ANALISIS SENTIMEN DESTINASI WISATA KABUPATEN GRESIK MENGGUNAKAN *LINEAR DISCRIMINANT ANALYSIS* (LDA) DAN *SUPPORT VECTOR MACHINE* (SVM)

SKRIPSI

UIN SUNAN AMPEL
S U R A B A Y A

Disusun Oleh:

MUHAMMAD HANAFI
H96219053

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA

2023

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama	:	MUHAMMAD HANAFI
NIM	:	H9629053
Program Studi	:	Sistem Informasi
Angkatan	:	2019

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: "ANALISIS SENTIMEN DESTINASI WISATA KABUPATEN GRESIK MENGGUNAKAN *LINEAR DISCRIMINANT ANALYSIS* (LDA) DAN *SUPPORT VECTOR MACHINE* (SVM)". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 21 September 2023

Yang menyatakan,

Muhammad Hanafi

H96219053

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

NAMA	:	MUHAMMAD HANAFI
NIM	:	H96219053
JUDUL	:	ANALISIS SENTIMEN DESTINASI WISATA KABUPATEN GRESIK MENGGUNAKAN *LINEAR DISCRIMINANT ANALYSIS* (LDA) DAN *SUPPORT VECTOR MACHINE* (SVM)

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 21 September 2023

Dosen Pembimbing 1	Dosen Pembimbing 2
Mujib Ridwan, S.Kom., M.T	Subhan Nooriansyah, M.Kom.
NIP. 198604272014031004	NIP. 199012282020121010

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Muhammad Hanafi ini telah dipertahankan di depan tim penguji skripsi
di Surabaya, 27 September 2023.

Mengesahkan
Dewan Penguji,

Dosen Penguji 1

Dwi Rolliawati, MT

NIP. 197909272014032001

Dosen Penguji 2

Dr. Eng. Anang Kunaefi, M. Kom

NIP. 197911132014031001

Dosen Penguji 3

Mujib Ridwan, S.Kom., M.T

NIP. 198604272014031004

Dosen Penguji 4

Subhan Nooriansyah, M.Kom.

NIP. 199012282020121010

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya

Saepul Hamdani, M.Pd.

NIP. 19650732000031002

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN
Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

UIN SUNAN AMPEL
S U R A B A Y A

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Muhammad Hanafi
NIM : H96219053
Fakultas/Jurusan : Sains dan Teknologi / Sistem Informasi
E-mail address : hanafim3000@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :
☑ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)
yang berjudul :

ANALISIS SENTIMEN DESTINASI WISATA KABUPATEN GRESIK DENGAN *LINEAR DISCRIMINANT ANALYSIS* (LDA) DAN *SUPPORT VECTOR MACHINE* (SVM)

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara ***fulltext*** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 4 Oktober 2023

Penulis

(Muhammad Hanafi)

ABSTRAK

ANALISIS SENTIMEN DESTINASI WISATA KABUPATEN GRESIK MENGGUNAKAN *LINEAR DISCRIMINANT ANALYSIS* (LDA) DAN *SUPPORT VECTOR MACHINE* (SVM)

Oleh:

Muhammad Hanafi

Pariwisata di Indonesia naik menjadi peringkat 32 di dunia. Berdasarkan Data Kunjungan Wisata Online (DAKUWISON) wisatawan yang berkunjung ke Kabupaten Gresik menurun dibandingkan tahun sebelumnya. Kebijakan PPKM telah dihilangkan namun tidak berdampak pada pengunjung wisata. Tujuan dari penelitian ini adalah analisis sentimen ulasan untuk mengetahui persepsi dari pengunjung terkait objek wisata yang ada di Kabupaten Gresik. Metode yang digunakan adalah *Support Vector Machine (SVM)* dengan *Linear Discriminant Analysis (LDA)*. Data yang diperoleh dari proses *web scraping* sebanyak 3460 ulasan. Hasil penelitian menunjukkan model SVM dengan LDA menghasilkan nilai *F1-score* 66% lebih baik dibandingkan dengan model SVM tanpa LDA yang menghasilkan nilai *F1-score* 53%. Penerapan LDA menurunkan kompleksitas waktu dalam prediksi sentimen. Hal ini, karena reduksi dimensi oleh LDA yang membuat lebih cepat melakukan prediksi. Model tersebut digunakan untuk klasifikasi sentimen berikutnya dan menghasilkan bahwa sentimen pengunjung atau masyarakat lebih cenderung ke positif. Berdasarkan hasil klasifikasi sentimen sebanyak 511 ulasan dihasilkan 89% ulasan positif, 7% ulasan negatif dan 4% ulasan netral. Frekuensi kata yang dihasilkan menunjukkan kondisi wisata Kabupaten Gresik bagus, dan bersih. Dari frekuensi kata sentimen negatif, 5% wisata mengindikasi harga tiket masuk mahal. Pernyataan ini didukung dengan adanya kata “tiket”, dan “masuk” dalam sentimen negatif.

Kata Kunci: Analisis Sentimen, Ulasan, *Support Vector Machine, Linear Discriminant Analysis*

ABSTRACT

SENTIMENT ANALYSIS OF TOURIST DESTINATIONS IN GRESIK DISTRICT USING LINEAR DISCRIMINANT ANALYSIS (LDA) AND SUPPORT VECTOR MACHINE (SVM)

By:

Muhammad Hanafi

Tourism in Indonesia rose to 32nd in the world. Based on Online Tourist Visit Data (DAKUWISON) tourists visiting Gresik Regency decreased compared to the previous year. The PPKM policy has been eliminated but has no impact on tourist visitors. The purpose of this research is to analyze the sentiment of reviews to find out the perceptions of visitors regarding tourist attractions in Gresik Regency. The method used is Support Vector Machine (SVM) with Linear Discriminant Analysis (LDA). The data obtained from the web scraping process is 3460 reviews. The results showed that the SVM model with LDA produced a 66% F1-score value better than the SVM model without LDA which produced a 53% F1-score value. The application of LDA reduces the time complexity in sentiment prediction. This is due to the dimensionality reduction by LDA which makes prediction faster. The model was used for the next sentiment classification and resulted in that the sentiment of visitors or the public is more likely to be positive. Based on the sentiment classification results of 511 reviews, 89% positive reviews, 7% negative reviews and 4% neutral reviews were generated. The resulting word frequency shows the condition of Gresik Regency tourism is good, and clean. From the frequency of negative sentiment words, 5% of tours indicate expensive entrance ticket prices. This statement is supported by the words "ticket", and "entry" in the negative sentiment.

Keywords**: **Sentiment Analysis, Reviews, Support Vector Machine, Linear Discriminant Analysis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1. Latar Belakang

Indonesia merupakan wilayah yang luas dan memiliki keragaman ekosistem yang tak ternilai. Memiliki sumber daya alam melimpah sehingga mampu meningkatkan perekonomian terutama dari sektor wisata (Erfina & Wardani, 2022). Data peringkat pariwisata Indonesia naik pesat pada 2022. Sebelumnya Indonesia berada pada urutan 44, selama 18 bulan, peringkat wisata Indonesia melesat ke urutan 32. Data peringkat tersebut dirilis oleh *World Economic Forum* pada Mei 2022. Sandiaga Salahuddin Uno sebagai Menteri Pariwisata dan Ekonomi Kreatif (Menparekraf), menjelaskan bahwa peringkat Indonesia mengalami kenaikan yang signifikan. Sebelumnya berada pada peringkat 44 naik ke peringkat 32, pada masa pandemi Covid-19. Hal ini, memberikan dampak pada sektor pariwisata Indonesia masuk dalam peringkat 8 di kawasan Asia Pasifik. Informasi tersebut diberikan dalam event *Weekly Press Briefing* di Jakarta, Gedung Sapta Pesona, pada tanggal Senin 30 Mei 2022 (Hendriyani, 2022).

Pada provinsi di Indonesia, yang paling sesuai untuk dikelola terutama pada sektor pariwisata adalah Provinsi Jawa Timur (Parameswari dkk., 2022). Secara geografis, letak Jawa Timur dari sebelah utara berbatasan dengan Laut Jawa, sebelah timur Selat Bali, dari samping selatan Samudra Hindia, dan sebelah barat berbatasan dengan Provinsi Jawa Tengah. Luas wilayahnya terbesar diantara 6 provinsi di Pulau Jawa, yaitu sebesar 47.922 km^2. Badan Pusat Statistik (BPS) mencatat, pariwisata domestik Indonesia masih didominasi oleh arus perjalanan dari Pulau Jawa. Salah satu yang tertinggi yaitu Jawa Timur dengan jumlah perjalanan tertinggi pada 2022 sebesar 198,91 juta perjalanan, atau 27,07% dari seluruh perjalanan wisatawan nusantara yang mencapai 734,86 juta perjalanan. Secara kalkulasi mengalami peningkatan 25,41% dibandingkan tahun sebelumnya (Santika, 2023). Kondisi ini disinyalir karena sudah meredanya pandemi Covid-19 dibandingkan sebelumnya.

Jawa Timur menjadi salah satu provinsi terkaya urutan kedua setelah DKI Jakarta berdasarkan nilai Produk Domestik Regional Bruto (PDRB) sebesar Rp 2.454.499 miliar dengan jumlah penduduk sekitar 40.878.800 jiwa (Saputra, 2022). Berdasarkan daerahnya, Kabupaten Gresik menjadi urutan pertama dengan PDRB 109.313.000 per kapita per tahun mengungguli 29 Kabupaten lain di Jawa Timur menjadikan Kabupaten Gresik terkaya urutan kedua.

Berdasarkan Data Kunjungan Wisata Online (DAKUWISON) tahun 2022, pengunjung wisata di Kabupaten Gresik baik wisatawan nusantara atau mancanegara mengalami masa naik turun dengan rata-rata 300-350 ribu pengunjung (*Data Kunjungan Wisata Online*, 2018). Berbeda dengan pada tahun 2023, terjadi penurunan drastis, tercatat pada bulan April hanya 100 ribu wisatawan. Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) oleh pemerintah secara resmi mencabut kebijakan tersebut. Pencabutan kebijakan tersebut tentunya akan memberikan dampak positif khususnya bagi sektor wisata. Namun sebaliknya, berdasarkan data, wisatawan yang berkunjung ke Kabupaten Gresik menurun dibandingkan tahun sebelumnya. Pengelola wisata tidak banyak melakukan perubahan terkait harga tiket, wahana, dan fasilitas.

Faktor lain bagi pengunjung adalah ulasan pengunjung lain yang telah mendatangi objek wisata yang akhirnya menjadi referensi terhadap pengunjung baru (Herlawati dkk., 2021). Rating dan ulasan dari objek wisata dapat dilihat dari platform lain, yaitu Google Maps. Google review menjadi salah satu aspek penilaian dalam era *big data* untuk mengumpulkan informasi wisatawan ke tempat yang telah dikunjungi (Haq, 2020). Dari ulasan-ulasan tersebut akan terdapat star rating dari lokasi, dengan begitu pengunjung akan menilai dari seberapa banyak atau besar rating tersebut mengindikasi bahwa wisata itu bagus. Namun, ulasan yang diberikan yang membentuk rating dari objek tersebut belum bisa dinilai sebagai hal yang positif atau negatif. Penulis akan bisa memberikan penilaian secara bebas yang menimbulkan kemungkinan ketidaksesuaian dapat terjadi (Hesay dkk., 2021). Perlu adanya pertimbangan dan pengukuran yang tepat. Dalam menganalisis kasus tersebut dibutuhkan suatu teknik pengolahan data dengan jumlah data berskala besar. Teknik yang dapat diimplementasikan dengan kasus pengolahan data masif adalah Analisis Sentimen (Herlawati dkk., 2021).

Analisis sentimen atau disebut juga opinion mining adalah sebuah studi komputasional mengenai beberapa hal meliputi pendapat dan emosi orang (Steven & Wella, 2020). Proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini disebut Analisis Sentimen (Khofifah dkk., 2022). Penerapan dalam sektor industri dan pariwisata sudah menjadi topik yang populer untuk analisis sentimen. Dalam menentukan keputusan dan pandangan terhadap suatu objek, analisis sentimen bisa digunakan (Ginantra dkk., 2022). Pada penelitian yang telah dilakukan oleh Utami dan Erfina, ulasan wisatawan di Google Maps dilakukan analisis sentimen terhadapnya untuk mengetahui rekomendasi tempat wisata di Bali (Utami & Erfina, 2022). Penelitian yang lain oleh Erfina dan Wardani, analisis sentimen dimanfaatkan untuk mengetahui universitas terbaik berdasarkan ulasan pengunjung atau mahasiswa (Erfina & Wardani, 2022).

Dalam melakukan analisis sentimen, terdapat beberapa metode yaitu *machine learning-based*, *lexicon-based* dan *hybrid* (Sharma dkk., 2020). Pendekatan atau algoritma dari *machine Iearning* terbagi menjadi *supervised Iearning* dan *unsupervised Iearning*. Pendekatan *supervised learning* digunakan dalam penyelesaian kasus klasifikasi dan regresi, dengan melibatkan atribut output yang telah ditentukan sebelumnya, selain penggunaan atribut input (Berry dkk., 2020). *Unsupervised learning* digunakan pada kasus clustering dan asosiasi. Pengolahan didasari pada dataset tanpa label yang kemudian dikelompokkan berdasarkan angka (Kumar dkk., 2020).

Metode lainnya yaitu *lexicon-based,* dalam penerapannya tidak terdapat proses pembelajaran, data berupa *corpus* dengan bobot nilai setiap kata masing-masing. Dalam metode ini, *corpus* yang diberikan akan berpengaruh terhadap hasil analisisnya, apabila terdapat kata atau penggunaan bahasa yang tidak baku *corpus* dapat disesuaikan (Najib dkk., 2019). Metode berikutnya adalah metode pendekatan *hybrid* atau gabungan. Pendekatan ini bekerja dengan menggabungkan pendekatan *machine learning* dan pendekatan *lexicon-based* (Gupta & Joshi, 2019).

Metode *supervised learning* pada dasarnya melakukan klasifikasi, salah satu contohnya melakukan klasifikasi terhadap sentimen dari teks yang

diimplementasikan pada beberapa algoritma klasifikasi seperti, *Decision Tree, Naïve Bayes Classifier,* dan *Support Vector Machine*s (SVM) (Sharma dkk., 2020). Dalam penerapannya, sentiment analisis menggunakan metode *Naïve Bayes Classifier* dan *Decision Tree* dalam penilaian tempat tujuan wisata mampu diimplementasikan, namun dengan tingkat akurasi yang rendah yaitu 73,33% dan 60,83 (Somantri & Dairoh, 2019). Algoritma klasifikasi SVM memiliki kemampuan lebih baik dari *Naive Bayes Classifier* dan *Logistic Regression* dalam memprediksi sentimen ulasan dengan akurasi 89% (Prasetyo & Hidayatullah, 2020). Namun, pada penelitian yang lain SVM dengan LDA *feature extraction* didapatkan penurunan akurasi sebesar 78.15%, penyesuaian *feature extraction* perlu diperhatikan (Abelard & Sibaroni, 2021). SVM memerlukan waktu training yang lama dan biaya komputasi mahal (Yue dkk., 2019), penerapan *Linear Discriminant Analysis* dapat mengurangi kompleksitas waktu dan penggunaan memori dalam klasifikasi teks (Kowsari dkk., 2019).

Penelitian ini berproses dengan memfokuskan pada analisis data ulasan atau review di Google Map. Data yang dikumpulkan berdasarkan dua parameter yaitu tempat lokasi Wisata di Kabupaten Gresik dan semua komentar setiap pengunjung terhadap tempat lokasi Wisata tersebut. Metode analisis digunakan terhadap data dengan sentimen analisis berbasis *Support Vector Machine* (SVM) dengan *Linear Discriminant Analysis* (LDA) sebagai pemodelan latih sentimen analisis. Hasil dari sentimen analisis ini dapat memberikan informasi sentimen pengunjung terhadap tempat wisata di Kabupaten Gresik, sehingga para pengelola wisata mendapatkan persepsi dari masyarakat terkait objek wisata yang ada di Kabupaten Gresik sebagai bahan pertimbangan dalam memutuskan solusi yang harus dilakukan untuk perkembangan wisata berikutnya.

1.2. Perumusan Masalah

1. Bagaimana implementasi metode *Support Vector Machine* (SVM) dan *Linear Discriminant Analysis* (LDA) untuk menganalisis sentimen terhadap ulasan lokasi wisata di Kabupaten Gresik?
2. Bagaimana persepsi masyarakat terhadap ulasan yang diberikan berdasarkan hasil analisis sentimen?

1.3. Batasan Masalah

1. Data diperoleh dari ulasan setiap lokasi pada Google Maps dengan rating 3 keatas, ulasan min 100 dan rentang waktu 3 tahun terakhir.
2. Tidak termasuk wisata religi dan wisata air.
3. Data yang digunakan sebagai analisis adalah teks Bahasa Indonesia.
4. Output dari klasifikasi sentimen akan dikelompokkan ke dalam sentimen positif, negatif, dan netral,

1.4. Tujuan Penelitian

1. Menerapkan metode *Support Vector Machine* (SVM) dan *Linear Discriminant Analysis* (LDA) untuk menganalisis sentimen terhadap ulasan lokasi wisata di Kabupaten Gresik.
2. Menganalisis persepsi masyarakat terhadap ulasan yang diberikan berdasarkan hasil analisis sentimen.

1.5. Manfaat Penelitian

1. Manfaat Akademis

   Penelitian ini dapat memberikan wawasan terhadap penggunaan pendekatan machine learning sebagai metode analisis sentimen dengan algoritma SVM dan LDA. Di samping itu, sentimen yang disiratkan masyarakat terhadap destinasi wisata akan teranalisis dalam penelitian ini, sehingga dapat diketahui bagaimana sudut pandang mereka.
2. Manfaat Praktis

   Manfaat dari segi umum dari penelitian ini dapat menjadi referensi ilmiah dalam penambahan variabel untuk membantu pengelola wisata untuk menentukan solusi pengembangan objek wisatanya.

BAB II
TINJAUAN PUSTAKA

2.1. Tinjauan Penelitian Terdahulu

Dalam sebuah penelitian memerlukan tinjauan penelitian-penelitian sebelumnya yang terikat dan relevan sebagai pilihan referensi dan pengembangan terhadap penelitian yang akan direncanakan. Tujuan dari tinjauan terhadap penelitian yang lalu agar dapat memberikan informasi bagaimana metode yang diberikan oleh peneliti sebelumnya sebagai solusi dalam menguraikan masalah yang seragam serta mengetahui hal yang menjadi perbedaan antara penelitian yang dilakukan terhadap penelitian sebelumnya. Pada Tabel 2.1 terangkum beberapa hasil tinjauan dari beberapa penelitian.

Tabel 2. 1 Penelitian Terdahulu

No.	Judul	Hasil	Pembeda
1	Analisis Sentimen Wisata Alun-Alun Kota Batu menggunakan Algoritma *Support Vector Machine* (Baihaqi dkk., 2022)	1. Bertujuan untuk menganalisis perspektif pengunjung terhadap objek wisata alun-alun kota Batu. 2. Data diperoleh sebanyak 240 dan diklasifikasi dengan *Support Vector Machine* dengan parameter Kernel rbf dan nilai C 50. 3. Hasil evaluasi mendapatkan nilai akurasi 89,58%, *recall* 89,48%, *precision* 90,73%, *f-measure* 89,45%	1. Pengumpulan data bersumber pada situs *TripAdvuisor* dengan menggunakan *tools webscraper.io* terbilang masih rendah 2. Menggunakan pembobotan TF-IDF
2	Analisis Sentimen Opini Pelanggan Terhadap Aspek Pariwisata Pantai Malang Selatan Menggunakan TF-IDF dan *Support Vector Machine* (Pratama dkk., 2018)	1. Bertujuan untuk menganalisis opini pada *Tripadvisor* objek wisata pantai Malang Selatan dari sudut pandang pelanggan. 2. Metode SVM dengan menggunakan TF-IDF sebagai *Term Weighting* dengan 43 objek pantai di Malang Selatan mendapatkan 674 data 3. Hasil perolehan evaluasi dari akurasi 85%, *Precision*	1. Penelitian ini bersumber hanya pada *TripAdvisor* dan hanya pada objek wisata pantai di Malang Selatan 2. Metode teks *preprocessing* tidak terdapat *normalization* sehingga berpengaruh terhadap jumlah

No.	Judul	Hasil	Pembeda
		85%, *Recall* 87%, dan *F1-Score* 85%.	*feature* 3. Menggunakan TF-IDF sebagai *term weighting*
3	Analisis Sentimen Masyarakat Indonesia Terhadap Metaverse Menggunakan Algoritma *Support Vector Machine* (Sumayah dkk., 2023)	1. Bertujuan untuk menganalisis sentimen dari tanggapan masyarakat Indonesia terhadap Metaverse 2. Sumber data dari twitter diperoleh 2504 data,, dengan klasifikasi SVM dan pembobotan TF-IDF, hasil akurasi 81%, *precision* 79%, *recall* 63%, *F1-score* 57%.	1. Penelitian ini membahas topik terkait metaverse 2. Metode yang digunakan SVM dengn pembobotan TF-IDF 3. Pelabelan dengan *scrip python,* namun dengan kamus sentimen (membuat) sendiri
4	Analisis Sentimen Review Wisatawan Pada Objek Wisata Ubud Menggunakan Algoritma *Support Vector Machine* (Suryawan dkk., 2023)	1. Bertujuan untuk menganalisis sentimen review objek wisata di Ubud pada situs TripAdvisor pada 2. Data diperoleh sebanyak 669 data. 3. Hasil metode klasifikasi SVM akurasi 84,01%, *recall* 89,83%, *precision* 90,40% dan *F1-score* 90,11%.	1. Penelitian ini dilakukan pada objek wisata Ubud bersumber hanya pada TripAdvisor 2. Pengumpulan data menggunakan *tools ParseHub* namun mendapat hasil kurang baik 3. Pembobotan menggunakan TF-IDF
5	Analisis Sentimen Menggunakan *Naive Bayes* Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps (Khofifah dkk., 2022)	1. Bertujuan untuk menganalisis sentimen review destinasi pantai di Karawang. 2. Hasil klasifikasi dengan *Naive Bayes* pada 5 pantai. 3 diantara 5 pantai mendapatkan komen positif pantai Pakis 65%, pantai Sedari 70% dan pantai Samudera Baru 85%.	1. Metode pengumpulan data dengan studi dokumentasi mengamati review objek wisata, ulasan yang didapatkan kurang maksimal 2. Analisis sentimen dilakukan dengan *rapid miner* dengan metode *naive bayes* perlu adanya penerapan metode analisis lainnya

Pada beberapa penelitian terdahulu telah dilakukan analisis sentimen terhadap wisata dan penerapan metode *Support Vector Machine* (SVM). Namun, dari beberapa penelitian tersebut belum membahas tentang pariwisata di Kabupaten Gresik, serta belum membahas mengenai metode *dimensional reduction* terhadap klasifikasi teks dengan *Linear Discriminant Analysis*. Sehingga kondisi serupa dapat membuat perbedaan antara penelitian yang akan dilakukan ini dengan penelitian terdahulu.

2.2. Dasar Teori

2.2.1. *Google Maps*

Peta virtual yang disediakan Google dapat diakses melalui browser atau aplikasi smartphone untuk mengetahui letak suatu area atau lokasi tertentu disebut Google maps. Google maps juga dapat menentukan jarak dan rute yang sesuai untuk menuju lokasi yang telah dipilih. Pada setiap lokasi pula terdapat ulasan dari setiap pengunjung yang berkontribusi mengisi komentar, pesan maupun kesan mereka rasakan selama berada di lokasi (Utami & Erfina, 2022). Tempat yang menampung poin-poin tersebut terangkum pada fitur Google Review yang biasanya terdapat pada seluruh alamat yang terdaftar dalam Google Maps (Haq, 2020).

Google review akan muncul ketika seseorang mencari dan mengklik salah satu lokasi. Bentuk dari review tersebut akan berupa ulasan dalam bentuk bintang atau rating dan ulasan kata atau komentar. Berdasarkan review tersebut dapat diasumsikan bahwa penilaian orang terhadap lokasi tersebut dengan sudut pandang beragam yang mereka dapatkan saat berada di lokasi. Dalam era big data, google review menjadi salah satu aspek penting khususnya dalam bidang pariwisata karean terdapat informasi penilaian dari wisatawan yang telah mengunjungi suatu objek tersebut.

2.2.2. Analisis Sentimen

Metode digital dalam mengidentifikasi, mengekstraksi dan mengelola sebuah data berupa teks dalam mengekspresikan jenis sentimen (Steven & Wella, 2020). Pendeteksian, analisis, dan evaluasi terhadap keadaan pikiran individu dalam berbagai masalah, peristiwa, atau layanan adalah cakupan dari analisis

sentimen. Tujuan dilakukan analisis sentimen adalah untuk menemukan dan mengidentifikasi sentimen dari sebuah kalimat dengan mengklasifikasikannya berdasarkan nilai polaritas.

Dalam cabang ilmu analisis sentimen memiliki 2 sudut pandang dengan cakupan yang berbeda. *Opinion mining* berfokus pada penilaian pendapat atau opini dan ekspresi, *emotional mining* berfokus pada penilaian emosi individu berdasarkan artikulasi atau pengucapan saat berkomunikasi (Steven & Wella, 2020). Penerapan dalam sektor industri dan pariwisata sudah menjadi topik yang umum. Klasifikasi yang dihasilkan dari analisis sentimen meliputi sentimen positif, negatif dan netral.

2.2.3. Pariwisata

Pariwisata adalah segala aktivitas yang difasilitasi dan pelayanan yang tersedia oleh pemerintah, Pemerintah Daerah, masyarakat, dan pengusaha saat berwisata di tempat tertentu (Ridho & Marseto, 2023). Wisata adalah serangkaian kegiatan yang diawali dengan perjalanan oleh seseorang atau sekelompok orang dengan tujuan berekreasi ke suatu tempat dalam masa sementara, dan pengembangan diri atau mengenal keunikan tempat tersebut.

Dengan berwisata, seseorang dapat memperoleh informasi baru terhadap dunia luar yang teramat luas. Mempelajari adat dan budaya dari setiap perbedaan yang ada. Seseorang dapat menilai keunikan yang diberikan sesuai dengan persepsi mereka. Ulasan tersebut yang menjadi ketertarikan terhadap orang lain.

2.2.4. Kabupaten Gresik

Luas daerah di Kabupaten Gresik sekitar 1.194 km² dengan jumlah penduduk 1.311.215 jiwa dan kepadatan 1.098 jiwa/km2. Secara geografis, letak Kabupaten Gresik pada sebelah timur ada Kota Surabaya dan Selat Madura, dari sisi barat Kabupaten Lamongan, Laut Jawa di utara, dan di posisi selatan ada Kabupaten Sidoarjo dan Mojokerto. Dalam sektor industri, Kabupaten Gresik dikenal oleh masyarakat, unggul dalam sektor tersebut (Ridho & Marseto, 2023).

Potensi wisata di Kabupaten Gresik ikut berkembang. Kota industri dan kota wisata religi adalah sebagian sebutan oleh orang-orang yang Kabupaten Gresik. Berkisar antara 30 objek wisata lebih terdapat di Kabupaten Gresik dengan jenis

kategori yang berbeda diantaranya, wisata alam, wisata budaya dan wisata minat khusus.

2.2.5. Data Preprocessing

Proses mempersiapkan data dengan menghilangkan beberapa komponen yang tidak diperlukan dalam analisis agar tidak terdapat gangguan saat pengolahan data ke proses selanjutnya disebut Data preprocessing (Larasati dkk., 2022). Manfaatnya adalah untuk memperoleh kumpulan kata yang penting dan relevan dalam analisis sentimen. Proses ini dilakukan untuk mengubah data atau teks menjadi siap untuk diolah agar meningkatkan keakuratan dalam pengklasifikasian dalam proses analisis (Oktafani & Prasetyaningrum, 2022). Data *preprocessing* terbagi menjadi beberapa proses yaitu:

1. Data *Cleaning*

   Atribut pada kalimat seperti seperti simbol, spasi berlebih, tanda baca, emotikon dan angka pada tahap ini akan dihilangkan.

2. *Case Folding*

   Proses mengonversi sebuah kalimat atau kata dalam dokumen menjadi huruf kecil (lowercase) (Abelard & Sibaroni, 2021).

3. *Tokenizing*

   Tokenizing adalah memisahkan kata pada kalimat yang disebut token atau kata tunggal kemudian untuk dapat diproses ke tahap selanjutnya (Ramadhan & Sibaroni, 2021).

4. *Normalization*

   Proses untuk mengubah kata asal atau tidak baku diubah sesuai dengan kaidah bahasa berdasarkan kamus adalah *normalization* (Larasati dkk., 2022).

5. *Stopword*

   Stopword adalah kata yang tidak deskriptif dan sering ada dalam jumlah besar yang tidak ada hubungannya dengan informasi yang dibutuhkan (Wardhana & Sibaroni, 2021).

6. *Stemming*

   Proses berikutnya, mengubah setiap kata yang memiliki imbuhan menjadi bentuk dasar dari kata tersebut dengan menghilangkan imbuhan seperti awalan, akhiran, dan konfiks (Bourequat & Mourad, 2021).

2.2.6. *TextBlob*

Library dalam Python yang memiliki kemampuan mengolah informasi tekstual yang bertujuan untuk mencadangkan akses dalam pemrosesan teks umum adalah *TextBlob* (Kaur & Sharma, 2020). *TextBlob* menyediakan API dasar untuk melakukan tugas pemrosesan bahasa alami umum seperti penandaan bagian ucapan, ekstraksi frasa kata benda, analisis sentimen, klasifikasi, dan terjemahan (Abiola dkk., 2023). Polaritas merepresentasikan sentimen yang termasuk positif, negatif dan netral berdasarkan batas nilainya. Setiap nilai polaritas dari 0 sampai +1 termasuk positif, 0 berarti netral dan -1 sampai 0 termasuk negatif (Abiola dkk., 2023).

Perhitungan nilai polaritas pada kalimat dihitung berdasarkan jumlah total polaritas setiap kata, dan dibagi dengan jumlah kata pada kalimat. Apabila terdapat kata negasi, *TextBlob* akan mengalikan polaritas kata dengan -0.5 (Fahmi dkk., 2020). *TextBlob* akan mengabaikan istilah yang tidak dikenalnya, dan akan mempertimbangkan kata dan ekspresi yang dapat diterapkan ekstrim dan titik tengah untuk sampai pada skor akhir. Rumus perhitungan nilai polaritas *TextBlob* sebagai berikut:

$$Polaritas = \frac{w_1 + w_2 + .,,. w_n}{s_n} \quad (2.1)$$

Keterangan:

w: nilai polaritas kata

s : jumlah kata

Pada *TextBlob*, pelabelan teks bahasa Indonesia menghasilkan hasil yang kurang maksimal, dibutuhkan tambahan *corpus* yang berisi kata-kata bahasa Indonesia. Hal tersebut, dapat diperoleh dengan metode *lexicon-based*. Di dalam kamus lexicon setiap kata memiliki bobot polaritas dan digunakan untuk klasifikasi atau pelabelan sentimen (Yerzi & Sibaroni, 2021). Kamus akan digunakan sebagai

fitur tambahan pada *TextBlob*. Kamus lexicon memiliki dataset dengan konsep nilai polaritas dari -5 sampai +5.

2.2.7. Validasi Data

Dari pelabelan otomatis akan dilakukan validasi data secara manual oleh individu. Validasi disertakan pada penelitian oleh manusia sangat dianjurkan, mesin bisa mengalami kesusahan saat memahami sebuah konteks, yang akan sangat berbeda dengan manusia yang memiliki penilaian emosional dan komunikasi yang kompleks (Lappeman dkk., 2020). Beberapa tantangan yang biasanya terdapat pada sentimen analisis seperti ambiguitas, konteks kalimat, dan slang akan selalu ada dalam setiap kalimat. Akurasi yang diperoleh dari analisis sentimen akan menjadi tantangan besar karena pilihan mesin dan manusia yang tidak selalu sama (Lai & Tan, 2019). Tingkat analisis sentimen masih belum bisa mencapai sempurna karena pengukuran subjektivitas dapat lebih terukur oleh seorang individu. Dari penelitian tertentu, menyatakan bahwa 70% - 80% derajat nilai kecocokan antara manusia dengan mesin (Lai & Tan, 2019).

2.2.8. *Judgement Sampling*

Penelitian analisis sentimen dapat menerapkan teknik *non-probability* (Lappeman dkk., 2020). *Judgement sampling* adalah salah satu jenis teknik *non-probability sampling*. *Judgement sampling* didasarkan pada pemilihan yang sesuai dengan kebutuhan tanpa adanya pilihan random. *Judgment sampling* digunakan untuk sampel dalam dimensi kecil dan diambil dari populasi dengan pemahaman terhadapnya baik dan jelas dalam pemilihan metode sampel (Al-Khalidi, 2023). Keunggulan yang terdapat pada *judgement sampling* yaitu cepat, sesuai untuk penelitian eksplorasi, dan tanpa biaya.

2.2.9. *Feature Extraction*

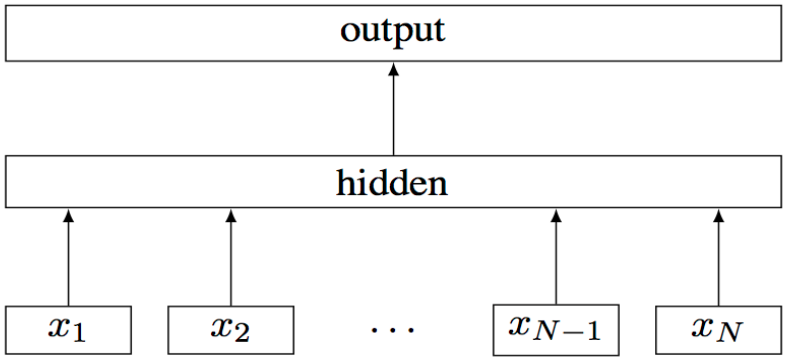
Feature Extraction sangat berpengaruh dalam kemampuan menghasilkan kinerja yang baik dalam machine learning (Cahyanti dkk., 2020). Salah satu model ekstraksi fitur yang dapat diterapkan adalah *FastText*. *FastText* adalah teknik vektorisasi teks yang dikenal sebagai *word embedding* yang dikembangkan oleh Facebook (ÇeliK & Koç, 2021). *Word embedding* dikenal sebagai representasi kata, berperan penting dalam menghasilkan vektor kata berkelanjutan berdasarkan arti

kata dalam dokumen. *Word embedding* menangkap informasi semantik dan sintaksis dari kata-kata yang digunakan untuk mengukur kesamaan kata, yang umumnya digunakan dalam tugas *Natural Language Processing* (NLP) (ÇeliK & Koç, 2021). *FastText* umumnya digunakan untuk menyelesaikan masalah klasifikasi kalimat dan representasi kata agar lebih efisien dan lebih cepat daripada metode *Word2Vec* dan *GloVe* (Agustiningsih dkk., 2022).

Word embedding FastText sebagian besar berfungsi untuk memecah kata menjadi sub-kata yang berbeda (postfix dan prefiks) (Mengistie & Kumar, 2021). Perbedaan antara *FastText* dan *Word2Vec* terdapat pada pendekatan representasi kata. *Word2Vec* menggunakan setiap kata sebagai satuan terkecil, sedangkan *FastText* menggunakan suku kata (N-Gram) sebagai satuan terkecil dengan panjang N terbentuk sesuai dengan panjang kata (Khomsah dkk., 2022). *FastText* menggunakan pendekatan berbasis Skip-gram di mana setiap kata direpresentasikan sebagai sekumpulan karakter N-gram. *FastText* menyesuaikan korpus teks yang disediakan dan membentuk model ruang vektor dimensi tinggi dengan tujuan agar vektor dari kata-kata yang mirip berdekatan. Fungsi *FastText* direpresentasikan dalam rumus berikut,

$$\sum_{t=1}^{T}\left[\sum_{c\in\mathcal{C}_t}\ell(s(w_t,\ w_c))+\sum_{n\in\mathcal{N}_{t,c}}\ell(-s(w_t,\ n))\right] \tag{2.2}$$

dengan *s* representasi skor, *w* merepresentasikan bobot, *l* merepresentasikan log(1 + e − x), dan n merepresentasikan jumlah kata dalam korpus (Agustiningsih dkk., 2022).



Gambar 2. 1 Model Arsitektur *FastText* (Hb dkk., 2018)

FastText melakukan pemetaan terhadap suku kata berdasarkan kosakata dan urutan karakter ($c_1 \ldots c_n$) ke dalam vektor *h*. Urutan karakter dari bahasa memungkingkan memberikan informasi dari makna sebuah kata. *FastText* mampu mengatasi permasalahan *out of vocabulary* dengan kemampuan dalam memberikan representasi kata yang tidak muncul di data latih (Nurdin dkk., 2020).

$$f_{subword}: \left(v(c_1.,,,.c_n)\right) \rightarrow h \quad (2.3)$$

2.2.10. *Linear Discriminant Analysis*

Pendekatan reduksi dimensi populer lainnya untuk langkah pra-pemrosesan dalam aplikasi data mining dan machine learning disebut *Linear Discriminant Analysis* (LDA) (Reddy dkk., 2020). Tujuan utama LDA adalah untuk memproyeksikan kumpulan data dengan jumlah fitur yang tinggi ke ruang yang berdimensi lebih kecil atau subruang i yang lebih kecil ($i \leq x - 1$) tanpa mengganggu informasi kelas dengan pemisahan kelas yang baik (Reddy dkk., 2020). Teknik ini umum diterapkan dalam kasus klasifikasi data dan reduksi dimensi. Penerapan LDA terhadap frekuensi kelas yang berbeda dan evaluasi terhadap data uij bisa sangat membantu.

Pendekatan untuk LDA adalah transformasi kelas-dependen dan kelas-independen terhadap rasio antara varians kelas ke dalam varians dan rasio keseluruhan varians ke dalam varians kelas. Misalkan $x_i \in \mathbb{R}^d$ yang berupa sampel *d*-dimensi dan $y_i \epsilon \{1.2.,,.c\}$ dikaitkan dengan target atau output, dengan *n* dilambangkan sebagai jumlah dokumen dan *c* sebagai jumlah kategori. Jumlah sampel pada setiap kelas dapat dihitung dengan rumus:

$$S_w = \sum_{l=1}^{c} S_l \quad (2.4)$$

2.2.11. *Synthetic Minority Over-sampling Technique*

Salah satu teknik sampling yang digunakan untuk masalah *imbalance* data dengan *upsampling*. SMOTE melakukan penyeimbangan data dengan menambahkan data baru pada kelas minoritas dari data buatan yang dihasilkan sehingga jumlah data pada kelas minoritas dan mayoritas menjadi seimbang (Kaope & Pristyanto, 2023). Penggunaan metode SMOTE terkadang dapat membuat sebuah model menjadi *overfitting* karena data duplikasi terhadap kelas minoritas menimbulkan data latih yang sama (Kasanah dkk., 2019). Data sintetik ditentukan

dengan diawali perhitungan jarak antar data pada data minoritas, kemudian penentuan nilai persentase dan k terdekat dengan menggunakan persamaan

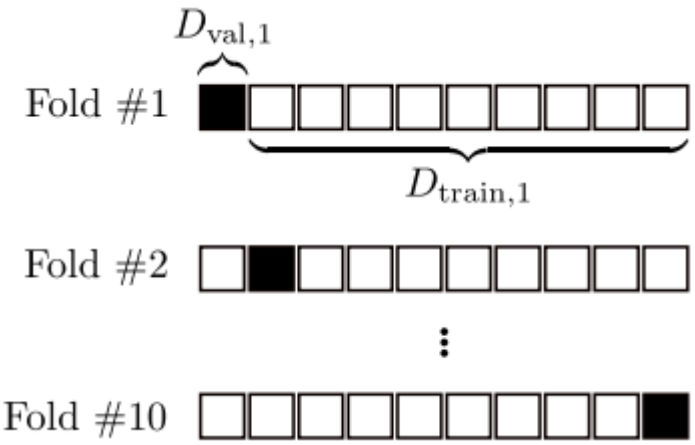
$$x_{syn} = x_i + (x_{knn} - x_i) \times \delta \quad (2.5)$$

Operator x_{syn} adalah sampel sintetik baru dari proses SMOTE, yang diperoleh dari x_i yang akan disintesis atau direplikasi dari sampel minoritas, x_{knn} adalah jumlah sampel tetangga yang akan digunakan untuk mensintesis sampel baru dari kelas minoritas, dan δ adalah nilai acak dari nol hingga satu (Kaope & Pristyanto, 2023).

2.2.12. *Cross Validation*

Pendekatan untuk mengeksplorasi sampel data pelatihan dan penilaian akurasi beberapa kali agar berpotensi meningkatkan hasil disebut *Cross validation*. Pada *cross validation* membentuk bagian dengan setiap sampel yang diulang beberapa kali dengan tujuan meningkatkan hasil dan analisis (A. Ramezan dkk., 2019). Pembagian data berlangsung secara bergantian dari himpunan data yang dibagi menjadi data uji dan data latih untuk proses selanjutnya. *K-fold cross validation* adalah salah satu jenis dalam *cross validation*..

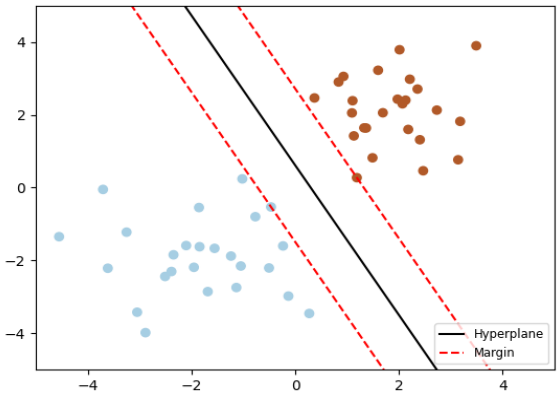
K-fold cross validation adalah metode yang digunakan dalam membagi data sesuai dengan nilai k atau *fold*. Metode ini berlangsung berulang-ulang dalam membagi data. Data terbagi menjadi data *training* dan data *testing.* Dalam pengujian evaluasi data dengan kondisi tertentu, *K-fold cross validation* dapat menjadi sebuah solusi dalam pembangunan model analisis sentimen dan menentukan *hyperparameter* (Ikegami & Darmawan, 2022). Model *k-fold cross validation* akan membagi data berdasarkan nilai yang ditentukan dan akan melakukan iterasi sesuai dengan nilai k yang telah diberikan (Arsi & Waluyo, 2021).



Gambar 2. 2 *10-Fold Cross Validation* (Berrar, 2019)

2.2.13. *Support Vector Machine*

Dalam *machine learning,* terdapat metode *supervised learning* yang digunakan dalam menganalisis data dan mengenali pola salah satunya *Support Vector Machine.* Metode ini merupakan metode yang sesuai dalam kasus klasifikasi. Konsep awal SVM berdasarkan dari permasalahan klasifikasi antara dua kelas positif dan negatif yang membutuhkan trainig set (Steven & Wella, 2020). Konsep klasifikasi dari SVM memaksimalkan batas-batas (margin) dengan *hyperplane* atau garis yang memisahkan dua kelas. Tingkat generalisasi yang tinggi dengan kemampuan dalam menemukan *hyperplane* yang optimal, berpengaruh pada tingkat akurasi yang diberikan.



Gambar 2. 3 Formasi *Hyperplane* SVM

Pada gambar 2.1, garis berwarna merah memisahkan dua kelas yaitu kelas +1 dan -1 disebut dengan hyperplane. Sedangkan, garis putus-putus biru menunjukkan margin, yang merupakan jarak terdekat ke hyperplane. Nilai hyperplane harus ditentukan terlebih dahulu untuk memaksimalkan nilai margin

(Abelard & Sibaroni, 2021). Dalam model SVM, sebagai data latih umumnya dari terbentuk dari fitur biasanya X_i dan nilai target atau label y_i (Pratama dkk., 2018). Dalam SVM, akan membentuk sebuah *classifier*, pembentukan tersebut digambarkan dengan rumus:

$$f(x_i) = \{\geq 0, y_i = +1, < 0, y_i = -1\} \quad (2.6)$$

Rumus dalam pembentukkan *hyperplane* sebagai berikut (Pratama dkk., 2018):

$$W.x + b = 0 \quad (2.7)$$

Keterangan:

W : nilai dalam vektor

b : nilai bias

X : data latih

Proses untuk memperoleh model yang optimal dengan menentukan parameter yang sesuai dalam meningkatkan kinerja model disebut *Hyperparameter Tuning* (Fitriansyah & Sibaroni, 2023). *Gridsearch* digunakan untuk menentukan *hyperparameter* yang paling optimal. *Hyperparameter* tersebut antara lain C, *gamma*, dan *kernel*. Pada umumnya, *support vector* dapat ditentukan berdasarkan fungsi *kernel* pada SVM (Suryawan dkk., 2023), yaitu:

1. *Linear* : $(x.x')$ (2.8)
2. *RBF* : $\exp(-\gamma||x - x'||^2)$ (2.9)
3. *Sigmoid*: $\tanh(\gamma(x.x^2) + r)$ (2.10)

Keterangan:

d: *degree* (derajat)

r: *coef*()

γ: parameter *gamma*

2.2.14. *Confusion Matrix*

Confusion Matrix digunakan pada saat pengukuran kinerja dan evaluasi dengan matriks prediksi dibandingkan dengan hasil prediksi sebenarnya (Ramadhan & Sibaroni, 2021). *Confusion Matrix* digambarkan dalam bentuk tabel yang menampilkan perbandingan dari jumlah data diklasifikasikan dengan benar dan diklasifikasikan dengan salah (Steven & Wella, 2020). Hasil perhitungan evaluasi akurasi, *Precision*, *Recall*, dan *F1-score* pada model analisis sentimen berawal dari

Confusion Matrix (Ikegami & Darmawan, 2022). Tabel 2.2 gambaran mengenai *Confusion Matrix* antara class aktual dan prediksi.

Tabel 2. 2 *Confusion Matrix* (Bourequat & Mourad, 2021)

Confusion Matrix		Kelas prediksi	
		Kelas benar	Kelas salah
Aktual	Kelas benar	*True Positive*	*False Negative*
	Kelas salah	*False Positive*	*True Negative*

2.2.15. Evaluasi

Pengukuran kinerja dari sebuah model yang dibangun perlu dilakukan. Perhitungan dilakukan berdasarkan yaitu *Accuracy, Precision, Recall, F-measure*, dan *Accuracy*. *Precision* merepresentasikan nilai keakuratan data terhadap hasil prediksi yang didapatkan dari model (Ramadhan & Sibaroni, 2021). *Recall* adalah rasio jumlah dokumen yang dipulihkan oleh pengklasifikasi terhadap jumlah total dokumen yang relevan (Wardhana & Sibaroni, 2021). *F-measure* atau *f1-score* mengukur rata-rata kombinasi harmonik antara *Precision* dan *Recall*. *Accuracy* adalah hasil penilain terkait klasifikasi yang tepat dilakukan oleh model yang telah dibangun. Rumus *Precision, Recall, f-measure,* dan *Accuracy* sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (2.11)$$

$$Recall = \frac{TP}{TP+FN} \quad (2.12)$$

$$F\text{-}1\ Score = 2\frac{Precision \times Recall}{Precision + Recall} \quad (2.13)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.14)$$

Keterangan:

TP atau *True Positives*: label positf pada data positif dan benar

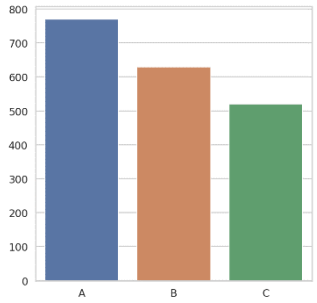
TN atau *True Negatives*: label negatif pada data negatif dan benar

FP atau *False Positives*: label positif pada data negatif dan salah

FN atau *False Negative*: label negatif pada data positif dan salah

2.2.16. Visualisasi Data

Teknik dalam mempresentasikan data dalam bentuk grafik atau gambar agar dapat memberikan pandangan tentang informasi yang lebih jelas dalam sebuah data adalah visualisasi data (Wahjoerini dkk., 2022). Visualisasi data bertujuan untuk membuat data lebih mudah dipahami, mengidentifikasi pola dan tren dalam data, serta membantu dalam pengambilan keputusan yang berbasis data (Guntara, 2023). *Bar chart* adalah grafik batang 2 dimensi dari 2 kelas atau lebih sebuah data. Pembentukkan grafik batang berdasarkan ukuran pada luas batang (panjang x lebar) (Muharni & Candra, 2022). Tinggi batang bervariasi berdasarkan total data setiap kategori dengan lebar batang yang sama..



Gambar 2. 4 *Bar Chart*

Jenis visualisasi column chart atau bar chart ini baik digunakan jika tujuan utama kita adalah melihat komposisi antar dimensi. *Wordcloud* adalah visualisasi dari daftar sebuah data dari bahasa atau teks yang memiliki bobot tertentu dengan digambarkan dari perhatian meningkat (Fahmi dkk., 2020). Penggambaran visual *Wordcloud* berdasarkan frekuensi kemunculan kata-kata tersebut dari suatu kumpulan teks pada dokumen. Ukuran huruf dari visualisasi mengimplementasikan frekuensi kemunculan kata tersebut. Ukuran kata yang yang besar merepresentasikan seberapa sering kata tersebut muncul, dan ukuran kata yang kecil sesuai dengan seberapa sedikit kata tersebut.

Gambar 2. 5 *Wordcloud* (Hendrawan dkk., 2022)

2.3. Integrasi Keilmuan

Penelitian yang akan dilakukan tentunya memiliki manfaat dari sudut pandang Islam diketahui dengan mengkonsultasikan dengan ahli di bidang tafsir Al-Qur'an terkait integrasi keilmuan yanng terkandung dalam bidang penelitian tersebut. Komunikasi berlangsung dan dilakukan dengan Ustadz H. Musyafa', Lc., M.Pd.I, M.Pd., di Masjid Al Wachid, Jl. Menanggal V No.70, Kec. Gayungan, Surabaya. Beliau merupakan tokoh masyarakat sekaligus Imam sholat Masjid Al Wachid. Beliau juga menjadi pengurus di Yayasan Lembaga Pendidikan Islam Al Hikmah Surabaya dan sebagai tenaga pengajar guru Sejarah Islam di SMA Al Hikmah Surabaya. Penelitian ini membahas mengenai analisis sentimen berdasarkan review pengunjung wisata. Menurut Ustadz H. Musyafa' ada beberapa ayat surat dan hadits yang berkaitan dengan penelitian ini, salah satunya QS. Al-'Asr [103]:3.

اِلَّا الَّذِيْنَ اٰمَنُوْا وَعَمِلُوا الصّٰلِحٰتِ وَتَوَاصَوْا بِالْحَقِّ ۙ وَتَوَاصَوْا بِالصَّبْرِ

"Kecuali orang-orang yang beriman dan beramal saleh serta saling menasihati untuk kebenaran dan kesabaran." (QS. Al-'Asr 103:3)

Berdasarkan ayat tersebut, seorang mukmin dianjurkan saling menasehati dalam kebenaran. Penelitian ini berhubungan dengan nasehat atau pendapat pengunjung terhadap objek wisata. Suatu pendapat pada dasarnya untuk memberikan opini terkait kondisi pada wisata tersebut dalam hal kebaikan. Firman Allah lainnya yang berkaitan dengan penelitian ini adalah QS. Ali 'Imron [3]:159.

فَبِمَا رَحْمَةٍ مِّنَ اللّٰهِ لِنْتَ لَهُمْۚ وَلَوْ كُنْتَ فَظًّا غَلِيْظَ الْقَلْبِ لَانْفَضُّوْا مِنْ حَوْلِكَۖ فَاعْفُ عَنْهُمْ وَاسْتَغْفِرْ لَهُمْ وَشَاوِرْهُمْ فِى الْاَمْرِۚ فَاِذَا عَزَمْتَ فَتَوَكَّلْ عَلَى اللّٰهِۗ اِنَّ اللّٰهَ يُحِبُّ الْمُتَوَكِّلِيْنَ

“Maka, berkat rahmat Allah engkau (Nabi Muhammad) berlaku lemah lembut terhadap mereka. Seandainya engkau bersikap keras dan berhati kasar, tentulah mereka akan menjauh dari sekitarmu. Oleh karena itu, maafkanlah mereka, mohonkanlah ampunan untuk mereka, dan bermusyawarahlah dengan mereka dalam segala urusan (penting). Kemudian, apabila engkau telah membulatkan tekad, bertawakallah kepada Allah. Sesungguhnya Allah mencintai orang-orang yang bertawakal.” (QS. Ali ‘Imron 3:159)

Ayat ini menjelaskan, dalam memutuskan segala sesuatu diwajibkan bagi para pemimpin agar bermusyawarah demi kemaslahatan bersama. Pendapat atau ulasan yang diberikan akan diolah terlebih dahulu sebelum diputuskan. Pengelola wisata akan mempertimbangkan dengan koleganya dengan musyawarah memutuskan keputusan terbaik.

Berdasarkan Kitab Hadits Al-Arbain An-Nawawiyah mengenai kewajiban mengingkari kemungkaran, Rasulullah *shallallahu ‘alaihi wa sallam* bersabda:

عَنْ أَبِي سَعِيْدٍ الخُدْرِيِّ رَضِيَ اللهُ عَنْهُ، قَالَ: سَمِعْتُ رَسُولَ اللهِ ﷺ يَقُوْلُ: « مَنْ رَأَى مِنْكُمْ مُنْكَراً فَلْيُغَيِّرْهُ بِيَدِهِ، فَإِنْ لَمْ يَستَطِعْ فَبِلِسَانِهِ، فَإِنْ لَمْ يَستَطِعْ فَبِقَلْبِهِ وَذَلِكَ أَضْعَفُ الإِيْمَانِ » رَوَاهُ مُسْلِمٌ

Dari Abu Sa’id Al-Khudri *radhiyallahu ‘anhu*, ia berkata, “Aku mendengar Rasulullah *shallallahu ‘alaihi wa sallam* bersabda, ‘*Barangsiapa dari kalian melihat kemungkaran, ubahlah dengan tangannya (kekuasaannya). Jika tidak bisa, ubahlah dengan lisannya. Jika tidak bisa, ingkarilah dengan hatinya, dan itu merupakan selemah-lemahnya iman.*” (HR. Muslim no. 49)

Hadits tersebut menjelaskan, setiap menemukan kemungkaran atau kekurangan sebaiknya diubah dengan kadar kekuasaan. Hal ini mengisyaratkan sebagai seorang pemimpin atau pengelola lebih baik melakukan perubahan berdasarkan arahannya tetapi tetap memperhatikan pertimbangan yang diberikan oleh timnya dan pendapat dari orang lain.

BAB III
METODE PENELITIAN

3.1. Jenis Penelitian

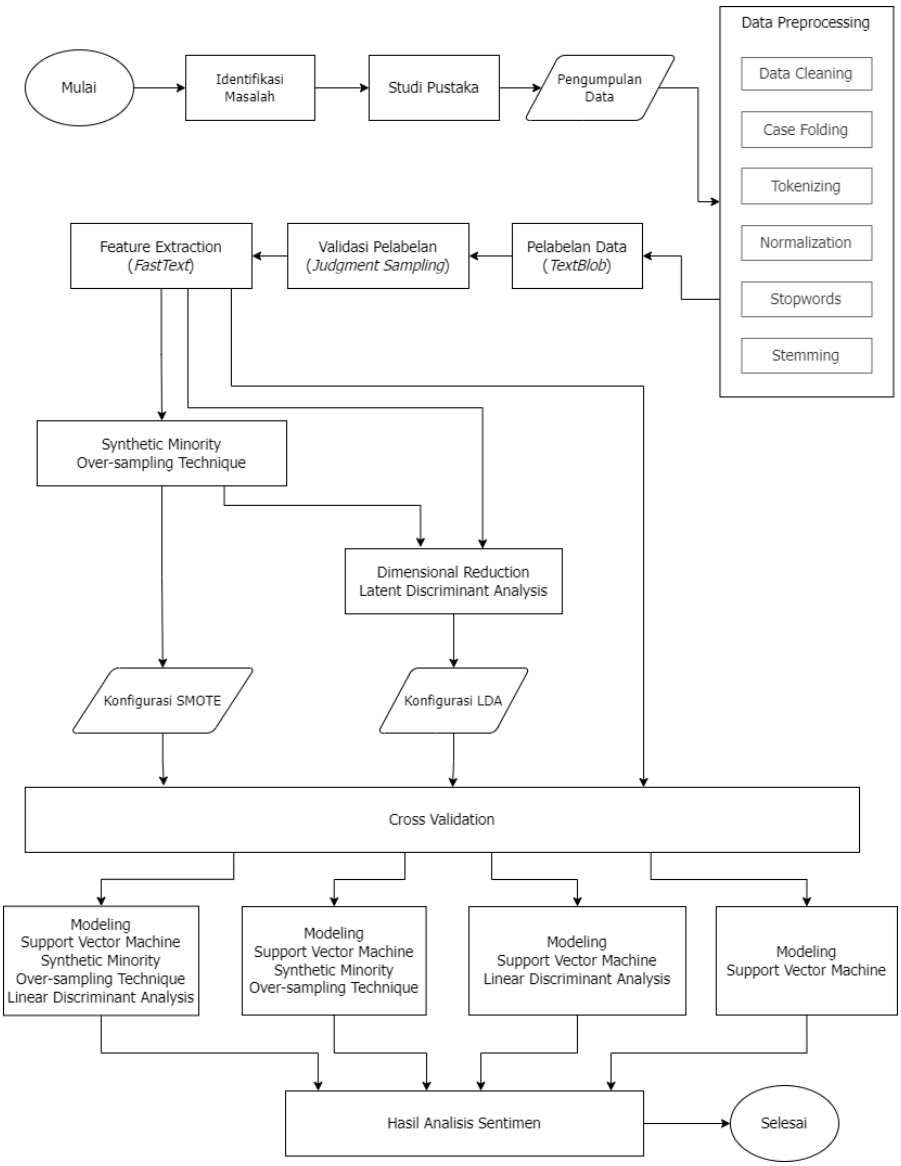
Metode penelitian yang diterapkan yaitu kuantitatif deskriptif. Dalam penerapannya, metode kuantitatif menggunakan angka atau jumlah data yang dikumpulkan dengan tujuan untuk menampilkan hasil data (Nurfajiah dkk., 2021). Tujuan penerapan metode tersebut adalah untuk menjelaskan dari sudut pandang pengukuran yang diperoleh secara detil dari yang diteliti.

3.2. Sumber Data

Pada penilitian yang akan dilakukan, menggunakan sumber data yang diperoleh dari *Google Maps*. Salah satu fiturnya *Google Review*. Data diambil menggunakan program *coding python* dengan melakukan Web *Scraping*. Penelitian ini menggunakan data sekunder sebagai sumber data. Data sekunder adalah sekumpulan data yang didapatkan dari bentuk informasi bermacam sumber, sehingga wawancara tidak perlu dilakukan oleh peneliti dalam mendapatkan data (Nurfajiah dkk., 2021). Penelitian ini mengambil data Google Review lokasi wisata dengan rentang waktu dari tahun 2020 sampai tahun 2023. Tahun 2020 dipilih karena awal terjadinya pandemi covid-19 dan sebagai pembanding antara tahun 2022 - 2023 yaitu pasca covid-19. Pada Desember 2022 di Indonesia, Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) telah diumumkan bahwasanya telah resmi dicabut oleh pemerintah, oleh sebab itu tahun 2022 dan 2023 dipilih sebagai data.

3.3. Metode Penelitian

Alur penelitian terangkum dalam Gambar 3.1 dengan beberapa Iangkah-Iangkah yang telah digambarkan dalam diagram.



Gambar 3. 1 Diagram Alir Penelitian

3.3.1. Identifikasi Masalah

Sebagai awal memulai penelitian, dilakukan identifikasi masalah berdasarkan objek penelitian. Berdasarkan Data Kunjungan Wisata Online (DAKUWISON) tahun 2023 terjadi penurunan wisatawan pada objek wisata di Kabupaten Gresik. Faktor lain bagi pengunjung adalah ulasan pengunjung lain yang telah mendatangi objek wisata dan menjadi referensi terhadap pengunjung baru. Masalah tersebut terkait dengan perlunya melakukan Analisis Sentimen untuk mengetahui perspektif pengunjung terhadap objek wisata di Kabupaten Gresik. Dengan harapan hasil tersebut dapat menjadi bahan pertimbangan dalam rangka memutuskan solusi yang perlu dilakukan oleh pengelola wisata.

3.3.2. Studi Pustaka

Tahap selanjutnya dilakukan studi pustaka untuk mencocokkan masalah dengan referensi penelitian terdahulu, menentukan dasar teori dari penelitian yang relevan, dan menerapkan beragam kaidah dalam menguraikan masalah dengan kasus serupa.

3.3.3. Pengumpulan Data

Setelah informasi diperoleh, tahap selanjutnya yang dilakukan pengumpulan data. Ulasan dari objek wisata di Kabupaten Gresik akan menjadi tempat pengambilan data. Teknik yang digunakan dalam pengambilan data yaitu web *scraping* menggunakan *library* Selenium dalam Python. Pada Tabel 3.1 dilampirkan sebagai contoh data yang terkumpulkan.

Tabel 3. 1 Hasil Pengumpulan Data

Data
Bagus tempatnya. Murah tiket dewasa 10k, anak2 5k.tempat sholat n toilet jg bersih. Kalau ksni enak sblum dhuhur jd bs fto2 asik ndak kepanasan 😀
Bagus, banyak spot foto yang instagramable, htm 15 rb dapat voucher untuk beli oleh2 senilai 5 rb, parkir mobil 10 rb
Ada hall yang ckup menampung 100 orang lebih, gazebo yang banyak, spot foto yg ok, namun masih panas kalau siang
Bagus.... Cmak kurang pepohann... Soalnya kalo siang puannnassss ngak ada tempat teduh.....

3.3.4. Data *Preprocessing*

1. Data *Cleaning*

Pembersihan data dengan menghilangkan atribut pada kalimat seperti seperti simbol, spasi berlebih, tanda baca, emotikon dan angka pada tahap ini.

2. *Case Folding*

Perubahan dari seluruh kalimat menjadi huruf kecil dengan tujuan agar proses penanganan terhadap teks menjadi sama

3. *Tokenizing*

Pemisahan teks dari kalimat menjadi per kata untuk mempermudah proses selanjutnya.

4. *Normalization*

Normalization digunakan untuk menormalkan kata. Pada tahap ini akan setiap kata yang awalnya asal atau tidak baku akan diubah sesuai dengan kaidah bahasa berdasarkan kamus *slangwords* yang diperoleh dari *kaggle* (Diandra, 2022).

5. *Stopwords*

Beberapa kata yang tidak diperlukan yang bisa mempengaruhi proses analisis akan dihilangkan. Pada tahap ini menerapkan *library* dari *python* sebagai penunjang proses yaitu NLTK. Proses ini akan mempertahankan kata yang penting dan bermakna. Adapun beberapa tambahan kata yang tidak relevan untuk analisis diantaranya: "yg","dg","rt","h","t","kah","j","nich","los","loss","nihh","annya","q","x","da","sak","ygy","tt","donk","donkk","doong","bossq","kar","nyaa","nya","ny", "ohh", "gess", "ges", "db", "si", "sii", "sih", "sihh", "x", "rekk", "rek", "reekk", "cus", "lo", "lho", "loh", "lohh", "gais", "lah", "lahh", "ny", "nya", "nyaa", "tuh", "tuhh", "weeh", "kd", "dech", "dehh", "deh", "l", "uy", "ngos", "oey", "oi", "yng", "hm", "los", "losss", "tul", "tuyuull", "yuul", "th", "b", "lt", "eh", "ehh", "sea", "gaess", "don", "dong", "sutle", "ta", "tak", "takk", "lur", "lurr", "hua", "ss", "lip", "mencle", "hem", "hiks", "wuuah", "ith", "v", "o", "masszeh", "ied", "puan", "deni", "cak", "cok", "tok", "yopi", "mashok", "oz", "brow", "eks", "woyo", "coy", "rw", "ft", "gwk", "sich", "jeddih", "kars", "ig", "heemm", "yach", "laah", "tep", "tk", "min", "ugk", "the", "over", "skuy", "uhm","gini".

6. *Stemming*

Proses berikutnya, mengubah setiap kata yang memiliki imbuhan menjadi bentuk dasar dari kata tersebut dengan menghilangkan imbuhan agar dapat lebih mudah untuk proses berikutnya. *Library* sastrawi digunakan dalam tahap ini.

3.3.5. Pelabelan Data

Setelah melalui *preprocessing*, akan terbentuk sebuah dataset. Data selanjutnya akan dilabelkan berdasarkan nilai polaritas. *TextBlob* berperan dalam menilai polaritas pada setiap kata yang selanjutnya akan dihitung berdasarkan rata-rata dari total nilai polaritas setiap kata pada kalimat dengan banyak kata. Kamus leksikon bahasa indonesia ditambahkan untuk menambah korpus dalam penilaian polaritas (Anasta, 2023). Perhitungan rata-rata polaritas yang didapatkan dari satu kalimat ulasan, dengan menerapkan rumus (2.1). Nilai rata-rata dari polaritas

kalimat tersebut akan digunakan sebagai tanda bahwa kalimat tersebut tergolong dalam label positif, negatif, dan netral.

Tabel 3. 2 Nilai Polaritas dan Label

Nilai Polaritas	Label
Score > 0	Positif
Score = 0	Netral
Score < 0	Negatif

3.3.6. Validasi Pelabelan

Setelah melakukan pelabelan data secara otomatis, selanjutnya dilakukan validasi manual dengan bantuan dari 3 pengelola atau karyawan wisata sebagai penentuan sentimen pada ulasan. Data yang akan dinilai merupakan data sampel yang dipilih berdasarkan teknik *non-probability* yaitu *judgement sampling*. Pengelola atau karyawan akan memberikan jawaban dalam menentukan sentimen yang terkandung dalam ulasan Berdasarkan hasil yang diperoleh, akan hitung akurasi *TextBlob* terhadap validasi pengelola dengan rumus (2.14).

3.3.7. Visualisasi Hasil Pelabelan

Dalam memvisualisasikan label yang telah ditentukan sebelumnya, menggunakan *bar chart* dan *wordcloud*. Perbandingan total banyak label positif, negatif dan netral akan terlihat dalam bentuk visualisasi *Bar chart*. Kata-kata yang sering digunakan dalam isi ulasan akan visualisasikan berdasarkan dari label positif dan negatif dengan *wordcloud*.

3.3.8. *Feature Extraction*

Metode *word embedding* diimplementasikan dalam vektorisasi teks dengan bantuan *FastText*. Setelah data dipreproses dan dilabelkan, lanjut ke proses vektorisasi. *FastText* diimplementasikan berdasarkan rumus (2.2) dengan memvektor setiap suku kata dari satu kata yang terkandung dalam sebuah kalimat. Representasi nilai vektor tersebut akan dapat dibaca oleh mesin dan dapat digunakan untuk proses pembelajaran sentimen analisis dengan SVM.

3.3.9. *Dimensional Reduction*

Pada tahap ini, data yang berbentuk vektorisasi angka akan dikurangi dimensinya. Pengurangan dimensi menerapkan rumus (2.4). Pada proses ini, hasil vektorisasi dan label data akan diterapkan dengan parameter *n_component* = 2. Pemilihan nilai *n_component* didasari dari *n-kelas* – 1 sesuai pada dokumentasi *sklearn*. Apabila komponen yang diberikan melebihi *n-kelas*, maka akan terjadi kesalahan dalam perhitungan matriks kovarians antar kelas. Output yang dihasilkan setelah proses yang dilalui akan membuat dimensi atau ukuran vektor menjadi semakin menyusut. Proses ini tidak akan menguraikan informasi yang terdapat sebelumnya.

3.3.10. *Cross Validation*

Pada tahap selanjutnya, akan dilakukan *k-fold cross validation* menerapkan nilai k=10. Penerapan proses ini diasumsikan untuk mengatasi apabila saat pelatihan data mengalami *overfitting*. Proses *hyperparameter tuning* juga dilakukan untuk menentukan *hyperparameter* yang optimal dengan *GridsearchCV*. Parameter SVM yang akan digunakan diantaranya,

Parameter	*Input*
Kernel	*RBF, Sigmoid*
C	1, 10, 100
Gamma	10, 1, 0.1

Hasil akhir akan berupa skor akurasi berdasarkan rumus (2.13) dan parameter yang optimal

3.3.11. *Modeling*

Tahapan ini mulai dilakukan pembangunan model atau *modelling*. Pembangunan model berdasarkan *hyperparameter tuning* yang dilakukan sebelumnya. Berikut penjelasan dari setiap tahapan pembangunan model:

1. Model training dilatih dengan SVM dan parameter yang telah diuji sebelumnya dengan membagi data latih dan data test 80:20. Terdapat 2 skenario dalam pelatihan data

- Skenario 1: data setelah dibagi, akan langsung ditraining dengan model dan parameter yang telah ditentukan. Model yang akan dilatih adalah model SVM tanpa LDA, dan model SVM dengan LDA
- Skenario 2: penerapan *upsampling* dengan metode SMOTE dilakukan pada label minoritas hingga setara label mayoritas. Data akan lebih banyak 2 kali lipat dari data awal sebelumnya. Selanjutnya akan ditraining dengan model dan parameter yang telah ditentukan. Pada skenario ini model yang dibuat adalah Model SVM dengan SMOTE dan Model SVM dengan SMOTE + LDA.

2. Evaluasi model dilakukan setelah model dilatih dengan mengaitkan sudut pandang, yaitu cenderung ke label positif, dan negatif agar diperoleh nilai ukur kemampuan pengklasifikasian berdasarkan label yang telah ditentukan. Metrik yang akan dinilai sebagai evaluasi yaitu *Precision* dengan rumus (2.11), *Recall* dengan rumus (2.12), *F1-score* (2.13), dan *Accuracy* (2.14). Hasil evaluasi dari SVM dan SVM dengan LDA akan dibandingkan.

3.3.12. Analisis Hasil Sentimen

Hasil pelatihan model yang sebelumnya dilakukan, akan diterapkan pada data prediksi. Model tersebut akan melakukan prediksi terhadap data dan secara otomatis menentukan sentimen dari ulasan. Dari hasil prediksi pada semua model akan dibandingkan dan dipilih yang cocok. Pembahasan selanjutnya dari analisis hasil sentimen pada model yang telah dipilih dengan menampilkan jumlah sentimen tiap label.

BAB IV
HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Proses *scraping* yang telah dilakukan dengan menggunakan *library selenium* dalam *python*. Data yang terkumpul berkisar diantara tahun 2020 – 2023 di bulan Mei. Total data yang sebanyak 3460, dengan rincian sebagai berikut:

Tabel 4. 1 Jumlah Ulasan Terkumpul

Tahun	Jumlah Ulasan
2020	338
2021	736
2022	2046
2023	340
Total	3460

Perolehan data pada tahun 2022 mengalami kenaikan yang cukup drastis. Pada tahun tersebut kasus Covid-19 mulai mereda, hal ini berpengaruh pada aktivitas pengunjung. Data yang telah diperoleh seperti pada tabel berikut, dengan berbagai komponen atau karakter. Hasil dari pengumpulan, dijadikan satu file untuk proses berikutnya.

Tabel 4. 2 Hasil Pengumpulan Data

Data
Dihari biasa kurang Ramai, masakannya enak tapi kalau bisa porsinya ditambahi y.. 😁😁. Coba kalau ada kolam renangnya, mungkin akan menambah antusias pengunjung.. 👍👍👍
Mirip selecta Batu. Hanya saja, kesadaran para pengunjung untuk menjaga kebersihan di area rest area sangat memprihatinkan. Tempat sampah yang disediakan kurang, dan kondisi toilet umum yang menjijikkan.
Masih dalam tahap pembangunan
Wisata yang sangat indah, dengan fasilitas yang terus dikembangkan, semoga bisa menjadi media untuk memperkerjakan masyarakat nya ✨🏠…

4.2. Data *Preprocessing*

Proses berikutnya dilakukan eksplorasi terhadap data dengan mengecek kolom kosong dan duplikat karakter. Data pada salah satu kolom yang tidak sesuai akan dihapus. Persebaran data setelah eksplorasi sebagai berikut:

Tabel 4. 3 Jumlah Ulasan Setelah Eksplorasi

Tahun	Jumlah Ulasan
2020	262
2021	648
2022	1872
2023	330
Total	3112

Berikutnya terhadap data dilakukan data *preprocessing* dengan memanfaatkan *Google Colab* sebagai sarana pemrosesan yang dibekali RAM 12.7 GB dan *disk* 107.7 GB. Beberapa tahapan dalam data *preprocessing* pada penelitian ini:

4.2.1. Data *Cleaning*

Proses awal dilakukan *cleaning* dengan modul *RegEx* pada python. Salah satu set fungsi pada modul *RegEx* memiliki cara kerja mengganti (dalam penelitian ini dihilangkan) atribut pada kalimat seperti seperti simbol, spasi berlebih, tanda baca, emotikon dan angka.

Tabel 4. 4 Proses Data *Cleaning*

Sebelum	Sesudah
Dihari biasa kurang Ramai, masakannya enak tapi kalau bisa porsinya ditambahi y.. 😁😁. Coba kalau ada kolam renangnya, mungkin akan menambah antusias pengunjung.. 👆👆👆	Dihari biasa kurang Ramai masakannya enak tapi kalau bisa porsinya ditambahi y Coba kalau ada kolam renangnya mungkin akan menambah antusias pengunjung
Mirip selecta Batu. Hanya saja, kesadaran para pengunjung untuk menjaga kebersihan di area rest area sangat memprihatinkan. Tempat sampah yang disediakan kurang, dan kondisi toilet umum yang menjijikkan.	Mirip selecta Batu Hanya saja kesadaran para pengunjung untuk menjaga kebersihan di area rest area sangat memprihatinkan Tempat sampah yang disediakan kurang dan kondisi toilet umum yang menjijikkan
Masih dalam tahap pembangunan	Masih dalam tahap pembangunan
Bagusss sebenarnya ..... Cmak parkirnya jauhhhhhh..... N tiap permainan bayrrr	Bagus sebenarnya Cmak parkirnya jauh N tiap permainan bayr

Dalam proses ini, akan mensejajarkan teks atau mengganti *newline* paragraf dengan spasi, menghapus spasi pada awal dan akhir teks dan perulangan karakter, seperti “baguusss” menjadi “bagus”.

4.2.2. *Case Folding*

Setelah data dibersihkan, data atau teks diubah ke *lowercase* atau huruf kecil. Proses tersebut dilakukan agar data diperlakukan secara sama dan mempermudah proses berikutnya.

Tabel 4. 5 Proses *Case Folding*

Sebelum	Sesudah
Dihari biasa kurang Ramai masakannya enak tapi kalau bisa porsinya ditambahi y Coba kalau ada kolam renangnya mungkin akan menambah antusias pengunjung	dihari biasa kurang ramai masakannya enak tapi kalau bisa porsinya ditambahi y coba kalau ada kolam renangnya mungkin akan menambah antusias pengunjung
Mirip selecta Batu Hanya saja kesadaran para pengunjung untuk menjaga kebersihan di area rest area sangat memprihatinkan Tempat sampah yang disediakan kurang dan kondisi toilet umum yang menjijikkan	mirip selecta batu hanya saja kesadaran para pengunjung untuk menjaga kebersihan di area rest area sangat memprihatinkan tempat sampah yang disediakan kurang dan kondisi toilet umum yang menjijikkan
Masih dalam tahap pembangunan	masih dalam tahap pembangunan
Bagus sebenarnya Cmak parkirnya jauh N tiap permainan bayr	bagus sebenarnya cmak parkirnya jauh n tiap permainan bayr

Pada ulasan yang diawali dengan huruf kapital atau huruf besar yang terdapat pada kalimat, akan diubah menjadi huruf kecil seluuruhnya.

4.2.3. *Tokenizing*

Teks selanjutnya akan dipisah dengan *library* NLTK. Pemisahan kata berdasarkan *whitespace* atau spasi per kata. Data teks akan berubah menjadi *array list* dari setiap *row* setelah proses *tokenizing* berlangsung.

Tabel 4. 6 Proses *Tokenizing*

Sebelum	Sesudah
dihari biasa kurang ramai masakannya enak tapi kalau bisa porsinya ditambahi y coba kalau ada kolam renangnya mungkin akan menambah antusias pengunjung	['dihari', 'biasa', 'kurang', 'ramai', 'masakannya', 'enak', 'tapi', 'kalau', 'bisa', 'porsinya', 'ditambahi', 'y', 'coba', 'kalau', 'ada', 'kolam', 'renangnya', 'mungkin', 'akan', 'menambah', 'antusias', 'pengunjung']
mirip selecta batu hanya saja kesadaran para pengunjung untuk menjaga kebersihan di area rest area sangat memprihatinkan tempat sampah yang disediakan kurang dan kondisi toilet umum yang menjijikkan	['mirip', 'selecta', 'batu', 'hanya', 'saja', 'kesadaran', 'para', 'pengunjung', 'untuk', 'menjaga', 'kebersihan', 'di', 'area', 'rest', 'area', 'sangat', 'memprihatinkan', 'tempat', 'sampah', 'yang', 'disediakan', 'kurang', 'dan', 'kondisi', 'toilet', 'umum', 'yang', 'menjijikkan']
masih dalam tahap pembangunan	['masih', 'dalam', 'tahap', 'pembangunan']

Sebelum	Sesudah
bagus sebenarnya cmak parkirnya jauh n tiap permainan bayr	['bagus', 'sebenarnya', 'cmak', 'parkirnya', 'jauh', 'n', 'tiap', 'permainan', 'bayr']

Pemisahan kata dari spasi tiap kalimat, apabila terdapat spasi berlebih, akan dinilai sebagai pemisah dan diubah menjadi token oleh NLTK, itulah mengapa pada saat *cleaning* data dilakukan dahulu sebelum *tokenizing.*

4.2.4. *Normalization*

Pada tahap ini, setiap kata yang tidak baku akan diubah sesuai dengan kaidah bahasa berdasarkan kamus *slangwords.* Proses ini menormalkan kata yang berguna dalam pelabelan data.

Tabel 4. 7 Proses *Normalization*

Sebelum	Sesudah
['dihari', 'biasa', 'kurang', 'ramai', 'masakannya', 'enak', 'tapi', 'kalau', 'bisa', 'porsinya', 'ditambahi', 'y', 'coba', 'kalau', 'ada', 'kolam', 'renangnya', 'mungkin', 'akan', 'menambah', 'antusias', 'pengunjung']	['hari', 'biasa', 'kurang', 'ramai', 'masakannya', 'enak', 'tapi', 'kalau', 'bisa', 'porsinya', 'ditambahi', 'ya', 'coba', 'kalau', 'ada', 'kolam', 'renangnya', 'mungkin', 'akan', 'menambah', 'antusias', 'pengunjung']
['mirip', 'selecta', 'batu', 'hanya', 'saja', 'kesadaran', 'para', 'pengunjung', 'untuk', 'menjaga', 'kebersihan', 'di', 'area', 'rest', 'area', 'sangat', 'memprihatinkan', 'tempat', 'sampah', 'yang', 'disediakan', 'kurang', 'dan', 'kondisi', 'toilet', 'umum', 'yang', 'menjijikkan']	['mirip', 'selecta', 'batu', 'hanya', 'saja', 'kesadaran', 'para', 'pengunjung', 'untuk', 'menjaga', 'kebersihan', 'di', 'area', 'istirahat', 'area', 'sangat', 'memprihatinkan', 'tempat', 'sampah', 'yang', 'disediakan', 'kurang', 'dan', 'kondisi', 'toilet', 'umum', 'yang', 'menjijikkan']
['masih', 'dalam', 'tahap', 'pembangunan']	['masih', 'dalam', 'tahap', 'pembangunan']
[['bagus', 'sebenarnya', 'cmak', 'parkirnya', 'jauh', 'n', 'tiap', 'permainan', 'bayr']	[['bagus', 'sebenarnya', 'cuma', 'parkirnya', 'jauh', 'dan', 'tiap', 'permainan', 'bayar']

Perubahan kata merujuk pada kamus yang diberikan, apabila di dalam kamus tidak terdapat kata non baku pada teks, perubahan pada tidak akan terjadi. Oleh karenanya penambahan kata baku pada kamus disesuaikan dengan data yang akan diproses.

4.2.5. *Stopwords*

Beberapa kata yang tidak diperlukan akan dihilangkan. Penerapan *library* NLTK digunakan dalam proses ini. Contoh *stopwords* yang terdapat pada NLTK seperti, ‘dan’, ‘yang’, ‘ini’, ‘itu’, ‘adalah’ dan sebagainya. Namun tidak semua kata yang tidak memiliki sentiment terkandung dalam *stopwords*, kata-kata ungkapan

tambahan seperti “masszeh”, "ied", "cak", "cok", "tok", "yopi", "mashok", "oz", "brow", "eks", "woyo", "coy”, "heemm", "yach", "laah", "tep", "tk", dan lain-lain, ditambahkan pada *stopwords.*

Tabel 4. 8 Proses *Stopwords*

Sebelum	Sesudah
['hari', 'biasa', 'kurang', 'ramai', 'masakannya', 'enak', 'tapi', 'kalau', 'bisa', 'porsinya', 'ditambahi', 'ya', 'coba', 'kalau', 'ada', 'kolam', 'renangnya', 'mungkin', 'akan', 'menambah', 'antusias', 'pengunjung']	['kurang', 'ramai', 'masakannya', 'enak', 'porsinya', 'ditambahi', 'coba', 'kolam', 'renangnya', 'menambah', 'antusias']
['mirip', 'selecta', 'batu', 'hanya', 'saja', 'kesadaran', 'para', 'pengunjung', 'untuk', 'menjaga', 'kebersihan', 'di', 'area', 'istirahat', 'area', 'sangat', 'memprihatinkan', 'tempat', 'sampah', 'yang', 'disediakan', 'kurang', 'dan', 'kondisi', 'toilet', 'umum', 'yang', 'menjijikkan']	['selecta', 'kesadaran', 'menjaga', 'kebersihan', 'memprihatinkan', 'sampah', 'disediakan', 'kurang', 'kondisi', 'toilet', 'menjijikkan']
['masih', 'dalam', 'tahap', 'pembangunan']	['pembangunan']
[['bagus', 'sebenarnya', 'cuma', 'parkirnya', 'jauh', 'dan', 'tiap', 'permainan', 'bayar']	['bagus', 'parkirnya', 'permainan']

Penambahan kata *stopwords*, bisa disesuaikan dengan mengecek frekuensi kata pada data. selanjutnya dipilih kata yang memang tidak dibutuhkan pada proses selanjutnya berdasarkan seberapa banyak frekuensinya.

4.2.6. *Stemming*

Proses berikutnya, mengubah setiap kata yang memiliki imbuhan menjadi bentuk dasar dari kata tersebut dengan menghilangkan imbuhan agar dapat lebih mudah untuk proses berikutnya. *Library* sastrawi digunakan dalam tahap ini.

Tabel 4. 9 Proses *Stemming*

Sebelum	Sesudah
['kurang', 'ramai', 'masakannya', 'enak', 'porsinya', 'ditambahi', 'coba', 'kolam', 'renangnya', 'menambah', 'antusias']	['kurang', 'ramai', 'masakan', 'enak', 'porsi', 'tambah', 'coba', 'kolam', 'renang', 'tambah', 'antusias']
['selecta', 'kesadaran', 'menjaga', 'kebersihan', 'memprihatinkan', 'sampah', 'disediakan', 'kurang', 'kondisi', 'toilet', 'menjijikkan']	['selecta', 'sadar', 'jaga', 'bersih', 'prihatin', 'sampah', 'sedia', 'kurang', 'kondisi', 'toilet', 'jijik']
['pembangunan']	['bangun']
[['bagus', 'sebenarnya', 'cuma', 'parkirnya', 'jauh', 'dan', 'tiap', 'permainan', 'bayar']	['bagus', 'parkir', 'main']

4.3. Pelabelan Data

Data yang telah selesai dalam rangkaian proses *preprocessing,* selanjutnya kata-kata yang terpisah sebelumnya dilakukan *join* untuk digabungkan kembali seperti kalimat. Dalam proses ini dilakukan pengecekan ulang terkait *null value* atau nilai kosong setelah proses *preprocessing* selesai. Proses pelabelan berdasarkan nilai polaritas kata dengan mengimplementasikan *library TextBlob.* Rumus perhitungan dan penentuan label sudah dilampirkan pada Bab 3.

Tabel 4. 10 Hasil Pelabelan Data

Ulasan	Polaritas	Label
kurang ramai masakan enak porsi tambah coba kolam renang tambah antusias	0.363636	positif
selecta sadar jaga bersih prihatin sampah sedia kurang kondisi toilet jijik	-0.018182	negatif
bangun	0.000000	netral
bagus parkir main	0.333333	positif

4.4. Validasi Pelabelan

Validasi manual dilakukan oleh pengelola wisata di Kabupaten Gresik. Hasil dari data yang sudah terlabelkan oleh *TextBlob* sebanyak 3012 data dengan rincian sebanyak 2462 data label positif, 356 data label negatif, dan 194 data label netral. Dari seluruh data diambil sampel dengan teknik *non-probability judgement sampling* sebagai validasi data. Jumlah sampel sebanyak 355 data dengan rincian 284 label positif, 48 label negatif, dan 23 netral.

Beberapa hasil yang dalam validasi pelabelan oleh pengelola wisata dengan angka 0 sebagai label netral, 1 sebagai label positif, 2 sebagai label negatif. Hasil tersebut dibandingkan dengan label *TextBlob* dengan tanda V artinya "Valid", atau sesuai karena hasil label oleh pengelola dan *TextBlob* sama, dan tanda I artinya "Invalid", atau tidak sesuai karena hasil label oleh pengelola dan *TextBlob* berbeda.

Tabel 4. 11 Validasi Pelabelan

Review	Pengelola			Label TextBlob	V/I
	I	II	III		
Bintang 4 dulu ya,,, tanamannya lebih dirawat lagi ,, pasti lebih bagus,,	1	1	0	1	V
Alhamdulillah pertama kali kesitu	1	1	1	1	V
Ya lumayan pas k sana kolam renang tidak ada airnya ...	0	2	2	0	I

Review	Pengelola			Label TextBlob	V/I
	I	II	III		
Jalan masuk manuju tempatnya masih sempit tapi masih bisa dilewati mobil	1	2	2	2	V
Tempat wisata di gresik banyak spot foto instagramable tiket masuk 10rb.tempat parkir mobil krng tertata rapi	0	1	1	1	V
Banyak area spot foto yang bagus bagus .. tiket murah .. cuma 15rb (free 1 pcs kripik/tukar kupon) .. Makanan warung murah" ..	1	1	0	1	V
memanjakan mata dengan melihat bunga2 dan tanaman hijau 😍 ….	1	0	1	1	V
- memanjakan mata - banyak tempat istirahat - makanan di lokasi harga standart - kalau panas mending jgn kesini, kurang bisa menikmati - cocok sekali untuk wisata keluarga	1	0	0	1	I
Tempat Wisata dengan keindahan alam yg sangat bagus dan menarik 👍👍👍👍👍 …	1	1	1	1	V

Berdasarkan penyesuaian hasil pelabelan yang diberikan oleh pengelola wisata dan dari *TextBlob* mendapatkan akurasi 81% dengan rincian sebagai berikut:

Tabel 4. 12 Akurasi Validasi

		TextBlob		
		Positif	Negatif	Netral
Pengelola	Positif	262	15	13
	Negatif	7	17	3
	Netral	15	16	7

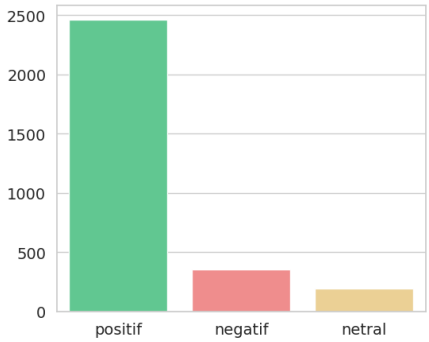
$$Accuracy = \frac{262 + 17 + 7}{262 + 7 + 15 + 17 + 15 + 16 + 7 + 13 + 3} = 0.81$$

Hasil tersebut sebagai acuan bahwa pelabelan yang dilakukan oleh mesin lewat *library python* yaitu *TextBlob* dapat dipertanggungjawabkan dengan kemungkinan proporsi prediksi dengan label asli sekitar 81%.

4.5. Visualisasi Hasil Pelabelan

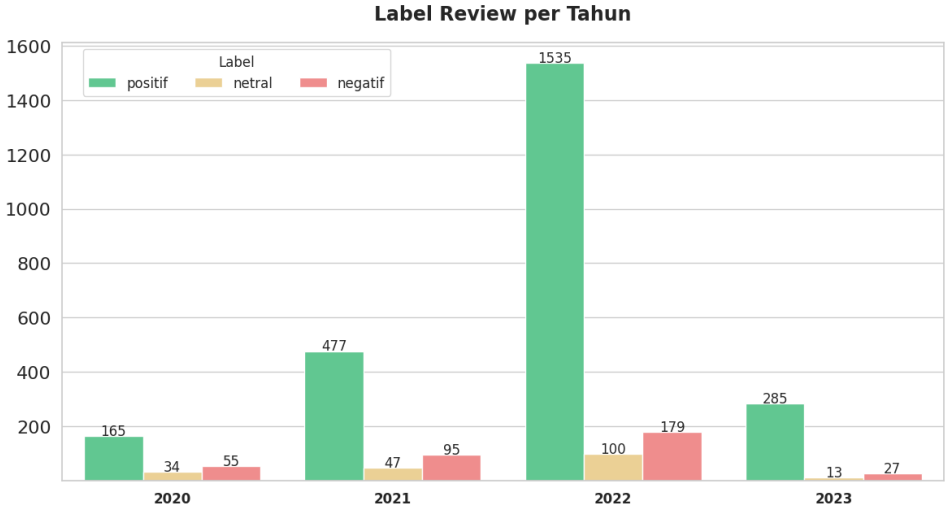
Data yang telah diberi label menggunakan *TextBlob* dipresentasikan dalam bentuk visual *bar chart* dan *wordcloud*. Secara persebaran jumlah per label dapat terlihat lebih jelas dalam bar chart. Pada Gambar 4. hasil visual menunjukkan label

data positif sangat mendominasi dibanding dengan label negatif dan netral. Rincian hasil pelabelan sebanyak 2462 data label positif, 356 data label negatif, dan 194 data label netral. Dari hasil ini terlihat bahwa persebaran label data tidak seimbang atau biasa disebut *imbalance*.



Gambar 4. 1 Hasil Pelabelan

Berdasarkan persebaran label per tahun, 2020 didapati 254 data, 2021 dengan 619 data, 2022 dengan 1814 data, 2023 dengan 325 data. Pada tahun 2022 mendominasi dengan jumlah review terbanyak. Pada tahun tersebut, kasus covid-19 sudah sedikit mereda, dan hal tersebut mempengaruhi aktivitas dari pengunjung yang membuat review pada wisata terjadi lonjakan. Persebaran label secara jelas sebagai berikut:



Gambar 4. 2 Hasil Label Berdasarkan Tahun

Visualisasi yang dihasilkan dari *wordcloud* positif terlihat beberapa kata yang paling dominan adalah “bagus”, “murah”, “bersih”, “keluarga”, “nyaman”, “indah”. Kata-kata dari *wordcloud* tersebut merepresentasikan bahwa tempat wisata

di Kabupaten Gresik dipandang bagus dan bersih, dengan tarif yang murah. Selain itu terdapat coba kata “keluarga” yang berarti cocok sebagai destinasi liburan bersama keluarga. Pada Gambar 4.3 ditampilkan beberapa kata yang termasuk.

Gambar 4. 3 *Wordcloud* Positif

Visualisasi berikutnya dari *wordcloud* negatif. Pada Gambar 4.4, beberapa kata yang dominan diantaranya, “panas", "kurang", "masuk", "parkir", "sampah", "kotor". Berdasarkan visualisasinya, kondisi umum cuaca khususnya di Kabupaten Gresik umumnya panas, terutama pada saat siang hari. Terdapat kata "sampah", dan "kotor" yang berarti masih terlihat dari tempat wisata yang dinyatakan kurang bersih dan terlihat sampah oleh pengunjung. Dan terkait kata “parkir, mungkin berkaitan dengan lahan ataupun penataan yang kurang semestinya.

Gambar 4. 4 *Wordcloud* Negatif

Pada visualisasi *wordcloud* netral, beberapa kata yang terhimpun yaitu "bagus", "panas", "kurang", "masuk", "tidak", "jalan", "bersih", "parkir", "bangun". Representasi kata-kata kemungkinan bermakna ambigu karena memang

label yang diberikan dinilai berdasarkan total nilai polaritas dan kata-kata ini yang awalnya bernilai positif atau negatif berganti menjadi netral karena gabungan nilai kata atau token yang lain. Alasan lain bisa didapat dari *library* dalam pelabelan yang digunakan belum memiliki korpus tersebut.

Gambar 4. 5 *Wordcloud* Netral

4.6. *Feature Extraction*

Vektorisasi kata dalam tahapan *feature extraction* menggunakan metode *word embedding* dengan melibatkan *library fasttext.* Dalam penerapannya, *fasttext* perlu melakukan pelatihan atau pembuatan model *word embedding* terlebih dahulu. Data yang dijadikan model dari dataset sebelumnya saat setelah proses *normalization*. Tujuannya agar model masih dapat memahami setiap keterkaitan kata dalam kalimat. Pelatihan model menggunakan parameter *vector*_size=300, *window*=5, dan min_*count*=1. Metode *skipgram* digunakan saat pelatihan.

Tabel 4. 13 Dimensi Vektor *Word Embedding*

Kata	Bobot Vektor
mahal	[-1.62349101e-02 -1.08257554e-01 -7.63800219e-02 -2.46672973e-01 -4.18619551e-02 -7.52320737e-02 -2.28508338e-02 2.44102761e-01....]
banget	[-4.65952009e-02 -1.07536539e-01 -3.43883969e-02 -2.08722338e-01 2.94592027e-02 -1.61106795e-01 5.40422201e-02 2.55411237e-01....]
tiketnya	[-5.06395753e-03 -1.12682573e-01 -5.68375066e-02 -2.24064335e-01 -2.63470076e-02 -1.16714239e-01 -3.32656840e-04 2.33217686e-01....]
sesuai	[-3.91320623e-02 -1.02338113e-01 -7.19698370e-02 -2.27308720e-01 2.49081780e-03 -1.07084498e-01 4.17233957e-03 2.38597691e-01....]
bagus	[-0.01574064 -0.131223 -0.05647508 -0.2058455 0.04474568 -0.22068912 0.06564671 0.25796372 0.31875953 -0.2018514....]

Setiap kata berdasarkan data, memiliki panjang dimensi vektor sebesar 300, sesuai dengan parameter yang diberikan. Hasil *vocabulary* yang tercatat pada model sebanyak 3986 kata. Selanjutnya dilakukan vektorisasi dari model *word embedding* yang dibuat terhadap data yang telah melalui preprocess. Dengan menggunakan modul pada *fasttext* yaitu *get_mean_vector* untuk menghitung rata-rata vektor pada kalimat sesuai dengan kata. Token atau kata yang sebelumnya memiliki nilai vektor akan dihitung rata-ratanya berdasarkan kata-kata yang terdapat dari kalimat. Apabila token tidak terdapat pada *vocabulary* model, nilai vektor tidak akan muncul dan diabaikan.

Tabel 4. 14 Vektorisasi *Word Embedding*

Normalisasi	Preprocess	Nilai Vektor
['mahal', 'banget', 'tiketnya', 'lt']	mahal tiket	[1.83229730e-03 -3.69379371e-02 -3.05171423e-02 -8.35390389e-02 -1.68148614e-02 -2.53048241e-02 -1.05776284e-02 7.98804760e-02....]
['tidak', 'sesuai', 'dengan', 'alamat']	tidak sesuai alamat	[-1.82894450e-02 -3.78937796e-02 -2.51321290e-02 -8.65281001e-02 3.46543919e-03 -4.34830673e-02 3.85852181e-03 9.33042169e-02....]
['sayang', 'fasilitas', 'umum', 'untuk', 'mengisi', 'seluler', 'tidak', 'ada']	fasilitas isi seluler tidak	[-1.39559107e-02 -4.02228348e-02 -2.63559744e-02 -8.75243619e-02 -5.00383321e-05 -4.36284542e-02 2.90480116e-03 9.26949680e-02....]
['pemandangan', 'dari', 'tangga', 'bagus', 'tempat', 'sepeda', 'air', 'juga', 'bagus']	pandang tangga bagus air	[-1.08295660e-02 -4.29774411e-02 -2.43364405e-02 -8.30833465e-02 1.47760450e-03 -5.16274273e-02 9.11834277e-03 9.31149125e-02....]
['tempat', 'bagus', 'dan', 'nyaman']	bagus nyaman	[-4.79991268e-03 -4.89871763e-02 -2.28010975e-02 -7.29774088e-02 2.01044679e-02 -8.48201513e-02 2.35946774e-02 9.26606059e-02....]
['bagus', 'kalau', 'malam', 'hari']	bagus malam	[-0.01258913 -0.04042022 -0.02294916 -0.08145761 0.00769187 -0.05813879 0.01428309 0.09544285....]
['bagus', 'banyak', 'spot', 'foto', 'yang', 'menarik', 'harga', 'ribu', 'dapat', 'voucer', 'untuk', 'beli', 'oleh', 'seperti', 'ribu', 'parkir', 'mobil', 'ribu']	bagus banyak foto tarik harga voucer parkir	[-0.00197906 -0.04406758 -0.0268376 -0.08085848 -0.00358149 -0.05404743 0.00753708 0.08718611 0.09511771 -0.06591768 -0.04603549 -0.02666262 -0.00765704 -0.04430808...]

4.7. *Dimensional Reduction*

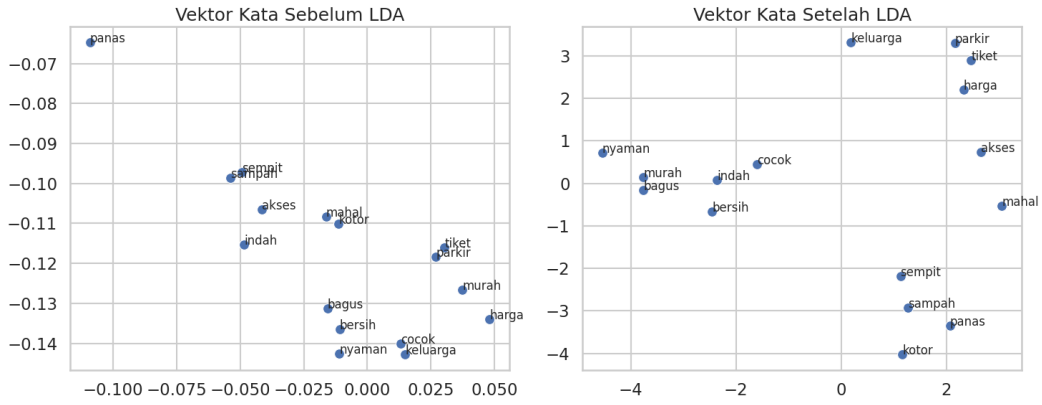
Setelah melalui vektorisasi, proses berikutnya dilakukan *dimesional reduction*. Proses ini melibatkan *library* dari python yaitu *Linear Discriminant*

Analysis atau LDA. Dari vektorisasi yang dilakukan sebelumnya sudah jelas dimensi vektor yang dibuat lumayan besar, maka proses ini diberlakukan. Penggunaan LDA diharapkan mengurangi kompleksitas waktu dan penggunaan memori dalam klasifikasi teks.

Tabel 4. 15 Hasil *Dimensionality Reduction* Dengan LDA

Preprocess	Nilai Vektor	LDA
mahal tiket	[2.70726206e-03 -3.51351425e-02 -2.54306123e-02 -7.29529858e-02 -3.43517400e-02 -1.79948024e-02 -2.71813665e-03 9.67339873e-02....]	[-1.16006283663432, 1.20231756702958]
tidak sesuai alamat	[-1.99843477e-02 -3.19549292e-02 -1.52914627e-02 -6.86511993e-02 -2.83875014e-03 -2.84010191e-02 -4.48028743e-03 1.11028820e-01....]	[-2.07791727974203, 2.66395543622202]
fasilitas isi seluler tidak	[-1.68385487e-02 -3.38327065e-02 -1.77545156e-02 -6.98467493e-02 -7.24523328e-03 -2.95195505e-02 -4.26708534e-03 1.11021094e-01....]	[-2.92599423607206, 2.17904381247681]
pandang tangga bagus air	[-1.40075376e-02 -3.56660970e-02 -1.92637444e-02 -6.68459907e-02 -8.66690092e-03 -3.66220511e-02 4.50883526e-05 1.12328157e-01...]	[0.09743813447600, 0.617208598269308]
bagus nyaman	[-5.03349071e-03 -4.35868539e-02 -2.53299773e-02 -6.51155114e-02 3.35126300e-03 -7.50591457e-02 1.38598848e-02 1.14290513e-01....]	[1.08478058999961, -0.11861751764211]
bagus malam	[-1.68488994e-02 -3.29934955e-02 -1.48201361e-02 -6.61741495e-02 -2.75033526e-03 -4.32174504e-02 5.64206205e-03 1.15728125e-01....]	[-0.08218208424129, -0.95042429838153]
bagus banyak foto tarik harga voucer parkir	[-3.12697631e-03 -3.96755524e-02 -2.62661614e-02 -6.96963519e-02 -2.05301624e-02 -4.59569581e-02 8.15425161e-03 1.07930966e-01...]	[0.47461380326610, -0.29999275565206]

Nilai vektor sebelumnya dengan dimensi sekian akan reduksi hingga sesuai yang ditentukan berdasarkan *n_component* yang telah diberikan. Pada penelitian ini menggunakan n_*component* = 2, yang membuat dimensi nilai vektor menjadi 2. Pemakaian LDA dapat mempengaruhi kedekatan vektor tiap kata pada model *word embedding*. Sebagai contoh berikut perbedaan vektor kata sebelum dan sesudah menggunakan LDA.



Gambar 4. 6 Vektor Kata Sebelum & Sesudah LDA

Perubahan visualisasi vektor didasari dari label dari setiap vektor. Oleh karenanya, dari setiap vektor akan membentuk sebuah kelas sesuai dengan label dari vektor tersebut.

4.8. *Cross Validation*

Proses berikutnya dilakukan *Cross validation* dengan metode *k-fold cross validation.* Nilai k =10 diberikan akan membagi 10 potongan data yang digunakan dengan proporsi data latih 90% dan data uji 10%. Setiap *fold* dari k akan melakukan pelatihan dan pengujian sebanyak 10 kali. Hasil setiap *fold* bisa bervariasi tergantung bagian data yang terpotong dalam pembagian proporsi sebelumnya. Dalam proses ini, *hyperparameter tuning* dengan metode *GridSearchCV* dilakukan. Dengan parameter-parameter yang sudah ditentukan sebelumnya dalam Bab 3.

Tabel 4. 16 Hasil *GridSearchCV*

Model	Parameter		
	Kernel	C	Gamma
SVM	RBF	100	10
SVM LDA	RBF	100	1
SVM SMOTE	RBF	100	10
SVM SMOTE LDA	RBF	1	10

Berdasarkan hasil *GridSearchCV*, terdapat beberapa perbedaan parameter pada setiap model. Dari hasil penentuan tersebut, berikutnya lanjut melakukan *Cross Validation Score* dengan menerapkan parameter yang terpilih dari metode *GridSearchCV*. *Scoring* pada *cross_val_score* menggunakan *F1-Macro* karena *imbalance* data.

Tabel 4. 17 *Cross Validation Score*

Fold	*Cross Validation Score*			
	SVM	SVM+LDA	SVM+SMOTE	SVM+SMOTE+LDA
1	0.44977188	0.74062619	0.82305418	0.91131132
2	0.47902639	0.77060932	0.80522372	0.91472014
3	0.52105509	0.74158551	0.80084171	0.93384108
4	0.46044487	0.72047242	0.82550609	0.92839691
5	0.61169066	0.81278233	0.8090671	0.92044334
6	0.4821534	0.80413089	0.79425049	0.89335439
7	0.47967153	0.81876484	0.79425705	0.91721335
8	0.47214486	0.88559671	0.82341891	0.9179073
9	0.55595188	0.80861678	0.7993305	0.91471969
10	0.52051282	0.7818057	0.83528702	0.91848636
Mean	0.50324234	0.78849907	0.81102368	0.91703939

Berdasarkan hasil rata-rata score, penerapan LDA pada model mengalami peningkatan. Model SVM dengan LDA meningkat dengan skor 0.78. Penggunaan SMOTE sebagai *upsampling* karena *imbalance* data memberikan dampak kenaikan *score* dengan hasil 0.81, dengan ditambah menggunakan LDA semakin meningkatkan nilai skor menjadi 0.91. Hasil skor ini sebagai acuan bagaimana seberapa model dapat memahami data.

4.9. *Modelling*

Pembuatan model menyesuaikan dengan beberapa model dengan parameter-parameter yang terpilih sebelumnya. Dalam tahap ini *split* data dilakukan dengan proporsi 80:20, 80% sebagai data training dan 20% sebagai data testing, yang berarti dari total data 3012 data, 2409 data digunakan sebagai data training dan 603 data sebagai data testing. Pelatihan setiap model sesuai skenario yang telah direncanakan.

4.9.1. Model *Training*

- Skenario 1

Tahap pelatihan model dilakukan dengan SVM sesuai dengan parameter pada Tabel 4.15. Data pelatihan berasal dari proses *cross validation* sebelumnya untuk mencegah overfitting. Sebelumnya, dilakukan perubahan terhadap label dengan menggunakan fungsi *LabelEncoder*(), dengan angka 0 sebagai netral, sebagai positif, dan 2 sebagai negatif. Pembagian label menggunakan metode *stratify* yang berguna agar persebaran dalam pembagian label merata.

1. SVM

   Proses dilakukan dengan model SVM dengan nilai fitur berdasarkan vektorisasi *word embedding* dan nilai target dari label. Dimensi fitur sejumlah 300 dengan baris sesuai pembagian porsi sebelumnya.

2. SVM + LDA

   Penerapan LDA mengubah dimensi vektor dari vektorisasi *word embedding*. Nilai fitur dari data testing akan berkurang dari sebelumnya menjadi 2. Dengan direduksinya dimensi akan berpengaruh pada waktu proses pelatihan dan penggunaan memori.

- Skenario 2

Pada skenario kedua, pemberlakuan *upsampling* diterapkan terhadap label minoritas. SMOTE dipilih untuk melakukan *upsampling* terhadap data. *Upsampling* dilakukan sebelum pembagian data dengan tujuan menyeimbangkan data terlebih dahulu. Persebaran awal data yaitu label positif 2562, negatif 358 dan netral 194, kemudian disampling agar persebarannya sama. Total data meningkat menjadi 7386 data, dengan setiap label berjumlah sama yaitu 2462. Proses berikutnya data dibagi dengan proporsi yang sama, dengan rincian 5908 data sebagai data training dan 1478 sebagai data testing.

1. SVM + SMOTE

   Dari cross validation, dengan menggunakan parameter yang optimal pada Tabel 4.15, selanjutnya pelatihan model dengan data yang sudah di upsampling. Pada proses cv berlangsung lebih lama dari model dengan data awal, karena jumlah data yang lebih besar.

2. SVM + SMOTE + LDA

   Pada proses ini, setelah data dibagi akan direduksi dimensinya dengan LDA, dimensi yang diberikan yaitu n_*component* = 2. Setelah dimensi fitur direduksi lanjut pelatihan model sesuai parameter sebelumnya. Reduksi dimensi dengan LDA sangat berpengaruh terhadap proses cv yang lebih cepat dibanding tanpa menggunakan LDA.

Berdasarkan hasil evaluasi training dari tiap model, model dengan LDA memiliki nilai akurasi lebih tinggi, dibanding dari model tanpa LDA. Pada model skenario 1, dengan *imbalance* data, menghasilkan metrik evaluasi training yang

cukup baik, terutama setelah penggunaan LDA. Skenario 2, evaluasi training tiap model sudah optimal. Data yang seimbang sangat mempengaruhi nilai metrik evaluasi training.

Tabel 4. 18 Evaluasi Training Model

Model	*Accuracy*	*Precision*	*Recall*	*F1-Score*
SVM	86%	86%	52%	59%
SVM+LDA	93%	91%	79%	84%
SVM+SMOTE	83%	83%	83%	83%
SVM+SMOTE+LDA	92%	92%	92%	92%

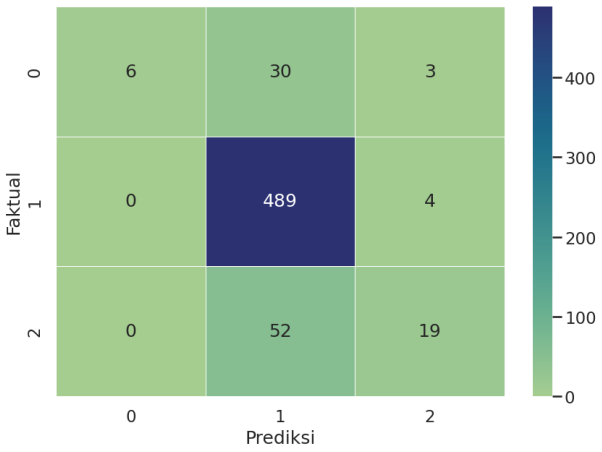
Hasil perhitungan *precision*, *recall*, dan *f1-score* menggunakan *macro average*, yang menghasilkan nilai metrik dalam tabel. Berdasarkan hasil evaluasi data latih, model SVM dengan LDA memberikan hasil yang cukup baik. Terjadi peningkatan dari setiap metrik evaluasi training.

4.9.2. Evaluasi

Performa dari setiap model diuji dari beberapa metrik evaluasi seperti *Precision, Recall*, *f-measure*, dan *Accuracy*. Hasil metrik evaluasi diperoleh berdasarkan nilai *confusion matrix* tiap model. Setiap label sentimen diganti dengan angka dengan angka 0 sebagai label netral, 1 sebagai label positif, 2 sebagai label negatif.

1. SVM

   Model awal yang telah dilatih, menghasilkan nilai *confusion matrix* seperti gambar berikut dengan 603 data sebagai testing.



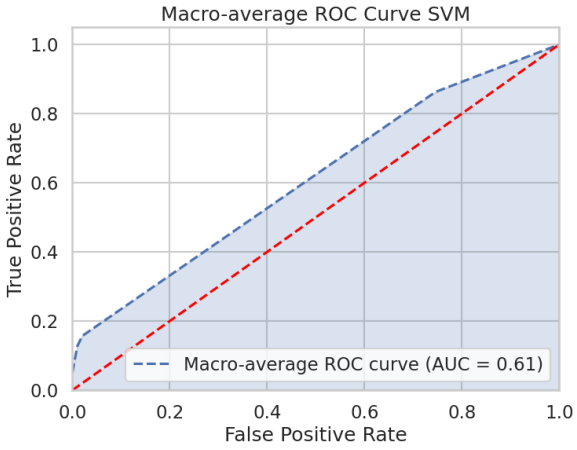
Gambar 4. 7 *Confusion Matrix* Model SVM

Warna pada diagram terlihat sangat kontras karena persebaran data yang tidak seimbang. Berdasarkan hasil tersebut, berpengaruh terhadap metrik-metrik evaluasi khususnya pada label netral dan negatif.

Tabel 4. 19 Evaluasi SVM Tiap Label

Label	Metrik Evaluasi		
	Precision	*Recall*	*F1-Score*
Netral	1.00	0.15	0.27
Positif	0.86	0.99	0.92
Negatif	0.73	0.27	0.39

Hasil akurasi model SVM diperoleh 85% pada hasil *cross validation score* sebelumnya diperoleh nilai rata-rata 0.50324234 berdasarkan nilai *F1-Macro*. Hasil ini belum bisa dibilang optimal karena nilai *Recall* dan *f1-score* masih rendah pada label netral dan negatif. Perhitungan metrik evaluasi lain menggunakan kurva ROC (*Receiving Operating Characteristic*). ROC akan mengilustrasikan sebuah grafik untuk mengevaluasi kinerja model dalam memisahkan kelas. Pembentukan kurva dibantu dengan nilai AUC (*Area Under The Curve*). Bentuk dari kurva ROC yang dihasilkan dari model ini

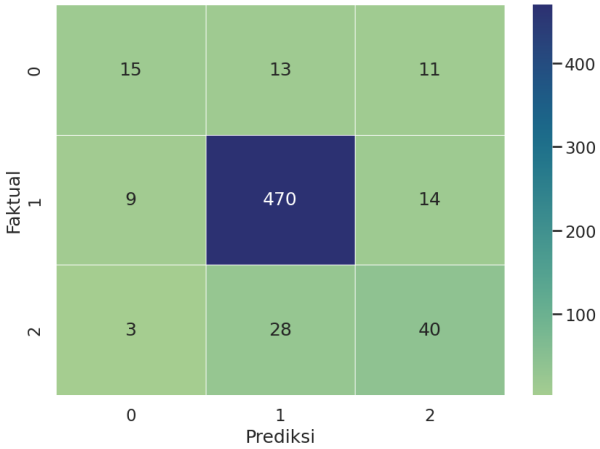


Gambar 4. 8 Kurva ROC Model SVM

Perhitungan AUC dari pembagian *True Positive Rate* atau *Recall* dan *False Positive Rate* atau *Fallout*. Rentang nilai AUC antara 0 sampai 1. Model menghasilkan nilai AUC 0.61, nilai tersebut masih tergolong rendah. Nilai ini didasarkan dari nilai *F1-Macro*.

2. SVM + LDA

Model selanjutnya, SVM dengan *dimensional reduction* LDA dengan jumlah data sama seperti model sebelumnya sebagai data testing. Berikut hasil *confusion matrix* model.



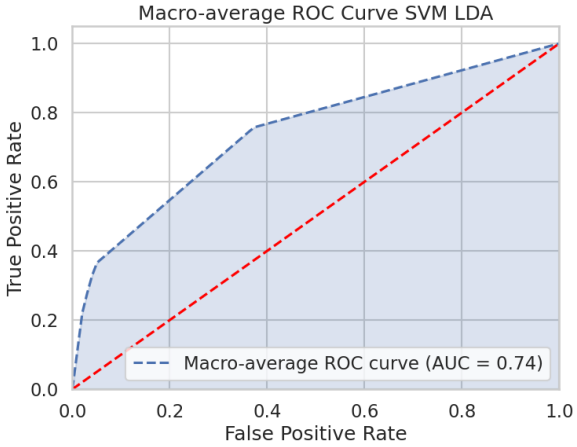
Gambar 4. 9 *Confusion Matrix* Model SVM+LDA

Hasil *confusion matrix* hampir mirip dengan model sebelumnya, warna pada setiap label begitu kontras karena data yang tidak seimbang. Metrik evaluasi yang diperoleh sedikit meningkat dengan model sebelumnya, meskipun pada label minoritas evaluasinya masih terbilang rendah.

Tabel 4. 20 Evaluasi SVM + LDA Tiap Label

Label	Metrik Evaluasi		
	Precision	*Recall*	*F1-Score*
Netral	0.56	0.38	0.45
Positif	0.92	0.95	0.94
Negatif	0.62	0.56	0.59

Peningkatan akurasi terjadi pada model, tercatat 87% dan pada hasil *cross validation score* memang sudah terjadi peningkatan, diperoleh nilai rata-rata 0.78849907 berdasarkan nilai F1-Macro. Hasil ini lebih baik jika dibandingkan tanpa menggunakan LDA, namun karena data yang tidak seimbang membuat hasil kurang pasti. Berikut kurva ROC dari model.

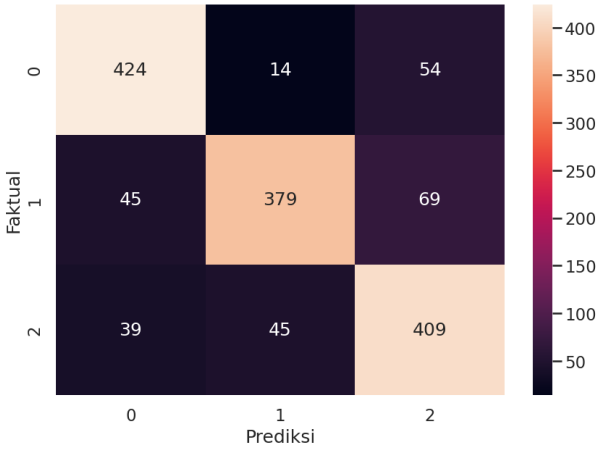


Gambar 4. 10 Kurva ROC Model SVM+LDA

Hasil nilai AUC meningkat dengan perolehan 0.74, model ini lebih baik dari model SVM tanpa LDA. Berdasarkan nilai tersebut dapat dikategorikan sebagai *Fair Classification.*

3. SVM + SMOTE

Model skenario 2, menggunakan *upsampling* terlebih dahulu pada data. 1478 data sebagai data training dengan pembagian label yang sama.



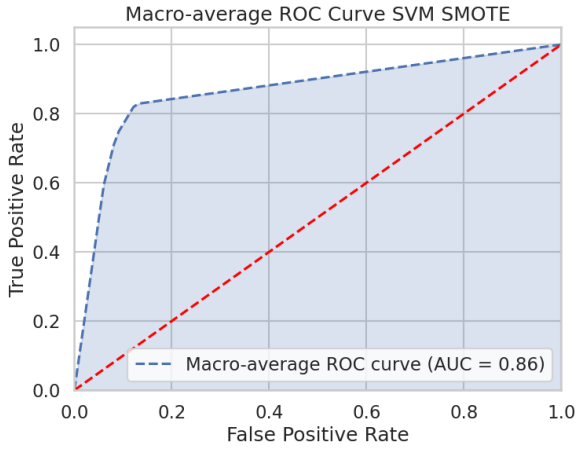
Gambar 4. 11 *Confusion Matrix* Model SVM + SMOTE

Hasil *confusion matrix* sangat berbeda, penerapan *upsampling* sangat berpengaruh. Metrik evaluasi pada setiap label minoritas meningkat.

Tabel 4. 21 Evaluasi SVM + SMOTE Tiap Label

Label	Metrik Evaluasi		
	Precision	*Recall*	*F1-Score*
Netral	0.83	0.86	0.85
Positif	0.87	0.77	0.81
Negatif	0.77	0.83	0.80

Secara akurasi memperoleh 82%, dengan nilai rata-rata *cross validation score* 0.81102368 berdasarkan nilai F1-Macro. Data yang seimbang menjadikan hasil evaluasi lebih optimal. Setiap nilai metrik lebih baik dibanding dengan model dari data awal. Berikut kurva ROC yang diperoleh.

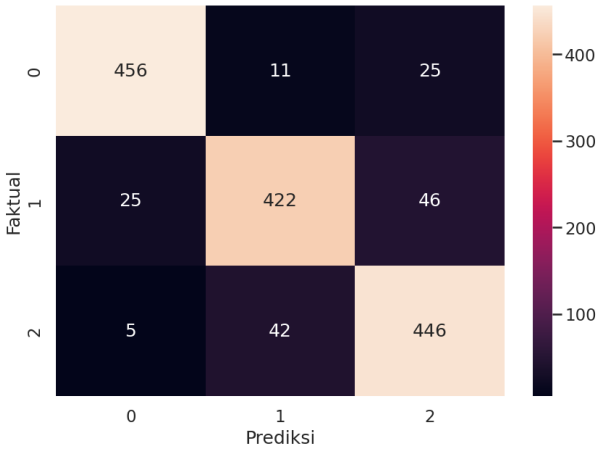


Gambar 4. 12 Kurva ROC Model SVM+SMOTE

Pada model skenario 2, penggunaan *upsampling* berpengaruh pada nilai AUC dengan menghasilkan 0.86. Hasil ini mengungguli 2 model skenario 1, dan tergolong pada kategori *Good Classification.*

4. SVM + SMOTE + LDA

Pada *confusion matrix*, model dengan *upsampling* dan LDA, memiliki hasil yang sangat baik. Dari model sebelumnya yang menerapkan *upsampling* yang memang sudah optimal diberikan LDA sehingga meningkatkan hasilnya.



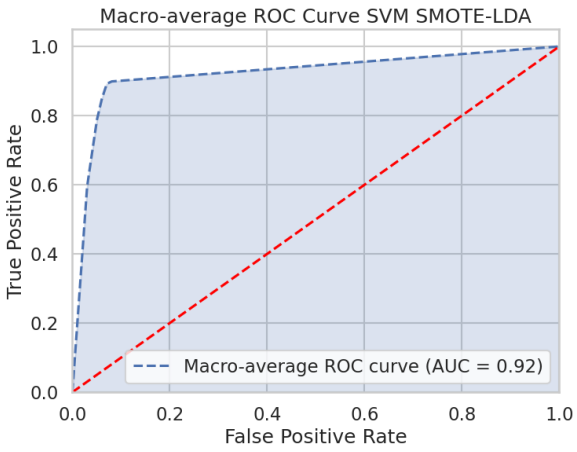
Gambar 4. 13 *Confusion Matrix* Model SVM+SMOTE+LDA

Hasil ini terbilang sangat baik dibandingkan model-model yang dilatih sebelumnya. Setiap metrik evaluasi pada label meningkat secara signifikan.

Tabel 4. 22 Evaluasi SVM + SMOTE + LDA Tiap Label

Label	Metrik Evaluasi		
	Precision	*Recall*	*F1-Score*
Netral	0.94	0.93	0.93
Positif	0.89	0.86	0.87
Negatif	0.86	0.90	0.88

Secara akurasi memperoleh 90%, dengan nilai rata-rata *cross validation score* 0.91703939 berdasarkan nilai F1-Macro. Dari *cross validation score* memang sudah terlihat nilainya yang tinggi, sehingga pada evaluasi model setiap metrik menghasilkan nilai yang baik. Berikut hasil kurva ROC dari model.



Gambar 4. 14 Kurva ROC Model SVM+SMOTE+LDA

Sesuai dengan model skenario 1, dengan LDA dapat meningkat nilai metrik evaluasi pada model. Hasil nilai AUC pada model ini mendapatkan 0.92. Nilai ini sangat tinggi dari model-model lain. Model dapat dikategorikan sebagai *Excellent Classification.*

Berdasarkan hasil evaluasi setiap model memiliki nilai metrik yang beragam. Dari hasil keseluruhan, model skenario 2 menghasilkan akurasi yang lebih baik namun perlu diperhatikan, tidak selamanya hasil dari nilai akurasi yang tinggi, membuat model lebih baik. Prediksi pada setiap model perlu dilakukan untuk mengetahui kebenaran dari hasil evaluasi model.

Tabel 4. 23 Hasil Metrik Evaluasi Model

Model	*Accuracy*	*Precision*	*Recall*	*F1-Score*
SVM	85%	86%	47%	53%
SVM+LDA	87%	70%	63%	66%
SVM+SMOTE	82%	82%	82%	82%
SVM+SMOTE+LDA	90%	90%	90%	90%

Untuk akurasi model SVM dengan LDA lebih baik dengan nilai 87%, dibandingkan dengan model SVM tanpa LDA 85%. Namun, penilaian berdasarkan akurasi masih belum dapat menandakan bahwa kinerja model baik, dikarenakan data yang digunakan dalam pelatihan merupakan *imbalance* data perlu metrik evaluasi lain sebagai penunjangnya. Tentunya nilai dari *precision, recall,* dan *f1-score* akan sangat berarti. Hasil perhitungan *precision, recall,* dan *f1-score* menggunakan *macro average,* yang menghasilkan nilai metrik dalam tabel.

Berdasarkan hasil evaluasi, nilai *precision* pada model SVM tanpa LDA lebih tinggi 86% dibandingkan dengan model SVM dengan LDA 70%, hal dikarenakan pada model SVM mengalami *overfit,* yang disebabkan dari *imbalance* data. Hasil *precision* pada label netral pada model SVM mendapatkan nilai 1.0, padahal data label netral minoritas. Berbeda dengan model SVM dengan LDA, yang memberikan hasil sesuai dengan keadaan dari data realnya. Hasil metrik evaluasi *recall,* model SVM menghasilkan 47% lebih tinggi model SVM dengan LDA 63%. Pada hasil *f1-score,* hasil rata-rata kombinasi harmonik *precision* dan *recall,* model SVM dengan LDA menghasilkan nilai *f1-score* 66%. Dari model SVM, karena nilai *recall* yang menurun, nilai *f1-score* menjadi 53%.

Pada model skenario 2, penerapan *upsampling* dengan SMOTE menghasilkan model yang baik. Pada SVM SMOTE akurasi menjadi 82% dengan evaluasi metrik yang merata dan model SVM SMOTE dengan LDA menghasilkan akurasi 90%. Terlihat pada hasil metrik evaluasi yang fit dan stabil. Pada hasil metrik evaluasi *precision, recall,* dan *f1-score,* pada model SVM SMOTE memperoleh nilai 82%, sedangkan pada model SVM SMOTE dengan LDA memperoleh nilai 90%. Data yang seimbang dapat memberikan hasil yang lebih baik.

4.10. Analisis Hasil Sentimen

Pada tahap ini setiap model yang telah dilatih dan dievaluasi digunakan untuk analisis sentimen. Prediksi dilakukan pada data baru yang di tahun 2023 dengan total data sebanyak 534 data ulasan. Sebelum melakukan prediksi, dilakukan *preprocessing* data terlebih dahulu. Tahapan *preprocessing* sama seperti pada proses sebelumnya. Hasil setelah preprocessing dilalui tercatat 511 data ulasan. Setelah itu dilakukan klasifikasi sentimen dengan model-model yang telah dibuat. Berikut hasil klasifikasi tiap model.

Tabel 4. 24 Hasil Klasifikasi Sentimen

Ulasan	SVM	SVM+LDA	SVM+ SMOTE	SVM+SMOTE+ LDA
Sudah lebih baik. Cuma catatan pada kran air bilas kolam nya jika terpakai semua aliran air kran nya jadi mengecil atau tidak keluar sama sekali.	1	2	1	1
Lumayan bagus, spot foto	1	1	1	1
Wisatanya cukup ramai parkirannya luas ada banyak sekali wahana sayang sekali harga tiket tiap wahana beda" .. Saran aja sih kenapa ndk dijadikan satu tiket aja di pintu masuk jd tdk usah antri lebih banyak lagi... N jg harga tiketnya kalau bisa lebih terjangkau lagi. Terima kasih🙏…	1	2	1	1
Murmer, makanan juga enak ,terjangkau jadi hemat dan gak bikin tongpes	1	1	1	1
Hotssss	1	2	1	1
Melali tipis2 lurrrrr.....	1	0	1	1
Biasa aja sih sepi bgt, kurang menarik, awal2 kesini rame makin kesini makin sepi.	1	2	1	1
untuk acara piknik bersama keluarga atau teman sangat cocok, tempatnya lumayan bersih dan terawat, ada gazebo dan sampahnya juga dipilah, mungkin kurang wahana bermain untuk anak-anak dan tempat wudhu kurang sedikit memadai	1	1	1	1
Jalannya sulit gabisa naik mobil	1	2	1	1

Total sentimen yang diperoleh dari prediksi tiap model tercantum dalam tabel berikut.

Tabel 4. 25 Total Prediksi Sentimen Tiap Model

Sentimen	Model			
	SVM	**SVM+LDA**	SVM+SMOTE	SVM+SMOTE+LDA
Netral	5 review	**22 review**	5 review	5 review
Positif	487 review	**456 review**	487 review	487 review
Negatif	19 review	**33 review**	19 review	19 review

Berdasarkan hasil prediksi, model SVM dengan LDA menghasilkan jumlah sentimen yang berbeda, dengan model-model lain. Pada Tabel 4. 24, hasil klasifikasi sentimen dari model SVM dengan LDA, lebih cocok memprediksi sentimen sesuai ulasan yang diberikan. Sesuai dengan hasil evaluasi pada Tabel 4.23, yang menyatakan bahwa model SVM kemungkinan *overfit*, sesuai dengan klasifikasi pada data prediksi, klasifikasi yang dihasilkan kurang tepat.

Pada model dengan *upsampling*, mengalami hal yang sama. Data sintetis yang diberikan oleh SMOTE mengalami *overgeneralization*, karena terjadi kemungkinan penyebaran data sintetis yang dibuat kepada wilayah kelas minoritas atau mayoritas (Wijayanti dkk., 2021). Kejadian ini yang menyebabkan hasil klasifikasi kurang tepat atau kinerjanya menurun. Kecepatan saat melakukan prediksi dari tiap model berbeda-beda. Berikut waktu yang dibutuhkan dari model melakukan prediksi sentimen.

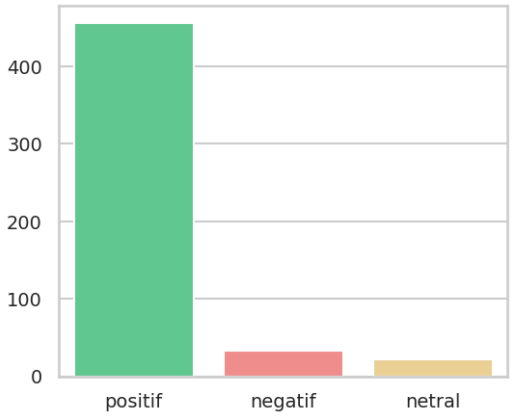
Tabel 4. 26 Waktu Prediksi Model

	SVM	**SVM+LDA**	SVM+SMOTE	SVM+SMOTE+LDA
Waktu Prediksi	80,34 ms	**14,866 ms**	119,079 ms	79,204 ms

Berdasarkan waktu tempuh dari setiap model, penerapan *dimensional reduction* dengan LDA dapat mengurangi kompleksitas waktu. Dari hasil tersebut, model SVM dengan LDA dipilih sebagai model untuk memprediksi sentimen dan analisis ulasan wisata di Kabupaten Gresik sehingga tujuan dari penelitian ini dapat terpenuhi.

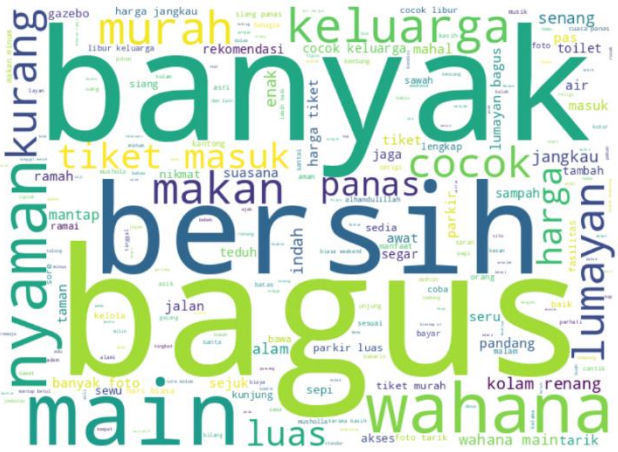
Berdasarkan hasil klasifikasi sentimen yang diperoleh dari model SVM dan LDA dengan 511 ulasan dihasilkan 89% data atau 456 ulasan positif, 7% data atau 33 ulasan negatif dan 4% data atau 22 ulasan netral. Dari hasil klasifikasi sentimen

tersebut, dapat disimpulkan bahwa klasifikasi sentimen terhadap ulasan lokasi wisata di Kabupaten Gresik cenderung memiliki sentimen positif.



Gambar 4. 15 Hasil Klasifikasi Sentimen Berdasarkan Model SVM+LDA

Pada *bar chart* terlihat sentimen positif sangat mendominasi. Ulasan ini termasuk terbaru pada tahun 2023. Berdasarkan visualisasi *wordcloud* terdapat sebagian kata yang termasuk dalam sentimen positif, negatif dan netral. Semakin tinggi frekuensi kemunculan kata, gambar yang terbentuk akan semakin besar begitupun sebaliknya.



Gambar 4. 16 *Wordcloud* Sentimen Positif

Visualisasi yang dihasilkan dari *wordcloud* sentimen positif terlihat beberapa kata yang paling dominan adalah “bagus”, ”banyak”, “bersih”, “wahana”, “nyaman”, “main“, keluarga”. Kata-kata dari *wordcloud* tersebut merepresentasikan bahwa umumnya lokasi wisata di Kabupaten Gresik terlihat bagus dan bersih. Selain itu, terdapat kata “keluarga” yang berarti cocok sebagai destinasi liburan bersama keluarga. Terdapat kata “wahana”, dan “main” yang

berarti dalam lokasi tersebut sudah terdapat wahana untuk bermain baik untuk kalangan dewasa atau anak-anak.

Gambar 4. 17 *Wordcloud* Sentimen Negatif

Visualisasi berikutnya dari *wordcloud* sentimen negatif, beberapa kata yang dominan diantaranya, “panas", "banyak", "tiket", "parkir", "mahal". Salah satu kata yang paling mencolok yaitu “panas”, tidak dapat dipungkiri memang kondisi umum cuaca di Kabupaten Gresik memang panas, terutama pada saat siang hari. Kata ‘tiket” mengindikasikan 2 kemungkinan, dari tiket masuk atau tiket wahana yang terlalu mahal menurut ulasan pengunjung. Dan terkait kata “parkir, berkaitan dengan lahan yang kurang luas ataupun penataan yang kurang semestinya.

Gambar 4. 18 *Wordcloud* Sentimen Netral

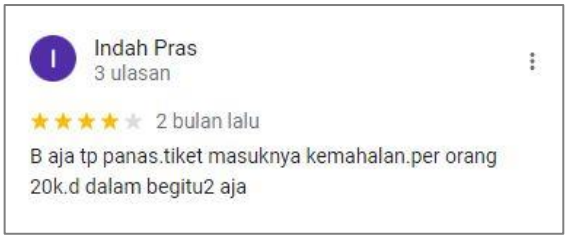
Pada *wordcloud* sentimen netral, terdapat beberapa kata yang terlihat diantaranya, “bagus", "kurang", "edu", "alami", "panas". Beberapa kata tersebut umumnya dapat tergolong pada sentimen positif, seperti “bagus”, “alami” dan sentimen negatif seperti “panas”, kurang”. Kondisi tersebut terjadi karena pada

pelabelan data sebelumnya menghitung nilai polaritas berdasarkan rata-rata dan menghasilkan nilai 0 yang berarti netral. Hal tersebut yang berpengaruh pada klasifikasi sentimen yang menyebabkkan sebagian kalimat diprediksi netral. Berikut hasil frekuensi kemunculan kata terbanyak pada setiap sentimen.

Tabel 4. 27 Frekuensi Kata Pada Sentimen

Frekuensi Kata					
Sentimen Positif		Sentimen Negatif		Sentimen Netral	
Kata	Frekuensi	Kata	Frekuensi	Kata	Frekuensi
bagus	130	panas	9	bagus	4
banyak	104	banyak	5	kurang	2
main	76	tiket	4	edu	2
wahana	74	parkir	4	alami	1
tiket	73	mahal	4	panas	1
keluarga	63	kurang	4	asyik	1
harga	56	siang	4	seruu	1
cocok	55	pagi	4	santera	1
bersih	52	jalan	3	versi	1
lumayan	51	masuk	3	kecewa	1

Berdasarkan frekuensi kata per sentimen, terdapat kata yang memiliki 2 arti dalam sudut pandang sentimen. Dari kata “tiket” mengindikasikan 2 pengertian namun pada dasarnya kata tersebut merujuk pada sebuah harga yang ditawarkan. Pernyataan tersebut didukung terdapatnya kata “harga” dalam sentimen. Beberapa wisata yang telah diambil ulasannya, dapat dikalkulasikan secara persentase 95% wisata menjual tiket dengan harga terjangkau, 5% wisata dengan harga yang relatif mahal. Harga tiket yang terjangkau merujuk pada tiket wahana karena terdapat kata “wahana” dalam sentimen positif. Untuk harga tiket yang mahal kemungkinan merujuk pada harga tiket masuk dari wisata di Kabupaten Gresik kurang sesuai untuk pengunjung. Pernyataan ini didukung dengan adanya kata “masuk” dalam sentimen negatif.



Gambar 4. 19 Ulasan Google Maps

Kata “panas” memiliki frekuensi tertinggi pada sentimen negatif yaitu 9 kali kemunculan kata, memang cuaca pada Kabupaten Gresik pastinya panas, terlebih lagi di saat siang hari. Indikasi waktu yang disyaratkan pada pagi dan siang karena terdapat frekuensi kata tersebut pada sentimen negatif.

Gambar 4. 20 Ulasan Google Maps

Pada frekuensi kata sentimen netral, setiap hasil kemunculan kata hanya terdapat satu kali kemunculan. Hal ini karena prediksi yang telah dilakukan, bergantung pada pelabelan data sebelumnya. Dasar pelabelan dilakukan dari rata-rata nilai polaritas pada kalimat yang menghasilkan nilai 0. Terdapat kata “bagus” & “kurang” yang seharusnya berada pada sentimen positif dan negatif. Berlaku juga terhadap hasil frekuensi kata lainnya yang kurang dari 2.

Dari pernyataan hasil analisis tersebut, dapat menjadi tambahan analisis terhadap pengelola wisata dalam pemilihan solusi peningkatan wisata di Kabupaten Gresik. Dengan memperhitungkan terkait tarif yang diberikan, dan memberikan banyak penghijauan atau tempat berteduh, agar pengunjung dari setiap wisata lebih optimal bahkan meningkat. Sentimen yang diberikan pengunjung memang lebih cenderung ke positif, dan dari hasil frekuensi, kondisi wisata dapat dibilang bagus, dan bersih. Keputusan untuk peningkatan kualitas tetap harus dipertahankan dari setiap wisata. Untuk komponen yang lain seperti “parkir”, “jalan” pasti dari setiap wisata punya rancangan tahap pembangunan terhadap hal itu, namun masih dalam tahap proses.

4.11. Diskusi

Pada latar belakang di Bab 1 berdasarkan penelitian sebelumnya menyatakan bahwa "penerapan *Linear Discriminant Analysis* dapat mengurangi kompleksitas waktu dalam klasifikasi teks" hasil penelitian ini sesuai dengan yang pernyataan tersebut bahwa model yang menerapkan LDA, dapat mengurangi waktu saat eksekusi, dalam hal ini saat pelaksanaan prediksi klasifikasi sentimen. Menurut analisa peneliti, mengurangnya waktu disebabkan karena reduksi dimensi yang diberikan, yang membuat fitur menjadi lebih sederhana dan akhirnya berimbas pada waktu eksekusi model. Akurasi yang dihasilkan dari model yang menerapkan LDA lebih meningkat dibandingkan model tanpa LDA, namun karena data yang *imbalance* membuat hasil presisi, *recall* dan *f1-score* model kurang sempurna.

BAB V
PENUTUP

Berikut merupakan kesimpulan dan saran yang dapat diberikan dari penelitian analisis sentimen destinasi wisata Kabupaten Gresik menggunakan *Linear Discriminant Analysis* (LDA) dan *Support Vector Machine* (SVM).

5.1. Kesimpulan

Dari proses penyelesaian penelitian, mulai dari awal sampai hasil diperoleh, dapat disimpulkan bahwa:

1. Model SVM dengan LDA dapat diimplementasikan dalam analisis sentimen. Proses awal melakukan data *preprocessing* dengan alur, *cleaning, case folding, tokenizing, normalization, stopwords, stemming.* Pelabelan data dengan *TextBlob*. Proses *word embedding* diawali dengan pembuatan model dengan library *FastText* yang selanjutnya model tersebut digunakan dalam vektorisasi kata. Berikutnya proses reduksi dimensi dengan LDA. Model dilatih dan menghasilkan *F1-score* 66% lebih baik dibandingkan model SVM tanpa LDA yang menghasilkan nilai *F1-score* 55%. Dari hasil klasifikasi, model SVM LDA memperoleh hasil yang lebih baik daripada model lainnya.
2. Hasil dari klasifikasi sentimen menunjukkan sentimen pengunjung terhadap lokasi wisata di Kabupaten Gresik cenderung positif. Jumlah data dengan sentimen positif 456 ulasan, sentimen negatif 33 ulasan, dan sentimen netral 22 ulasan. Model klasifikasi yang digunakan adalah model SVM dengan LDA.

5.2. Saran

Hasil dari penelitian yang dijelaskan masih belum sempurna, dan terdapat kekurangan. Oleh karenanya, berikut saran untuk pengembangan penelitian di masa yang akan datang:

1. Metode *deep learning* bisa diterapkan pada analisis sentimen berikutnya dengan menerapkan LDA.

2. Model *word embedding* pada dasarnya memahami konteks keterkaitan kata, maka data yang baik, dan besar akan menghasilkan model yang baik pula. Untuk hal ini bisa menggunakan *pre-trained* model *word embedding* atau melatih model dari korpus yang lebih besar dan dengan Metode vektorisasi kata yang lain seperti GloVe dan Word2Vector.
3. Terdapat beberapa kata yang dihapus oleh *stopwords* seperti "kurang", "banyak", dan "tidak". Pengecualian akan berpengaruh pada hasil label, namun berpengaruh terhadap frekeunsi kata. Pelabelan data sebaiknya dilakukan berdasarkan dari ahli agar hasil klasifikasi yang diberikan lebih sesuai.
4. Penerapan SMOTE pada data teks membuat hasil prediksi tidak akurat. Penanganan *imbalance* data terkhusus data teks dapat melakukan penerapan metode lain seperti *text augmentation*.

DAFTAR PUSTAKA

A. Ramezan, C., A. Warner, T., & E. Maxwell, A. (2019). Evaluation of Sampling and Cross-Validation Tuning Strategies for Regional-Scale Machine Learning Classification. *Remote Sensing*, *11*(2), 185. https://doi.org/10.3390/rs11020185

Abelard, A. R., & Sibaroni, Y. (2021). Multi-Aspect Sentiment Analysis On Netflix Application Using Latent Dirichlet Allocation and Support Vector Machine Methods. *JURNAL INFOTEL*, *13*(3), 128–133. https://doi.org/10.20895/infotel.v13i3.670

Abiola, O., Abayomi-Alli, A., Tale, O. A., Misra, S., & Abayomi-Alli, O. (2023). Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser. *Journal of Electrical Systems and Information Technology*, *10*(1), 1–20.

Agustiningsih, K. K., Utami, E., & Alsyaibani, M. A. (2022). Sentiment Analysis of COVID-19 Vaccines in Indonesia on Twitter Using Pre-Trained and Self-Training Word Embeddings. *Jurnal Ilmu Komputer Dan Informasi*, *15*(1), 39–46. https://doi.org/10.21609/jiki.v15i1.1044

Al-Khalidi, M. R. (2023). Perancangan Sistem Indikator Penilaian Kinerja Karyawan Divisi Produksi Pada PT Kurniawan Sejati Dengan Menggunakan Metode Analytical Hierarchy Process Dan Rating Scale. *Industrial Engineering Online Journal*, *12*(1), Article 1. https://ejournal3.undip.ac.id/index.php/ieoj/article/view/37448

Anasta, D. (2023). *Sentiment Analysis Review of The Sayurbox App*. https://www.kaggle.com/code/dindaanasta/sentiment-analysis-review-of-the-sayurbox-app

Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM).

Jurnal Teknologi Informasi dan Ilmu Komputer, *8*(1), 147. https://doi.org/10.25126/jtiik.0813944

Baihaqi, G. F., Ratnawati, D. E., & Hanggara, B. T. (2022). Analisis Sentimen Wisata Alun-Alun Kota Batu menggunakan Algoritma Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *6*(12), 6010–6018.

Berrar, D. (2019). Cross-Validation. Dalam *Encyclopedia of Bioinformatics and Computational Biology* (hlm. 542–545). Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.20349-X

Berry, M. W., Mohamed, A., & Yap, B. W. (Ed.). (2020). *Supervised and Unsupervised Learning for Data Science*. Springer International Publishing. https://doi.org/10.1007/978-3-030-22475-2

Bourequat, W., & Mourad, H. (2021). Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine. *International Journal of Advances in Data and Information Systems*, *2*(1), 36–44. https://doi.org/10.25008/ijadis.v2i1.1216

Cahyanti, F. E., Adiwijaya, & Faraby, S. A. (2020). On The Feature Extraction For Sentiment Analysis of Movie Reviews Based on SVM. *2020 8th International Conference on Information and Communication Technology (ICoICT)*, 1–5. https://doi.org/10.1109/ICoICT49345.2020.9166397

ÇeliK, Ö., & Koç, B. C. (2021). Classification of Turkish News Texts with TF-IDF, Word2vec and Fasttext Vector Model Methods. *Deu Muhendislik Fakultesi Fen ve Muhendislik*, *23*(67), 121–127. https://doi.org/10.21205/deufmd.2021236710

Data Kunjungan Wisata Online. (2018). https://dakuwison.gresikkab.go.id/

Diandra, D. (2022). *Analisis Sentimen Ulasan MyXL dengan SVM*. https://kaggle.com/code/dimasdiandraa/analisis-sentimen-ulasan-myxl-dengan-svm

Erfina, A., & Wardani, N. R. (2022). Analisis Sentimen Perguruan Tinggi Termewah Di Indonesia Menurut Ulasan Google Maps Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Manajemen Informatika & Sistem Informasi (MISI)*, *5*(1), 77–85.

Fahmi, A., Ramadhan, I., & Agussalim. (2020). Analisis Sentiment Masyarakat Selama Bulan Ramadhan Dalam Menghadapi Pandemi Covid-19. *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, *1*(1).

Fitriansyah, A. R., & Sibaroni, Y. (2023). Analisis Sentimen Terhadap Pembangunan Kereta Cepat Jakarta—Bandung Pada Media Sosial Twitter Menggunakan Metode SVM dan GloVe Word Embedding. *e-Proceeding of Engineering*, *10*(2), 1713–1723.

Ginantra, N. L. W. S. R., Yanti, C. P., Prasetya, G. D., Sarasvananda, I. B. G., & Wiguna, I. K. A. G. (2022). Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, *11*(3), 205–215. https://doi.org/10.23887/janapati.v11i3.49450

Guntara, R. G. (2023). Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab. *ULIL ALBAB : Jurnal Ilmiah Multidisiplin*, *2*(6), 2091–2100. https://doi.org/10.56799/jim.v2i6.1578

Gupta, I., & Joshi, N. (2019). Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic. *Journal of Intelligent Systems*, *29*(1), 1611–1625. https://doi.org/10.1515/jisys-2019-0106

Haq, F. U. (2020). Penggunaan Google Review Sebagai Penilaian Kepuasan Pengunjung Dalam Pariwisata. *Tornare*, *2*(1), 10. https://doi.org/10.24198/tornare.v2i1.25826

Hb, B. G., Ravi, V., M, A. K., & KP, S. (2018, Maret 21). *Distributed Representation using Target Classes: Bag of Tricks for Security and Privacy Analytics*.

Hendrawan, I. R., Utami, E., & Hartanto, A. D. (2022). Analisis Perbandingan Metode Tf-Idf dan Word2vec pada Klasifikasi Teks Sentimen Masyarakat Terhadap Produk Lokal di Indonesia. *Smart Comp :Jurnalnya Orang Pintar Komputer*, *11*(3), 497–503.

Hendriyani, I. G. A. D. (2022, Mei 30). *SIARAN PERS: Indeks Kinerja Pariwisata Indonesia Raih Peringkat ke-32 Besar Dunia Menurut WEF*. Kemenparekraf/Baparekraf RI. https://www.kemenparekraf.go.id/berita/siaran-pers-indeks-kinerja-pariwisata-indonesia-raih-peringkat-ke-32-besar-dunia-menurut-wec

Herlawati, H., Handayanto, R. T., Atika, P. D., Khasanah, F. N., Yusuf, A. Y. P., & Septia, D. Y. (2021). Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine. *Jurnal Komtika (Komputasi dan Informatika)*, *5*(2), 153–163. https://doi.org/10.31603/komtika.v5i2.6280

Hesay, I. K., Indiriati, & Adinugroho, S. (2021). Analisis Sentimen Ulasan Pengunjung Simpang Lima Gumul Kediri menggunakan Metode BM25 dan Neighbor-Weighted K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *5*(7), 3160–3169.

Ikegami, A., & Darmawan, I. D. M. B. A. (2022). Analisis Sentimen dan Pemodelan Topik Ulasan Aplikasi Noice Menggunakan XGBoost dan LDA. *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, *1*(1), 325–336.

Kaope, C., & Pristyanto, Y. (2023). The Effect of Class Imbalance Handling on Datasets Toward Classification Algorithm Performance. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *22*(2), 227–238. https://doi.org/10.30812/matrik.v22i2.2515

Kasanah, A. N., Muladi, M., & Pujianto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *3*(2), 196–201. https://doi.org/10.29207/resti.v3i2.945

Kaur, C., & Sharma, Dr. A. (2020). Sentiment Analysis of Tweets on Social Issues using Machine Learning Approach. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(4), 6303–6311. https://doi.org/10.30534/ijatcse/2020/310942020

Khofifah, W., Rahayu, D. N., & Yusuf, A. M. (2022). Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps. *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, *16*(4), 28–38. https://doi.org/10.35969/interkom.v16i4.192

Khomsah, S., Ramadhani, R. D., & Wijaya, S. (2022). The Accuracy Comparison Between Word2Vec and FastText On Sentiment Analysis of Hotel Reviews. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *6*(3), 352–358. https://doi.org/10.29207/resti.v6i3.3711

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text Classification Algorithms: A Survey. *Information*, *10*(4), 150. https://doi.org/10.3390/info10040150

Kumar, L. A., Jayashree, L. S., & Manimegalai, R. (Ed.). (2020). *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications: AISGSC 2019*. Springer International Publishing. https://doi.org/10.1007/978-3-030-24051-6

Lai, V., & Tan, C. (2019). On Human Predictions with Explanations and Predictions of Machine Learning Models: A Case Study on Deception Detection. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 29–38. https://doi.org/10.1145/3287560.3287590

Lappeman, J., Clark, R., Evans, J., Sierra-Rubia, L., & Gordon, P. (2020). Studying social media sentiment using human validated analysis. *MethodsX*, *7*, 100867. https://doi.org/10.1016/j.mex.2020.100867

Larasati, L., Nabilla, S., & Haryanto, E. (2022). Sentiment Analysis Untuk Review Destinasi Wisata Unggulan Gunung Kidul Menggunakan Metode Lexicon Dan Pivot. *Indonesian Journal of Business Intelligence (IJUBI)*, *5*(2), 102. https://doi.org/10.21927/ijubi.v5i2.2604

Mengistie, T. T., & Kumar, D. (2021). Deep Learning Based Sentiment Analysis On COVID-19 Public Reviews. *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 444–449. https://doi.org/10.1109/ICAIIC51459.2021.9415191

Muharni, S., & Candra, A. (2022). *Buku Modul Visualisasi Data Menggunakan Data Studio*.

Najib, A. C., Irsyad, A., Qandi, G. A., & Rakhmawati, N. A. (2019). Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter. *Fountain of Informatics Journal*, *4*(2), 41. https://doi.org/10.21111/fij.v4i2.3573

Nurdin, A., Anggo Seno Aji, B., Bustamin, A., & Abidin, Z. (2020). Perbandingan Kinerja Word Embedding Word2vec, Glove, Dan Fasttext Pada Klasifikasi Teks. *Jurnal Tekno Kompak*, *14*(2), 74–79. https://doi.org/10.33365/jtk.v14i2.732

Nurfajiah, A., Hartati, T., & Amalia, D. R. (2021). Integrated Library System untuk Meningkatkan Efektivitas Layanan Perpustakaan Dengan Menggunakan Metode Algoritma Apriori (Studi Kasus: Perpustakaan Kabupaten Cirebon). *Journal of Information Technology*, *3*(1), 39–44.

Oktafani, M., & Prasetyaningrum, P. T. (2022). Implementasi Support Vector Machine Untuk Analisis Sentimen Komentar Aplikasi Tanda Tangan Digital. *Jurnal Sistem Informasi Dan Bisnis Cerdas (SIBC)*, *15*(1), 10–19.

Parameswari, P. L., Astuti, I., & Ariestya, W. W. (2022). Implementasi Metode AHP Pada Sistem Pendukung Keputusan Pariwisata Jawa Timur. *Jurnal Teknoinfo*, *16*(1), 40. https://doi.org/10.33365/jti.v16i1.1401

Prasetyo, D. B., & Hidayatullah, A. F. (2020). Identifikasi Dual Sentimen Terhadap Ulasan Objek Wisata di Daerah Istimewa Yogyakarta. *AUTOMATA*, *1*(1).

Pratama, Y. T., Bachtiar, F. A., & Setiawan, N. Y. (2018). Analisis Sentimen Opini Pelanggan Terhadap Aspek Pariwisata Pantai Malang Selatan Menggunakan TF-IDF Dan Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *2*(12), 6244–6252.

Ramadhan, V. G., & Sibaroni, Y. (2021). Sentiment Analysis of Public Opinion Related to Rapid Test Using LDA Method. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(4), 672–679. https://doi.org/10.29207/resti.v5i4.3139

Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access*, *8*, 54776–54788. https://doi.org/10.1109/ACCESS.2020.2980942

Ridho, D. M., & Marseto. (2023). Analisis Pengaruh Pariwisata Terhadap Pendapatan Asli Daerah Di Kabupaten Gresik Tahun 2010-2022. *Jurnal Ekonomi Bisnis Manajemen Prima*, *4*(2), 72–82.

Santika, E. F. (2023). *Wisatawan Jawa Timur Mendominasi Arus Pariwisata Domestik 2022 | Databoks*. https://databoks.katadata.co.id/datapublish/2023/05/05/wisatawan-jawa-timur-mendominasi-arus-pariwisata-domestik-2022

Saputra, R. R. (2022, Agustus 30). *Daftar 10 Provinsi Terkaya di Indonesia, Nomor 7 Ditempati IKN Nusantara*. iNews.ID. https://kaltim.inews.id/berita/provinsi-terkaya-di-indonesia

Sharma, D., Sabharwal, M., Goyal, V., & Vij, M. (2020). Sentiment Analysis Techniques for Social Media Data: A Review. Dalam A. K. Luhach, J. A. Kosa, R. C. Poonia, X.-Z. Gao, & D. Singh (Ed.), *First International Conference on Sustainable Technologies for Computational Intelligence* (Vol. 1045, hlm. 75–90). Springer Singapore. https://doi.org/10.1007/978-981-15-0029-9_7

Somantri, O., & Dairoh, D. (2019). Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, *5*(2), 191. https://doi.org/10.26418/jp.v5i2.32661

Steven, C., & Wella, W. (2020). The Right Sentiment Analysis Method of Indonesian Tourism in Social Media Twitter. *IJNMT (International Journal of New Media Technology)*, *7*(2), 102–110. https://doi.org/10.31937/ijnmt.v7i2.1732

Sumayah, S., Sembiring, F., & Jatmiko, W. (2023). Analisis Sentimen Masyarakat Indonesia Terhadap Metaverse Menggunakan Algoritma Support Vector Machine. *Jurnal Teknik Informatika (JUTIF)*, *4*(1), 143–150.

Suryawan, I. W. B., Utami, N. W., & Fredlina, K. Q. (2023). Analisis Sentimen Review Wisatawan Pada Objek Wisata Ubud Menggunakan Algoritma Support Vector Machine. *Jurnal Informatika Teknologi dan Sains*, *5*(1), 133–140. https://doi.org/10.51401/jinteks.v5i1.2242

Utami, D. S., & Erfina, A. (2022). Analisis Sentimen Objek Wisata Bali Di Google Maps Menggunakan Algoritma Naive Bayes. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, *6*(1), 418–427.

Wahjoerini, W., Pamurti, A. A., & Prabowo, D. (2022). Pelatihan Pembuatan Visualisasi Data Spasial Bagi Siswa SMA Walisongo Semarang.

SELAPARANG: Jurnal Pengabdian Masyarakat Berkemajuan, *6*(3), 1126. https://doi.org/10.31764/jpmb.v6i3.9331

Wardhana, J. A., & Sibaroni, Y. (2021). Aspect Level Sentiment Analysis on Zoom Cloud Meetings App Review Using LDA. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(4), 631–638. https://doi.org/10.29207/resti.v5i4.3143

Wijayanti, N. P. Y. T., N. Kencana, E., & Sumarjaya, I. W. (2021). SMOTE: Potensi Dan Kekurangnya Pada Survei. *E-Jurnal Matematika*, *10*(4), 235–240. https://doi.org/10.24843/MTK.2021.v10.i04.p348

Yerzi, F. S., & Sibaroni, Y. (2021). Analisis Sentimen Terhadap Kebijakan Pemerintah Dalam Menangani Covid-19 Dengan Pendekatan Lexicon Based. *e-Proceeding of Engineering*, *8*(5), 11354–11366.

Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A Survey of Sentiment Analysis in Social Media. *Knowledge and Information Systems*, *60*(2), 617–663. https://doi.org/10.1007/s10115-018-1236-4

LAMPIRAN

Lampiran A. Surat Permohonan Izin Penelitian

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Dr. Ir. H. Soekarno No. 682, Gunungayar, Surabaya
E-Mail : saintek@uinsby.ac.id Website : http://uinsa.ac.id/fst

UIN SUNAN AMPEL
S U R A B A Y A

Nomor : B - 3170 /Un.07/07/D1/PP.00.9/07/2023
Lampiran : -
Perihal : Validasi Label

Yth,
Eduwisata Lontar Sewu
Desa Hendrosari, Kecamatan Menganti, Kabupaten Gresik, Jawa Timur 61174
Di Tempat

Assalamu'alaikum Wr. Wb.

Sehubungan dengan program peningkatan kompetensi dan ketrampilan mahasiswa pada Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Ampel Surabaya Bidang Penelitian, bersama ini Dekan menyampaikan bahwa mahasiswa dengan identitas sebagai berikut:

Nama	:	MUHAMMAD HANAFI
Tempat, Tgl lahir	:	GRESIK, 5 Mei 2000
NIM	:	H96219053
Semester/Prodi	:	8 / Sistem Informasi
Alamat	:	DESA HENDROSARI

bermaksud melakukan penelitian pada tanggal 11 Juli 2023 dengan judul Analisis Sentimen Destinasi Wisata Kabupaten Gresik Menggunakan Linear Discriminant Analysis (LDA) dan Support Vector Machine (SVM). Oleh karena itu, kami mohon kepada Pimpinan **Eduwisata Lontar Sewu** untuk berkenan memberikan izin, demi kelancaran penelitian yang bersangkutan.

Demikian permohonan izin ini, dan atas kerjasamanya kami sampaikan terimakasih.

Wassalamu'alaikum Wr. Wb.

Surabaya, 10 Juli 2023
An. Dekan
Wakil Dekan Akademik dan Kelembagaan

Dr. Moh. Hafiyusholeh, M.Si., M.PMat.
NIP. 198002042014031001

ISO 9001 2015 CERTIFIED ISO 21001 2018 CERTIFIED

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Dr. Ir. H. Soekarno No. 682, Gunungayar, Surabaya
E-Mail : saintek@uinsby.ac.id Website : http://uinsa.ac.id/fst

Nomor : B - 3215 /Un.07/07/D1/PP.00.9/07/2023
Lampiran : -
Perihal : Validasi Label

Yth,
Wisata Alam GOSARI (WAGOS)
Desa Gosari, Kecamatan Ujungpangkah, Kabupaten Gresik, Jawa Timur 61154
Di Tempat

Assalamu'alaikum Wr. Wb.

Sehubungan dengan program peningkatan kompetensi dan ketrampilan mahasiswa pada Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Ampel Surabaya Bidang Penelitian, bersama ini Dekan menyampaikan bahwa mahasiswa dengan identitas sebagai berikut:

Nama	:	MUHAMMAD HANAFI
Tempat, Tgl lahir	:	GRESIK, 5 Mei 2000
NIM	:	H96219053
Semester/Prodi	:	8 / Sistem Informasi
Alamat	:	DESA HENDROSARI

bermaksud melakukan penelitian pada tanggal 13 Juli 2023 dengan judul Analisis Sentimen Destinasi Wisata Kabupaten Gresik Menggunakan Linear Discriminant Analysis (LDA) dan Support Vector Machine (SVM). Oleh karena itu, kami mohon kepada Pimpinan **Wisata Alam GOSARI (WAGOS)** untuk berkenan memberikan izin, demi kelancaran penelitianyang bersangkutan.

Demikian permohonan izin ini, dan atas kerjasamanya kami sampaikan terimakasih.

Wassalamu'alaikum Wr. Wb.

Surabaya, 12 Juli 2023
An. Dekan
Wakil Dekan Akademik dan Kelembagaan

Dr. Moh. Hafiyusholeh, M.Si., MPMat.
NIP. 198002042014031001

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Dr. Ir. H. Soekarno No. 682, Gunungayar, Surabaya
E-Mail : saintek@uinsby.ac.id Website : http://uinsa.ac.id/fst

Nomor : B - 3216 /Un.07/07/D1/PP.00.9/07/2023
Lampiran : -
Perihal : Validasi Label

Yth,
Wisata Alam SETIGI (SELO TIRTO GIRI)
Desa Sekapuk, Kecamatan Panceng, Kabupaten Gresik, Jawa Timur 61154
Di Tempat

Assalamu'alaikum Wr. Wb.

Sehubungan dengan program peningkatan kompetensi dan ketrampilan mahasiswa pada Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Ampel Surabaya Bidang Penelitian, bersama ini Dekan menyampaikan bahwa mahasiswa dengan identitas sebagai berikut:

Nama	:	MUHAMMAD HANAFI
Tempat, Tgl lahir	:	GRESIK, 5 Mei 2000
NIM	:	H96219053
Semester/Prodi	:	8 / Sistem Informasi
Alamat	:	DESA HENDROSARI

bermaksud melakukan penelitian pada tanggal 13 Juli 2023 dengan judul Analisis Sentimen Destinasi Wisata Kabupaten Gresik Menggunakan Linear Discriminant Analysis (LDA) dan Support Vector Machine (SVM). Oleh karena itu, kami mohon kepada Pimpinan **Wisata Alam SETIGI (SELO TIRTO GIRI)** untuk berkenan memberikan izin, demi kelancaran penelitianyang bersangkutan.

Demikian permohonan izin ini, dan atas kerjasamanya kami sampaikan terimakasih.

Wassalamu'alaikum Wr. Wb.

Surabaya, 12 Juli 2023
An. Dekan
Wakil Dekan Akademik dan Kelembagaan

Dr. Moh. Hafiyusholeh, M.Si., M.PMat.
NIP. 198002042014031001

Lampiran B. Surat Balasan Permohonan Penelitian

PEMERINTAH KABUPATEN GRESIK
KECAMATAN MENGANTI
DESA HENDROSARI
Jl. Protokol Hendrosari-Menganti-Gresik,Telp. (031) 7992439
email : desahendrosari@gmail.com
Kode Pos 61174

Nomor : 145/365/437.111.22/2023
Lampiran : -
Perihal : Surat Balasan Validasi Data

Dengan Hormat,

Berdasarkan Surat dari Universitas Islam Negeri Sunan Ampel Surabaya, Fakultas Sains dan Teknologi Nomor : B-3170/Un.07/07/D1/PP.00.9/07/2023, Perihal Validasi Label Data pengunjung Edu Wisata Lontar Sewu, Desa Hendrosari Kecamatan Menganti Kabupaten Gresik.

Dengan ini kami selaku Pemerintah Desa Hendrosari Kecamatan Menganti Kabupaten Gresik memberikan ijin kepada :

NO	NIM	NAMA	PROGRAM STUDI
1.	H96219053	MUHAMMAD HANAFI	Sistem Informasi

Untuk melaksanakan Penelitian dengan judul **" Analisis Sentimen Destinasi Wisata Kabupaten Gresik Menggunakan *Linear Discriminant Analysis (LDA)* dan *Support Vector Machine (SVM)* "** di Edu Wisata Lontar Sewu, Desa Hendrosari Kecamatan Menganti Kabupaten Gresik.

Demikian Surat ini diberikan dan untuk digunakan sebagaimana mestinya.

Hendrosari, 12 Juli 2023
'a.n. Kepala Desa Hendrosari
Sekretaris Desa

KEPALA DESA HENDROSARI

ARIFIN, ST

Nomor : **0024/WAG/XI/2023**

Perihal : **Surat Balasan Permohonan Izin Penelitian**

Kepada Yth.
Dewan Fakultas Sains Dan Teknologi
UIN Sunan Ampel Surabaya
di
Surabaya

Dengan hormat,

Memperhatikan surat dari Dekan Fakultas Sains dan Teknologi UIN Sunan Ampel nomor: B - 3215 /Un.07/07/D1/PP.00.9/07/2023 tanggal 12 Juli 2023, perihal perizinan tempat penelitian dalam rangka penyusunan skripsi mahasiswa atas nama Muhammad Hanafi dengan judul, Analisis Sentimen Destinasi Wisata Kabupaten Gresik Menggunakan Linear Discriminant Analysis (LDA) dan Support Vector Machine (SVM).

Perlu kami sampaikan beberapa hal sebagai berikut:

1. Pada dasarnya kami tidak keberatan, maka kami dapat mengizinkan pelaksanaan penelitian tersebut di tempat kami.
2. Izin melakukan penelitian diberikan untuk keperluan akademik.
3. Waktu pengambilan data harus dilakukan di waktu hari kerja.

Demikian surat balasan dari kami, atas perhatian kami ucapkan terima kasih.

Gresik, 26 Juli 2023
Ketua Pengelola Wisata
Alam Gosari (WAGOS)

MISBAHUDDAWAM

Wisata Alam Gosari
wagos_gosari
walamgosari@gmail.com
Desa Gosari - Ujungpangkah - Gresik 6115

Badan Usaha Milik Desa
BUMDES SEKAPUK

1. Unit Multijasa
2. Unit PAM
3. Unit Tambang
4. Unit Wisata Setigi
5. Unit kebersihan Desa
6. Unit Sarana & Prasarana Olahraga Terpadu

SURAT KETERANGAN BALASAN PERMOHONAN IZIN PENELITIAN
Nomor: A.5-5.3/435/BUMDes/2023

Yang bertandatangan di bawah ini:

Nama	: Asjudi
Jabatan	: Direktur
Unit Kerja	: Badan Usaha Milik Desa "BUMDES SEKAPUK"

Dengan ini menerangkan dengan sebenarnya bahwa :

Nama	: MUHAMMAD HANAFI
NIM	: H96219053
Program Studi	: Sistem Informasi
Instansi	: UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA

Telah kami setujui untuk mengadakan penelitian di unit usaha BUMDesa Sekapuk dengan judul penelitian "Analisis Sentimen Destinasi Wisata Kabupaten Gresik Menggunakan Linear Discriminant Analysis (LDA) dan Support Vector Machine (SVM)".

Demikian surat keterangan ini dibuat dengan sebenarnya agar dapat digunakan sebagaimana mestinya.

Sekapuk, 24 Juli 2023

Yang Menerangkan,
Direktur BUMDes Sekapuk

BADAN USAHA MILIK DESA BUMDES SEKAPUK UJUNGPANGKAH GRESIK

ASJUDI

Kantor : Jalan Prof Buya Hamka. Graha Artha No 1 Sekapuk Ujungpangkah Gresik Telp . 081216694613
bumdessekapuk@gmail.com @bumdes.sekapuk Bumdes Sekapuk Ujungpangkah
#nyataberdayaguna

Lampiran C. Dokumentasi Pelaksanaan Validasi Pelabelan