

# Analisis Sentimen Ulasan Aplikasi Tripadvisor Dengan Metode Support Vector Machine, K-Nearest Neighbor, Dan Naive Bayes

Antonius Mbay Ndapamuri<sup>1</sup>, Danny Manongga<sup>2</sup>, Ade Iriani<sup>3</sup>

<sup>1,2,3</sup> Universitas Kristen Satya Wacana, Jl. Diponegoro No.52-60, Salatiga, Indonesia

Email: 972019007@student.uksw.edu<sup>1</sup>, danny.manongga@uksw.edu<sup>2</sup>, ade.iriani@uksw.edu<sup>3</sup>

**Abstrack** – This study aims to classify reviews on the Tripadvisor application found on the Google Play store based on Word Cloud and Visual Network Explorer with the Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), and Naïve Bayes algorithms. The research stage begins with data collection of 1,000 data obtained using the Web Scraping technique, then the data preprocessing stage and continues with the classification model using SVM, K-NN, and Naïve Bayes. Next, the evaluation of the model with the results of the SVM obtained the best accuracy of 89,8%. The final stage of visual analysis uses Visual Word Cloud and Visual Network Explorer to obtain user decision information in giving positive and negative reviews. Complaints often appear such as bad applications, not bad applications, delete applications, and application errors, please fix the application, the application makes the cellphone slow because it updates frequently. To overcome this problem, it is hoped that the Tripadvisor application will further improve application performance by updating in terms of application usage to overcome the problems experienced by users.

**Keywords** - Word Cloud, Network Explorer, Support Vector Machine, K-Nearest Neighbor, Naive Bayes.

**Intisari** – Penelitian ini bertujuan untuk mengklasifikasikan review pada aplikasi *Tripadvisor* yang terdapat pada *Google Playstore* berbasis *Word Cloud* dan *Visual Network Explorer* dengan algoritma Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), dan Naïve Bayes. Tahapan penelitian dimulai dengan pengumpulan data sebanyak 1.000 data yang diperoleh dengan menggunakan teknik Web Scraping, kemudian tahap Preprocessing data dan dilanjutkan dengan model klasifikasi menggunakan SVM, K-NN, dan Naïve Bayes. Berikutnya evaluasi model dengan hasil SVM memperoleh akurasi terbaik 89,8%. Tahap terakhir analisis visual menggunakan *Visual Word Cloud* dan *Visual Network Explorer* untuk mendapatkan informasi keputusan pengguna dalam memberikan *Review* positif dan negatif. Keluhan yang sering muncul seperti aplikasi jelek, aplikasi lumayan, hapus aplikasi, aplikasi error, aplikasinya tolong diperbaiki, aplikasi membuat hp jadi lemot karna sering update. Untuk mengatasi masalah tersebut diharapkan pihak aplikasi Tripadvisor agar lebih meningkatkan kinerja aplikasi dengan melakukan update dari segi penggunaan aplikasi agar mengatasi masalah yang dialami para pengguna.

Kata Kunci - *Word Cloud*, *Network Explorer*, *Support Vector Machine*, *K-Nearest Neighbor*, *Naive Bayes*.

## I. PENDAHULUAN

Di era digital Indonesia, teknologi informasi sangat penting bagi pertumbuhan perusahaan. Teknologi informasi akan sangat penting untuk operasi transaksi normal, berkala, dan insidental, termasuk pertukaran barang dan layanan komputer, serta untuk memberikan banyak informasi dengan cepat dan akurat [1]. Pendidikan, bisnis, pemerintahan, dan komunikasi sosial hanyalah beberapa bidang kehidupan manusia yang telah berubah akibat pertumbuhan

teknologi internet [2]. Perkembangan dunia teknologi saat ini telah mengalami perubahan yang sangat signifikan dari berbagai bidang, baik dari segi ekonomi, budaya, social, dan Pendidikan. Dengan adanya media *online* sangat memudahkan masyarakat dalam mengakses informasi. Salah satu informasi yang dicari oleh para pengguna internet adalah informasi mengenai destinasi objek wisata. Seiring perkembangan zaman kebutuhan manusia pun semakin meningkat khususnya dalam memanfaatkan media *online* untuk mengakses lokasi objek wisata yang akan dikunjungi.

Wisatawan merupakan salah satu variabel yang memberikan dampak signifikan terhadap pertumbuhan ekonomi suatu daerah untuk meningkatkan devisa negara, meningkatkan pendapatan di suatu daerah, meningkatkan kesempatan kerja, dan menghindari penyalahgunaan sumber daya alam [3]. Sebelum mengunjungi suatu destinasi wisata, pengunjung perlu mengetahui keadaan di sana. Hal ini dibuktikan melalui evaluasi atau testimoni yang ditinggalkan oleh pengunjung suatu lokasi[4]. Agar dapat dipahami bahwa wisatawan adalah orang yang melakukan perjalanan ke suatu lokasi atau karena alasan tertentu [5]. Didasari oleh tingginya minat terhadap pariwisata yang merupakan salah satu industri penyumbang pertumbuhan perekonomian nasional. Hal ini dapat meningkatkan permintaan akan informasi dan memotivasi penyedia layanan untuk menciptakan teknologi baru guna menganalisis data dan menawarkan informasi dengan cepat dan efisien. Sebagian besar wisatawan dapat mencari informasi tentang destinasi objek wisata yang ingin mereka kunjungi menggunakan aplikasi *tripadvisor*. Aplikasi *tripadvisor* merupakan salah satu aplikasi terpercaya dan terkenal di dunia yang menawarkan review detail mulai dari tempat wisata sampai dengan hotel yang tepat untuk disewa. Adanya berbagai *review* pada aplikasi *tripadvisor* ini sangat bermanfaat bagi wisatawan karena dapat menjadi tambahan informasi mengenai tempat tujuan, transportasi, dan juga akomodasi.

Tujuan dari penelitian ini adalah untuk mengklasifikasi review sentimen positif dan negatif pada aplikasi *tripadvisor* berbasis *word cloud* dan *network explorer* dengan algoritma Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), Naïve Bayes. Hasil dari penelitian ini akan menjadi gambaran sentimen terhadap aplikasi *tripadvisor* apakah opini para pengguna lebih cenderung berkomentar positif atau negatif dan peneliti melakukan perbandingan dari ketiga algoritma yang digunakan untuk menunjukkan hasil akurasi, presisi, recall, dan f1-score serta menampilkan visual data yang mempengaruhi tingkat kepuasan para pengguna.

## II. SIGNIFIKANSI STUDI

### A. Penelitian Terdahulu

Beberapa penelitian terdahulu yang menjadi acuan dalam melakukan penelitian ini seperti yang ditunjukkan pada tabel.

TABEL I  
PENELITIAN TERDAHULU

Nomor	Penulis	Penelitian Terdahulu
1	Acep Saepulrohman[6]	Analisis Sentimen Kepuasan Pengguna Aplikasi WhatsApp Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine. Hasil klasifikasi menunjukkan bahwa tingkat akurasi untuk kepuasan pengguna aplikasi Whatsapp berdasarkan algoritma Naïve Bayes adalah sebesar 70,40%, dan Support Vector Machine adalah sebesar 77,00%.
2	Nuraine Herlinawati[7]	Analisis Sentimen Zoom Cloud Meetings Di Play Store Menggunakan Naïve Bayes Dan Support Vector Machine. Hasil dari pengujian menggunakan Naïve Bayes memperoleh nilai akurasi 74,37% dan Support Vector Machine memperoleh nilai akurasi 81,22%.

3	Ari Putra Wibowo [8]	Komparasi Metode Naïve Bayes Dan K-Nearest Neighbor Terhadap Analisis Sentimen Pengguna Aplikasi Peduli Lindungi. Hasil dari penelitian menggunakan algoritma Naïve Bayes di peroleh hasil akurasi 70,46% dan untuk algoritma K-Nearest Neighbor di peroleh akurasi 73,33%.
4	Susanti Fransiska[9]	Sentiment Analysis Provider by.U on <i>Google Play Store</i> Reviews with TF-IDF and Support Vector Machine (SVM) Method. Hasil klasifikasi sentimen menggunakan Support Vector Machine (SVM) dan TF-IDF sebagai ekstraksi fitur. TF-IDF+SVM dengan 5-Fold Validation menghasilkan akurasi yang cukup baik dengan akurasi rata-rata 84,7%, presisi 84,9%, recall 84,7%, dan f sebesar 84,8%.

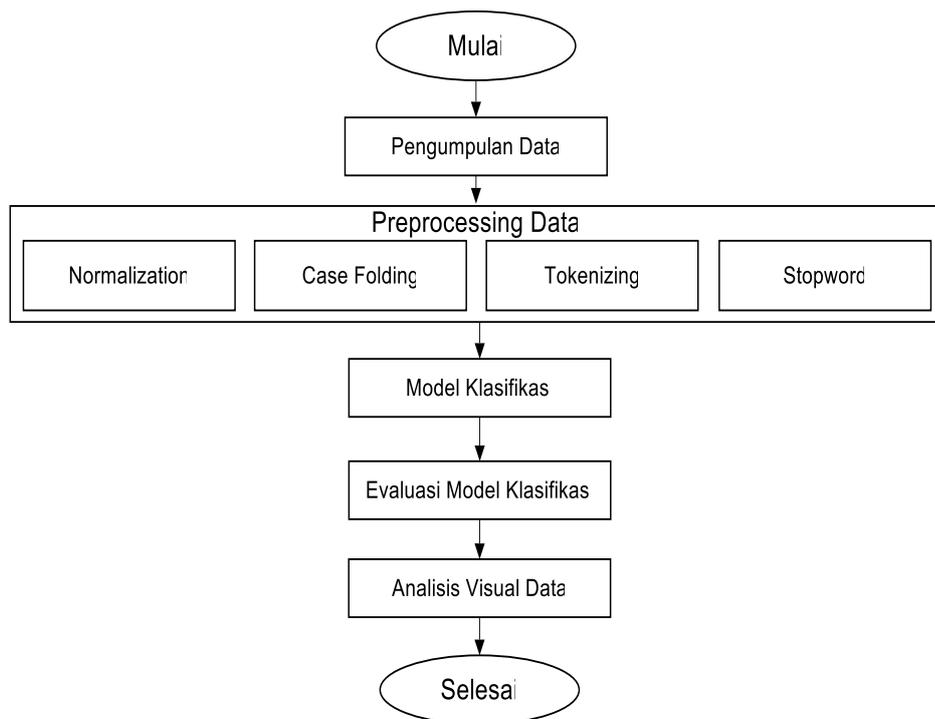
### B. Tinjauan Pustaka

Sentimen Analisis atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif [10]. Analisis sentimen adalah kegiatan menganalisis pendapat, opini, sikap, atau perasaan tentang suatu produk, subjek, atau isu tertentu untuk mengidentifikasi apakah mengandung sentimen positif atau negatif. Analisis sentimen adalah prosedur yang digunakan untuk menentukan isi suatu dataset berupa opini atau pandangan (sentimen) berupa teks tentang suatu isu atau peristiwa yang bersifat positif, negatif, atau netral. Analisis sentimen adalah salah satu domain dari Natural Language Processing (NLP) [11]. Analisis sentimen, juga dikenal sebagai ekstraksi data otomatis, pemrosesan, dan pemahaman data tekstual, adalah prosesnya. Menganalisis opini publik atau data opini untuk menyimpulkan data yang diperiksa sesudahnya[12] Analisis sentimen adalah kegiatan menganalisis pendapat, opini, sikap, atau perasaan tentang suatu produk, subjek, atau isu tertentu untuk mengidentifikasi apakah mengandung sentimen positif atau negatif.

Support Vector Machine (SVM) merupakan metode yang efektif untuk mengklasifikasikan dan meregresikan berbagai masalah linier dan non-linier adalah mesin vektor. Penggunaan aplikasi ini memiliki keuntungan karena memiliki dimensi pemanfaatan fungsi kernel yang cukup tinggi[9]. K-Nearest Neighbor (K-NN) merupakan algoritma klasifikasi untuk mengelompokkan data baru berdasarkan jarak suatu data baru ke beberapa data atau tetangga terdekatnya. Algoritma K-NN bekerja dengan cara mencari jumlah kpola (diantara semua pola latih yang ada pada semua kelas) yang paling dekat dengan pola masukan, kemudian menentukan kelas dengan keputusan berdasarkan jumlah pola yang paling banyak diantara kpola tersebut [14]. Klasifikasi Nave Bayes (NB), yang sering digunakan, adalah teknik pembelajaran berdasarkan teori Bayesian dan membuat asumsi kuat bahwa dampak nilai atribut pada kelas tertentu tidak bergantung pada kelas tersebut atau dipengaruhi oleh atribut lainnya. Teori Bayes adalah teori klasifikasi statistik yang menggunakan data historis untuk meramalkan karakteristik kelas dari anggota probabilitas[15].

### C. Metode Penelitian

Penelitian ini melakukan analisis review pengguna dari aplikasi *tripadvisor* pada *google play store* dengan menggunakan metode Support Vector Machine, K-Nearest Neighbor dan Naïve Bayes dengan tahapan alur penelitian yang ditunjukkan pada gambar 1.



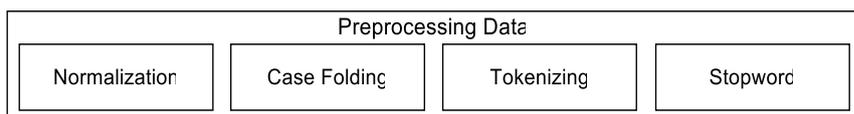
Gambar 1. Tahapan Alur Penelitian

1. *Pengumpulan Data*

Proses pengumpulan data dilakukan dengan web scraping dan menggunakan Bahasa pemrograman python yang telah tersedia pada *google colab*. *Review* pengguna aplikasi *tripadvisor* pada *google playstore* merupakan data yang diambil sebagai data yang diperlukan untuk proses analisis sentimen. Proses scraping data hanya focus dalam kategori *newest* (data terbaru), dari *review* data yang telah terkumpul dilakukan pelabelan klasifikasi yaitu kelas positif dan kelas negatif. Dataset yang dikumpulkan adalah data ulasan yang ada pada aplikasi *tripadvisor* sebanyak 1000 dataset dan dibagi menjadi dua bagian yaitu dataset positif berjumlah 600 data, dataset negatif berjumlah 400 data.

2. *Preprocessing Data*

Pada proses ini dilakukan preprocessing data terlebih dahulu agar sesuai bentuk ejaan sehingga dapat di proses dengan mudah oleh *text mining*. *Text mining* merupakan proses menganalisis pencarian teks yang menjadi sebuah informasi penting dalam mencari hasil yang spesifik. Tahapan preprocessing data ditunjukkan pada gambar 2.



Gambar 2. Tahap Preprocessing Data

3. *Model Klasifikasi*

Langkah selanjutnya dilakukan pemodelan klasifikasi yang diimplementasikan menggunakan *tools orange* dan tiga algoritma machine learning yaitu SVM, K-NN dan Naive Bayes. Untuk mendapatkan hasil accuracy, precision, recall, dan f1- score terbesar antara tiga proses pemodelan klasifikasi yang digunakan, tahap pemodelan klasifikasi ini diselesaikan secara terpisah. Penelitian ini menggunakan kumpulan data training untuk pemodelan klasifikasi, data training yang digunakan sebanyak 80% dari total dataset.

4. *Evaluasi Model Klasifikasi*

Tujuan dari fase terakhir penelitian ini adalah untuk menilai seberapa baik model klasifikasi machine learning yang digunakan pada fase sebelumnya. Pengukuran penilaian ini digunakan untuk membandingkan efisiensi dan kinerja dari tiga model klasifikasi machine learning yang digunakan. Confusion matrix adalah salah satu metode yang digunakan untuk menilai keefektifan model klasifikasi machine learning. Confusion matrix adalah matriks 2 x 2 untuk merangkum seluruh hasil klasifikasi benar dan salah yang ditampilkan pada Tabel 1. Empat variabel pengukuran dibuat dari kombinasi data true positive (TP), false positive (FP), false negative (FN), dan true negative (TN), berdasarkan hasil dari confusion matrix, dapat menghasilkan nilai accuracy, nilai precision, nilai recall, dan f1-score .

TABEL II  
CONFUSION MATRIX

		Klasifikasi Predicted	
		Negatif	Positif
Klasifikasi Actual	Negatif	True positive (TP)	False positive (FP)
	Positif	False negative (FN)	True negative (TN)

a. Accuracy

Tingkat akurasi klasifikasi model ditunjukkan dengan akurasi. Untuk mengukur kinerja model, data uji harus secara akurat disiapkan atau dipahami sebagai sejauh mana nilai yang diproyeksikan cocok dengan nilai sebenarnya. Rumus perhitungan accuracy sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} * 100\% \tag{1}$$

b. Precision

Precision adalah sejauh mana hasil dan data prediksi model akurat atau tingkat di mana sistem bereaksi terhadap permintaan informasi pengguna dan menghasilkan hasil. Rumus perhitungan precision sebagai berikut:

$$Precision = \frac{TP}{TP+FP} * 100\% \tag{2}$$

c. Recall

Recall menguraikan keberhasilan model dalam inovasi informasi. Ini juga dapat disebut sebagai hitungan dokumen teks penting. Rumus perhitungan recall sebagai berikut:

$$Recall = \frac{TP}{TP+FN} * 100\% \tag{3}$$

d. F1-Score

F1-Score adalah perbandingan rata-rata dari precision dan recall. F1-Score diperoleh dengan gambaran kinerja performa algoritma untuk berbagai ukuran dataset yang berbeda dan menghasilkan analisis yang jauh lebih baik. Rumus perhitungan f1-score sebagai berikut:

$$F1 - score = \frac{2*Precision*Recall}{Precision+Recall} * 100\% \tag{4}$$

D. *Analisis Visual Data*

Selanjutnya dilakukan proses visual data dengan menggunakan *network explorer* dan *word cloud*. *Network Explorer* merupakan jaringan dari kata yang paling banyak berhubungan dan seberapa dekat hubungan itu terjadi, semakin sering kata terhubung maka semakin tebal garis

frekuensi dari kata tersebut. *Word Cloud* adalah kata yang paling banyak muncul, semakin sering satu kata digunakan, maka semakin besar ukuran kata ditampilkan pada word cloud.

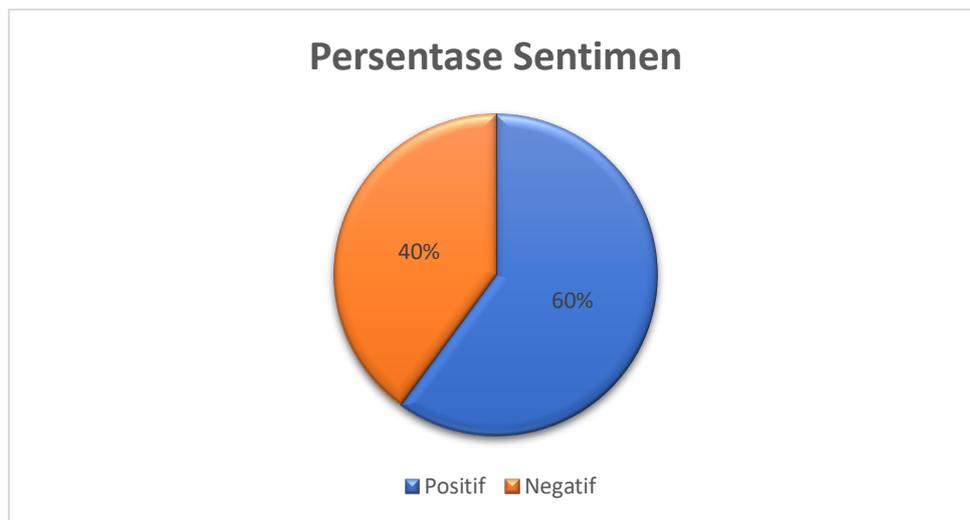
### III. HASIL DAN PEMBAHASAN

#### A. Pengumpulan Data

Pada tahap pengumpulan data ini dilakukan proses web scraping untuk mengumpulkan dataset yang didapat dari ulasan aplikasi *tripadvisor* pada situs *google playstore*. Hasil dari scrapping data ulasan pada *tripadvisor* ditunjukkan pada gambar dibawah ini.

No	UserName	Date	review	Rating	Sentimen
1	Davin Putra	7/2/2022 10:18	sangat tidak memuaskan, pembatalan dilakukan sepihak dalam waktu yg mepet dan posisi sudah dekat hotel dan jg sudah di bayar penuh didepan.kecewa se	1	negatif
2	romansa dota	5/3/2022 19:10	baru instal udah un instal lagi masa gek ada idr rupiah cuman dollar gmna si!!! lu ga cinta Indonesia ga usah buka usaha di negri kami	1	negatif
3	Dodi Fitrah	4/19/2022 17:52	baru dowload mudah lancar	1	negatif
4	Fahmi Fahmi	1/31/2022 5:39	saya sudah susah payah dengan aplikasi ini aplikasi ini sangat jelek	1	negatif
5	Gabriella Zilvania	11/16/2021 1:26	banyak sekali ulasan atau komentar yang di pending di tahan dan tidak dimunculkan kenapa padahal ulasan yang positif itu sangat berperan penting terhadap	1	negatif
6	Kadek Amanta	2/28/2023 4:48	sangat mudah untuk mencari tempat liburan	5	positif
7	TAUFIK BUDY SANTOSO	2/19/2023 18:56	sangat membantu dalam pemilihan akomodasi	5	positif
8	schmad Usmara	2/10/2023 9:29	bagus dan rencan sangat baik	5	positif
9	zuli amanah	2/1/2023 9:05	aplikasi yang sangat membantu	5	positif
10	Salmiati Iq Izzat	12/17/2022 13:49	kenapa belum bisa di pakai	4	negatif
11	warouw fabian	10/25/2022 7:04	bagus	4	positif
12	Nabila Hasanah	7/7/2022 7:27	bagus	4	positif
13	Olvin Skw	6/22/2023 5:01	mantap	4	positif
14	Doju Store	4/24/2022 2:32	aplikasi yang cukup membantu para wisatawan	4	positif
15	salma abudan	12/12/2022 7:12	aplikasinya susah dan membingungkan	3	negatif
16	D.A Channel	10/24/2022 3:51	kasih bintang dulu, kalau nanti trip saya rame melalui aplikasi ini saya kasih bintang lima	3	negatif
17	Novend Christian Saga	12/9/2021 5:29	sudah tidak bisa pesan tiket pesawat awal nya senang dengan aplikasi ini tapi setelah fitur nya berkurang merasa tidak berguna ini aplikasi	3	negatif
18	Alekius Alex	5/30/2021 20:14	baru percobaan jika informasi yang disajikan relevan dengan apa yang dicari baru bintangnya ditambah	3	negatif
19	Samson Harati	21/1/2021 14:33	sangat membantu	3	positif
20	M IQBAL ALKAREEM	11/23/2020 4:18	baru mencoba kalo bagus aku tambahn bintang	3	positif
21	winda fryn_	12/14/2022 1:47	tolong dengan sangat saya sudah ulasan beberapa hotel dan masa tidak ada yang masuk semua, tolong dibantu ya	2	negatif
22	Diani Irdanylita	11/19/2022 10:02	belum berhasil membuat listing, ditolak terus	2	negatif
23	Rohati Nurhayati	4/3/2022 2:44	kemaren tidak bisa buka google saya jadi bingung	2	negatif
24	Putri Bae	7/6/2021 10:18	ada yang mengerti tidak cara pake nya	2	negatif
25	ANNISA ANNISA	6/12/2021 23:45	saya tidak tahu apa ini	2	negatif
26	Supo supardi Supardi	4/6/2021 23:49	terangkan layernya	2	negatif
27	Ghalapulla Reena	3/18/2021 3:30	tertera harga murah, ketika pembayaran harga tidak sesuai dengan yang tertera. saat melakukan pembayaran, kode pembayaran yang diberi tidak terbaca di ri	2	negatif
28	Intan Juntak	2/2/2021 6:32	pembayaran nya susah	2	negatif
29	ROHIMINS37 ROHIMINS	12/26/2020 5:16	tidak tahu pake juga saya	2	negatif
30	Nana Puspita	10/25/2022 0:06	masih belajar menggunakan aplikasinya	4	positif

Gambar 3. Tahap Pengumpulan Data



Gambar 4. Persentase Sentimen Pengguna Aplikasi *Tripadvisor*

Dataset yang dikumpulkan adalah data ulasan aplikasi *tripadvisor* pada *google playstore* dengan jangka waktu 7 januari 2018 hingga 28 februari 2023 sebanyak 1000 data yang di peroleh menggunakan web scraping. Data yang di ambil adalah data dengan rating bintang 1, 2, 3, 4 dan 5 dan kemudian sentimen para pengguna dikelompokan menjadi dua bagian, yaitu sentimen positif sebanyak 600 data ulasan dan sentimen negatif sebanyak 400 data ulasan. Selanjutnya dilakukan anotasi secara manual oleh dua anatator dan menghasilkan persentase positif sebanyak 60% dan persentase negatif sebanyak 40%. Gambar 4, diatas menunjukkan bahwa hasil dari anotasi sentimen pengguna pada ulasan aplikasi *tripadvisor*, lebih banyak bersentimen positif dibandingkan dengan yang bersentimen negatif.

*B. Preprocessing Data*

Data ulasan pada aplikasi *tripadvisor* yang telah dikumpulkan berjumlah 1000 data, diantaranya data sentimen positif berjumlah 600 data dan sentimen negatif berjumlah 400 data dan kemudian dilakukan preprocessing. Kata yang digunakan pada ulasan pengguna merupakan data teks dalam bentuk belum terstruktur sehingga perlu di olah menjadi informasi yang bermanfaat. Adapun tahapan yang telah dilakukan dalam preprocessing data yaitu proses normalization, case folding, tokenizing, dan stopword, yang ditunjukkan dalam tabel.

1. Normalization merupakan proses untuk memperbaiki ejaan kata yang salah menjadi kata yang sesuai. Contohnya “aplikasi yg cukup membantu thanks *tripadvisor*” menjadi “aplikasi yang cukup membantu terimakasih *tripadvisor*”.

TABEL III  
PROSES NORMALIZATION

Sebelum Normalization	Sesudah Normalization
<p><u>A</u>pk yg sangat bermanfaat <u>u</u>/ mencari <u>i</u>nfo traveling terutama <u>t</u>tg hotel. Kalo bisa ditambahkan <u>o</u>psi supaya pihak hotel membaca <u>r</u>eview pengunjung. Sangat <u>B</u>agos, <u>M</u>akenye Saya Kasih BINTANG lima. Semoga Aplikasi <u>T</u>ripAdvisor Semakin Lebih JAYA <u>A</u>nd Semakin di MAJU Ke Depannya.</p>	<p>Aplikasi yang sangat bermanfaat untuk mencari informasi traveling terutama tentang hotel. Kalo bisa ditambahkan pilihan supaya pihak hotel membaca ulasan pengunjung. Sangat bagus, makanya saya kasih bintang lima. Semoga aplikasi <i>tripadvisor</i> semakin lebih jaya dan semakin maju ke depannya.</p>

2. Case Folding merupakan proses perubahan huruf besar menjadi huruf kecil, proses ini dilakukan untuk mengelompokan istilah kata-kata yang memiliki arti yang sama. Contohnya “APLIKASI yang CUKUP membantu Terimakasih *tripadvisor*” menjadi “aplikasi yang cukup membantu thanks *tripadvisor*”.

TABEL IV  
PROSES CASE FOLDING

Sebelum Case Folding	Sesudah Case Folding
<p><u>A</u>pk yg sangat bermanfaat u/ mencari info traveling terutama <u>t</u>tg hotel. <u>K</u>alo bisa ditambahkan opsi supaya pihak hotel membaca review pengunjung. Sangat <u>B</u>agos, <u>M</u>akenye <u>S</u>aya <u>K</u>asih BINTANG lima. Semoga <u>A</u>plikasi <u>T</u>ripAdvisor <u>S</u>emakin <u>L</u>ebih JAYA <u>A</u>nd Semakin di <u>M</u>AJU <u>K</u>e <u>D</u>epannya.</p>	<p>aplikasi yang sangat bermanfaat untuk mencari informasi traveling terutama tentang hotel. kalo bisa ditambahkan pilihan supaya pihak hotel membaca ulasan pengunjung. sangat bagus, makanya saya kasih bintang lima. semoga aplikasi <i>tripadvisor</i> semakin lebih jaya dan semakin maju ke depannya.</p>

3. Tokenizing merupakan proses memisahkan kalimat untuk menjadi beberapa bagian kata, tujuan proses ini yaitu mendapatkan bagian-bagian kata yang dijadikan nilai dari sebuah matriks dokumen. Contohnya “Aplikasi yang cukup membantu terimakasih *tripadvisor*” menjadi “Aplikasi”, “yang”, “cukup”, “membantu”, “terimakasih”, “*tripadvisor*”.

TABEL V  
PROSES TOKENIZING

Sebelum Tokenizing	Sesudah Tokenizing
<p>Aplikasi yang sangat bermanfaat untuk mencari informasi traveling terutama tentang hotel. kalo bisa ditambahkan pilihan supaya pihak hotel membaca ulasan pengunjung. Sangat bagus, makanya saya kasih bintang lima. semoga aplikasi <i>tripadvisor</i> semakin lebih jaya dan semakin maju ke depannya.</p>	<p>“Aplikasi”, “yang”, “sangat”, “bermanfaat”, “untuk”, “mencari”, “informasi”, “traveling”, “terutama”, “tentang”, “hotel”. “kalo”, “bisa”, “ditambahkan”, “pilihan”, “supaya”, “pihak”, “hotel”, “membaca”, “ulasan”, “pengunjung”. “Sangat”, “bagus”, “makanya”, “saya”, “kasih”, “bintang” “lima”. “semoga”, “aplikasi”, “<i>tripadvisor</i>”, “semakin”, “lebih”, “jaya”, “dan”, “semakin” “maju”, “ke”, “depannya”.</p>

4. Stopword merupakan proses menghilangkan kata yang tidak berpengaruh dalam suatu kalimat tetapi tidak mengurangi informasi dari sebuah kalimat itu jika dihilangkan. Contohnya “Aplikasi yang cukup membantu terimakasih *tripadvisor*” menjadi “Aplikasi”, “membantu”, “terimakasih”, “*tripadvisor*”.

TABEL VI  
PROSES STOPWORD

Sebelum Stopword	Sesudah Stopword
Aplikasi yang sangat bermanfaat untuk mencari informasi traveling terutama tentang hotel. kalo bisa ditambahkan pilihan supaya pihak hotel membaca ulasan pengunjung. Sangat bagus, makanya saya kasih bintang lima. semoga aplikasi <i>tripadvisor</i> semakin lebih jaya dan semakin maju ke depannya.	“Aplikasi”, “bermanfaat”, “mencari”, “informasi”, “travelling”, “tentang”, “hotel”. “ditambahkan”, “pilihan”, “supaya”, “pihak”, “hotel”, “membaca”, “ulasan”, “pengunjung”. “Bagus”, “makanya”, “kasih”, “bintang”, “lima”. “aplikasi”, “ <i>tripadvisor</i> ”, “semakin”, “jaya”, “semakin”, “maju”.

### C. Pembagian Dataset Untuk Model Klasifikasi

Pada tahap ini dilakukan klasifikasi sentimen positif dan negatif dengan pemodelan data yang dibagi menjadi dua bagian yaitu data training dan data testing dan selanjutnya diklasifikasikan menggunakan algoritma Support Vector Machine, K-Nearest Neighbor dan Naïve Bayes, pembagian data training dan data testing merupakan data yang sudah memiliki label kelas. Tabel 5 menunjukkan hasil dari pembagian data yang menjadi 80% : 20% skenario.

TABEL VII  
DATA TRAINING DAN DATA TESTING

Aplikasi	Klasifikasi	Data Train 80%	Data Test 20%	Jumlah
<i>Tripadvisor</i>	Positif	500	100	600
	Negatif	300	100	400
	Total	800	200	1000

Setelah dilakukan pembagian pemodelan klasifikasi data training dan data testing selanjutnya dilakukan perbandingan tingkat aAccuracy, precision, recall, dan f1-score menggunakan confusion matrix untuk membantu membangun model klasifikasi terbaik dalam proses pengujian dan dilakukan Predictions untuk menentukan perkiraan data yang benar dan salah dari hasil proses klasifikasi.

### D. Hasil Evaluasi Model Klasifikasi

Model klasifikasi SVM, K-NN, dan Naïve Bayes yang sudah dilakukan selanjutnya dilakukan evaluasi menggunakan confusion matrix untuk mendapat hasil dengan nilai accuracy, precision, recall, dan f1-score terbaik antara tiga algoritma pemodelan yang digunakan. Gambar 5, adalah confusion matrix model SVM. Actual values adalah data dengan ulasan aktual dan predicated values adalah data dengan ulasan prediksi. Dari ulasan yang digunakan dalam proses pengujian untuk sentimen positif dan negatif, menghasilkan empat kombinasi sebagai berikut.

		Predicted		Σ
		negatif	positif	
Actual	negatif	89.9 %	10.3 %	319
	positif	10.1 %	89.7 %	481
Σ		297	503	800

Gambar 5. Confusion Matrix Model SVM

1. *True positive* mendapat nilai dengan persentase 89,9%, sebanyak 267 data ulasan dan mampu memprediksi benar sebagai label positif menggunakan model SVM.
2. *False positive* mendapat nilai dengan persentase 10,3%, sebanyak 52 data ulasan dan mampu memprediksi salah sebagai label positif menggunakan model SVM.
3. *False negative* mendapat nilai dengan persentase 10,1%, sebanyak 30 data ulasan dan mampu memprediksi salah sebagai label negatif menggunakan model SVM.
4. *True negative* mendapat nilai dengan persentase 89,7%, sebanyak 451 data ulasan dan mampu memprediksi dengan benar sebagai label negatif menggunakan model SVM.

Gambar. 6 adalah confusion matrix model K-NN. Actual values adalah data dengan ulasan aktual dan predicated values adalah data dengan ulasan prediksi. Dari ulasan yang digunakan dalam proses pengujian untuk sentiment positif dan negatif, menghasilkan empat kombinasi sebagai berikut.

		Predicted		Σ
		negatif	positif	
Actual	negatif	84.0 %	16.0 %	319
	positif	7.3 %	92.7 %	481
Σ		303	497	800

Gambar 6. Confusion Matrix Model K-NN

1. *True positive* mendapat nilai dengan persentase 84,0%, sebanyak 268 data ulasan dan mampu memprediksi benar sebagai label positif menggunakan model K-NN.
2. *False positive* mendapat nilai dengan persentase 16,0%, sebanyak 51 data ulasan dan mampu memprediksi salah sebagai label positif menggunakan model K-NN.
3. *False negative* mendapat nilai dengan persentase 7,3%, sebanyak 35 data ulasan dan mampu memprediksi salah sebagai label negatif menggunakan model K-NN.
4. *True negative* mendapat nilai dengan persentase 92,7%, sebanyak 446 data ulasan dan mampu memprediksi dengan benar sebagai label negatif menggunakan model K-NN.

Gambar. 7 adalah confusion matrix model Naïve Bayes. Actual values adalah data dengan ulasan aktual dan predicated values adalah data dengan ulasan prediksi. Dari ulasan yang digunakan dalam proses pengujian untuk sentiment positif dan negatif, menghasilkan empat kombinasi sebagai berikut.

		Predicted		Σ
		negatif	positif	
Actual	negatif	90.0 %	10.0 %	319
	positif	12.7 %	87.3 %	481
Σ		348	452	800

Gambar 7. Confusion Matrix Model Naïve Bayes

1. *True positive* mendapat nilai dengan persentase 90,0%, sebanyak 287 data ulasan dan mampu memprediksi benar sebagai label positif menggunakan model Naïve Bayes.
2. *False positive* mendapat nilai dengan persentase 10,0%, sebanyak 32 data ulasan dan

- mampu memprediksi salah sebagai label positif menggunakan model Naïve Bayes.
3. *False negative* mendapat nilai dengan persentase 12,7%, sebanyak 61 data ulasan dan mampu memprediksi salah sebagai label negatif menggunakan model Naïve Bayes.
  4. *True negative* mendapat nilai dengan persentase 87,3%, sebanyak 420 data ulasan dan mampu memprediksi dengan benar sebagai label negatif menggunakan model Naïve Bayes.

Dari pengukuran evaluasi model menggunakan confusion matrix diperoleh perbandingan nilai accuracy, precision, recall, dan f1-score dari masing-masing model klasifikasi, yang terdiri dari model klasifikasi SVM, K-NN dan Naïve Bayes.

TABEL VIII  
PERBANDINGAN HASIL ACCURACY, PRECISION, RECALL, DAN F1-SCORE

Model Klasifikasi	SVM	KNN	Naïve Bayes
Accuracy	89,8%	89,0%	88,6%
Precision	89,7%	84,0%	90,0%
Recall	89,9%	53,5%	87,6%
F1-Score	89,8%	65,3%	88,8%

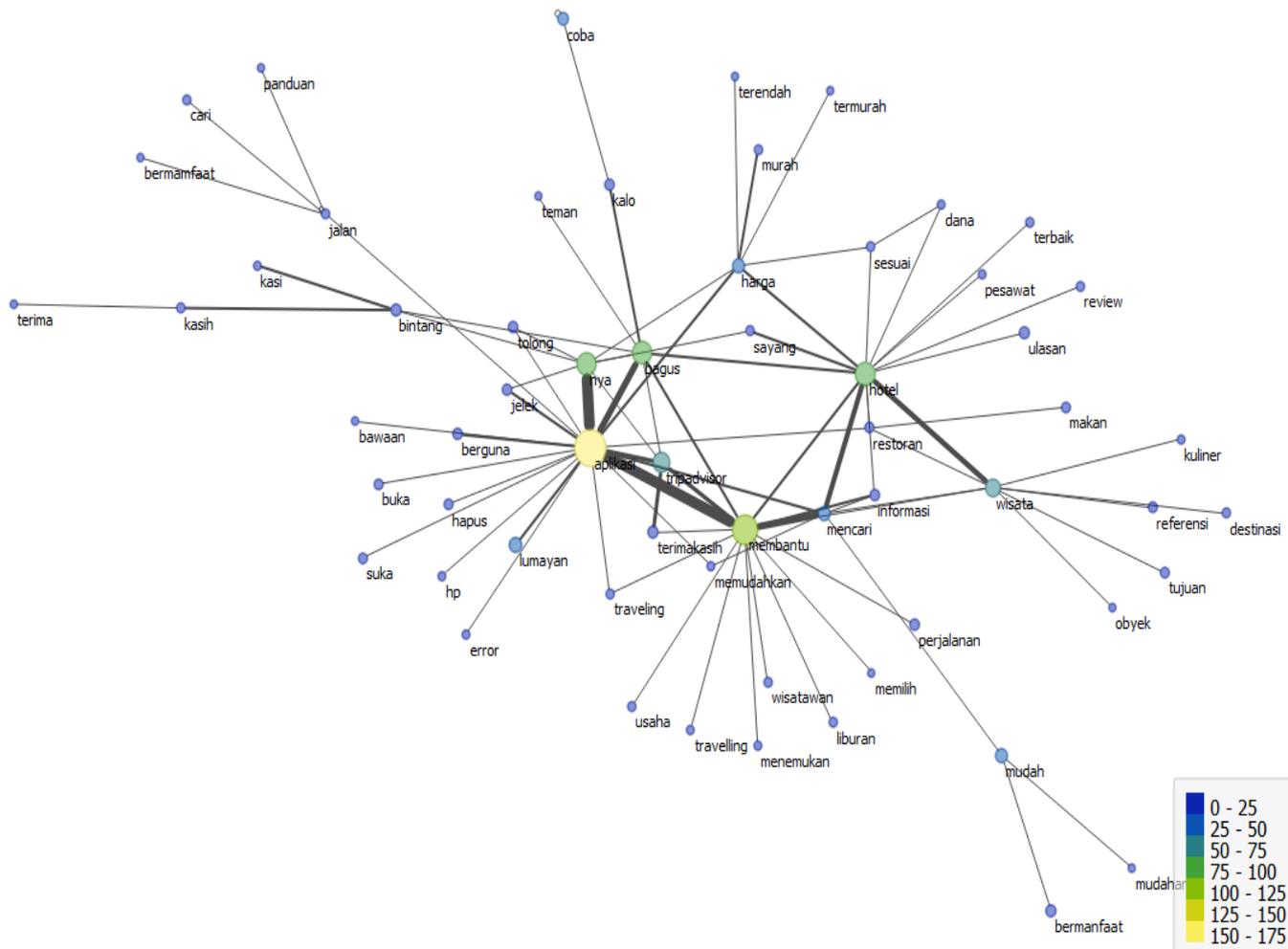
Hasil pengukuran dari masing-masing model klasifikasi pada tabel 7, menunjukkan bahwa tingkat akurasi menggunakan model SVM jauh lebih unggul dengan nilai akurasi tertinggi sebesar 89,8% sedangkan model klasifikasi K-NN memiliki akurasi sebesar 89,02% dan model klasifikasi naïve bayes memiliki nilai akurasi paling rendah sebesar 88,65% dibandingkan dengan dua model klasifikasi lainnya. Model klasifikasi naïve bayes jauh lebih unggul dengan nilai precision tertinggi sebesar 0,90, model klasifikasi SVM memiliki nilai precision sebesar 0,89, dan model klasifikasi K-NN memiliki nilai precision paling rendah sebesar 0,84 dibandingkan dengan dua model klasifikasi lainnya. Model klasifikasi SVM jauh lebih unggul dengan nilai recall tertinggi sebesar 0,89 sedangkan model klasifikasi naïve bayes memiliki nilai recall sebesar 0,87 dan model klasifikasi K-NN memiliki nilai precision paling rendah sebesar 0,53 dibandingkan dengan dua model klasifikasi lainnya. Model klasifikasi SVM jauh lebih unggul dengan nilai recall tertinggi sebesar 0,89 sedangkan model klasifikasi naïve bayes memiliki nilai recall sebesar 0,88 dan model klasifikasi K-NN memiliki nilai precision paling rendah sebesar 0,63 dibandingkan dengan dua model klasifikasi lainnya. Sehingga dapat disimpulkan bahwa SVM merupakan model klasifikasi terbaik dengan performa tertinggi dalam memprediksi kelas.

TABEL IX  
PERBANDINGAN HASIL PENELITIAN SEBELUMNYA DAN MODEL USALAN

Author	Metode	Accuracy	Precision	Recall	F1-Score
Acep Saepulrohman	SVM	77,00%	-	-	-
	Naïve Bayes	70,40%	-	-	-
Nuraine Herlinawati	SVM	81,22%	-	-	-
	Naïve Bayes	74,37%	-	-	-
Ari Putra Wibowo	K-NN	73,33%	-	-	-
	Naïve Bayes	70,46%	-	-	-
Susanti Fransiska	SVM	84,70%	84,9%	84,7%	84,8%
	SVM	89,80%	89,7%	89,9%	89,8%
Antonius Mbay Ndapamuri	K-NN	89,00%	84,0%	53,5%	65,3%
	Naïve Bayes	88,60%	90,0%	87,6%	88,8%

Dari hasil pengukuran pada tabel 7 dan tabel 8 menunjukkan perbandingan tingkat performa pada masing-masing metode yang digunakan antara lain yaitu Acep Saepulrohman dan Nuraine Herlinawati menggunakan dua metode yaitu SVM menghasilkan nilai akurasi 77,00%, dan 81,22% meningkat menjadi 89,8%, dan Naïve Bayes menghasilkan nilai akurasi 70,40%, dan 74,37% meningkat menjadi 88,65%. Penelitian yang dilakukan oleh Ari Putra Wibowo





Gambar 9. Hasil Visual Network Explorer

Hasil visual data pada gambar. 8 dan gambar. 9, menunjukkan banyaknya kata dengan sentimen positif seperti “aplikasi”, “membantu”, “hotel”, “bagus”, “*tripadvisor*”, “wisata”, “mencari”, “mudah”, “terimakasih”, “informasi”, “bermanfaat”, “perjalanan”, “lengkap”, “berguna”. Komentar yang kata-kata dengan frekuensi saling terhubung pada sentimen positif yaitu, “aplikasi sangat membantu”, “aplikasi sangat bagus”, “aplikasi sangat bermanfaat mencari informasi wisata”, “terimakasih *tripadvisor*”. Hal memperlihatkan bahwa para pengguna merasakan manfaat positif ketika sedang menggunakan aplikasi *tripadvisor*.

Hasil visual data pada gambar. 8 dan gambar. 9, juga menunjukkan banyaknya kata dengan sentimen negatif seperti, “aplikasi”, “lumayan”, “harga”, “coba”, “kecewa”, “error” “jelek”, “tolong”, “susah”, “update”, “download”. Komentar yang kata-kata dengan frekuensi saling terhubung pada sentimen negatif yaitu, “aplikasi jelek “aplikasi lumayan”, “hapus aplikasi”, “aplikasi error”, “aplikasinya tolong diperbaiki”, “aplikasi membuat hp jadi lemot karna sering update”. Hal ini memperlihatkan kendala dan keluhan para pengguna ketika sedang menggunakan aplikasi *tripadvisor*. keluhan yang sering muncul pada komentar dengan sentimen negatif menunjukkan bahwa performa aplikasi *tripadvisor* yang kurang baik. Untuk mengatasi masalah yang dirasakan para pengguna diharapkan pihak dari aplikasi *tripadvisor* agar lebih meningkatkan kinerja aplikasi dengan melakukan update versi terbaru yang lebih baik dari segi penggunaan aplikasi agar mengatasi masalah yang di alami para pengguna.

#### IV. KESIMPULAN

Dataset yang dikumpulkan adalah data review aplikasi *tripadvisor* pada *google playstore* periode 7 Januari 2018 sampai dengan 28 februari 2023 sebanyak 1000 data diperoleh dengan menggunakan web scraping. Hasil scrapping data menunjukkan bahwa 600 (60%) pengguna mengomentari data ulasan secara positif, dibandingkan dengan 400 (40%) komentar negatif. Pada pengujian evaluasi model klasifikasi sentimen, diperoleh hasil dengan akurasi menggunakan model SVM sebesar 89,8%, model K-NN sebesar 89,02%, dan Naïve Bayes sebesar 88,65%. Sehingga disimpulkan bahwa nilai dengan akurasi terbaik dalam pemodelan klasifikasi pada *review* aplikasi *tripadvisor* yaitu menggunakan algoritma SVM dengan nilai akurasi sebesar 89,8%. Hasil pengukuran menggunakan visual *word cloud* dan visual *network explorer* menghasilkan faktor penyebab yang mempengaruhi keputusan pengguna dalam memberikan *review* positif dan negatif. Faktor penyebab yang mempengaruhi keputusan pengguna untuk memberikan *review* sentimen positif seperti, aplikasi sangat membantu, aplikasi sangat bagus, aplikasi sangat berguna dalam mencari informasi wisata, dan terima kasih kepada *tripadvisor*. Sebaliknya faktor penyebab yang mempengaruhi keputusan pengguna untuk memberikan *review* sentimen negatif seperti, aplikasi sangat jelek, aplikasi lumayan, menghapus aplikasi yang bikin berat hp, aplikasi error, dan terjadi bug, tolong diperbaiki, dan jangan terlalu sering update.

#### REFERENSI

- [1] S. Sri Utami, "Pengaruh Teknologi Informasi dalam Perkembangan Bisnis (Setyaningsih Sri Utami) PENGARUH TEKNOLOGI INFORMASI DALAM PERKEMBANGAN BISNIS."
- [2] Y. Hapsari *et al.*, "Opinion Mining Terhadap Toko Online Di Media Sosial Menggunakan Algoritma Naïve Bayes (Studi Kasus: Akun Facebook Dugal Delivery)," vol. 03, no. 02, 2018.
- [3] M. T. Dewi, A. Herdiani, and D. S. Kusumo, "Multi-Aspect Sentiment Analysis Komentar Wisata TripAdvisor dengan Rule-Based Classifier (Studi Kasus : Bandung Raya)."
- [4] R. Y. Hayuningtyas and R. Sari, "ANALISIS SENTIMEN OPINI PUBLIK BAHASA INDONESIA TERHADAP WISATA TMII MENGGUNAKAN NAÏVE BAYES DAN PSO," *Jurnal TECHNO Nusa Mandiri*, vol. 16, no. 1, p. 37, 2019, [Online]. Available: <http://nusamandiri.ac.id/>
- [5] R. Wulan Sari and D. Hartama, *Seminar Nasional Sains & Teknologi Informasi (SENSASI) Data Mining: Algoritma K-Means Pada Pengelompokkan Wisata Asing ke Indonesia Menurut Provinsi*. 2018. [Online]. Available: <http://seminar-id.com/semnas-sensasi2018.htmlPage|322>
- [6] A. Saepulrohman, S. Saepudin, and D. Gustian, "Analisis Sentimen Kepuasan Pengguna Aplikasi Whatsapp Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine," *is The Best Accounting Information Systems and Information Technology Business Enterprise this is link for OJS usf@*, vol. 6, no. 2, pp. 91–105, Dec. 2021, doi: 10.34010/aisthebest.v6i2.4919.
- [7] N. Herlinawati *et al.*, "ANALISIS SENTIMEN ZOOM CLOUD MEETINGS DI PLAY STORE MENGGUNAKAN NAÏVE BAYES DAN SUPPORT VECTOR MACHINE," 2020.
- [8] A. P. Wibowo, W. Darmawan, A. Stmik, and W. Pratama, "KOMPARASI METODE NAÏVE BAYES DAN K-NEAREST NEIGHBOR TERHADAP ANALISIS

- SENTIMEN PENGGUNA APLIKASI PEDULILINDUNGI,” 2022, [Online]. Available: <http://ejournal.stmik-wp.ac.id>
- [9] S. Fransiska and A. Irham Gufroni, “Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method,” *Scientific Journal of Informatics*, vol. 7, no. 2, pp. 2407–7658, 2020, [Online]. Available: <http://journal.unnes.ac.id/nju/index.php/sji>
- [10] B. Liu, “Sentiment Analysis and Opinion Mining,” Morgan & Claypool Publishers, 2012.
- [11] S. Fanissa, M. A. Fauzi, and S. Adinugroho, “Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking,” 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [12] A. Erfina and Y. H. Putra, “Irony Sentence Detection Techniques Using Fuzzy Historical Classifier,” in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Nov. 2019. doi: 10.1088/1757-899X/662/6/062004.
- [13] D. Joko Haryanto, L. Muflikhah, and M. Ali Fauzi, “Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion,” 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [14] F. S. Pamungkas and I. Kharisudin, “Analisis Sentimen dengan SVM,” vol. 4, pp. 628–634, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [15] A. Fauzi, A. N. Rais, M. Faittullah Akbar, and W. Gata, “46 Seminar Nasional Teknologi Informasi Universitas Ibn Khaldun,” 2018.
- [16] Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, 3(1), 50-59. doi:<https://doi.org/10.35314/isi.v3i1.335>