

## **Text Mining Classification Opini Publik Terhadap Provider di Indonesia**

**Said Thaufik Rizaldi<sup>1</sup>, Abdal Al Khairi<sup>2</sup>, Mustakim<sup>3</sup>**

<sup>1,2,3</sup>Program Studi Sistem Informasi, Fakultas Sains dan Teknologi  
Universitas Islam Negeri Sultan Syarif Kasim Riau

<sup>1,2,3</sup>Puzzle Research Data Technology (Predatech), Fakultas Sains dan Teknologi  
Universitas Islam Negeri Sultan Syarif Kasim Riau  
Jl. HR. Soebrantas Km. 18 Panam Pekanbaru Riau

e-mail: <sup>1</sup>11753101376@students.uin-suska.ac.id, <sup>2</sup>abdalkhairi96@gmail.com,  
<sup>3</sup>mustakim@uin-suska.ac.id

### **Abstrak**

*Opini adalah central dari aktivitas manusia yang memiliki pengaruh pada perilaku manusia dimana pada setiap segala keputusan, maka seseorang akan mencari pendapat dari orang lain. Perkembangan sarana media sosial seperti Twitter digunakan seperti individu untuk mendukung proses pengambilan keputusan sehingga dilakukan teknik sentimen analisis menggunakan Text Mining Classification untuk dapat mengidentifikasi sentimen yang berasal dari opini masyarakat pengguna kartu provider yang merupakan jenis data yang tidak terstruktur. Berdasarkan hal tersebut, penelitian ini melakukan klasifikasi dengan menggunakan Text Mining dengan menggunakan algoritma Naïve Bayes Classifier (NBC) dengan menggunakan 10 percobaan pada K-Fold Cross Validation. Hasil penelitian ini menunjukkan bahwa K=4 menghasilkan nilai akurasi terbaik pada penelitian ini sebesar 68,85%.*

**Kata kunci:** Text Mining, Klasifikasi, Opini Publik, Naïve Bayes, Fold Cross Validation

### **Abstract**

*The opinion is the center of human activity that influences human behavior wherein every decision, and someone will seek the opinion of others. The development of social media facilities such as Twitter is used as an individual to support the decision-making process so that a sentiment analysis technique is carried out using Text Mining Classification to be able to identify sentiments that come from the public opinion of provider card users, which is a type of unstructured data. This study performs a classification using Text Mining using the Naïve Bayes Classifier (NBC) algorithm using ten experiments on K-Fold Cross Validation. The results of this study indicate that K = 4 produces the best accuracy value in this study of 68.85%.*

**Keywords:** Text Mining, Classification, Public Opinion, Naïve Bayes, Fold Cross Validation

### **1. Pendahuluan**

Media sosial merupakan sarana bagi penggunaannya untuk berinteraksi secara *online* dalam bentuk tulisan, gambar dan audio-visual melalui perangkat komputer [1]. Twitter merupakan sosial media yang cukup populer di dunia yang memungkinkan penggunaannya untuk membagikan pesan yang disebut dengan *tweet* melalui berbagai platform [2]. Jumlah *tweet* yang dikirimkan setiap harinya mencapai 350 juta lebih per hari. Hal ini, digunakan oleh organisasi untuk memantau pergerakan pesaing melalui opini penggunaannya melalui *tweet* yang dikirimkan terintegrasi dengan merek mereka [3]. Selain itu dari sisi pengguna, interaksi dan konten melalui saluran – saluran resmi provider mempengaruhi keputusan untuk membeli merek dan produk tertentu berdasarkan *user-generated content* oleh pengguna – pengguna yang lain [4].

Salah satu provider yang memiliki jumlah pengguna terbesar di Indonesia adalah Telkomsel dengan jumlah pelanggan aktif sebesar 170 juta lebih [5]. Salah satu sarana yang digunakan Telkomsel untuk melakukan pemasaran adalah dengan menggunakan Twitter. Jumlah pengikut dari akun Twitter resmi Telkomsel adalah 1,5 juta lebih pada akun @Telkomsel dimana *tweet* tersebut berisikan *review* terhadap produk-produknya dan

permasalahan lain seperti kualitas jaringan pada provider tersebut [6]. Untuk meningkatkan kualitas layanan telekomunikasi salah satunya melakukan analisis terhadap *review* pengguna untuk menganalisis sentimen dari sebuah data *text* yang dapat membantu dalam menentukan kebijakan untuk memberikan layanan yang lebih baik [2]. Dari potensi dan permasalahan tersebut dengan memanfaatkan metode analisis sentimen dapat dilakukan untuk mengetahui opini publik terhadap layanan provider tersebut.

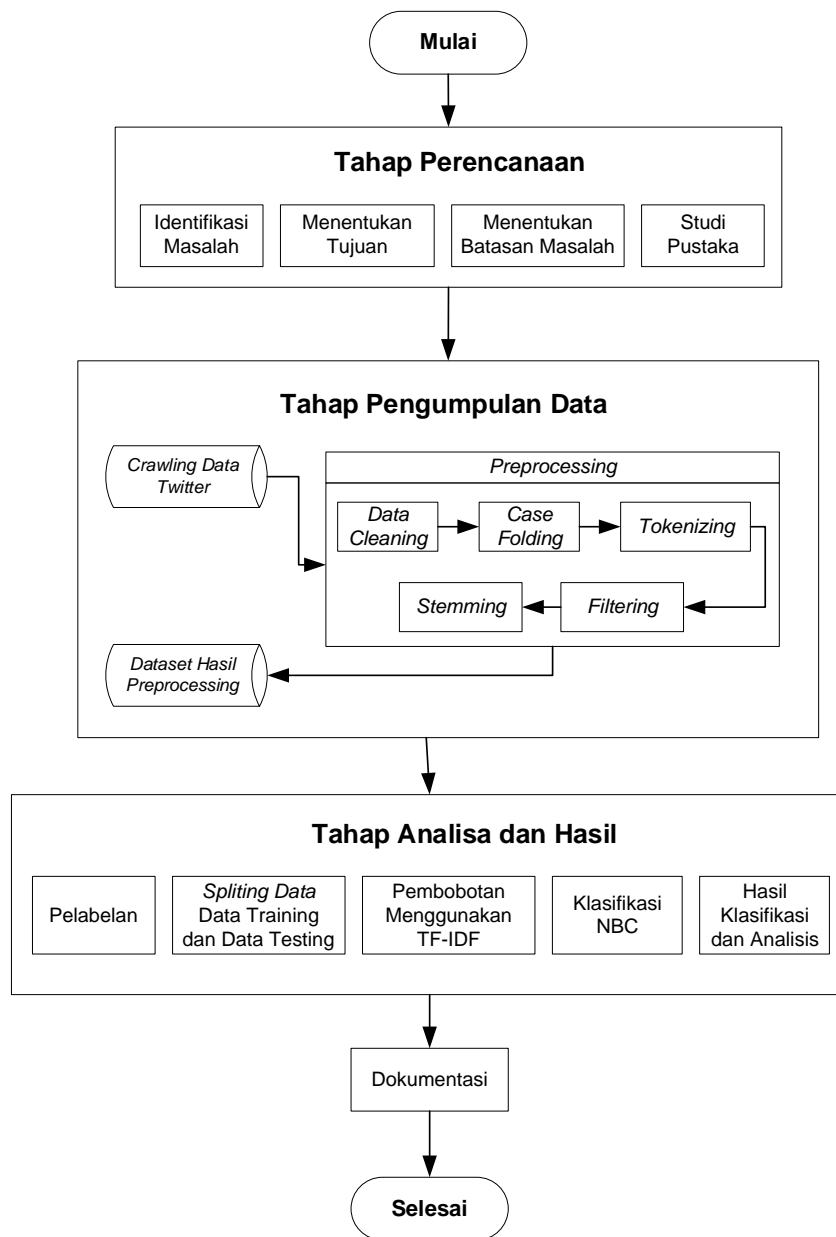
Berbagai jenis metode yang dilakukan untuk menganalisis sentimen yang disebut dengan *text mining* diantaranya adalah *classification* yang dapat dilakukan guna menggali *knowledge* dari teks [7]. *Text mining classification* adalah bidang ilmu komputer yang dapat melakukan pembelajaran secara mesin terhadap pendapat dan sikap dari pengguna terhadap layanan produk dari tulisan atau teks untuk menentukan sentimen penulis [8].

Penelitian terkait dengan *text mining classification* pernah dilakukan oleh Samah, Badarudin, Ibrahim, dan kawan-kawan pada tahun 2021 tentang Analisis Sentimen reputasi penyedia layanan komunikasi di Twitter dengan menggunakan algoritma Naïve Bayes Classifier. Penelitian ini menghasilkan akurasi yang cukup signifikan yakni sebesar 89% dari hasil prediksi model sebesar 129.7 ribu lebih label negatif dan 89,9 ribu lebih dengan label positif. Namun, pada penelitian ini masih terdapat label yang tidak dapat diprediksi kelasnya [4]. Selanjutnya, penelitian terkait *text mining* yang diterapkan untuk mengetahui Analisis Sentimen pada pengguna jasa ojek berbasis aplikasi *online* dengan menggunakan algoritma Naïve Bayes Classifier menghasilkan akurasi sebesar 80% lebih. Namun pada penelitian ini terdapat *missing* terhadap mekanisme uji fitur yang tidak sesuai. Pada penelitian ini memberikan rekomendasi dengan menambah *data training* untuk meningkatkan evaluasi pada hasil akurasi [9]. Sebagian besar penelitian lainnya menggunakan teknik *text mining classification* untuk melakukan Analisis Sentimen dengan menggunakan salah satunya menggunakan algoritma Naïve Bayes Classifier yang dinilai efektif [10] [11] [12].

Berdasarkan ulasan pendahuluan didukung oleh penelitian yang dilakukan oleh peneliti yang meneliti topik sejenis sebelumnya [4] [9] [10] [11] [12], penelitian ini akan menerapkan *text mining classification* terhadap opini publik terkait dengan penyediaan layanan kartu provider dengan menggunakan algoritma Naïve Bayes Classifier dimana penelitian ini menghasilkan analisis sentimen dari opini publik terkait dengan pengembangan layanan kartu provider.

## 2. Metode Penelitian

Pada Tahap Perencanaan dilakukan Identifikasi Masalah berupa fenomena yang terjadi pada layanan kartu provider sehingga menghasilkan opini publik terutama di sosial media. Adapun tujuan dari penelitian ini sesuai dengan identifikasi masalah yang dibangun yakni melakukan *text mining classification* dengan menggunakan algoritma Naïve Bayes Classifier berdasarkan penelitian terdahulu terhadap kajian Studi Pustaka. Tahapan yang selanjutnya yakni Tahap Pengumpulan Data. Tahap Pengumpulan Data yakni dilakukan *crawling data* dari platform Twitter yang berisikan data *text* terkait opini terhadap layanan kartu provider khususnya pada akun Twitter resmi @telkomsel pada tahun 2020. Setelah proses *crawling*, selanjutnya dilakukan proses *Data Cleaning*, *Case Folding*, *Tokenizing*, *Stemming* dan *Filtering* untuk mendapatkan data bersih untuk selanjutnya dilakukan tahapan Pelabelan, proses pelabelan data adalah dengan menggunakan metode yang pernah dilakukan oleh Salton dan kawan-kawan dengan melabelkan menjadi 3 kelas yakni positif, netral dan negatif [13]. Selanjutnya proses *Splitting Data*, Pembobotan TF IDF dan Klasifikasi menggunakan algoritma Naïve Bayes Classifier. Berikut adalah model *flowchart* Metodologi Penelitian pada Gambar 1.



Gambar 1. Metodologi Penelitian

### 2.1. Text Mining Classification

*Text Mining* adalah metode *Data Mining* yang dapat mengidentifikasi *knowledge* dari kumpulan teks untuk tujuan tertentu [14]. *Text Mining* bertujuan untuk memperoleh kata-kata pada dokumen yang merepresentasikan isi dari dokumen tersebut sehingga dapat menghasilkan sebuah pola yang berguna antar dokumen [15]. Adapun sebagian besar proses *Text Mining Classification* yakni proses *preprocessing* mengubah kalimat *unstructured* menjadi *structured* lalu proses klasifikasi yang dilakukan pelabelan yang dilakukan berdasarkan kata kunci yang ditetapkan berdasarkan kemiripan antar *text* [7] [15].

## 2.2. Text Preprocessing

*Text Preprocessing* adalah proses yang dilakukan sebelum data dapat diolah untuk dilakukan seleksi atau *filter* untuk mendapatkan kata yang penting namun tetap mempertahankan karakter dari subjek teks tersebut. Adapun selain mendapatkan kata penting pada subjek teks, *Text Preprocessing* efisien untuk mengurangi fitur yang berdimensi tinggi dan mengurangi *noise* sehingga akan mengefisiensi waktu pemrosesan data dan meningkatkan akurasi [16]. Diantara proses *Text Preprocessing* diantaranya lain *Case Folding*, *Tokenizing*, *Stemming* dan *Filtering* [17] [18].

## 2.3. Term Frequency-Inverse Document Frequency (TF-IDF)

*Term Frequency-Inverse Document Frequency* atau *TF-IDF* adalah metode yang mengkombinasikan antara dua teknik *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Metode TF didasarkan suatu suku dengan nilai frekuensi suku yang lebih tinggi dianggap lebih penting daripada suku yang nilai yang lebih rendah. Selanjutnya metode IDF adalah metode yang melengkapi TF karena memiliki kekurangan terhadap pengabaian frekuensi pengumpulan dimana mekanisme dari metode IDF adalah istilah yang muncul lebih sedikit dianggap lebih penting dari pada banyaknya istilah yang muncul [19]. Berikut adalah rumus dari TF-IDF pada Persamaan (1) sebagai berikut :

$$TF - IDF(t, d, D) = TF(t, d) \cdot IDF(t, d, D) \quad (1)$$

Dimana,  $TF - IDF(t, d, D)$  mewakili bobot *term*  $t$  dari dokumen  $d$  pada *corpus*  $D$ , sedangkan  $TF(t, d)$  mewakili nilai *TF* dari *term*  $t$  pada dokumen  $d$  [19]

## 2.4. Twitter

Twitter merupakan platform *micro-blogging* dengan jangkauan yang dapat menghubungkan banyak orang di dunia sehingga Twitter menjadi bagian kehidupan rutin digital secara individu. Twitter dengan terhubungan ke individu yang beragam dan partisipasi aktif pengguna menjadikannya dasar kualitatif dan kuantitatif untuk menganalisis sentimen, sehingga beberapa penelitian *Soft Computing* (SC) menjadikan Twitter sebagai studi kasus dalam menganalisis kompleksitas dari data Twitter [20].

## 2.5. Provider Seluler

Provider Seluler adalah layanan penyelenggara komunikasi yang mengangkut informasi secara elektronik, sebagai penyelenggara jasa telekomunikasi. Istilah ini mencakup perusahaan publik dan swasta dalam bisnis telekomunikasi, internet, kabel, satelit, dan layanan terkelola [21]. Adapun provider penyedia layanan penyelenggara komunikasi di Indonesia adalah salah satunya Telkomel (PT. Telkomsel Seluler) [22].

## 2.5. Opini Publik

Opini adalah *central* dari aktivitas manusia yang memiliki pengaruh pada perilaku manusia. Dimana pada setiap segala keputusan, maka seseorang akan mencari pendapat dari orang lain. Perkembangan sarana media sosial seperti Twitter digunakan seperti individu untuk mendukung proses pengambilan keputusan sehingga dilakukan teknik sentimen analisis. Analisis sentimen pada opini secara sederhana dapat didefinisikan sebagai sentimen positif atau negatif, sudut pandang, emosi atau apresiasi tentang sesuatu bagi seseorang [23].

## 2.6. Naïve Bayes Classifier (NBC)

Naïve Bayes Classifier (NBC) adalah algoritma yang digunakan untuk mendapatkan probabilitas tertinggi untuk mengklasifikasikan data uji pada kategori yang paling sesuai [17]. NBC adalah didasarkan pada teorema dasar yakni Teori Bayes yang dapat memprediksi data yang akurat [24]. NBC memiliki persamaan sebagai berikut pada Persamaan (2)

$$P(y_k|x_a) = \frac{P(y_k) \cdot P(x_a|y_k)}{P(x_a)} \quad (2)$$

Dimana, peluang munculnya  $x_a$  pada kelas kategori  $Y_k$   $P(y_a|x_k)$ , dikalikan dengan probabilitas kategori pada kelas  $P(y_k)$  yang akan menghasilkan distribusi dokumen  $P(x_a)$  [17].

### 2.6. K-Fold Cross Validation

K-Fold Cross Validation adalah metode validasi dimana menyilang (*cross*) data pada nilai  $k$  menjadi ukuran yang sama. Metode ini digunakan untuk menguji akurasi pada satu teknik klasifikasi atau yang bersifat prediksi. Berikut pada Persamaan (3) dinyatakan untuk menghitung kinerja klasifikasi

$$accuracy = \frac{\Sigma true\ classification}{\Sigma testing\ data} \times 100\% \quad (3)$$

Dimana, akurasi adalah jumlah klasifikasi yang benar dibagi jumlah daripada data testing dikali dengan seratus persen [25].

## 3. Hasil dan Pembahasan

### 3.1. Pengumpulan Data

Berdasarkan metodologi penelitian pada Gambar 1. Maka dilakukan pengumpulan data awal dengan metode *crawling* di Twitter, data yang akan diambil adalah berhubungan dengan kata kunci pada Kartu Provider pada penelitian ini yakni @Telkomsel sehingga total data yang dapat digunakan untuk dijadikan *dataset* penelitian adalah sebesar 1.500 data dari 3.000 data yang didapat dengan pembagian masing-masing sebesar 500 data untuk kelas positif, negatif dan netral. Adapun atribut yang akan diambil untuk penelitian berupa *tweets* pada *username* dengan data sebagai berikut pada Tabel 1. Data Awal *Crawling*.

Tabel 1. Data Awal *Crawling*

No.	Tweets
1	@primadaadam @Telkomsel Nitip ya kan Siapa tau ada yg mau beli pulsa Tsel yuk aku jual
2	@Telkomsel jaringan g stabil trus dbuat vidcall kerjaan ptah2 oke lah tp ini lngsung putus ntap emng
3	@Telkomsel Udah bisa min hehehe makasih responnya
4	Pantes, kepalaku sampai keluar tanduk karena jaringan super lemot. Padahal rumah, tetanggan sama kantor walikota
5	@lmsskim @Telkomsel Kuota aplikasi edukasinya dapat dari mana kak?
...	...
3000	@radenrauf @Telkomsel Mana sempat gebetannya keburu digebet orang :)

### 3.2. Cleaning Data

Pada proses *Cleaning Data* dilakukan proses seperti *case folding*, menghilangkan karakter yang mengandung huruf, simbol dan angka serta menghapus kata yang mengandung *username* pengguna maupun *username* kata kunci. Berikut adalah hasil *Cleaning Data* pada Tabel 2.

Tabel 2. Hasil *Cleaning*

No.	Tweets
1	nitip ya kan siapa tau ada yg mau beli pulsa tsel yuk aku jual
2	jaringan g stabil trus dbuat vidcall kerjaan ptah oke lah tp ini lngsung putus ntap emng
3	udah bisa min hehehe makasih responnya
4	pantes kepalaku sampai keluar tanduk karena jaringan super lemot padahal rumah tetanggan sama kantor walikota
5	kuota aplikasi edukasinya dapat dari mana kak
...	...
3000	mana sempat gebetannya keburu digebet orang

### 3.3. Pelabelan Data

Pada proses Pelabelan Data adalah memberikan label terhadap data Twitter yang tidak memiliki label menjadi dapat dipelajari atau memiliki label (*supervised learning*) [26]. Berikut adalah data yang telah diberi label pada Tabel 3. Hasil Pelabelan.

Tabel 3. Hasil Pelabelan

No.	Tweets	Sentimen
1	nitip ya kan siapa tau ada yg mau beli pulsa tsel yuk aku jual	Netral
2	jaringan g stabil trus dbuat vidcall kerja ptah oke lah tp ini lngsung putus ntap emng	Netral
3	udah bisa min hehehe makasih responnya	Netral
4	pantes kepalaku sampai keluar tanduk karena jaringan super lemot padahal rumah tetanggan sama kantor walikota	Negatif
5	kuota aplikasi edukasinya dapat dari mana kak	Netral
...	...	...
3000	mana sempat gebetannya keburu digebet orang	Netral

### 3.4. Text Preprocessing

Berdasarkan pada Gambar 1. Metodologi Penelitian. Preprocessing yang dilakukan pada penelitian ini diantaranya *tokenizing*, *stemming* dan *filtering*. Pada proses *filtering* dilakukan *stopword* yakni memilah dan menghilangkan kata hubung dengan menggunakan penelitian *stoplist* milik Tala tahun 2003 [27]. Berikut adalah hasil *Text Preprocessing* yang dilakukan pada Tabel 4.

Tabel 4. Hasil Preprocessing

No.	Tweets
1	nitip tau beli pulsa tsel yuk jual
2	jaring g stabil trus dbuat vidcall kerja ptah oke lngsung putus ntap emng
3	udah hehehe makasih respon
4	pantes kepala tanduk jaring super lot rumah tetanggan kantor walikota
5	kuota aplikasi edukasi
...	...
3000	gebetannya keburu digebet orang

### 3.5. Proses TF-IDF

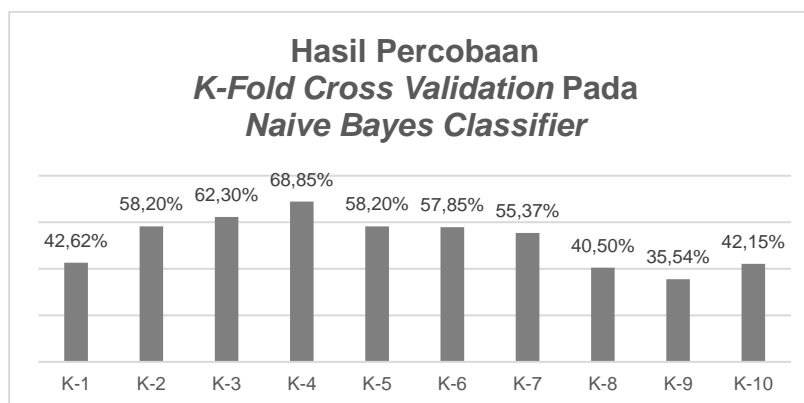
Setelah dilakukan proses *preprocessing* pada data penelitian, selanjutnya dilakukan proses TF-IDF dimana pada penelitian ini awalnya memiliki 3000 *record* data, setelah dilakukan proses ini tereduksi menjadi 1500 *record* data dimana menghilangkan *term* yang tidak memiliki nilai yang signifikan dalam mendukung penelitian. Berikut adalah hasil dari Proses TF-IDF pada Tabel 5.

Tabel 5. Proses TF-IDF

No	Term						
	aja	banget	bantu	beli	cek	...	youtube
1	0,000	0,000	0,461	0,000	0,000	...	0,000
2	0,000	0,000	0,000	0,000	0,000	...	0,000
3	0,000	0,000	0,000	0,000	0,000	...	0,000
4	0,000	0,000	0,000	0,000	0,000	...	0,000
5	1,000	0,000	0,000	0,000	0,000	...	0,000
...	...	...	...	...	...	...	...
1500	0,000	0,000	0,000	0,000	0,000	...	0,000

### 3.6. K-Fold Cross Validation pada Algoritma NBC

Setelah dilakukan proses TF-IDF, selanjutnya dilakukan klasifikasi dengan menggunakan algoritma NBC. Untuk menguji performa dari algoritma tersebut pada penelitian, dilakukan evaluasi terhadap algoritma tersebut dengan menggunakan K-Fold Cross Validation. Pada penelitian ini digunakan eksperimen parameter *k* sebanyak 1 hingga 10 untuk melakukan optimasi *k* yang terbaik pada penelitian ini. Berikut adalah Hasil Percobaan K-Fold Cross Validation pada NBC pada Gambar 2.



Gambar 2. Hasil Percobaan K-Fold Cross Validation

Berikut pada Gambar 2. Ditunjukkan Hasil Percobaan K-Fold Cross Validation pada NBC dimana menghasilkan percobaan  $k=4$  dengan nilai akurasi 68,85% sebagai percobaan dengan nilai akurasi yang tertinggi pada penelitian ini, sedangkan  $k=9$  menunjukkan perbedaan sebesar 33% lebih sebagai percobaan dengan nilai akurasi yang lebih terendah pada penelitian ini.

#### 4. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan bahwa penggunaan metode *Text Mining Classification* menggunakan algoritma Naïve Bayes Classifier (NBC) pada data Opini Publik Kartu Provider di Twitter menghasilkan nilai parameter  $k=4$  sebagai nilai akurasi yang terbaik pada penelitian ini yakni sebesar 68,85%. Sedangkan nilai parameter yang lain menghasilkan nilai 30% relatif lebih rendah yang dihasilkan pada penelitian ini. Saran terkait dengan penelitian ini adalah agar untuk penelitian kedepannya dilakukan dengan menggunakan lebih banyak data dan perbandingan terhadap beberapa algoritma pada klasifikasi sehingga menghasilkan akurasi yang lebih baik.

#### Daftar Pustaka

- [1] S. Siddiqui dan T. Singh, "Social Media its Impact with Positive and Negative Aspects," *Int. J. Comput. Appl. Technol. Res.*, vol. 5, no. 2, hal. 71–75, 2016, doi: 10.7753/ijcatr0502.1006.
- [2] A. Mustafa, S. A., dan S. Sohail, "Sentiment Classification of Twitter Data Belonging to Saudi Arabian Telecommunication Companies," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 1, 2017, doi: 10.14569/IJACSA.2017.080150.
- [3] H. Bagheri dan M. J. Islam, "Twitter Sentiment Analysis," *Int. J. Eng. Res. Technol.*, vol. 8, no. 03, hal. 1–2, 2017, doi: 10.31219/osf.io/6xc4y.
- [4] K. A. F. A. Samah, I. M. Badarudin, S. Ibrahim, N. A. Mocketar, dan L. S. Riza, "Naïve Bayes Twitter Sentiment Analysis In Visualizing The Reputation Of Communication Service Providers: During Covid-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 5, hal. 1753–1764, Apr 2021, doi: 10.17762/turcomat.v12i5.2176.
- [5] S. T. Wahyuni, "Effect of Service Quality and Switching Barriers on Customer Satisfaction and Its Impact on Customer Loyalty User Telkomsel for Online Learning During Pandemy," *Int. Conf. Bus. Soc. Sci.*, no. 1, hal. 1033–1042, 2020.
- [6] B. K. Viva, "Telkomsel Jadi Operator Seluler dengan Pelanggan Terbanyak di Indonesia," *katadata.co.id*, 2021. <https://databoks.katadata.co.id/datapublish/2021/09/22/telkomsel-jadi-operator-seluler-dengan-pelanggan-terbanyak-di-indonesia> (diakses Okt 04, 2021).
- [7] F. Fatimah, R. Agustiansyah, dan A. Musnansyah, "Analisis Sentimen Terhadap Pemilihan Gubernur Provinsi Sulawesi Selatan Tahun 2018 Menggunakan Data Twitter Dan Metode Sentiment Lexicon," *e-Proceeding Eng.*, vol. 6, no. 1, hal. 1928–1934, 2019.
- [8] M. Farhadloo dan E. Rolland, "Fundamentals of sentiment analysis and its applications," *Stud. Comput. Intell.*, vol. 639, hal. 1–24, 2016, doi: 10.1007/978-3-319-30319-2\_1.
- [9] D. G. Nugroho, Y. H. Chrisnanto, dan A. Wahana, "Analisis Sentimen Pada Jasa Ojek Online ... (Nugroho dkk.)," *Pros. SNST Ke - 7*, vol. 7, hal. 156–161, 2016.
- [10] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, dan W. Gata, "Sentiment Analysis on E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine," *J. Ilmu Komput. dan Inf.*, vol.

- 13, no. 2, hal. 109–122, 2020, doi: 10.21609/jiki.v13i2.885.
- [11] M. Sodanil, "Multi-Language Sentiment Analysis for Hotel Reviews," *MATEC Web Conf.*, vol. 75, hal. 5–8, 2016, doi: 10.1051/matecon/20167503002.
- [12] R. I. Permatasari, M. A. Fauzi, P. P. Adikara, dan E. D. L. Sari, "Twitter Sentiment Analysis of Movie Reviews using Ensemble Features Based Naïve Bayes," *3rd Int. Conf. Sustain. Inf. Eng. Technol. SIET 2018 - Proc.*, hal. 92–95, 2018, doi: 10.1109/SIET.2018.8693195.
- [13] G. Salton dan C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manag.*, vol. 24, no. 5, hal. 513–523, 1988.
- [14] A. Priyanto dan M. R. Ma'arif, "Implementasi Web Scrapping dan Text Mining untuk Akuisisi dan Kategorisasi Informasi dari Internet (Studi Kasus: Tutorial Hidroponik)," *Indones. J. Inf. Syst.*, vol. 1, no. 1, hal. 25–33, 2018, doi: 10.24002/ijis.v1i1.1664.
- [15] H. Andreas, N. Andreas, P. Gerhard, dan A. Fraunhofer, "A brief survey of text mining," in *LDV Forum-GLDV Journal for Computational Linguistics and Language Technology*, 2005, vol. 20, no. 1, hal. 19–62.
- [16] W. Chen, Z. Xu, X. Zheng, Q. Yu, dan Y. Luo, "Research on sentiment classification of online travel review text," *Appl. Sci.*, vol. 10, no. 15, 2020, doi: 10.3390/AP10155275.
- [17] R. Novendri, A. S. Callista, D. N. Pratama, dan C. E. Puspita, "Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes," *Bull. Comput. Sci. Electr. Eng.*, vol. 1, no. 1, hal. 26–32, 2020, doi: 10.25008/bcsee.v1i1.5.
- [18] M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, dan H. A. Gozali, "Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, 2020, doi: 10.1088/1757-899X/874/1/012017.
- [19] Z. Jiang, B. Gao, Y. He, Y. Han, P. Doyle, dan Q. Zhu, "Text Classification Using Novel Term Weighting Scheme-Based Improved TF-IDF for Internet Media Reports," *Math. Probl. Eng.*, vol. 2021, no. ii, 2021, doi: 10.1155/2021/6619088.
- [20] A. Kumar dan A. Jaiswal, "Systematic literature review of sentiment analysis on Twitter using soft computing techniques," *Concurr. Comput. Pract. Exp.*, vol. 32, no. 1, hal. 1–29, 2020, doi: 10.1002/cpe.5107.
- [21] K. A. F. A. Samah, A. A. S. A. Sani, N. Sabri, S. Ibrahim, A. F. A. Fadzil, dan L. S. Riza, "Visualizing communication of service providers reputation during COVID-19 pandemic: A conceptual model," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1.4 Special Issue, hal. 558–568, 2020, doi: 10.30534/ijatcse/2020/7891.42020.
- [22] A. R. Susanti, T. Djatna, dan W. A. Kusuma, "Twitter's sentiment analysis on GSM services using Multinomial Naïve Bayes," *Telkomnika (Telecommunication Comput. Electron. Control.)*, vol. 15, no. 3, hal. 1354–1361, 2017, doi: 10.12928/TELKOMNIKA.v15i3.4284.
- [23] B. Keith Norambuena, E. F. Lettura, dan C. M. Villegas, "Sentiment analysis and opinion mining applied to scientific paper reviews," *Intell. Data Anal.*, vol. 23, no. 1, hal. 191–214, 2019, doi: 10.3233/IDA-173807.
- [24] Mustakim *et al.*, "Data Sharing Technique Modeling for Naive Bayes Classifier for Eligibility Classification of Recipient Students in the Smart Indonesia Program," *J. Phys. Conf. Ser.*, vol. 1424, no. 1, 2019, doi: 10.1088/1742-6596/1424/1/012009.
- [25] F. Tempola, M. Muhammad, dan A. Khairan, "Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, hal. 577, 2018, doi: 10.25126/jtiik.201855983.
- [26] S. A. El Rahman, F. A. Alotaibi, dan W. A. Alshehri, "Sentiment Analysis of Twitter Data," *2019 Int. Conf. Comput. Inf. Sci. ICCIS 2019*, 2019, doi: 10.1109/ICCISci.2019.8716464.
- [27] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," *M.Sc. Thesis, Append. D*, vol. pp, hal. 39–46, 2003.