

KOMPARASI METODE FEATURE SELECTION TEXT MINING PADA PERMASALAHAN KLASIFIKASI KELUHAN PELANGGAN INDUSTRI TELEKOMUNIKASI MENGGUNAKAN SMOTE DAN NAÏVE BAYES

COMPARISON OF TEXT MINING FEATURE SELECTION METHODS IN THE CLASSIFICATION OF CUSTOMER COMPLAINTS IN THE TELECOMMUNICATION INDUSTRY USING SMOTE AND NAÏVE BAYES

Siti Fauziah¹, Dedi Dwi Saputra², Risca Lusiana Pratiwi³, Mochammad Rizky Kusumayudha⁴

^{1,2,3,4}Fakultas Teknologi Informasi, ^{1 & 3}Sistem Informasi, ²Ilmu Komputer, Informatika⁴

Universitas Nusa Mandiri dan Universitas Siber Indonesia

Email: siti.suz@nusamandiri.ac.id

Abstrak

Twitter merupakan salah satu media sosial yang populer di Indonesia. Tidak hanya masyarakat biasa yang menggunakan media sosial ini, banyak juga perusahaan yang memanfaatkan Twitter sebagai sarana promosi dan pendekatan kepada para pelanggannya. Termasuk Tri Indonesia yang mempunyai official akun di Twitter, salah satunya adalah @3CareIndonesia. Para pengguna Tri Indonesia memanfaatkan akun official ini sebagai sarana berbagi opini tentang pengalaman menggunakan provider ini dalam bentuk komentar di Twitter @3CareIndonesia. Penelitian ini dilakukan untuk mengukur seberapa akurat algoritma SMOTE dan Naïve Bayes dalam mengetahui sentimen pada komentar para pelanggan Tri Indonesia lalu mengkategorikannya kedalam komplain dan bukan komplain dengan menggunakan algoritma Naïve Bayes. Proses pengambilan data menggunakan metode crawling data menggunakan aplikasi Rapidminer. Kemudian dari data yang didapat dilakukan preprocessing tingkat 1 menggunakan Gataframework dan preprocessing tingkat 2 menggunakan Rapidminer. Hasil dari penelitian ini menunjukkan bahwa kombinasi dari SMOTE dan Naïve Bayes dapat menghasilkan pemodelan yang cukup baik dengan nilai accuracy sebesar 81.85%, precision sebesar 80.44%, recall sebesar 83.10% dan AUC sebesar 0.817.

Kata Kunci: Rapidminer, Naïve Bayes, Gataframework, Twitter, Sentiment Analysis

Abstract

Twitter is one of the popular social media platforms in Indonesia. It is not only used by ordinary people, but many companies also utilize Twitter to promote and

reach out to their customers. One of these companies is Tri Indonesia, which has an official Twitter account @3CareIndonesia. Users of Tri Indonesia utilize this official account to share their opinions and experiences using this service by commenting on Twitter @3CareIndonesia. This research was conducted to measure the accuracy of the SMOTE and Naïve Bayes algorithms in determining the sentiment of comments from Tri Indonesia customers and categorizing them into complaints and non-complaints using the Naïve Bayes algorithm. The data was collected through data crawling using the Rapidminer application. Then, the obtained data underwent level 1 preprocessing using the Gataframework and level 2 preprocessing using Rapidminer. This research indicates that the combination of SMOTE and Naïve Bayes can generate a fairly good model with an accuracy of 81.85%, precision of 80.44%, recall of 83.10%, and an AUC of 0.817.

Keywords: Rapidminer, Naïve Bayes, Gataframework, Crawling Data, Sentiment Analysis

PENDAHULUAN

Media sosial adalah sebuah sarana yang dimanfaatkan oleh masyarakat Indoensia untuk mengakses, memberi dan tempat diskusi seputar kritik dan saran. Twitter salah satu media social yang sering dikunjungi untuk mendiskusikan dengan cara memberi saran dan kritik [1]. Twitter merupakan platform media social berupa alat komunikasi online yang banyak digunakan oleh masyarakat Indonesia, ada sekitar 24 juta pengguna Twitter di Indonesia pada awal 2023 [2]. Selain itu banyak juga perusahaan yang memanfaatkan twitter sebagai media promosi dan interaksi secara langsung kepada para pengguna nya [3]. Tweet merupakan istilah dari sebuah tulisan pengguna Twitter yang diunggah pada akunnya [4]. Fitur tweet membuat para pengguna twitter dapat memberikan opini, dan berdiskusi di twitter. Tweet merupakan sebuah teks atau informasi

yang diberikan oleh pengguna dengan cara memposting, yang mempunyai batasan 149 karakter. Menggunakan sebuah fitur tweet ini pengguna twitter dapat berbagi opini dan ulasan yang berisikan berbagai macam ungkapan pengguna sebagai pola masa yang akan datang, yang dimana kita para pengguna twitter dapat membuah hasil jajakn pendapat [5]. Komentar adalah serangkaian penuturan kata yang berupa Informasi, opini, atau laporan dari pengamatan mengenai suatu objek, termasuk ulasan yang mungkin mengandung pandangan baik atau pandangan negatif, dan digunakan untuk melakukan evaluasi [6]. Seperti komentar yang telah di lakukan oleh pengguna Tri Indonesia terhadap layanan 3CareIndonesia yang berisikian berbagai macam text pertanyaan, kritik, dan pujian terhadap Tri

Indonesia. Pengguna melakukan tweet kepada @3CareIndonesia melalui platform Twitter dengan berbagai macam bahasa yang telah diutarakan. Komentar yang telah diberikan oleh pengguna atau pelanggan Tri Indonesia pada saat ini banyak sekali pemberian komentar yang negatif merujuk kepada keluhan pengguna mengenai terjadinya kelambatan pada jaringan Tri Indonesia, hilangnya sinyal secara tiba-tiba, dan gangguan kepada sim-card Tri Indonesia. Oleh karena itu sangat cocok untuk dijadikan sebuah penelitian yang berkaitan dengan pendapat berupa kritik dan saran untuk layanan @3CareIndonesia tersedia di platform twitter untuk di jadikan sebuah data yang akan dikelola, untuk dilakukan analisis sentiment dengan teks yang berupa interaksi antara pengguna dengan layanan @3CareIndonesia.

Analysis sentiment atau opinion mining adalah proses untuk mengidentifikasi dan mengenali atau mengkategorikan sebuah teks yang berupa emosi pengguna untuk layanan apapun seperti film, masalah produk, acara, atau attribute apapun yang mengandung opini positif, negatif, atau netral [7]. Analisis sentiment dapat digunakan untuk menentukan apakah sentiment yang diekspresikan seseorang dalam teks bermakna positif, negatif, atau netral, menggunakan sebuah algoritma tertentu. Dalam penelitian ini menggunakan algoritma SMOTE dan Naïve Bayes untuk metode penelitian, dan teknik Machine Learning sehingga

dapat mengolah data berupa teks dan melakukan sentiment analysis terhadap komentar yang ada pada akun Twitter @3CareIndonesia menggunakan metode text mining dengan algoritma SMOTE dan Naïve Bayes. Metode dan algoritma ini kami ingin membagi data yang didapatkan melalui platform Twitter @3CareIndonesia ke dalam dua kategori yaitu, complain dan not complain. Sumber untuk pengambilan data berupa kritik atau saran dari tweets para pengguna Twitter di akun official layanan Tri Indonesia yaitu, @3CareIndonesia yang pengambilan datanya menggunakan Twitter API dengan bantuan tools Rapidminer.

Penelitian ini mempunyai beberapa referensi dari Studi sebelumnya termasuk penelitian yang telah dilakukan oleh Tanthy T. Widowati dan Mujiono Sadikin menunjukkan bahwa algoritma Naïve Bayes dapat digunakan untuk analisis sentimen dengan 2 kategori dan mendapatkan hasil yang baik dengan akurasi 91.48%, presisi 89.28%, recall 91.58% [8]. Lalu, Dedi D. Saputra dan team, dalam penelitiannya bahwa penambahan algoritma SMOTE dapat mengoptimasi algoritma Naïve Bayes sehingga menaikkan akurasi yang di dapat dalam penelitian ini menjadi 78.63% dari sebelumnya 57.06%, sebab algoritma SMOTE dapat menangani masalah pada

keseimbangan dataset [9]. Juga penelitian yang dilakukan oleh Budi K. Rahmat dan team, dalam penelitiannya yang dilakukan bahwa kombinasi algoritma Naïve Bayes dan SMOTE dapat digunakan untuk meneliti sentimen terhadap suatu produk berdasarkan opini atau review pengguna produk di twitter @Telkomsel dengan total 1000 dataset, dan mendapatkan hasil yang baik [10]. Penelitian ini ditujukan untuk mengetahui seberapa akurat penggunaan algoritma Naïve Bayes dan SMOTE dalam menangani kasus analisis sentimen terhadap data yang bersumber dari Twitter @3CareIndonesia dengan total 2000 dataset dan mengkategorikannya kedalam komplain dan bukan komplain.

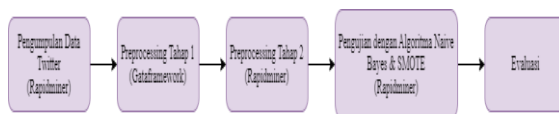
LANDASAN TEORI

Text Mining

Text Mining adalah proses ekstraksi informasi atau pengetahuan dari teks atau dokumen yang tidak terstruktur [11]. Hal ini sering digunakan dalam analisis sentimen, pengelompokan, atau klasifikasi teks.

Feature Selection

Feature Selection adalah proses yang melibatkan penghapusan kata-kata yang dianggap tidak memiliki relevansi [11]. Tindakan seleksi fitur dapat



diinterpretasikan sebagai upaya untuk

mengurangi kompleksitas dimensi dalam sebuah koleksi teks dengan tujuan untuk meningkatkan akurasi dan efektivitas dalam proses klasifikasi.

Klasifikasi

Klasifikasi adalah tugas pemodelan di mana entitas atau data diklasifikasikan ke dalam kategori atau label tertentu berdasarkan fitur-fitur yang ada [12]. Dalam penelitian ini, klasifikasi digunakan untuk mengkategorikan keluhan pelanggan ke dalam kategori yang sesuai.

SMOTE (*Synthetic Minority Over-sampling Technique*)

SMOTE adalah metode oversampling yang digunakan untuk menangani ketidakseimbangan kelas dalam masalah klasifikasi [13]. Ini menghasilkan sampel sintesis untuk kelas minoritas agar distribusi kelas lebih seimbang.

Naïve Bayes

Teori Naïve Bayes menunjukkan kemampuan klasifikasi yang mirip dengan Decision Trees dan Neural Networks [14]. Faktanya, algoritma Naïve Bayes menunjukkan tingkat akurasi dan kecepatan yang tinggi ketika diterapkan pada basis data yang besar [15].

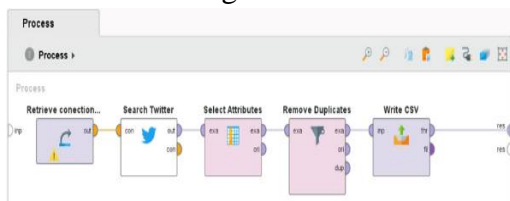
METODE PENELITIAN

Gambar 1. Tahapan Penelitian

(Fauziah, 2023)

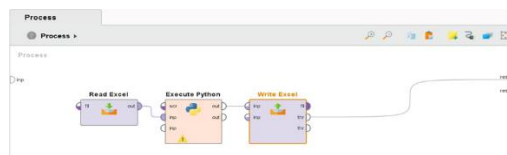
Penjelasan Tahapan yang dilakukan pada penelitian ini sebagai berikut:

1. Pengumpulan Data Twitter menggunakan RapidMiner
Pengumpulan data pada penelitian ini dilakukan dengan metode crawling data menggunakan aplikasi RapidMiner. Untuk mendapatkan akses data Twitter, terlebih dahulu dilakukan login menggunakan akun Twitter yang dimiliki dan melakukan request API pada Twitter untuk mendapatkan kode aksesnya. Berikut adalah tools yang peneliti gunakan dalam aplikasi RapidMiner untuk crawling data



Gambar 2. Proses Crawling Data Menggunakan Aplikasi RapidMiner (Fauziah, 2023)

2. Preprocessing Tahap 1 Menggunakan RapidMiner dan Gataframework
Preprocessing tahap 1 dilakukan dengan aplikasi RapidMiner dan juga Web Service dari Gataframework dan menggunakan bahasa pemrograman Python. Berikut tools yang peneliti gunakan dalam aplikasi RapidMiner



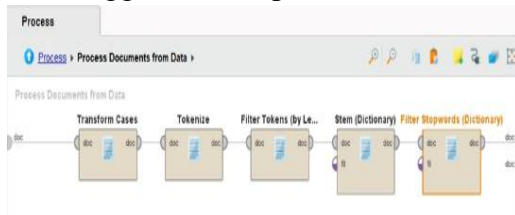
Gambar 3. Preprocessing Gataframework (Fauziah, 2023)

Preprocessing yang dilakukan pada Gataframework sebagai berikut:

- a. @Anotation Removal: Proses pada tahap ini adalah menghilangkan simbol @ dan juga username pada dataset.
- b. Transformation: Remove URL: Proses pada tahap ini adalah menghilangkan URL yang terdapat pada dataset.
- c. # (Hash Tag) Removal: Proses pada tahap ini adalah menghilangkan # (Hashtag) yang ada pada dataset
- d. Remove Tokenize Regex: Proses pada tahap ini adalah menghilangkan tanda baca, simbol dan karakter khusus yang terdapat pada dataset.
- e. Indonesian Stemming: Proses pada tahap ini adalah menghilangkan kata kata ber imbuhan yang ada pada data set dan mengubahnya menjadi kata dasar sesuai Bahasa Indonesia yang ada pada database Gataframework.
- f. Indonesian Stopword: Proses pada tahap ini adalah menghilangkan kata kata sambung dalam Bahasa Indonesia atau kata kata yang memerlukan kata lain agar memiliki makna

seperti dari, ada, di. Kata yang di hilangkan sesuai dengan kamus yang ada dalam database Gataframework.

3. Preprocessing Tahap 2 Menggunakan RapidMiner



Gambar 4. Preprocessing Tahap 2 (Fauziah, 2023)

Tahapan ini adalah preprocessing data terakhir sebelum data dilakukan pengujian dengan algoritma Naïve Bayes dan SMOTE. Berikut tools yang digunakan dalam aplikasi RapidMiner untuk preprocessing tahap 2:

Penjelasan tools yang digunakan sebagai berikut:

- Transform Cases:** Tahapan ini mengubah huruf kapital yang ada pada dataset menjadi huruf kecil.
- Tokenize:** Pada tahapan ini dilakukan pemecahan satu kalimat menjadi satuan kata. Kata yang sudah terpisah ini biasa disebut dengan token.
- Filter Tokens (By Length):** Pada tahapan ini dilakukan pemilihan token (kata) yang akan digunakan berdasarkan panjang kata tersebut. Dalam penelitian ini kata (token) yang akan digunakan adalah kata yang memiliki panjang 3 – 24 huruf.
- Stem (Dictionary):** Pada tahapan ini dilakukan pelengkapan stemming yang sudah dilakukan pada tahap 1

dengan Gataframework. Melakukan pembenaran ejaan untuk kata-kata yang disingkat pada dataset.

- Filter Stopwords (Dictionary):** Pada tahapan ini dilakukan pelengkapan yang sudah dilakukan pada tahap 1 dengan Gataframework. Menambahkan stopwords yang tidak digunakan untuk penilaian sentimen untuk kategori komplain dan bukan komplain seperti : nama, agama, daerah, dst.

4. Pengujian dengan Algoritma Naïve Bayes dan SMOTE

Pada tahap ini dilakukan pengujian dataset yang telah di preprocessing dengan kombinasi algoritma Naïve Bayes dan SMOTE untuk mendapatkan hasil kualifikasi yang terbaik yang akan digunakan untuk proses analisis sentimen dari dataset.

5. Evaluasi

Pada tahapan ini dilakukan perbandingan dari hasil hasil yang didapatkan selama tahap tahap pengujian yang dilakukan. Akan ditampilkan hasil dari perbandingan tersebut dengan tabel yang berisis accuracy, precision, recall dan AUC yang didapat dari setiap kombinasi pengujian.

HASIL DAN PEMBAHASAN

Hasil dari penelitian yang dilakukan berdasarkan metode penelitian yang sudah dijelaskan sebelumnya sebagai berikut:

1. Pengumpulan Data

Data yang dipakai dalam penelitian ini adalah data Tweet yang berada di akun twitter @3CareIndonesia. Data yang berasal dari komentar pada akun twitter @3CareIndonesia, dan sampel data yang digunakan adalah 2000 data komentar yang ada sampai 27 Maret 2023. Lalu dilakukan pelabelan untuk data yang didapat dengan sentimen komplain dan bukan komplain. Berikut adalah hasil yang didapatkan:



Gambar 5. Hasil dari crawling data dan pelabelan (Fauziah, 2023)

2. Preprocessing Tahap 1



Gambar 6. Preprocessing tahap 1 menggunakan Web Service Gataframework (Fauziah, 2023)

Hasil dari preprocessing tahap 1 menggunakan Web Service Gataframework sebagai berikut:

a. @Annotation Removal

Tabel 1. Hasil Proses Annotation Removal dalam Gataframework

Data teks	@Annotation Removal
@3CareIndonesia iyaa saya sudah cek iyaa saya sudah cek kak, dan daerah saya kak, dan daerah saya saya termasuk daerah yg saya termasuk daerah yg daerah yg bisa menggunakan menggunakan kuota lokal, tp knpa kuota lokal, tp knpa ttep gabisa dipkee itu ttep gabisa dipkee itu yg saya pertanyakan yg saya pertanyakan https://t.co/tf9wo0ya https://t.co/Tf9wO 0YAkK #3CareIndonesia	iyaa saya sudah cek kak, dan daerah saya termasuk daerah yg bisa menggunakan kuota lokal, tp knpa ttep gabisa dipkee itu yg saya pertanyakan https://t.co/tf9wo0ya kk #3CareIndonesia

Sumber: Fauziah, 2023

b. Transformation: Remove URL

Tabel 2. Hasil Proses Transformation: Remove URL dalam Gataframework

@Anotation Removal	Transformation: Remove URL
iyaa saya sudah cek kak, dan daerah saya saya termasuk daerah yg daerah yg bisa menggunakan menggunakan kuota lokal, tp knpa ttep gabisa dipkee itu ttep gabisa dipkee itu yg saya pertanyakan yg saya pertanyakan #3CareIndonesia	iyaa saya sudah cek kak, dan daerah saya termasuk daerah yg bisa menggunakan kuota lokal, tp knpa ttep gabisa dipkee itu yg saya pertanyakan #3CareIndonesia

<https://t.co/tf9wo0yakk>
#3CareIndonesia

Sumber: Fauziah, 2023

c. Hasil Proses Hash Tag removal menggunakan Gataframwork Tabel 3. Transformation: Remove URL dalam Gataframework

<i>Transformation: Remove URL</i>	<i># (Hash Tag) Removal</i>
iyaa saya sudah cek kak, dan daerah saya termasuk daerah yg bisa menggunakan kuota lokal, tp knpa ttep gabisa dipkee itu yg saya pertanyakan	iyaa saya sudah cek kak, dan daerah saya termasuk daerah yg bisa menggunakan kuota lokal, tp knpa ttep gabisa dipkee yg saya pertanyakan

Sumber: Fauziah, 2023

d. Hasil Proses Hash Tag removal menggunakan Gataframwork

Tabel 4. Hasil Proses Remove Tokenize Regex menggunakan Gataframework

<i># (Hash Tag) Removal</i>	<i>Remove Tokenize Regex</i>
iyaa saya sudah cek kak, dan daerah saya termasuk daerah yg bisa menggunakan kuota lokal, tp knpa ttep gabisa dipkee itu yg saya pertanyakan	iyaa saya sudah cek kak dan daerah saya termasuk daerah yg bisa menggunakan kuota lokal tp knpa ttep gabisa dipkee itu yg saya pertanyakan

Sumber: Fauziah, 2023

e. Indonesian Stemming

Tabel 5. Hasil Proses Indonesian Stemming menggunakan Gataframework

<i>Remove Tokenize Regex</i>	<i>Indonesian Stemming</i>
iyaa saya sudah cek kak dan daerah saya termasuk daerah yg bisa menggunakan kuota lokal tp knpa ttep gabisa dipkee itu yg saya pertanyakan	iyaa saya sudah cek kak dan daerah saya masuk daerah yg bisa guna kuota lokal tp knpa ttep gabisa dipkee itu yg saya tanya

Sumber: Fauziah, 2023

f. Indonesian Stopwords Removal

Tabel 6. Hasil Proses Indonesian Stopwords Removal menggunakan Gataframework

<i>Indonesia Stemming</i>	<i>Indonesian Stopwords Removal</i>
iyaa saya sudah cek kak dan daerah saya masuk daerah yg bisa guna kuota lokal tp knpa ttep gabisa dipkee itu yg saya tanya	iyaa cek daerah masuk daerah kuota lokal ttep gabisa dipkee

Sumber: Fauziah, 2023

3. Indonesian Stopwords Removal

Hasil dari preprocessing tahap 2 melanjutkan preprocessing tahap 1 menggunakan aplikasi RapidMiner dan menambahkan stemming dan stopword yang dibuat oleh peneliti, sebagai berikut:

a. Trasform Cases

Tabel 7. Hasil Proses Transform Cases menggunakan RapidMiner

<i>Indonesia Stemming</i>	<i>Indonesian Stopwords Removal</i>
---------------------------	-------------------------------------

iyaa saya sudah cek kak dan daerah saya masuk daerah yg bisa guna kuota lokal tp knpa ttep gabisa dipkee itu yg saya tanya	iyaa cek daerah masuk daerah kuota lokal ttep gabisa dipkee
--	---

Sumber: Fauziah, 2023

b. Tokenize

Tabel 8. Hasil Proses Tokenize menggunakan RapidMiner

<i>Trasform Cases</i>	<i>Tokenize</i>
iyaa cek daerah masuk daerah lokal ttep gabisa dipkee	iyaa cek daerah masuk daerah kuota lokal ttep gabisa dipkee

Sumber: Fauziah, 2023

c. Filter Token (By Length)

Tabel 9. Hasil Filter Token (By Length) menggunakan RapidMiner

<i>Tokenize</i>	<i>Filter Token (By Length)</i>
iyaa cek daerah masuk daerah kuota lokal ttep gabisa dipkee	iyaa cek daerah masuk daerah kuota lokal ttep gabisa dipkee

Sumber: Fauziah, 2023

d. Stemming (Menggunakan kamus yang dibuat oleh peneliti)

Tabel 10. Hasil dari Stemming (dengan kamus yang dibuat peneliti) menggunakan RapidMiner

<i>Filter Token (By Length)</i>	<i>Stemming</i>
iyaa cek daerah masuk daerah kuota lokal ttep gabisa dipkee	iya cek daerah masuk kuota lokal tetap gabisa pakai

Sumber: Fauziah, 2023

e. Stopwords (Menggunakan kamus yang dibuat peneliti)

Tabel 11. Hasil dari Stopword (dengan kamus yang dibuat peneliti) menggunakan RapidMiner

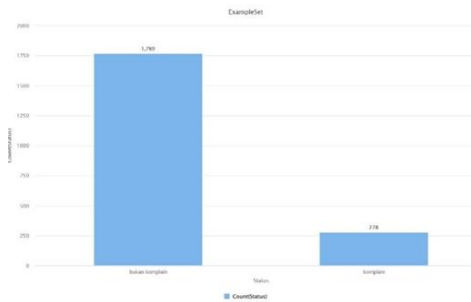
<i>Stemming</i>	<i>Stopword</i>
iyaa cek daerah masuk daerah kuota lokal ttep gabisa dipkee	cek masuk daerah kuota lokal tetap gabisa pakai

Sumber: Fauziah, 2023

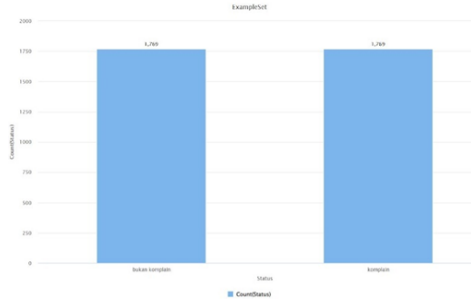
4. Penggunaan metode SMOTE

Penggunaan metode SMOTE dapat mengatasi masalah keseimbangan data agar klasifikasi yang dihasilkan dapat lebih baik. Seperti dataset yang digunakan dalam penelitian ini yang memiliki ketidak seimbangan data antara komplain dan bukan komplain. Perbedaan yang terlihat setelah

penggunaan metode SMOTE sebagai berikut:



Gambar 6. Dataset sebelum menggunakan metode SMOTE (Fauziah, 2023)



Gambar 7. Dataset setelah menggunakan metode SMOTE (Fauziah, 2023)

5. Evaluasi

Setelah semua tahapan dilakukan sebelumnya, tahap ini dilakukan perhitungan hasil kualifikasi dari semua kombinasi yang dilakukan oleh peneliti. Hasil dari evaluasi akan digambarkan dengan tabel sebagai berikut:

Tabel 12. Hasil Evaluasi

Kombinasi Metode	Algoritma	Hasil
<i>Preprocessing</i> Tahap 1 + <i>Preprocessing</i> tahap 2 (tanpa <i>Stemming</i> dan <i>Stopword</i>)	<i>Naïve Bayes</i>	<i>Accuracy</i> : 57.06% <i>Precision</i> : 16.41% <i>Recall</i> : 53.09% <i>AUC</i> : 0.475

<i>Preprocessing</i> tahap 1 + <i>Preprocessing</i> tahap 2 (tanpa <i>Stemming</i> dan <i>Stopword</i>)	<i>Naïve Bayes</i> + <i>SMOTE</i>	<i>Accuracy</i> : 78.63% <i>Precision</i> : 70.15% <i>Recall</i> : 99.77% <i>AUC</i> : 0.555
<i>Preprocessing</i> tahap 1 + <i>Preprocessing</i> tahap 2 (hanya menggunakan <i>Stemming</i> tanpa <i>Stopword</i>)	<i>Naïve Bayes</i> + <i>SMOTE</i>	<i>Accuracy</i> : 78.31% <i>Precision</i> : 70.83% <i>Recall</i> : 99.32% <i>AUC</i> : 0.589
<i>Preprocessing</i> tahap 1 + <i>Preprocessing</i> tahap 2 (menggunakan <i>Stemming</i> dan <i>Stopword</i>)	<i>Naïve Bayes</i> + <i>SMOTE</i>	<i>Accuracy</i> : 81.85% <i>Precision</i> : 81.25% <i>Recall</i> : 82.92% <i>AUC</i> : 0.817

Sumber: Fauziah, 2023

KESIMPULAN

Kesimpulan yang didapatkan dari penelitian ini adalah penggunaan Gataframework dalam melakukan preprocessing sangat membantu untuk mengurangi noise pada dataset dan melakukan pembersihan data sebelum diproses ke tahap berikutnya. Selain itu penambahan algoritma SMOTE juga dapat membantu dalam permasalahan keseimbangan dataset seperti pada penelitian ini. Dengan menambahkan algoritma SMOTE, hasil dari kualifikasi yang didapat menjadi lebih baik, terbukti dengan kenaikan accuracy menjadi 78.63%

dari sebelum menggunakan SMOTE hanya 57.06%. Kenaikan tinggi accuracy, precision, recall dan AUC juga terlihat saat penambahan Stopword yang dibuat oleh peneliti, karena dengan membuang kata kata yang tidak bersangkutan di dalam dataset untuk penelitian ini dapat meningkatkan hasil dari kualifikasi. Dan yang harus diperhatikan saat pelabelan dataset yang akan digunakan, jika dataset yang dilabeli dengan tidak benar akan menurunkan hasil dari proses kualifikasi. Hasil paling baik yang didapatkan dalam penelitian ini adalah penggabungan antara preprocessing tahap 1, preprocessing tahap 2, dan kombinasi dari SMOTE dan Naïve Bayes dengan accuracy sebesar 81.85%, precision 82.25%, recall 82.92% dan AUC 0.817. Dan dapat disimpulkan bahwa penggunaan kombinasi algoritma Naïve Bayes dan SMOTE cukup akurat dalam melakukan klasifikasi terhadap dataset yang digunakan pada penelitian ini.

DAFTAR PUSTAKA

- [1] F. F. Mailoa, “Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia,” *J. Inf. Syst. Public Heal.*, vol. 6, no. 1, p. 44, 2021, doi: 10.22146/jisph.44455.
- [2] DataReportal, “Digital 2023: Indonesia,” 2023. [https://datareportal.com/reports/digital-2023-indonesia?rq=Digital 2023%3A Indonesia](https://datareportal.com/reports/digital-2023-indonesia?rq=Digital%2023%3A%20Indonesia)
- [3] Kurnia, I. Purnamasari, and D. D. Saputra, “Analisis Sentimen Dengan Metode Naïve Bayes, SMOTE Dan Adaboost Pada Twitter Bank BTN,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 2, pp. 235–242, 2023, doi: 10.35870/jtik.v7i3.707.
- [4] F. Syah, H. Fajrin, A. N. Afif, M. R. Saeputra, D. Mirranty, and D. D. Saputra, “Analisa Sentimen Terhadap Twitter IndihomeCare Menggunakan Perbandingan Algoritma Smote, Support Vector Machine, AdaBoost dan Particle Swarm Optimization,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 1, pp. 53–58, 2023, doi: 10.35870/jtik.v7i1.686.
- [5] P. Mehta and S. Pandya, “A review on sentiment analysis methodologies, practices and applications,” *Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 601–609, 2020.
- [6] R. R. A. Siregar, Z. U. Siregar, and R. Arianto, “Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor,” *Kilat*, vol. 8, no. 1, pp. 81–92, 2019, doi: 10.33322/kilat.v8i1.421.

- [7] A. Alsaeedi and M. Z. Khan, "A study on sentiment analysis techniques of Twitter data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 2, pp. 361–374, 2019, doi: 10.14569/ijacsa.2019.0100248.
- [8] T. T. Widowati and M. Sadikin, "Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.
- [9] D. D. Saputra *et al.*, "Optimization Sentiments of Analysis from Tweets in myXLCare using Naive Bayes Algorithm and Synthetic Minority over Sampling Technique Method," *J. Phys. Conf. Ser.*, vol. 1471, no. 1, 2020, doi: 10.1088/1742-6596/1471/1/012014.
- [10] B. Kurniawan, A. Suwarisman, I. Afriyanti, A. Wahyudi, and D. D. Saputra, "Analisis Sentimen Complain dan Bukan Complain pada Twitter Telkomsel dengan SMOTE dan Naive Bayes," *J. JTik (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 1, pp. 106–113, 2023, doi: 10.35870/jtik.v7i1.691.
- [11] I. J. T. Gurning, P. P. Adikara, and R. S. Perdana, "Analisis Sentimen Dokumen Twitter menggunakan Metode Naive Bayes dengan Seleksi Fitur GU Metric," ... *Teknol. Inf. dan Ilmu ...*, vol. 7, no. 5, 2023, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12665%0Ahttps://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/12665/5759>
- [12] S. Febriani and H. Sulistiani, "Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5," *89Jurnal Teknol. dan Sist. Inf.*, vol. 2, no. 4, pp. 89–95, 2021.
- [13] A. Hardoni, D. P. Rini, and S. Sukemi, "Integrasi SMOTE pada Naive Bayes dan Logistic Regression Berbasis Particle Swarm Optimization untuk Prediksi Cacat Perangkat Lunak," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 233, 2021, doi: 10.30865/mib.v5i1.2616.
- [14] K. Setiawan, R. Beni, Burhanuddin, A. Budi Paryanti, and F. Fauzi, "Komparasi Metode Naive Bayes Dan Support Vector Machine Menggunakan Particle Swarm Optimization Untuk Analisis Sentimen Mobil Esemka Jisamar (Journal of Information System , Applied , Management , Accounting

and Research) p-ISSN : 2598-8700 (Printed) J,” *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 4, no. 3, pp. 102–111, 2020, [Online]. Available: <http://journal.stmikjayakarta.ac.id/index.php/jisamarTelp.+62-21-3905050>

- [15] R. Fahlapi *et al.*, “Analisa Sentimen Vaksinasi Covid-19 Dengan Metode Support Vector Machine Dan Naïve Bayes Berbasis Teknik Smote,” *J. Inform. Kaputama*, vol. 6, no. 1, pp. 57–63, 2022, doi: 10.59697/jik.v6i1.136.