

# Analisis Media Sosial Penyedia Layanan Internet Menggunakan Algoritma XGBOOST

Jonathan Iskandar<sup>1</sup>, Viny Christanti Mawardi<sup>2</sup>, Janson Hendryli<sup>3</sup>

Teknik Informatika, Fakultas Teknologi Informasi

Universitas Tarumanagara

Jakarta, Indonesia

e-mail: <sup>1</sup>jonathaniskandar@outlook.com, <sup>2</sup>viny@fti.untar.ac.id, <sup>3</sup>jansonh@fti.untar.ac.id

## Abstrak

Pada era revolusi industri 4.0 ini, internet telah berkembang menjadi hal vital yang berpengaruh pada aktivitas masyarakat baik dari segi sosial maupun ekonomi. Hal ini terasa signifikan di Indonesia dilihat dari meningkatnya jumlah pengguna media sosial, sehingga membuat Indonesia menjadi salah satu negara dengan pengguna internet terbesar di dunia. Provider internet bermunculan semakin banyak untuk memenuhi kebutuhan pengguna. Pelayanan yang diberikan dapat dinilai oleh masyarakat dengan berbagai cara seperti menggunakan media sosial. Saat ini banyak aplikasi yang digunakan untuk membantu menganalisa kepuasan pelanggan dari media sosial. Analisa tersebut dapat digunakan untuk meningkatkan pelayanan provider dan membantu masyarakat mengetahui kualitas provider. Pada penelitian ini dilakukan penelitian untuk membuat aplikasi yang dapat membantu pengguna internet di Indonesia untuk mengetahui informasi provider mana yang menyediakan layanan internet yang baik berdasarkan komentar dari pengguna. Penelitian ini dilakukan dengan mengumpulkan data teks melalui API yang disediakan oleh Twitter dan diubah menjadi vektor menggunakan TF-IDF kemudian komentar teks diklasifikasi menggunakan algoritma XGBoost. Penelitian ini telah menghasilkan sebuah aplikasi berbasis web yang dapat menunjukkan hasil klasifikasi sentimen secara real-time mengenai penyedia layanan internet di Indonesia. Akurasi terbaik dari metode ini adalah 89.3% dengan 263 data yang diambil dari 9 provider.

**Kata kunci:** Analisis sentimen, Twitter, TF-IDF, XGBoost, Penyedia Layanan Internet.

## Abstract

In this era of the industrial revolution 4.0, the internet has developed into a vital thing that affects people's activities both socially and economically. This is significant in Indonesia, seen from the increasing number of social media users, thus making Indonesia one of the countries with the largest internet users in the world. More and more internet providers have sprung up to meet user needs. The services provided can be assessed by the community in various ways, such as using social media. Currently, many applications are used to help analyze customer satisfaction from social media. This analysis can be used to improve service providers and help the public know the quality of providers. In this study, research was conducted to create applications that can help internet users in Indonesia to find out which provider information provides good internet services based on comments from users. This research was conducted by collecting text data through the API provided by Twitter and converted into vectors using TF-IDF then the text comments were classified using the XGBoost algorithm. This research has produced a web-based application that can show the results of sentiment classification in real-time regarding internet service providers in Indonesia. The best accuracy of this method is 89.3% with 263 data taken from 9 providers.

**Keywords:** Sentiment analysis, Twitter, TF-IDF, XGBoost, Internet Service Provider.

## 1. Pendahuluan

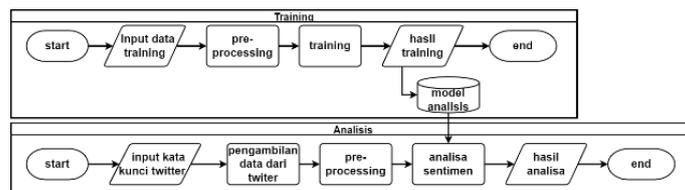
Pada saat ini, internet telah berkembang menjadi hal vital yang berpengaruh pada aktivitas masyarakat baik dari segi sosial maupun ekonomi. Berdasarkan survei yang dilakukan oleh APJII media sosial menjadi alasan untuk seseorang menggunakan internet [1] dan rata-rata waktu yang digunakan dalam berinternet adalah sebanyak 8 jam ke atas [2]. Internet telah menjadi kebutuhan yang tidak terpisahkan dari kehidupan rumah tangga masyarakat sehari-hari, sehingga kestabilan jaringan internet yang baik telah menjadi keharusan dalam memilih penyedia layanan internet. Media sosial merupakan wadah bagi perusahaan untuk menerima ulasan dari pelanggan, serta media sosial dapat mempermudah layanan *Customer Service* [3], dimana dapat juga menjadi sarana bagi para pelanggan dalam menyampaikan keluhan kepada perusahaan sehingga menjadi alternatif dari *Customer Service*. Perancangan sistem analisis media sosial ini dimaksudkan untuk membantu pengguna internet di Indonesia dalam melihat sentimen pengguna media

sosial berbasis teks terhadap penyedia layanan internet populer, guna membantu dalam pengambilan keputusan. Analisis media sosial dilakukan untuk mendapatkan bagaimana pendapat ataupun pandangan dari pelanggan terhadap perusahaan-perusahaan penyedia layanan internet. Proses analisis dengan klasifikasi algoritma XGBoost merupakan algoritma pengembangan dari *Gradient Boosting* untuk membangun model dengan prediksi yang lebih kuat sehingga memiliki performa baik dalam proses klasifikasi [4]. Sebelumnya penelitian terhadap layanan *e-payment service provider* menghasilkan akurasi dari nilai *recall* untuk K-NN dan XGBoost sebanyak 85,2% dan 82,8%, dan untuk *precision* Naïve Bayes mendapat akurasi paling tinggi sebanyak 72% [5]. Hal serupa dilakukan oleh Rofiqoh et. Al [6], dalam jurnalnya yang meneliti sentiment analisis terhadap penyedia layanan seluler di Indonesia, yang dimana datasetnya akan digunakan pada perancangan ini. Berdasarkan penelitian sebelumnya, diciptakanlah rancangan penelitian ini yang memuat sentiment analisis terhadap penyedia layanan internet di Indonesia dalam bentuk aplikasi berbasis web yang lebih mudah diakses oleh masyarakat umum.

**2. Metode Penelitian**

Pada penelitian ini dilakukan beberapa tahapan dalam menyusun aplikasi berbasis web. Tahapan pembuatan aplikasi ini terdiri dari analisis, pengumpulan data, perancangan diagram alir dan antar muka dan pengujian sistem. Analisis dilakukan untuk memilih aplikasi sosial media yang akan dianalisis dan memilih metode yang akan digunakan dalam pembuatan aplikasi. Pengumpulan data dilakukan untuk membangun model dari metode XGBoost. Perancangan diagram alir dan antar muka digunakan sebagai rancangan untuk menghasilkan aplikasi. Pengujian sistem dilakukan untuk memeriksa apakah metode sudah dapat digunakan sesuai dengan tujuan sistem yaitu menghasilkan hasil analisis.

Sistem yang dirancang merupakan sebuah sistem media sosial analisis dengan berbasis web menggunakan media sosial Twitter sebagai sumber data dan menggunakan Python sebagai bahasa pemrograman. Sistem memungkinkan untuk menerima kata kunci secara manual dan juga menggunakan tombol yang secara otomatis mengambil akun Twitter penyedia layanan internet sebagai kata kunci.

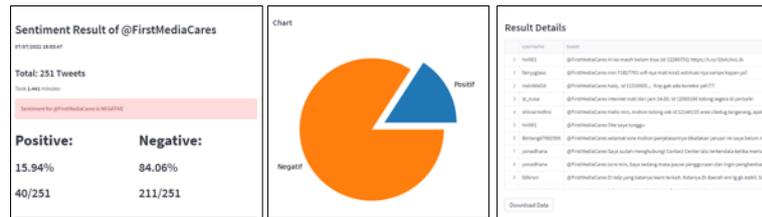


Gambar 3. Diagram Alir

Sistem yang dirancang adalah sistem untuk melakukan *request* data dan analisis sentimen terkandung dalam tweet yang terdapat pada media sosial Twitter. Pengambilan tweet dari Twitter dilakukan dengan menggunakan library tweepy untuk mengakses API milik Twitter dengan menggunakan bahasa pemrograman Python. Dengan menggunakan API yang telah disediakan oleh Twitter, sistem dapat mengambil informasi mengenai tweet berdasarkan data yang terdapat pada basis data Twitter. Pada gambar 1 dapat dilihat rancangan diagram alir yang terbagi dalam dua proses, yaitu proses Training dimana data latih dimasukkan ke dalam sistem untuk dilakukan pembersihan data melalui tahap pra-pemrosesan dan dilakukan proses pelatihan model dan disimpan dalam penyimpanan model analisis. Lalu diikuti proses Analisis untuk sistem menerima masukkan dari pengguna dari tombol yang disediakan dan mengambil data yang memenuhi parameter masukkan menggunakan API Twitter. Setelah data dari API Twitter didapatkan, data melewati proses pembersihan pada tahap pra-pemrosesan lalu digunakan dalam proses analisis bersama dengan model yang telah didapatkan dari proses latih sebelumnya untuk mendapatkan hasil.



Gambar 2. Rancangan tampilan utama aplikasi



Gambar 3. Rancangan tampilan hasil analisis provider

Sistem aplikasi yang dirancang merupakan aplikasi yang berbasis pada web. Dimana membutuhkan spesifikasi rancangan sistem yang terdiri dari bagian home yang digunakan untuk memilih salah satu penyedia layanan internet yang telah yang akan dianalisis seperti pada gambar 2. Bagian hasil analisis dan detil analisis merupakan tampilan yang dapat dilihat pada Gambar 3. Pengguna dapat mengetahui hasil analisis dari unggahan suatu provider yang didapatkan dari sistem dalam bentuk persentase positif dan negatif dan pie chart. Sistem juga akan menampilkan detil teks sentimen yang diperoleh dalam bentuk list kumpulan unggahan untuk suatu provider.

### 2.1. Pra-pemrosesan

Pra-pemrosesan merupakan tahap persiapan sebelum data yang didapatkan diproses oleh sistem untuk pelatihan sistem maupun proses analisis. Pemrosesan yang dilakukan pada tahap ini antara lain *case folding*, *filtering*, dan *tokenizing*.

*Case folding* merupakan proses dimana mengubah atau menghapus bagian-bagian yang tidak relevan dalam proses karena tidak mempengaruhi dalam pemrosesan teks. Filterisasi atau *filtering* merupakan proses pengambilan kata-kata yang penting dan menghapus kata yang kurang penting, memiliki frekuensi tinggi, dan tidak memiliki makna agar sistem dapat lebih menitikberatkan pada kata yang lebih penting. *Tokenizing* atau tokenisasi merupakan tahap membagi kata-kata pada kalimat menjadi *token-token* sehingga mempersiapkan setiap kata untuk memiliki nilainya sendiri.

### 2.2. Representasi Kata

Pada tahap pembuatan representasi kata, digunakan modul TF-IDF untuk mengubah kata-kata yang telah diproses dalam tahap sebelumnya menjadi bentuk vektor yang merepresentasikan kata tersebut berdasarkan hubungan antar kata yang terdapat pada data latih. TF-IDF adalah pembobotan yang dipergunakan dalam penentuan bobot dari suatu data dalam *text mining* [7]. Nilai TF-IDF bertambah sebanding dengan jumlah munculnya istilah dalam dan bergantung pada jumlah dokumen dalam korpus yang memiliki istilah tersebut. TF-IDF dapat dihitung dengan mengalikan TF dan IDF [8] yang dapat dilihat pada persamaan (1).

$$tdidf(t, d, D) = tf(t, d) \times idf(t, D) \quad (1)$$

TF atau *Term frequency*,  $tf(t, d)$ , merupakan banyaknya frekuensi kata  $t$  di dalam dokumen  $d$ . IDF atau *Inverse document frequency*,  $idf(t, D)$ , adalah informasi seberapa sering istilah  $t$  muncul dalam keseluruhan dokumen seperti pada persamaan (2) [8]. Dengan nilai  $D$  merupakan banyak dokumen dan  $t'$  sebagai banyaknya dokumen yang mengandung istilah  $t$ .

### 2.3. Algoritma XGBoost

Pada pelatihan modul algoritma XGBoost vektor dimasukkan ke dalam pohon keputusan, pada pohon keputusan dibentuk model yang menentukan sentimen yang terbentuk dari vektor hasil relasi dari kata satu dengan kata lainnya.

Proses XGBoost dimulai dengan menentukan tabel nilai prediksi awal dan parameter seperti *maximum depth* untuk menentukan banyaknya turunan dari akar, *learning rate* untuk menentukan banyaknya nilai yang didapatkan dari proses latih, *gamma* untuk menentukan *hyperparameter pseudo-regularization* dalam *gradient boosting*, *minimum child weight* untuk menentukan nilai minimum dari suatu daun, dan *base score* untuk menentukan skor awal [9].

Dari tabel nilai prediksi dicari ambang batas akar dengan menghitung nilai  $x$  rata-rata dua titik berdekatan untuk menentukan nilai pemisah daun, selanjutnya nilai dari dua daun tersebut digunakan untuk mencari nilai *similarity* dengan menggunakan persamaan (2) dengan  $\hat{Y}_i$  merupakan residual ke- $i$ ,  $\lambda$  merupakan regresi lambda, dan  $F_i$  merupakan probabilitas ke- $i$  [10].

$$Similarity = \frac{(\sum \hat{Y}_i)^2}{\sum [F_{i-1} \times (1 - F_{i-1})] - \lambda} \quad (2)$$

Setelah didapatkan nilai *similarity* dari kedua daun, *gain* didapatkan dengan menggunakan menjumlahkan kedua daun dan dikurangi oleh *similarity*. Setelah *gain* dari semua pohon dua titik berdekatan didapatkan, diambil pohon dengan *gain* paling besar dan menghitung besarnya *output value* dari masing-masing daun dengan menggunakan persamaan (3) untuk membentuk pohon akhir model 1 dengan  $\hat{Y}_i$  merupakan residual ke- $i$ ,  $\lambda$  merupakan regresi lambda, dan  $F_i$  merupakan probabilitas ke- $i$  [10].

$$\text{Output Value} = \frac{(\sum \hat{Y}_i)}{\sum [F_{i-1} \times (1 - F_{i-1})] - \lambda} \quad (3)$$

Untuk membentuk perhitungan nilai prediksi model ke- $n$  atau  $F_n(x)$  digunakan persamaan (4) untuk membentuk tabel prediksi  $F_n(x)$  dengan  $h_i(x)$  merupakan nilai probabilitas ke- $i$  [10].

$$F_n(x) = \frac{1}{1 + EXP(-x)} \left( \left[ \frac{h_0(x)}{1 - h_0(x)} \right] + \sum_{i=1}^n [\eta \times h_i(x)] \right) \quad (4)$$

Prediksi uji menggunakan modul pengujian data uji yang dihasilkan dibandingkan dengan nilai sentimen yang benar terdapat pada data latih, hasil perbandingan digunakan untuk proses pengukuran akurasi yang dilakukan oleh modul.

Setelah modul sudah terbentuk dan mendapatkan nilai akurasi, modul digunakan pada aplikasi untuk proses analisa data tweet dari Twitter yang sudah melewati proses pra-pemrosesan. Hasil dari proses analisa didapatkan setelah data Twitter melewati modul analisis sentimen dan dikeluarkan berupa persentase dari nilai sentimen positif dan negatif, dan juga berbentuk diagram. Selain mengeluarkan data keseluruhan, hasil juga dikeluarkan dalam bentuk tabel untuk masing-masing tweet.

### 3. Hasil dan Pembahasan

Setelah sistem selesai dirancang dan dibuat, dilakukan pengujian terhadap program yang telah dibuat. Proses pengujian dilakukan pada model sistem analisis sosial media dilakukan dengan menerapkan algoritma XGBoost dan dengan tujuan untuk mengetahui apakah sistem yang telah dibuat telah memenuhi rancangan sistem. Pengujian terhadap model dilakukan untuk mengetahui kemampuan sistem dalam menampilkan hasil dari proses analisis tweet terhadap penyedia layanan internet yang telah dipilih. Pengujian ini dilakukan untuk memastikan semua fitur dapat berjalan dengan baik sesuai dengan yang diinginkan.

XGBoost adalah sebuah metode yang menghasilkan sebuah model tree dari data latih yang disediakan. Untuk memperoleh model tersebut perlu dilakukan pelatihan menggunakan data yang sudah diberi label secara manual. Data latih berisi kalimat sentiment yang umum digunakan di media sosial. Pada penelitian ini dilakukan pengumpulan data dari dua buah sumber, data latih dari dataset tweet sentiment cellular service provider yang sudah diberi label secara manual yang diambil pada bulan Januari 2020 sebanyak 300 tweet [6]. Data latih kedua diperoleh dari pengumpulan data twitter 9 provider pada bulan Juni 2022 sebanyak 263 data. Data yang dikumpulkan akan digunakan untuk proses representasi kata sebagai vektor setelah melewati tahap pra-pemrosesan.

#### 3.1. Hasil Pengujian Pengambilan Tweet

Pada pengujian pengambilan tweet dilakukan pengambilan tweet menggunakan API Twitter ke dalam sistem yang nantinya digunakan dalam proses analisa sentimen. Tabel 1 menggambarkan waktu yang dibutuhkan saat pengambilan data dilakukan berdasarkan jumlah tweet yang diinginkan dan jumlah tweet yang dapat diperoleh dalam 7 hari terakhir. Rata-rata waktu yang diperlukan sistem untuk mengambil 250 tweet dengan periode tidak terbatas adalah selama 1.7 menit. Waktu penarikan data dengan API membutuhkan waktu yang relatif cukup lama sehingga akan mempengaruhi proses pelatihan untuk melakukan sosial media analisis.

Pada penelitian ini juga dilakukan pengujian penarikan data Twitter dengan API. Pengujian dilakukan dengan menarik sebanyak 10 tweet yang sesuai dengan penyedia layanan internet yang dipilih secara *realtime*. Berdasarkan tabel 1 dapat dilihat bahwa jumlah data provider Groovy yang ditarik hanya terdiri dari 1 tweet, hal ini disebabkan karena dalam 7 hari terakhir hanya ada 1 tweet yang muncul untuk provider Groovy. Apabila analisis dilakukan dengan mengambil tweet berdasarkan jumlah kuantitas tweet, proses analisis akan menggunakan waktu yang terlalu lama. Namun apabila analisis dilakukan dengan membatasi rentang waktu tertentu dengan mengambil tweet periode terakhir akan lebih cepat namun perolehan tweetnya sedikit. Sehingga apabila ada provider yang tidak mendapatkan komentar dalam 7 hari terakhir tidak dapat menghasilkan analisis yang diinginkan.

Tabel 3. Waktu Pengambilan Setiap Provider (Menit) dan Jumlah Tweet

Provider	Waktu untuk 250 Tweet	Jumlah Tweet dalam 7 hari
IndiHome	1.650	10
First Media	1.695	10
Biznet	1.690	10
My Republic	1.639	10
MNC Play	1.656	10
Transvision	1.630	10
CBN	1.746	10
Megavision	1.805	6
Groovy	1.732	1
Rata-rata Waktu	1.700	

### 3.2. Hasil Pengujian dengan XGBoost

Pengujian ini merupakan pengujian algoritma XGBoost untuk menentukan akurasi terbesar yang digunakan dalam sistem. Pengujian metode XGBoost ini dilakukan dengan dua dataset. Pengujian pertama adalah pengujian untuk dengan menggunakan 300 data tweet dari dataset dari penelitian [6] yang terdiri dari 139 sentimen positif dan 161 sentimen negatif. Sedangkan pengujian kedua adalah pengujian menggunakan 263 data, terdiri dari 113 sentimen positif dan 150 sentimen negatif yang dikumpulkan secara langsung pada bulan Juni 2022.

Pengujian data training dilakukan sebanyak 5 kali untuk membangun model yang dapat dilihat pada tabel 2. Pengujian mendapatkan nilai tertinggi untuk dataset dari penelitian [6] adalah 83.5%. Sedangkan untuk dataset yang diambil dari 9 provider menghasilkan akurasi tertinggi sebesar 89.3%. Berdasarkan pengujian ini dapat dilihat bahwa metode XGBoost dapat mengklasifikasi data tweet dengan baik ke kategori positif dan negatif.

Tabel 2. Rata-rata Akurasi Pengujian

Pengujian	Dataset [6]				Data 9 provider				
	Akurasi	F1 Score	Precision	Recall	Pengujian	Akurasi	F1 Score	Precision	Recall
1	0.817	0.815	0.826	0.817	1	0.859	0.857	0.866	0.859
2	0.810	0.807	0.817	0.810	2	0.848	0.846	0.852	0.843
3	<b>0.835</b>	0.835	0.838	0.835	3	0.814	0.804	0.853	0.814
4	0.833	0.829	0.844	0.833	4	<b>0.893</b>	0.892	0.899	0.893
5	0.819	0.813	0.860	0.819	5	0.886	0.885	0.889	0.886

Hasil akhir dari pengujian ini diketahui bahwa sistem analisis sosial media dengan menggunakan algoritma XGBoost dapat menentukan sentimen yang terdapat pada tweet terhadap penyedia layanan internet melalui sosial media Twitter. Metode XGBoost yang membangun model dalam bentuk struktur tree dapat mengenali susunan kata bukan hanya berdasarkan jenis kata apakah ada kata sentimen yang bersifat umum seperti baik, bagus, jelek atau lainnya. Metode ini memprediksi kata berdasarkan model susunan kata yang terbentuk dari data training. Sehingga kalimat dengan pola-pola yang beragam dapat dikenali sebagai salah satu kelas sentimen yang tidak bergantung hanya satu kata saja secara independen.

Hasil dari model ini digunakan untuk melakukan analisis data uji sebanyak 60 data tes yang ditarik secara real time atau sebanyak 48 tweet berada pada sentimen yang seharusnya. Dalam hasil pengujian klasifikasi sentimen dengan XGBoost didapatkan hasil akurasi sebesar 80%. Label yang paling banyak salah diklasifikasikan adalah label negatif menjadi positif yaitu sebanyak 9 label negatif dikategorikan menjadi positif, sedangkan label positif menjadi negative hanya sebanyak 3 tweet.

Tabel 3. Hasil Prediksi

No	Teks	Hasil	Benar/Salah
1	harga jual jumlah per rupiah dari kilobyte internet memang juara mahal nya sih tsel vs	positive	Salah
2	asik juga kencang begini	positive	Benar
3	jarigan lelet karena gangguan atau rusak ya	negative	Benar

Pada Tabel 3 dapat dilihat pada proses pengukuran akurasi ditemukan kesalahan pelabelan pada kalimat 1 yang mendapat label "positif" karena terdapat 2 kata dengan 2 sentimen yang berbeda seperti kata "juara" yang bermakna positif dan "mahalnya" yang bermakna negatif. Terdapatnya dua makna ganda tersebut menyebabkan kesalahan dalam memberikan klasifikasi sentiment. Sehingga sistem masih belum dapat menganalisis kalimat sentimen yang memiliki kata sentimen ganda walaupun sebenarnya memiliki makna tersirat bahwa sebenarnya pernyataan tersebut bermakna negatif. Seperti kata "juara" pada kalimat tersebut sebenarnya memberikan makna bahwa biaya internet dari provider tersebut sangat mahal.

Pada kalimat 2 dan 3 sistem berhasil mengklasifikasikan kalimat tersebut ke kelas sentimen yang seharusnya. Pada kalimat 2 kata “asik” dan “kencang” berhasil mendapat makna positif walaupun kata “asik” ditulis tidak baku dan kata “kencang” memberikan penegasan yang sesuai dengan makna positif. Pada kalimat 3 kata “lelet”, “gangguan” dan “rusak” memberikan makna negatif walaupun tidak menggambarkan perasaan dari si penulis. Ketiga kata tersebut merupakan kata sifat, kata benda dan kata sifat yang menunjukkan sesuatu yang arahnya negatif.

#### 4. Kesimpulan

Berdasarkan analisa dan hasil dari tahap pengujian terhadap sistem Sosial Media Analisis Layanan Provider Internet Menggunakan Algoritma XGBoost, maka diperoleh kesimpulan, yaitu:

1. Aplikasi dapat melakukan analisis menggunakan API Twitter dengan mengambil data secara *realtime* berdasarkan 7 hari unggahan terakhir.
2. Pada penelitian ini metode XGBoost berhasil melakukan klasifikasi 300 data tweet dengan nilai akurasi terbaik sebesar 83.5% dan 263 data tweet dengan nilai akurasi terbaik 89.3%.
3. Metode XGBoost sudah berhasil mengenali kalimat sentimen dalam berbagai kelas kata seperti kata sifat, kata kerja, kata benda yang menggambarkan perasaan pengguna. Namun metode ini masih belum dapat mengenali sentimen yang memiliki makna kiasan atau makna ganda.

Penelitian ini telah menghasilkan sebuah aplikasi untuk menganalisis layanan provider internet sesuai dengan kebutuhan pengguna. Namun hasil akurasi yang masih 80% untuk data uji masih perlu terus diperbaiki dan dikembangkan agar menghasilkan hasil yang lebih baik. Kalimat pengguna yang masih mengandung makna ganda atau bukan makna sebenarnya perlu diatasi agar tidak salah dianalisis. Pengenalan makna ganda perlu dikenali dengan berbagai cara menggunakan metode natural language untuk mendeteksi kata sebelum dan sesudahnya agar kalimat dikenali secara utuh.

#### Daftar Pustaka

- [1] APJII, “Laporan Survei Internet APJII 2019–2020 (Q2),” 2020.
- [2] S. Kemp, “DIGITAL 2021: INDONESIA,” *DataReportal*, Feb. 11, 2021.
- [3] A. O. Siagian, R. Martiwi, and N. Indra, “Kemajuan Pemasaran Produk Dalam Memanfaatkan Media Sosial Di Era Digital,” *Jurnal Pemasaran Kompetitif*, vol. 3, no. 3, p. 44, Jun. 2020, doi: 10.32493/jpkpk.v3i3.4497.
- [4] C. Bentéjac, A. Csörgö, and G. Martínez-Muñoz, “A comparative analysis of gradient boosting algorithms,” *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/s10462-020-09896-5.
- [5] D. A. Al-Qudah, A. M. Al-Zoubi, P. A. Castillo-Valdivieso, and H. Faris, “Sentiment Analysis for e-Payment Service Providers Using Evolutionary eXtreme Gradient Boosting,” *IEEE Access*, vol. 8, pp. 189930–189944, 2020, doi: 10.1109/ACCESS.2020.3032216.
- [6] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features,” vol. 1, no. 12, pp. 1725–1732, 2017. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] F. F. Abdulloh and I. R. Pambudi, “Analisis Sentimen Pengguna Youtube Terhadap Program Vaksin Covid-19,” *CSRID (Computer Science Research and Its Development Journal)*, vol. 13, no. 3, p. 141, Nov. 2021, doi: 10.22303/csrid.13.3.2021.141-148.
- [8] A. N. Rohman, R. D. Handayani, R. D. Y. P., and K. Kusriani, “DETEKSI EMOSI MEDIA SOSIAL MENGGUNAKAN TERM FREQUENCY- INVERSE DOCUMENT FREQUENCY,” *CSRID (Computer Science Research and Its Development Journal)*, vol. 11, no. 3, p. 140, Mar. 2021, doi: 10.22303/csrid.11.3.2019.140-148.
- [9] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” Mar. 2016, doi: 10.1145/2939672.2939785.
- [10] R. H. Hama Aziz and N. Dimililer, “SentiXGboost: enhanced sentiment analysis in social media posts with ensemble XGBoost classifier,” *Journal of the Chinese Institute of Engineers*, vol. 44, no. 6, pp. 562–572, Aug. 2021, doi: 10.1080/02533839.2021.1933598.