

ANALISIS SENTIMEN TERHADAP TELKOMSEL DAN XL BERBASIS MACHINE LEARNING PADA DATA TWITTER

SENTIMENT ANALYSIS TOWARD ISP TELKOMSEL AND XL ON TWITTER USING MACHUNE LEARNING CLASSIFICATION

Trisiwi Indra Cahyani¹, Windu Gata², Dedi Dwi Saputra³, Haffifah Bella Novitasari⁴,
Hernawati⁵

^{1,2,3,5}Universitas Nusa Mandiri, ⁴Kominfo Republik Indonesia
14002458@nusamandiri.ac.id

ABSTRACT

In Indonesia internet users reach more than 200 million users. Telkomsel and XL are compete to be number one internet service provider. Social media Twitter makes users more honest in giving reviews. User feedback will become word of mouth (WoW) recommendations. In this study aims to determine public views of Telkomsel and XL providers based on tweet data on Twitter in July and August 2022. The dataset collected from Twitter using Twitter API with the keywords "XL Internet", "Telkomsel Internet", "MyXL", and "MyTelkomsel" and obtained as many as 17.543 data. Then dataset will be case folding, tokenized, normalized, stopword removal, stemming, and TF-IDF weighting process. Classification model uses Maximum Entropy, Multinomial Naïve Bayes, and Complement Naïve Bayes. To test ability to generalize, 10-Fold Cross Validation was carried out for each model. The result show that ME method is better than MNB and CNB as accuracy value 84.11%, 81.53%, and 79.95%.

Keywords : Sentiment Analysis, Telkomsel, XL, Internet Service Provider, Classification

ABSTRAK

Di Indonesia pengguna internet mencapai lebih dari 200 juta pengguna. Telkomsel dan XL bersaing untuk menjadi penyedia layanan internet nomor satu. Media sosial Twitter membuat pengguna lebih jujur dalam memberikan review. Umpan balik pengguna akan menjadi rekomendasi dari mulut ke mulut (WoW). Pada penelitian ini bertujuan untuk mengetahui pandangan masyarakat terhadap provider Telkomsel dan XL berdasarkan data tweet di Twitter pada bulan Juli dan Agustus 2022. Dataset dikumpulkan dari Twitter menggunakan Twitter API dengan kata kunci "XL Internet", "Telkomsel Internet", "MyXL", dan "MyTelkomsel" dan diperoleh sebanyak 17.543 data. Kemudian dataset akan dilakukan case folding, tokenized, normalized, stopword removal, stemming, dan proses pembobotan TF-IDF. Model klasifikasi menggunakan Entropy Maksimum, Multinomial Naïve Bayes, dan Complement Naïve Bayes. Untuk menguji kemampuan menggeneralisasi, dilakukan 10-Fold Cross Validation untuk masing-masing model. Hasil menunjukkan bahwa metode ME lebih baik dari MNB dan CNB dengan nilai akurasi 84,11%, 81,53%, dan 79,95%.

Kata Kunci : Analisis Sentimen, Telkomsel, XL, Layanan Internet, Klasifikasi

PENDAHULUAN

Perkembangan Teknologi informasi membutuhkan dukungan internet untuk kelancaran informasi. Operator seluler menggunakan jaringan koneksi seluler seperti 3G atau 4G untuk terhubung dengan internet. XL dan Telkomsel merupakan *Internet Service Provider (ISP)* atau Penyedia Layanan Internet yang memiliki kecepatan tertinggi di Indonesia nomor satu dan dua (Opensignal, 2022). Kecepatan tertinggi XL dalam mengunduh mencapai 18,7 Mbps dan mengunggah 6,7

Mbps sedangkan kecepatan tertinggi Telkomsel dalam mengunduh mencapai 16,3 Mbps dan mengunggah 7,7 Mbps. Lebih dari 200 juta atau setara dengan 77% penduduk di Indonesia sudah menggunakan Internet (APJII, 2022). Terjadi peningkatan sebesar 20 persen pengguna internet dari tahun 2020 ke tahun 2022. Masyarakat Indonesia sering menggunakan internet untuk *video conference, e-learning, video streaming, e-commerce*, dan lain-lain.

Twitter adalah layanan media sosial yang dimiliki oleh Twitter Inc., yang menawarkan *microblog* berbasis teks / tulisan. Terdapat 250 karakter yang dapat dituliskan oleh pengguna yang bernama *tweet* (Husada & Toba, 2020). Tweets dapat dilihat secara umum, namun pengguna dapat membatasi dengan hanya memperlihatkan kepada *followers*. Pengguna yang mengikuti pengguna lain dinamakan dengan *followers*. Berbeda dengan media sosial lain, yang mengutamakan produk mereka sebagai nilai jual, Twitter lebih mengutamakan komunitas orang sebagai daya tarik pengguna. Sehingga, pengguna Twitter bertumbuh dengan cepat dan saling mengomentari suatu kejadian secara *real time*. Twitter menjadi salah satu media sosial yang paling populer digunakan oleh masyarakat Indonesia (Toy, Sari, & Cholissodin, 2021). Pada umumnya twitter digunakan sebagai laporan *update* dan penggalangan dana oleh organisasi sosial ketika terjadi bencana alam atau *force majeure*. Namun, Twitter juga menjadi aduan utama ketika menyukai atau memperlakukan produk oleh *brand* atau perusahaan. Perusahaan dapat memanfaatkan Twitter untuk mengali lebih dalam mengenai pandangan publik terhadap kualitas, manfaat, dan kepuasan layanan.

Melalui ulasan dari pengguna di Twitter, dapat digunakan menjadi alat efektif dan efisien dalam menemukan informasi terhadap suatu produk atau jasa. Ulasan pengguna akan membangun rekomendasi berdasarkan *word of mouth* (WoM) yang menjadi peran utama dalam pengambilan keputusan pengguna lain. WoM melibatkan pengguna dalam berbagi *image brand* berdasarkan reaksi atau opini terhadap produk atau jasa yang ditawarkan. Untuk mengekstraksi informasi pada setiap ulasan diperlukan analisis sentimen yang menerima *feedback* (pendapat) dari konsumen berdasarkan ulasan pengguna di Twitter (Sipayung, Maharani, & Zefanya, 2016). Analisis sentimen diaplikasikan

untuk mengelompokkan ulasan ke dalam *feedback* positif, negatif, dan netral (Darwis, Siskawati, & Abidin, 2021). Dengan melalui *feedback* yang akurat, perusahaan dapat secara cepat dan tepat untuk memperbaiki produknya atau mengambil kebijakan lain yang sesuai dengan data yang didapatkan.

Dalam melakukan analisis sentimen dapat dilakukan dengan dua metode yaitu metode *leticon-based* dan *learning-based*. Metode *leticon-based* digunakan berdasarkan kamus kata, dalam studi (Choudhary & Choudhary, 2018), menggunakan metode ini untuk menganalisis data teks Twitter dalam mengulas pembelian barang. Menggunakan *learning-based* dapat mempermudah menganalisis sentimen dengan membangun sebuah model. Penelitian (Darwis, Siskawati, & Abidin, 2021) menggunakan *Naïve Bayes* untuk menganalisis sentimen BMKG di Twitter dengan membagi sentimen ke dalam opini positif, negatif, dan netral menghasilkan akurasi sebesar 69.97% dengan hampir 80% merupakan sentimen positif. Studi serupa menggunakan *Naïve Bayes* pada (Amelia, Damansah, Prastiwi, & Purbaya, 2022) yang melakukan analisis sentimen terhadap drama Korea menghasilkan nilai akurasi sebesar 69%. Metode *Naïve Bayes* menjadi metode yang sering digunakan dalam analisis sentimen dikarenakan mampu menghasilkan kinerja yang cukup baik (Singgalen, 2021). Pembobotan TF-IDF mampu meningkatkan kinerja akurasi dari *machine learning* (Hidayat, Garno, & Ridha, 2021). Menggunakan pendekatan *Multinomial Naïve Bayes*, penelitian Ratnawati (2018) melakukan analisis sentimen berdasarkan ulasan penonton film pada Twitter dengan evaluasi *K-Fold Cross Validation* menghasilkan akurasi rerata 82,4%. Penelitian Fernandes (2022) menggunakan *Maximum Entropy* untuk mengklasifikasi ulasan di twitter terhadap layanan Indosat Ooredoo menghasilkan akurasi sebesar 86.21% dan AUC sebesar 0.968%. Penggunaan *Multinomial* dan

Maximum Entropy mengalami kerugian ketika dataset yang digunakan tidak seimbang antar kelas sehingga *Complement Naïve Bayes* menjadi salah satu solusi (Sutrisno, Jondri, & Lhaksana, 2021). Dalam studi yang dilakukan oleh Sutrisno, *et. all* (2021) *Complement Naïve Bayes* memiliki performa cukup baik dalam mengatasi over sampling pada data yang tidak seimbang. Dari penelitian yang sudah ada sebelumnya, penulis mengambil metode *machine learning* menggunakan algoritma *Maximum Entropy*, *Complement Naïve Bayes*, dan *Multinomial Naïve Bayes* untuk mengklasifikasikan sentimen, sehingga penelitian ini bertujuan untuk membandingkan 3 metode Algoritma berdasarkan hasil klasifikasi pada 2 provider internet, Telkomsel dan xl dengan sumber data teks dari twitter berdasarkan tweet pengguna dalam kualitas penggunaan internet Telkomsel dan XL

METODE

Pengumpulan Data

Pada Penelitian ini, menggunakan kata kunci “Telkomsel Internet”, “XL Internet”, “MyTelkomsel”, dan “MyXL” pada tweet berbahasa Indonesia pada Bulan Juli dan Agustus 2022 yang ditarik dari Twitter. Scraping dilakukan dengan memanfaatkan *snsrape* untuk mendapatkan id tweet yang mempunyai kata kunci. Dari id tweet yang didapatkan kemudian akan dicari teks dari tweet menggunakan *Tweepy* yang terhubung dengan API Twitter dan menghasilkan 17.543 data terkait provider Telkomsel dan XL

Preprocessing

Data yang dikumpulkan kemudian dilakukan tahapan preprocessing dengan langkah-langkah sebagai berikut:

1. *Case Folding* : Mengubah semua karakter dalam dokumen menjadi huruf kecil (*lowercase*) untuk menyeragamkan bentuk huruf

dikarenakan dalam penulisan tweet terdapat perbedaan huruf yang sebenarnya mempunyai arti sama seperti “Baik”, “BAIK”, dan “baik”.

2. *Tokenisasi* : Tokenisasi merupakan pemotongan sebuah kalimat menjadi kata atau token. Pemisahan kata dapat dilakukan dengan mengatur mode split untuk menjadikan untaian kata tunggal. Misalnya kalimat “baik hati” menjadi dua token yaitu token “baik” dan “hati”.
3. *Cleaning* : Dalam sebuah tweet, pengguna sering memberikan link terkait dengan topik yang dibicarakan, namun link tersebut diperpendek oleh twitter sehingga satu link yang sama akan menjadi short link berbeda pada tweet setiap pengguna yang mengakibatkan kualitas tweet menjadi berkurang. Selain link atau URL, juga mention pengguna lain yang tidak memiliki pengaruh dalam ulasan pengguna. Sehingga perlu dilakukan pembersihan mulai URL, mention (@<user_name>), emotikon, angka (1 - 9), simbol (./?><) , hashtag (#), dan retweet.
4. *Normalisasi Kata* : Dalam menuliskan tweet terkadang pengguna menggunakan Bahasa yang dia mengerti dibandingkan menggunakan Bahasa baku misalnya menggunakan kata gua sebagai kata ganti aku atau menggunakan kosakata daerah dan luar negeri untuk menuliskan pandangannya. Sehingga perlu dilakukan perubahan kata dari semua kata tidak baku menjadi kata baku. Normalisasi kata menggunakan kamus kata yang dibuat oleh Putra (2022).
5. *Menghapus Data Duplikat* : Untuk membuat data menjadi lebih baik maka lebih baik menggunakan data yang bervariasi sehingga data yang terduplikat akan dihapus. Terkadang dalam menuliskan padangan sering kali diganggu oleh buzzer atau pengguna yang mengampanyekan sebuah cuitan sama namun dituliskan oleh banyak pengguna secara sistematis yang

- membuat pandangan menjadi terlalu bias.
6. *Stopword Removal* : Tahapan ini menghapus kata-kata yang sering muncul namun tidak memiliki pengaruh terhadap sentimen seperti penunjuk waktu, kata tanya, kata sambung, dan kata yang tidak dapat berdiri sendiri. Korpus kata yang digunakan merupakan korpus dari Sastrawi dan NLTK.
 7. *Stemming* : Menemukan kata dasar dari kata imbuhan diperlukan untuk mengurangi dimensi pada data. Dengan membuang imbuhan awalan dan akhiran pada kata menggunakan Algoritma Nazief dan Adriani menghasilkan kata dasar sehingga dimensi menjadi lebih kecil (Herlingga, Prisma, Prehanto, & Dermawan, 2020). Misalnya kata dasar dari “menemukan” dan “ditemukan” yaitu “temu”. Algoritma Nazief dan Adriani sudah disediakan oleh *library* Sastrawi, sebuah *library* yang berfokus untuk menstemming Bahasa Indonesia yang dikembangkan dari PHP menjadi Python.
 8. *Pembobotan TF-IDF* : Hasil dari proses stemming kemudian dilakukan perhitungan pembobotan menggunakan TF-IDF yang menghasilkan nilai dalam rentang 0-1 pada setiap token atau fitur kata yang digunakan (Randhika, Young, Suryadibrata, & Mandala, 2021). Pembobotan TF-IDF menggunakan *library* Sk-Learn yang disediakan untuk bahasa pemrograman python.

Model Klasifikasi

Maximum Entropy merupakan Algoritma klasifikasi probabilistik yang termasuk pada model eksponensial, dengan berdasar pada nilai entropi. *Maximum Entropy* dapat memanfaatkan data kontinu dan kategorial sehingga efisien namun mengalami kerugian pada fitur yang independen (Shabily, Adikara, & Fauzi, 2019).

Multinomial Naïve Bayes merupakan metode yang efisien dengan berdasarkan

pada *teorema Bayes*. Model ini sangat sensitif terhadap fitur independen sekaligus perhitungan yang dilakukan lebih cepat (Randhika, Young, Suryadibrata, & Mandala, 2021). Namun, MNB sering kali mengalami kegagalan ketika terjadi ketimpangan jumlah data setiap kelas sehingga muncullah perhitungan *Complement Naïve Bayes* yang berperforma baik pada data tidak seimbang (Seref & Bostanci, 2019).

Evaluasi

Validasi model menggunakan K Fold Cross Validation, dimana proses ini akan membagi data secara acak ke dalam 10 bagian. Proses pengujian dimulai dengan pembentukan model dengan data pada bagian pertama sebagai data test dan sisanya sebagai data train (Tapikab, Djahi, & Widiastuti, 2019). Proses ini dilakukan sampai semua bagian telah dijadikan data test. Hasil dari proses ini merupakan ukuran dari performa Algoritma klasifikasi yang dipakai. Performa setiap model akan diukur menggunakan nilai akurasi. Akurasi didapatkan dari Confusion Matrix.

Tabel 1. Confusion Matrix

	Actual Positive		Actual Negative	
Predicted Positive	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)
Predicted Negative	False Positive (FP)	True Positive (TP)	True Negative (TN)	False Negative (FN)

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

HASIL DAN PEMBAHASAN

Pengumpulan Data

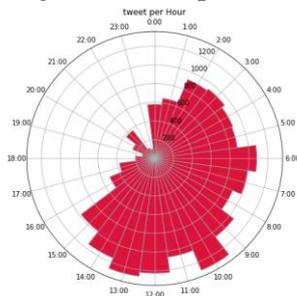
Data dikumpulkan dari media sosial Twitter menggunakan kata kunci “Telkomsel Internet”, “XL Internet”, “MyTelkomsel”, dan “MyXL” dalam Bahasa Indonesia. Tweet diambil pada bulan Juli dan Agustus 2022 menggunakan API Twitter dan mendapatkan 17.543 data. Setiap kata kunci mendapatkan data sejumlah pada Tabel 2.

Tabel 2. Jumlah Data setiap Kata Kunci

No	Kata Kunci	Jumlah
1	Telkomsel Internet	3.855
2	myTelkomsel	8.942

3	XL Internet	3.720
4	MyXL	1.026

Karakteristik pengguna dalam bermain media sosial Twitter mampu bertahan lebih dari dua jam setiap harinya dan jam masa aktif seharian atau 24 jam penuh. Namun pada pengumpulan data terkait provider internet menunjukkan hal berbeda. Jumlah data yang dikumpulkan berdasarkan jam terlihat pada Gambar 1.



Gambar 1. Jumlah Data Berdasarkan Jam
Jika diperhatikan, pengguna sering mencuitkan terkait provider internet pada jam 02.00 s.d 15.00.

Preprocessing

Data yang dikumpulkan, kemudian dilakukan proses preprocessing. Tahap pertama dilakukan penghapusan data duplikat, simbol, angka, dan mengubah huruf kapital menjadi huruf kecil. Hasil proses ini menyisakan data sebanyak 16.051. Selanjutnya data diberikan label secara manual dan sampel hasil pelabelan terdapat pada Tabel 3.

Tabel 3. Pelabelan Data

No	Tweet	Label
1	aplikasi myxl itu sebal erar gangguan	Netral
2	haha kakak mohon maaf atas ketidaknyamanan kamu mengenai jaringan yang dialami xl axiata akan senantiasa meningkatkan kualitas dan pelayanan lagi ke depannya untuk pengecekan lebih detail pesan langsung layanan e care kami di akun twitter iya terima kasih	Positif
3	admin admin ini kenapa tayang nyala sinyalnya jadi lambat iya biasanya cepat di daerah bekasi mustika jaya	Negatif
4	malam kakak saat ini telkomsel melakukan penyesuaian pada spesifikasi dan benefit paket tapi tenang saja kakak karena di aplikasi mytelkomsel masuk ada penawaran paket disney hotstar loh terima kasih arie	Positif
5	tidak internet telkomsel dan wifi indhome kenapa deh jelek sekali ditanya yang lain ternyata gitu juga	Negatif

6	halo kakak mohon maaf iya untuk keluhan tidak bisa akses aplikasi mytelkomsel silakan infokan nomor hp lokasi detail waktu dan tanggal kejadian melalui inbox agar dibantu pengecekan lebih lanjut terimakasih bara	Netral
---	---	--------

Hasil pelabelan kemudian akan dilakukan stopword, stemming, dan normalisasi. Normalisasi menggunakan kamus Putra. Stopword menggunakan stopword Sastrawi dan NLTK Bahasa Indonesia. Stemming menggunakan Nazief dan Andriani pada Library Stemming Sastrawi. Hasil preprocessing kemudian dilakukan pembobotan menggunakan TF-IDF. Hasil sebelum dan sesudah preprocessing terangkum pada Tabel 4.

Tabel 4. Hasil Preprocessig

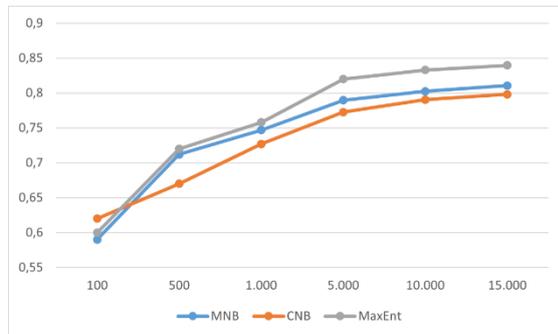
No	Sebelum	Sesudah
1	@yoriyogss @myXL Iya mas sama saya juga begitu. Kuota mepet udah gitu pulsa kepotong gatau karena apa	iya kuota sedikit pulsa potong
2	hai gais kali aja ada yg pemakai telkomsel jugaa, boleh saran, tips atau apaa gt terkait data internet 😊 gue selalu nyari buat sebulan tp kalo ada ga banyak" bgt GB nyaa krn di rumah masih ada wifi. sayang kalo kebuangg, tp mahal bgt hikkss.. apakahh kalian ada yg tauu?	makai telkomsel saran data internet cari satu bulan banyak gb rumah masih wifi sayang buang mahal
3	Weekend waktunya untuk keluarga, dari mulai bercanda sambil cerita seputar kegiatannya masing-masing, bisa bikin komunikasi antar keluarga semakin erat. Sama seperti XL akan selalu #AdaUntukIndonesia dengan jaringan tercepat yang bisa dinikmati bersama keluarga. #WeAreMore https://t.co/Th6aAXmyp	waktu keluarga canda cerita kegiatan komunikasi keluarga makin erat xl jaringan cepat nikmat keluarga.

Kinerja Model Berdasarkan Jumlah Data

Algoritma model klasifikasi Multinomial Naive Bayes, Complement Naive Bayes, dan Maximum Entropy akan dilakukan pengujian 10-Fold Cross Validation menggunakan Dataset yang dikumpulkan. Pengujian dilakukan menggunakan sebagian data dari Dataset untuk mengetahui kinerja model dalam jumlah data yang berbeda. Hasil pengujian dirangkum pada Tabel 5 dan grafik perubahan akurasi pada Gambar 2.

Tabel 5. Kemampuan Model

<i>Sum</i>	<i>Acc MNB</i>	<i>Acc CNB</i>	<i>Acc ME</i>
100	0.5900	0.6200	0.6000
500	0.7120	0.6700	0.7200
1.000	0.7470	0.7270	0.7580
5.000	0.7896	0.7726	0.8200
10.000	0.8023	0.7904	0.8329
15.000	0.8106	0.7981	0.8397



Gambar 2. Jumlah Data Berdasarkan Jam
Berdasarkan Gambar 2, keseluruhan model mengalami peningkatan setiap kali terjadi pertambahan data yang digunakan. Model *Maximum Entropy* mengalami peningkatan paling tinggi jika dibandingkan model *Multinomial Naïve Bayes*, dan *Complement Naïve Bayes*.

Kinerja Model Klasifikasi

Ketiga model kemudian dilakukan pengujian menggunakan keseluruhan dataset dengan mengganti nilai K pada K-Fold Cross Validation. Nilai K yang digunakan adalah 5, 10, dan 15. Hasil percobaan telah dirangkum pada Tabel 6.

Tabel 6. Kemampuan Model

<i>metrix</i>	<i>5 Fold</i>	<i>10 Fold</i>	<i>15 Fold</i>
Acc MNB	0.8122	0.8153	0.8155
Pre MNB	0.8165	0.8204	0.8204
Acc CNB	0.7993	0.7995	0.7999
Pre CNB	0.8168	0.8185	0.8192
Acc ME	0.8407	0.8411	0.8410
Pre ME	0.8412	0.8411	0.8424

Jika diperhatikan, dalam percobaan ini perbedaan nilai K tidak memiliki pengaruh yang signifikan. Sehingga analisis perlu dilanjutkan untuk melihat setiap bagian pada 10-Fold Cross Validation. Hasil pengujian terdapat pada Tabel 7.

Tabel 7. Kemampuan Model pada 10 Fold

<i>Fold</i>	<i>Acc MNB</i>	<i>Acc CNB</i>	<i>Acc ME</i>
1	0.8101	0.8169	0.8325
2	0.8006	0.8224	0.8355
3	0.8019	0.8081	0.8455
4	0.7981	0.8100	0.8467
5	0.7844	0.8019	0.8436
6	0.8150	0.8237	0.8424
7	0.7869	0.8056	0.8268
8	0.7882	0.8125	0.8368
9	0.8000	0.8299	0.8536
10	0.8106	0.8224	0.8486

Berdasarkan akurasi model yang didapatkan, dihitung standar deviasi menggunakan fungsi *stdev* dari *library statistics* dan didapatkan nilai standar deviasi 0.0091 untuk akurasi *Multinomial Naïve Bayes*, 0.0105 untuk akurasi *Complement Naïve Bayes*, dan 0.0081 untuk *Maximum Entropy*. Nilai standar deviasi yang kecil ini, menunjukkan bahwa ketiga model sudah baik dalam menggeneralisasi data dan bekerja dengan baik pada setiap bagian dari dataset.

Model klasifikasi akan diuji coba dengan dataset bulan Juli 2022 dan dataset bulan Agustus 2022. Data dibagi menggunakan 10-Fold Cross Validation. Hasil pengujian dirangkum pada Tabel 8.

Tabel 8. Kemampuan Model pada 10 Fold

<i>ISP</i>	<i>Acc MNB</i>	<i>Acc CNB</i>	<i>Acc ME</i>
Telkomsel	0.8101	0.7979	0.8323
XL	0.8388	0.8221	0.8533

Berdasarkan Tabel 8 dapat dilihat bahwa model lebih mudah memahami dan memiliki kinerja yang lebih baik pada XL daripada dataset Telkomsel. Perbedaan kinerja model pada metrik akurasi berkisar 1-7%. Untuk telkomsel, dapat lihat model *Complement Naïve Bayes* lebih baik dibandingkan model *Multinomial Naïve Bayes*, berbeda dengan XL model *Complement Naïve Bayes* lebih rendah dibandingkan model *Multinomial Naïve Bayes*. Hasil pelabelan setiap sentiment dilihat pada Tabel 9.

Tabel 9. Sentimen setiap provider

<i>Sentimen</i>	<i>Telkomsel</i>	<i>XL</i>
Netral	4898	5393
Positif	650	439

Negatif	2344	2387
---------	------	------

Ringkasan

Model percobaan klasifikasi yang digunakan, yaitu *Maximum Entropy*, *Multinomial Naïve Bayes*, dan *Complement Naïve Bayes*. Model yang mempunyai kinerja paling baik terdapat pada *Maximum Entropy* dengan akurasi sebesar 0.8411. Dalam penelitian ini model *Complement Naïve Bayes* tidak lebih baik dibandingkan *Multinomial Naïve Bayes*, namun perbedaan kinerja kedua model tidak terlalu signifikan. Ketiga model telah mampu untuk mengklasifikasikan ketiga label untuk provider telkomsel dan xl, dan kinerja yang didapatkan telah memuaskan dikarenakan mampu kinerja model diatas 80%.

Jumlah data yang akan digunakan mempengaruhi kemampuan model dalam melakukan klasifikasi. Semakin sedikit data yang digunakan model akan memungkinkan model untuk bekerja lebih cepat namun dengan kinerja yang tidak optimal. Sebaliknya, dengan memperbanyak data akan memiliki kemungkinan untuk model berkinerja dengan baik, namun kebutuhan waktu akan semakin meningkat selaras dengan jumlah data.

Penggunaan evaluasi k-fold bertujuan untuk meyakinkan bahwa model mampu menggeneralisasi dengan baik tanpa mempertimbangkan penempatan data yang ada. Dari penelitian yang sudah dilakukan, penggunaan nilai K menghasilkan kinerja yang tidak berbeda signifikan atau di bawah 2% dengan K tertinggi pada nilai 10 dan 15. Menggunakan K yang berbeda kinerja *Maximum Entropy* masih terbaik dan *Complement Naïve Bayes* memiliki kinerja yang terendah, namun tidak berbeda signifikan dengan *Multinomial Naïve Bayes*.

Mengevaluasi setiap bagian fold pada pembagian k-fold dilakukan untuk mengetahui kinerja model dalam menggeneralisasi hasil yang sama

walaupun dengan titik berbeda. Berdasarkan pengujian yang sudah dilakukan dengan mengevaluasi setiap fold pada 10-Fold Cross Validation, diketahui bahwa kinerja ketiga model tidak berbeda signifikan dengan standar deviasi di bawah 0.1 pada setiap Fold

SIMPULAN

Berdasarkan penelitian yang sudah dilakukan mengenai analisis sentimen terkait provider internet Telkomsel dan XL pada bulan Juli dan Agustus 2022 di Indonesia yang dilakukan pengujian menggunakan tiga model, yaitu *Maximum Entropy*, *Multinomial Naïve Bayes*, dan *Complement Naïve Bayes*, maka dapat diambil kesimpulan bahwa hasil pengujian klasifikasi terbaik pada *Maximum Entropy*. Kinerja klasifikasi algoritma *Multinomial Naïve Bayes*, *Complement Naïve Bayes*, dan *Maximum Entropy* menghasilkan akurasi rerata sebesar 81.53%, 79.55%, dan 84.11%. Secara rata-rata model *Complement* mempunyai kinerja yang lebih rendah dibandingkan *Multinomial*, hal ini dikarenakan standar deviasi *Complement* lebih tinggi dibandingkan *Multinomial* yang mempengaruhi model dalam hasil kinerja klasifikasi setiap fold pada 10 Fold Cross Validation. Dikarenakan model sudah mampu mencapai akurasi diatas 80% dan standar deviasi dibawah 2% maka ketiga model mampu bekerja dengan baik untuk mengklasifikasikan data teks provider Telkomsel dan XL sehingga dapat digunakan sebagai model untuk analisis sentimen.

DAFTAR PUSTAKA

Amelia, R., Damansah, Prastiwi, N., & Purbaya, M. (2022). Implementasi Algoritma Naïve Bayes terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea

- pada Twitter. *JURIKOM : Jurnal Riset Komputer*, 9(2), 338-343.
- APJII. (2022). *Indonesia Digital Outlook 2022*. Jakarta: Asosiasi Penyelenggara Jasa Internet Indonesia.
- Choudhary, M., & Choudhary, P. (2018). Sentiment Analysis of Text Reviewing Algorithm using Data Mining. *2018 Inter. Conf. on Smart Sys. and Inventive Tech. (ICSSIT)*.
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naïve Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *Jurnal Tekno Kompak*, 15(1), 131-145.
- Fernandes, G., & Lina, I. (2022). Sentiment Analysis on Twitter Using Maximum Entropy : a Case Study on Indosat Ooredoo. *E-Komtek*, 6(1), 27-35.
- Herlingga, A., Prisma, P., Prehanto, D., & Dermawan, D. (2020). Algoritma Stemming Nazief & Adriani dengan Metode Cosine Similarity untuk Chatbot Telegram Terintegrasi dengan E-layanan. *JINACS : Journal of Informatics and Computer Science*, 2(1), 19-26.
- Hidayat, T., Garno, & Ridha, A. (2021). Analisis Sentimen Permindahan Ibu Kota pada Twitter dengan Metode Support Vector Machine. *Jurnal Ilmu Komputer*, 14(2).
- Husada, I., & Toba, H. (2020). Pengaruh Metode Penyeimbangan Kelas Terhadap Tingkat Akurasi Analisis Sentimen pada Tweets Berbahasa Indonesia. *Jurnal Teknik Informatika dan Sistem Informasi*, 6(2), 400-413.
- Opensignal. (2022). *Laporan Pengalaman Jaringan Seluler Indonesia*. Opensignal Limited.
- Putra, G., Swastika, W., & Irawan, P. (2022). Perbandingan Particle Swarm Optimization dengan Genetic Algorithm dalam Feature Selection untuk Analisis Sentimen pada Permendikbudristek PPKS-LPT. *JEPIN*, 8(3), 412-421.
- Randhika, M., Young, J., Suryadibrata, A., & Mandala, H. (2021). Implementasi Algoritma Complement dan Multinomial Naïve Bayes Classifier pada Klasifikasi Kategori Berita Media Online. *Ultimatics : Jurnal Teknik Informatika*, 13(1).
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film pada Twitter. *Jurnal Inovtek Polbeng*, 3(1), 50-60.
- Seref, B., & Bostanci, E. (2019). Performance Comparison of Naïve Bayes and Complement Naïve Bayes Algorithm. *2019 IEEE 6th International Conference on Electrical and Electronics Engineering (ICEEE)*. USA.
- Shabily, A., Adikara, P., & Fauzi, M. (2019). Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan metode Maximum Entropy. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(5), 4204-4209.
- Singgalen, Y. (2021). Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Systematic Literature Review. *Jurnal of Information System and Informatics*, 3(2), 278-302.
- Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Informasi (JSI)*, 8(2), 958 - 965.
- Sutrisno, F. K., Jondri, & Lhaksana, K. (2021). Analisis Sentimen Destinasi Wisata Kuliner di Twitter Menggunakan TF-IDF dan Complement Naive Bayes pada Dataset Tidak Seimbang. *Proceeding of Engineering*, (pp. 4115-4123).
- Tapikab, H., Djahi, B., & Widiastuti, T. (2019). Klasifikasi Spam E-mail Menggunakan Metode Transformed

Complement Naïve Bayes (TCNB).

J-ICON : J. Komputer dan Informatika, 7(1), 21-26.

Toy, K. S., Sari, Y. A., & Cholissodin, I. (2021). Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(11), 5068-5074.