

Sentiment Analysis Berbasis Algoritma Naïve Bayes Classifier untuk Identifikasi Persepsi Masyarakat Terhadap Produk / Layanan Perusahaan

Affandy¹, Oktania Nandiyati²

Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang
Jl.Imam Bonjol No.207 Semarang, telp. (024) 3569196
e-mail: ¹affandy@dsn.dinus.ac.id, ²oktaniandyt@gmail.com

Diterima: 11 Mei 2019; Direvisi: 30 Mei 2020; Disetujui: 31 Mei 2020

Abstrak

Twitter merupakan layanan microblogging terpopuler di Indonesia, dengan jumlah pengguna mencapai hampir 23 juta. Di era big data seperti saat ini tweets atau “kicauan” dari pelanggan, pemerhati, konsumen potensial, atau masyarakat dari pengguna produk atau layanan dari suatu perusahaan akan sangat membantu perusahaan dalam mengetahui lanskap industri dan konsumen, sehingga dapat menentukan langkah strategis yang akan menyumbangkan kontribusi terhadap pertumbuhan perusahaan. Namun demikian pemanfaatan data dari media sosial Twitter terkendala dengan sejumlah kesulitan teknis dalam proses pengumpulannya, pengolahannya, serta analisisnya. Secara spesifik penelitian ini menerapkan algoritma Naïve Bayes Classifier dalam proses analisa sentimen pada data tweets ke dalam sebuah prototipe aplikasi yang ditujukan untuk memudahkan perusahaan/organisasi mengetahui persepsi masyarakat terhadap produk atau layanan yang dimiliki. Algoritma NBC dipilih karena algoritma ini mampu melakukan klasifikasi dengan baik walaupun menggunakan data training yang kecil, namun memiliki akurasi dan kecepatan proses yang tinggi untuk penanganan data training yang besar. Dari hasil evaluasi didapati prototipe berjalan dengan baik dimana keyword yang dimasukkan akan mentrigger Twitter API untuk melakukan crawling selanjutnya proses mining dapat terpantau di tiap tahapan dan di akhir proses, sistem akan menunjukkan nilai akhir level sentiment serta representasi log hasil perhitungan dalam bentuk chart dengan kurun waktu tertentu.

Kata kunci: twitter, sentiment analysis, klasifikasi, naïve bayes classifier, prototype

Abstract

Twitter is the most popular microblogging service in Indonesia, with nearly 23 million users. In the era of big data such as the current tweets from customers, observers, potential consumers, or the community of users of products or services of a company will greatly help companies in knowing the industrial and consumer landscape, so as to determine strategic plans that will contribute to the company's growth. However, the use of data from social media such as Twitter is hampered by a number of technical difficulties in the process of collecting, processing, and analysing. Specifically, this research applies the Naïve Bayes Classifier algorithm in the process of sentiment analysis of tweets data into a prototype application that is intended to make it easier for companies / organizations to know people's perceptions of their products or services. The NBC algorithm was chosen because this algorithm is able to do a good classification even though it uses small training data, but has high accuracy and process speed for handling large training data. From the evaluation results found a prototype running well where the keywords entered will

trigger the Twitter API to crawl the data then the mining process can be monitored at each stage and at the end of the process, the system will show the final sentiment level values and the representation of the calculation results log in a chart form over a certain period of time.

Keywords: twitter, sentiment analysis, classification, naïve bayes classifier, prototype

1. PENDAHULUAN

Riset dari Hootsuite, lembaga manajemen media sosial, dan We Are Social, perusahaan agensi global, di awal tahun 2020 menyebutkan bahwa hingga saat ini potensi pengguna media sosial di Indonesia mencapai 160 juta pengguna (kisaran lebih dari 59% dari total penduduk Indonesia) [1]. Angka ini meningkat cukup signifikan, mencapai 17%, dari jumlah pengguna di tahun sebelumnya. Sementara itu hasil analisa dan statistik *web traffic* di Indonesia diketahui 4 dari 10 website yang paling sering diakses hingga Februari 2020 adalah situs media sosial yang diantaranya adalah: Youtube, Facebook, Instagram, dan Twitter [2]. Hal ini menunjukkan bahwa kepemilikan akun dalam media sosial menjadi suatu kebutuhan bagi siapapun yang beraktifitas dengan memanfaatkan teknologi internet. Media sosial sering kali dimaknai sebagai platform teknologi interaktif dengan media komputer dan internet yang memfasilitasi penciptaan dan pendistribusian informasi, ide, hobi atau minat, dan berbagai bentuk ekspresi lainnya melalui jaringan dan komunitas virtual [3].

Salah satu model media sosial yang berkembang saat ini adalah platform *microblogging*, media *broadcast online* yang memungkinkan penggunanya untuk bertukar elemen dari konten yang berukuran kecil atau pendek yang dapat berbentuk kalimat ringkas, gambar tunggal, ataupun link dari suatu video. Pesan dengan ukuran kecil ini sering juga disebut dengan *micropost*. Karena ukurannya yang relatif kecil inilah yang menjadikan situs penyedia *microblogging* ini menjadi populer (Twitter, Mastodon, Amino, dan sebagainya). Twitter merupakan layanan *microblogging* terpopuler di Indonesia, dengan jumlah mencapai hampir 23 juta pengguna, dengan peningkatan mencapai dua kali lipatnya dari lima tahun sebelumnya [4]. Menurut data yang telah dirilis oleh Twitter pada akhir tahun 2016 lalu, dikatakan bahwa jumlah *tweets* yang dihasilkan oleh pengguna Twitter Indonesia sepanjang tahun 2016 mencapai 4.1 miliar *tweets* [5].

Jumlah *tweets* yang terekam dalam bentuk teks digital dengan volume yang sangat besar ini tentunya menyimpan pengetahuan yang sangat berharga, utamanya bagi dunia bisnis. Di era *big data* seperti saat ini *tweets* atau “kicauan” dari pelanggan, pemerhati, konsumen potensial, atau masyarakat dari pengguna produk atau layanan dari suatu perusahaan akan sangat membantu perusahaan dalam mengetahui lanskap industri dan konsumen, sehingga dapat menentukan langkah strategis yang akan menyumbangkan kontribusi terhadap pertumbuhan perusahaan. Penilaian dari publik terkait dengan kualitas produk atau layanan, fitur yang ditawarkan, tingkat kepuasan, dan berbagai hal lainnya akan sangat membantu perusahaan dalam menentukan karakteristik konsumen, ekspektasi konsumen, dan persepsi konsumen terhadap citra perusahaan. Namun demikian pemanfaatan data dari media sosial khususnya Twitter terkendala dengan sejumlah kesulitan teknis dalam proses pengumpulannya, pengolahannya, serta analisisnya [6].

Di dalam bidang ilmu komputasi, analisa sentimen dari opini masyarakat yang dituangkan dalam bentuk teks termasuk dalam bidang kajian *text mining* dengan sub-kajian *opinion mining*. Fungsi utama dari analisis sentimen adalah menghasilkan kelompok polaritas dari teks yang ada pada sekumpulan dokumen yang selanjutnya diidentifikasi apakah dokumen mengandung pendapat yang bersifat positif, negatif, ataupun netral [6]. Sejumlah algoritma telah banyak diterapkan dalam kasus analisa sentimen yang diantaranya Naïve Bayes Classifier (NBC), Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), Decision Tree, dan sebagainya. [7, 8, 9, 10]. K-NN merupakan algoritma yang paling umum digunakan namun demikian K-NN memiliki kelemahan apabila menangani *data training* dalam jumlah yang kecil. Sebaliknya dengan SVM yang memiliki performa baik pada *data training* yang kecil namun demikian konsistensi performa SVM menurun seiring dengan peningkatan jumlah data [11, 12]. Solusi

- mengumpulkan *tweets*, melakukan pre-processing dari *tweets* yang terkumpul, melakukan proses klasifikasi opini dari *tweets*, dan merepresentasikan statistik dari hasil klasifikasi.
2. Pemahaman data, proses pengumpulan data dilakukan dengan teknik *crawling* dimana semua *tweets* yang relevan dengan keyword yang ditentukan akan ditarik untuk dijadikan sebagai dataset. Proses *crawling* menggunakan library *TwitterOAuth()* dan fungsi *get('search/tweets')* yang disediakan dari Twitter API. Dari dataset yang didapatkan selanjutnya akan dibagi menjadi *data training* dan *data testing*.
 3. Persiapan data, *tweets* yang terkumpul umumnya dalam bentuk tidak terstruktur atau semi terstruktur. Proses konversi terhadap data perlu dilakukan sebelum proses klasifikasi dijalankan, adapun proses tersebut meliputi: (a) merubah semua teks menjadi lower case dan menghapus semua karakter yang bukan huruf (case folding), (b) merubah struktur kalimat menjadi kata atau *term* (*tokenizing*), (c) merubah setiap kata menjadi kata dasar (*stemming*), proses *stemming* dilakukan dengan menggunakan algoritma *Enhanced Confix-Stripping* (ECS), dan (d) menghapus semua kata hubung, kata depan, kata ganti, dan *keyword* pencarian data saat proses pengumpulan data dilakukan (*stopword removal*). Daftar kata *stopword* yang digunakan pada penelitian ini bersumber dari Tala [14].
 4. Penentuan model mining, fungsi mining yang akan dijalankan dalam penelitian adalah fungsi klasifikasi, dimana data yang terkumpul akan dikelompokkan ke dalam 3 label sentimen yang meliputi positif, negatif, dan normal. Dari proses kajian literatur penelitian disimpulkan bahwa proses klasifikasi akan menggunakan algoritma Naïve Bayes Classifier (NBC). Secara umum teori Bayes merupakan teori perbandingan peluang dari hipotesa dengan rumus umum sebagai berikut:

$$P(C|O) = \frac{P(O|C) \times P(C)}{P(O)} \quad (1)$$

Dimana:

$P(C O)$	= Probabilitas hipotesis berdasarkan kondisi
$P(O C)$	= Probabilitas berdasarkan kondisi pada hipotesis
C	= Hipotesis data merupakan kelas spesifik
O	= Data dengan kelas yang belum diketahui
$P(O)$	= Probabilitas O

Pada proses klasifikasi sentimen formula NBC adalah sebagai berikut

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} P(V_j) \prod P(x_1|V_j) \quad (2)$$

Dimana V_{MAP} merupakan probabilitas tertinggi dari semua kelas yang diujikan. Sedangkan $P(V_j)$ dan $P(x_1|V_j)$ dilakukan pada saat training menggunakan persamaan 3 dan 4 sebagai berikut:

$$P(V_j) = \frac{|docsj|}{|docstraining|} \quad (3)$$

$$P(x_1|V_j) = \frac{ni+1}{n+|kosakata|} \quad (4)$$

Dimana

$ docsj $	= jumlah dokumen yang memiliki kategori V_j
$ docstraining $	= jumlah dokumen dari semua kategori yang digunakan dalam proses training
N	= frekuensi kemunculan kata x_1 dalam dokumen yang memiliki kelas atau kategori V_j
$ kosakata $	= jumlah kata dari semua kategori yang digunakan dalam proses training
ni	= jumlah frekuensi kemunculan setiap kata

5. Proses evaluasi model, hasil klasifikasi selanjutnya akan diuji menggunakan pendekatan *confusion matrix*, dimana rasio data aktual akan dibandingkan dengan data hasil klasifikasi dari model. Instrumen evaluasi klasifikasi diukur dengan menggunakan ukuran *accuracy*, *precision*, *recall* dan *f-measure*.
6. *Deployment*, pada tahapan ini data *keyword* yang diinputkan, proses klasifikasi, evaluasi, dan representasi pengetahuan yang didapat di-*deploy* melalui sebuah prototipe aplikasi yang dirancang dengan tujuan memudahkan perusahaan atau organisasi dalam proses identifikasi persepsi publik terhadap suatu produk/layanan yang diinputkan sebagai *keyword*.

3. HASIL DAN PEMBAHASAN

Paparan berikut ini menjelaskan proses pengembangan prototipe aplikasi dengan mengimplementasikan algoritma NBC untuk mendapatkan representasi sentimen atau persepsi masyarakat yang didapat dari teks pada Twitter dari masukan berupa *keyword* yang menggambarkan produk atau layanan, atau nama brand atau nama organisasi. Tahapan pengembangan hingga evaluasi aplikasi dipaparkan mengikuti software proses model prototyping.

3.1. Analisa Kebutuhan Sistem

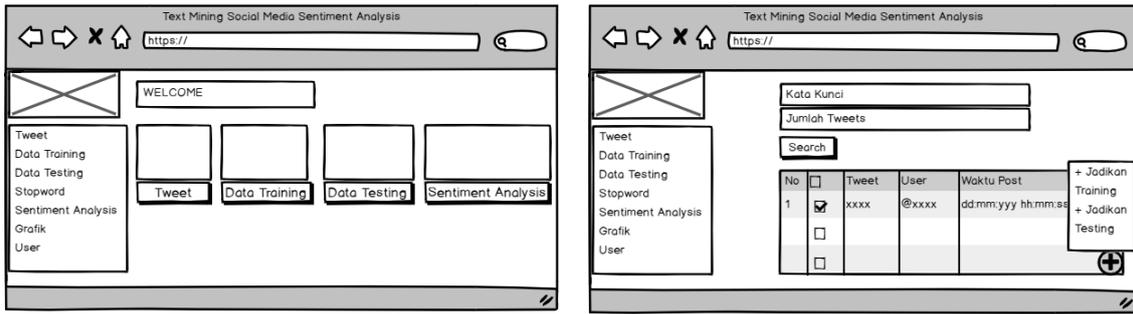
Secara garis besar kebutuhan sistem dikelompokkan ke dalam dua kategori, yakni kebutuhan *functional* dan *non-functional*. Paparan rinci kebutuhan tersebut dapat dilihat pada tabel 1 di bawah ini:

Tabel 1. Daftar kebutuhan fungsional dan non-fungsional dari sistem

ID	Deskripsi Kebutuhan Fungsional
R1	User dapat memasukkan <i>keyword</i> sesuai kebutuhan yang kemudian oleh system digunakan untuk mengumpulkan semua <i>tweets</i> yang relevan dengan <i>keyword</i> tersebut.
R2	User dapat membatasi jumlah <i>tweets</i> yang dikumpulkan sesuai kebutuhan.
R3	User dapat mereview hasil klasifikasi tiap-tiap <i>tweets</i> dengan memperhatikan nilai probabilitas positif, negatif, dan netral serta hasil akhir klasifikasi sentimen.
R4	User dapat menyimpan hasil analisa dalam bentuk <i>log</i> sehingga dapat digunakan untuk monitoring perkembangan sentimen dari waktu ke waktu.
R5	User dapat melihat representasi hasil akhir dari klasifikasi sentimen dan perkembangan sentimen dalam bentuk <i>chart</i> sederhana.
ID	Deskripsi Kebutuhan Non-Fungsional
R6	Rancangan <i>user interface</i> mudah dipahami dengan proses interaksi yang simpel.
R7	Output dalam bentuk <i>chart</i> yang jelas dan kontras untuk membedakan masing-masing kelas dalam sentimen
R8	Proses klasifikasi berlangsung dalam tempo waktu tidak melebihi 5 menit untuk dataset dengan jumlah record antara 300 – 500 <i>tweets</i> .

3.2. Modelling Quick Design

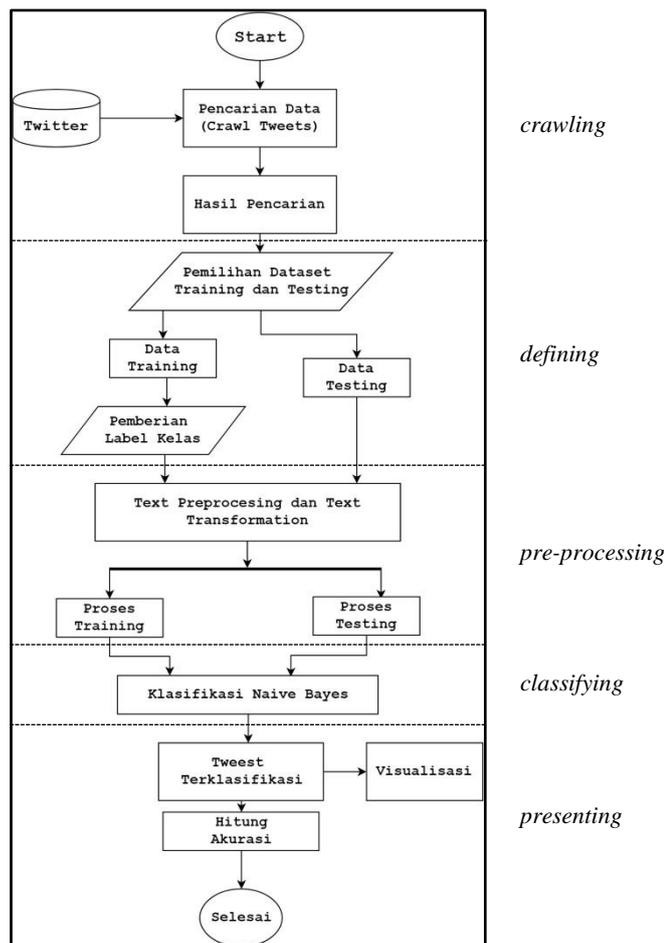
Pada tahapan ini dikembangkan model desain yang fokus pada representasi aspek-aspek software yang tampak oleh user yang meliputi *interface layout*, bentuk keluaran, dan kontrol interaksi. Berikut ini adalah beberapa desain utama pada prototipe aplikasi analisa sentimen dari media sosial Twitter ditunjukkan pada gambar 2.



Gambar 2. Rancangan *interface* untuk prototipe aplikasi sentimen analisis

3.3. Pengembangan Prototipe

Pada tahapan ini spesifikasi sistem yang telah ditentukan diubah menjadi unit kerja yang lebih teknis. Sejumlah *tool* pengembang seperti *html*, *php*, *css*, dan *javascript* digunakan untuk konstruksi user interface, kontrol, dan implementasi algoritma NBC. Untuk pengelolaan basis data sistem memanfaatkan *mysql*, sedangkan *framework* yang digunakan dalam pembuatan sistem adalah *framework laravel*. Secara umum struktur program terbangun dalam 5 bagian yang meliputi: pengumpulan data *tweets* (*crawling*), pendefinisian dataset (*testing* dan *training*), *preprocessing* dan transformasi text, proses klasifikasi, visualisasi dan evaluasi, sebagaimana tampak pada gambar 3.



Gambar 3. Struktur prototipe program aplikasi analisa sentimen

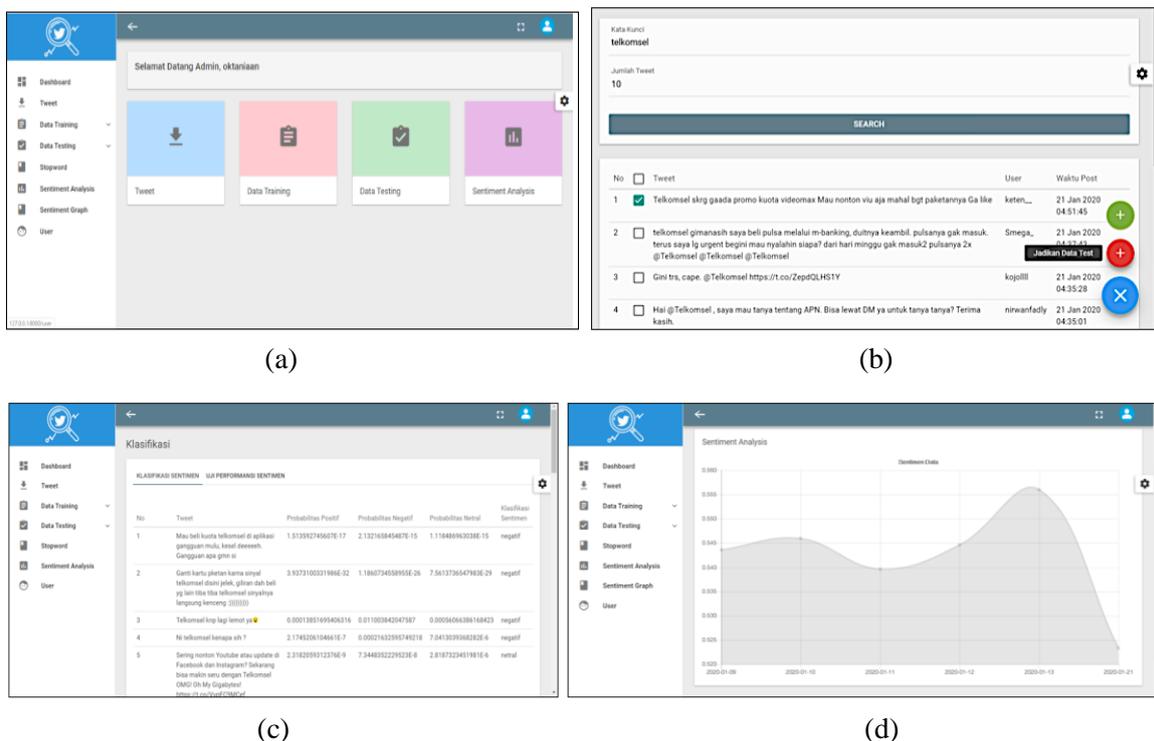
Berikut ini adalah hasil keluaran dari pengembangan prototipe pada modul *crawling* dan transformasi teks Twitter pada gambar 4. Keluaran ini merupakan hasil olahan dari bagian *pre-processing* sebagaimana tampak pada struktur program pada gambar 4

Username	Tweet	Tanggal	Tweets	Transformation
Anggijendriadi	Min @ [redacted] kenapa jaringan [redacted] 2 minggu belakangan ini lemot banget di gedung menara karya, mohon diperbaiki dong min. Terima kasih.	2019-10-04	Min @ [redacted] kenapa jaringan [redacted] 2 minggu belakangan ini lemot banget di gedung menara karya, mohon diperbaiki dong min. Terima kasih.	[jaring] [minggu] [belakang] [lemot] [gedung] [menara] [karya] [mohon] [baik] [terima] [kasih]
Cepots	Ini tiap bulan bayar mahal kok internet gue lemot bgt, sinyalnya hilang-hilangan @ [redacted] @ [redacted] ?	2019-10-04	Ini tiap bulan bayar mahal kok internet gue lemot bgt, sinyalnya hilang-hilangan @ [redacted] @ [redacted] ?	[bayar] [mahal] [internet] [lemot] [sinyal] [hilang]
Pikhacuwu	Kalo mau upgrade kartu telkomsel yg 3G ke 4G bayar berapa ya gengs?	2019-10-04	Kalo mau upgrade kartu telkomsel yg 3G ke 4G bayar berapa ya gengs?	[kalo] [kartu] [bayar] [berapa]
pusstyle07	Guys yg pake [redacted], ada yg bisa rekomendasiin paketan internet yang murah meriah dan banyak menguntungkan ga??	2019-10-04	Guys yg pake [redacted], ada yg bisa rekomendasiin paketan internet yang murah meriah dan banyak menguntungkan ga??	[rekomendasi] [paket] [internet] [murah] [riah] [untung]
Nurulhanifaaaa	Terima kasih [redacted] hehehe promo disaat yg sangat tepat	2019-10-04	Terima kasih [redacted] hehehe promo disaat yg sangat tepat	[terima] [kasih] [promo] [saat] [tepat]
f_rahmadani67	Asik [redacted] promonya mantab 50gb ☺	2019-10-04	Asik [redacted] promonya mantab 50gb ☺	[asik] [promo] [mantab]

Gambar 4. Keluaran dari proses *pre-processing* dan *text transformation*

3.4. Deployment dan Feedback

Di akhir proses pengembangan sistem selanjutnya memasuki tahapan pengujian, dimana proses verifikasi dan validasi dilakukan dengan mengacu pada daftar spesifikasi kebutuhan yang telah ditentukan dan logika program melalui skenario testing program yang telah dibuat.



Gambar 5. Kondisi prototipe saat dijalankan
 (a) tampilan awal, (b) hasil *crawling*, (c) proses klasifikasi, (d) representasi hasil klasifikasi

Dari hasil evaluasi didapati prototipe yang tampak pada gambar 5 berjalan dengan baik dimana *keyword* yang dimasukkan akan mentrigger Twitter API untuk melakukan *crawling* selama tujuh hari sebelumnya pada gambar 5.b. Sementara itu masukkan untuk jumlah *tweets* yang

dikelola akan mengontrol jumlah *tweets* yang akan ditampilkan pada tabel view. Tahapan proses *mining* dapat terpantau pada bagian/sheet *Text-PreProcessing*, *Text Transformation*, *Unigram*, dan *Unigram Frequency*. Demikian juga dengan proses klasifikasi, sistem secara detail dapat menampilkan hasil klasifikasi dari suatu teks *tweets* berikut dengan nilai probabilitasnya yang dikalkulasi mengikuti rumus persamaan 3 dan 4 dan tampak pada gambar 5.c.

Sebagai contoh apabila dari hasil *preprocessing* didapatkan term [parah] [paket] [mahal] [sinyal] [hilang] maka $P(V_j)$ pada setiap kategori dapat dihitung dengan menggunakan rumus persamaan 3.

1. Probabilitas $P(C_{positif}) \rightarrow P(C_{positif}) = \frac{|2|}{|6|} = 0.333$
2. Probabilitas $P(C_{negatif}) \rightarrow P(C_{negatif}) = \frac{|2|}{|6|} = 0.333$
3. Probabilitas $P(C_{netral}) \rightarrow P(C_{netral}) = \frac{|2|}{|6|} = 0.333$

Untuk selanjutnya proses identifikasi dalam penentuan kelas mana yang paling tepat untuk potongan *tweets* tersebut perhitungan dilakukan dengan mengacu pada rumus persamaan 4.

1. Probabilitas data *testing* terhadap kelas Positif

P(parah Positif): $(0+1)/(8+29) = 0.0270$	P(sinyal Positif): $(0+1)/(8+29) = 0.0270$
P(paket Positif): $(0+1)/(8+29) = 0.0270$	P(hilang Positif): $(0+1)/(8+29) = 0.0270$
P(mahal Positif): $(0+1)/(8+29) = 0.0270$	
2. Probabilitas data *testing* terhadap kelas Negatif

P(parah Negatif): $(0+1)/(17+29) = 0.0217$	P(sinyal Negatif): $(1+1)/(17+29) = 0.0435$
P(paket Negatif): $(0+1)/(17+29) = 0.0217$	P(hilang Negatif): $(1+1)/(17+29) = 0.0435$
P(mahal Negatif): $(1+1)/(17+29) = 0.0435$	
3. Probabilitas data *testing* terhadap kelas Netral

P(parah Netral): $(0+1)/(10+29) = 0.0256$	P(sinyal Netral): $(0+1)/(10+29) = 0.0256$
P(paket Netral): $(1+1)/(10+29) = 0.0513$	P(hilang Netral): $(0+1)/(10+29) = 0.0256$
P(mahal Netral): $(0+1)/(10+29) = 0.0256$	

Setelah pada data *testing* probabilitas pada setiap *term* terhadap setiap kelas telah diketahui. Maka dapat dilanjutkan dengan menghitung probabilitas dari $P(x_i|V_j)P(V_j)$ untuk menentukan kelas yang tepat pada data *testing*.

1. Kelas Positif atau $P(\text{Positif}|\text{testing})$

$$= P(C_{positif}) \times P(\text{parah}|\text{Positif}) \times P(\text{paket}|\text{Positif}) \times P(\text{mahal}|\text{Positif}) \times P(\text{sinyal}|\text{Positif}) \times P(\text{hilang}|\text{Positif})$$

$$= 0.333 \times 0.0270 \times 0.0270 \times 0.0270 \times 0.0270 \times 0.0270$$

$$= 4.78 \times 10^{-9}$$
2. Kelas Negatif atau $P(\text{Negatif}|\text{testing})$

$$= P(C_{negatif}) \times P(\text{parah}|\text{Negatif}) \times P(\text{paket}|\text{Negatif}) \times P(\text{mahal}|\text{Negatif}) \times P(\text{sinyal}|\text{Negatif}) \times P(\text{hilang}|\text{Negatif})$$

$$= 0.333 \times 0.0217 \times 0.0217 \times 0.0435 \times 0.0435 \times 0.0435$$

$$= 1.29 \times 10^{-8}$$
3. Kelas Netral atau $P(\text{Netral}|\text{testing})$

$$= P(C_{netral}) \times P(\text{parah}|\text{Netral}) \times P(\text{paket}|\text{Netral}) \times P(\text{mahal}|\text{Netral}) \times P(\text{sinyal}|\text{Netral}) \times P(\text{hilang}|\text{Netral})$$

$$= 0.333 \times 0.0256 \times 0.0513 \times 0.0256 \times 0.0256 \times 0.0256$$

$$= 7.34 \times 10^{-9}$$

Dari proses perhitungan tampak bahwa probabilitas tertinggi ada pada kelas Negatif (1.29×10^{-8}) maka *tweets* “parah paket mahal sinyal hilang” diklasifikasikan sebagai sentimen negatif.

Di akhir proses, sistem akan menunjukkan nilai akhir dari proses sentimen analysis berdasarkan data yang telah di-*crawling*. Apabila dikehendaki user dapat menyimpan *log* hasil perhitungan dan menampilkannya dalam bentuk *chart* dengan kurun waktu tertentu tampak pada gambar 5.d.

4. KESIMPULAN DAN RISET LANJUTAN

Berdasarkan hasil uraian pada pembahasan dan sejumlah temuan dalam proses pengembangan penelitian ini, didapati bahwa prototipe aplikasi dapat bekerja dengan baik sesuai dengan kebutuhan sistem. Perusahaan atau organisasi dapat dengan mudah mengidentifikasi persepsi masyarakat terkait dengan produk/layanan yang sedang diamati. Proses *crawling*, *preprocessing* dan representasi hasil dapat dengan mudah dimanfaatkan walau hanya dengan literasi komputasi yang minim. Dari sejumlah percobaan didapati penerapan algoritma NBC pada analisa sentimen dari data *tweets* ini memiliki rerata akurasi presisi sebesar 88%, dengan pencapaian *recall* 90% serta *F1 score* sebesar 89%. Proses *crawling* sendiri secara umum bergantung pada kualitas koneksi internet, dari sejumlah percobaan proses *crawling* yang menghasilkan rerata 100 *tweets* membutuhkan rerata waktu kurang dari 5 menit, demikian juga untuk proses klasifikasinya.

Sejumlah kelemahan terkait dengan riset ini yang membuka kesempatan untuk dapat dilakukannya perbaikan sebagai kelanjutan riset diantaranya adalah perlunya mendapatkan tool atau strategi yang memungkinkan durasi periode *crawling* melebihi 7 hari, penyusunan algoritma pencarian yang dapat mengkombinasikan *keyword* pencarian dengan sejumlah notasi logika agar *query* pencarian lebih fleksibel dengan ruang lingkup hasil yang lebih tajam. Representasi hasil analisa sentimen dalam bentuk *chart* hanya memaparkan rasio sentimen positif dan netral terhadap keseluruhan *tweets* yang didapat, detail jumlah untuk masing-masing sentimen belum tergambarkan dengan jelas sehingga perlu ada perbaikan dalam penyediaan informasi tersebut.

DAFTAR PUSTAKA

- [1] K., Simon, "Digital 2020: Indonesia" 2020. [Online]. Available: <https://datareportal.com/reports/digital-2020-indonesia> [Accessed 03 03 2020].
- [2] Similarweb LTD, "Top sites ranking for all categories in Indonesia" 2020. [Online]. Available: <https://www.similarweb.com/top-websites/indonesia> [Accessed 04 03 2020].
- [3] Obar, Jonathan A.; Wildman, Steve (2015). "Social media definition and the governance challenge: An introduction to the special issue". *Telecommunications Policy*. 39 (9): 745–750. doi:10.1016/j.telpol.2015.07.014. SSRN 2647377.
- [4] Statista, "Number of Twitter users in Indonesia from 2014 to 2019" 2020. [Online]. Available: <https://www.statista.com/statistics/490548/twitter-users-indonesia/> [Accessed 04 03 2020]
- [5] Carley, Kathleen & Malik, Momin & Kowalchuck, Michael & Pfeffer, Juergen & Landwehr, Peter. (2015). Twitter Usage in Indonesia. 10.13140/RG.2.1.2163.9925.
- [6] A., Alamsyah, "(Big) Data Analytics for Economics, Business and Management: A Social Network Approach," dalam In Workshop Big Data Puslitbang Aptika dan IKP, 2015.
- [7] J. Ipawati, Kusri dan E. T. Luthfi, "Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen," Indonesian Journal on Networking and Security, vol. VI, no. 1, pp. 28-36, 2017.
- [8] R. Sari, "Komparasi Algoritma Support Vector Machine, Naïve Bayes Dan C4.5 Untuk Klasifikasi SMS," Indonesian Journal on Computer and Information Technology, vol. II, no. 2, pp. 7-13, 2017.
- [9] Hermanto, S. J. Kuryanti dan S. N. Khasanah, "Comparison of Naïve Bayes Algorithm, C4.5 and Random Forest for Service Classification Ojek Online," Journal Publications & Informatics Engineering Research, vol. III, no. 2, pp. 266-274, 2019.

-
- [10] Y. Aggraini, Sucipto dan R. Indriati, "Cyberbullying Detection Modelling at Twitter Social Networking," *Information System*, vol. VI, no. 2, pp. 113-118, 2018.
- [11] S. K. Lidya, O. S. Sitompul dan S. Efendi, "Sentimen Analisis pada Teks Bahasa Indonesia Menggunakan Support Vector Machine dan K-Nearest Neighbor," dalam *Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA)*, Yogyakarta , 2015.
- [12] R. N. Devita, H. W. Herwanto dan A. P. Wibawa , "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Bahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer* , vol. 5, no. 4, pp. 427-434, 2018 .
- [13] Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (pp. 29-39). London, UK: Springer-Verlag.
- [14] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," *Institute for Logic, Language and Computation Universiteit Van Amsterdam The Netherlands*, 2003.
-