



UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER

Kampus I: Jl. Harsono RM No. 67, Ragunan, Pasar Minggu, Jakarta Selatan 12550
Telepon: (021) 27808121 – 27808882
Kampus II: Jl. Raya Perjuangan, Marga Mulya, Bekasi Utara, Jawa Barat, 17142
Telepon: (021) 88955882, Fax.: (021) 88955871
Web: fasilkom.ubharajaya.ac.id, E-mail: fasilkom@ubharajaya.ac.id

SURAT TUGAS

Nomor: ST/127/XI/2021/B-FASILKOM-UBJ

1. Dasar: Kalender Akademik Ubhara Jaya Tahun Akademik 2021/2022.
2. Dalam rangka mewujudkan Tri Dharma Perguruan Tinggi untuk Dosen di Universitas Bhayangkara Jakarta Raya maka dihimbau untuk melakukan Penelitian.
3. Sehubungan dengan hal tersebut di atas, maka Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya menugaskan:

NO.	NAMA	NIDN	JABATAN	KETERANGAN
1.	Herlawati, S.Si., M.M., M.Kom.	0311097302	Dosen Tetap Prodi Informatika	Sebagai Penulis Kedua
2.	Prima Dina Atika, S.Kom., M.Kom.	0311037107	Dosen Tetap Prodi Informatika	Sebagai Penulis Ketiga
3.	Fata Nidaul Khasanah, S.Kom., M.Eng.	0327059202	Dosen Tetap Prodi Informatika	Sebagai Penulis Keempat
4.	Ajif Yunizar Pratama Yusuf, S.Si., M.Eng.	0328068603	Dosen Tetap Prodi Informatika	Sebagai Penulis Kelima

Membuat Artikel Ilmiah dengan judul **“Analisis Sentimen Pada Situs Google Review Dengan Naïve Bayes dan Support Vector Machine”** dengan menerima LoA pada tanggal 28 November 2021 untuk dipublikasikan di media Jurnal Komtika (Komputasi dan Informatika), Vol. 5, No. 2, November 2021.

4. Demikian penugasan ini agar dapat dilaksanakan dengan penuh rasa tanggung jawab.

Bekasi, 29 November 2021
DEKAN FAKULTAS ILMU KOMPUTER

Dr. Tyastuti Sri Lestari, S.Si., M.M.
NIP. 1408206

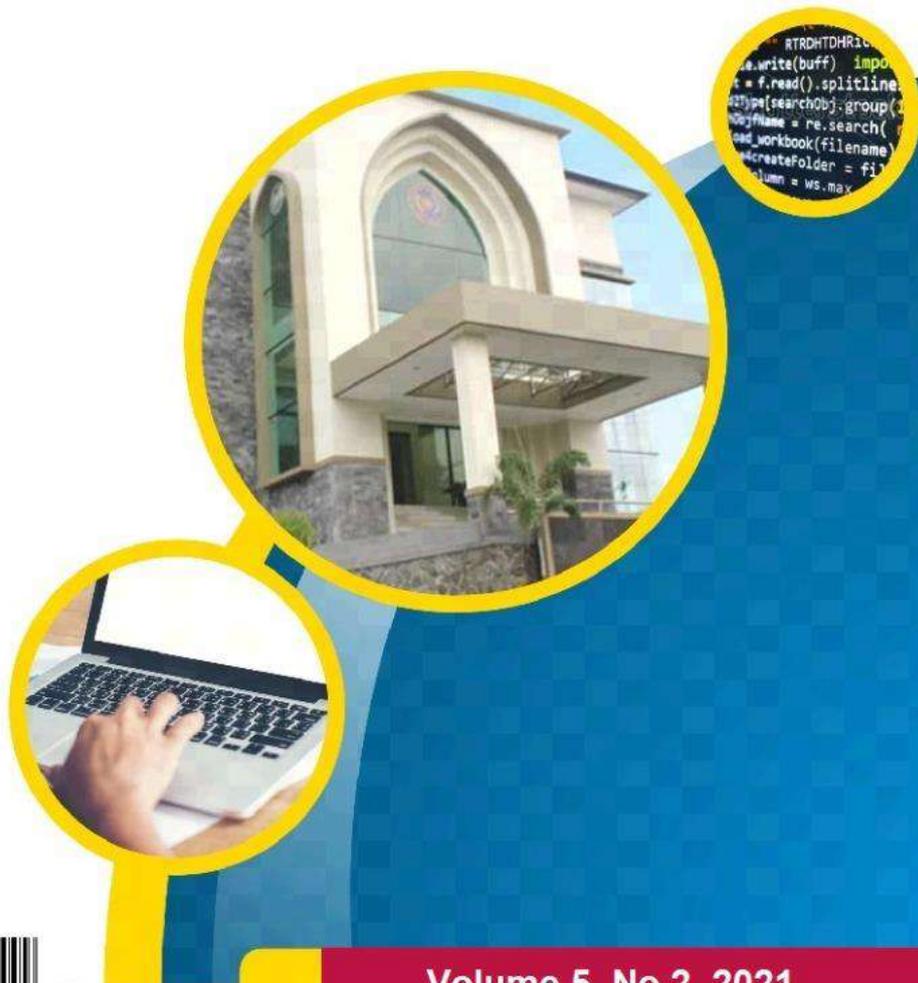
COVER
JURNAL KOMTIKA (KOMPUTASI DAN INFORMATIKA)
VOL 5 NO 2 (2021) NOVEMBER 2021



eISSN 2580-734X
ISSN 2580-2852

JURNAL KOMTIKA

Komputasi dan Informatika



Volume 5, No.2, 2021

DAFTAR ISI
JURNAL KOMTIKA (KOMPUTASI DAN INFORMATIKA)
VOL 5 NO 2 (2021) NOVEMBER 2021

Articles

[Prediction of University Student Performance Based on Tracer Study Dataset Using Artificial Neural Network](#)

 72 - 82

 Zahrina Aulia Adriani, Irma Palupi

 PDF

 Read Statistic: 84

[Instance Selection dengan Naïve Bayes pada Klasifikasi Kanker Serviks](#)

 83 - 91

 Fari Katul Fikriah

 PDF

 Read Statistic: 78

[Implementasi Metode Forward Chaining, Certainty Factor dan Dempster Shafer pada Sistem Pakar Diagnosis Penyakit Gigi dan Mulut](#)

 92 - 102

 Siti Nurajizah, Ita Yulianti, Elin Panca Saputra, Rani Kurnia Dewi

[Rancang Bangun Sistem Penilaian Produk Virtual Reality Berbasis Website Di PT.Shinta VR](#)

 103 - 111

 Ridho Bramasta, Lisda Fitriana Masitoh

 PDF

 Read Statistic: 105

[Penerapan Data Mining Association Rule Menggunakan Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel](#)

 112 - 121

 Guntoro Guntoro, Charles Parmonangan Hutabarat

 PDF

 Read Statistic: 112

[Sistem Pakar Diagnosa Penyakit pada Hewan Kucing Berbasis Web](#)

 122 - 131

 Faiz Zaki Ramadhan, Gilang Aditya, Purnama Dileon Yamora Nainggolan, Faisal Dharma Adhinata

 PDF

 Read Statistic: 133

[Pengembangan Sistem Pemantauan dan Deteksi Serangan pada Ekosistem Rumah Cerdas](#)

 132 - 139

 Dedy Hariyadi, Chanief Budi Setiawan, Barep Setiyadi



 Read Statistic: 155

[Sistem Pendukung Keputusan Penilaian Kinerja Tenaga Kependidikan \(TENDIK\) Dengan Menggunakan Metode SMARTER](#)

 140 - 152

 Dito Putro Utomo, Bister Purba



 Read Statistic: 85

[Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine](#)

 153 - 163

 Rahmadya Trias Handayanto, Herlawati Herlawati, Prima Dina Atika, Fata Nidaul Khasanah, Ajif Yunizar Pratama Yusuf, Dwi Yoga Septia

DEWAN REDAKSI
JURNAL KOMTIKA (KOMPUTASI DAN INFORMATIKA)
VOL 5 NO 2 (2021) NOVEMBER 2021

[Home](#) / [Editorial Team](#)

Editorial Team

Editor in Chief

[Maimunah](#), SCOPUS ID: [57212675206](#), Universitas Muhammadiyah Magelang, Indonesia

Managing Editor

[Pristi Sukmasetya](#), SCOPUS ID: [57189213600](#), Universitas Muhammadiyah Magelang, Indonesia

Editorial Board

[Deki Satria](#), SCOPUS ID: [57200214510](#), BRI University, Indonesia

[Esti Suryani](#), SCOPUS ID: [57191243845](#), Universitas Sebelas Maret, Indonesia

[Herlawati](#), SCOPUS ID: [55613443500](#), Universitas Bhayangkara Jakarta Raya, Indonesia

Analisis Sentimen Pada Situs Google Review dengan *Naïve Bayes* dan *Support Vector Machine*

Rahmadya Trias Handayanto ¹, Herlawati ^{2*}, Prima Dina Atika ², Fata Nidaul Khasanah ²,
Ajif Yunizar Pratama Yusuf ², Dwi Yoga Septia ²

¹ Fakultas Teknik, Universitas Islam 45

² Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya

*email: herlawati@ubharajaya.ac.id

DOI: <https://doi.org/10.31603/komtika.v5i2.6280>

Received: 25-11-2021, Revised: 26-11- 2021, Accepted: 28-11-2021

ABSTRACT

Tourism is the sources of income which is influenced by customer satisfaction. One way to know customer satisfaction is feedback, one of which is a review using an application. One of the feedback applications is Google Review. Such applications are have been widely used, for example in this study in this case study, Summarecon Mal Bekasi, can reach 60,000 comments. To find out the sentiment of the large number of comments, it is necessary to use computational tools. The current research applies sentiment analysis using the Naïve Bayes method and the Support Vector Machine. Data retrieval is done by web scrapping technique. Furthermore, the comment data is processed by pre-processing and labelling using the Lexicon dictionary. The process of applying sentiment analysis is carried out to determine whether the comments are positive or negative. In this study, the accuracy of the Naïve Bayes and Support Vector Machine methods in conducting sentiment analysis on the Summarecon Mal Bekasi review with a data of 2,143 comments with an accuracy for Naïve Bayes and Support Vector Machine 80.95% and 100% respectively. A Jason-style application is built to show the implementation in Flask framework.

Keywords: *Sentiment Analysis, Summarecon Mal Bekasi, Naïve Bayes, Lexicon, RStudio..*

ABSTRAK

Sektor pariwisata adalah salah satu sumber pemasukan negara yang dipengaruhi oleh kepuasan pelanggan. Salah satu cara mengetahui kepuasan pelanggan adalah umpan balik, salah satunya adalah review dari aplikasi. Salah satu aplikasi umpan balik adalah Google Review. Aplikasi yang berupa ulasan ini banyak digunakan, misalnya untuk studi kasus dalam penelitian ini, Summarecon Mal Bekasi, dapat mencapai 60.000 komentar. Untuk mengetahui sentimen dari komentar-komentar yang banyak tersebut perlu menggunakan alat bantu komputasi. Penelitian menerapkan *sentiment analysis* menggunakan metode *Naïve Bayes* dan *Support Vector Machine*. Pengambilan data dilakukan dengan teknik *web scrapping* kemudian dilakukan *pre-processing* dan pelabelan menggunakan kamus Lexicon. Penerapan analisis sentimen dilakukan untuk menentukan komentar tersebut bersifat positif atau negatif. Dalam penelitian ini dapat diketahui akurasi metode *Naïve Bayes* dan *Support Vector Machine* dalam melakukan analisis sentimen pada ulasan Summarecon Mal Bekasi dengan data sebanyak 2.143 komentar dengan akurasi untuk *Naïve Bayes* dan *Support Vector Machine* berturut-turut 80,95% dan 100%. Sebuah prototipe berbasis Jason pada penelitian ini diimplementasikan dengan framework Flask.

Kata-kata kunci: *Analisis Sentimen, Summarecon Mal Bekasi, Naïve Bayes, Support Vector Machine.*

PENDAHULUAN

Hampir tiap kota di Indonesia menjadikan pariwisata sebagai sumber pendapatan daerah. Salah satunya adalah, Kota Bekasi yang telah melakukan pembangunan di berbagai bidang diantaranya adalah pariwisata, selain industri yang sudah sejak lama berkembang. Menurut data dari Badan Pusat Statistik Provinsi Jawa Barat (<http://jabar.bps.go.id>), Kota Bekasi mengalami peningkatan jumlah kunjungan wisatawan. Dalam rentang tahun 2017-2019 Kota Bekasi mempunyai jumlah pengunjung sebanyak 1.958.551 wisatawan. Sebanyak 1.899.915 adalah pengunjung nusantara dan 58.636 lainnya adalah pengunjung mancanegara.

Karena terus berkembangnya Kota Bekasi, salah satu hal yang menjadi daya tarik pelaku bisnis adalah semakin menariknya usaha pembangunan mal-mal sebagai pusat pembelanjaan, media rekreasi, maupun sebagai media hiburan. Beberapa mal dapat dijumpai di Kota Bekasi seperti Grand Mal, Bekasi Cyber Park, dan lain-lain yang dapat dijumpai di aplikasi seperti Google Map [1].

Semakin banyaknya mal yang ada, maka semakin ketat juga persaingan antar pelaku bisnis tersebut. Berbagai usaha untuk menarik pengunjung juga harus dilakukan agar industri ini tetap bertahan dan berkembang. Sebagai salah satu faktor penting bagi pengunjung yang ingin melakukan kunjungan ke suatu mal adalah berupa ulasan pengunjung lain yang telah melakukan atau pernah berkunjung sebelumnya yang dapat digunakan sebagai referensi. Ulasan yang terdapat pada *Google Review* juga dapat menjadi pertimbangan untuk menentukan pilihan mal yang terbaik. Selain untuk menjadi bahan referensi pengunjung, ulasan pada *Google Review* juga dapat menjadi bahan evaluasi dari pihak mal agar dapat melakukan perbaikan demi kenyamanan dan kepuasan para pengunjung.

Banyaknya ulasan dari para pengunjung maka membutuhkan waktu yang lama untuk mengulas satu per satu. Oleh karena itu untuk mengumpulkan ulasan dan menganalisis dengan jumlah data berskala besar diperlukan suatu teknik. Salah satu teknik yang dapat dilakukan untuk menyelesaikan masalah ini adalah Analisis Sentimen.

Analisis sentimen merupakan studi komputasi dari opini-opini, sentimen, serta emosi yang diekspresikan dalam teks untuk menghasilkan kesimpulan sentimen yang didapatkan yaitu positif atau negatif [2]. Analisis sentimen digunakan untuk memudahkan pengguna pada proses memahami sentimen sehingga dapat melakukan penentuan keputusan pada suatu objek [3].

Penelitian ini bermaksud mengetahui sentimen dari data ulasan pengunjung Mal di Bekasi lewat aplikasi *Google Review*. Pengklasifikasian sentimen pada ulasan Mal di Bekasi akan meneliti tingkat akurasi metode klasifikasi *Naive Bayes* dan *Support Vector Machine*. Hasil klasifikasi juga diuji dengan survei langsung kepada masyarakat di kota Bekasi.

Terdapat beberapa penelitian sebelumnya yang membahas tentang pengaplikasian algoritma *Naive Bayes* dalam melakukan analisis sentimen ulasan atau *review* tentang penggunaan suatu aplikasi atau ulasan mengenai suatu tempat. *Naive Bayes*, *Random Forest*, dan *Support Vector Machine* merupakan metode untuk menganalisis sentimen ulasan pengguna aplikasi Ruang Guru pada *Google Play Store*. Tujuan penelitian ini untuk menentukan sentimen ulasan yang diberikan pelanggan terhadap aplikasi tersebut apakah negatif atau positif. Hasil akurasi untuk metode *Naive Bayes* pada penelitian ini yaitu sebesar 94,16%. [4]

Penggunaan metode *Naïve Bayes* dalam menyelesaikan masalah mengenai ekstraksi opini masyarakat terhadap opini film pada aplikasi Twitter [5]. Pada penelitiannya, peneliti menggunakan data *tweet* yang mengandung *hashtag* #judul film. Data yang diambil dan digunakan peneliti dimulai bulan Januari 2015 sampai dengan Desember 2015 sebanyak 500 kalimat. Hasil akurasi dari pengujian ini dibagi menjadi lima iterasi, dimana akurasi iterasi pertama sebanyak 90%, iterasi kedua sebanyak 72%, iterasi ketiga sebanyak 82%, iterasi keempat sebanyak 80%, dan iterasi kelima sebanyak 88%.

Penelitian tentang klasifikasi pada ulasan hotel-hotel di Singapura pada situs *TripAdvisor* menggunakan metode *Naïve Bayes*. Dimana pada situs tersebut, data hotel yang diperoleh sebanyak 33 restoran. Pada proses klasifikasi sentimen, peneliti menggunakan data sebanyak 1000 data sebagai data latih yang diperoleh dari situs *Kaggle*. Hasil dari penelitian ini divisualisasikan oleh framework RShiny yang menampilkan *wordcloud* atau kata apa saja yang paling banyak muncul pada ulasan serta histogram atau grafik yang menampilkan hasil frekuensi data. Akurasi pada penelitian ini diperoleh sebesar 73,33%. [6]

METODE DAN DATA

Pengumpulan data pada penelitian ini mengambil data dari ulasan mengenai Summarecon Mall Bekasi pada situs *Google Review* dan pendapat masyarakat Desa Pantai Hurip mengenai Summarecon Mal Bekasi yang diambil menggunakan angket.

Google Review

Google review merupakan salah satu dari fitur Google maps. Fitur sederhana yang mampu berdampak besar ini digunakan untuk memberi nilai angka dan ulasan berupa teks dan foto pada sebuah tempat. Di era industry 4.0 yang ditandai dengan tersedianya *big data*, pengguna dengan mudah memberikan nilai sesuai dengan faktor-faktor kepuasan pengunjung.

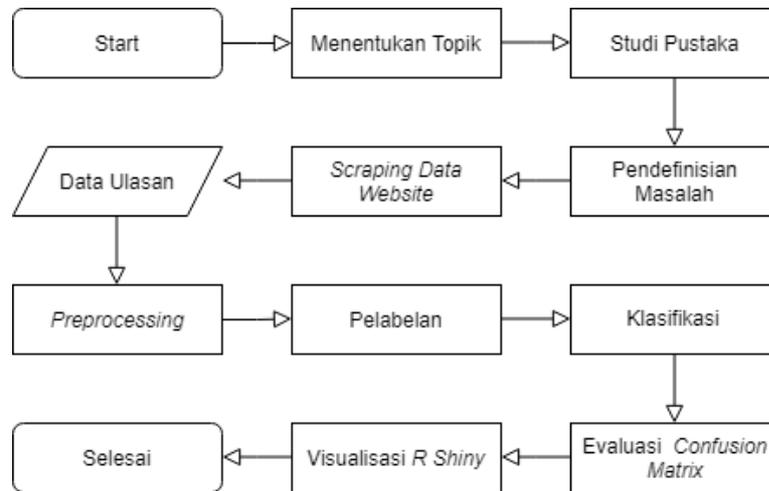
Data Summarecon Mall Bekasi (SMB) yang dicuplik dengan aplikasi *web scrapping* merupakan data yang akan dianalisis berasal dari data ulasan pada situs *Google Review*. SMB merupakan salah satu mal di Bekasi. Aplikasi tersebut ditambahkan sebagai *extension* pada *web browser* chrome dengan nama *Data Miner*.

Metode Machine Learning

Dalam penelitian ini, proses analisis data menggunakan *Microsoft Excel 2010* dan *Python*. Metode analisis data yang dilakukan diantaranya yaitu: 1) *Collection Data*, digunakan untuk memperoleh data ulasan mengenai Summarecon Mall Bekasi yang berasal dari situs *Google Review*. 2) *Preprocessing*, tahap ini untuk mengubah data menjadi data yang terstruktur dan membersihkan data dari *noise*, 3) Pembobotan Kata, dari pelabelan data kemudian dibagi dua kelas sentimen yaitu sentimen positif dan sentimen negatif. 4) Klasifikasi *Support Vector Machine* dan *Naïve Bayes* dipakai untuk pengklasifikasian ulasan berdasarkan ulasan positif dan ulasan negatif, serta mengetahui tingkat akurasi, dan 5) Evaluasi *Confusion Matrix*, untuk mengetahui tingkat akurasi *Support Vector Machine* dan *Naïve Bayes*.

Kerangka Penelitian

Langkah-langkah yang dilakukan dalam penelitian ini ditunjukkan dalam Gambar 1.



Gambar 1. Kerangka Penelitian

Setelah menentukan topik, melakukan studi pustaka mengenai penelitian-penelitian terdahulu sebagai dasar penelitian. *Scraping* data website bertujuan untuk mengambil data yang berasal dari ulasan mengenai Summarecon Mall Bekasi dalam *Google Review*. Pengambilan data ini menggunakan *software RStudio* dengan versi R.4.0.4. Data mentah yang sudah berhasil diambil kemudian disimpan dengan format *Comma Separate Value (CSV)*. Untuk klasifikasi perlu proses awal yang membersihkan data dari kata-kata yang tidak penting. Tahap ini dikenal dengan nama *preprocessing*.

Setelah melalui tahap *preprocessing* untuk pelabelan data dapat dilakukan analisis sentimen. Prosesnya menggunakan kamus *Lexicon* dan dilakukan secara otomatis dengan cara menghitung skor sentimen dan program *R*. Kelas sentimen dilakukan dengan cara selisih antara skor jumlah kata (+) dengan skor jumlah kata (-) dalam setiap kalimat yang diulas [7]. Kata-kata yang memperoleh skor $\leq 0 \rightarrow (-)$, sebaliknya kata-kata yang memperoleh skor $> 0 \rightarrow (+)$

Mempelajari pola data menggunakan data latih dilakukan untuk proses klasifikasi. Data tersebut yang berisi *training data* ulasan (+) dan *training data* ulasan (-) digunakan oleh algoritma *Naïve Bayes* dan *Support Vector Machine* untuk pengecekan pola data berdasarkan tipe data per kelas. Hasil dari proses yang dikenal dengan nama *Machine Learning* ini kemudian dapat diuji menggunakan data uji untuk mengetahui akurasi. Hasil akurasi terbaik akan diterapkan dalam satu prototipe aplikasi berbasis web untuk memprediksi sentimen sebuah ulasan/komentar. Framework yang digunakan adalah *Flask* dengan bahasa *Python*.

Shiny merupakan salah satu *framework* dalam *R* untuk membangun aplikasi berbasis website untuk visualisasi dari hasil analisis sentimen. Terdapat dua bentuk visualisasi yaitu *Word cloud* dan *histogram*. Untuk menampilkan kata-kata yang sering muncul besar kecilnya menampilkan banyaknya kemunculannya dalam data, sehingga kata terbanyak akan ditampilkan dalam ukuran yang lebih besar dan sebaliknya maka digunakan *word cloud*. 2) *Histogram* merupakan hasil dari banyaknya data yang ditampilkan dalam bentuk grafik batang, dan tampilan diagram batang yang semakin tinggi diberikan untuk data yang kemunculannya semakin banyak, sebaliknya, ukuran batang akan semakin pendek jika kemunculannya semakin sedikit [6].

Text Mining

Ciri-ciri berdimensi tinggi, ada *noise*, dan struktur teks yang tidak bagus pada data merupakan ciri pada teks ketika dilakukan *text mining*, maka dilakukan *text preprocessing* terlebih dahulu dengan cara penyeleksian kata, untuk memperoleh jenis data yang siap diolah oleh data mining. Untuk memiliki arti yang lebih sempit, maka kata-kata tersebut diubah menjadi bagian yang lebih kecil lagi [9].

Data Collection

Teknik *web scraping* dilakukan untuk pengambilan data. Proses mengambil sebuah dokumen semi-terstruktur dari internet, biasanya berupa halaman web berbentuk bahasa markup seperti HTML atau XHTML, lalu menganalisis dokumen tersebut untuk diperoleh suatu data tertentu dari halaman tertentu untuk dipakai kepentingan lain disebut *Web Scraping* [10]. *Web scraping* sering dikenal sebagai *screen scraping*. [11].

Preprocessing

Preprocessing merupakan tahap untuk membangun satu *set* fitur yang relevan dari *text* pada dokumen. Semua *set* fitur dipilih untuk koleksi dokumen yang disebut *representational* model dan setiap dokumen diwakili oleh nilai *numeric vector*. Tujuan dari *text preprocessing* adalah untuk mengekstrak *vector feature* berkualitas tinggi untuk setiap dokumen, sehingga *values* yang diambil adalah *values* yang bernilai tinggi atau penting [12].

Pembobotan Kata

Pembobotan kata merupakan proses pemberian bobot untuk setiap kata yang terdapat dalam sebuah dokumen. [13] Proses pembobotan kata dilakukan secara otomatis dengan cara menghitung skor sentimen menggunakan kamus *Lexicon* dan dilakukan menggunakan program *R*. Selisih skor jumlah kata positif dengan skor jumlah kata negatif dalam setiap kalimat ulasan adalah cara menentukan kelas sentimen [7]. Jika Kalimat yang memiliki skor $> 0 \rightarrow$ kelas positif sedangkan jika kalimat yang memiliki skor $\leq 0 \rightarrow$ kelas negatif,

Setelah pelabelan data maka selanjutnya data terbagi dua yaitu data latih dan data uji yang dibagi secara manual dengan cara mengacak keseluruhan data dari tiap masing-masing kelas.

Naïve Bayes

Naïve Bayes merupakan salah satu pengklasifikasi statistik dimana pengklasifikasinya dapat melakukan prediksi probabilitas keanggotaan kelas dari suatu data tuple yang akan masuk ke dalam kelas tertentu sesuai dengan perhitungan probabilitas [9].

Naïve Bayes dapat melakukan prediksi probabilitas keanggotaan kelas suatu data *tuple* yang akan masuk ke dalam suatu kelas, sesuai perhitungan probabilitasnya. Keuntungan metode ini adalah adanya jumlah data pelatihan yang kecil untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi [14].

Rumus *Naïve Bayes* secara umum yaitu [9]:

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

Keterangan:

X = Data dengan kelas yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$ = Probabilitas hipotesis H (prior prob.)

$P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut

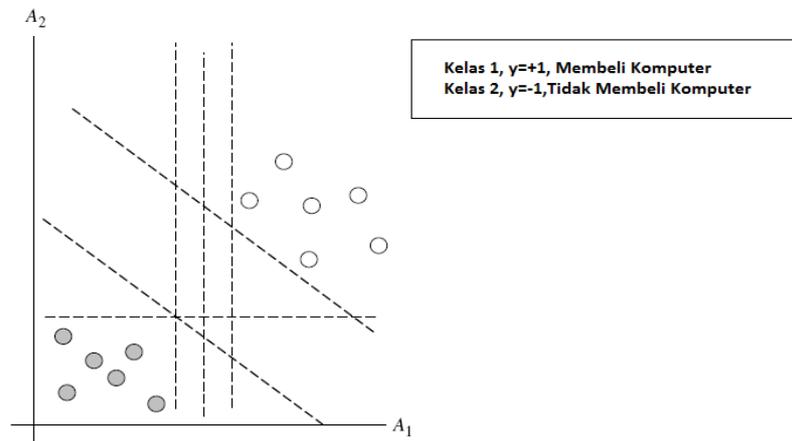
$P(X)$ = Probabilitas dari X

Aturan Bayes yaitu:

Apabila $P(h_1|x) < P(h_2|x)$, maka x diklasifikasikan sebagai h_2 . Pernyataan $P(h_1|x)$ menyatakan bahwa probabilitas hipotesis h_1 berdasarkan kondisi x terjadi, begitu pula dengan h_2 . Oleh karena itu dapat diperoleh klasifikasi dari x sesuai dengan probabilitas terbesar diantara probabilitas x terhadap semua kelas [9].

Support Vector Machine

Support Vector Machine menggunakan support vector, yaitu titik-titik dengan jarak terjauh yang memisahkan satu kelas dengan kelas lainnya. Gambar 2 memperlihatkan bagaimana Support Vector Machine memisahkan dua kelas data.



Gambar 2. Support Vector Machine

Pada Support Vector Machine Hyperplane pemisah diterapkan dengan rumusan:

$$w_0 + w_1x_1 + w_2x_2 = 0 \quad (2)$$

sehingga, setiap titik yang terletak di atas hyperplane pemisah memenuhi:

$$w_0 + w_1x_1 + w_2x_2 > 0 \quad (3)$$

maka, setiap titik yang terletak di bawah hyperplane pemisah memenuhi:

$$w_0 + w_1x_1 + w_2x_2 < 0 \quad (4)$$

Evaluasi

Proses pengklasifikasian suatu baris data dalam proses evaluasi ada 4 kemungkinan yang akan terjadi True positive (TP), False negative (FN), False positive (FP) dan True negative (TN) [15]. Tabel 1 disajikan confusion matrix yang merupakan hasil dari klasifikasi biner pada suatu dataset.

Tabel 1. Confusion Matrix

Kelas Prediksi	Kelas Aktual	
	Sentimen Positif	Sentimen Negatif
Sentimen Positif	TP	FN
Sentimen Negatif	FP	TN

Selain *confusion matrix*, pengevaluasian metode *Naïve Bayes* juga dilakukan dengan cara menguji akurasi atau ketepatan metode *Naïve Bayes* dalam melakukan klasifikasi analisis sentimen yang dihitung dengan persamaan 2.

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi Tepat}}{\text{Total Jumlah Prediksi}} \quad (2)$$

Langkah-langkah yang ditempuh untuk menerapkan metode klasifikasi sentimen antara lain: *text preprocessing*, pembobotan kata, pembuatan data uji dan data latih, klasifikasi, evaluasi, dan implementasi (visualisasi dengan *RShiny* dan web-app dengan Flask).

HASIL DAN PEMBAHASAN

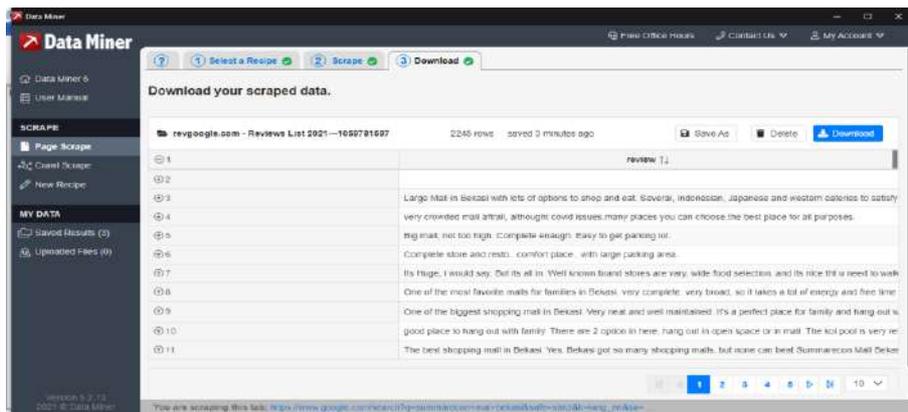
Penarikan Data

Proses penarikan data dilakukan dengan mengambil data ulasan pada *Google Review* untuk analisis sentimen, dan data yang telah diambil sebelumnya tentang pendapat masyarakat mengenai Summarecon Mall Bekasi.

Web Scraping

Pengambilan data untuk analisis sentimen diambil menggunakan teknik *web scraping*. Atribut yang diperlukan untuk *web scraping* diantaranya sebagai berikut: a) *Web Browser Microsoft Edge* versi 91.0.864.48; b) *Extension Data Miner* versi 5.2.73, c) Koneksi internet (*online*).

Untuk mendapatkan data ulasan Summarecon Mall Bekasi, dilakukan terlebih dahulu dengan menuliskan *keyword* “Summarecon Mall Bekasi” pada situs Google di kolom pencarian, hingga diperoleh halaman ulasan Summarecon Mall Bekasi seperti terlihat pada Gambar 3.



Gambar 3. Scraping menggunakan Data Miner

Gambar 3 menunjukkan tampilan ulasan pengunjung Summarecon Mall Bekasi yang diambil/*scrap* menggunakan Data Miner.

Evaluasi

Setelah dilakukan proses pelabelan data, selanjutnya adalah proses klasifikasi dengan metode *Naïve Bayes* dan *Support Vector Machine*. Untuk mengetahui performa dari metode klasifikasi yang digunakan maka perlu dilakukan pengujian dan evaluasi. Dari klasifikasi hasilnya akan ditampilkan dalam bentuk *confusion matrix*. Data yang digunakan pada *confusion matrix* adalah *data testing* yang berjumlah 714 data. *Confusion matrix* terdapat dua kelas yaitu *prediction* dan *actual*. Model *confusion matrix* ditunjukkan pada Tabel 1.

Tabel 1. Hasil *Confusion Matrix*

Kelas Prediksi	Kelas Aktual	
	Sentimen Positif	Sentimen Negatif
Sentimen Positif	225	132
Sentimen Negatif	4	353

Berdasarkan penelitian, diperoleh ulasan Summarecon Mall Bekasi pada kelas *actual* positif dan diklasifikasikan dengan benar (*True Positive / TP*) sebanyak 353 ulasan. Sedangkan hasil klasifikasi salah (*False Positive / FP*) pada kelas *actual* positif adalah sebanyak 132 ulasan. Untuk hasil klasifikasi benar pada kelas *actual* negatif (*True Negative / TN*) terdiri dari 225 ulasan, sedangkan hasil klasifikasi salah pada kelas *actual* negatif (*False Negative*) adalah 4 ulasan. Dari hasil *confusion matrix* maka dilakukan perhitungan nilai akurasi dengan persamaan 2 sehingga diperoleh akurasi 0,8095 (80,95%).

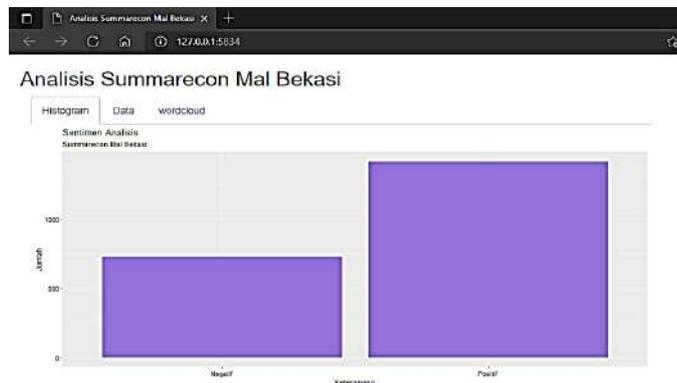
```
(base) D:\AI\flask>python svm_linear.py
Results for SVC(kernel=linear)
Training time: 0.106708s; Prediction time: 0.018937s
positive: {'precision': 1.0, 'recall': 1.0, 'f1-score': 1.0, 'support': 329}
negative: {'precision': 1.0, 'recall': 1.0, 'f1-score': 1.0, 'support': 139}
```

Gambar 4. Hasil Klasifikasi Support Vector Machine

Gambar 4 menunjukkan akurasi *Support Vector Machine* melebihi *Naïve Bayes* dengan *presisi*, *recall*, dan *f1-score* seluruhnya 100%. Sehingga untuk implementasi *Support Vector Machine* layak dipilih.

Implementasi

Visualisasi digunakan pada penelitian ini agar dapat dilihat hasil dari pengklasifikasian analisis sentimen mengenai ulasan Summarecon Mall Bekasi. Proses visualisasi menggunakan framework khusus yang akan mengeluarkan output berupa halaman web bernama RShiny. Gambar 5 menunjukkan rekapitulasi sentimen terhadap Summarecon Mall Bekasi.



Gambar 5. Visualisasi Histogram dengan RShiny

Halaman *wordcloud* menampilkan kumpulan kata-kata yang paling sering muncul pada ulasan dan ukuran dari setiap kata menggambarkan banyaknya kata tersebut muncul. Semakin

besar ukuran kata, maka semakin banyak dan sering juga kata tersebut ada pada ulasan Summarecon Mall Bekasi. Visualisasi *wordcloud* dapat dilihat pada Gambar 6.



Gambar 6. *Wordcloud*

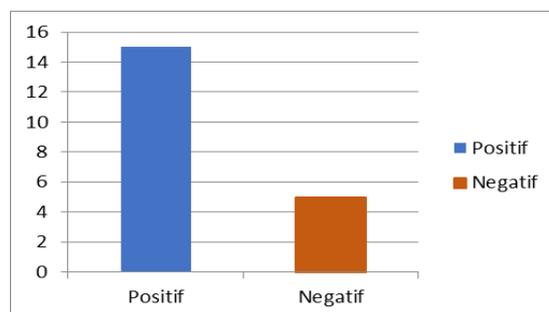
Sebuah aplikasi berbasis Web dengan framework Flask ditunjukkan pada Gambar 7. Dengan masukan sebuah komentar, aplikasi akan mengirimkan output berformat Jason berupa sentiment positif atau negatif yang dapat digunakan oleh aplikasi lain, misalnya PHP-MySQL.



Gambar 7. Implementasi Aplikasi Web

Perbandingan Data Ulasan Google Review dengan Pendapat Masyarakat

Gambar 8 menunjukkan hasil survei terhadap masyarakat. Dari 20 responden yang telah mengisi angket mengenai Summarecon Mal Bekasi dan sudah diakumulasikan serta dimasukkan kedalam kategori positif dan negatif, dapat diketahui bahwa 15 responden berpendapat positif, dan 5 responden berpendapat negatif.



Gambar 8. Grafik Kategori Pendapat Masyarakat Mengenai Summarecon Mall Bekasi

Berdasarkan data yang sudah diolah menggunakan teknik analisis sentimen dan data pendapat masyarakat, opini positif terhadap Summarecon Mall Bekasi masih terlihat lebih tinggi dibandingkan dengan opini negatif.

Dari 2143 data ulasan yang sudah diambil pada situs *Google Review* dan sudah diklasifikasikan menjadi sentimen positif dan negatif, 66% diantaranya berkomentar positif dan 34% berkomentar negatif.



Gambar 9. Perbandingan Data *Google Review* dan Pendapat Masyarakat

Sedangkan untuk data pendapat masyarakat Desa Pantai Hurip, dari 20 responden yang telah mengisi angket yang berisikan pernyataan-pernyataan mengenai Summarecon Mall Bekasi, 75% diantaranya berpendapat positif dan 25% berpendapat negatif.

KESIMPULAN

Berdasarkan hasil pembahasan maka dapat disimpulkan bahwa teknik *web scraping* dapat digunakan untuk mendapatkan data yang besar dari halaman *website* secara otomatis sehingga mempermudah proses pengambilan data. Data ulasan mengenai Summarecon Mall Bekasi pada situs *Google Review* cenderung memuaskan. Hasil pengujian menunjukkan *Support Vector Machine* layak diimplementasikan mengingat akurasi yang lebih baik dari *Naïve Bayes*. Untuk riset berikutnya pada bidang *Natural Language Processing (NLP)*, khususnya sentiment analysis, perlu menerapkan *Deep Learning* baik untuk review lokasi wisata maupun bidang lain seperti kesehatan, pendidikan, pemerintahan, dan lain-lain.

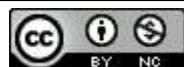
UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada LPPM Universitas Islam 45 Bekasi dan LPPMP Universitas Bhayangkara Jakarta Raya, koresponden survey, dan pihak lain yang telah mendukung penelitian ini. Penulis juga mengucapkan terimakasih kepada para reviewer yang telah memberikan saran-sarannya.

DAFTAR PUSTAKA

- [1] Emahlia and M. Baiquni, "Distribusi dan pola pusat perbelanjaan skala besar secara spasial di kota bekasi," *Geogr. UGM*, vol. 6, no. 3, pp. 1–15, 2017.
- [2] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool, 2012.
- [3] R. Cahyani and P. P. Adikara, "Analisis Sentimen terhadap Ulasan Hotel menggunakan Boosting Weighted Extreme Learning Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 8, pp. 7767–7773, 2019.

- [4] E. Fitri, “Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine,” *J. Transform.*, vol. 18, no. 1, p. 71, 2020, doi: 10.26623/transformatika.v18i1.2317.
- [5] F. Ratnawati, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter,” *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018, doi: 10.35314/isi.v3i1.335.
- [6] V. A. Permadi, “Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura,” *J. Buana Inform.*, vol. 11, no. 2, p. 140, 2020, doi: 10.24002/jbi.v11i2.3769.
- [7] S. R. Kurniasari, “s,” p. 121, 2018.
- [8] F. U. Haq, “Penggunaan Google Review Sebagai Penilaian Kepuasan Pengunjung Dalam Pariwisata,” *Tornare*, vol. 2, no. 1, p. 10, Jan. 2020, doi: 10.24198/tornare.v2i1.25826.
- [9] J. Han and M. Kamber, “Data Mining: Data Mining Concepts and Techniques,” Dec. 2006, doi: 10.1109/ICMIRA.2013.45.
- [10] M. Turland, *php/architect’s Guide to Web Scraping with PHP*. Marco Tabini & Associates Inc, Canada, 2010.
- [11] A. Josi, L. A. Abdillah, and Suryayusra, “Penerapan teknik web scraping pada mesin pencari artikel ilmiah,” 2014.
- [12] M. Perovšek, J. Kranjc, T. Erjavec, B. Cestnik, and N. Lavrač, “TextFlows: A visual programming platform for text mining and natural language processing,” *Sci. Comput. Program.*, vol. 121, pp. 128–152, Jun. 2016, doi: 10.1016/j.scico.2016.01.001.
- [13] G. Wilson, R. Devillers, and O. Hoeber, “Fuzzy Logic Ranking for Personalized Geographic Information Retrieval,” 2013, pp. 111–123.
- [14] A. V. Febrianti, “Analisis Sentimen Data Ulasan Pengunjung Objek Wisata Lawang Sewu Kota Semarang,” 2020.
- [15] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

Journal Profile

Jurnal Komtika (Komputasi dan Informatika)
 eISSN : 2580734X | pISSN :
 Universitas Muhammadiyah Magelang

Sinta S4
 Sinta Score

Indiced by GARUDA

5 H-Index
 5 I5-Index

52 Citations
 52 5 Year Citations



2018 2019 2020 2021

Sinta Accreditations

Citation Statistics

20
10
0

2016 2019 2020 2021 2022

Search:

Page 1 of 3 | Total Records : 23

Publications	Citation
Fuzzy Inference System (FIS) dengan Metode Tsukamoto dan Mamdani dalam Menentukan Kelayakan Kenaikan Gaji Karyawan	9

Penerbit:
 Universitas Muhammadiyah
 Magelang

Website | Editor URL

Address:

SERTIFIKAT

Kementerian Riset dan Teknologi/
 Badan Riset dan Inovasi Nasional

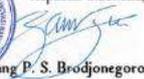



Petikan dari Keputusan Menteri Riset dan Teknologi/
 Kepala Badan Riset dan Inovasi Nasional
 Nomor 200/M/KPT/2020
 Peringkat Akreditasi Jurnal Ilmiah Periode III Tahun 2020
 Nama Jurnal Ilmiah
Jurnal Komtika (Komputasi dan Informatika)
 E-ISSN: 2580734X
 Penerbit: Universitas Muhammadiyah Magelang

Ditetapkan sebagai Jurnal Ilmiah

TERAKREDITASI PERINGKAT 4

Akreditasi Berlaku selama 5 (lima) Tahun, yaitu
 Volume 2 Nomor 1 Tahun 2018 sampai Volume 6 Nomor 2 Tahun 2022
 Jakarta, 23 Desember 2020
 Menteri Riset dan Teknologi/
 Kepala Badan Riset dan Inovasi Nasional
 Republik Indonesia,





Bambang P. S. Brodjonegoro