

Implementasi Scraping Data Untuk Sentiment Analysis Pengguna Dompert Digital Dengan Menggunakan Algoritma Machine Learning

Implementation of Data Scraping for Sentiment Analysis of Digital Wallet Users Using Machine Learning Algorithm

Eka Yuniar¹, Dwi Safiroh², Dian Wahyuningsih³

^{1,2}Sistem Informasi, STMIK PPKIA Pradnya Paramita

³Teknologi Informasi, STMIK PPKIA Pradnya Paramita

eka@stimata.ac.id, utsalinah@stimata.ac.id, dian@stmata.ac.id

Received:	Revised:	Accepted:	Available online:
21.03.2022	15.04.2022	21.04.2022	30.04.2022

Abstrak: Analisis media sosial merupakan alat yang efektif untuk memahami sikap, preferensi, dan pendapat orang. Bagi sebuah perusahaan, analisis media sosial dapat membantu perusahaan untuk mengambil keputusan tentang kebutuhan, sikap, opini atau tren tentang calon pelanggan atau calon pelanggan. Hasil penelitian Wearesocial Hootsuite menunjukkan bahwa dari 150 juta orang Indonesia, 88% aktif menggunakan media sosial. Alhasil, banyak informasi yang dulunya sulit didapat kini sangat mudah didapat. Dengan banyaknya data, pendataan dapat dipermudah dengan menggunakan sistem pendataan otomatis. Penggunaan Machine Learning Algorithm sebagai metode digunakan untuk mengetahui sejauh mana tingkat pencarian dari data media sosial yang dikumpulkan untuk mendapatkan hasil yang akurat terkait perubahan perilaku, gaya hidup, dan aktivitas masyarakat dalam menghadapi pandemi ini. Scraping data dilakukan dengan menggunakan data media sosial Twitter dengan berbagai hashtag yang mendukung opinion mining yang terdapat di masyarakat. Hasil yang diperoleh adalah analisis pengguna gopay memiliki tingkat sentimen positif sebesar 79,6%, sentimen negatif sebesar 20,4% dari total 250 data. Sementara itu, pengguna Linkaja memiliki tingkat sentimen positif sebesar 62,1%, komentar negatif 37,9% dari 250 data komentar yang diambil. Hasil pengujian juga dilakukan proses penghitungan tingkat akurasi dengan recall dan presisi yaitu 87% untuk Gopay dan 89% untuk LinkAja.

Kata kunci: Dompert digital, Scrapping, Gopay, LinkAja

Abstract: Analysis of social media is an effective tool for understanding people's attitudes, preferences and opinions. For a company, social media analysis can help companies to make decisions about needs, attitudes, opinions or trends about potential customers or potential customers. Hootsuite's Wearesocial research results show that out of 150 million Indonesians, 88% are actively using social media. As a result, a lot of information that was once difficult to obtain is now very easy to obtain. With so much data, data collection can be made easier by using an automated data collection system. The use of Machine Learning Algorithm as a method is used to find out how far the level of search from social media data collected is to get accurate results related to changes in behavior, lifestyle, and community activities in dealing with this pandemic. Scraping data is done using Twitter social media data with various hashtags that support opinion mining contained in the community. The results obtained are the analysis of gopay users having a positive sentiment level of 79,6%, negative sentiment of 20,4% from total of 250 data. Meanwhile, Linkaja users have a positive sentiment level of 62,1%, 37,9% negative comments from 250 data comments taken. The test results also carried out a process of calculating the level of accuracy with recall and precision, namely 87% for Gopay and 89% for LinkAja.

Keywords: Digital wallet, Scrapping, Gopay, LinkAja

1. PENDAHULUAN

Media sosial atau sosial media atau sosmed adalah sebuah platform digital yang memungkinkan penggunanya saling berkomunikasi atau membagikan konten baik berupa tulisan, foto, video dan lain-lain. Dengan adanya sosmed ini seolah mempersempit jarak antar manusia, kita dapat bertukar kabar dan saling merekomendasikan sesuai sesuai minat kepada orang lain.

Bagi kita pengguna sosial media pasti akan merasakan dampak dalam pemakaiannya, antar lain: meningkatkan keuntungan dari penjualan produk, meningkatkan kredibilitas perusahaan atau organisasi, menjalin kerjasama bisnis secara luas. Meski dalam penggunaannya pasti dampak negatif pun tidak dapat kita hindari. Namun dengan adanya media sosial kita bisa mengambil manfaat yang dapat kita ambil antara lain: mendapatkan informasi secara aktual, mampu berkomunikasi secara real-time, meningkatkan brand awareness, meningkatkan trafik dan peringkat pencarian serta membangun relasi dengan cepat dan luas.

Dengan beberapa dampak dan manfaat yang dapat kita ambil dari media sosial, saat ini banyak perusahaan yang berlomba-lomba menggunakan media sosial ini untuk memenuhi berbagai kebutuhan informasi, namun dengan semakin berkembangnya data yang beredar di media sosial tentunya kita harus dapat memilih dan memilah data atau informasi apa saja yang kita perlukan.

Semakin berkembangnya media sosial yang digunakan untuk transaksi baik berupa produk dan jasa maka kebutuhan pada pembayaran yang cepat dan mudah pun semakin tinggi. Biasanya kita hanya dapat melakukan transaksi pembayaran melalui cash dan atau transfer bank, namun hari ini telah banyak bermunculan dompet digital yang memudahkan transaksi.

Bank Indonesia menyebutkan transaksi nontunai mengalami peningkatan sebesar 24% pada tahun 2018 lalu. Meskipun transaksi di pasar masih didominasi oleh cash (90%), tetapi pertumbuhan transaksi nontunai tersebut dinilai baik oleh pihak regulator.

Tren positif tersebut disambut baik oleh pelaku industri untuk terus menggenjalkan penggunaan layanan digital melalui cara yang unik, mulai dari promosi hingga kolaborasi dengan sektor lain. Contoh kerja sama tersebut, antara lain Go-Pay dengan Alfamart, OVO dengan Tokopedia dan Grab, serta DANA dengan Ramayana.

Analisis terhadap media sosial adalah alat yang efektif untuk memahami sikap, preferensi dan opini masyarakat. Bagi suatu perusahaan, analisis media sosial dapat membantu perusahaan untuk membuat keputusan mengenai kebutuhan, sikap, pendapat atau trend tentang pelanggan atau calon pelanggan potensial. Hasil riset Wearesocial Hootsuite menunjukkan bahwa dari 150 juta penduduk Indonesia 88% nya aktif menggunakan media sosial [1]. Hasilnya, banyak informasi yang dulu nya sulit untuk didapatkan, kini sangat mudah untuk didapat. Dengan banyaknya data tersebut, pengumpulan data dapat dipermudah dengan menggunakan sebuah sistem pengumpulan data otomatis.

Algoritma Machine Learning telah banyak digunakan sebagai metode untuk mengetahui seberapa jauh tingkat penelusuran dari data social media yang dikumpulkan untuk mendapatkan hasil [2] [3] yang akurat terkait dengan perubahan perilaku, gaya hidup, serta aktivitas masyarakat dalam menghadapi pandemic ini. Scraping data dilakukan dengan menggunakan data social media twitter dengan berbagai hastag yang mendukung opinion mining yang terdapat dalam masyarakat.

Tujuan dari penelitian ini adalah untuk mendapatkan sentimentil pengguna terhadap dompet digital yang beredar saat ini sehingga memudahkan perusahaan untuk pengambilan keputusan.

2. KAJIAN LITERATUR

2.1 Analisis Sentimen

Analisis Sentimen sebagai salah satu menganalisis opini, menilai dan mengetahui penerimaan pengguna atas sebuah produk, isu dan topik tertentu studi untuk [1]. Klasifikasi dari hasil analisis sentiment dapat digunakan untuk memprediksi sentiment polarity yang diperoleh dari pengguna [2]. analisis sentiment telah dimanfaatkan diberbagai bidang mulai dari Kesehatan [3], politik [4] hingga ekonomi [5], kemampuannya dalam mendukung pembuatan keputusan [6], mulai dari keputusan bisnis sampai dengan keputusan politik.

Dalam rangka mendapatkan analisis sentimen pengguna terhadap dompet digital, dapat melalui beberapa tahapan, diantaranya adalah mempersiapkan data.

Sentimen analisis dapat diklasifikasikan kedalam kelas sentiment bersifat positif, negative, dan netral [10].

- 1) Sentimen positif : Menurut Kamus Besar Bahasa Indonesia merupakan reaksi/ sikap yang meningkatkan nilai seseorang atau sesuatu.
- 2) Sedangkan Sentimen Negatif : merupakan reaksi atau sikap yang menurunkan nilai seseorang atau sesuatu.

2.2 Scraping

Proses mempersiapkan data dalam penelitian ini berkaitan dengan data dari media sosial, dimana data ini merupakan sumber daya yang paling dinamis berkaitan perilaku manusia, kondisi yang sedang terjadi di masyarakat. Namun data yang berasal dari media sosial tersebut tidak akan dapat memberikan manfaat apabila belum diolah. Aktifitas mengolah data, atau lebih tepatnya mengekstraksi data dari web atau media sosial guna mengetahui pola yang tersembunyi didalamnya disebut dengan Scraping [7]

Teknik Scraping merupakan salah satu bagian dalam analisis sentimen terdiri dari beberapa tahapan, yaitu :

- Web scraper akan diberi sejumlah url untuk dimuat sebelum proses scraper dilakukan.
- Selanjutnya scraper memuat kode HTML laman yang dimaksud.
- Berdasarkan data yang berhasil dimuat pada tahapan sebelumnya, scraper akan melakukan ekstraksi data sesuai dengan kebutuhan data yang diperlukan oleh pengguna.
- Tahapan terakhir adalah mengumpulkan data hasil ekstraksi kedalam 1 format, CSV, bahkan JSON supaya dapat digunakan sebagai API.

Proses selanjutnya setelah data diekstrak dengan salah satu Teknik scraping, dilanjutkan dengan pembersihan dan pemrosesan data.

2.3 Text Preprocessing

Text preprocessing merupakan tahap awal dimana data teks dengan noise akan dikurangi agar dapat diolah lebih lanjut, untuk mengurangi noise tersebut. Tahap text preprocessing ini mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi knowledge discovery sistem text mining. Data tekstual akan diproses dalam beberapa tahapan text preprocessing yaitu case folding, tokenisasi, filtering dan stemming. Dengan dilakukannya text preprocessing akan terbentuk dataset bersih, dataset yang terbentuk dari proses ini akan memudahkan dalam pemrosesan sistem. [12]

- 1) *Case Folding Case* : dilakukan untuk mengubah setiap karakter didalam teks menjadi huruf kecil. Tidak semua kata dalam teks konsisten dalam penggunaan huruf kapital disinilah tujuan dilakukan case folding untuk mengkonversi setiap karakter dalam kata menjadi huruf kecil
- 2) *Filtering* : Filtering merupakan proses dalam *text preprocessing* setelah tokenisasi, filtering dilakukan untuk mengambil kata penting hasil tokenisasi. Proses *filtering* dalam membuang kata-kata yang tidak digunakan atau *stopword* terdapat dalam *bag of words*.
- 3) *Stopword* merupakan daftar kata-kata yang tidak mempresentasikan isi dari suatu dokumen teks, *stopword* dilakukan untuk meghilangkan kata atau term yang tidak memiliki arti. Daftar stoplist akan dibuat sebelum melakukan proses *stopword removal*, jika kata-kata terdapat dalam daftar stoplist, maka kata tersebut akan dihapus, sehingga kata-kata yang tersisa akan dianggap kata yang mencirikan isi suatu dokumen.
- 4) *Stemming* : merupakan proses mengubah kata menjadi bentuk dasarnya. *Stemming* dilakukan untuk menyeragamkan bentuk kata. Tujuan dari proses *stemming* adalah menghilangkan imbuhan-imbuhan baik itu berupa *prefiks*, *sufiks*, maupun *konfiks* yang ada pada setiap kata *Stemming* dalam penelitian ini dilakukan berdasarkan aturan morfologi bahasa Indonesia.

2.4 Twitter

Aktifitas pengolahan data dimulai dengan penentuan sumber data mana yang akan dimanfaatkan. Pada penelitian ini sumber data yang digunakan berasal dari media sosial Twitter, aplikasi ini merupakan salah satu media sosial yang populer dimanfaatkan oleh masyarakat, hal ini ditunjukkan oleh data dari laporan Statista yang menyatakan bahwa pengguna Twitter mencapai 18,45 juta pada awal tahun 2022 [8].

2.5 Algoritma Naïve Bayes

Naïve bayes merupakan salah satu metode yang digunakan pada text mining yang sederhana tetapi memiliki keakuratan yang tinggi dalam mengklasifikasi. [14]

Dalam algoritma naïve bayes setiap dokumen dipresentasikan dengan pasangan atribut “ $x^1, x^2, x^3, \dots, x^n$ ” dimana x^1 adalah kata pertama, x^2 adalah kata kedua dan seterusnya. Probabilitas dari semua dokumen (posterior probability) suatu dokumen pada suatu kategori dapat dihitung dengan menggunakan persamaan berikut :

$$V_{MAP} = \underset{V_j \in V}{\arg \max} \frac{P(X_1, X_2, \dots, X_n | V_j) P(V_j)}{P(X_1, X_2, \dots, X_n)} \pi \tag{1}$$

Klasifikasi teks menggunakan naïve bayes dilakukan dengan memaksimalkan nilai dari persamaan, asumsi kemandirian bersyarat yang "naif" memegang peranan. Untuk penyebut $P(X_1, X_2, \dots, X_n | V_j)$ nilainya konstan atau sama untuk semua kategori (V_j). Sehingga persamaan dapat ditulis sebagai berikut,

$$V_{MAP} = \underset{V_j \in V}{\text{arg max}} P(X_1, X_2, \dots, X_n | V_j)P(V_j) \tag{2}$$

Persamaan diatas dapat disederhanakan menjadi,

$$V_{MAP} = \underset{V_j \in V}{\text{arg max}} \prod_{i=1}^n P(X_i | V_j)P(V_j) \tag{3}$$

Keterangan:

V_j = Kategori tweet $j = 1,2,3, \dots, n$

Dimana dalam penelitian ini $j1$ = kategori *tweet* sentiment positif, $j2$ = kategori *tweet* sentiment negative, dan $j3$ = kategori *tweet* sentiment *netral*.

$P(X_i | V_j)$ = Probabilitas x_i pada kategori V_j

$P(V_j)$ = Probabilitas dari V_j

Untuk nilai dan dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut:

$$P(V_j) = \frac{|docs_j|}{|contoh|} \tag{4}$$

$$P(X_i | V_j) = \frac{n_k + 1}{n + |kosakata|} \tag{5}$$

Keterangan :

$|docs_j|$ = Jumlah dokumen setiap kategori j .

$|contoh|$ = Jumlah dokumen dari semua kategori

n_k = Jumlah frekuensi kemunculan kata

n = Jumlah frekuensi kemunculan kata dari setiap kategori

$|kosakata|$ = Jumlah semua kata dari semua kategori [12]

2.6 Dataset

Dataset yang digunakan berasal dari twitter berupa data tweet sebagai data tekstual. *Dataset* merupakan data latih dan data uji, dari data ini akan dihitung akurasi berdasarkan persentase jumlah data latih dan data uji. Tahap pertama adalah melakukan pelatihan terhadap data opini yang telah diketahui kategorinya, dan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

2.7 Recall dan Precision

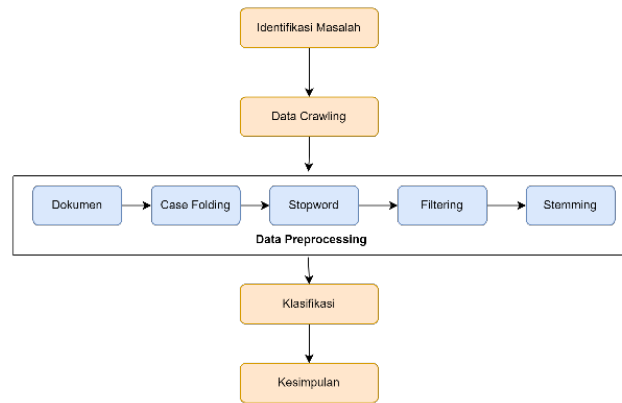
Recall dan *precision* digunakan untuk mengukur tingkat efektivitas sistem temu kembali informasi. Nilai dari *recall* dan *precision* biasanya ditunjukkan dalam persen (%). [12]

$$\text{Recall} = \frac{\text{jumlah dokumen relevan yang terpanggil (a)}}{\text{jumlah dokumen relevan yang ada di dalam database (a + c)}} \times 100 \tag{6}$$

$$\text{Precision} = \frac{\text{jumlah dokumen relevan yang terpanggil (a)}}{\text{jumlah dokumen yang terpanggil dalam pencarian (a + b)}} \times 100 \tag{7}$$

3. METODE PENELITIAN

Dalam penelitian ini akan dilakukan perhitungan tentang sentimen pengguna dompet digital yang dilakukan dalam beberapa tahap (Gambar 1), yaitu:



Gambar 1. Metode Penelitian

3.1 Identifikasi Masalah

adalah melakukan identifikasi masalah. Permasalahan yang akan diangkat adalah sentiment pengguna dompet digital pada pengguna Gopay dan LinkAja. Dilakukan pengumpulan data tweet dari responden melalui API twitter.

3.2 Data Crawling

Data yang akan digunakan adalah data ekstraksi dari twitter untuk pengguna gopay dan linkaja. Dataset yang digunakan adalah data tweet yang berasal dari akun resmi dari @GopayIndonesia dan @LinkAja, sejumlah masing-masing 250 data yang diambil pada tanggal 1 Mei 2022.

Tabel 1. Data tweet Gopay yang diambil

No	tweets
1	@adlimanes Hai kak, kami mohon maaf atas kendala yang kamu alami. Agar dapat kami lakukan pengecekan lebih lanjut kami mohon kesediaan kamu untuk menginformasikan nomor HP dan email yang terdaftar via DM ya. Terima kasih^ast
2	Yuk cek semua fotonya buat cari tau promonya apa aja. Syarat dan ketentuan berlaku. Cek info lengkapnya di https://t.co/rwk6uP9AvO https://t.co/kpJFgITrzb
3	Udah gajian atau belum, pakai GoPay #SemuaJadiBisa bayar ini itu lebih hemat di mana-mana! Ada promo DISKON & CASHBACK hingga 70% yang bisa kamu dapetin, buat belanja kebutuhan, jajan camilan, hingga pesan taksi Bluebird di aplikasi MyBluebird.
4	@enouadio Hai kak, kami mohon maaf atas kendala yang kamu alami. Agar dapat kami lakukan pengecekan lebih lanjut kami mohon kesediaan kamu untuk menginformasikan nomor HP yang terdaftar dan screenshot bukti top up yang kamu lakukan melalui DM ya. Terima kasih^shl
5	@liviay00 Hai kak, kami mohon maaf atas kendala yang kamu alami. Agar dapat kami lakukan pengecekan lebih lanjut kami mohon kesediaan kamu untuk menginformasikan nomor HP yang terdaftar dan screenshot bukti transaksi kamu melalui DM ya. Terima kasih^shl

3.3 Data Preprocessing

1. Case Folding

Pada tahap ini, dilakukan seleksi komentar yang mengandung kata atau hastag, dan retweet, karena komentar tweet akan mengganggu dalam proses Analisis sentiment tweet.

Tabel 2. Tabel karakter yang dihapus

No	Karakter yang dihapus	Arti karakter
1	https://[w\.\V]*	Hyperlink
2	:lt;[nbsp]>	#lt;nbsp,gt,&

3	@ #)\w+	Dan
4	[^A-Za-z0-9\s\-\]	Selain huruf, spasi, strip
5	(\-\)	-
6	\n	Enter
7	\s{2,}	Spasi lebih dari 2
8	^rt.*	Retweet

2. Stopword

Selanjutnya yang dilakukan adalah membuat Daftar stoplist akan dibuat sebelum melakukan proses stopword removal, jika kata-kata terdapat dalam daftar stoplist, maka kata tersebut akan dihapus, sehingga kata-kata yang tersisa akan dianggap kata yang mencirikan isi suatu dokumen. Adapun kata-kata yang kami pilih antara lain saya, sampai, yang, di, dan, dll.

Tabel 3. Hasil stopword Gopay

No	tweet	Stopwords	remove
0	34 kata	25 kata	9 kata
1	18 kata	16 kata	2 kata
2	36 kata	30 kata	6 kata
:	:	:	:
146	14 kata	10 kata	4 kata

Tabel 4. Hasil stopword LinkAja

No	tweet	Stopwords	remove
0	37 kata	32 kata	5 kata
1	10 kata	10 kata	0 kata
2	39 kata	36 kata	3 kata
:	:	:	:
139	38 kata	31 kata	7 kata

3. Filtering

Dilakukan untuk membersihkan kata-kata hasil *stopwords*. Berikut ini contoh dari proses filtering. Berikut ini proses hasil filtering :

Tabel 5. Hasil filtering LinkAja

hai	kak	mohon	maaf
atas	kendala	Kamu	alami
dapat	lakukan	pengecekan	lebih
lanjut	mohon	kesediaan	kamu
Menginformasikan	nomor	hp	Email
Terdaftar	Via	Dm	Terima
kasih			

4. Stemming

Dilakukan untuk membersihkan kata-kata hasil *stemming*. Berikut ini contoh dari proses filtering. Berikut ini proses hasil stemming :

Tabel 6. Hasil stemming LinkAja

mohon	maaf	kendala	Kamu
alami	dapat	lakukan	pengecekan
lebih	lanjut	mohon	kesediaan
kamu	informasi	nomor	hp
Email	daftar	Via	Dm
Terima	kasih		

4. HASIL DAN PEMBAHASAN

3.4 Klasifikasi

Klasifikasi dilakukan dengan memasukkan data latih *positif* sejumlah 1.492 kata *positive* dan *negative* sejumlah 2.956 kata *negative* dengan menggunakan algoritma Naïve bayes.

Tabel 7. Hasil Klasifikasi Gopay

No	tweet	Positif	Negatif
0	suka nabung hasil enggak pernah liat dua mungk...	0.074074	0.259259
1	biaya tetap sama cukup bayar 1x bulan jumlah b...	0.125000	0.166667
2	mbah ceunahyang aja sampe ikut pake pas udah t...	0.034483	0.034483

Tweet yang diujikan dilakukan prediksi dari data latih yang telah diinput sehingga didapat hasil perhitungan per masing-masing tweet lebih dominan *positif* atau *negatif*. Pada Tabel 7 merupakan hasil klasifikasi gopay sentiment analisis yang telah dilakukan perhitungan sebelumnya

Tabel 8. Hasil Klasifikasi LinkAja

No	tweet	Positif	Negatif
0	bingung gimana cara habis livin poin mu mendin...	0.093750	0.156250
1	hai kak terima kasih hubung linkaja informasi ...	0.137931	0.137931
2	siapa mau hadiah kumpul sekarang khusus buat s...	0.161290	0.032258

Demikian pula pada Tabel 8 telah didapat hasil sentiment analisis per masing-masing tweet yang diujikan. Berdasarkan hasil dari proses klasifikasi memperoleh hasil akurasi sebagai berikut :

Tabel 9. Hasil Klasifikasi LinkAja

Data	Positif	Negatif	Akurasi
Gopay	79,6 %	20,4 %	87 %
LinkAja	62,1 %	37,9 %	89 %

5. KESIMPULAN

Berdasarkan hasil pengujian didapatkan bahwa tingkat sentiment positif dari twitter resmi Gopay memiliki kecenderungan positif yang lebih tinggi yaitu sebesar 79,6 % dibandingkan dengan akun twitter resmi dari LinkAja yaitu 62,1%. Sedangkan untuk sentiment negative untuk akun gopay yaitu sebesar 20, 4% dan LinkAja 37,9%. Hasil pengujian juga dilakukan proses perhitungan tingkat akurasi dengan recall dan precision yaitu sebesar 87% untuk Gopay dan 89% untuk LinkAja.

REFERENSI

[1] A. D. Riyanto, "Andi.Link," Hootsuite (Indonesian Digital Report), 2020. [Online]. Available: <https://andi.link/hootsuite-we-are-social-indonesian-digital-report-2020/>. [Accessed 2022].

[2] X. Liu, A. Nourbakhsh and Q. Li, "Reuters tracer: Toward automated news production using large scale social media data," in *IEEE International Conference on Big Data (Big Data)*, Boston, USA, 2017.

[3] Q. Li, Q. Zhang, L. Si and Y. Liu, "Rumor Detection on Social Media: Datasets, Methods and," *EMNLP*, 2019.

- [4] E. D. B. I. Ronny Julianto, "Analisis Sentimen Layanan Provider Telepon Seluler pada Twitter menggunakan Metode Naïve Bayesian Classification," *Journal of Big Data Analytic and Artificial Intelligence*, pp. 24-30, 2017.
- [5] X. N. J.-T. S. Sinno Jialin Pan, "Cross-domain sentiment classification via spectral feature alignment," in *International World Wide Web Conference Committee*, 2010.
- [6] A. Zunic, P. Corcoran and I. Spasic, "Sentiment Analysis in Health and Well-Being: Systematic Review," *JMIR Medical Informatics*, 2020.
- [7] W. Budiharto and M. Meiliana, "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis," *Journal of Big Data*, 2018.
- [8] Saura, J. R. Saura, P. P. Snchez and A. Grilo, "Detecting Indicators for Startup Business Success: Sentiment Analysis Using Text Data Mining," *Sustainability*, 2019.
- [9] H. Kaur, V. Mangat and Nidhi, "A survey of sentiment analysis techniques," in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* , Palladam, India, 2017.
- [10] F. a. D. D. A. Nur, "Pengembangan Aplikasi Sentiment Analysis Menggunakan Metode Naïve Bayes (Studi Kasus Sentiment Analysis dari media twitter)," *Seminar Nasional Sistem Informasi Indonesia, 2-3 November 2015* , vol. 2015, pp. 335-340, 2015.
- [11] M. V. S. P. M. & A. G. M. Jain, "Data extraction and sentimental analysis from "twitter" using web scrapping," *nternational Journal of Engineering and Advanced Technology*, 2019.
- [12] H. S. T. Lian Ardiani, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *Jurnal Sistem dan Teknologi Informasi*, Vols. Vol. 8, No. 2, April 2020 , no. DOI: 10.26418/justin.v8i2.36776, pp. 183-190, 2020.
- [13] "Leading countries based on number of Twitter users as of January 2022," January 2022. [Online]. Available: <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>.
- [14] Y. K. S. Derick Iskandar, "PERBANDINGAN AKURASI KLASIFIKASI TINGKAT KEMISKINAN ANTARA ALGORITMA C 4.5 DAN NAÏVE BAYES," *Jurnal Ilmiah NERO* , vol. Vol 2. No. 1, pp. 37-43, 2015.
- [15] Weare.social, "datareportal," datareportal, 11 FEBRUARY 2021. [Online]. Available: <https://datareportal.com/reports/digital-2021-indonesia>. [Accessed 30 May 2022].