihat hasil sentimen terhadap topik Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) menggunakan data Twitter. *Dataset* yang digunakan merupakan *dataset* baru Twitter yang diambil setelah pembuatan model selesai dilakukan dimana dataset baru didapatkan sebanyak 644 *tweet*. Klasifikasi dilakukan dengan menggunakan model *naive bayes classifier* yang telah di bangun dengan menggunakan variasi dengan nilai terbaik yaitu 9:1. Hasil dari Klasifikasi didapatkan 644 data *tweet* yang digunakan terdapat sebesar 29.0% data berlabel negatif dan 71.0% data berlabel positif.



GAMBAR 5 Analisis Sentimen



GAMBAR 6 Diagram Persebaran Data Sentimen

Hasil klasifikasi yang dimaksud ditunjukkan pada Gambar 5. Pada Gambar 5 diketahui hasil analisis sentimen yang dilakukan terhadap *dataset* Twitter mengenai Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) didapatkan diketahui bahwa *dataset* Twitter yang digunakan memiliki sentimen negatif yang lebih banyak dibanding sentimen positif yang menunjukkan bahwa sentimen masyarakat terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) adalah negatif. penulis menampilkan Persebaran data sentimen Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) yang telah di dapat selama penelitian dilakukan.

Diagram persebaran data yang dimaksud ditunjukkan pada Gambar 6. Pada Gambar 6 diketahui persebaran data yang diperoleh selama penelitian ini dilakukan. Sumbu y pada diagram menunjukkan jumlah *tweet*, kemudian sumbu x merupakan data keterangan waktu *tweet*. Data *training* dan *testing* diambil mulai dari tanggal 01 April 2022 – 03 Juni 2022 kemudian data analisis diambil dari tanggal 21 Desember 2022-22 Desember 2022, pada hari pertama *crawling* data menunjukkan data sentimen diangka 40 sentimen negatif dan positif, kemudian sentimen meningkat pada tanggal 10 April 2022 sampai 22 Desember 2022 di mana peningkatan yang signifikan berada pada tanggal 21 Desember di mana sentimen negatif memiliki jumlah *tweet* kurang lebih 250 data.

KESIMPULAN DAN SARAN

Kesimpulan

Hasil analisis sentimen terhadap topik “Analisis Sentimen Media Sosial Twitter Pada Kasus Pemberlakuan Pembatasan Kegiatan Masyarakat Dengan Menggunakan Metode *Naïve Bayes Classifier*” dari 644 data *tweet* didapatkan 71,0% sentimen negatif dan 29,0% sentimen positif. Nilai akurasi yang didapat dari model yang telah dibangun adalah 0,84, nilai presisi 0,85, nilai *recall* 0,84 dan nilai *F1-Score* 0,84. Dari diagram persebaran data yang telah di buat dapat menyimpulkan bahwa sentimen pada hari pertama *crawling* data menunjukkan data sentimen diangka 40 sentimen negatif dan positif, kemudian data sentimen menurun pada tanggal 03 April sampai 09 April dan meningkat pada tanggal 10 April 2022 sampai 06 Juni 2022 puncak kenaikan sentimen ada pada tanggal 21 Desember 2022 di mana data sentimen menunjukkan jumlah data kurang lebih 250 sentimen negatif..

Saran

Pada tahap *Crawling* data Twitter perlu dilakukan pengambilan data secara berkala agar sentimen serta diagram persebaran data yang dibuat mendapatkan hasil yang lebih baik lagi. Perlu lebih banyak langkah diperlukan dalam tahap *pre-processing* untuk membuat *dataset* lebih mudah diproses. Pada tahap *training* data perlu dilakukan penambahan data lebih banyak sehingga kemungkinan mendapatkan hasil evaluasi yang lebih bagus.

REFERENSI

- [1] E. M. O. N. Taofik Krisdiyanto, “Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers,” *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 7, no. 1, pp. 32–37, 2021, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/coreit/article/view/12945>
- [2] M. W. A. Putra, Susanti, Erlin, and Herwin, “Analisis Sentimen Dompot Elektronik Pada Twitter Menggunakan Metode Naïve Bayes Classifier,” *IT J. Res. Dev.*, vol. 5, no. 1, pp. 72–86, 2020, doi: 10.25299/itjrd.2020.vol5(1).5159.
- [3] D. G. Nugroho, Y. H. Chrisnanto, and A. Wahana, “Analisis Sentimen Pada Jasa Ojek Online ... (Nugroho dkk.),” pp. 156–161, 2015.
- [4] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, “Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm,” *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.
- [5] M. Christianto, J. Andjarwirawan, and A. Tjondrowiguno, “Aplikasi analisa sentimen pada komentar berbahasa Indonesia dalam objek video di website YouTube menggunakan metode Naïve Bayes classifier,” *J. Infra*, vol. 8.1, pp. 255–259, 2020.
- [6] D. Rusdaman and D. Rosiyadi, “Analisa Sentimen Terhadap Tokoh Publik Menggunakan Metode Naïve Bayes Classifier Dan Support Vector Machine,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 4, no. 2, pp. 230–235, 2019.