



# Analisis Sentimen Komentar Konsumen Industri Jamu di Media Sosial menggunakan *Artificial Neural Network* dan *K-Nearest Neighbor*

Daniel Kurniawan\*, Hindriyanto Dwi Purnomo, Ade Iriani

Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana

Naskah masuk: 21 November 2023; Diterima untuk publikasi: 14 Maret 2024  
DOI: 10.21456/vol14iss3pp210-223

## Abstract

Phytopharmaceutical plants have become one of the main commodities contributing significantly to the economy through their use in the pharmaceutical, cosmetic, and health industries. However, behind this economic potential, traditional herbal medicine businesses often face challenges, particularly in promotion and brand identity. Social media platforms like Instagram have now introduced unique features to support business and marketing, primarily by providing in-depth information about herbal products and offering opportunities for businesses to receive feedback from consumers. Comments on social media are valuable but often unstructured; hence, sentiment analysis is necessary to organize and categorize this data. By combining comment data with information from Google Trends, cause-and-effect relationships from comments during specific periods can be identified using path analysis. This research aims to analyze consumer comments on the Sidomuncul company's Instagram platform, with the hope of benefiting the company and advancing herbal medicine products. The methods used in this study include Artificial Neural Network (ANN) and K-nearest neighbor (KNN) to classify comments into positive, negative, and neutral categories. Both methods show satisfactory results in classification, with an average accuracy of 0.887 for ANN and 0.874 for KNN. However, the ROC curve for the KNN model indicates a relatively low AUC value in classifying negative comments, at 0.598.

**Keywords:** Sentiment Analysis; Artificial Neural Network; KNN; Path Analysis; Herbal Medicine Products

## Abstrak

Tanaman biofarmaka telah menjadi salah satu komoditas utama yang memberikan kontribusi signifikan terhadap perekonomian melalui penggunaannya dalam industri obat-obatan, kosmetik, dan kesehatan. Namun, dibalik potensi perekonomian tersebut usaha jamu tradisional sering menghadapi masalah dan salah satunya pada promosi dan identitas merek dagang. Platform media sosial seperti *Instagram* kini telah memperkenalkan fitur-fitur unik untuk mendukung bisnis dan pemasaran utamanya dalam memberikan informasi mendalam tentang produk jamu serta memberikan kesempatan bagi pelaku bisnis untuk menerima umpan balik dari konsumen. Komentar di media sosial memiliki nilai penting namun seringkali tidak terstruktur, oleh karena itu analisis sentimen diperlukan untuk mengorganisir dan mengkategorikan data tersebut. Dengan menggabungkan data komentar dengan informasi dari *Google Trends*, hubungan sebab-akibat dari komentar dalam periode tertentu dapat diidentifikasi menggunakan analisis jalur. Penelitian ini bertujuan untuk menganalisis komentar konsumen terhadap perusahaan Sidomuncul di platform media sosial *Instagram*, dengan harapan memberikan manfaat bagi perusahaan serta kemajuan produk obat herbal itu sendiri. Metode yang digunakan pada penelitian ini meliputi *Artificial Neural Network* (ANN) dan *K-Nearest Neighbor* (KNN) untuk mengklasifikasikan komentar menjadi komentar positif, negatif, dan netral. Kedua metode menunjukkan hasil yang memuaskan dalam mengklasifikasi, yakni rata-rata akurasi sebesar 0.887 pada metode ANN dan 0.874 pada metode KNN. Meski demikian, kurva ROC pada model KNN menunjukkan nilai AUC yang cukup rendah dalam mengklasifikasi komentar yang bernilai negatif, yakni 0.598.

**Kata kunci :** Analisis Sentimen; *Artificial Neural Network*; *K-Nearest Neighbor*; Analisis Jalur; Produk Obat Herbal

## 1. Pendahuluan

Tanaman biofarmaka khususnya yang dimanfaatkan untuk pengobatan sudah dikenal oleh

masyarakat lintas generasi tidak hanya di Indonesia tetapi juga di negara-negara lain seperti China, Thailand, India, dan Myanmar karena tanaman tersebut lebih terjangkau dan mudah ditemukan

\*) *Corresponding author:* 972020016@student.uksw.edu

(Wibowo *et al.*, 2019). Di Indonesia, tanaman biofarmaka menjadi salah satu komoditas besar dan banyak berkontribusi dalam perekonomian diantaranya tanaman jahe, kunyit, dan temulawak. Badan Pusat Statistik (BPS) menyebutkan tanaman biofarmaka sebagai tanaman yang bermanfaat untuk obat-obatan, kosmetik dan kesehatan yang dikonsumsi atau digunakan dari bagian tanaman yang berupa daun, bunga, buah dan umbi (rimpang) atau akar. Dari tahun ke tahun sejak 2016 hingga 2022 produksi tiga (3) komoditas tersebut mengalami pasang surut terlihat dari Tabel 1. mengenai produksi biofarmaka (dalam ribu ton) yang diambil dari katalog hortikultura BPS tahun 2022.

Tabel 1. Produksi Biofarmaka Indonesia tahun 2018-2022 (dalam ribu ton)

Biofarmaka	2019	2020	2021	2022
Jahe	174.38	183.52	307.24	247.45
kunyit	190.91	193.58	184.82	196.5
temulawak	29.64	26.74	32.28	28.1

Obat Herbal (OH) di Jawa secara garis besar dapat dibagi menjadi tiga kategori produk: jamu, OH terstandar, dan fitofarmaka (Nawiyanto, 2016; Amalia *and* Aprianingsih, 2017). Jenis tanaman untuk OH dan memiliki nilai kegunaan yang tinggi di wilayah kabupaten Sambas antara lain adalah sirih, kunyit kuning, sirsak, bawang merah, kalimao, jahe merah (Pranaka *et al.*, 2020). Di wilayah lain seperti Yogyakarta tepatnya di daerah Taman Nasional Gunung Merapi diidentifikasi 50 spesies tumbuhan obat dan 3 diantaranya memiliki aktivitas antioksidan yang tinggi (Nurwijayanto *et al.*, 2020). Di daerah Baturaden dan Subang terdapat 11 jenis tanaman yang juga dimanfaatkan sebagai OH untuk pengobatan diabetes mellitus (Utaminigrum *et al.*, 2020). Hal ini menjelaskan setiap wilayah di Indonesia memiliki potensi tanaman obat dan dapat dimanfaatkan untuk OH.

Beberapa manfaat yang bisa didapatkan dari mengkonsumsi tanaman obat antara lain untuk pengobatan kategori antibiotik, cacar, keremut, mencerdaskan otak, sakit telinga, usus buntu, demam, sakit perut, hipertensi, penyakit kulit, memperlancar BAK, dan masuk angin (Pranaka *et al.*, 2020). Beberapa tanaman obat yang dijadikan sebagai ramuan jamu juga dapat diberikan pada penderita batu saluran kemih yang mengalami perubahan lebih baik pada fungsi fisik, peranan fisik, rasa nyeri, kesehatan umum, fungsi sosial, energi, peranan emosi dan kesehatan jiwa setelah perawatan selama 8 minggu (Nisa *et al.*, 2021). Selain itu OH juga dipakai untuk pengobatan hipertensi (Amir *et al.*, 2022) hingga yang paling parah seperti gagal ginjal dan diabetes (Prabawa *and* Fitriani, 2020).

Semakin berkembangnya teknologi untuk memperoleh informasi dan faktor lain seperti daya ekonomi membuat masyarakat mengalami perubahan

kebiasaan untuk beralih dari pengobatan modern pada pengobatan tradisional dengan memanfaatkan OH (Siregar *et al.*, 2020). Pulau Jawa adalah tempat bagi banyak produsen obat-obatan tradisional, baik skala kecil, industri rumahan dan skala besar, industri berbasis perusahaan (Nawiyanto, 2016). Pola tataguna hutan rakyat dengan agroforestri herbal dapat memberikan sumbangan pendapatan pada rumah tangga petani baik jangka pendek maupun jangka panjang (Arinah *et al.*, 2021). Dalam skala industri kecil hingga besar OH dapat dikembangkan menjadi minuman *sachet* dengan bahan baku laos dan kunyit (Nuzuliyah, 2018). Pengolahan tanaman obat menjadi OH juga memiliki prospek dengan mengeksport dan memasarkannya secara internasional seperti negara Malaysia, Singapura, Brunei, Korea, Mongolia, Hongkong, Rusia, Arab Saudi, Australia, dan Amerika Serikat (Nawiyanto, 2016).

Pengolahan dan pemasaran OH tidak selalu mengalami jalan mulus tetapi juga muncul masalah-masalah yang perlu diperhatikan. Permasalahan yang terkait dengan pemasaran produk OH selain pada pendistribusian yaitu dalam membangun identitas atau merek dagang dan promosi (Prabawa *and* Fitriani, 2020). Promosi produk dengan menggunakan iklan berperan penting dan dapat mendorong konsumen untuk melakukan pembelian karena semakin menarik iklan, calon pembeli juga akan tergiur untuk membeli produk (Lorinda *and* Amron, 2023). Membangun bisnis baru dengan menggunakan produk jamu memang tidak mudah tapi dengan konsep *Product Service System*, pelaku bisnis dapat menciptakan proses interaksi langsung dengan konsumen (Amalia *and* Aprianingsih, 2017). Hal ini dirasa perlu karena calon pembeli juga akan mendapatkan informasi yang lebih mendalam mengenai produk jamu dan pelaku bisnis juga dapat memperoleh keuntungan dengan menerima umpan balik dari konsumennya. Disinilah peran media sosial (MS) mampu memfasilitasi hubungan antara pelaku bisnis dan konsumennya.

*Instagram* dan *Facebook* yang merupakan perusahaan yang bergerak di bidang teknologi baru-baru ini meluncurkan fitur unik yang dapat mendukung pelaku usaha baik kecil hingga besar (Herzallah *et al.*, 2022). Merk atau *brand* dapat ditingkatkan juga melalui MS melalui kata dari mulut MS untuk mendorong saling bercerita, membagikan ulasan, dan merujuk teman-teman lain di MS (Tuten *and* Perotti, 2019). Pengguna aktif MS salah satunya *Instagram* didominasi oleh perempuan dan yang sering melakukan pembelian adalah kalangan milenial (Muhammad *and* Hartono, 2021). Hal ini cukup menarik karena generasi muda bahkan pemilik status pendidikan tinggi kurang memiliki pengetahuan dan pengalaman dalam memanfaatkan tanaman obat untuk pengobatan penyakit (Utaminigrum *et al.*, 2020). Maka dari itu MS juga dapat memberikan fasilitas untuk melakukan edukasi

dan promosi untuk meningkatkan penggunaan produk jamu (Mustofa *et al.*, 2022). Adanya sebuah komunitas pecinta herbal juga akan memberikan dampak pada peningkatan kesehatan pada keluarga dan masyarakat (Adi *et al.*, 2023). *Instagram* menjadi salah satu *platform* penting untuk mengamati perilaku, tren, dan pola konsumsi dalam segmen masyarakat melalui fitur-fitur uniknya. Para peneliti dapat menggunakan *Instagram* untuk mengamati komunitas khusus dan mendapatkan wawasan dari interaksi di kolom komentar tentang topik tertentu, yang memungkinkan mereka mengumpulkan data dan memahami lebih dalam perilaku serta preferensi pengguna.

Pentingnya *Instagram* sebagai *platform* komunikasi sosial menekankan pentingnya memahami perasaan pengguna dalam kolom komentar. Oleh karena itu, penggunaan analisis sentimen menjadi penting untuk menganalisis dan memahami pandangan serta perasaan yang tercermin dalam konten *Instagram*. Teknik analisis sentimen telah banyak diaplikasikan dalam berbagai bidang seperti kesehatan, pendidikan, industri, pemasaran, manajemen layanan, dan lain sebagainya. *Twitter* sebagai salah satu MS yang dimanfaatkan oleh (Tüzemen *et al.*, 2023) untuk meninjau sikap pengguna *twitter* terhadap vaksin yang beredar pada masa pandemi. Data *twitter* tersedia dalam beragam jenis bahasa dan (Hameed *et al.*, 2023) menggunakan model AraBERT untuk membantu dalam menganalisis teks dalam bahasa arab. Studi Iddrisu *et al.*, (2023) menawarkan sebuah *framework* yang dapat dipadukan dengan pembelajaran mesin lainnya untuk mengenali komentar yang terkesan sarkasme pada industri penerbangan. Analisis sentimen telah memakai beberapa metode pembelajaran mesin seperti *naïve bayes*, *support vector machine*, knn dan menghasilkan keluaran yang bervariasi dari sumber opini publik yang tersedia (Mufidah *et al.*, 2022; Hidayat *et al.*, 2021; Faesal *et al.*, 2020; Muttaqien *et al.*, 2022).

Implementasi metode *Neural network* (NN) yang dilakukan oleh Wei, *et al.*, (2022) pada teknologi pencetakan 3D PolyJet, telah menunjukkan kemampuan NN dalam mengatasi tantangan yang kompleks dalam memperkirakan warna objek yang akan dicetak dengan tingkat akurasi yang lebih tinggi. NN juga didukung oleh fungsi optimasi yang dapat meningkatkan efisiensi pelatihan dengan cara adaptif untuk memodifikasi inisialisasi awal (Occhipinti *et al.*, 2022). Selain itu, metode lain yang umum digunakan untuk menganalisis sentimen adalah *K-Nearest Neighbor* (KNN), seperti yang dilakukan dalam pemanfaatan teknologi di kota cerdas Pekanbaru. Dalam penelitian oleh Anam *et al.* (2021), KNN digunakan untuk menganalisis komentar dan menjadikannya sebagai bahan evaluasi bagi pemerintah kota. Sun *et al.*, (2022) menyebutkan metode KNN dan NN memberikan

hasil terbaik pada proses pembelajaran dan pada data uji mengenai produksi biodiesel dengan minyak sayur.

Penelitian ini bertujuan untuk menganalisis komentar atau umpan balik dari konsumen perusahaan Sidomuncul pada MS *Instagram* melalui akun resmi produk penjualannya. Data sentimen kemudian diklasifikasikan dengan nilai positif, nilai negatif dan nilai netral. Data sentimen yang telah didapatkan kemudian dipadukan dengan data tren pada periode 2019 hingga 2022 untuk melihat korelasinya dengan memakai metode analisis jalur. Pada bagian kedua dipaparkan literatur mengenai sentimen analisis dan metode NN dan KNN yang digunakan. Alur dari proses sentimen analisis serta hasil pembahasan berupa perbandingan antara metode NN dan KNN ada di bagian ketiga dan keempat. Bagian kelima berisi kesimpulan yang diharapkan akan memberi manfaat dan kontribusi untuk perusahaan dan OH itu sendiri.

## 2. Kerangka Teoritis

### 2.1. Analisis Sentimen

Berbagai macam situs di internet kini banyak berkembang dan kaya akan informasi demikian halnya informasi pada MS yang cukup sulit untuk diolah secara manual. Analisis sentimen (AS) merupakan salah satu cara dalam proses mengolah dan menganalisis data teks yang tidak berstruktur seperti pada situs *Tripadvisor* sebagai situs penyedia layanan ulasan di bidang pariwisata (Morama *et al.*, 2022). Adapun situs bernama *twitter* yang menyediakan fasilitas bagi pengguna untuk saling bertukar pendapat secara *online*. *Twitter* memiliki sejumlah besar data mengenai perasaan atau pendapat pengguna mengenai topik vaksin yang muncul pada masa pandemi dan dengan menggunakan AS dapat memberikan gambaran umum pendapat mengenai vaksin Covid-19 yang telah beredar di beberapa negara (Tüzemen *et al.*, 2023). Mengingat pengguna internet yang beragam dan tidak hanya di Indonesia, untuk mengolah dan menganalisis teks juga perlu memerhatikan bahasa yang dipergunakan. Aspek bahasa menjadi penting karena kecenderungan literasi mengenai sentimen analisis mengarah pada teks dalam Bahasa Inggris dan literasi dalam bahasa lain seperti Bahasa Arab jarang ditemukan salah satunya karena bahasa tersebut juga sulit untuk dipelajari (Hameed *et al.*, 2023). Hasil keluaran dari AS terkadang sulit untuk dipahami apalagi bisa pendapat tersebut terkesan sarkasme sehingga perlu upaya awal dalam mengenali kalimat sarkasme (Iddrisu *et al.*, 2023). Alamoodi *et al.*, (2021) menjabarkan tahapan-tahapan umum dalam proses AS yakni sebagai berikut:

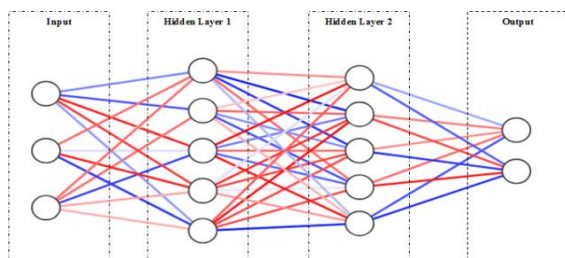
1) *Platform* MS: tahapan dimulai dengan memilih *platform* MS yang akan dipergunakan secara

online seperti Facebook, Twitter, Instagram, Youtube, Reddit, Wikipedia.

- 2) Pengumpulan data: tahapan untuk mengumpulkan data dengan kata kunci tertentu atau dengan hashtag untuk mendapat informasi seperti pada tweets, posts, teks berita atau kalimat.
- 3) Pre-processing: data yang diambil belum tentu seluruhnya data terstruktur sehingga tahapan ini berfokus untuk mempersiapkan data untuk fase berikutnya seperti ekstraksi fitur, tokenisasi, dan pembersihan data.
- 4) Analisis data: proses pengolahan data sesuai dengan tujuan tertentu seperti identifikasi polaritas, analisis sentimen, dan analisis frekuensi.
- 5) Hasil keluaran: keluaran dari AS dapat berupa evaluasi performa melalui skor *F-measure*, *Recall*, *Precision*, *Area Under ROC Curve*, *Receiver Operating Characteristic (ROC) Curve* (Iddrisu et al., 2023).

## 2.2. Artificial Neural Network

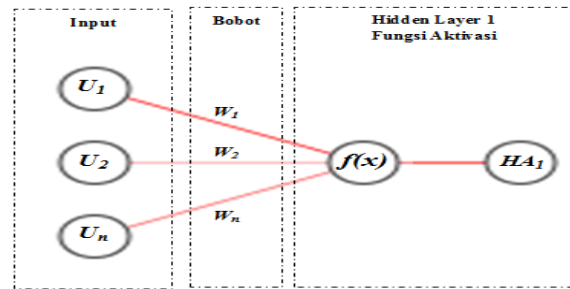
Artificial Neural Network (ANN) atau Neural Network (NN) merupakan percabangan dari Artificial Intelligence (AI) untuk memecahkan masalah pada sistem nonlinier dan Multilayer Perceptron (MLP) menjadi salah satu jenisnya (Gardner and Dorling, 1998). Selayaknya sistem neuron pada manusia, MLP terdiri dari serangkaian node yang terhubung satu sama lain dan diilustrasikan seperti pada Gambar 1. Setiap node menerima input dan terhubung dengan node lain sehingga membentuk sebuah jaringan. Setiap koneksi memiliki bobot masing-masing dan menghasilkan node output melalui sebuah fungsi aktivasi. Hal ini memungkinkan NN untuk memodelkan hubungan yang bersifat nonlinier antara variabel input dan respon output (Wei et al., 2022). Metode MLP memiliki struktur yang terdiri dari layer input, layer tersembunyi, dan layer output seperti yang diilustrasikan pada Gambar 1.



Gambar 1. Model Artificial Neural Network

Layer tersembunyi memiliki jumlah variasi yang berbeda dari hanya satu layer, dua layer seperti pada ilustrasi atau bahkan lebih banyak layer. MLP merupakan metode pembelajaran mesin yang terpadu sehingga diperlukan pelatihan dengan waktu yang cukup lama karena melibatkan banyak data latihan untuk memahami aturan klasifikasi (Lu et al., 1996).

Jumlah optimal node yang dipakai untuk setiap layer tersembunyi bisa bervariasi tergantung pada kompleksitas masalah yang akan dipecahkan. Bila jumlah node pada setiap lapisan terlalu kecil maka algoritma akan sulit memperoleh hasil klasifikasi yang baik dan semakin banyak jumlah node, kemampuan memproses informasi akan naik namun akan memperlambat proses pelatihan. Gambar 2 berikut ini adalah model sederhana dari MLP dengan memakai satu lapisan perceptron.



Gambar 2. Model sederhana MLP

Alur proses pelatihan pada metode MLP menggunakan suatu algoritma latihan dikenal dengan backpropagation. Tujuan pelatihan ini untuk mencari nilai bobot yang paling kecil menghasilkan kesalahan. Pelatihan diawali dengan menginisialisasikan nilai bobot koneksi ( $w_n$ ) antar node secara acak dengan nilai relatif kecil. Kemudian sinyal pengajar diberikan input ( $u_i$ ) dan output ( $d_i$ ) seperti pada persamaan (1) dan (2).

$$U_i = U_1, U_2, \dots, U_n \quad (1)$$

$$d_i = \begin{cases} +1 & u_i \in A \\ -1 & u_i \in B \end{cases} \quad (2)$$

$$y = f \left( \sum_{i=0}^n w_i u_i \right) \quad (3)$$

Nilai  $u_i$  dan  $w_i$  yang telah diberikan selanjutnya dilakukan perhitungan dengan persamaan (3), dimana  $y$  merupakan fungsi aktivasi yang diperoleh dari total jumlah hasil kali  $u_i$  dengan  $w_i$  dan akan dijadikan perbandingan dengan target atau keluaran yang sebenarnya. Bilamana hasil fungsi aktivasi tidak sesuai target, maka dilakukan penyesuaian terhadap  $w_i$  hingga diperoleh hasil yang sama atau yang paling mendekati, baru kemudian pelatihan dapat diakhiri (Liu, 2022).

Tugas yang dapat dilakukan pada model dapat dipakai pada permasalahan yang cukup kompleks misalkan seperti meramalkan jumlah tersangka penyalahgunaan narkoba pada studi (Pratiwi et al. 2019), memprediksi indeks harga komoditas makanan oleh (Zahara and Sugianto, 2021), dan

memprediksi curah hujan setiap bulannya (Sunardi *et al.*, 2020).

### 2.3. K-Nearest Neighbor

K-Nearest Neighbor (KNN) adalah algoritma pembelajaran mesin yang melakukan prediksi hubungan antara data yang tidak terlihat dan data yang telah ada sebelumnya, dan berdasarkan prediksi tersebut, mengklasifikasikan data baru ke dalam kategori yang sudah ada dan paling sesuai (Sun *et al.*, 2022). Dengan kata lain, sebuah data akan mempunyai kategori yang mirip dengan data atau tetangga lain di daerah sekitarnya (Kurniawan and Saputra, 2019).

Salah satu tantangan utama pada metode KNN terdapat dalam menentukan nilai  $k$  yang optimal, karena ini dapat menghasilkan perubahan besar dalam hasil. Selain itu, perhitungan jarak *Euclidean* antara setiap titik data dalam dataset pelatihan dapat memerlukan sumber daya komputasi yang signifikan. Meski demikian KNN memiliki kemudahan dalam hal penerapannya pada berbagai masalah tanpa banyak kesulitan. Algoritma ini juga cenderung toleran terhadap gangguan dalam data yang digunakan untuk pelatihan. Kecepatan eksekusi algoritma K-NN yang baik, dan interpretasi hasil yang mudah, menjadikannya efektif bahkan ketika diterapkan pada dataset yang cukup besar (Bansal *et al.*, 2022). Dari kelebihan dan kekurangan tersebut metode KNN dipilih. Metode KNN cukup umum digunakan pada bidang lain misalnya untuk kemajuan bidang pendidikan baik dari jenjang menengah (Kurniawan and Saputra, 2019) hingga jenjang perguruan tinggi (Hozairi *et al.*, 2021). Pada umumnya langkah awal untuk menggunakan metode KNN adalah dengan menentukan nilai  $k$  atau jumlah tetangga terdekat dari titik tengahnya. Kemudian dilanjutkan dengan mengukur seberapa jauh jarak antar data latih menggunakan fungsi jarak seperti fungsi *Euclidean* atau fungsi *Manhattan*. Persamaan fungsi *Euclidean* (4) dan *Manhattan* (5), berikut ini menunjukkan bagaimana jarak antara data  $x_i$  dan  $y_i$  dalam data latih.

$$Euc_{Distance} = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (4)$$

$$Man_{Distance} = \sum_{i=1}^k |x_i - y_i| \quad (5)$$

### 2.4. Evaluasi Model

Evaluasi model adalah proses dalam mengevaluasi kemampuan model dari algoritma pembelajaran dalam memecahkan masalah atau mengelola data (Afrinanda *et al.*, 2023). Evaluasi pada model menggunakan *true positive* (TP), *false*

*positive* (FP), *true negative* (TN), *false negative* (FN) sebagai *confusion matrix* untuk mengukur performa algoritma pembelajaran mesin. Hasil dari *confusion matrix* dipergunakan dalam perhitungan untuk memperoleh *accuracy*, *precision*, *recall* dan skor F1 sebagai evaluasi performa algoritma pembelajaran yang diterapkan (Onita, 2023). Skor *accuracy* digunakan untuk mengukur kebenaran dari hasil prediksi sesuai dengan kategori yang diberikan dimana *accuracy* adalah hasil pembagian jumlah prediksi yang benar dengan jumlah seluruh prediksi (Rustam *et al.*, 2020). Skor dengan nilai 1 adalah skor tertinggi dan sebaliknya, nilai 0 adalah skor terendah. Rumus perhitungan *accuracy*, *precision*, *recall*, dan skor F1 ditunjukkan pada persamaan (6), (7), (8), (9) seperti berikut (Rustam *et al.*, 2020).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$precision = \frac{TP}{TP + FP} \quad (7)$$

$$recall = \frac{TP}{TP + FN} \quad (8)$$

$$skor\ F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (9)$$

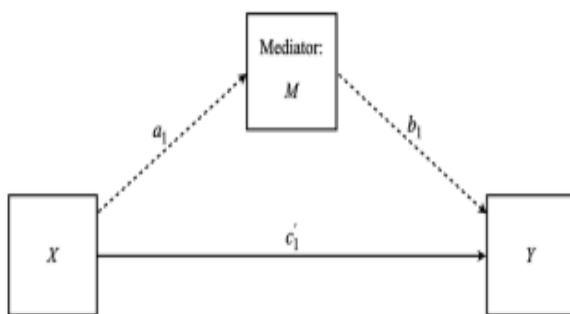
### 2.5. Google trends

*Google Trends* (GT) berperan untuk menganalisis contoh pencarian pada *platform Google* untuk menghitung jumlah pencarian yang terjadi dalam periode waktu tertentu. Data GT dapat dikumpulkan berdasarkan wilayah dan periode, dan disajikan sebagai data deret waktu untuk menunjukkan perubahan minat dan pendapat di kalangan pengguna internet (Wu *et al.*, 2019). Proses normalisasi data pencarian terhadap topik atau disebut *Index Google Trends* ditetapkan dengan nilai angka antara 0 hingga 100, dengan 0 mewakili volume pencarian yang sangat rendah atau tidak ada dalam hasil sedangkan 100 mewakili volume pencarian tertinggi pada sebuah topik (Cervellin *et al.*, 2017). GT dapat digunakan untuk memprediksi dan memantau suatu topik misalnya penyebaran penyakit flu (Mavragani and Ochoa, 2019). Namun, perlu diingat bahwa GT memiliki keterbatasan karena hanya mengandalkan volume pencarian dan kueri web, sehingga tidak dapat menjamin kebenaran informasi di web. Meskipun demikian, data GT memungkinkan opsi kueri yang terkait dengan fenomena atau topik tertentu, misalnya permintaan akan suatu produk (Bulczak, 2021). Fitur-fitur yang dimiliki GT adalah Beranda, Jelajahi, Sedang trending, Pencarian setahun, dan Langganan. Fitur pertama “Beranda” memberikan beberapa pilihan berita mengapa sebuah topik yang dicari sedang tren. Fitur “Jelajah” melakukan eksplorasi minat *online* pada sebuah kata

kunci dari pencarian lalu memberikan gambaran visual menurut periode waktu dan wilayah tertentu, dan topik berita terkait mengenai kata kunci. Fitur “Sedang trending” memperlihatkan topik berita yang sedang tren secara harian dan *real-time* pada wilayah pengguna GT. Fitur “Pencarian setahun” memperlihatkan beberapa kategori dan topik terkait yang sedang tren pada tahun dan wilayah secara khusus. Fitur “Langganan” memberikan layanan berbayar dengan melakukan pencarian topik atau hal-hal yang sedang tren di suatu wilayah secara harian atau mingguan kemudian disampaikan melalui surat elektronik.

### 2.6. Analisis Jalur

Analisis jalur termasuk dalam metode regresi dipakai dan dimanfaatkan untuk menguji korelasi dua atau lebih variabel yang menimbulkan hubungan sebab-akibat. Hasil pemodelan merupakan beberapa variabel independen yang memengaruhi beberapa variabel dependen yang dapat bersifat langsung ataupun tidak langsung (Maqbool and Sheikh, 2022). Gambar 3. merupakan contoh model analisis jalur dari Jollineau dan Bowen (2023) yang menunjukkan bagaimana hubungan antara variabel X (variabel independen) memengaruhi variabel Y (variabel dependen) secara langsung didukung dengan anak panah garis tidak putus. Terdapat juga hubungan tidak langsung antara variabel X dan variabel Y yang ditunjukkan oleh anak panah dengan garis putus yang melalui variabel M.

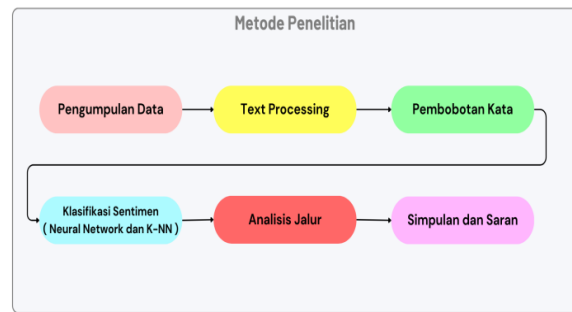


Gambar 3. Model sederhana analisis jalur

## 3. Metode

### 3.1. Alur Penelitian

Alur atau tahapan-tahapan yang dilakukan dalam penelitian ini digambarkan dengan menggunakan diagram alir dengan maksud agar lebih mudah dipahami mengenai langkah-langkah yang dilakukan. Pada Gambar 4. menunjukkan enam tahapan yang dilakukan dalam penelitian. Dalam mendukung tahapan-tahapan yang dilakukan akan digunakan beberapa perangkat lunak seperti *Orange Data Mining* (ODM) untuk mengolah data dan pembuatan model klasifikasi serta *RStudio* untuk membuat model dalam analisis jalur.



Gambar 4. Diagram Alur Penelitian

### 3.2. Pengumpulan Data

Data komentar diperoleh melalui *web scraping* atau ekstraksi data dari halaman *online* dan menyimpannya dalam bentuk file (Zhao, 2017). *Website* atau halaman *online* yang dimaksud adalah *MS Instagram* dengan pemilik akun resmi penjualan pada perusahaan Sidomuncul dan data dalam bentuk *time series* dari GT pada periode waktu lima (5) tahun terakhir sejak bulan Agustus 2018. Kedua data tersebut diproses dan diolah untuk memperoleh hubungan yang terkait antara komentar yang diterima oleh akun resmi penjualan perusahaan Sidomuncul dengan data tren pada periode bulan Januari 2018 hingga Desember 2022.

### 3.3. Text Processing

*Text Processing* adalah tahapan mula-mula yang dilakukan dalam proses *opinion mining* seperti data *tweet* yaitu dengan menyaring kata dalam komentar seperti entitas *tweet*, nama pengguna, tanda baca, angka, *link URL*, emotikon, dan objek lainnya yang tidak dipergunakan untuk proses pada *machine learning* (Faesal et al., 2020). Pada studi Parasati et al., (2020) dan Anam et al., (2021) yang pernah dilakukan, tahapan-tahapan dalam *text processing* terdiri:

- 1) Menghilangkan imbuhan pada kata dasar atau *Stemming*;
- 2) *Case folding* atau merubah teks dalam format huruf kecil;
- 3) Penghapusan karakter kosong atau *Whitespace*;
- 4) *stop removal*;
- 5) Penghapusan nomor pada data teks
- 6) Penghapusan tanda baca atau *punctuation*;
- 7) Pemecahan kalimat atau *tokenizing*.

### 3.4. Pembobotan Kata

TF-IDF merupakan metode yang digunakan untuk memberikan nilai pada kata kunci, dan membantu dalam identifikasi dan klasifikasi pada dokumen tertentu. TF-IDF merupakan kombinasi multiplikasi metode *Term Frequency* (TF) atau jumlah kata yang sering muncul pada dokumen dan metode *Inverse Document Frequency* (IDF) yang memberikan bobot yang lebih rendah untuk kata-kata yang sering muncul dan memberikan bobot yang

lebih besar untuk kata-kata yang jarang muncul (Qaiser and Ali, 2018).

### 3.5. Model dan Konfigurasi Klasifikasi Sentimen

Dalam penelitian, klasifikasi menggunakan metode *machine learning* yakni ANN dan KNN untuk memperoleh polaritas komentar positif, negatif, dan netral konsumen. Untuk memperoleh hasil klasifikasi yang optimal, data latih akan menggunakan beberapa model dengan perpaduan yang berbeda. Setiap model memiliki kombinasi dan variasi pada jumlah node, fungsi aktivasi, dan fungsi optimasi namun tetap memiliki dua layer tersembunyi yang ditunjukkan Tabel 2.

Tabel 2. Konfigurasi model ANN

Model	Hidden Layer	Jumlah Node	Fungsi Aktivasi	Fungsi Optimasi
ANN1	2	100	Identity	Adam
ANN2	2	100	Identity	L-BFGS-B
ANN3	2	100	ReLU	Adam
ANN4	2	100	ReLU	L-BFGS-B
ANN5	2	200	Identity	Adam
ANN6	2	200	Identity	L-BFGS-B
ANN7	2	200	ReLU	Adam
ANN8	2	200	ReLU	L-BFGS-B

Kombinasi node terbaik akan dipakai dalam melakukan klasifikasi pada seluruh data komentar yang telah dikumpulkan. Hal sama juga dilakukan pada pengaturan jumlah  $k$ , matriks, dan bobot metode KNN terhadap data latih yang ditunjukkan Tabel 3.

Tabel 3. Konfigurasi model KNN

Model	Jumlah $k$	Matriks	Bobot
KNN1	3	Euclidean	Distance
KNN2	3	Euclidean	Uniform
KNN3	3	Manhattan	Distance
KNN4	3	Manhattan	Uniform
KNN5	5	Euclidean	Distance
KNN6	5	Euclidean	Uniform
KNN7	5	Manhattan	Distance
KNN8	5	Manhattan	Uniform

## 4. Hasil dan Pembahasan

### 4.1. Data Penelitian

Data penelitian diperoleh dari proses *web scraping* pada akun resmi penjualan perusahaan Sidomuncul dan GT ditunjukkan oleh Tabel 4. dan Gambar 5.

Tabel 4. Komentar pada posting akun penjualan perusahaan.

No	Komentar
1	bisa dikonsumsi ibu menyusui?
2	@ummuakbarazka Halo. Produk jamu serbuk Gemuk Sehat ini tidak kami sarankan untuk dikonsumsi oleh ibu menyusui. Salam.
3	ada efek samping tidak?
4	@ummuakbarazka Halo. Produk jamu Gemuk Sehat memberi manfaat untuk melancarkan pencernaan dan meningkatkan nafsu makan. Namun cukup dikonsumsi selama 2 bulan saja, setelahnya dijeda agak lama baru dapat dikonsumsi kembali. Salam.

No	Komentar
5	bisa dibeli dimana?
6	@ummuakbarazka halo. Silakan pesan melalui <a href="http://www.sidomunculstore.com">www.sidomunculstore.com</a> ya. Salam.
7	Semoga dimurahkan rezeki
8	Ini berapa mbak ... saya gejala diabetes
9	Jasa renovasi interior dan exterior. Tolong hubungi. Terima kasih
10	Berapa mbak?

Komentar yang dikumpulkan melalui akun resmi penjualan berjumlah 1906 komentar yang memiliki kategori berbeda antara komentar positif yang menunjukkan kata kunci mengajak, *mention*, rasa terima kasih, dukungan, doa serta rasa penasaran komentator, komentar negatif yang menunjukkan kata kunci yang buruk pada produk atau layanan, dan komentar netral yang diberikan oleh administrator atau komentator yang memberikan informasi. Data komentar masih bersifat tidak seimbang sehingga perlu dilakukan proses penyeimbangan atau dikenal dengan *resampling*. Metode *resampling* akan menyesuaikan dataset untuk mengatasi ketidakseimbangan pelabelan dan SMOTE merupakan salah satunya jenis metode *resampling* (Padurariu and Breaban, 2019). Proses SMOTE dilakukan pada tahapan *pre-processing* yakni setelah pembobotan kata.

Gambar 5 menunjukkan fluktuasi tren yang terjadi pada kurun waktu 5 tahun terakhir mengenai topik penelitian. *Index Google Trends* memberikan skor 0 untuk tren pencarian terendah dan skor 100 untuk tren pencarian tertinggi pada waktu tertentu (Mavragani and Ochoa, 2019). Dari kurva tersebut tampak bahwa periode tren pencarian mencapai skor tertinggi ada pada sekitar bulan Juli - Agustus 2021 dan skor tren pencarian terendah ditemukan pada periode bulan Agustus 2018.



Gambar 5. Data Google Trends periode 2018 – 2023

### 4.2. Text Processing

Data komentar yang akan dipakai untuk model klasifikasi maupun prediksi akan melalui tahap *text processing* yaitu dengan menghilangkan daftar kata *stopword*, tanda baca, mengubah setiap kata menjadi huruf kecil, dan *stemming* (Ramadhani and Fajarianto, 2020). Dalam tahapan ini data komentar yang berasal dari Bahasa Inggris seperti *thanks* atau *your welcome* telah dirubah dalam bentuk Bahasa Indonesia dan beberapa kata yang disingkat telah dirubah ke bentuk awal. Proses *text processing* yang dilakukan adalah mengubah setiap kata menjadi bentuk huruf kecil atau disebut *case folding*,

menghilangkan kata yang bersifat umum seperti “dan, atau, ke, di, dan sebagainya” atau disebut *stopwording*, menghilangkan kata yang merujuk sebagai alamat *website*, menghilangkan tanda baca atau *punctuation*, menghilangkan nomor, menghilangkan tagar, dan melakukan pemecahan kalimat dalam kata per kata atau *tokenization*.

#### 4.3. Hasil Pembobotan

Pembobotan atau pemberian nilai pada setiap kata dalam teks dilakukan dengan mempertimbangkan frekuensi dan pentingnya kata tersebut dalam sebuah teks atau dokumen (Afrinanda *et al.*, 2023). Pembobotan akan menggunakan metode TF-IDF sebagai algoritma yang efisien dan mudah untuk mencari kata-kata dalam sebuah dokumen (Qaiser and Ali, 2018). Cuplikan hasil yang diperoleh dalam pembobotan kata pada data komentar ditunjukkan pada Tabel 5.

Tabel 5. Pembobotan Komentar

Komentar	Pembobotan
admin punya agen atau distributor wilayah manado Sulawesi Utara	admin=3.49232, agen=5.94332, distributor=5.60685, utara=5.761, manado=7.55276, sulawesi=6.85961, wilayah=6.45415
Mantap ini Coba buka ah dari HP	ah=5.25227, buka=3.38047, coba=5.94542, hp=5.94542, mantap=4.84681
Sidomuncul tidak pernah kecewa sama produknya tapi saya cukup kecewa dengan pelayanan karyawannya	karyawannya=6.85961, kecewa=12.9083, pelayanan=6.45415, produknya=5.15487, sidomuncul=1.89677
Ini harganya berapa Tutup botolnya susah dibuka	harganya=4.71955, botolnya=6.16647, dibuka=6.45415, tutup=5.47332, susah=5.94332

Seperti yang telah disebutkan sebelumnya, proses SMOTE dilakukan pada tahap *pre-processing* setelah tahap pembobotan. Data latih yang sebelumnya sudah dibagi yakni 80% dari jumlah keseluruhan data kemudian disubstitusi dengan proses SMOTE untuk menangani ketidakseimbangan data. Hasil dari tahapan ini adalah bahwa setiap komentar positif, komentar negatif, memiliki jumlah yang sama, yaitu 1257 komentar. Sehingga model klasifikasi yang dihasilkan akan menangani sejumlah 3771 komentar.

#### 4.4. Model Klasifikasi

Pada penelitian ini digunakan dua metode klasifikasi yaitu ANN dan KNN. Salah satu hal yang perlu diperhatikan metode ANN adalah mengenai jumlah node atau neuron pada setiap layer tersembunyi. Jumlah node yang dibutuhkan akan

tergantung pada kompleksitas data, jumlah *noise*, dan kuantitas data pelatihan. Jika node dalam setiap lapisan terlalu kecil, maka algoritma akan gagal mencapai nilai minimal dalam pelatihan tetapi jika terlalu banyak kinerja juga akan menjadi buruk (Gardner and Dorling, 1998). Pengaturan yang pertama dilakukan yakni membuat beberapa model kemudian melakukan inialisasi kombinasi parameter seperti jumlah node pada layer tersembunyi, fungsi aktivasi, dan fungsi optimasi. Dari variasi model yang telah dibuat, selanjutnya dilakukan tahapan pelatihan dan validasi menggunakan metode *10-folds cross validation*. Dari perbandingan beberapa model kombinasi yang dibuat, hasil optimal terdapat pada model ANN8 dengan kombinasi *neuron* 200 pada layer tersembunyi pertama, 200 pada layer tersembunyi kedua, fungsi aktivasi ReLU, dan fungsi optimasi L-BFGS-B. Hasil pelatihan dan validasi data latih dengan model ANN8 diperoleh rata-rata akurasi sebesar 0,977. Hasil pelatihan dan validasi pada model yang lain ditunjukkan di Tabel 6.

Tabel 6. Hasil data latih model ANN

Model	Data Latih	
	AUC	Acc
ANN1	0.993	0.971
ANN2	0.991	0.976
ANN3	0.994	0.969
ANN4	0.988	0.973
ANN5	0.994	0.972
ANN6	0.990	0.973
ANN7	0.994	0.974
ANN8	0.990	0.977

Metode KNN dimana  $k$  dalam KNN mengacu pada jumlah tetangga dari titik data baru dan menentukan nilai yang sesuai dengan  $k$  menjadi proses yang terpenting (Kurniawan and Saputra, 2019). Hasil pelatihan data latih dan validasi dengan *10-folds cross validation* pada variasi model dengan jumlah  $k$ , fungsi jarak antara *Euclidean* atau *Manhattan* dan perhitungan bobot jarak antara tipe *uniform* dan *distance* ditunjukkan pula pada Tabel 7. Dari perbandingan data latih pada Tabel 7, hasil optimal didapatkan pada model KNN3 dengan jumlah  $k$  yaitu 3, fungsi jarak *Manhattan* dan bobot yang dipakai adalah tipe *Distance*. Nilai akurasi rata-rata hasil pelatihan dan validasi pada model KNN3 adalah 0,967.

Tabel 7. Hasil data latih model KNN

Model	Data Latih	
	AUC	Acc
KNN1	0.983	0.953
KNN2	0.982	0.950
KNN3	0.987	0.967
KNN4	0.987	0.964
KNN5	0.988	0.947
KNN6	0.987	0.941
KNN7	0.990	0.965
KNN8	0.989	0.960



Melalui kedua tabel yang telah disajikan yakni Tabel 6. dan Tabel 7. kemudian diambil masing-masing satu model yang memiliki akurasi paling tinggi untuk dijadikan sebagai pengaturan *default* pada proses klasifikasi. Model pertama adalah metode ANN yang terdiri dari dua layer tersembunyi dengan jumlah neuron masing-masing 200 serta fungsi aktivasi *ReLU* dan optimasi L-BFGS-B. Jumlah *k* yang dipakai pada model KNN adalah 3 dengan fungsi jarak *Manhattan* dan tipe bobot *Distance*. Hasil evaluasi dari masing-masing model seperti *accuracy*, *recall*, *precision*, dan F1 kemudian dikomparasikan seperti yang ditunjukkan pada Tabel 8. Model ANN dan KNN dalam Tabel 8. menunjukkan hasil pelatihan yang tidak terlalu jauh berbeda diantara kedua model dimana rata-rata akurasi model ANN8 sebesar 0,977 dan skor AUC sebesar 0,990 sedangkan skor rata-rata akurasi model KNN3 sebesar 0,967 dan skor AUC sebesar 0,987.

Tabel 8. Evaluasi Model

Model	AUC	Acc	F1	Precision	Recall
ANN8	0.990	0.977	0.977	0.978	0.977
KNN3	0.987	0.967	0.967	0.967	0.967

Evaluasi hasil prediksi terhadap data latih untuk model ANN8 juga bisa dilihat menggunakan *confusion matrix* seperti pada Gambar 6. Hasil penyeimbangan data dengan metode SMOTE didapati 3771 komentar dan hasil dari data latih didapatkan 1255 komentar adalah *True Negatif*, 1247 komentar adalah *True Netral*, 1182 komentar adalah *True Positif*, 11 komentar *False Negatif*, 65 komentar *False Netral*, serta 11 komentar *False Positif*.

		Predicted			Σ
		Negatif	Netral	Positif	
Actual	Negatif	1255	0	2	1257
	Netral	1	1247	9	1257
	Positif	10	65	1182	1257
Σ		1266	1312	1193	3771

Gambar 6. *Confusion Matrix* model ANN8

*Confusion matrix* untuk model KNN3 ditunjukkan pada Gambar 7. dimana dari 3771 komentar untuk data latih didapatkan 1257 komentar merupakan *True Negatif*, 1205 komentar *True Netral*, 1183 komentar *True Positif*, 8 komentar *False Negatif*, 66 komentar *False Netral* dan 52 komentar *False Positif*.

		Predicted			Σ
		Negatif	Netral	Positif	
Actual	Negatif	1257	0	0	1257
	Netral	0	1205	52	1257
	Positif	8	66	1183	1257
Σ		1265	1271	1235	3771

Gambar 7. *Confusion Matrix* model KNN3

Dari *confusion matrix* yang telah ditunjukkan dari Gambar 6. dan Gambar 7., ketika model ANN8 dipakai dalam klasifikasi maka model mengalami kesulitan dalam memprediksi komentar netral. Model KNN3 yang dipakai untuk klasifikasi komentar juga mengalami kesulitan dalam memprediksi komentar netral dan positif.

Model yang telah dibuat dengan menggunakan data latih kemudian diimplementasikan pada data uji untuk dilakukan proses prediksi dalam menangani komentar dengan kategori komentar positif, negatif, dan netral. Komentar dengan kata kunci yang memiliki kesan mendukung, memberi doa, mengajak, perasaan penasaran atau melakukan pujian menjadi target untuk kelas positif. Komentar dengan kata kunci yang mengandung kesan kekecewaan pada layanan, produk, stok atau isu menjadi target untuk kelas negatif. Komentar yang mengandung kata kunci tanggapan jawaban atau spam ditargetkan untuk kelas netral. Hasil klasifikasi terhadap data uji dengan menggunakan model ANN8 didapati sebanyak 310 data terklasifikasi komentar positif, 3 komentar negatif, dan 68 komentar netral. Data uji model KNN3 diklasifikasi dan diperoleh komentar positif sejumlah 306 komentar, komentar negatif sebanyak 3 komentar, dan komentar netral sebanyak 72 komentar. Dataset yang digunakan pada data uji memperlihatkan bahwa komentar didominasi oleh komentar positif. Hal ini menunjukkan bahwa konten promosi yang dilakukan akun penjualan resmi “@sidomunculstore” mampu menarik perhatian dan respon yang positif. Hal ini penting karena upaya dan aktivitas promosi dapat memberikan citra yang baik bagi perusahaan (Al Saed *et al.*, 2020).

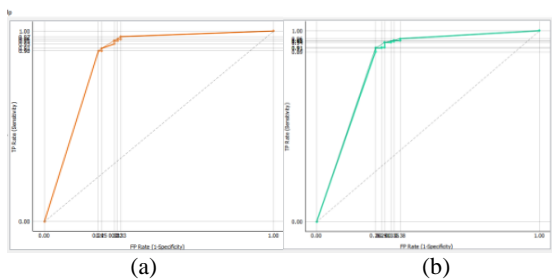
Tabel 9. memuat perbandingan antara skor akurasi label positif, negatif, dan netral dari hasil pelatihan data latih dan hasil prediksi data uji dengan menggunakan model ANN8 dan KNN3. Berdasarkan Tabel 9. didapatkan bahwa skor akurasi pada pelatihan data latih tidak selalu diperoleh skor akurasi yang sama dengan skor akurasi prediksi pada data uji. Seperti pada pelatihan klasifikasi komentar positif model ANN8 didapatkan skor akurasi 0,977 tetapi setelah model digunakan pada data uji akurasi menunjukkan skor 0,887.

Tabel 9. Perbandingan hasil klasifikasi data latih dan data uji

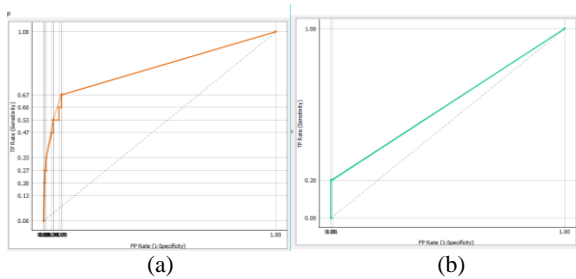
Klasifikasi	Evaluasi	Data Latih		Data Uji	
		ANN8	KNN3	ANN8	KNN3
Positif	AUC	0.985	0.979	0.853	0.833
	Acc	0.977	0.967	0.887	0.874
Negatif	AUC	0.998	0.999	0.803	0.598
	Acc	0.997	0.998	0.887	0.874
Netral	AUC	0.988	0.983	0.903	0.889
	Acc	0.980	0.969	0.887	0.874

Untuk mendukung perbandingan hasil data uji dalam melakukan klasifikasi dengan model ANN8 dan KNN3, berikut adalah kurva ROC yang digunakan untuk menunjukkan sejauh mana

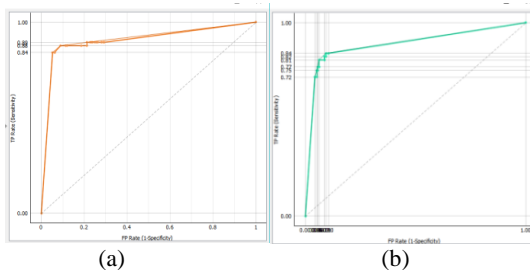
kemampuan model dalam memprediksi sebuah kelas. Evaluasi model melalui kurva ROC dengan kisaran penilaian dari 0,7 hingga 1.0 menunjukkan tingkat performa dari cukup hingga sangat baik (Afrinanda *et al.*, 2023). Gambar 8., Gambar 9., dan Gambar 10. merupakan grafik kurva ROC yang diperoleh dari data uji dimana gambar bagian (a) mewakili model ANN8 dan gambar bagian (b) mewakili model KNN3. Gambar 8. merupakan kurva ROC untuk prediksi kelas positif, Gambar 9. adalah kurva ROC untuk kelas negatif, dan Gambar 10. adalah kurva ROC untuk kelas netral. Dari ketiga kurva ROC tersebut model ANN8 dan KNN3 sama-sama baik dalam mengenali kelas positif, dan kelas netral namun model KNN3 kurang dalam menangani komentar dengan kelas negatif.



Gambar 8. Kurva ROC untuk kelas positif (a) ANN, (b) KNN



Gambar 9. Kurva ROC untuk kelas negatif (a) ANN, (b) KNN



Gambar 10. Kurva ROC untuk kelas netral (a) ANN, (b) KNN

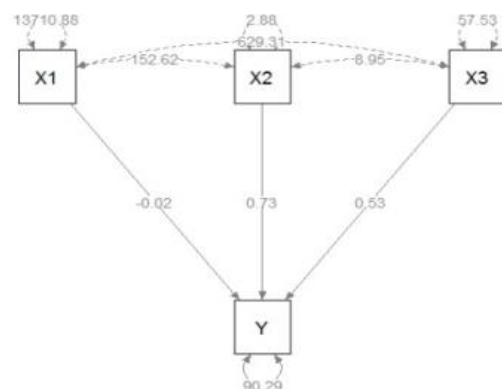
Model ANN8 dan KNN3 menunjukkan performa yang baik dalam klasifikasi terhadap data uji dilihat dari nilai akurasi rata-rata sebesar 0,887 pada model ANN8 dan 0,874 pada model KNN3. Meskipun rata-

rata akurasi menunjukkan model mampu melakukan klasifikasi dengan baik, namun kurva ROC pada model KNN3 menunjukkan skor AUC yang rendah yakni 0,598 dalam mengklasifikasi komentar negatif. Model KNN3 yang diimplementasikan pada komentar positif menunjukkan skor AUC 0,833 dan pada komentar netral menunjukkan skor 0,889. Model ANN8 yang diimplementasikan pada komentar positif diperoleh skor AUC 0,853, pada komentar negatif diperoleh skor 0,803, dan pada komentar netral diperoleh skor 0,903. Model ANN8 dan KNN3 menunjukkan bahwa model mampu mengklasifikasikan komentar dengan baik, tetapi model KNN3 gagal diimplementasikan pada komentar negatif (Afrinanda *et al.*, 2023).

#### 4.5. Analisis Jalur

Tahapan dilanjutkan dengan melakukan analisis hubungan antara komentar positif, negatif, dan netral dengan tren yang memanfaatkan data *time series* GT mengenai perusahaan Sidumuncul pada periode 2019 hingga 2022. Analisis jalur seperti pada (Mavragani and Ochoa, 2019) dipakai untuk menemukan hubungan sebab-akibat dan dengan menggunakan data hasil sentimen dipadukan dengan data GT diharapkan dapat terlihat perilaku dari konsumen OH pada MS *Instagram*. Data sentimen yang sudah dikelompokkan dari tahun 2019 hingga 2022 menunjukkan bahwa data komentar pada tahun 2020 memiliki jumlah komentar yang terbanyak dimana pada tahun tersebut adalah ketika masa pandemi masih berlangsung. Disaat seperti inilah konsumen aktif dalam MS *Instagram* dan waktu tersebut bisa dimanfaatkan perusahaan untuk melakukan promosi *online*.

Data sentimen komentar positif, negatif, dan netral kemudian diolah dengan fungsi regresi untuk memperoleh hubungan pada setiap variabel seperti diagram analisis jalur pada Gambar 11. Diagram menggambarkan hubungan antara variabel bebas X1 yaitu komentar positif, X2 yaitu komentar negatif, dan X3 yaitu komentar netral dengan Y yaitu tren dalam GT.



Gambar 11. Diagram Plot Analisis Jalur

Diagram pada Gambar 11. menunjukkan bahwa setiap variabel memiliki hubungan secara langsung dan tidak langsung. Skor koefisien X1 dengan Y adalah -0,2; skor koefisien X2 dengan Y sebesar 0,73; dan skor koefisien X3 dengan Y sebesar 0,53. Dari ketiga hubungan langsung antar variabel menunjukkan bahwa X2 yakni komentar negatif dan X3 yakni komentar netral menjadi variabel yang memengaruhi variabel Y atau tren pencarian GT. Hal tersebut menunjukkan bahwa data tren GT dari tahun 2019 hingga 2022 mengenai Sidomuncul dipengaruhi komentar negatif dan komentar netral.

## 5. Kesimpulan

Analisis sentimen sudah dilakukan terhadap umpan balik atau interaksi melalui komentar yang ada pada akun penjualan resmi *Instagram* milik perusahaan Sidomuncul. Ketersediaan data komentar memperlihatkan konsumen ikut terlibat dalam interaksi yang dilakukan oleh perusahaan. Metode yang digunakan pada penelitian yakni ANN dan KNN. Kedua metode menunjukkan performa klasifikasi yang baik dengan rata-rata akurasi mencapai 0,887 pada metode ANN dan 0,874 pada metode KNN. Meskipun begitu, evaluasi dengan kurva ROC menunjukkan performa metode KNN dalam mengidentifikasi komentar negatif tergolong rendah dengan skor AUC 0,598. Hasil klasifikasi data uji dengan metode ANN diperoleh komentar positif sebanyak 310 komentar, komentar negatif sebanyak 3, komentar netral sebanyak 68, dan dengan metode KNN diperoleh komentar positif sebanyak 306, komentar negatif sebanyak 3, serta komentar netral sebanyak 72. Komentar yang didominasi oleh komentar positif bukan berarti memberikan pengaruh terhadap tren. Hasil analisis jalur menunjukkan komentar negatif dan komentar netral yang memberikan pengaruh terhadap tren Sidomuncul di tahun 2019 hingga 2022.

## Daftar Pustaka

- Adi, L.T., Shaluhiah, Z., Widjanarko, B., 2023. Pemberdayaan Komunitas Herbal Medicine Class di Tangerang Selatan dalam Meningkatkan Kesehatan Keluarga. *Perilaku Dan Promosi Kesehatan: Indonesian Journal of Health Promotion and Behavior*, 5(1), 53-61. <https://doi.org/10.47034/ppk.v5i1.6849>
- Afrinanda, R., Efrizoni, L., Agustin, W., Rahmiati, 2023. Hybrid Model for Sentiment Analysis of Bitcoin Prices using Deep Learning Algorithm. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 22(2), 309-324. <https://doi.org/10.30812/matrik.v22i2.2640>
- Al Saed, R., Upadhyaya, A., Saleh, M.A., 2020. Role of Airline Promotion Activities in Destination Branding: Case of Dubai vis-à-vis Emirates Airline. *European Research on Management and Business Economics*, 26(3), 121-126. <https://doi.org/10.1016/j.edeen.2020.07.001>
- Alamoodi, A.H., Zaidan, B.B., Zaidan, A.A., Albahri, O.S., Mohammed, K.I., Malik, R.Q., Almahdi, E.M., Chyad, M.A., Tareq, Z., Albahri, A.S., Hameed, H., Alaa, M., 2021. Sentiment Analysis and its Applications in Fighting COVID-19 and Infectious Diseases: A Systematic Review. *Expert Systems with Applications*, 167, 114155. <https://doi.org/10.1016/j.eswa.2020.114155>
- Amalia, F. A., Aprianingsih, A., 2017. Business Model of Jamu as Indonesian Traditional Herbal Medicine in New Economy. *Asian Journal of Technology Management*, 10(1), 19-29. <https://doi.org/10.12695/ajtm.2017.10.1.3>
- Amir, A., Rantesigi, N., Agusrianto, A., 2022. Seduhan Bawang Putih Terhadap Penurunan Tekanan Darah pada Pasien Hipertensi: A Literature Review. *Poltekita: Jurnal Ilmu Kesehatan*, 16(1), 113-117. <https://doi.org/10.33860/jik.v16i1.685>
- Anam, M.K., Pikir, B.N., Firdaus, M.B., Erlinda, S., Agustin., 2021. Penerapan Naïve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen dan Pemerintah. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 21(1), 139-150. <https://doi.org/https://doi.org/10.30812/matrik.v21i1.1092>
- Arinah, H., Andayani, W., Purwanto, R.H., 2021. Analisis Finansial Hutan Rakyat Pola Agroforestri Herbal di Desa Gerbosari Kabupaten Kulon Progo. *Jurnal Ilmu Kehutanan*, 15(2), 137-146.
- Bansal, M., Goyal, A., Choudhary, A., 2022. A Comparative Analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in Machine Learning. *Decision Analytics Journal*, 3, 100071. <https://doi.org/10.1016/j.dajour.2022.100071>
- Bulczak, G.M., 2021. Use of Google Trends to Predict the Real Estate Market: Evidence from the United Kingdom. *International Real Estate Review*, 24(4), 613-631. <https://doi.org/10.53383/100332>
- Cervellin, G., Comelli, I., Lippi, G., 2017. Is Google Trends a Reliable Tool for Digital Epidemiology? Insights from Different Clinical Settings. *Journal of Epidemiology and Global Health*, 7(3), 185-189. <https://doi.org/10.1016/j.jegh.2017.06.001>
- Faesar, A., Muslim, A., Ruger, A.H., Kusriani, 2020. Sentimen Analisis pada Data Tweet Pengguna Twitter Terhadap Produk Penjualan Toko Online Menggunakan Metode K-Means. *MATRIK: Jurnal Manajemen, Teknik Informatika dan*

- Rekayasa Komputer*, 19(2), 207-213.  
<https://doi.org/https://doi.org/10.30812/matrik.v19i2.640>
- Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14-15), 2627-2636.  
[https://doi.org/10.1016/S1352-2310\(97\)00447-0](https://doi.org/10.1016/S1352-2310(97)00447-0)
- Hameed, R.A., Abed, W.J., Sadiq, A.T., 2023. Evaluation of Hotel Performance with Sentiment Analysis by Deep Learning Techniques. *International Journal of Interactive Mobile Technologies (IJIM)*, 17(9), 70-87.  
<https://doi.org/10.3991/ijim.v17i09.38755>
- Herzallah, D., Muñoz-Leiva, F., Liebana-Cabanillas, F., 2022. Drivers of Purchase Intention in Instagram Commerce. *Spanish Journal of Marketing-ESIC*, 26(3), 168-188.  
<http://dx.doi.org/10.1108/SJME-03-2022-0043>
- Hidayat, E.Y., Hardiansyah, R.W., Affandy, 2021. Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network. *Jurnal Nasional Teknologi dan Sistem Informasi*, 7(2), 108-118.  
<https://doi.org/10.25077/TEKNOSI.v7i2.2021.108-118>
- Hozairi, Anwari, Alim, S., 2021. Implementasi Orange Data Mining untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes. *Network Engineering Research Operation*, 6(2), 133-144.  
<https://doi.org/10.21107/nero.v6i2.237>
- Iddrisu, A.M., Mensah, S., Boafu, F., Yeluripati, G.R., Kudjo, P., 2023. A sentiment analysis framework to classify instances of sarcastic sentiments within the aviation sector. *International Journal of Information Management Data Insights*, 3(2), 100180.  
<https://doi.org/10.1016/j.ijime.2023.100180>
- Jollineau, S.J., Bowen, R.M., 2023. A Practical Guide to using Path Analysis: Mediation and Moderation in Accounting Research. *Journal of Financial Reporting*, 8(1), 11-40.  
<https://doi.org/10.2308/jfr-2021-004>
- Kurniawan, D., Saputra, A., Penerapan K-Nearest Neighbour dalam Penerimaan Peserta Didik dengan Sistem Zonasi. *Jurnal Sistem Informasi Bisnis*, 9(2), 212-219.  
<https://doi.org/10.21456/vol9iss2pp212-219>
- Liu, H., 2022. Optimal Selection of Control Parameters for Automatic Machining Based on BP Neural Network. *Energy Reports*, 8, 7016-7024.  
<https://doi.org/https://doi.org/10.1016/j.egy.2022.05.038>
- Lorinda, I.P., Amron, 2023. Pengaruh Kualitas Produk, Iklan dan Citra Merek terhadap Keputusan Pembelian Produk Tolak Angin Sidomuncul di Kota Semarang. *MBIA*, 22(1), 53-64.  
<https://doi.org/10.33557/mbia.v22i1.2238>
- Lu, H., Setiono, R., Liu, H., 1996. Effective Data Mining using Neural Networks. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 957-961.  
<https://doi.org/10.1109/69.553163>
- Maqbool, A., Sheikh, N.A., 2022. Impact of Financial Decisions on Firm Performance: Path Analysis Approach. *South Asian Journal of Management Sciences*, 16(1), 116-125.
- Mavragani, A., Ochoa, G., 2019. Google Trends in Infodemiology and Infection: Methodology Framework. *JMIR Public Health and Surveillance*, 5(2), 1-15.  
<https://doi.org/10.2196/13439>
- Morama, H.C., Ratnawati, D.E., Arwani, I., 2022. Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(4), 1702-1708.
- Mufidah, F.S., Winarno, S., Alzami, F., Udayanti, E.D., Sani, R.R., 2022. Analisis Sentimen Masyarakat Terhadap Layanan ShopeeFood Melalui Media Sosial Twitter dengan Algoritma Naïve Bayes Classifier. *Journal of Information System*, 7(1), 14-25.  
<https://doi.org/10.33633/joins.v7i1.5883>
- Muhammad, F., Hartono, S., 2021. Marketplace Analysis of Purchase Decision Factors for Instagram Social Media Users. *Journal of Intelligence Studies in Business*, 11(3), 42-56.
- Mustofa, F.I., Baiquni, F., Triyono, A., Wijayanti, E., Wahyono, S., 2022. Pengetahuan, Sikap dan Praktik Masyarakat dalam Penggunaan Jamu untuk Meningkatkan Daya Tahan Tubuh Selama Pandemi Covid-19 di Indonesia. *Jurnal Tumbuhan Obat Indonesia*, 15(1), 57-68.  
<https://doi.org/10.22435/jtoi.v15i1.6034>
- Muttaqien, D.D., Tibyani, Hartono, P.P., 2022. Implementasi Support Vector Machine pada Analisis Sentimen mengenai Bantuan Sosial di Era Pandemi Covid-19 pada Pengguna Twitter. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(1), 163-171.
- Nawiyanto., 2016. Modernizing Traditional Medicines in Java: Regulations, Production and Distribution Network. *Paramita: Historical Studies Journal*, 26(2), 119-133.  
<https://doi.org/10.15294/paramita.v26i2.7175>
- Nisa, U., Astana, P.R.W., Jannah, W.D.M., Triyono, A., Ardiyanto, D., Zulkarnain, Z., Fitriani, U., Novianto, F., 2021. Kualitas Hidup Pasien Batu Saluran Kemih Yang Menggunakan Ramuan Jamu Di Klinik Jejaring Saintifikasi Jamu. *Jurnal Tumbuhan Obat Indonesia*, 14(1), 87-98.  
<https://doi.org/10.22435/jtoi.v14i1.4365>

- Nurwijayanto, A., Na'iem, M., Syahbudin, A., Wahyuono, S., 2020. Eksplorasi Potensi Antioksidan Tumbuhan Obat dari Taman Nasional Gunung Merapi Yogyakarta. *Jurnal Tumbuhan Obat Indonesia*, 13(1), 25-31.
- Nuzuliyah, L., 2018. Analisis Nilai Tambah Produk Olahan Tanaman Rimpang (Added Value Analysis of Rhizome Product). *Jurnal Teknologi Dan Manajemen Agroindustri*, 7(1), 31-38. <https://doi.org/10.21776/ub.industria.2018.007.01.4>
- Occhipinti, A., Rogers, L., Angione, C., 2022. A pipeline and comparative study of 12 machine learning models for text classification. *Expert Systems with Applications*, 201, 117193. <https://doi.org/10.1016/j.eswa.2022.117193>
- Onita, D., 2023. Active Learning Based on Transfer Learning Techniques for Text Classification. *IEEE Access*, 11, 28751–28761. <https://doi.org/10.1109/ACCESS.2023.3260771>
- Padurariu, C., Breaban, M.E., 2019. Dealing with Data Imbalance in Text Classification. *Procedia Computer Science*, 159, 736-745. <https://doi.org/10.1016/j.procs.2019.09.229>
- Parasati, W., Bachtiar, F.A., Setiawan, N.Y., 2020. Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naive Bayes Classifier. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(4), 1090–1099
- Prabawa, H.W., Fitriani, A.D., 2020. Mempertahankan Eksistensi Jamu Tradisional melalui Perubahan Desain Pengemasan dan Pemasaran. *DEDIKASI: Community Service Reports*, 2(1), 35–46. <https://doi.org/10.20961/dedikasi.v2i1.35848>
- Pranaka, R.N., Yusro, F., Budiastutik, I., 2020. Pemanfaatan Tanaman Obat oleh Masyarakat Suku Melayu di Kabupaten Sambas. *Jurnal Tumbuhan Obat Indonesia*, 13(1), 1–24. <https://doi.org/10.22435/jtoi.v13i1.1887>
- Pratiwi, P.G., Putra, I.K.G.D, Putri, D.P.S., 2019. Peramalan Jumlah Tersangka Penyalahgunaan Narkoba Menggunakan Metode Multilayer Perceptron. *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)*, 7(2) 143-150. <https://doi.org/10.24843/JIM.2019.v07.i02.p06>
- Qaiser, S., Ali, R., 2018. Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*, 181(1), 25-29. <https://doi.org/10.5120/ijca2018917395>
- Ramadhani, N., Fajarianto, N., 2020. Sistem Informasi Evaluasi Perkuliahan dengan Sentimen Analisis Menggunakan Naive Bayes dan Smoothing Laplace. *Jurnal Sistem Informasi Bisnis*, 10(2), 228-234. <https://doi.org/10.21456/vol10iss2pp228-234>
- Rustam, F., Mehmood, A., Ahmad, M., Ullah, S., Khan, D.M., Choi, G.S., 2020. Classification of Shopify App User Reviews Using Novel Multi Text Features. *IEEE Access*, 8, 30234-30244. <https://doi.org/10.1109/ACCESS.2020.2972632>
- Soedarsono, D.K., Mohamad, B., Adamu, A.A., Pradita, K.A., 2020. Managing Digital Marketing Communication of Coffee Shop Using Instagram. *International Journal of Interactive Mobile Technologies (IJIM)*, 14(5), 108-118. <https://doi.org/10.3991/ijim.v14i05.13351>
- Sun, X., Opulencia, M.J.C., Alexandrovich, T.P., Khan, A., Algarni, M., Abdelrahman, A., 2022. Modeling and Optimization of Vegetable Oil Biodiesel Production with Heterogeneous Nano Catalytic Process: Multi-Layer Perceptron, Decision Regression Tree, and K-Nearest Neighbor Methods. *Environmental Technology and Innovation*, 27, 102794. <https://doi.org/10.1016/j.eti.2022.102794>
- Sunardi, Yudhana, A., Muflih, G.Z., Sistem Prediksi Curah Hujan Bulanan Menggunakan Jaringan Saraf Tiruan Backpropagation. *Jurnal Sistem Informasi Bisnis*, 10(2), 155-162. <https://doi.org/10.21456/vol10iss2pp155-162>
- Siregar, R.S., Hadiguna, R.A., Kamil, I., Nazir, N., 2020. Permintaan dan Penawaran Tanaman Obat Tradisional di Provinsi Sumatera Utara. *Jurnal Tumbuhan Obat Indonesia*, 13(1), 50-60.
- Tuten, T., Perotti, V., 2019. Lies, Brands and Social Media. *Qualitative Market Research: An International Journal*, 22(1), 5-13. <http://dx.doi.org/10.1108/QMR-02-2017-0063>
- Tüzemen, S., Barış-Tüzemen, Ö., Çelik, A.K., 2023. Sentiment Analysis and Machine Learning Classification of COVID-19 Vaccine Tweets: Vaccination in the Shadow of Fear-Trust Dilemma. *Informatica*, 47(1), 73-82. <https://doi.org/10.31449/inf.v47i1.4055>
- Utamingrum, W., Nofrianti, Hartanti, D., 2020. Ethnomedicinal Survey of Traditional Antidiabetic Plants in Baturradan and Sumbang. *Medisains*, 18(2), 43-51. <http://dx.doi.org/10.30595/medisains.v18i2.7169>
- Wei, X., Zou, N., Zeng, L., Pei, Z., 2022. PolyJet 3D printing: Predicting color by multilayer perceptron neural network. *Annals of 3D Printed Medicine*, 5, 100049. <https://doi.org/https://doi.org/10.1016/j.stlm.2022.100049>
- Wibowo, A., Diniawati, E., Sein, T.T., 2019. Analysis of Traditional Health Care in Three Primary Health Care in West Java Province, Indonesia, 2018. *Kesmas*, 14(1), 43-50. <https://doi.org/10.21109/kesmas.v14i1.2700>
- Wu, D., Xu, Z., Bach, S., 2019. Predicting and Forecasting Avocado Sales using Google Trends. *Journal of Marketing Analytics*, 11(4), 629-641.
- Zahara, S., Sugianto, 2021. Prediksi Indeks Harga

Konsumen Komoditas Makanan Berbasis Cloud Computing Menggunakan Multilayer Perceptron. *JOINTECS (Journal of Information Technology and Computer Science)*, 6(1), 21-28. <https://doi.org/10.31328/jointecs.v6i1.1702>

Zhao, B., 2017. Web Scraping. *Schintler, L., McNeely, C. (eds) Encyclopedia of Big Data. Springer, Cham*, 1-3. [https://doi.org/10.1007/978-3-319-32001-4\\_483-1](https://doi.org/10.1007/978-3-319-32001-4_483-1)