

# Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier

Sifa Melina Salsabila<sup>1)\*</sup>, Aang Alim Murtopo<sup>2)</sup>, Nurul Fadhillah<sup>3)</sup>

<sup>1)2)3)</sup>Teknik Informatika, STMIK YMI Tegal

<sup>1)</sup>[melina.sifa25@gmail.com](mailto:melina.sifa25@gmail.com), <sup>2)</sup>[aang.alim@gmail.com](mailto:aang.alim@gmail.com), <sup>3)</sup>[nurulfadillah18@gmail.com](mailto:nurulfadillah18@gmail.com)

## Abstrak :

Tokopedia adalah *e-commerce* dengan pengunjung situs web terbanyak, mencapai 157 juta kunjungan pada kuartal pertama tahun 2022. Meskipun menempati urutan pertama dan bintang 5 menerima nilai paling banyak, itu pasti ada yang memberikan komentar negatif. Untuk membuktikan analisis dan pengujian itu maka memerlukan metode analisis dan klasifikasi komentar pelanggan dalam beberapa sentimen positif dan negatif. Analisis ini menggunakan metode *Naïve Bayes Classifier* terhadap komentar pelanggan tentang aplikasi Tokopedia di *Playstore*. Performa yang dihasilkan dari hasil percobaan yang dilakukan oleh *rapid evaluator* pada 143 data pengujian menghasilkan nilai akurasi 95,10%, dengan *class recall: 91,46%* (*positive class: Positive*) sedangkan *class precision: 100.00%* (*positive class: Positive*) dan nilai *AUC: 0.999* (*positive class: Positive*)

**Kata kunci :** Analisis Sentimen, Pelanggan, Tokopedia, *Naïve Bayes Classifier*.

## PENDAHULUAN

Menurut Kementerian Dalam Negeri melalui data administrasi Umum Kedudukan Sipil dan jejaring sosial, penduduk Indonesia pada tahun 2021 akan mencapai 273 juta orang (Prof Zudan Arif Fakhrullah, 2022) Ini adalah pasar yang berkembang pesat bagi pengguna ponsel. Menurut perusahaan riset Data Reportal, pada Januari 2022, total dari telepon genggam yang terhubung mencapai 370,1 juta. (Rizal, 2022) Dengan angka yang telah disebutkan, maka bisa jadi Indonesia menduduki urutan ke-4 di dunia dengan pengguna ponsel teraktif setelah Negara China, India, dan Amerika Serikat.

Data reportal mengungkapkan angka pengguna perangkat seluler di Republik Indonesia kini telah mencapai sekitar 133,3% dari jumlah seluruh penduduk atau sekitar 370,1 juta jiwa, (Rizal, 2022) Dari data yang telah didapatkan menjadi kesempatan yang sangat bagus untuk perkembangan pengusaha dalam ataupun luar negeri, lalu setelahnya berusaha untuk mewujudkan dan mengembangkan aplikasi yang berguna di kehidupan sehari-hari. Untuk memenangkan pasar di dunia bisnis dalam *e-commerce*, pengusaha harus bersaing untuk mengadakan inovasi dan kreativitas yang baru. Tokopedia merupakan *e-commerce* lokal ternama yang telah mengeluarkan aplikasi yang bisa dibuka oleh penjual dan pembeli melalui aplikasi *mobile*. Tokopedia mengeluarkan aplikasi resmi untuk android guna untuk membuka kesempatan yang lebih luas Kembali. Langkah ini termasuk pembuktian bahwa Tokopedia termasuk berkembang dengan sangat cepat. (Avi Tejo Bhaskoro, 2014)

Penelitian ini memberikan solusi untuk menentukan persentase respon dan komentar pelanggan aplikasi berdasarkan respon pelanggan terhadap Tokopedia, dimana komentar tersebut akan menjadi masukan untuk meningkatkan kualitas layanan, kualitas produk dan operasional internal perusahaan. Penelitian ini memiliki keuntungan memberi keterangan pada pihak manajemen Tokopedia agar selalu konsisten dalam memberikan layanan terhadap pelanggan sehingga pelanggan tetap percaya, yang akhirnya bisa memberikan keuntungan yang sesuai dengan target yang telah ditetapkan.

Tujuan dari penelitian ini adalah mengetahui hasil *sentiment analysis* komentar masyarakat terhadap Tokopedia dan mengetahui hasil akurasi dari klasifikasi *sentiment analysis*.

## TINJAUAN PUSTAKA

Beberapa penelitian yang lalu terkait dengan *sentiment analysis*, diantaranya penelitian dari Yane Marita Febrianti, Indriati, dan Agus Wahyu Widodo pada Percobaannya menghasilkan nilai *precision*, nilai *recall*, dan hasil *accuracy* pada analisis sentimen tanpa tahap *preprocessing*. Hasil analisis yang baik ternyata dengan menggunakan tahap *preprocessing* yakni menghasilkan nilai *precision 76%* dan *0,76* dengan nilai *recall*. (Yane Marita Febrianti, 2018) Sedangkan menurut penelitian Santoso, Valonia & Virginia, Gloria & Lukito, Yuan,

\*Sifa Melina Salsabila<sup>1)</sup>, Aang Alim Murtopo<sup>2)</sup>, Nurul Fadhillah<sup>3)</sup>



penelitiannya dilakukan berdasarkan hasil pertimbangan pembicara yang mengarah pada analisis emosional, baik dari emosi positif, negatif dan netral yang telah diambil dari hasil pengisian angket oleh mahasiswa dengan metode SVM dan diperoleh hasil penelitian dari segi akurasi. Nilai tertinggi diperoleh dari kumpulan data dengan  $k = 6$  dan akurasi yang diperoleh sebesar 67,83%.(Santoso<sup>1</sup> et al., 2017)

### Analisis Sentiment

Fungsi analisis sentimen adalah menggolongkan tiap kutub teks pada berbagai sumber internet maupun media sosial yang berbentuk dokumen atau kalimat kemudian ditentukan apakah kata tersebut termasuk kategori positif, netral ataupun negatif. Analisis sentiment juga bisa menunjukkan emosi kesedihan, kegembiraan, atau kemarahan(Hirst, n.d.) Langkah pertama yang harus dilakukan selama proses dokumen adalah membagi satu karakter menjadi kata-kata atau token, yang biasa dikenal sebagai "encoding". Tokenisasi adalah proses memisahkan kalimat menjadi bagian kata demi kata. Bagian kata ini disebut token.(Mulia, 2021)

### Metode Naïve Bayes Classifier

*Naïve Bayes Classifier* (NBC) adalah metode dengan klasifikasi yang sesuai dengan teorma bayes, Pada metode *naïve bayes* pengklasifikasiannya berdasarkan probabilitas yang bisa jadi ada dimasa depan. Berikut ini persamaan (1) yang digunakan:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (1)$$

Keterangan:

- X = peubah acak
- C = kelas
- $P(C|X)$  = probabilitas C didasarkan pada kondisi X
- $P(X|C)$  = probabilitas X dibawah kondisi C
- $P(C)$  = probabilitas C
- $P(X)$  = probabilitas X

Perhitungan yang dibutuhkan dalam NBC yaitu probabilitas tiap kelas, persamaan (2) mencari probabilitas tiap kelas dinotasikan sebagai berikut:

$$P(C) = \frac{N_c}{N} \quad (2)$$

Keterangan:

- $P(C)$  = probabilitas C
- $N_c$  = jumlah total kelas C
- N = jumlah total keseluruhan kelas

### Tokopedia

Tokopedia adalah *e-commerce* dengan jumlah pengunjung situs *web* tertinggi, mencapai 157 juta pengunjung pada kuartal pertama tahun 2022 (Sfr, 2022). Namun peringkat tertinggi dan mendapat 5 bintang adalah nilai tertinggi, mungkin masih ada yang memberi umpan balik negatif dan memberikan umpan balik netral tentang aplikasi Tokopedia ini, faktor kepercayaan komunitas ini memainkan peran penting dalam pertumbuhan merek. Karena biasanya barang yang telah kita beli pada *e-commerce* bisa tidak sama seperti gambar yang diiklankan dan juga saat diterima dalam kondisi baik, keadaan rusak, bahkan tidak sesuai dengan yang telah dipilih. Mendaftar di Tokopedia mudah dan gratis. Setelah berjalan, [www.tokopedia.com](http://www.tokopedia.com) menjadi *online shop* dengan perkembangan yang begitu cepat di Indonesia meskipun terbilang sangat muda, dari segi jumlah member, jumlah penjual, dari jumlah produk hingga jumlah pembelian. . dan menjual penawaran setiap hari. (Arbaini, 2020)

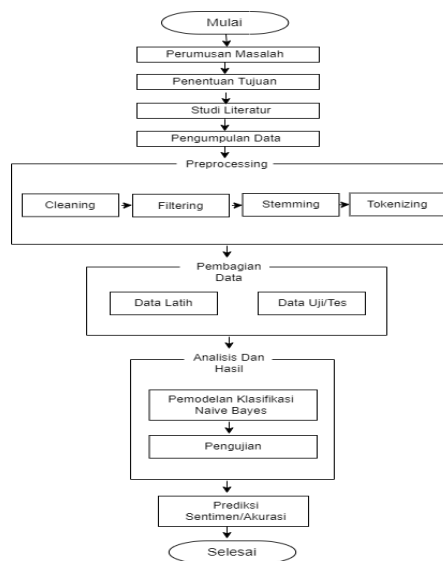
## METODE PENELITIAN

### Tahapan Penelitian

Tahapan penelitian adalah serangkaian bagan yang menggambarkan metode penelitian. Berikut adalah gambar dari tahapan penelitian yang digunakan.

\*Sifa Melina Salsabila<sup>1)</sup>, Aang Alim Murtopo<sup>2)</sup>, Nurul Fadhillah<sup>3)</sup>





Gambar 1. Tahapan Penelitian

Berdasarkan gambar diatas maka bisa dijelaskan sebagai berikut:

1. Perumusan Masalah

Pada perumusan masalah maka diperlukan untuk menentukan lingkungan permasalahan yang akan diteliti. Rumusan masalah merupakan cara untuk membuka atau menjawab suatu masalah.

2. Penentuan Tujuan

Fungsi dari penetapan tujuan adalah untuk lebih memperjelas apa tujuan dari penelitian ini. Tujuan dari penelitian ini adalah untuk mengetahui hasil analisis sentimen opini publik di Tokopedia dan untuk mengetahui hasil akurasi dari pengklasifikasian menggunakan metode *Naive Bayes Classifier*.

3. Studi Literatur

*Studi Literatur* digunakan untuk mengidentifikasi proses penelitian yang telah terjadi di bidang analisis sentimen, tinjauan pustaka ini juga dilakukan untuk meninjau penilaian yang ada dan digunakan untuk mengidentifikasi studi kasus dan metode digunakan penulis dalam proses penelitian.

4. Pengumpulan Data

Pengumpulan data ini diperoleh dari aplikasi Tokopedia di *Google Play Store*. Kemudian, penulis mengamati ulasan pelanggan terhadap aplikasi Tokopedia. Banyak sekali *review* yang diberikan oleh pelanggan aplikasi Tokopedia, untuk itulah penulis mempelajari rating *review* publik di *Google Play Store*.

5. Preprocessing

*Preprocessing* merupakan langkah penting untuk data dalam proses eksplorasi. Tujuan dari *preprocessing* adalah merubah data dari setiap sumber ke dalam format yang mudah dipahami dan juga mengatur format data agar lebih baik (Latifah, 2018) ada beberapa tahapan yang dilalui pada proses *preprocessing* diantaranya:

- a. **Cleaning.** digunakan untuk menghapus beberapa karakter yang bukan termasuk dari huruf abjad. Contohnya menghapus *emoticon*, menghapus karakter angka, *hastag*(#), *mention* (@), dan menghapus *URL* dan *link* yang ada pada komentar.
- b. **Filtering** menghapus tanda baca dan kata kunci yang telah diterapkan juga menghapus beberapa kata yang tidak penting yang telah ditetapkan sebelumnya.
- c. **Stemming.** yaitu merubah kata dasar dari kata yang berimbuhan.
- d. **Tokenizing.** yaitu memisah kata per kata dari spasi yang telah ditemukan.

6. Pembagian Data

Data akan dipisahkan menjadi 2 bagian yaitu: data *training* dan data *testing*. Dimana data *training* itu lebih besar dibandingkan dengan data *testing*.

7. analisis dan hasil

Pada tahap ini, metode yang digunakan untuk mengklasifikasian yaitu metode *Naive Bayes Classifier*. Gambaran dari klasifikasi ini berdasarkan teorema *Bayes* (statistik *Bayesian*) dengan asumsi independensi kuat. Aturan yang lebih rinci untuk basis model adalah "model memiliki sifat kebebasan". (Apriani et al., 2019).

\*Sifa Melina Salsabila<sup>1)</sup>, Aang Alim Murtopo<sup>2)</sup>, Nurul Fadhilah<sup>3)</sup>



## HASIL PENELITIAN DAN DISKUSI

### Pengelompokkan Data Training dan Data Testing

Mengelompokkan data yang akan digunakan yaitu data *training* dan data *testing*. Data tersebut dikumpulkan menggunakan teknik *scrapping* pada aplikasi Tokopedia di *Google Playstore*. Pertama adalah dengan membuka aplikasi Tokopedia di halaman *Google Playstore* lalu menarik data dari komentar pelanggan aplikasi tersebut menggunakan *Google Colab*. Untuk data *training*, 80% dari 714 komentar dipilih secara acak, kemudian ada 411 komentar positif dan 303 komentar negatif. Dan data *testing* dipilih sebanyak 20% dari 714 komentar dengan penggabungan data yang sama dengan data training untuk dilakukan proses perbandingan kata secara otomatis melalui aplikasi yang dibuat. Pada tabel 1 akan di tampilkan rincina jumlah data *training*:

Tabel 1. Rincian Jumlah Data *Training*

Data	Sentimen Positif	Sentimen Negatif	Total Sentimen
Komentar Pengguna	411	303	714

### Proses Analisis menggunakan metode *Naïve Bayes Classifier*

Pembobotan pada data *training* dilakukan pada setiap proses klasifikasi. Pembobotan dilakukan dengan menghitung probabilitas kata pada label positif, dan negatif. Hasilnya kemudian dijadikan sebagai contoh dalam menentukan sentimen positif dan negatif. Dalam menentukan sentimen itu dengan memperhitungkan probabilitas pada data *testing* dengan menguji pada probabilitas data *training*. Pada tahap ini digunakan algoritma *naïve bayes*.

Tabel 2. Data Pelanggan Tokopedia

Username	Coment	Label
Akhyar Nasution	Mantap	Positive
Damori Aris	Bagus	Positive
Pyan putra289	aduh kok naik siii jadi minimal ribu jadi mikir kali deh kalo mo belanja onlin	Negative
Tommy Wotan	tokopedia terbaik	Positive

Saat menerapkan model *naive Bayes*, pertama-tama kita temukan probabilitas yang diasumsikan untuk setiap kelas P(H). Asumsi yang ada adalah pelanggan memberikan komentar positif dan negatif. Berikut perhitungan probabilitas seperti tabel di bawah ini:

Tabel 3. Probabilitas Kelas

	Kelas	
	Positif	Negatif
Positif	411	303
P(Positif)	0,575	P(Negatif) 0,424
	$\frac{411}{714}$	$\frac{303}{714}$

Ketika probabilitas setiap hipotesis sudah diketahui, langkah berikutnya adalah untuk menghitung probabilitas dari kondisi yang diberikan (probabilitas X) berdasarkan probabilitas setiap hipotesis (probabilitas H) bisa dilihat pada tabel di bawah ini:

Tabel 4. Probabilitas Kata

Kosa Kata	Dokumen	Positive	Negative
Tokopedia	120	76	69
Bagus	133	37	17
Mantap	99	47	0
Belanja	81	51	40
Ongkir	64	23	61
Gratis	51	22	40

Untuk menentukan nilai setiap kelas data *testing* menggunakan tabel probabilitas kata. Contoh data *testing* ada pada tabel dibawah ini sebagai berikut:

\*Sifa Melina Salsabila<sup>1)</sup>, Aang Alim Murtopo<sup>2)</sup>, Nurul Fadhillah<sup>3)</sup>

Tabel 5. Contoh Data *Testing*

<i>Username</i>	<i>Coment</i>	<i>label</i>
Faris Ihsan Albana	belanja di tokopedia mantap banyak <i>discount</i>	<i>Positive</i>
Siti Zainah	kok tidak bisa cod nya	<i>Negative</i>

Setelah diketahui probabilitas setiap kata, maka selanjutnya menjumlah probabilitas akhir diseluruh kelas. Perhitungan tiap kata

Kata Tokopedia

$$P(\text{Tokopedia}|\text{Positif}) = \frac{76+1}{411+714} = 0,0684$$

$$P(\text{Tokopedia}|\text{Negatif}) = \frac{69+1}{303+714} = 0,0688$$

Kata Bagus

$$P(\text{Bagus}|\text{Positif}) = \frac{37+1}{411+714} = 0,0337$$

$$P(\text{Bagus}|\text{Negatif}) = \frac{17+1}{303+714} = 0,0176$$

Kata Mantap

$$P(\text{Mantap}|\text{Positif}) = \frac{47+1}{411+714} = 0,0426$$

$$P(\text{Mantap}|\text{Negatif}) = \frac{0+1}{303+714} = 9,832$$

Kata Belanja

$$P(\text{Belanja}|\text{Positif}) = \frac{51+1}{411+714} = 0,0462$$

$$P(\text{Belanja}|\text{Negatif}) = \frac{40+1}{303+714} = 0,0403$$

Kata Ongkir

$$P(\text{Ongkir}|\text{Positif}) = \frac{23+1}{411+714} = 0,0213$$

$$P(\text{Ongkir}|\text{Negatif}) = \frac{61+1}{303+714} = 0,0609$$

Kata Gratis

$$P(\text{Gratis}|\text{Positif}) = \frac{22+1}{411+714} = 0,0204$$

$$P(\text{Gratis}|\text{Negatif}) = \frac{40+1}{303+714} = 0,0403$$

$$P(X|\text{Positif}) = P(\text{tokopedia}|\text{Positif}) \times P(\text{bagus}|\text{positif}) \times P(\text{mantap}|\text{positif}) \times P(\text{belanja}|\text{positif}) \times P(\text{ongkir}|\text{positif}) \times P(\text{gratis}|\text{positif})$$

$$= 0,0684 \times 0,0337 \times 0,0426 \times 0,0462 \times 0,0213 \times 0,0204$$

$$= 1,971275608032192$$

$$P(X|\text{Negatif}) = P(\text{tokopedia}|\text{negatif}) \times P(\text{bagus}|\text{negatif}) \times P(\text{mantap}|\text{negatif})$$

$$\times P(\text{belanja}|\text{negatif}) \times P(\text{ongkir}|\text{negatif}) \times P(\text{gratis}|\text{negatif})$$

$$= 0,0688 \times 0,0176 \times 9,832 \times 0,0403 \times 0,0609 \times 0,0403$$

$$= 1,177525608564265$$

Untuk mendapatkan probabilitas akhir, maka nilai yang sudah didapatkan kemudian dimasukkan.

$$P(X|\text{Positif}) * P(\text{Positif})$$

$$= 0,575 \times 1,971275608032192$$

$$= 1,13348347461851$$

$$P(X|\text{Negatif}) * P(\text{Negatif})$$

$$= 0,424 \times 1,177525608564265$$

$$= 0,4992708580312484$$

### Hasil Pengukuran Akurasi

Pengujian dilakukan untuk memutuskan apakah perhitungan yang dilakukan di atas benar menggunakan metode *Naïve Bayes*. Pengujian dilakukan dengan menggunakan data *testing* terpilih, kemudian data *testing* dibandingkan dengan data *training* yang telah diketahui sebelumnya. Hasil eksperimen yang dilakukan menghasilkan nilai eksak dengan menggunakan *Confusion Matrix* dari *RapidMiner*.

*Confusion Matrix*

*Confusion matrix* adalah bentuk matrik yang terdiri dari *true positive* dan *true negative*, kemudian diuji menggunakan data *testing* untuk menunjukan nilai yang benar. Berikut ini adalah gambar hasil perhitungannya:

\*Sifa Melina Salsabila<sup>1)</sup>, Aang Alim Murtopo<sup>2)</sup>, Nurul Fadhillah<sup>3)</sup>

accuracy: 95.10%

	true Negatif	true positif	class precision
pred. Negatif	61	7	89.71%
pred. positif	0	75	100.00%
class recall	100.00%	91.46%	

Gambar 2. Hasil Akurasi Data

### KESIMPULAN

Kesimpulan dari penelitian ini bahwa metode *Naïve Bayes* Classifier terbukti bisa untuk menganalisis secara otomatis. Performa yang telah dikerjakan dengan pengujian data *real-time* menggunakan aplikasi bawaan dan cepat, setiap kata dikategorikan berdasarkan persepsi positif dan negatif. Performa yang dihasilkan pada pengujian yang dilakukan oleh *rapid evaluator* pada 143 data pengujian menghasilkan nilai akurasi 95,10%, dengan *class recall*: 91,46% (*positive class*: *Positive*) sedangkan *class precision*: 100.00% (*positive class*: *Positive*) dan nilai *AUC*: 0.999 (*positive class*: *Positive*)

Saran dari penulis untuk menambahkan beberapa metode lain, seperti memilih objek geografis untuk mengumpulkan informasi untuk lebih mengurangi keraguan secara signifikan untuk meningkatkan akurasi menggunakan algoritma *Naïve Bayes*. Menambahkan data kamus pada saat melakukan tahap *stopword*, seperti pada komentar pelanggan di *Google Playstore*, masih banyak contoh penggunaan bahasa yang tidak baku.

### REFERENSI

- Apriani, R., Gustian, D., Program, S., Sistem, I., Putra, U. N., Indonesia, S., Raya, J., Kaler, C., 21, N., & Sukabumi, K. (2019). ANALISIS SENTIMEN DENGAN NAÏVE BAYES TERHADAP KOMENTAR APLIKASI TOKOPEDIA. In *Jurnal Rekayasa Teknologi Nusa Putra* (Vol. 6, Issue 1).
- Arbaini, P. (2020). *Jurnal Bisnis dan Manajemen PENGARUH CONSUMER ONLINE RATING DAN REVIEW TERHADAP KEPUTUSAN PEMBELIAN PADA PENGGUNA MARKETPLACE TOKOPEDIA* (Vol. 7, Issue 1). <http://jurnal.unmer.ac.id/index.php/jbm>
- Avi Tejo Bhaskoro. (2014, June 12). *Aplikasi Mobile untuk Android*. DailySocial. <https://dailysocial.id/post/tokopedia-android>
- Hirst, G. (n.d.). *Sentiment Analysis and Opinion Mining*. <http://www.morganclaypool.com/toc/hlt/1/1>
- Latifah, E. F. U. (2018). *PERBANDINGAN KINERJA MACHINE LEARNING*.
- Mulia, R. K. (2021). *TEXT MINING ANALYSIS DAN SENTIMENT ANALYSIS DENGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER*.
- Prof Zudan Arif Fakhrullah. (2022, February 24). *273 Juta Penduduk Indonesia Terupdate Versi Kemendagri*. Direktorat Jenderal Kependudukan Dan Pencatatan Sipil Kementerian Dalam Negeri Republik Indonesia. <https://dukcapil.kemendagri.go.id/berita/baca/1032/273-juta-penduduk-indonesia-terupdate-versi-kemendagri>
- Rizal. (2022, February 22). *Berapa Banyak Jumlah Perangkat Seluler di Indonesia Tahun Ini?* INFO KOMPUTER. <https://infokomputer.grid.id/read/123154830/berapa-banyak-jumlah-perangkat-seluler-di-indonesia-tahun-ini?page=all>
- Santoso<sup>1</sup>, V. I., Virginia<sup>2</sup>, G., & Lukito<sup>3</sup>, Y. (2017). *PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE* (Vol. 14, Issue 2).
- Sfr. (2022, May 30). *Tokopedia Jadi Jawara e-Commerce RI Kuartal I 2022*. CNN Indonesia. <https://www.cnnindonesia.com/ekonomi/20220530133716-92-802688/tokopedia-jadi-jawara-e-commerce-ri-kuartal-i-2022>
- Yane Marita Febrianti. (2018). ANALISIS SENTIMEN PADA ULASAN “LAZADA” BERBAHASA INDONESIA MENGGUNAKAN K-NEAREST NEIGHBOR (K-NN). *Repository.Ub.Ac.Id, Analisis Sentimen Pada Ulasan Lazada Berbahasa Indonesia Menggunakan K-Nearest Neighbor(K-NN)*.

\*Sifa Melina Salsabila<sup>1</sup>, Aang Alim Murtopo<sup>2</sup>, Nurul Fadhilah<sup>3</sup>

