

# Penerapan SVM untuk Menganalisis Sentimen Masyarakat Terhadap Wisata Blitar

Vina Febriyani<sup>1</sup>, Harliana<sup>\*2</sup>, Fatra Nonggala Putra<sup>3</sup>, RDR Yusron<sup>4</sup>, Tito Prabowo<sup>5</sup>

<sup>1,2,3,4</sup> Ilmu Komputer, Fakultas Ilmu Eksakta, Universitas Nahdlatul Ulama Blitar

E-mail: <sup>1</sup>vinafeb3007@gmail.com, <sup>\*2</sup>harliana@unublitar.ac.id, <sup>3</sup>putra.fatra08@gmail.com, <sup>4</sup>rdryusron@gmail.com, <sup>5</sup>titoprabowo@unublitar.ac.id

## ARTICLE INFO

### Article history:

Submitted:  
July 13, 2024

Accepted:  
July 18, 2024

Published:  
August 1, 2024

## ABSTRACT

Blitar, known as Patria City and Proclaimer City, has various interesting tourist destinations covering history, culture, nature, culinary and modern tourism. The rapid development of social media has changed the way people get information and express their opinions, so sentiment analysis from social media has become important for understanding public opinion. This research aims to analyze public sentiment towards tourism in Blitar using the SVM algorithm with data obtained from social media Twitter. The research results show that the SVM algorithm is effective in classifying sentiment from tweet data. The classification results produce a positive sentiment class and a negative sentiment class, with the test results of visitor reviews of tourism in Blitar being more dominant with positive sentiment reviews. Apart from that, testing carried out using k-fold cross validation also produced the best accuracy at fold 10

## ABSTRAK

Blitar, yang dikenal dengan sebutan Kota Patria dan Kota Proklamator, memiliki berbagai destinasi wisata yang menarik yang mencakup sejarah, budaya, alam, kuliner, dan wisata modern. Perkembangan pesat media sosial telah mengubah cara masyarakat mendapatkan informasi dan mengekspresikan opini mereka, sehingga analisis sentimen dari media sosial menjadi penting untuk memahami opini publik. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap wisata di Blitar menggunakan algoritma SVM dengan data yang diperoleh dari media sosial Twitter. Hasil penelitian menunjukkan bahwa algoritma SVM efektif dalam mengklasifikasikan sentimen dari data tweet, Hasil klasifikasi menghasilkan kelas sentimen positif dan kelas sentimen negatif, dengan hasil pengujian ulasan pengunjung terhadap wisata di Blitar lebih dominan ke ulasan sentimen bernilai positif. Selain itu pengujian yang dilakukan dengan menggunakan k-fold cross validation juga menghasilkan akurasi terbaik berada pada fold 10

### Keywords:

*Sentiment analysis, SVM, social media, twitter, Blitar tourism*

### Kata Kunci:

Analisis sentimen, SVM, sosial media, twitter, wisata Blitar

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



### Corresponding Author:

Harliana,  
Program Studi Ilmu Komputer, Fakultas Ilmu Eksakta, Universitas  
Nahdlatul Ulama Blitar  
Jalan Masjid No.22 Kauman, Jawa Timur, Indonesia.  
Email: harliana@unublitar.ac.id

## 1. PENDAHULUAN

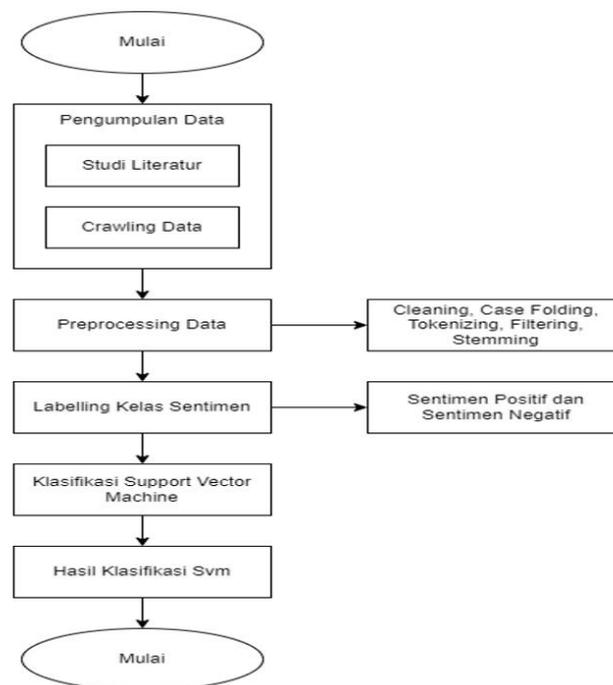
Secara geografis Blitar merupakan salah satu Kota di Jawa Timur yang berada di bawah kaki Gunung Kelud yang berbatasan langsung dengan Kabupaten Kediri, Kabupaten Malang, serta Kabupaten Tulungagung[1]. Dari sisi udara, Kota Blitar memiliki suhu udara yang relatif lebih sejuk dibandingkan kota lainnya karena memiliki RTH (Ruang Terbuka Hijau) sekitar 18% atau sekitar 561,473Ha dari total wilayahnya[2]. Menurut data DISPARBUDPORA, saat ini Kabupaten dan Kota Blitar memiliki kurang lebih 71 obyek wisata yang terbagi menjadi: wisata kuliner yang berada di jalan lawu[3]; wisata cagar budaya seperti Kebun Kopi De Karanganyar dan Candi Penataran[4]; wisata sejarah seperti goa umbul tok[5], bukit pertapaan[6], istana gebang dengan segala tradisinya[7]; wisata alam seperti pantai gondo mayit[8], pantai peh pulo, sumur amber [9], sirah kencong[10] serta beberapa tempat wisata yang instagramable sekaligus edukatif seperti Blitar Park[11] dan kampung coklat[12]. TIC (*Tourism Information Center*) dan BAPPARDA merupakan salah satu fasilitas yang dimiliki oleh DISPARBUDPORA untuk mempromosikan obyek wisata yang ada di Blitar melalui media instagram. Namun sayangnya kedua strategi ini masih dirasakan kurang maksimal karena tidak dapat memberikan *feedback* dari masyarakat yang telah berkunjung ke destinasi wisata yang ditampilkan. Berbeda halnya ketika pengguna menceritakan pengalamannya saat berkunjung ke destinasi wisata menggunakan media sosial *twitter*, dimana pengguna lain dapat memberikan responnya melalui pendapat, saran, opini, maupun kritikan mengenai topik yang sedang dibahas[13]. Melalui *tweet-tweet* inilah pengguna lain dapat mengetahui sentimen yang muncul mengenai topik yang sedang dibahas apakah bersifat negatif ataukah positif.

Berdasarkan hal tersebut maka akan dilakukan analisis sentimen masyarakat terhadap destinasi wisata yang berada di Blitar dengan Algoritma *Support Vektor Machine* (SVM) melalui media sosial *twitter*. Respon yang realtime, cepatnya informasi yang tersebar[14] serta adanya fitur trending topik menjadi salah satu alasan peneliti menggunakan *twitter* daripada media sosial lain. Selain itu *twitter* juga mampu menghasilkan 383juta *tweet* perhari dan menjadi media sosial kelima yang paling banyak dikunjungi di tahun 2022[14]. Adapun tujuan dari penelitian ini yaitu mengetahui akurasi yang dihasilkan oleh Algoritma SVM dalam menilai persepsi masyarakat terhadap destinasi wisata di Blitar melalui *tweet-tweet* yang dituliskan di *twitter*. Selanjutnya analisis sentimen ini hanya akan dikategorikan menjadi 2 yaitu, sentimen positif dan sentimen negatif. Penerapan media sosial *twitter* untuk sentimen analisis sudah banyak dilakukan guna melihat respon masyarakat, salah satunya terkait isu politik terhadap pelaksanaan Pemilu saat pandemi Tahun 2020[15], Pemilu Tahun 2019[16], Pemilu Tahun 2024[17], serta kinerja KPK tahun 2018 sampai dengan 2019[18]. Dalam bidang pendidikan sentimen analisis juga pernah dilakukan untuk mengetahui respon masyarakat terkait pelaksanaan sekolah daring selama pandemi[19], kegiatan kurikulum merdeka belajar tingkat SD dan SMP[20], serta penerapan kebijakan zonasi sekolah bagi calon siswa baru[21]. Sedangkan pada bidang pariwisata, sentimen analisis juga pernah dilakukan untuk melihat respon masyarakat terhadap wisata pantai di Kabupaten Karawang dengan dataset yang digunakan adalah ulasan dari Google Map melalui pendekatan Naive Bayes[22], serta ulasan obyek wisata di Bali dengan Algoritma Naive Bayes berdasarkan hasil review Wisatawan di Google Map[23]. Berbeda dengan penelitian terdahulu, pada penelitian ini akan menggunakan Algoritma SVM karena algoritma ini mampu menghasilkan akurasi yang tinggi yaitu 89,58% melalui pengujian cross validation dengan nilai  $k=8$ [24], selain itu SVM juga memiliki kecenderungan tingkat akurasi yang lebih besar bila dibandingkan dengan algoritma Naive Bayes dan KNN meskipun sudah dilakukan optimasi dengan Algoritma PSO pada ketiganya[25]. Perhitungan konsep klasifikasi yang jelas dan lengkap secara matematis pada Algoritma SVM[26] juga menjadi alasan tersendiri bagi penulis dalam memilih Algoritma ini

## 2. METODE PENELITIAN

Gambar 1 merupakan gambaran mengenai penelitian yang akan dilakukan. Awal penelitian akan dimulai dengan melakukan pengumpulan data melalui studi literatur dan *crawling* data. Seperti penelitian lainnya studi literatur akan dilakukan dengan mempelajari berbagai literatur mengenai Algoritma SVM serta bagaimana mengolah dan menganalisis opini masyarakat menjadi sentimen yang positif ataupun negatif berdasarkan dataset yang didapatkan melalui suatu jurnal. Kemudian akan dilanjutkan dengan melakukan *crawling* data dengan mengambil data dari *twitter*. Selanjutnya akan dilakukan *preprocessing* data yaitu memilih teks pada *tweet* yang didapatkan guna mengurangi jumlah kata yang tidak diperlukan untuk tahap analisis. Pada tahapan *preprocessing*, akan dilakukan tahapan lain seperti *cleaning*, *case folding*, *tokenizing*, *filtering* dan *stemming*. Pada tahap *cleaning dataset* akan dimasukkan kedalam sistem kemudian sistem akan melakukan proses pembersihan dokumen (*cleaning*) guna mengurangi *noise* terhadap karakter dan kata yang tidak diperlukan. Pada tahap *case folding*, sistem akan merubah teks pada dokumen menjadi huruf kecil guna mempermudah analisis. Selanjutnya akan dilakukan *tokenizing* yaitu merubah teks menjadi kata ataupun frasa yang lebih kecil yang dibatasi dengan titik ataupun titik koma. Selanjutnya akan dilakukan proses *filtering* yaitu menghapus duplikat dari data yang tidak relevan. Dan yang terakhir dalam proses *preprocessing* yaitu proses *stemming* yaitu merubah kata dalam teks menjadi kata dasar yang sama.

Langkah selanjutnya yaitu proses *labelling* terhadap kelas yang terbentuk. Seperti yang telah dijabarkan diatas bahwa penelitian ini hanya akan membentuk sentimen positif (ketika nilai yang dihasilkan  $> 0$ ) dan sentimen negatif (dengan nilai yang dihasilkan  $< 0$ ). Selanjutnya akan dilakukan proses klasifikasi dengan Algoritma SVM terhadap dataset dokumen yang telah dihasilkan. Dataset yang diklasifikasi akan dibagi menjadi 2 yaitu data latih dan data uji dimana rasio perbandingan keduanya adalah 80:20. Hal ini berarti 80% dari total dataset akan digunakan sebagai data latih dan 20% dari total dataset akan digunakan sebagai data uji. Yang perlu diperhatikan sebelum melakukan proses klasifikasi dengan SVM adalah bahwa data harus sudah melalui proses labelling dan dirubah menjadi bentuk numerik agar sistem dapat memprosesnya. Selanjutnya akan dilakukan pengujian dan evaluasi terhadap hasil klasifikasi yang didapatkan. Pengujian akan dilakukan melalui teknik validasi K-Fold Cross validation dengan nilai k 3, 5, 7, dan 10.



Gambar 1. Tahapan penelitian yang dilakukan

### 3. HASIL DAN PEMBAHASAN

Mengacu pada Gambar 1, maka pengumpulan data pada penelitian ini akan dimulai dengan crawling data dengan menggunakan python dan google colab, dimana dataset yang digunakan dalam penelitian ini berasal dari 972 tweet dengan kata kunci "wisata blitar ", "wisata kota blitar ", "wisata kabupaten blitar " dan "Destinasi wisata blitar ". Dataset dikumpulkan dari awal tahun 2020 sampai dengan januari 2024. Gambar 2 merupakan hasil crawling data yang tersimpan dengan format csv.

A	B	C	D	E	F	G	H	I	J	K	L	M
1	created_at_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id	si_conversat	username	tweet_url	
2	Fri Apr 07 1,64E+18	@noisette_douce Malang wkwk... Banyak wisata di situ, makanan jg enak2. Terus ke Madiun jg recommended.	0	2	0	0	1	in	1,01E+18	1,64E+18	wanitassa	https://twitter.com/war
3	Fri Apr 07 1,64E+18	Hutan Pinus Loji Blitar, Wisata Sejuk Alami di Lereng Gunung Kelud https://t.co/w55U7H47 https://t.co/YqWQ	0	0	0	0	in	7,78E+17	1,64E+18	TheLuxury	https://twitter.com/The	
4	Fri Apr 07 1,64E+18	3. Bagus lagi kalo ada trans jatim yg antar kota/kabupaten dalam satu karesidenan kediri raya ini, misal kedi	0	1	0	0	in	1,18E+18	1,64E+18	kemrunyu	https://twitter.com/kem	
5	Fri Apr 07 1,64E+18	3. Bagus lagi kalo ada trans jatim yg antar kota/kabupaten dalam satu karesidenan kediri raya ini, misal kedi	0	1	0	0	in	1,18E+18	1,64E+18	kemrunyu	https://twitter.com/kem	
6	Fri Apr 07 1,64E+18	3. Bagus lagi kalo ada trans jatim yg antar kota/kabupaten dalam satu karesidenan kediri raya ini, misal kedi	0	1	0	0	in	1,18E+18	1,64E+18	kemrunyu	https://twitter.com/kem	
7	Fri Apr 14 1,65E+18	35 Tempat Wisata Blitar Menarik Dikunjungi Saat Liburan! https://t.co/h6Oo6bPIOH	0	0	0	0	in	2,4E+08	1,65E+18	detikTravel	https://twitter.com/det	
8	Fri Apr 14 1,65E+18	#35 Tempat Wisata Blitar Menarik Dikunjungi Saat Liburan! - detikTravel #Colomadu #Karanganyar https://t.c	0	0	0	0	in	3,3E+09	1,65E+18	Colomadu	https://twitter.com/Col	
9	Fri Apr 14 1,65E+18	Fish Garden Blitar, Objek Wisata dengan Berbagai Spot Bertema Ikan https://t.co/DYALrHrKX https://t.co/AbYU	0	0	0	0	in	7,78E+17	1,65E+18	TheLuxury	https://twitter.com/The	
10	Fri Apr 15 1,51E+18	Ketua Dewan Perwakilan Rakyat Daerah (DPRD) Kabupaten Blitar Suwito Saren Saboto meminta Pemerintah Kal	0	1	0	0	in	1,5E+08	1,51E+18	PDI_Perju	https://twitter.com/PDI	
11	Fri Apr 21 1,65E+18	Anggota Satgas KamselTibcarlantas Operasi Ketupat Semeru 2023 Satsabhara Polres Blitar melaksanakan live	0	0	0	0	in	1,25E+18	1,65E+18	blitarlanti	https://twitter.com/blit	
12	Fri Apr 28 1,65E+18	Wisata Kampung Coklat di Blitar Alami Kebakaran Dahsyat https://t.co/Ox89PEI645	0	0	0	0	in	1,17E+18	1,65E+18	IdHeadlin	https://twitter.com/IdH	
13	Fri Apr 28 1,65E+18	Laporan situasi di tempat Wisata Blitar Park oleh petugas pos pengamanan terpadu Polres Blitar #polisindone	0	0	0	1	in	2,4E+09	1,65E+18	HUMASRei	https://twitter.com/HUM	
14	Fri Apr 28 1,65E+18	Giat pengamanan kawasan wisata pantai pangsi ds.tumpakkepuh kec. Bakung kab. Blitar https://t.co/AkVA1q85	0	0	0	1	in	1,65E+18	1,65E+18	sebakunj	https://twitter.com/sek	
15	Fri Apr 29 1,52E+18	8 Wisata Menarik di Kabupaten Blitar, Cocok Dikunjungi Saat Libur Lebaran https://t.co/dGCSGvZgfw	0	0	0	0	in	1,50E+18	1,52E+18	KlikTimes	https://twitter.com/Klik	
16	Fri Aug 18 1,69E+18	Jumat (18/8/2023) bertempat di Wisata Kesambi Trees Park, Kapolres Blitar Kota AKBP DANANG SETIYO P. S., S.	0	0	0	0	in	1,55E+18	1,69E+18	PolsekPor	https://twitter.com/Pol	
17	Fri Aug 18 1,69E+18	Jumat (18/8/2023) bertempat di Wisata Kesambi Trees Park, Kapolres Blitar Kota AKBP DANANG SETIYO P. S., S.	0	0	0	0	in	1,55E+18	1,69E+18	PolsekPor	https://twitter.com/Pol	
18	Fri Aug 25 1,69E+18	Polsek Sananwetan Polres Blitar Kota, Melaksanakan kegiatan Jumat Curhat Bersama Masyarakat yang tergab	0	0	0	1	in	1,49E+18	1,69E+18	Polsek_Sa	https://twitter.com/Pol	
19	Fri Aug 25 1,69E+18	Polsek Sananwetan Polres Blitar Kota, Melaksanakan kegiatan Jumat Curhat Bersama Masyarakat yang tergab	0	0	0	1	in	1,49E+18	1,69E+18	Polsek_Sa	https://twitter.com/Pol	
20	Fri Dec 01 1,73E+18	Ke Blitar yuk, Wisata Edukasi Candi, Belajar Kisah Ramayana dan Krisnayana Dari Relief yang Ada Di Candi Pi	0	0	0	0	in	1,51E+18	1,73E+18	terpantau	https://twitter.com/terp	
21	Fri Dec 01 1,73E+18	Ke Blitar yuk, Wisata Edukasi Mengunjungi Candi Kotes Atau Candi Papoh I dan II. Serasa Kembali ke Zaman N	0	0	0	0	in	1,51E+18	1,73E+18	terpantau	https://twitter.com/terp	
22	Fri Dec 01 1,73E+18	Utas singkat terkait wisata sejarah ke 'Museum Bung Karno', Blitar, Jawa Timur, Jumat (1/12). Let's go!	0	1	0	1	in	8,51E+17	1,73E+18	olic_27	https://twitter.com/olic	
23	Fri Dec 01 1,73E+18	Pasca ziarah wisata, waktu uduh mau dzuhur, waktunya sholat.Jumat. Di sini lah, Masjid Ar-Rahman, Kota Blita	0	1	0	0	in	8,51E+17	1,73E+18	olic_27	https://twitter.com/olic	
24	Fri Dec 01 1,73E+18	Pasca ziarah wisata, waktu uduh mau dzuhur, waktunya sholat.Jumat. Di sini lah, Masjid Ar-Rahman, Kota Blita	0	1	0	0	in	8,51E+17	1,73E+18	olic_27	https://twitter.com/olic	

Gambar 2. Hasil crawling data

Hasil *crawling* data terdiri dari 12 kolom dengan *created\_at* adalah waktu posting tweet oleh pengguna, *id\_str* adalah nomor *id twitter*, *full\_text* adalah isi tweet, *language* adalah bahasa yang digunakan, *user\_id* adalah nomor *id* pengguna tweet, *username* adalah nama pengguna twett, dan *tweet\_url* adalah link tweet. Selanjutnya data tersebut akan digunakan untuk proses preprocessing data pertama yaitu cleaning dimana pada tahapan ini akan dilakukan penghapusan data dan atribut yang dianggap tidak penting seperti hastag, mention, symbol, emoticon, angka, tanggal dan tautan link. Table 1 adalah hasil dari proses cleaning ini.

Table 1. Hasil proses *cleaning*

Sebelum	Sesudah
Wisata Alam Air Terjun Tirto Galuh Kabupaten Blitar Hidden Paradise #TempatWisata#belajarbahasainggris #visitpare https://t.co/XDvtvpTOLy	Wisata Alam Air Terjun Tirto Galuh Kabupaten Blitar Hidden Paradise
Blitar Park, Wisata Ramah Anak dengan Wahana Permainan Seru, Cocok untuk Liburan Akhir Pekan https://t.co/WMAtJoxgUo	Blitar Park, Wisata Ramah Anak dengan Wahana Permainan Seru, Cocok untuk Liburan Akhir Pekan

Selanjutnya akan dilakukan proses case folding, dimana hasil dari proses ini menggantikan kolom *full\_text* atau dataset asli dengan huruf kecil melalui perintah persamaan (1).

$$\text{text\_df}[\text{'full\_text'}] = \text{text\_df}[\text{'full\_text'}].\text{str}.\text{lower}() \quad \dots \text{pers (1)}$$

Pada proses tokenizing, teks yang dihasilkan tadi akan dipecah menjadi sebuah kata / frasa yang bertujuan untuk mencari makna dari kata tersebut. Hasil dari proses tokenizing tersebut terdapat pada tabel 2.

Table 2. Hasil proses *tokenizing*

Sebelum	Sesudah
wisata alam air terjun tirto galuh kabupaten blitar hidden paradise	['wisata', 'alam', 'air', 'terjun', 'tirto', 'galuh', 'kabupaten', 'blitar', 'hidden', 'paradise']
blitar park wisata ramah anak dengan wahana permainan seru	['blitar', 'park', 'wisata', 'ramah', 'anak', 'dengan', 'wahana', 'permainan', 'seru']

cocok untuk liburan akhir pekan

'permainan', 'seru', 'cocok', 'untuk', 'liburan', 'akhir', 'pekan']

Tahapan selanjutnya yaitu proses filtering, dimana sebelum melakukan filtering akan dilakukan pemilihan dataset yang memiliki kesamaan antar kolom atau duplikat dataset, maka dari itu akan dilakukan proses duplikat dataset dengan source code persamaan (2) sehingga menghasilkan dataset 648 data ulasan. Proses filtering ini akan menggunakan library NLTK untuk mendapatkan stopword yang tidak perlu tersebut seperti "dan", "yang", "ini", "itu" dan sebagainya.

```
text_df.drop_duplicates(subset="full_text", keep = 'first', inplace = true) ... pers (2)
```

Tahapan stemming, akan dilakukan perubahan kata menjadi kata dasar dengan menghilangkan imbuhan (awalan, sisipan dan akhiran) dengan bantuan *library* sastrawi. Selanjutnya akan dilakukan proses labelling data dengan Lexicon, dimana jumlah kata yang ditemukan dalam kamus lexicon akan menentukan skor yang diberikan kepada kata-kata negatif dan positif[27]. Jumlah dataset yang akan diproses pada labelling berkurang menjadi 648 dataset, hal ini disebabkan karena selama preprocessing, data yang terduplikasi, tweet tidak relevan, tweet yang tidak memiliki makna, akan di hapus untuk menghindari bias dalam analisis atau proses selanjutnya. Meskipun jumlah data berkurang, kualitas dan relevansi data yang tersisa akan meningkat dan memberikan hasil analisis yang lebih akurat. Hasil dari proses labelling ini terdapat pada Tabel 3.

Table 3. Hasil proses *labelling*

Tweet	Labelling
wisata elit sambang pantai gondo mayit blitar	Positif
wisata air terjun lawean kawisari hidden gems blitar cantik tawan hati	Negatif
jalur wisata kota blitar rawan macet polres blitar laku rekayasa lintas	Negatif
pariwisata bangkit bulan kunjung wisata kabupaten blitar capai ribu unjung	Negatif
wisata alam kebun teh sirah kencong blitar	Positif
kompleks wisata budaya candi natar blitar	Negatif

Berdasarkan Tabel 3, maka ada beberapa alasan mengapa setiap tweet diberi label positif atau negatif:

- Tweet: "wisata elit sambang pantai gondo mayit blitar" Labeling: Positif  
Alasan: Tweet ini menunjukkan bahwa pengalaman dari wisatawan terhadap wisata tersebut yang berkualitas dan menyenangkan.
- Tweet: "jalur wisata kota blitar rawan macet polres blitar laku rekayasa lintas" Labeling: Negatif  
Alasan: Tweet ini mencakup keluhan dari wisatawan terhadap jalur wisata kota Blitar yang menimbulkan kemacetan sehingga mengganggu kenyamanan pengunjung atau wisatawan.
- Tweet: "wisata alam kebun teh sirah kencong blitar" Labeling: Positif  
Alasan: Tweet ini menunjukkan bahwa tempat wisata memiliki keindahan alam yang asri dan menyegarkan.

Berdasarkan pelabelan data dari 648 data ulasan menghasilkan 357(55.09%) data ulasan untuk sentimen positif, dan sentimen negatif sebanyak 291(44.91%) data ulasan.

### Proses SVM

Pada tahap ini, proses klasifikasi dokumen akan dimulai dengan data dari hasil pembobotan *Lexicon* yang telah dilakukan sebelumnya. Untuk memulai proses ini, metode *Support Vector Machine* membutuhkan fitur untuk proses datanya. Setelah melakukan proses labelling data dengan menggunakan *Lexicon*, data akan dibagi menjadi data latih dan data uji yaitu 20% data uji dan 80% data latih. Setelah melakukan proses *split* data dengan pembagian data uji 20% dan data latih 80% maka diperoleh data 518 untuk data latih dan 130 untuk data uji. Pembagian data uji dan data latih merupakan langkah penting dalam proses pelatihan model *Machine Learning*. Tujuannya adalah membagi kumpulan data menjadi dua dataset, dataset untuk melatih model data (data latih) dan satu dataset lagi untuk menguji model (data uji). Untuk menilai kinerja model pada data yang belum pernah dilihat sebelumnya, pembagian ini dapat memberikan gambaran yang lebih realistis tentang bagaimana performa model berfungsi pada data baru. Tabel 4 data data latih yang akan digunakan.

Table 4. Data latih

No	Tweet
1	istana sakura wisata hits nuansa jepang blitar
2	destinasi wisata air terjun blitar segar jiwa
3	jalur wisata kota blitar rawan macet polres blitar laku rekayasa lintas
4	area wisata kampung coklat kabupaten blitar hangus lalap api bakar akibat korsleting listrik ruang produksi tugas butuh jam takluk api

- 
- 5 blitar masjid arrahman salah wisata religi kota blitar wajib kunjung blitar kalah masjid andal surabaya  
 6 wisata populer blitar jawa timur recommended libur
- 

Setiap baris pada Tabel 4 berisi teks ulasan atau komentar yang berisi berbagai topik tentang wisata Blitar, seperti pengalaman wisatawan, informasi tentang lokasi, serta masalah dan solusi yang terkait dengan wisata tersebut. Data ini mencakup berbagai topik dan sentimen yang memberikan gambaran menyeluruh tentang pengalaman dan pandangan wisatawan serta aspek lain yang relevan dengan wisata Blitar. Data ini digunakan untuk melatih model pembelajaran mesin atau analisis teks untuk memahami, mengklasifikasikan, atau menganalisis informasi terkait wisata di wilayah Blitar. Sedangkan untuk data uji terdapat pada Tabel 5.

Table 4. Data latih

No	Tweet
1	pantai tambakrejo wisata pantai blitar selatan jawa timur
2	wisata bukit teletubbies blitar hits teletubbiesnya
3	wisata asik camping hutan pinus loji blitar
4	puncak kawan wisata alam blitar instagenic adem tenang
5	goa kedungkrombang salah destinasi wisata alam blitar wajib kunjung
6	indah wisata konservasi tangkar rusa maliran blitar

Dataset yang telah melalui proses pembagian data akan melakukan proses klasifikasi menggunakan algoritma *Support Vector Machine*. Pengujian data dengan klasifikasi SVM (*Support Vector Machine*) adalah langkah penting dalam proses Machine Learning untuk menilai kinerja model data yang telah dilatih. Penelitian ini menggunakan bahasa pemrograman Python, yang memiliki *library Sklearn* yang dapat digunakan untuk mengimplementasikan SVM. Pada pengujian data akan dilakukan proses mengubah data teks menjadi representasi numerik dengan menggunakan fungsi *CountVectorizer* dalam bentuk matriks array. Tujuan vektorisasi agar model masih dapat memahami setiap keterkaitan kata dalam kalimat. Output yang dihasilkan oleh proses fungsi *CountVectorizer* berupa array dua dimensi. Nilai dari matriks vektor tersebut adalah frekuensi setiap teks yang muncul. Token atau kata yang sebelumnya mempunyai nilai vektor akan dirata-rata berdasarkan kata pada kalimat tersebut. Jika token tidak ada dalam model kosakata, nilai vektor tidak akan ditampilkan dan diabaikan.

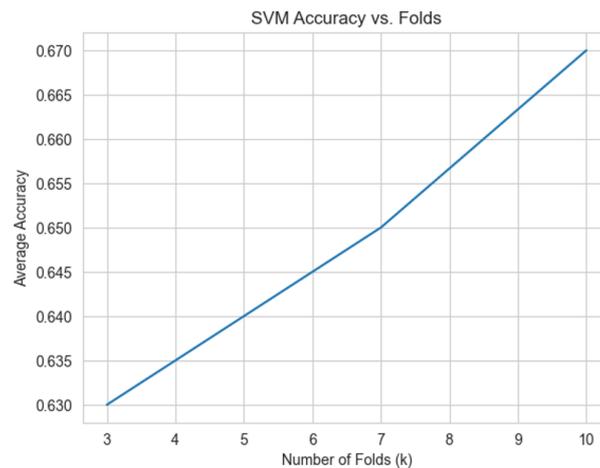
### Hasil Analisis Sentimen

Pada proses pengujian k-fold cross validation. Data training akan diacak menjadi beberapa bagian sesuai dengan nilai fold yang diberikan yaitu nilai 3, 5, 7, dan 10, yang terlihat pada table 5.

Table 5. Parameter SVM dengan evaluasi k-fold *cross validation*

Nilai K-Fold	Hasil Akurasi
K=3	0.6370367880987587
K=5	0.6428117998506347
K=7	0.6544401544401544
K=10	0.6736048265460031

Berdasarkan Tabel 5, diketahui bahwa nilai fold 10 akan menghasilkan nilai akurasi terbaik bila dibandingkan dengan fold 3, 5, dan 7. Hal ini disebabkan karena proses pengujian diulang 10 kali untuk mendapatkan validasi yang akurat, sehingga dapat membantu dan mengurangi variabel dalam estimasi performa model dan menghasilkan estimasi yang lebih stabil dan akurat bila dibandingkan dengan fold yang lebih sedikit. Selain itu nilai fold yang semakin meningkat juga akan menghasilkan nilai akurasi yang dihasilkan. Penggunaan nilai fold yang terlalu sedikit, seperti 2 atau 5, dapat menghasilkan estimasi varians yang sangat bias dan rendah yang mungkin tidak mencerminkan performa model. Dimana hasil pengujian ini terlihat pada gambar 3. Berdasarkan hal tersebut maka pemilihan validasi silang sebanyak 10 kali lipat untuk menguji data adalah yang terbaik karena dapat memberikan estimasi performa yang lebih handal dan stabil sekaligus memastikan bahwa model akan diuji secara menyeluruh pada kumpulan data yang berbeda.



Gambar 3. Hasil akurasi SVM dengan fold yang berbeda

Selanjutnya untuk mengetahui kata-kata yang paling sering digunakan dari dataset yang digunakan, maka akan digunakan WordCloud untuk mengetahui sentimen negatif dan positif terkait wisata Blitar. Hasil dari WordCloud ini ditunjukkan pada gambar 5



Gambar 5. Hasil WordCloud sentiment negatif

Berdasarkan gambar 5 maka kata kunci dengan sentimen negatif dari teks yang didapatkan yaitu:

- “Wisata”: kata kunci ini muncul secara terus menerus, yang menunjukkan bahwa banyak percakapan atau ulasan negatif tentang wisata di Blitar.
- “Blitar”: nama daerah yang sering muncul, sehingga menjadi fokus dari berbagai masalah negatif.
- “Pantau Arus”: kata ini menunjukkan bahwa orang berbicara tentang memantau arus lalu lintas.
- “Lantas”: kata ini menunjukkan bahwa masalah lalu lintas sering menjadi perhatian.

Sedangkan untuk beberapa kata yang paling banyak digunakan dalam visualisasi wordcloud sentimen positif adalah “wisata”, “pantai”, “blitar”, “desa”, “kampung”, dan lainnya. Visualisasi ini menunjukkan bahwa tempat wisata Blitar terkenal dengan wisata pantai yang berada didesa atau perkampungan yang berada di daerah Blitar. Gambar 6 merupakan tampilan dari hasil WordCloud sentiment positif yang didapatkan.



Gambar 6. Hasil WordCloud sentiment positif

Berdasarkan Gambar 6 maka, beberapa kata yang termasuk sentiment positif adalah:

- a. Dominasi kata “Blitar” dan “Wisata”: kata-kata “Blitar dan “Wisata” kembali muncul dalam jumlah yang signifikan, menandakan bahwa kedua kata tersebut menjadi elemen utama dalam teks yang memiliki sentimen positif, hal ini menunjukkan bahwa banyak review atau komentar positif mengenai wisata Blitar.
- b. Tempat wisata populer: kata-kata seperti “Kampung Coklat”, “Pantai Blitar”, “Pantai Serang”, Dan “Makam Karno” ditampilkan dengan ukuran yang besar, menunjukkan bahwa tempat-tempat tersebut sering dalam konteks positif.
- c. Jenis Wisata dan Aktivitas: berbagai jenis wisata di Blitar yang mendapatkan ulasan positif dari pengunjung digambarkan dengan istilah seperti “wisata edukasi”, “wisata alam”, wisata religius” dan “wisata kampung”. Selain itu, aktivitas seperti "kunjung", "libur", "bakar", dan "tarik" juga dominan, sehingga menunjukkan bahwa aktivitas tersebut populer dikalangan wisatawan.
- d. Kontribusi dari Polres dan Satlantas: munculnya kata-kata seperti “polres blitar” dan “satlantas menunjukan apresiasi terhadap keamanan dan manajemen lalu lintas yang mendukung pengalaman pengunjung yang positif.
- e. Kata-kata yang berkaitan dengan lokasi: Kata-kata seperti "kabupaten blitar", "kota blitar", "jawa timur", "lokasi", dan "jalan" menunjukkan bahwa ulasan positif tidak hanya berfokus pada tempat wisata tetapi juga wilayah yang lebih luas dan infrastruktur yang mendukung.
- f. Aspek sosial dan keluarga: Kata "keluarga" muncul dengan cukup banyak, menunjukkan bahwa Blitar adalah tempat yang bagus untuk dikunjungi bersama keluarga. Kata "masyarakat" juga muncul, menunjukkan bahwa orang dapat memiliki hubungan yang baik dengan orang-orang di sekitarnya.
- g. Atribut yang Positif: Di Blitar, kata-kata seperti "baik", "bagus", "cocok", "populer", dan "rekomendasi" sering dikaitkan dengan hal-hal yang baik.

*Word cloud* ini menunjukkan bahwa pengunjung melihat tempat wisata Blitar dengan baik. Kata-kata yang muncul menunjukkan tempat-tempat yang populer, aktivitas yang disukai, dan fitur keamanan dan infrastruktur yang mendukung kunjungan yang menyenangkan.

#### 4. KESIMPULAN

Berdasarkan analisis sentimen wisata pada media sosial twitter menggunakan SVM didapatkan beberapa kesimpulan bahwa model klasifikasi SVM dapat digunakan untuk mengklasifikasi sentimen wisata Blitar pada ulasan Twitter dengan menggunakan 648 tweet, yang terdiri dari 357 ulasan positif dan 291 ulasan negatif. Hasil klasifikasi menghasilkan kelas sentimen positif dan kelas sentimen negatif, dengan hasil pengujian ulasan pengunjung terhadap wisata di Blitar lebih dominan ke ulasan sentimen bernilai positif. Selain itu pengujian yang dilakukan dengan menggunakan k-fold cross validation juga menghasilkan akurasi terbaik berada pada fold 10 dengan nilai akurasi 0.6736048265460031. Hasil analisis sentimen menunjukkan bahwa sebagian besar pengunjung yang menulis *tweet* dengan *keyword* “wisata blitar” akan memberikan ulasan yang cenderung baik mengenai wisata yang berada di Blitar.

#### DAFTAR PUSTAKA

- [1] T. Widodo, “Hubungan Tutupan Lahan Terhadap ketersediaan Air di Kecamatan Kanigoro Kabupaten Blitar,” *Briliant J. Ris. dan Konseptual*, vol. 5, no. 4, p. 851, 2020, doi: 10.28926/briliant.v5i4.535.
- [2] D. Vebri Rukmana, D. Kunto Nurkukuh, and D. Wismoro, “Efektivitas Fungsi Ekologis Taman Kota Blitar Berdasarkan Persepsi Masyarakat,” *Matra*, vol. 1, no. 1, pp. 94–104, 2020.
- [3] M. Billah, S. Rusmawati, and A. R. Utomo, “Pemberdayaan Masyarakat Melalui Destination Branding Kawasan Wisata Kuliner Lawu Kauman Kota Blitar,” *Citiz. J. Ilm. Multidisiplin Indones.*, vol. 2, no. 3, pp. 444–452, 2022, doi: 10.53866/jimi.v2i3.122.
- [4] S. Sugiarto, E. Triandini, P. F. Nuryananda, A. F. E. Yonce, C. A. Putra, and A. M. Rizki, “Pendampingan Pemetaan dan Pengemasan Wisata Cagar Budaya di Kabupaten Blitar,” *J. Pengabd. Multidisiplin*, vol. 2, no. 3, pp. 74–78, 2022, doi: 10.51214/japamul.v2i3.299.
- [5] R. E. Wulandari, Y. Yatmin, and A. Budiarto, “Goa Umbul Tuk Sebagai Tempat Wisata Bersejarah di Blitar Selatan,” in *Prosiding SEMDIKJAR*, 2022, vol. 5, pp. 709–715.
- [6] A. N. Annisa *et al.*, “BUDIKDAMBER Sebagai Upaya Pengembangan Potensi Ekowisata Bukit Pertapaan, Desa Bagelenan, Kabupaten Blitar,” *Media Gizi Kemas*, vol. 11, no. 2, pp. 8–9, 2022.
- [7] S. Febrianti Rahayu, “Makna Grebeg Pancasila Sebagai Upaya Pelestarian Nilai Pancasila di Kota Blitar,” *Sumbula J. Stud. Keagamaan, Sos. dan Budaya*, vol. 7, no. 1, pp. 169–179, 2022, doi:

- 10.32492/sumbula.v7i1.4877.
- [8] A. Zunaidi *et al.*, “Upaya Menambah Daya Tarik Objek Wisata Melalui Rancangan Spot Foto Pantai Pasetran Gondo Mayit Blitar,” *J. Ilm. Pangabdhi*, vol. 8, no. 2, pp. 81–86, 2022, doi: 10.21107/pangabdhi.v8i2.16550.
- [9] A. Z. Nisak, N. A. Rohmah, and R. M. I. Indana, “Pengelolaan Obyek Wisata Sumur Amber di Desa Kandangan Kec Srengat Kab Blitar,” *Panor. J. Kaji. Pariwisata*, vol. 1, no. 2, 2024, doi: 10.34833/panorama.v1i2.3016.
- [10] R. W. Apsari, E. N. Billah, and N. Insani, “Dampak Covid-19 Terhadap Pengelolaan Agrowisata Perkebunan Teh Sirah Kencong Kabupaten Blitar sebagai Obyek Wisata Berkelanjutan,” *EDUTOURISM J. Tour. Res.*, vol. 2, no. 02, pp. 61–72, 1970, doi: 10.53050/ejtr.v2i02.139.
- [11] D. W. S. Rahayu and A. D. S. Hamidah, “Analisis Daya Tarik Atmosfer Instagramable untuk Menarik Minat Kunjung Kembali Wisatawan di Blitar Park,” *J. Simki Econ.*, vol. 7, no. 1, pp. 31–40, 2024, doi: 10.29407/jse.v7i1.480.
- [12] A. Andriani and T. Sa’adah, “Peran Strategi Diferensiasi dalam Meningkatkan Jumlah Wisatawan (Studi Kasus Kampung Coklat Blitar),” *Al-Muraqabah J. Manag. Sharia Bus.*, vol. 1, no. 1, pp. 65–78, 2021, doi: 10.30762/almuraqabah.v1i1.145.
- [13] A. Putra and R. Latifah, “Analisis Sentimen Pengguna Twitter Terhadap Aplikasi Pinjaman Online Menggunakan Metode Support Vector Machine,” *Semin. Nas. Penelit. LPPM UMJ*, pp. 1–7, 2022.
- [14] N. A. Salsabila, “Analisis Sentimen Pada Media Sosial Twitter Terhadap Tokoh Gus Dur Menggunakan Metode Naive Bayes dan SVM,” Universitas Islam Negeri Syarif Hidayatullah, 2022.
- [15] A. Muzaki and A. Witanti, “Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive Bayes Classifier,” *J. Tek. Inform.*, vol. 2, no. 2, pp. 101–107, 2021, doi: 10.20884/1.jutif.2021.2.2.51.
- [16] S. Juanita, “Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes,” *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 552, 2020, doi: 10.30865/mib.v4i3.2140.
- [17] R. Vindua and A. U. Zailani, “Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, p. 479, 2023, doi: 10.30865/jurikom.v10i2.5945.
- [18] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, “Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia,” *EduTic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [19] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, “Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning,” *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
- [20] A. R. Maulana, S. H. Wijoyo, and Y. T. Mursityo, “Sekolah Dasar Dan Sekolah Menengah Pada Media Sosial Twitter Dengan Menggunakan Metode Word Embedding Dan Long Short-Term Memory Networks (LSTM),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 3, pp. 523–530, 2023, doi: 10.25126/jtiik.2023106977.
- [21] R. Nooraeni, A. B. Safiruddin, A. F. Afifah, K. D. Agung, and N. N. Rosyad, “Analisis Sentimen Publik terhadap Sistem Zonasi Sekolah Menggunakan Data Twitter dengan Metode Naive Bayes Classification,” *Fakt. Exacta*, vol. 12, no. 4, p. 315, 2020, doi: 10.30998/faktorexacta.v12i4.5205.
- [22] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, “Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps,” *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 16, no. 4, pp. 28–38, 2022, doi: 10.35969/interkom.v16i4.192.
- [23] S. D. Utami and A. Erfina, “Analisis Sentimen Objek Wisata Bali Di Google Maps Menggunakan Algoritma Naive Bayes,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, pp. 418–427, 2022.
- [24] G. Fikri Baihaqi, D. E. Ratnawati, and B. T. Hanggara, “Analisis Sentimen Wisata Alun-Alun Kota Batu menggunakan Algoritma Support Vector Machine,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. Desember, pp. 6010–6018, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [25] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [26] F. Bei and S. Sudin, “Analisis Sentimen Aplikasi Tiket Online Di Play Store Menggunakan Metode Support Vector Machine (Svm),” *Sismatik*, vol. 01, no. 01, pp. 91–97, 2021.
- [27] A. R. Ismail and Raden Bagus Fajriya Hakim, “Implementasi Lexicon Based Untuk Analisis Sentimen Dalam Menentukan Rekomendasi Pantai Di DI Yogyakarta Berdasarkan Data Twitter,” *Emerg. Stat. Data Sci. J.*, vol. 1, no. 1, pp. 37–46, 2023, doi: 10.20885/esds.vol1.iss.1.art5.