

Identifikasi Kebutuhan Masyarakat Nusa Tenggara Barat pada Pandemi Covid-19 di Media Sosial dengan Metode *Crawling*

(Requirements Identification for NTB People in pandemic covid-19 at Social Media Using Crawling Method)

Jian Budiarto^{[1]*}

^[1]Rekayasa Perangkat Lunak, Universitas Bumigora Mataram

E-mail: jian.budiarto@gmail.com

KEYWORDS:

Social Media Sentiment, covid-19 pandemic, NTB requirements

ABSTRACT

People can write to the government about what they want through social media in the freedom of information era. Especially in the Covid-19 pandemic, the public's desire for government attention is getting higher. The purpose of this activity is to collect community needs at the beginning of the pandemic. The period of information retrieval is from March 30 to April 5, 2020. The activity uses the crawling method, which is gathering information with browser tools. The implementation stage starts from taking information with selenium, identifying posts and comments with Html tags, calculating the emergence of issues with N-gram NLP, and analyzing sentiment with Naive Bayes and Support-Vector Machine. The results showed that the community did not have panic to need foods with a total reaction of 172. The community was more worried about the spread of outbreaks from outside NTB with a reaction of 1,421.

KATA KUNCI:

Sentimen Media Sosial, Pandemi covid19, Kebutuhan NTB

ABSTRAK

Pada era kebebasan informasi ini, masyarakat dapat menulis kebutuhan yang diinginkan kepada pemerintah melalui media sosial. Terutama pada kondisi pandemi covid-19, keinginan masyarakat untuk mendapat perhatian pemerintah semakin tinggi. Tujuan penelitian ini adalah untuk menghimpun kebutuhan masyarakat pada awal masa pandemi. Periode pengambilan informasi dilakukan pada tanggal 30 maret sampai dengan 5 april 2020. Kegiatan menggunakan metode crawling yaitu penghimpunan informasi dengan alat bantu browser. Tahapan pelaksanaan dimulai dari pengambilan informasi dengan selenium, identifikasi posting dan komentar dengan Tag HTML, menghitung kemunculan isu dengan N-gram NLP, dan analisis sentimen dengan naive bayes dan Support-Vector Machine. Hasil penelitian diketahui bahwa masyarakat tidak memiliki kepanikan terhadap kebutuhan bahan pokok dengan jumlah reaksi 172. Masyarakat lebih khawatir terhadap faktor peredaran wabah dari luar NTB dengan reaksi 1.421.

I. PENDAHULUAN

Dalam membangun *Public Issue and Trust Management*, Biro Humas dan Protokol Provinsi Nusa Tenggara Barat bertugas menghimpun isu dan aspirasi publik tentang Gubernur dan Wakil Gubernur NTB yang tersebar melalui media sosial (lokal dan nasional). Namun, untuk memperoleh informasi tersebut secara cepat, tepat dan lengkap tidak dapat dilakukan oleh tim Humas & Protokol secara manual. Tim tersebut membutuhkan aplikasi *artificial intelligence* yang dapat menyajikan data tersebut secara otomatis dan *real time* setiap hari.

Tujuan kegiatan ini untuk melakukan identifikasi kebutuhan masyarakat NTB di masa pandemi covid-19. Hal ini dalam mendukung program NTB Gemilang di lingkungan Biro Humas dan Protokol Provinsi Nusa Tenggara Barat. Kegiatan ini menggabungkan metode *crawler* dilengkapi dengan klasifikasi *naive bayes* dan *support vector machine* pada sentimen analisis.

Data mining atau penemuan pengetahuan adalah prosedur menggunakan teknik statistik dan berbasis pengetahuan untuk menganalisis data dengan pola tambang yang memiliki makna dari kumpulan data

yang luas dan mengubahnya menjadi informasi bermanfaat [1]. Pengumpulan informasi tersebut menggunakan metode *crawling*. *Crawling* adalah cara untuk mendapatkan konten informasi yang berada pada website[2]. Teknik pengumpulan informasi dapat menghimpun data secara keseluruhan maupun secara spesifik fokus pada data tertentu [3]. Pada awalnya metode *crawler* digunakan oleh peramban web untuk melakukan pengisian indeks mereka. Kendala yang dihadapi dalam melakukan teknik *crawling* struktur data yang besar dan kurangnya kontrol terpusat terhadap konten. Marc merekomendasikan cara membangun dan mengoperasikan *crawler* dengan performa tinggi [4].

Salah satu pendekatan metode statistika yang dapat melakukan pengkategorian adalah klasifikasi. Klasifikasi merupakan salah satu teknik *data mining* yang digunakan untuk membangun suatu model dari sampel data yang belum terklasifikasi untuk digunakan mengklasifikasi sampel data baru ke dalam kelas-kelas yang sejenis. Klasifikasi termasuk ke dalam *supervised learning* karena menggunakan sekumpulan data untuk dianalisis terlebih dahulu, kemudian pola dari hasil analisis tersebut digunakan untuk pengklasifikasian data uji [5]. Proses klasifikasi data terdiri dari pembelajaran dan klasifikasi. Pada pembelajaran data *training* dianalisis menggunakan algoritma klasifikasi, selanjutnya pada klasifikasi digunakan data testing untuk memastikan tingkat akurasi. Terdapat beberapa algoritma pada metode yaitu *Naive bayes* dan *Support Vector Machine* (SVM).

Naive Bayes adalah salah satu algoritma teknik data mining yang menerapkan teori Bayes dalam klasifikasi dimana algoritma ini mengasumsikan bahwa atribut suatu objek bersifat independen atau bebas. Metode Naive Bayes telah banyak digunakan dalam penelitian mengenai *text mining*, beberapa kelebihan Naive Bayes diantaranya adalah algoritma sederhana tapi memiliki akurasi yang tinggi [6]. Teknik *Support-Vector Machine* (SVM) bersumber dari teori pembelajaran statistik dan telah menunjukkan hasil empiris yang menjanjikan dalam berbagai aplikasi praktis dari pengenalan digit tulisan tangan sampai kategorisasi teks. SVM juga bekerja sangat baik pada data dengan banyak dimensi dan menghindari kesulitan dari permasalahan dimensionalitas [7]. Dengan mengkombinasikan 2 metode tersebut diharapkan

mendapatkan hasil yang sesuai dengan kebutuhan pemerintah daerah Nusa Tenggara Barat.

Naive bayes dan SVM telah digunakan pada berbagai kebutuhan antara lain sebagai *spam filtering* [8], *spasial prediction* [9], *liver prediction* [10], identifikasi seismic dan nuklir[11], Sedangkan peneliti lain lebih banyak menggunakan naive bayes dan SVM pada sentimen analisis seperti analisis *sentimen hate speech* [12] , membandingkan SVM dan naive bayes wiktologi [13] , deteksi sentimen di Twitter, perbandingan SVM dan naive bayes[14] , klasifikasi kepribadian [15] dan sentimen analisis Facebook status menggunakan naive bayes [16].

Data mining atau penemuan pengetahuan adalah prosedur menggunakan teknik statistik dan berbasis pengetahuan untuk menganalisis data dengan pola tambang yang memiliki makna dari kumpulan data yang luas dan mengubahnya menjadi informasi bermanfaat [17]. Menurut Mirkin data mining didefinisikan sebagai suatu proses untuk mencari pola dari sekumpulan data yang terdapat di dalam database untuk kemudian dianalisis sehingga menghasilkan suatu informasi tertentu untuk dapat dimanfaatkan pada proses selanjutnya [18].

II. METODOLOGI

A. Metode Crawling

Pengambilan informasi dengan menggunakan teknologi *crawler* dapat menghimpun data secara otomatis. *Crawler* yang dibutuhkan menyesuaikan dengan karakteristik media sosial masing-masing.

Adapun langkah umum dalam pengambilan informasi *crawling* sebagai berikut.



Gbr. 1 Alur kerja metode crawling

1) Mendefinisikan target alamat situs yang diproses

Pada tahap ini data alamat situs ditentukan terlebih dahulu. Dalam kasus ini alamat situs yang digunakan adalah media sosial lokal NTB di Facebook, Instagram, dan Twitter.

2) Pencarian informasi (*crawling*)

Sebuah komputer yang telah terkoneksi dengan internet ditanamkan sebuah program crawler. Program ini menggunakan sumber kode *Selenium* [19] untuk melakukan otomatisasi. Pada waktu yang telah dijadwalkan, program tersebut akan berjalan dan membuka situs yang telah ditentukan sebelumnya. Tugas dari program tersebut adalah mengidentifikasi postingan dan komentar menggunakan struktur *Tag Html*.

Konten dan artikel yang telah didapatkan diolah kembali dengan menggunakan *N-gram* pada *Natural Language Processing* untuk menghitung jumlah kemunculan isu di hari tersebut. Kemunculan isu dikategorikan dalam beberapa kata yang telah ditentukan seperti masker, pelabuhan, bandara dan sebagainya. Cara menghitung kategori masih dilakukan dengan cara sederhana. Jika konten atau komentar berisi kata yang ditentukan, maka konten masuk ke kategori tersebut.

3) Menentukan sentimen dengan menggunakan SVM dan Naive Naves.

Tingkat prioritas tindakan dan pelaksanaan kebijakan ditentukan oleh sentimen yang diberikan masyarakat. Keluhan oleh masyarakat akan memberikan sentimen negatif terhadap pemerintah. Oleh karena itu, penting untuk mengetahui sentimen masyarakat di postingan dan komentar media sosial milik pemerintah dan *influencer* di NTB. Sentimen masyarakat dapat diketahui dengan menggunakan SVM dan Naive Bayes. Proses pembelajaran sentimen dalam kosakata bahasa indonesia menggunakan metode SVM [20]. Dataset tersebut dapat diunduh pada sumber berikut [21]. Dengan demikian, proses pengenalan sentimen tidak perlu dilakukan kembali.

Dalam menentukan sentimen metode SVM dilengkapi dengan metode Naive Bayes. Postingan dan komentar yang tidak dapat dideteksi sentimennya oleh SVM dapat diperbaiki oleh metode Naive Bayes.

B. Lokasi Pelaksanaan

Data yang digunakan merupakan data primer. Data diperoleh secara langsung dari hasil pengambilan informasi komentar dari sosial media seperti Facebook, Instagram dan Twitter berbagai akun sosial milik Dinas Provinsi, Official Gubernur dan Wakil Gubernur, Influencer dari pers media.

Pengambilan informasi dibedakan menjadi beberapa kategori antara lain: komentar netizen pada status sosial media Dinas Provinsi, komentar netizen terhadap status Gubernur dan Wakil Gubernur, dan komentar netizen terhadap status pers media (*influencer*).

C. Waktu Pengambilan Informasi

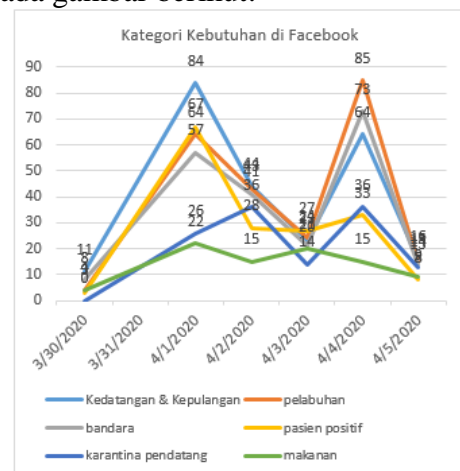
Informasi diambil pada periode 30 Maret 2020 sampai dengan 5 April 2020. Periode ini adalah periode awal covid-19 masuk ke wilayah NTB. Waktu pengambilan dimulai pukul 00.00 sampai 21.00 WITA setiap harinya. Kepanikan terjadi di media sosial, sehingga kecepatan dalam memperoleh informasi dibutuhkan pemerintah daerah dalam mengambil keputusan.

III. HASIL DAN PEMBAHASAN

Kebutuhan masyarakat NTB yang diperoleh dari hasil *crawler* media sosial dapat dijelaskan berdasarkan *platform* masing-masing sebagai berikut.

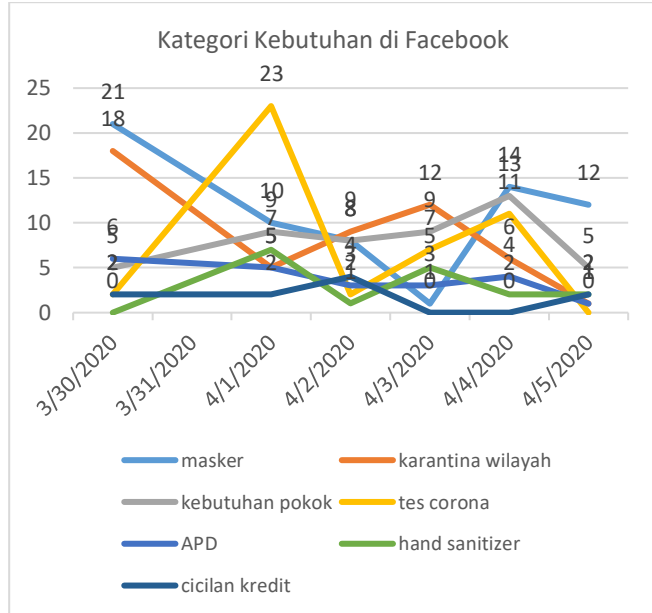
A. Kebutuhan di Media Sosial Facebook.

Pada media sosial Facebook data diambil pada periode 30 maret 2020 sampai dengan 5 April 2020 setiap pukul 21.00 WITA. Data telah dikategorikan seperti pada gambar berikut.



Gbr. 2 Kategori kebutuhan tertinggi di Facebook

Pada media sosial Facebook, kategori “kedatangan dan kepulangan warga” menjadi perhatian tertinggi selama periode awal pandemi dengan total keseluruhan sebesar 242 reaksi masyarakat. Hal ini diikuti oleh kategori “pelabuhan” dengan jumlah reaksi 235 kali, sedangkan “bandara” dengan 214 reaksi.

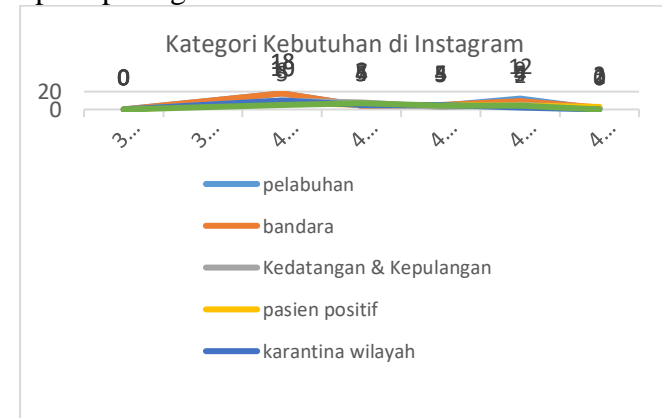


Gbr. 3 Kategori kebutuhan terendah di Facebook

Pada Gbr 3 diketahui kebutuhan masyarakat dalam jumlah reaksi rendah adalah kebutuhan “masker” sebesar 66 reaksi. Diikuti oleh kebutuhan akan “karantina wilayah” sebesar 51 reaksi. Sedangkan kategori “kebutuhan pokok” mendapatkan jumlah reaksi 49.

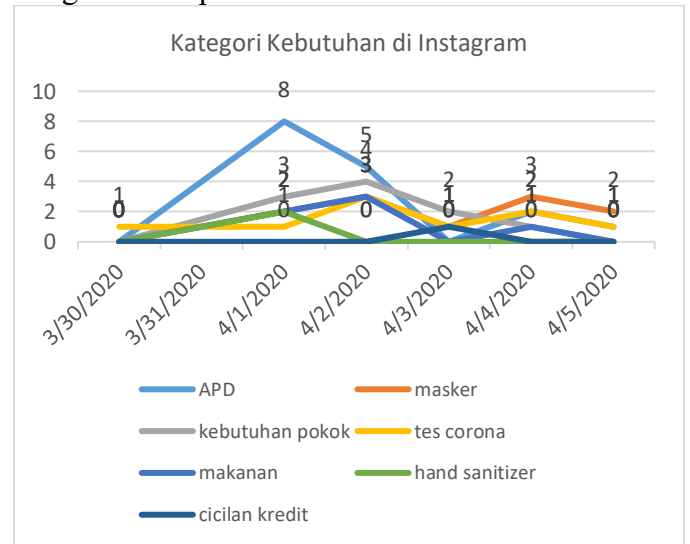
B. Kebutuhan di Media Sosial Instagram

Pada media sosial Instagram, data diambil pada periode 30 maret 2020 sampai dengan 5 April 2020 setiap pukul 21.00 WITA. Data telah dikategorikan seperti pada gambar berikut.



Gbr. 4 Kategori kebutuhan tertinggi di Instagram

Berdasarkan Gbr 4, kebutuhan masyarakat tertinggi ada pada kategori “pelabuhan” dengan jumlah reaksi 39 kali. Hal ini diikuti dengan kategori “bandara” dengan jumlah reaksi sebesar 38. Sedangkan kategori “kedatangan dan kepulangan warga” mendapatkan reaksi 28 kali.

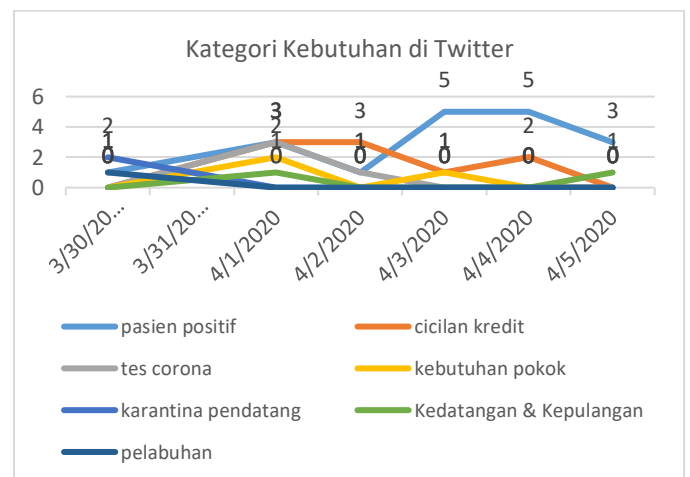


Gbr. 5 Kategori kebutuhan terendah di Instagram

Dalam kategori reaksi rendah kategori “APD” mendapatkan reaksi sebesar 16 kali. Hal ini diikuti oleh kategori “masker” dengan jumlah reaksi 11 kali, sedangkan kategori “kebutuhan pokok” mendapatkan reaksi sebesar 10.

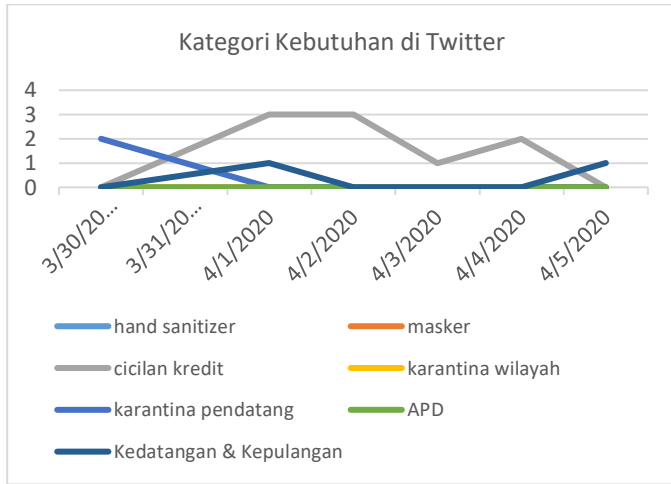
C. Kebutuhan di Media Sosial Twitter

Pada media sosial Twitter, percakapan tentang pandemi di NTB masih rendah. Hal ini karena media sosial ini belum banyak digunakan oleh masyarakat NTB. Data percakapan di Twitter telah dikategorikan sebagai berikut.



Gbr. 6 Kategori kebutuhan tertinggi di Twitter

Di media sosial Twitter, kategori percakapan masyarakat “pasien positif” memiliki reaksi tertinggi dengan jumlah sebesar 18. Hal ini diikuti oleh kategori “cicilan kredit” dengan jumlah reaksi sebesar 9 kali. Sedangkan kategori “tes corona” mendapatkan reaksi 4 kali.



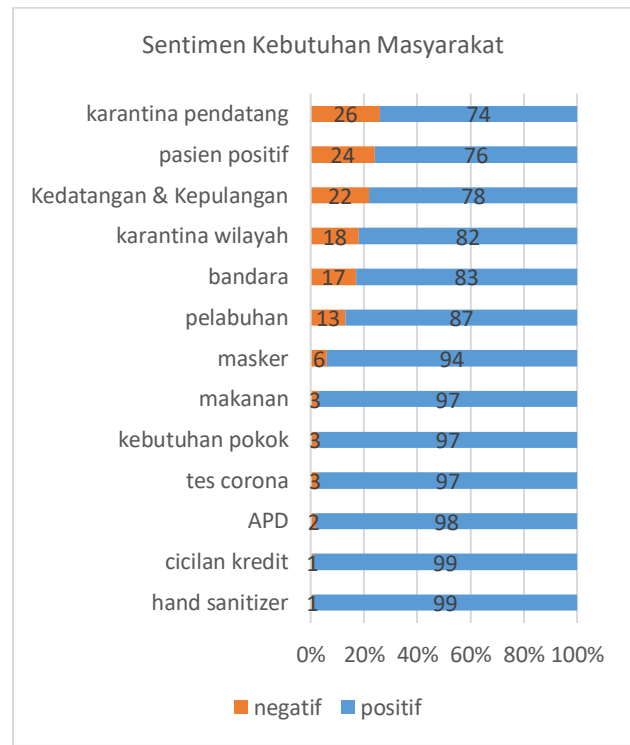
Gbr. 7 Kategori kebutuhan terendah di Twitter

Kebutuhan masyarakat lainnya seperti “hand sanitizer”, “masker”, “cicilan kredit”, “karantina wilayah” masih dibawah 5 kali reaksi setiap harinya.

Berdasarkan data yang telah disajikan pada setiap media sosial, urutan popularitas masyarakat Nusa Tenggara Barat dalam menyampaikan aspirasinya adalah media Facebook, Instagram, dan Twitter. Hal ini dapat terjadi karena masyarakat menganggap bahwa sosok Gubernur Nusa Tenggara Barat sangat aktif di media sosial Facebook. Masyarakat memberikan komentar pada kolom komentar Facebook dan berharap bahwa informasi tersebut dapat langsung diterima oleh Gubernur. Oleh karena itu, masalah yang terjadi segera dapat terselesaikan dengan baik.

D. Sentimen Kebutuhan Berdasarkan Kategori

Prioritas pelaksanaan penanggulangan pandemi merekomendasikan menggunakan sentimen sebagai tolak ukur. Dengan adanya sumsi bahwa kebutuhan masyarakat yang mendesak atau tidak optimalnya pelayanan semasa pandemi menyebabkan munculnya sentimen negatif. Berikut kami tampilkan sentimen masyarakat berdasarkan kategori.



Gbr. 8 Sentimen berdasarkan kategori

Pada gambar 7 diketahui bahwa kategori “karantina pendatang” menjadi sentimen negatif tertinggi dengan nilai 26%. Selanjutnya diikuti oleh kategori “pasien positif” dengan jumlah sebesar 24%. Sedangkan kekhawatiran masyarakat akan “kedatangan dan kepulangan warga” mendapatkan sentimen negatif sebesar 22%.

Mengetahui sentimen terhadap masing-masing kategori bukan satu-satunya indikator dalam menentukan prioritas pelaksanaan penanggulangan pandemi. Di sisi lain, ada indikator lain seperti jumlah reaksi yang telah dibahas sebelumnya.

IV. KESIMPULAN

Berdasarkan jumlah pengambilan informasi kebutuhan masyarakat pada periode 30 Maret 2020 – 5 April 2020 di seluruh sosial media dapat dihimpun sebagai berikut.

TABEL I
REKAPITULASI JUMLAH KEBUTUHAN MASYARAKAT PADA AWAL PANDEMI

No	Kebutuhan	Jumlah	Sentimen
1	pelabuhan	275	87%
2	Kedatangan & Kepulangan	272	78%
3	bandara	252	83%
4	pasien positif	209	76%
5	karantina pendatang	147	74%

No	Kebutuhan	Jumlah	Sentimen
6	makanan	91	97%
7	masker	77	94%
8	karantina wilayah	73	82%
9	kebutuhan pokok	62	97%
10	tes corona	58	97%
11	APD	38	98%
12	cicilan kredit	20	99%
13	hand sanitizer	19	99%

Berdasarkan data pada tabel diatas, diketahui bahwa kebutuhan masyarakat di media sosial lebih cenderung pada proteksi dan penyebaran virus covid19 dari luar NTB. Hal ini dengan jumlah reaksi masyarakat NTB yang menyebutkan bersifat proteksi dalam kategori Pelabuhan (275), Kepulauan & Kedatangan (272), Bandara (252), Karantina Pendatang (147) dan sebagainya. Total reaksi pada kecenderungan ini adalah 1.421 reaksi.

Sedangkan masyarakat tidak terlalu khawatir tentang kebutuhan pokok masyarakat Makanan (91), dan Masker (77), Kebutuhan Pokok (62) dan *Hand Sanitizer* (19) dengan total keseluruhan pada kecenderungan ini sebesar 172 reaksi. Masyarakat tidak terlalu khawatir akan adanya kelangkaan bahan pokok. Hal ini tentu sangat positif, sehingga di masyarakat tidak terjadi gejala *Panic Buying*.

Langkah terakhir dalam menimbang prioritas penanggulangan kebijakan harus dikombinasikan dengan sentimen. Berikut kami lampirkan 5 rekomendasi prioritas penanggulangan bencana pandemi berdasarkan kategori.

TABEL II
REKOMENDASI PRIORITAS PENANGGULANGAN
BENCANA DI AWAL PANDEMI

No	Kebutuhan	Jumlah	Sentimen
1	pelabuhan	275	87%
2	Kedatangan & Kepulauan	272	78%
3	bandara	252	83%
4	pasien positif	209	76%
5	karantina pendatang	147	74%
6	karantina wilayah	73	82%

Pada tabel diatas merupakan prioritas penanggulangan bencana berdasarkan urgenitas di media sosial. Alasan yang digunakan untuk menentukan prioritas tersebut adalah jumlah reaksi masyarakat yang besar. Hal ini menandakan isu tersebut berdampak luas kepada masyarakat. Selain itu, 6 rekomendasi yang telah dipilih memiliki

sentimen dibawah 90%. Hal ini dikhawatirkan dapat mengganggu kondisi masyarakat.

UCAPAN TERIMA KASIH

Terima kasih kami ucapkan kepada Biro Humas dan Protokol Setda Provinsi NTB yang telah memberikan kesempatan untuk melakukan penelitian dan pengabdian.

REFERENSI

- [1] S. Tufféry, *Data Mining and Statistics for Decision Making*. John Wiley & Sons, 2011.
- [2] S. Raghavan dan H. Garcia-Molina, "Crawling the Hidden Web," 2000. <http://ilpubs.stanford.edu:8090/456/> (diakses February 13, 2021).
- [3] G. Pant, P. Srinivasan, dan F. Menczer, "Crawling the Web," dalam *Web Dynamics: Adapting to Change in Content, Size, Topology and Use*, M. Levene dan A. Poulouvasilis, Ed. Berlin, Heidelberg: Springer, 2004, hlm. 153–177.
- [4] M. Najork dan A. Heydon, "High-Performance Web Crawling," dalam *Handbook of Massive Data Sets*, J. Abello, P. M. Pardalos, dan M. G. C. Resende, Ed. Boston, MA: Springer US, 2002, hlm. 25–45.
- [5] D. Sartika, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian | JATISI (Jurnal Teknik Informatika dan Sistem Informasi)," Diakses: Feb 13, 2021. [Daring]. Tersedia pada: <http://jurnal.mdp.ac.id/index.php/jatisi/article/view/78>.
- [6] D. Olson dan D. Delen, "Advanced Data Mining Techniques," 2008, doi: 10.5860/choice.45-6838.
- [7] I. Rish, "An Empirical Study of the Naïve Bayes Classifier," *IJCAI 2001 Work Empir Methods Artif Intell*, vol. 3, Jan 2001.
- [8] W. Feng, J. Sun, L. Zhang, C. Cao, dan Q. Yang, "A support vector machine based naive Bayes algorithm for spam filtering," dalam *2016 IEEE 35th International Performance Computing and Communications Conference (IPCCC)*, Des 2016, hlm. 1–8, doi: 10.1109/PCCC.2016.7820655.
- [9] B. T. Pham, D. Bui, I. Prakash, dan M. B. Dholakia, "Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS," vol. 10, no. 1, hlm. 10, 2016.
- [10] D. S. Vijayarani dan M. S. Dhayanand, "Liver Disease Prediction using SVM and Naïve Bayes Algorithms."
- [11] L. Dong, X. Li, dan G. Xie, "Nonlinear Methodologies for Identifying Seismic Event and Nuclear Explosion Using Random Forest, Support Vector Machine, and Naive Bayes Classification," *Abstract and Applied Analysis*, 2014.

- <https://www.hindawi.com/journals/aaa/2014/459137/> (diakses January 28, 2020).
- [12] G. A. Buntoro, "ANALISIS SENTIMEN HATESPEECH PADA TWITTER DENGAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE," *J. Din. Inform.*, vol. 5, no. 2, Sep 2016, Diakses: Jan 28, 2020. [Daring]. Tersedia pada: <http://ojs.upy.ac.id/ojs/index.php/dinf/article/view/975>.
- [13] S. Hassan, M. Rafi, dan M. S. Shaikh, "Comparing SVM and naïve Bayes classifiers for text categorization with Wikitology as knowledge enrichment," dalam *2011 IEEE 14th International Multitopic Conference*, Des 2011, hlm. 31–34, doi: 10.1109/INMIC.2011.6151495.
- [14] S. Rana dan A. Singh, "Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques," dalam *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, Okt 2016, hlm. 106–111, doi: 10.1109/NGCT.2016.7877399.
- [15] B. Y. Pratama dan R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM," dalam *2015 International Conference on Data and Software Engineering (ICoDSE)*, Nov 2015, hlm. 170–174, doi: 10.1109/ICoDSE.2015.7436992.
- [16] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, dan J. Caro, "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning," dalam *IISA 2013*, Jul 2013, hlm. 1–6, doi: 10.1109/IISA.2013.6623713.
- [17] D. S. Jasim, "Data Mining Approach and Its Application to Dresses ...," *moam.info*. <https://moam.info/data-mining-approach-and-its-application-to-dresses-599343111723ddcb690daa67.html> (diakses February 13, 2021).
- [18] B. Mirkin, "Data analysis, mathematical statistics, machine learning, data mining: Similarities and differences," dalam *2011 International Conference on Advanced Computer Science and Information Systems*, Des 2011, hlm. 1–8.
- [19] "SeleniumHQ Browser Automation." <https://www.selenium.dev/> (diakses Feb 14, 2021).
- [20] Jan Kristanto Wibisono dan M. S. Drs. Edi Winarko, "OPINION MINING PADA TWITTER UNTUK BAHASA INDONESIA DENGAN METODE SUPPORT VECTOR MACHINE DAN METODE BERBASIS LEXICON," Thesis, [Yogyakarta]: Universitas Gadjah Mada, 2013.
- [21] J. Kristanto, *jankristanto/mythesis*. 2017.