

**ANALISIS SENTIMEN OPINI MASYARAKAT PENGGUNA TWITTER
TERHADAP PARIWISATA LAMPUNG MENGGUNAKAN
*SUPPORT VECTOR MACHINE DAN NAIVE BAYES***

(Skripsi)

**Oleh
ARFINA SHELLA MEILANY
1817051005**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

ABSTRAK

ANALISIS SENTIMEN OPINI MASYARAKAT PENGGUNA TWITTER TERHADAP PARIWISATA LAMPUNG MENGGUNAKAN *SUPPORT VECTOR MACHINE* DAN *NAIVE BAYES*

Oleh

Arfina Shella Meilany

Pesatnya perkembangan teknologi membentuk gaya hidup serba digital. Perangkat elektronik sering digunakan sebagai alat yang dapat digunakan untuk mengakses informasi. Penyebarluasan informasi seperti referensi, berita, hingga tren terkini dapat dengan mudah dilakukan melalui media sosial. Salah satu media sosial yang paling populer digunakan sebagai sarana penyediaan berbagai sumber data penelitian, informasi, dan opini publik adalah Twitter. Dengan *tweet* menjadi gambaran proses mendapatkan informasi tersebut dan menganalisis opini dengan memanfaatkan salah satu riset yaitu analisis sentimen. Analisis sentimen adalah proses untuk mengevaluasi pendapat pengguna media sosial seperti Twitter. Pendapat atau penilaian yang disampaikan oleh pengguna Twitter terkait topik suatu objek atau peristiwa yang baru-baru ini terjadi dan salah satu topik populer yang diperbincangkan oleh masyarakat Lampung terkait Pariwisata Lampung. Pariwisata Lampung merupakan sektor salah satu provinsi tujuan wisata di Indonesia yang memiliki berbagai fasilitas yang disediakan. Dalam penelitian ini, kami membandingkan dua algoritma *Support Vector Machine* (SVM) dan *Naive Bayes* dalam analisis sentimen menggunakan data dari Twitter terkait Pariwisata Lampung. Berdasarkan hasil penelitian dengan menggunakan *Support Vector Machine* (SVM) menunjukkan hasil kinerja yang lebih baik dengan nilai akurasi sebesar 94,11% dibandingkan dengan *Naive Bayes* yang memiliki nilai akurasi sebesar 93,78%.

Kata Kunci: Analisis Sentimen, *Support Vector Machine* (SVM), *Naive Bayes*, Twitter, Pariwisata Lampung.

ABSTRACT

SENTIMENT ANALYSIS OF SUPPORT VECTOR MACHINE AND NAIVE BAYES FOR LAMPUNG TOURISM BASED ON TWITTER

By

Arfina Shella Meilany

The rapid development of technology forms an all-digital lifestyle. Electronic devices are often used as a tool that can be used to access information. Dissemination of information such as references, news, to the latest trends can easily be done through social media. One of the most popular social media used as a means of providing various sources of research data, information, and public opinion is Twitter. With tweets, it becomes an illustration of the process of getting that information and analyzing opinions by utilizing one of the researches, namely sentiment analysis. Sentiment analysis is a process to evaluate the opinions of social media users such as Twitter. Opinions or assessments submitted by Twitter users regarding the topic of an object or event that recently occurred and one of the popular topics discussed by the people of Lampung regarding Lampung Tourism. Lampung tourism is a sector of one of the tourist destination provinces in Indonesia which has various facilities provided. In this study, we compare two Support Vector Machine (SVM) and Naive Bayes algorithms in sentiment analysis using data from Twitter regarding Lampung Tourism. Based on the results of research using the Support Vector Machine (SVM) shows better performance results with an accuracy value of 94.11% compared to Naive Bayes which has an accuracy value of 93,78%.

Keywords: Sentiment Analysis, Support Vector Machine (SVM), Naive Bayes, Twitter, Lampung Tourism.

**ANALISIS SENTIMEN OPINI MASYARAKAT PENGGUNA TWITTER
TERHADAP PARIWISATA LAMPUNG MENGGUNAKAN
SUPPORT VECTOR MACHINE DAN *NAIVE BAYES***

Oleh
ARFINA SHELLA MEILANY
1817051005

Skripsi
Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA KOMPUTER

Pada
Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung



FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022

Judul Skripsi : **ANALISIS SENTIMEN OPINI MASYARAKAT
PENGGUNA TWITTER TERHADAP PARIWISATA
LAMPUNG MENGGUNAKAN *SUPPORT VECTOR
MACHINE* DAN *NAIVE BAYES***

Nama Mahasiswa : **Arfina Shella Meilany**

Nomor Induk Mahasiswa : **1817051005**


Program Studi : **S1 Ilmu Komputer**

Jurusan : **Ilmu Komputer**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**




1. **Komisi Pembimbing**


Prof. Drs. Admi Syarif, Ph.D.
NIP. 19670103 199203 1 003


Yunda Heningtyas, M.Kom.
NIP. 19890108 201903 2 014

2. **Ketua Jurusan Ilmu Komputer**


Didik Kurniawan, S.Si., M.T.
NIP 19800419 200501 1 004

MENGESAHKAN

1. Tim Penguji

Ketua : Prof. Drs. Admi Syarif, Ph.D.



**Penguji I
Sekertaris : Yunda Heningtyas, M.Kom.**



**Penguji II
Penguji Pembahas : Favorisen Lumbanraja, Ph.D.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T.
NIP 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : 23 Desember 2022

PERNYATAAN

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi saya yang berjudul "**Analisis Sentimen Opini Masyarakat Pengguna Twitter Terhadap Pariwisata Lampung Menggunakan *Support Vector Machine* Dan *Naive Bayes***" merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang di skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila dikemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung, 24 Januari 2023



ARFINA SHELLA MEILANY

NPM. 1817051005

RIWAYAT HIDUP



Penulis dilahirkan pada 26 Mei 2000 di Bandar Lampung sebagai anak pertama dari tiga bersaudara dengan Ayah bernama Lutfi Cahyono Adi dan Ibu bernama Ari Puspita.

Penulis menyelesaikan Pendidikan formal pertama kali di Taman Kanak-kanak (TK) Kartika II-26 Bandar Lampung pada tahun 2006, Pendidikan dasar di Sekolah Dasar (SD)

Negeri 2 Calang, Aceh Jaya pada tahun 2012, Sekolah Menengah Pertama (SMP) Negeri 1 Dringu, Probolinggo pada tahun 2015 dan Sekolah Menengah Atas (SMA) Negeri 1 Meulaboh, Aceh Barat yang diselesaikan pada tahun 2018.

Pada tahun 2018, penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SNMPTN. Selama menjadi mahasiswa beberapa kegiatan yang telah dilakukan penulis antara lain:

1. Menjadi anggota aktif bidang Medinfo di Organisasi Himpunan Mahasiswa Ilmu Komputer (HIMAKOM) pada periode 2018/2019.
2. Menjadi Asisten Dosen Jurusan Ilmu Komputer untuk mata kuliah Keamanan Sistem Informasi pada periode 2020/2021.
3. Pada bulan Februari 2021 penulis melaksanakan Kerja Praktik (KP) di STMIK Dharma Wacana Metro, Kota Metro.

4. Pada bulan September 2021 penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Sumur Putri, Kecamatan Teluk Betung Selatan, Kota Bandar Lampung.

PERSEMBAHAN

Alhamdulillahillobbilamin

Puji dan syukur kepada Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga saya dapat menyelesaikan skripsi ini. Sholawat serta salam saya sanjungkan kepada Nabi Muhammad SAW.

Aku persembahkan karya ini kepada:

Kedua Orang Tuaku Tercinta

Sebagai tanda terima kasihku kepada Papa dan Mama yang tercinta. Terima kasih senantiasa memberikan yang terbaik. Terima kasih melantunkan doa yang selalu menyertaiku. Terima kasih telah mendidik dan membesarkanku dengan kasih sayang. Terima kasih selalu mendukungku dalam segala pilihank dan semua pengorbanan yang belum bisa terbalaskan. Terima kasih Papa dan Mama.

Adikku Putri Ramadhani Arum Sari dan R.P. Ananda Wijaya P. Adi

Terima kasih telah memberikan semangat, dukungan, dan doa.

Seluruh Keluarga Besar Ilmu Komputer 2018

Yang selalu memberikan semangat dan dukungan

Almamater Tercinta, Universitas Lampung

MOTTO

“Rahasia untuk maju adalah memulai.”

(Mark Twain)

“Seorang pemenang adalah pemimpi yang tidak pernah menyerah.”

(Nelson Mandela)

“Bersemangatlah atas hal-hal yang bermanfaat bagimu. Minta tolonglah pada Allah, jangan engkau lemah.”

(HR. Muslim)

“Pendidikan adalah salah satunya kunci untuk membuka dunia ini, serta paspor untuk menuju kebebasan.”

(Oprah Winfrey)

SANWACANA

Puji syukur atas kehadiran Allah SWT, karena berkat rahmat dan hidayahnya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Opini Masyarakat Pengguna Twitter Terhadap Pariwisata Lampung Menggunakan *Support Vector Machine* Dan *Naive Bayes*” yang merupakan salah satu syarat untuk memperoleh gelar Sarjana Komputer di Jurusan Ilmu Komputer, FMIPA, Universitas Lampung.

Dalam proses penulisan skripsi dapat berjalan dikarenakan terdapat banyak pihak yang membantu. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kedua Orang tua tercinta, Papa Lutfi Cahyono Adi dan Mama Ari Puspita, adikku yaitu Putri Ramadhani Arum Sari dan R.P.Ananda Wijaya P.Adi yang selalu mendoakan, memberikan dukungan, dan kasih sayang yang membuat penulis bersemangat untuk mengerjakan skripsi agar dapat berjalan lancar.
2. Bapak Prof. Drs. Admi Syarif, Ph.D selaku Dosen Pembimbing I yang telah memberikan motivasi, semangat, candaan, kritik, saran, dan bimbingan dalam menyelesaikan skripsi ini.
3. Ibu Yunda Heningtyas, M.Kom. selaku Dosen Pembimbing II yang telah memberikan saran, motivasi dan bimbingan dalam menyelesaikan skripsi ini.
4. Bapak Favorisen R. Lumbanraja, Ph.D. selaku Pembahas dalam skripsi ini. Terima kasih atas saran dan masukan dalam skripsi ini.

5. Bapak Ardiansyah, S.Kom., M.Kom. selaku dosen pembimbing akademik yang telah memberikan bimbingan selama setiap semester.
6. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T. selaku dekan FMIPA, Universitas Lampung.
7. Bapak Didik Kurniawan, S.Si., M.T. selaku Ketua Jurusan Ilmu Komputer, FMIPA, Universitas Lampung.
8. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc. selaku Sekretaris Jurusan Ilmu Komputer, FMIPA, Universitas Lampung.
9. Bapak dan Ibu Dosen Jurusan Ilmu Komputer, FMIPA, Universitas Lampung yang telah memberikan ilmu dan pengalaman bagi penulis.
10. Ibu Ade Nora Maela, Mas Nofal, dan Bang Zainuddin yang telah membantu segala urusan administrasi dan segala jenis izin di Jurusan Ilmu Komputer.
11. Sahabat serta teman seperjuangan dari awal masuk Ilmu Komputer Arafia Isnayu Akaf. Terima kasih selalu memberikan semangat, motivasi, mendengarkan keluh kesah, berjuang bersama, dan menemani suka maupun duka.
12. Teman-temanku Atika Indah Nuraini, Elshinta Milenia, Yasmin Hasna, Amara Indah Pancarani, Bobby Malela Hutagalung, Muhammad Nur Ashiddiqi, Pandi Barep Arianza, dan Ridho Alrafi. Terima kasih atas bantuan, motivasi, dukungan dan selalu ada menemani penulis berjuang hingga mendapat gelar S.Kom.
13. Rizki Aris Pratama. Terima kasih selalu menemani serta memberikan semangat dukungan dan doa.
14. Teman-teman Ilmu Komputer 2018. Terima kasih telah menjadi rekan seperjuangan pada saat kuliah.

15. Semua pihak yang telah banyak membantu dalam proses penyelesaian skripsi ini yang belum disebutkan satu persatu.

Penulis menyadari bahwa dalam penulisan skripsi ini masih terdapat banyak kekurangan karena keterbatasan pengalaman serta pengetahuan penulis. Semoga skripsi ini memberikan manfaat bagi pihak yang membaca.

Bandar Lampung, 24 Januari 2023



Arfina Shella Meilany

NPM. 1817051005

DAFTAR ISI

	Halaman
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
DAFTAR KODE PROGRAM	xvi
I. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan.....	5
II. TINJAUAN PUSTAKA	7
2.1 Kecerdasan Buatan.....	7
2.2 <i>Machine Learning</i>	7
2.3 <i>Text Mining</i>	8
2.4 Analisis Sentimen	9
2.5 Twitter.....	9
2.6 <i>K-Fold Cross Validation</i>	10
2.7 <i>Support Vector Machine</i>	11
2.8 <i>Naive Bayes</i>	12
2.9 Evaluasi Model Hasil Pengujian	13
2.10 Python.....	14
2.11 Jupyter Notebook.....	15
2.12 Pariwisata Lampung.....	15
2.13 Penelitian Terdahulu	17
III. METODOLOGI PENELITIAN	19
3.1 Waktu dan Tempat Penelitian	19
3.2 Perangkat Penelitian.....	19

3.2.1 Perangkat Keras (<i>Hardware</i>).....	20
3.2.2 Perangkat Lunak (<i>Software</i>).....	20
3.3 Tahapan Penelitian	20
3.3.1 Studi Literatur	21
3.3.2 Pengumpulan Data	21
3.3.3 <i>Preprocessing</i> Data	25
3.3.4 Pembagian Data	31
3.3.5 Klasifikasi <i>Support Vector Machine</i>	35
3.3.6 Klasifikasi <i>Naive Bayes</i>	35
3.3.7 Evaluasi.....	36
IV. HASIL DAN PEMBAHASAN	37
4.1 Klasifikasi <i>Support Vector Machine</i> (SVM)	37
4.2 Klasifikasi <i>Naive Bayes</i>	38
4.3 <i>Training</i> Klasifikasi	39
4.4 Evaluasi Hasil Klasifikasi	40
V. SIMPULAN DAN SARAN	60
5.1 Simpulan	60
5.2 Saran	60
DAFTAR PUSTAKA	62

DAFTAR GAMBAR

Gambar	Halaman
1. Ilustrasi <i>K-Fold Cross Validation</i>	10
2. Ilustrasi <i>Support Vector Machine</i>	11
3. Perencanaan Waktu Penelitian.	19
4. Tahapan Penelitian.	20
5. Pengumpulan Data.	22
6. Hasil data <i>labeling manual</i>	23
7. Dataset yang digunakan pada penelitian.	23
8. Alur <i>Preprocessing Data</i>	26
9. Hasil Tahap <i>Case Folding</i>	28
10. Hasil Tahap <i>Removing Stopword</i>	29
11. Hasil Tahap <i>Stemming</i>	30
12. Hasil Tahap <i>Tokenizing</i>	30
13. Hasil <i>Vectorizing</i>	31
14. Hasil Perbandingan Akurasi SVM dan <i>Naive Bayes</i>	55
15. Hasil Perbandingan Presisi SVM dan <i>Naive Bayes</i>	56
16. Hasil Perbandingan <i>Recall</i> SVM dan <i>Naive Bayes</i>	57
17. Hasil Perbandingan <i>F1-Score</i> SVM dan <i>Naive Bayes</i>	58

DAFTAR TABEL

Tabel	Halaman
1. <i>Confusion Matrix</i>	13
2. Penelitian Terdahulu.....	17
3. Pengumpulan Data Twitter.....	22
4. Sentimen Positif.....	24
5. Sentimen Negatif.....	25
6. Pembagian Data Latih dan Data Uji.....	31
7. Kinerja Metode SVM.....	40
8. <i>Confusion Matrix SVM fold-1</i>	41
9. <i>Confusion Matrix SVM fold-2</i>	41
10. <i>Confusion Matrix SVM fold-3</i>	42
11. <i>Confusion Matrix SVM fold-4</i>	43
12. <i>Confusion Matrix SVM fold-5</i>	43
13. <i>Confusion Matrix SVM fold-6</i>	44
14. <i>Confusion Matrix SVM fold-7</i>	44
15. <i>Confusion Matrix SVM fold-8</i>	45
16. <i>Confusion Matrix SVM fold-9</i>	45
17. <i>Confusion Matrix SVM fold-10</i>	46
18. Kinerja Metode <i>Naive Bayes</i>	47
19. <i>Confusion Matrix Naive Bayes fold-1</i>	47
20. <i>Confusion Matrix Naive Bayes fold-2</i>	48
21. <i>Confusion Matrix Naive Bayes fold-3</i>	48
22. <i>Confusion Matrix Naive Bayes fold-4</i>	49
23. <i>Confusion Matrix Naive Bayes fold-5</i>	49
24. <i>Confusion Matrix Naive Bayes fold-6</i>	50
25. <i>Confusion Matrix Naive Bayes fold-7</i>	51
26. <i>Confusion Matrix Naive Bayes fold-8</i>	51
27. <i>Confusion Matrix Naive Bayes fold-9</i>	52
28. <i>Confusion Matrix Naive Bayes fold-10</i>	52

29. Perbandingan Kinerja Klasifikasi SVM dan <i>Naive Bayes</i>	54
30. Waktu Komputasi.	59

DAFTAR KODE PROGRAM

Kode Program	Halaman
1. Menghilangkan <i>New Line</i>	26
2. Menghilangkan <i>Emoji</i>	26
3. Menghilangkan <i>Hyperlink</i>	26
4. Menghilangkan <i>Username</i>	27
5. Menghilangkan Angka dan Simbol.	27
6. Menghilangkan Multi spasi.	27
7. Mengubah <i>Tweet</i> Menjadi Huruf Kecil.	27
8. Menghilangkan <i>Stopword</i>	28
9. Mengubah Kata Menjadi Bentuk Dasar.	29
10. <i>Tokenizing</i> dan <i>Vectorizing</i>	30
11. Pembagian Data <i>fold-1</i>	32
12. Pembagian Data <i>fold-2</i>	32
13. Pembagian Data <i>fold-3</i>	32
14. Pembagian Data <i>fold-4</i>	33
15. Pembagian Data <i>fold-5</i>	33
16. Pembagian Data <i>fold-6</i>	33
17. Pembagian Data <i>fold-7</i>	34
18. Pembagian Data <i>fold-8</i>	34
19. Pembagian Data <i>fold-9</i>	34
20. Pembagian Data <i>fold-10</i>	35
21. Model SVM.	37
22. <i>Confusion Matrix SVM</i>	38
23. Model <i>Naive Bayes</i>	38
24. <i>Confusion Matrix Naive Bayes</i>	39
25. <i>Training</i> Klasifikasi SVM.....	39
26. <i>Training</i> Klasifikasi <i>Naive Bayes</i>	40

I. PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi yang semakin pesat menyebabkan mudahnya penyebaran informasi melalui media sosial. Perkembangan tersebut membentuk pola hidup serba digital. Berbagai inovasi baru diciptakan untuk memudahkan masyarakat mendapat informasi dalam memenuhi kebutuhan masyarakat tersebut. Salah satu inovasinya adalah penggunaan *smartphone* untuk mendapatkan suatu informasi di berbagai media sosial. Saat ini media sosial memiliki peran besar dalam menyebarkan informasi kepada masyarakat. Salah satu media sosial yang cukup populer digunakan di Indonesia adalah Twitter. Pengguna Twitter pada akhir 2021 mencapai 16,32 juta (Utami and Erfina, 2021). Twitter merupakan media sosial yang dapat digunakan pengguna untuk mempublikasikan ide serta mencari berbagai informasi tentang hiburan, pendidikan, politik, dan lainnya. Ide yang dipublikasikan di Twitter merupakan pendapat pengguna mengenai suatu objek ataupun peristiwa yang terjadi (Putri and Romli, 2021). Salah satu opini Twitter yang mengalami cukup banyak perbincangan mengenai perkembangan sektor pariwisata di Indonesia. Hampir seluruh masyarakat menggunakan media sosial untuk beropini sesuai pendapat mereka sendiri (Sodik and Kharisudin, 2021).

Pariwisata merupakan sektor yang cukup penting dalam pertumbuhan ekonomi negara. Indonesia memiliki banyak objek pariwisata yang menarik dan sangat beragam. Banyaknya wisatawan lokal maupun mancanegara menikmati indahnnya pemandangan wisata yang berada di

Indonesia. Peningkatan wisatawan dapat meningkatkan perekonomian terutama tempat daerah tujuan wisata tersebut. Pada awal tahun 2020, wisatawan Indonesia mengalami penurunan yang signifikan. Hal ini disebabkan karena Indonesia merupakan salah satu negara pandemi Covid 19. Berbagai kebijakan dilakukan seperti melakukan protokol kesehatan, kelengkapan dokumen sebagai kelengkapan dalam perjalanan (Susanti, Sedyono, and Sembiring, 2016).

Pariwisata di Indonesia menjadi bagian dari pembangunan nasional secara sistematis dan terencana. Pengembangan sektor pariwisata dilakukan dengan mengembangkan produk dan jasa serta kualitas pada tempat wisata tersebut. Dalam melakukan manajemen pengelolaan wisata, diperlukan inovasi serta perawatan yang cukup. Strategi yang dilakukan dengan mengembangkan fasilitas penunjang wisata sehingga kepuasan masyarakat terhadap pengunjung wisata meningkat. Sektor pariwisata Provinsi Lampung merupakan salah satu Provinsi tujuan wisata di Indonesia. Lampung memiliki pantai, museum, wahana, dan lainnya (Sumastuti, Prabowo, and Violinda, 2021).

Berbagai wisata dan fasilitas yang ada di Lampung membuat masyarakat beropini mengenai kepuasan yang mereka dapatkan. Salah satu media sosial untuk mendapatkan opini tersebut melalui Twitter. Analisis komentar masyarakat yang terdapat di Twitter sangat sulit, dikarenakan data-data di media sosial tidak mudah dalam mendeteksi menjadi sentimen. Komentar pada tweet tersebut merupakan opini masyarakat yang mereka buat dengan ejaan kata-kata serta tanda baca yang mereka buat sendiri. Dengan demikian membuat dalam mereka beropini cukup salah dalam penyampaiannya perlu dikoreksi (Ardianto et al., 2020). Komentar yang terdapat di Twitter atau data di Twitter tersebut merupakan sebuah bahan penting yang dipilih dan menarik karena sebagian besar masyarakat menyampaikan pendapatnya melalui sosial media termasuk pada Twitter, dengan Twitter masyarakat bebas

beropini mengenai dukungan, keluh kesah, kritik dan saran mengenai perkembangan pariwisata yang ada di Lampung ini, bahkan dapat memberikan gagasannya mengenai kebijakan yang telah dilakukan sektor pariwisata Lampung.

Mengetahui seberapa puas masyarakat dalam menikmati pariwisata yang berada di Lampung apakah yang diberikan dapat dianggap positif atau negatif, penelitian ini menggunakan analisis sentimen (Widowati and Sadikin, 2021). Analisis sentimen merupakan metode dalam memperoleh data serta memahami suatu data dari platform di internet (Que, Iriani, and Purnomo, 2020). Analisis sentimen memfokuskan analisis dari *review* teks sebagai prediksi gambaran perasaan para netizen pada suatu kasus atau suatu objek di media sosial terutama Twitter (Tineges, Triayudi, and Sholihati, 2020). Tujuan dari sentimen analisis adalah mengevaluasi dari opini pengguna Twitter (Amalia, Umaidah, and Mayasari, 2021).

Berdasarkan latar belakang diatas, penelitian ini akan diklasifikasikan kedalam sentimen menggunakan algoritma SVM dan *Naive Bayes*. Hasil dari metode SVM dan *Naive Bayes* dievaluasi dengan *Confusion Matrix* untuk menghitung *accuracy*, *precision*, *recall*, dan *f1 score*. Menarik sebuah kesimpulan dari hasil evaluasi yang akan menghasilkan perbandingan dengan metode SVM dan *Naive Bayes*. SVM merupakan salah satu algoritma menghasilkan proses pembelajaran suatu masalah yang diterjemahkan untuk mencari garis untuk memisahkan kedua kelompok tersebut. *Naive Bayes* merupakan suatu yang dapat memprediksi peluang dimasa depan berdasarkan pengalaman sebelumnya (Darmawan, Kustian, and Rahayu, 2018).

Penelitian ini bertujuan untuk mengklasifikasikan dengan topik *tweet* pariwisata dengan metode SVM dan *Naive Bayes* dan dapat menyajikan penjabaran analisis sentimen dari kedua metode tersebut dengan baik

sehingga dapat memberikan informasi perkembangan khususnya untuk pariwisata Lampung.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan sebelumnya, maka rumusan masalah dalam penelitian ini adalah perbandingan metode *Support Vector Machine* dan *Naive Bayes* dalam mengevaluasi opini masyarakat pengguna twitter terhadap pariwisata Lampung.

1.3 Batasan Masalah

Dalam penelitian ini diberikan batasan-batasan masalah sebagai berikut:

- 1.3.1 Sumber opini tweet yang digunakan hanya berbahasa Indonesia.
- 1.3.2 Data tweet yang digunakan berhubungan dengan kata kunci Pariwisata Lampung.
- 1.3.3 Algoritma yang digunakan yaitu *Support Vector Machine* dan *Naive Bayes*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

- 1.4.1 Menerapkan algoritma *Support Vector Machine* dan *Naive Bayes* untuk melakukan analisis sentimen terhadap Pariwisata Lampung.
- 1.4.2 Menguji dan membandingkan mengenai hasil akurasi, presisi, *recall*, *f1-score*, dan performansi kecepatan algoritma *Support Vector Machine* dan *Naive Bayes*.

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut:

- 1.5.1 Membantu dalam menganalisis sentimen dengan metode *Support Vector Machine* dan *Naive Bayes*.
- 1.5.2 Mengetahui keunggulan dalam analisis sentimen antara algoritma *Support Vector Machine* dan *Naive Bayes*.

1.6 Sistematika Penulisan

Pada penulisan skripsi ini dibagi dalam tiga bagian sebagai berikut :

a. Bagian Awal

Pada bagian awal skripsi ini, berisi halaman judul, halaman pengesahan, motto dan persembahan, kata pengantar, abstrak, daftar isi, daftar tabel, daftar gambar, dan daftar lampiran.

b. Bagian Pokok

Pada bagian pokok penulisan skripsi ini, terdiri dari lima bab yaitu :

BAB I. PENDAHULUAN

Bab ini menguraikan pendahuluan yang berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II. TINJAUAN PUSTAKA

Bab ini menjelaskan tentang landasan teori yang berkaitan dengan topik yang berhubungan dengan sentimen analisis dengan menggunakan metode *Support Vector Machine* dan *Naive Bayes*.

BAB III. METODE PENELITIAN

Bab ini menjelaskan tentang prosedur yang dilakukan dalam penelitian ini meliputi tahapan penelitian, pengumpulan data, serta gambaran yang akan dikerjakan.

BAB IV. HASIL DAN PEMBAHASAN

Bab ini berisi hasil dari pengolahan data yang dilakukan dengan menggunakan metode SVM dan *Naive Bayes*.

BAB V. PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian ini dan disertai saran yang diberikan apabila ingin melanjutkan penelitian dengan tema tersebut.

c. Bagian Akhir

Pada bagian akhir skripsi berisi daftar pustaka sebagai acuan atau referensi penulisan yang mendukung pada isi skripsi ini.

II. TINJAUAN PUSTAKA

2.1 Kecerdasan Buatan

Kecerdasan Buatan atau *Artificial Intelligence* merupakan kemampuan yang memiliki konsep sebagai alat bantu yang dapat berpikir dan bekerja seperti manusia. Kecerdasan buatan bertujuan dalam membantu menemukan sebuah permasalahan atau pengambilan keputusan dan mencari solusi untuk menyelesaikan permasalahan tersebut. Bagian utama kecerdasan buatan adalah pengetahuan yang diperoleh melalui pendidikan dan pengalaman. Kecerdasan buatan menghasilkan baik media maupun teori kecerdasan yang dinyatakan dalam bahasa program komputer (Fauzan, 2020).

2.2 *Machine Learning*

Machine Learning merupakan suatu cabang Ilmu Komputer yang berhubungan dengan *Artificial Intelligence* dan studi algoritma komputer yang bekerja dengan mengumpulkan data dan membandingkan suatu data yang memiliki ukuran skala besar. *Machine Learning* digunakan sebagai *tools* yang mampu dalam klasifikasi jenis sentimen. *Machine Learning* mempelajari algoritma dan model statistik. *Machine Learning* dapat membantu algoritma yang digunakan agar bekerja secara otomatis. Salah satu yang dikembangkan oleh *Machine Learning* adalah analisis sentimen (*sentiment analysis*) (Santoso, Gata, and Paryanti, 2019).

2.3 *Text Mining*

Text mining merupakan proses dalam mengambil informasi dan pengetahuan dari berbagai sumber data yang berkualitas tinggi dari suatu teks. Sumber data pada *text mining* berdasarkan data teks yang memiliki sifat tidak terstruktur. Data teks yang digunakan seperti *tweet* di Twitter, artikel, dan lainnya. Data teks yang tidak terstruktur biasanya berupa teks, video, audio, foto dan lainnya (Fitriyah, Warsito, and Maruddani, 2020).

Dalam proses *text mining*, sebuah teks dokumen atau data mentah yang digunakan dipersiapkan terlebih dahulu sebelum di proses selanjutnya. *Text preprocessing* disebut juga mempersiapkan teks dokumen atau data mentah. Proses tersebut bertujuan pengubahan bahasa alami yaitu mengubah data yang tidak terstruktur menjadi terstruktur. Pada *text mining* terdapat tahapan *text preprocessing* (Husna, 2020) sebagai berikut.

- a. *Cleaning* merupakan suatu proses dalam membersihkan informasi teks dari bagian yang tidak diperlukan.
- b. *Case Folding* merupakan suatu bentuk *text preprocessing* menyeragamkan kata didalam teks menjadi huruf kecil (*lowercase*).
- c. *Removing Stopword* merupakan suatu proses dengan menghilangkan kata yang tidak diperlukan dalam data yang didapatkan.
- d. *Stemming* merupakan suatu bentuk *text preprocessing* dalam proses transformasi dari semua kata menjadi kata dasar.
- e. *Tokenizing* merupakan suatu bentuk *text preprocessing* dalam proses memecah kalimat menjadi satuan kata, dan seperti angka, simbol, dan tanda baca dihilangkan.

2.4 Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan proses dalam mengolah data dan mengekstrak data secara otomatis pada kalimat opini. Proses analisis dapat dilakukan dengan pendekatan *machine learning*. Analisis sentimen bertujuan untuk mempublikasikan suatu ide penelitian, yang akan mempresentasikan sebuah analisis terkait dengan pengklasifikasian sebuah data berupa teks dengan menggunakan metode SVM dan *Naive Bayes*. Analisis sentimen mengelompokkan kalimat atau pendapat berdasarkan popularitas teks didalamnya. Popularitas tersebut merupakan pendapat yang memiliki aspek positif yang berkaitan dengan hal baik atau keadaan normal dan aspek negatif yang berkaitan dengan hal buruk atau keadaan yang tidak diinginkan. Analisis sentimen pada Twitter menjadi suatu alat yang dapat menganalisis dari persepsi masyarakat. Pendekatan dengan analisis pada Twitter ini berfokus pada identifikasi sentimen tweet setiap individu (Que, Iriani, and Purnomo, 2020).

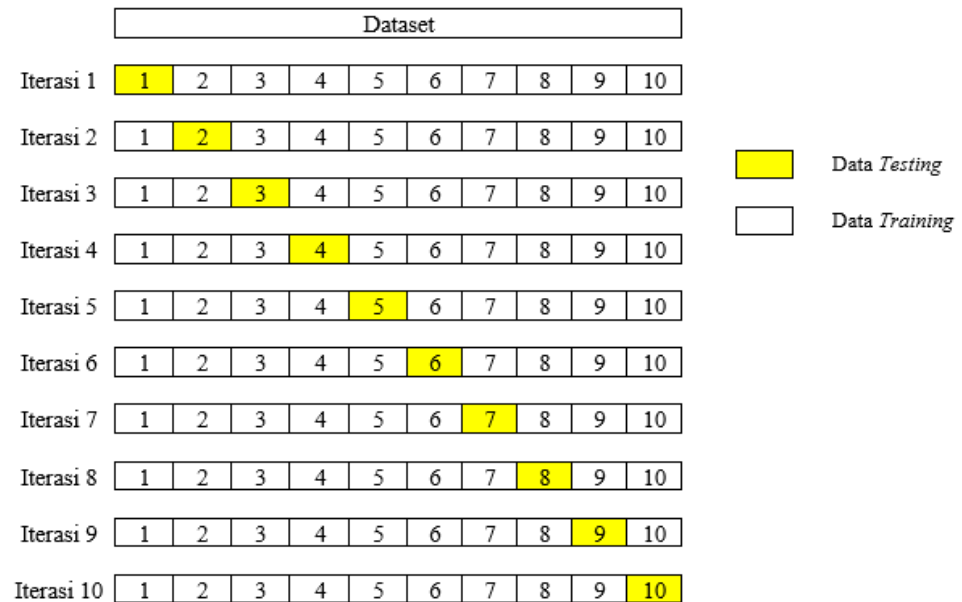
2.5 Twitter

Twitter merupakan platform media sosial yang populer di Indonesia. Twitter dapat memposting atau mempublikasi tweet opini masyarakat tentang hiburan, layanan, dan lainnya. Pengguna twitter dapat menyampaikan opini pribadinya bahkan dengan twitter dapat sebagai ladang berbisnis. Twitter memiliki tujuan untuk berinteraksi sosial, mudah dalam pencarian informasi, mendapatkan informasi, dokumentasi dan berbagi informasi (Fikri, Sabrila, and Azhar, 2020).

Twitter juga merupakan platform yang digunakan untuk melakukan tugas menganalisis sentimen. Pengguna Twitter membuat status atau yang sering disebut dengan *tweet*. Status tersebut pada Twitter setiap hari memposting suatu teks akan bertambah banyak (Kristiyanti et al., 2019).

2.6 *K-Fold Cross Validation*

K-Fold Cross Validation merupakan salah satu cara validasi atau menemukan parameter terbaik dari sebuah *dataset*. Pada *K-Fold Cross Validation* data dibagi menjadi k kelompok yang memiliki ukuran yang sama. Selanjutnya, subset dibagi menjadi data latih (*data training*) dan data uji (*data testing*). Iterasi dari setiap kelompok tersebut adalah data uji (*data testing*), sedangkan kelompok yang tidak berperan di dalam data uji merupakan kelompok dari data latih (*data training*). Data latih juga berperan sebagai model dan data testing diklasifikasikan serta dihitung keakuratan model atau nilai *error* berdasarkan model tersebut. Tujuan dari *cross validation* adalah memberikan solusi dengan membagi data menjadi *fold* untuk menguji model atau validasi data di beberapa titik *cross validation* (Nasution and Hayaty, 2019). Berikut merupakan ilustrasi proses *k-fold cross validation* dapat dilihat pada Gambar 1.



Gambar 1. Ilustrasi *K-Fold Cross Validation*.

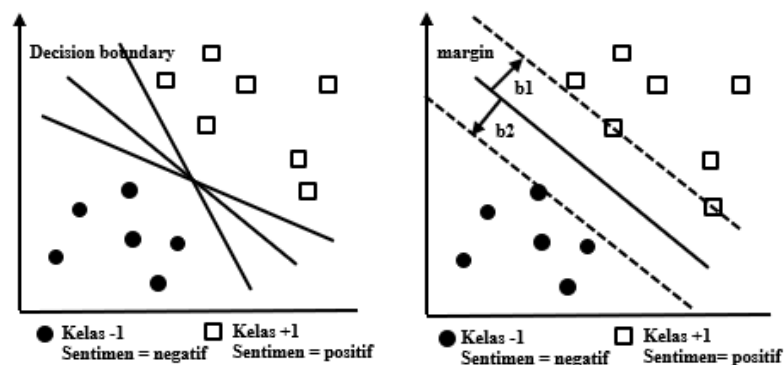
Sumber: (Nasution and Hayaty, 2019).

Jika dilakukan *k-fold* dengan $k = 10$, maka *dataset* dibagi menjadi 10 kelompok. Iterasi pertama yaitu kelompok 1 merupakan data uji, sedangkan kelompok 2-10 merupakan data latih. Selanjutnya, iterasi kedua yaitu kelompok 2 merupakan data uji, sedangkan kelompok 1 dan 3-10 merupakan data latih. Berlaku hingga iterasi berikutnya sampai iterasi kelima (Peryanto, Yudhana, and Umar, 2020).

2.7 Support Vector Machine

Support Vector Machine (SVM) merupakan suatu metode untuk melakukan prediksi atau klasifikasi pada suatu kelas berdasarkan pola. Prinsip kerja dari data SVM adalah dengan mencari garis pembatas (*hyperplane*) yang optimal dengan memaksimalkan dan memisahkan antara kelas yang berbeda yaitu positif dan negatif (Fitriyah, Warsito, and Maruddani, 2020).

Kelebihan dari SVM ini adalah bekerja dengan baik pada data, mendapatkan solusi optimal, rasio yang berkonversi rendah, dan memiliki kemampuan menghindari kesulitan dari masalah. Kelemahan pada SVM memiliki permasalahan pemilihan pada parameter yang sesuai dan berpengaruh terhadap hasil akurasi pada algoritma SVM tersebut. SVM memiliki tingkat akurasi lebih baik dibandingkan *Naive Bayes* (Kristiyanti et al., 2019).



Gambar 2. Ilustrasi *Support Vector Machine*.

Sumber: (Fitriyah, Warsito, and Maruddani, 2020).

Pada grafik pertama garis *Support Vector Machine* memungkinkan melakukan set data, sedangkan pada grafik kedua garis pemisah dengan margin maksimum dekat dengan *hyperplane*. Objek-objek yang paling dekat dengan *hyperplane* disebut dengan *Support Vector*. Untuk mendapatkan hasil klasifikasi yang baik, *hyperplane* harus dengan margin terbaik dan optimal (Santoso, Virginia, and Lukito, 2017).

2.8 *Naive Bayes*

Naive Bayes merupakan metode pengklasifikasian probabilitas yang paling sederhana dan banyak digunakan. *Naive Bayes* menerapkan teori *Bayes* dalam klasifikasi bahwa atribut suatu objek bersifat bebas. Pada klasifikasi ini berdasarkan probabilitas kata kunci dalam membandingkan data latih dan data uji. Proses dalam metode ini menyatakan terdapat atau tidaknya fitur pada kelas yang tidak berhubungan dengan kelas yang sama. Pada *Naive Bayes* juga dibutuhkan data awal berupa data pelatihan untuk mengambil suatu keputusan. Probabilitas yang dihitung pada metode ini dapat berdasarkan kata yang terdapat pada suatu dokumen (Sodik and Kharisudin, 2021).

Kelebihan dari *Naive Bayes* adalah paling sederhana, cepat, dan memiliki akurasi yang cukup tinggi. Dalam penggunaan metode ini membutuhkan jumlah data yang kecil untuk menentukan estimasi dalam proses pengklasifikasian. Metode *Naive Bayes* dalam proses klasifikasi memiliki dua tahap yaitu tahap pelatihan dilakukan proses analisis dengan sampel dokumen berupa pemilihan kata yang mungkin muncul dan tahap klasifikasi dilakukan dari nilai kategori suatu dokumen berdasarkan dokumen yang diklasifikasikan. Perhatikan rumus *Naive Bayes* sebagai berikut (Fikri, Sabrila, and Azhar, 2020).

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

Keterangan:

- X = Data dengan kelas yang belum diketahui
 H = Hipotesis data X merupakan kelas spesifik
 $P(H|X)$ = Probabilitas hipotesis data H berdasarkan kondisi X
 $P(H)$ = Probabilitas hipotesis data H
 $P(X|H)$ = Probabilitas X berdasarkan kondisi hipotesis H
 $P(X)$ = Probabilitas dari X.

2.9 Evaluasi Model Hasil Pengujian

Dalam keakuratan suatu data diperlukan validasi suatu model. Tujuan tahap evaluasi adalah mengetahui kegunaan pemodelan sebelumnya yang telah dilakukan. Hasil perbandingan pengujian dengan melakukan *Confusion Matrix*. *Confusion Matrix* merupakan *dataset* yang memiliki dua kelas yaitu positif dan negatif (Widowati and Sadikin, 2021). Bentuk *Confusion Matrix* dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Matrix* (Susanti, Sedyono, and Sembiring, 2016).

<i>Confusion Matrix</i>		<i>Prediction Class</i>	
		Negatif	Positif
<i>Actual Class</i>	Negatif	TN	FP
	Positif	FN	TP

Keterangan:

TN : *True Negative* merupakan data sentimen negatif yang sudah diklasifikasikan dengan hasil negatif.

FN : *False Negative* merupakan data sentimen positif yang sudah diklasifikasikan dengan hasil negatif.

FP : *False Positive* merupakan data sentimen negatif yang sudah diklasifikasikan dengan hasil positif.

TP : *True Positive* merupakan data sentimen positif yang sudah diklasifikasikan dengan hasil positif.

Pada hasil *confusion matrix* dari metode SVM dan *Naive Bayes* dengan klasifikasi data untuk menghitung dan melihat performa hasil perbandingan kedua metode tersebut dengan rumus sebagai berikut sebagai berikut (Susanti, Sedyono, and Sembiring, 2016).

$$2.9.1 \quad accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

Accuracy merupakan suatu rumus memprediksi kedekatan *True* (Positif dan Negatif) dengan nilai aktual keseluruhan data tersebut.

$$2.9.2 \quad precision = \frac{TP}{TP+FP} \times 100\%$$

Precision merupakan suatu rumus dalam tingkat ketepatan memprediksi *True Positive* dengan nilai keseluruhan data yang diprediksi positif.

$$2.9.3 \quad recall = \frac{TP}{TP+FN} \times 100\%$$

Recall merupakan suatu rumus dalam mengukur prediksi *True Positive* dengan nilai keseluruhan data yang diprediksi negatif.

$$2.9.4 \quad f1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

F1 score merupakan suatu rumus dalam perbandingan rata-rata dari *precision* dan *recall*.

2.10 Python

Python merupakan bahasa pemrograman yang dibuat oleh Guido van Rossum yang berorientasi pada objek dengan tingkat tinggi. Fitur di dalam python adalah mendukung dalam pemrograman yang berorientasi objek, imperatif, dan fungsional. Berikut ini merupakan beberapa *library* pada python sebagai berikut (Fahmi et al., 2020).

a. Tweepy

Tweepy merupakan *library* yang digunakan dalam mengambil data di Twitter dengan mengakses Twitter API.

b. Pandas

Pandas merupakan *library* yang digunakan untuk menganalisis dan mengolah data terstruktur.

c. *Snsrape*

Snsrape merupakan program yang digunakan untuk menyaring layanan sosial media yang peruntukan untuk mencari hal-hal yang diperlukan seperti profil pengguna, tagar, atau pencarian.

d. *Numerical Python*

Numerical Python merupakan *library* yang digunakan dalam proses komputasi dalam operasi vektor dan matriks.

e. CSV

Comma Separated Value (CSV) merupakan *library* yang digunakan dalam menyimpan data dengan format .csv.

2.11 Jupyter Notebook

Jupyter notebook merupakan aplikasi yang berbasis *web open source* berfungsi sebagai tempat membuat dan berbagi dokumen python. Pada dokumen tersebut terdapat kode yang dapat meningkatkan suatu tampilan dari *line code* dengan memadukan input, *output code*, *comment*, text, *image*, maupun persamaan pada matematika (Nofiyanti, Endah, and Oki Nur Haryanto, 2021).

2.12 Pariwisata Lampung

Provinsi Lampung saat ini merupakan salah satu provinsi yang potensial untuk bisa mengembangkan potensi dalam sektor pariwisatanya. Provinsi Lampung juga merupakan pintu gerbang Pulau Sumatera, dengan demikian Provinsi Lampung memiliki banyak tempat wisata dari berbagai kabupaten maupun kota. Wisata yang paling diminati adalah pantainya, seperti Pantai Teluk Kiluan, Pantai Duta Wisata, Pantai Mutun dan lainnya. Pantai sering dikunjungi wisatawan lokal bahkan mancanegara. Kebijakan yang dilakukan yaitu mengutamakan pembangunan pariwisata dan pengembangan objek-objek wisata yang berada di Provinsi Lampung. Semakin besar pengunjung, maka semakin besar juga penerimaan daerah yang bersumber dari wisatawan. Dinas Kebudayaan dan Pariwisata Lampung merancang kota agar membuat

daya tarik wisatawan seperti wisata kuliner, Museum Lampung, dan lainnya serta menargetkan wisatawan adalah wisatawan lokal (Mardihartono and Evadiani, 2019).

2.13 Penelitian Terdahulu

Tabel 2. Penelitian Terdahulu.

No	Komponen Penelitian Terdahulu				
	Judul Penelitian	Penulis	Metode	Data	Hasil
1	Uji Perbandingan Akurasi Analisis Sentimen Pariwisata menggunakan Algoritma <i>Support Vector Machine</i> dan <i>Naive Bayes</i>	(Susanti, Sedyono, and Sembiring, 2016)	Algoritma <i>Support Vector Machine</i> dan <i>Naive Bayes</i>	Pengambilan data menggunakan Foursquare yang diberikan sentimen positif dan negatif secara manual.	Pada penelitian ini menganalisis opini masyarakat pada suatu tempat wisata di Jawa Tengah yaitu mengenai Candi Prambanan dengan <i>dataset</i> sebanyak 350 opini memiliki akurasi dengan metode <i>Naive Bayes</i> sebesar 65.78%, presisi sebesar 65.97% dan <i>recall</i> sebesar 96.39%. Data metode SVM akurasi sebesar 76.48%, presisi sebesar 87.25%, dan <i>recall</i> sebesar 80.60%. Secara garis besar menandakan bahwa nilai dengan metode SVM lebih baik daripada metode <i>Naive Bayes</i> dengan selisih 10.96%.
2	Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi <i>Naive Bayes</i> dan <i>Support Vector Machine</i>	(Dwianto and Sadikin, 2021)	Algoritma <i>Support Vector Machine</i> dan <i>Naive Bayes</i>	Pengambilan data menggunakan data Twitter yang diberikan sentimen positif dan negatif secara manual.	Pada penelitian ini data yang digunakan berasal dari Twitter dengan <i>dataset</i> 1000 tweet GojekIndonesia dan 1000 tweet GrabID. Hasil analisis sentimen dengan metode <i>Naive Bayes</i> menghasilkan <i>class precision</i> sebesar 66.57%, <i>class recall</i> sebesar 57.14%, <i>accuracy</i> sebesar 84.08% dan SVM adalah <i>class precision</i> sebesar 54.78%, <i>class recall</i> sebesar 52.46%, <i>accuracy</i> sebesar 69.50%. Disimpulkan SVM lebih baik dibandingkan <i>Naive Bayes</i> .
3	Analisis Sentimen Perusahaan Listrik Negara Cabang	(Tuhuteru and Iriani, 2018)	<i>Support Vector Machine</i> dan <i>Naive Bayes</i>	Pengambilan data menggunakan data Twitter dan	Pada sentimen ini menggunakan data <i>training</i> sebanyak 491 yang kelasnya dibagi secara manual.

Tabel 2. Lanjutan.

No	Komponen Penelitian Terdahulu				
	Judul Penelitian	Penulis	Metode		Hasil
	Ambon Menggunakan Metode <i>Support Vector Machine</i> dan <i>Naive Bayes</i>			pembagian data dilakukan 8 kali dengan nilai <i>fold</i> pada <i>cross validation</i> yang sama.	Berdasarkan hasil pengujian metode SVM memiliki tingkat akurasi sebesar 81.67%, sedangkan <i>Naive Bayes</i> memiliki tingkat akurasi sebesar 67.20%. Dapat disimpulkan metode SVM lebih unggul dari metode <i>Naive Bayes</i> .
4	Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma <i>Support Vector Machine</i> dan <i>Naive Bayes</i>	(Setiawan, Utami, and Sudarmawan, 2021)	<i>Support Vector Machine</i> dan <i>Naive Bayes</i>	Pengambilan data menggunakan data Twitter dan pembagian data dengan 5 <i>fold</i> atau 5 mekanisme pengujian	Pada penelitian ini menggunakan dataset sebanyak 350 data tweet. Hasil pengujian pada penelitian ini dengan menggunakan <i>Naive Bayes</i> menghasilkan nilai akurasi sebesar 81.20%, waktu 9.00 detik, <i>recall</i> sebesar 79.60%, dan presisi sebesar 79,40%. Sedangkan SVM menghasilkan nilai akurasi sebesar 85%, waktu 31.60 detik, <i>recall</i> sebesar 84%, dan presisi sebesar 83,60%. Disimpulkan hasil terbaik dan lebih unggul dengan metode SVM.
5	Analisis Sentimen Opini Terhadap Vaksin Covid-19 pada Media Sosial Twitter Menggunakan <i>Support Vector Machine</i> dan <i>Naive Bayes</i>	(Fitriana, Utami, and Al Fatta, 2021)	<i>Support Vector Machine</i> dan <i>Naive Bayes</i>	Pengambilan data menggunakan data Twitter dan pembagian data dengan 5 <i>fold</i> atau 5 mekanisme pengujian	Data yang digunakan pada penelitian ini sebesar 1000 data tweet <i>keyword</i> vaksin. Hasil perbandingan performa dengan metode SVM memiliki akurasi sebesar 90.47%, dengan presisi sebesar 90.23%, dan <i>recall</i> sebesar 90.78%. Sedangkan <i>Naive Bayes</i> memiliki akurasi sebesar 88.64%, presisi, sebesar 87.32%, dan <i>recall</i> sebesar 88.13%. Berdasarkan hasil tersebut performa SVM lebih unggul dibandingkan <i>Naive Bayes</i> .

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada bulan Februari 2022 hingga bulan Oktober 2022. Penelitian ini dilakukan di Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung yang berada di Jalan Soemantri Brojonegoro No. 1 Gedung Meneng, Bandar Lampung.

Berikut adalah ringkasan dari alokasi penelitian ini sebagai berikut:

Kegiatan	Jadwal Pelaksanaan Penelitian																																							
	Februari				Maret				April				Mei				Juni				Juli				Agustus				September				Oktober							
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4				
Studi Literatur dan Penyusunan Bab I-III																																								
Pengumpulan data																																								
Seminar Usul																																								
<i>K-Fold Cross Validation</i>																																								
Klasifikasi SVM dan <i>Naive Bayes</i>																																								
Penyusunan Bab IV-V																																								
Seminar Hasil																																								

Gambar 3. Perencanaan Waktu Penelitian.

3.2 Perangkat Penelitian

Perangkat penelitian sebagai alat pendukung yang digunakan dalam penelitian ini adalah sebagai berikut.

3.2.1 Perangkat Keras (*Hardware*)

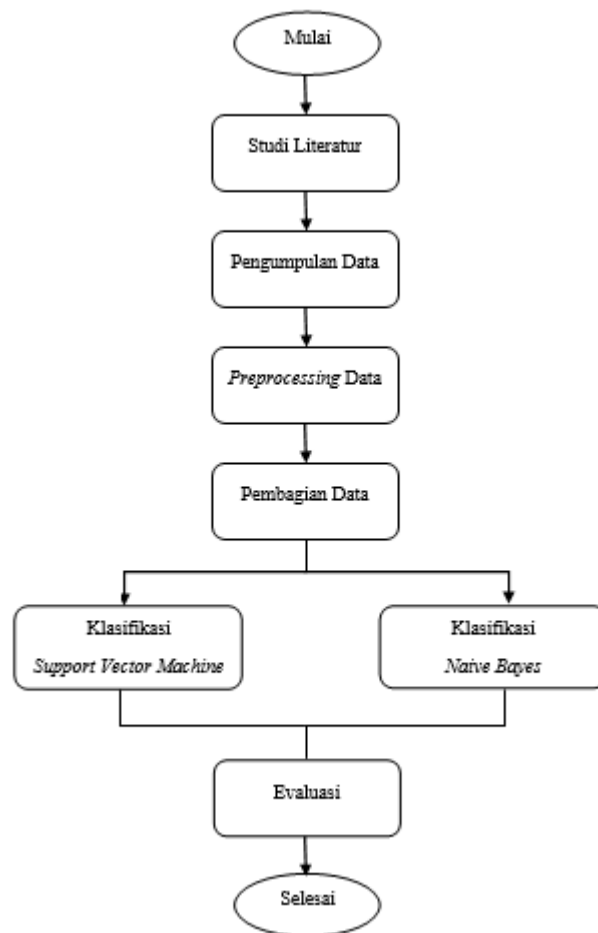
Laptop Acer Aspire E 14 dengan spesifikasi RAM 4 GB, SSD 120GB, dan *Processor Intel Core i5 7th Gen*.

3.2.2 Perangkat Lunak (*Software*)

- a. Sistem Operasi *Windows 10 Pro* 64-bit sebagai sistem operasi laptop.
- b. *Anaconda Navigator*, *Jupyter Notebook 3.2*, dan *Microsoft Excel* sebagai pendukung perangkat lunak dalam proses analisis sentimen.

3.3 Tahapan Penelitian

Tahapan penelitian ini terdapat pada gambar dibawah ini:



Gambar 4. Tahapan Penelitian.

3.3.1 Studi Literatur

Pada penelitian ini, tahap pertama yang dilakukan adalah studi literatur untuk mencari referensi dan pembelajaran dari berbagai jurnal, buku, maupun internet. Studi literatur ini bertujuan membantu mendapatkan solusi dari permasalahan yang akan dihadapi. Informasi yang dicari berkaitan dengan teori dari analisis sentimen opini masyarakat terhadap pariwisata Lampung di Twitter menggunakan klasifikasi *Support Vector Machine* dan *Naive Bayes*. Studi literatur juga berkaitan dengan penelitian terdahulu yang digunakan sebagai referensi dan perbandingan dalam penelitian ini dapat dilihat pada Tabel 2.

3.3.2 Pengumpulan Data

Tahap kedua penelitian ini adalah pengumpulan data *tweet* langsung yang diambil melalui data teks dari Twitter dengan proses menggunakan *library snsrape* berupa data format json yang dikonversikan menjadi file csv. Data teks berupa *tweet* yang dengan *keyword* pariwisata Lampung. Data yang dikumpulkan sebanyak 900 data teks yang telah mendapatkan nilai sentimen.

Penambahan nilai sentimen positif pada data teks dilakukan secara manual yaitu kalimat yang memiliki unsur positif, bersifat membangun, dan tidak ada unsur kebencian. Penambahan nilai sentimen negatif yaitu kalimat yang memiliki unsur negatif dan mengkritik yang mengarah menjatuhkan. Hasil dari pemberian nilai sentimen setiap tahunnya pada tiga tahun terakhir yaitu pada 29 April 2019 sampai 28 April 2022. Pada saat pengumpulan data, selanjutnya dapat dilakukan pengolahan data agar menjadi sebuah data yang akan memudahkan pada saat proses analisis sentimen.

Berikut merupakan tahap pengumpulan data.

- Membuka Anaconda Navigator
- Membuka text editor Jupyter Notebook kemudian mengetik “pip install snsrape”
- Selanjutnya melakukan proses *scraping* dengan mengisi kata kunci atau *keyword* Pariwisata Lampung dengan waktu yang diperlukan
- Kemudian secara otomatis file tersimpan dalam folder dengan format json.

Berikut merupakan proses dari *scraping* pengumpulan data dapat dilihat pada Gambar 5.

Mengambil Data dari Twitter

```
In [27]: # Membuat variabel kosong untuk menyimpan data
tweets = []

• 29 April 2019 - 28 April 2020

In [3]: # Mengambil data dari tanggal 29 April 2019 sampai 28 April 2020 dengan keyword "pariwisata Lampung"
for i,tweet in enumerate(sntwitter.TwitterSearchScrapper('pariwisata lampung since:2019-04-29 until:2020-04-28').get_items()):
    if tweet.content not in tweets:
        tweets.append(tweet.content)

• 29 April 2020 - 28 April 2021

In [16]: # Mengambil data dari tanggal 29 April 2020 sampai 28 April 2021 dengan keyword "pariwisata Lampung"
for i,tweet in enumerate(sntwitter.TwitterSearchScrapper('pariwisata lampung since:2020-04-29 until:2021-04-28').get_items()):
    if tweet.content not in tweets:
        tweets.append(tweet.content)

• 29 April 2021 - 28 April 2022

In [29]: # Mengambil data dari tanggal 29 April 2021 sampai 28 April 2022 dengan keyword "pariwisata Lampung"
for i,tweet in enumerate(sntwitter.TwitterSearchScrapper('pariwisata lampung since:2021-04-29 until:2022-04-28').get_items()):
```

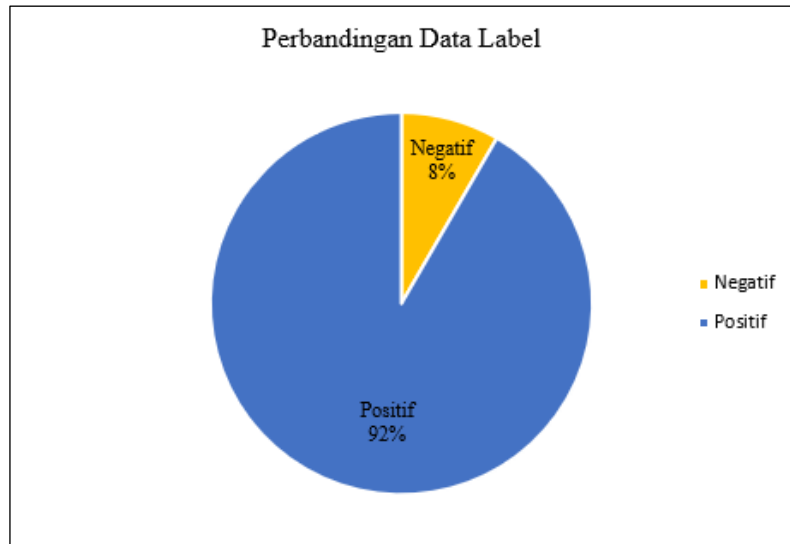
Gambar 5. Pengumpulan Data.

Data yang telah dikumpulkan dan diberi nilai sentimen dapat dilihat pada Tabel 3.

Tabel 3. Pengumpulan Data Twitter.

Kelas	Banyak Data
Positif	826
Negatif	74

Pemberian sentimen secara manual menghasilkan 826 *tweet* positif dan 74 *tweet* negatif atau menghasilkan sebesar 92% sentimen positif dan 8% sentimen negatif. Perbandingan data label dapat dilihat pada Gambar 6.



Gambar 6. Hasil data *labeling manual*.

Pengumpulan data dilakukan pada 29 April 2019 sampai 28 April 2022 untuk kata kunci yaitu ‘Pariwisata Lampung’. Berikut merupakan data yang dikumpulkan yang dapat dilihat pada Gambar 7.

Text	Sentiment
0:episode baru untuk podkes superioritas ilusif pengembangan masyarakat desa pariwisata bahari dan tsunami lampung;	1
1:selamat siang hari ini dilaksanakan pelantikan dan pengambilan sumpah pejabat pengawas di lingkungan dinas pariwisata dan ekonomi kreatif provinsi lampung;	1
2:kami juga pegiat pariwisata di lampung kena imbas wabah virus corona;	0
3:kami tidak sedang berperang melawan di tempat tempat wisata;	0
4:selamat malam himbauan sekretaris dinas pariwisata dan ekonomi kreatif provinsi lampung berdasarkan instruksi bapak gubernur arinal djunaedi dan ibu wakil gubernu;	
5:selamat malam video ajakan bapak kepala dinas pariwisata dan ekonomi kreatif provinsi lampung mengajak masyarakat lampung untuk selalu menjaga kesehatan dan	
6:pasti banyak yang tahu apa aja sih pariwisata yang ada di lampung;	1
7:melakukan penyemprotan desinfektan di wisata pemandian way belerang kalianda lampung selatan tindakan ini dilakukan dalam rangka mencegah penyebaran corona	
8:dinas pariwisata pesisir barat ikuti rapat terkait usaha pencegahan dan kesiapsiagaan covid melalui vicon dengan dinas pariwisata kab kota se provinsi lampung sehub	
9:selamat siang ini sedang berlangsung teleconference antara kepala dinas pariwisata dan ekonomi kreatif provinsi lampung dengan seluruh kepala dinas pariwisata se	
10:ni di wisata lampung kurang bagus pelayanannya;	0
11:from disporapar lamar himbauan kepada seluruh pelaku dan pengelola destinasi pariwisata hotel penginapan objek wisata rumah makan restoran cafe tempat hiburan	
12:rapat pembahasan konsultan masterplan dan feasibility study kawasan terintegrasi pariwisata bakauheni di ruang rapat kepala bappeda provinsi lampung kamis;	1
13:dirgahayu lampungku mari bersinergi untuk lampung yang berjaya maju terus pariwisata lampung yang berjaya;	1
14:kepala dinas kepemudaan olahraga dan pariwisata kabupaten lampung barat beserta jajaran mengucapkan dirgahayu provinsi lampung mari tingkatkan daya saing d	
15:assalamualaikum waarahmatullahi wabarakatuh selamat siang terkait dengan himbauan gubernur lampung bupati lampung barat dan kebijakan sektor pariwisata lai	
16:buka lampung wedding fair ibu riana sari arinal bidik pengembangan pariwisata dan ekonomi kreatif melalui wedding expo;	1
17:jihan mengatakan dalam rapat kerja raker dengan menteri pariwisata sudah menyampaikan hal serupa yakni agar lampung menjadi prioritas utama pembangunan ui	

Gambar 7. *Dataset* yang digunakan pada penelitian.

Hasil analisis kata-kata positif berupa *tweet* yang berkaitan dengan pariwisata Lampung dapat dilihat pada Tabel 4.

Tabel 4. Sentimen Positif.

<i>Tweet</i>	Jumlah <i>Tweet</i>	Nilai Sentimen
Pengembangan pariwisata bahari	60	Positif
Pelantikan dan pengambilan sumpah pejabat pengawas lingkungan	4	Positif
Pariwisata Lampung kena imbas covid-19	34	Positif
Himbauan menjaga kesehatan kepada masyarakat	57	Positif
Pengenalan pariwisata Lampung	41	Positif
Penyemprotan desinfektan	5	Positif
Himbauan pencegahan covid-19	32	Positif
Mendukung pariwisata Lampung	62	Positif
Himbauan pengelolaan destinasi pariwisata	47	Positif
Icon baru pariwisata Lampung	13	Positif
Dokumentasi kegiatan pariwisata Lampung	43	Positif
Penyambutan pariwisata kwartir cabang lampung	3	Positif
Strategi pemasaran pariwisata Lampung	44	Positif
Meninjau bandara di Lampung	2	Positif
Pengenalan baju adat Lampung	7	Positif
Pariwisata Lampung masuk ke dalam pariwisata nasional	6	Positif
Pantai favorit Lampung	25	Positif
Menjadikan ibu kota pariwisata, kota teraman, dan kota pendidikan berkarakter	16	Positif
Peminat pariwisata Lampung	64	Positif
Kunjungan wisatawan pada <i>smart city</i> Bandar Lampung	9	Positif
Keseruan acara yang diadakan pariwisata Lampung	27	Positif
Memajukan pariwisata	35	Positif
Beberapa hotel di Lampung	13	Positif
Perencanaan kawasan pasar seni di pkor way halim	10	Positif
Wisata kuliner Lampung	15	Positif
Membangun infrastruktur jalan zona pariwisata	29	Positif
Kawasan pariwisata di bakauheni	16	Positif
Respon baik masyarakat terhadap destinasi wisata Lampung	37	Positif
Penanaman mangrove dan bersih pantai	11	Positif
Unsur pariwisata Lampung di lukisan mural underpass unila	2	Positif

Tabel 4. Lanjutan.

<i>Tweet</i>	Jumlah <i>Tweet</i>	Nilai Sentimen
Pengelolaan limbah	2	Positif
Kain tapis Lampung	5	Positif
Fasilitas investasi perdagangan dan pariwisata	24	Positif
Peningkatan sumber daya pariwisata dan ekonomi kreatif	26	Positif

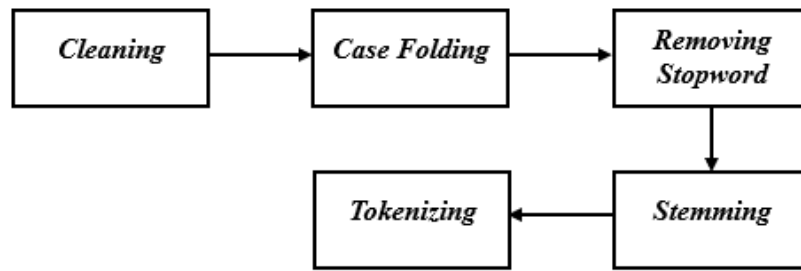
Hasil analisis kata-kata positif berupa *tweet* yang berkaitan dengan pariwisata Lampung dapat dilihat pada Tabel 5.

Tabel 5. Sentimen Negatif.

<i>Tweet</i>	Jumlah <i>Tweet</i>	Nilai Sentimen
Kurangnya soal pelayanan wisata	10	Negatif
Fasilitas tempat wisata Lampung	11	Negatif
Tidak ada pengaruh promosi acara yang dilakukan	7	Negatif
Dibiarkan wisata bahari	5	Negatif
Perawatan yang kurang terhadap destinasi wisata	9	Negatif
Banyaknya sampah di sekitar tempat wisata	6	Negatif
Terlalu banyak pengunjung	3	Negatif
Terdapat calo tiket di tempat wisata	1	Negatif
Terkendala sinyal	2	Negatif
Kurang memadai infrastruktur jalan	2	Negatif
Tiket mahal di tempat wisata	1	Negatif
Macet di tempat wisata pada saat hari libur	2	Negatif
Kebun binatang kurang terawat	3	Negatif
Kekurangan air	1	Negatif
Janji mengembangkan sektor pariwisata	2	Negatif
Tidak mematuhi protokol kesehatan	6	Negatif
Begal	3	Negatif

3.3.3 *Preprocessing Data*

Tahap ketiga dalam penelitian adalah *preprocessing* data untuk membuat data terstruktur agar memudahkan perhitungan dan tahap penelitian yang dilakukan selanjutnya. Berikut merupakan alur *preprocessing* yang dapat dilihat pada Gambar 8.



Gambar 8. Alur *Preprocessing* Data.

Berikut merupakan tahap-tahap *Preprocessing* sebagai berikut:

a. *Cleaning*

Cleaning merupakan proses dalam penghapusan karakter untuk mengurangi kesalahan yang terjadi. *Cleaning* dilakukan untuk menghilangkan *username* (Kode Program 1), *hyperlink* (Kode Program 2), emoji (Kode Program 3), *new line* (Kode Program 4), angka (Kode Program 5), dan simbol (Kode Program 6). *Cleaning* memiliki tujuan dalam mempermudah proses komputasi dan dapat meringankan waktu pengerjaan.

```

tweets['Text'].replace('\n', ' ',
regex=True, inplace=True)
  
```

Kode Program 1. Menghilangkan *New Line*.

```

tweets['Text'].replace(emoji, ' ',
regex=True, inplace=True)
  
```

Kode Program 2. Menghilangkan *Emoji*.

```

tweets['Text'].replace(link, ' ',
regex=True, inplace=True)
tweets['Text'].replace('https', '', regex=
True, inplace=True)
  
```

Kode Program 3. Menghilangkan *Hyperlink*.

```
tweets['Text'].replace(username, '', regex=
True, inplace=True)
```

Kode Program 4. Menghilangkan *Username*.

```
tweets['Text'].replace(number_symbols, '
', regex=True, inplace=True)
```

Kode Program 5. Menghilangkan Angka dan Simbol.

```
tweets['Text'].replace(spaces, ' ',
regex=True, inplace=True)
tweets['Text'] =
tweets['Text'].str.strip()
tweets.drop(tweets.index[tweets['Text'] ==
''], inplace=True)
```

Kode Program 6. Menghilangkan Multi spasi.

b. *Case Folding*

Proses *Case Folding* dilakukan dengan mengubah text menjadi huruf kecil (*lowercase*). Berikut dapat dilihat pada Kode Program 7.

```
tweets['Text'] = tweets['Text'].str.lower()
```

Kode Program 7. Mengubah *Tweet* Menjadi Huruf Kecil.

Berikut merupakan hasil dari proses *case folding* yang dapat dilihat pada Gambar 9.

tweets['Text'][0]
'episode baru untuk podkes superioritas ilusif pengembangan masyarakat desa pariwisata bahari dan tsunami lampung'
tweets['Text'][1]
'selamat siang hari ini dilaksanakan pelantikan dan pengambilan sumpah pejabat pengawas di lingkungan dinas pariwisata dan ekonomi kreatif provinsi lampung'
tweets['Text'][2]
'kami juga pegiat pariwisata di lampung kena imbas wabah virus corona'
tweets['Text'][3]
'kami tidak sedang berperang melawan di tempat tempat wisata'
tweets['Text'][4]
'selamat malam himbauan sekretaris dinas pariwisata dan ekonomi kreatif provinsi lampung berdasarkan instruksi bapak gubernur a rinal djunaedi dan ibu wakil gubernur'

Gambar 9. Hasil Tahap *Case Folding*.

c. *Removing Stopword*

Removing Stopword merupakan proses penghapusan kata-kata yang tidak diperlukan di dalam data yang didapatkan sehingga dapat meningkatkan hasil akurasi yang optimal. Berikut dapat dilihat pada Kode Program 8.

```

stopword_factory = StopWordRemoverFactory()
stopword=
stopword_factory.create_stop_word_remover()
for i,text in enumerate (tweets['Text']):
tweets.loc[i,          'Text']          =
stopword.remove(text)

```

Kode Program 8. Menghilangkan *Stopword*.

Berikut merupakan hasil dari proses *removing stopwords* yang dapat dilihat pada Gambar 10.

tweets['Text'][0]
'episode baru podkes superioritas ilusif pengembangan masyarakat desa pariwisata bahari tsunami lampung'
tweets['Text'][1]
'selamat siang hari dilaksanakan pelantikan pengambilan sumpah pejabat pengawas lingkungan dinas pariwisata ekonomi kreatif provinsi lampung'
tweets['Text'][2]
'pegiat pariwisata lampung kena imbas wabah virus corona'
tweets['Text'][3]
'sedang berperang melawan tempat tempat wisata'
tweets['Text'][4]
'selamat malam himbauan sekretaris dinas pariwisata ekonomi kreatif provinsi lampung berdasarkan instruksi bapak gubernur arinal djunaidi ibu wakil gubernur'

Gambar 10. Hasil Tahap *Removing Stopword*.

d. *Stemming*

Stemming merupakan proses dalam mengubah atau mengembalikan suatu kata menjadi bentuk kata dasar atau memiliki arti dan makna yang sama dengan kata dasar. Berikut dapat dilihat pada Kode Program 9.

```
stemming_factory = StemmerFactory()
stemmer = stemming_factory.create_stemmer()
for i, text in enumerate (tweets['Text']):
    tweets.loc[i, 'Text'] =
stemmer.stem(text)
```

Kode Program 9. Mengubah Kata Menjadi Bentuk Dasar.

Berikut merupakan hasil dari proses *stemming* yang dapat dilihat pada Gambar 11.

```

tweets['Text'][0]
'episode baru podkes superioritas ilusif kembang masyarakat desa pariwisata bahari tsunami lampung'

tweets['Text'][1]
'selamat siang hari laksana lantik ambil sumpah jabat awas lingkung dinas pariwisata ekonomi kreatif provinsi lampung'

tweets['Text'][2]
'giat pariwisata lampung kena imbas wabah virus corona'

tweets['Text'][3]
'sedang perang lawan tempat tempat wisata'

tweets['Text'][4]
'selamat malam himbauan sekretaris dinas pariwisata ekonomi kreatif provinsi lampung dasar instruksi bapak gubernur arinal djun
aidi ibu wakil gubernur'

```

Gambar 11. Hasil Tahap *Stemming*.

e. *Tokenizing dan Vectorizing*

Tokenizing merupakan proses dalam memisahkan kata-kata yang tersusun dalam sebuah kalimat yang akan dipisah menjadi kata tunggal. *Vectorizing* yang dilakukan pada penelitian ini menggunakan *CountVectorizer*. Berikut dapat dilihat pada Kode Program 10.

```

def tokenize(text):
    tknzs = TweetTokenizer()
    return tknzs.tokenize(text)

vectorizer =
CountVectorizer(tokenizer=tokenize)

```

Kode Program 10. *Tokenizing dan Vectorizing*.

Berikut merupakan hasil dari proses *tokenizing* yang dapat dilihat pada Gambar 12.

```

tokenize(tweets['Text'][0])
['episode',
'baru',
'podkes',
'superioritas',
'ilusif',
'kembang',
'masyarakat',
'desa',
'pariwisata',
'bahari',
'tsunami',
'lampung']

```

Gambar 12. Hasil Tahap *Tokenizing*.

Berikut merupakan hasil dari proses *vectorizing* yang dapat dilihat pada Gambar 13.

```
In [67]: x = vectorizer.fit_transform(tweets['Text'])
print(x.toarray()[0:1])
```

Gambar 13. Hasil *Vectorizing*.

3.3.4 Pembagian Data

Pada pembagian data menggunakan *10 – fold cross validation* , tahap ini dibagi menjadi dua yaitu data latih dan data uji. Data latih yang dibagi bertujuan untuk tahapan seleksi sedangkan data uji bertujuan untuk pengujian klasifikasi. Pembagian data dilakukan sebanyak 90% sebagai data latih dan 10% sebagai data uji secara berulang sebanyak 10 kali. Pembagian data latih dan data uji dapat dilihat pada Tabel 6.

Tabel 6. Pembagian Data Latih dan Data Uji.

Fold	Data Latih	Data Uji
1	90 – 899	0 – 89
2	0 -89, 180 - 899	90 – 179
3	0 – 179, 270 - 899	180 – 269
4	0 – 269, 360 – 899	270 – 359
5	0 – 359, 450 - 899	360 - 449
6	0 – 449, 540 - 899	450 - 539
7	0 – 539, 630 - 899	540 - 629
8	0 – 629, 720 - 899	630 - 719
9	0 – 719, 810 - 899	720 - 809
10	0 - 809	810 - 899

Pembagian data latih dan data uji menggunakan 10 – *fold cross validation*. Kode program data latih dan data uji *fold-1* dapat dilihat dengan Kode Program 11.

```
X_train_1 = tweets.loc[90:899, "Text"]
Y_train_1 = tweets.loc[90:899, "Sentiment"]
X_test_1 = tweets.loc[0:89, "Text"]
Y_test_1 = tweets.loc[0:89, "Sentiment"]
```

Kode Program 11. Pembagian Data *fold-1*.

Kode program data latih dan data uji *fold-2* dapat dilihat dengan Kode Program 12.

```
X_train_2 = pd.concat([tweets.loc[0:89,
"Text"], tweets.loc[180:899, "Text"]],
ignore_index=True)
Y_train_2 = pd.concat([tweets.loc[0:89,
"Sentiment"], tweets.loc[180:899,
"Sentiment"]], ignore_index=True)
X_test_2 = tweets.loc[90:179, "Text"]
Y_test_2 = tweets.loc[90:179, "Sentiment"]
```

Kode Program 12. Pembagian Data *fold-2*.

Kode program data latih dan data uji *fold-3* dapat dilihat dengan Kode Program 13.

```
X_train_3 = pd.concat([tweets.loc[0:179,
"Text"], tweets.loc[270:899, "Text"]],
ignore_index=True)
Y_train_3 = pd.concat([tweets.loc[0:179,
"Sentiment"], tweets.loc[270:899,
"Sentiment"]], ignore_index=True)
X_test_3 = tweets.loc[180:269, "Text"]
Y_test_3 = tweets.loc[180:269, "Sentiment"]
```

Kode Program 13. Pembagian Data *fold-3*.

Kode program data latih dan data uji *fold-4* dapat dilihat dengan Kode Program 14.

```
X_train_4 = pd.concat([tweets.loc[0:269,
"Text"], tweets.loc[360:899, "Text"]],
ignore_index=True)
Y_train_4 = pd.concat([tweets.loc[0:269,
"Sentiment"], tweets.loc[360:899,
"Sentiment"]], ignore_index=True)
X_test_4 = tweets.loc[270:359, "Text"]
Y_test_4 = tweets.loc[270:359, "Sentiment"]
```

Kode Program 14. Pembagian Data *fold-4*.

Kode program data latih dan data uji *fold-5* dapat dilihat dengan Kode Program 15.

```
X_train_5 = pd.concat([tweets.loc[0:359,
"Text"], tweets.loc[450:899, "Text"]],
ignore_index=True)
Y_train_5 = pd.concat([tweets.loc[0:359,
"Sentiment"], tweets.loc[450:899,
"Sentiment"]], ignore_index=True)
X_test_5 = tweets.loc[360:449, "Text"]
Y_test_5 = tweets.loc[360:449, "Sentiment"]
```

Kode Program 15. Pembagian Data *fold-5*.

Kode program data latih dan data uji *fold-6* dapat dilihat dengan Kode Program 16.

```
X_train_6 = pd.concat([tweets.loc[0:449,
"Text"], tweets.loc[540:899, "Text"]],
ignore_index=True)
Y_train_6 = pd.concat([tweets.loc[0:449,
"Sentiment"], tweets.loc[540:899,
"Sentiment"]], ignore_index=True)
X_test_6 = tweets.loc[450:539, "Text"]
Y_test_6 = tweets.loc[450:539, "Sentiment"]
```

Kode Program 16. Pembagian Data *fold-6*.

Kode program data latih dan data uji *fold-7* dapat dilihat dengan Kode Program 17.

```
X_train_7 = pd.concat([tweets.loc[0:539,
"Text"], tweets.loc[630:899, "Text"]],
ignore_index=True)
Y_train_7 = pd.concat([tweets.loc[0:539,
"Sentiment"], tweets.loc[630:899,
"Sentiment"]], ignore_index=True)
X_test_7 = tweets.loc[540:629, "Text"]
Y_test_7 = tweets.loc[540:629, "Sentiment"]
```

Kode Program 17. Pembagian Data *fold-7*.

Kode program data latih dan data uji *fold-8* dapat dilihat dengan Kode Program 18.

```
X_train_8 = pd.concat([tweets.loc[0:629,
"Text"], tweets.loc[720:899, "Text"]],
ignore_index=True)
Y_train_8 = pd.concat([tweets.loc[0:629,
"Sentiment"], tweets.loc[720:899,
"Sentiment"]], ignore_index=True)
X_test_8 = tweets.loc[630:719, "Text"]
Y_test_8 = tweets.loc[630:719, "Sentiment"]
```

Kode Program 18. Pembagian Data *fold-8*.

Kode program data latih dan data uji *fold-9* dapat dilihat dengan Kode Program 19.

```
X_train_9 = pd.concat([tweets.loc[0:719,
"Text"], tweets.loc[810:899, "Text"]],
ignore_index=True)
Y_train_9 = pd.concat([tweets.loc[0:719,
"Sentiment"], tweets.loc[810:899,
"Sentiment"]], ignore_index=True)
X_test_9 = tweets.loc[720:809, "Text"]
Y_test_9 = tweets.loc[720:809, "Sentiment"]
```

Kode Program 19. Pembagian Data *fold-9*.

Kode program data latih dan data uji *fold-10* dapat dilihat dengan Kode Program 20.

```
X_train_5 = tweets.loc[0:809, "Text"]
Y_train_5 = tweets.loc[0:809, "Sentiment"]
X_test_5 = tweets.loc[810:899, "Text"]
Y_test_5 = tweets.loc[810:899, "Sentiment"]
```

Kode Program 20. Pembagian Data *fold-10*.

3.3.5 Klasifikasi *Support Vector Machine*

Klasifikasi menggunakan algoritma *Support Vector Machine*. Tahap ini dilakukan ketika semua data sudah terkumpul yang diambil dari *tweet* di Twitter dan penerapannya dalam pengelompokan sebuah teks. Pengujian SVM menggunakan *kernel* yang digunakan yaitu *kernel Radial Basis Function*. Indikator parameter yang dilakukan pada SVM, diantaranya:

- a. Nilai akurasi (*accuracy*);
- b. Nilai presisi (*precision*);
- c. Nilai kemampuan kembali suatu data (*recall*);
- d. *F1-score*.

3.3.6 Klasifikasi *Naive Bayes*

Klasifikasi menggunakan algoritma *Naive Bayes*. Tahap ini dilakukan ketika semua data sudah terkumpul yang diambil dari *tweet* di Twitter dan penerapannya dalam sebuah model yang berdasarkan teorema *Bayes*. Pengujian *Naive Bayes* menggunakan model yang digunakan yaitu *Complement Naive Bayes*. Sama halnya seperti SVM, Indikator parameter yang dilakukan pada *Naive Bayes*, diantaranya:

- a. Nilai akurasi (*accuracy*);
- b. Nilai presisi (*precision*);
- c. Nilai kemampuan kembali suatu data (*recall*);
- d. *F1-score*.

3.3.7 Evaluasi

Pada tahap terakhir ini menampilkan perbandingan hasil yang dilakukan dengan algoritma SVM dan *Naive Bayes*. Evaluasi menggunakan *confusion matrix* yang mempresentasikan prediksi yang benar, kondisi dari data yang dihasilkan dan menghitung kualitas dari prediksi algoritma tersebut.

V. SIMPULAN DAN SARAN

5.1 Simpulan

Pada penelitian ini didapatkan hasil dan diambil simpulan adalah sebagai berikut:

- a. Analisis sentimen terhadap pariwisata Lampung berhasil dilakukan dengan menggunakan *Support Vector Machine* dan *Naive Bayes*. Sebanyak 92% opini masyarakat terhadap pariwisata Lampung termasuk kedalam sentimen positif dan 8% termasuk kedalam sentimen negatif.
- b. Hasil akurasi terbesar didapatkan oleh *Support Vector Machine* (SVM) dengan rata-rata akurasi sebesar 94,11%, presisi sebesar 95,72%, *recall* sebesar 97,90%, dan *f1-score* sebesar 96,77%. *Naive Bayes* menghasilkan rata-rata akurasi sebesar 93,78%, presisi sebesar 95,91%, *recall* sebesar 97,30%, dan *f1-score* sebesar 96,59%. Performansi rata-rata waktu komputasi yang dihasilkan algoritma *Naive Bayes* selama 0,771 detik sedangkan SVM selama 14,492 detik. Hal ini membuktikan bahwa algoritma *Naive Bayes* lebih cepat daripada SVM.

5.2 Saran

Pada penelitian ini terdapat saran yang diberikan untuk penelitian selanjutnya adalah sebagai berikut:

- a. Mengembangkan analisis sentimen dengan menggunakan metode klasifikasi lainnya seperti *Deep Learning* seperti *Recurrent Neural Network* (RNN) untuk dikombinasikan dan didapatkan nilai akurasi yang lebih baik dari penelitian ini.

- b. Menambahkan *dataset* yang lebih baik. Tidak hanya mengambil data dari Twitter tetapi dapat mengambil data yang lainnya seperti Instagram dan Facebook agar dapat meningkatkan performa dari klasifikasi analisis sentimen yang dilakukan.
- c. Analisis data tidak dilakukan secara manual dan bahasa yang digunakan tidak hanya bahasa Indonesia.

DAFTAR PUSTAKA

- Amalia, Bintang Sifa, Yuyun Umidah, And Rini Mayasari. 2021. "Analisis Sentimen Review Pelanggan Restoran Menggunakan Algoritma Support Vector Machine Dan K-Nearest Neighbor." *SITEKIN: Jurnal Sains, Teknologi Dan Industri* 19(1): 28–34.
- Ardianto, Rian. 2020. "Sentiment Analysis On E-Sports For Education Curriculum Using Naive Bayes And Support Vector Machine." *Jurnal Ilmu Komputer Dan Informasi* 13(2): 109–22.
- Darmawan, Agus, Nunu Kustian, And Wanti Rahayu. 2018. "Implementasi Data Mining Menggunakan Model SVM Untuk Prediksi Kepuasan Pengunjung Taman Tabebuya." *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)* 2(3): 299.
- Dwianto, Enos, And Mujiono Sadikin. 2021. "Analisis Sentimen Transportasi Online Pada Twitter Menggunakan Metode Klasifikasi Naïve Bayes Dan Support Vector Machine." *Format : Jurnal Ilmiah Teknik Informatika* 10(1): 94.
- Fahmi, Ardan Et Al. 2020. "Analisis Sentiment Masyarakat Selama Bulan Ramadhan Dalam Menghadapi Pandemi Covid-19." *Jurnal Informatika Dan Sistem Informasi (Jifosi)* 1(1): 608–17.
- Fauzan, Ivan. 2020. "Artificial Intelligence (AI) Pada Proses Pengawasan Dan Pengendalian Kepegawaian - Sebuah Eksplorasi Konsep Setelah Masa Pandemi Berakhir." *Jurnal Civil Service* 14(1): 31–42.
- Fikri, Mujaddid Izzul, Trifebi Shina Sabrila, And Yufis Azhar. 2020. "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Twitter." *Smatika Jurnal* 10(02): 71–76.

- Fitriana, Frizka, Ema Utami, And Hanif Al Fatta. 2021. "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 Pada Media Sosial Twitter Menggunakan Support Vector Machine Dan Naive Bayes." *Jurnal Komtika (Komputasi Dan Informatika)* 5(1): 19–25.
- Fitriyah, Nur, Budi Warsito, And Di Asih I Maruddani. 2020. "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm." *Jurnal Gaussian* 9(3): 376–90.
- Husna, Rizqa El. 2020. "Analisis Klasifikasi Sentimen Pada Twitter Mengenai Netflix Yang Diblokir Oleh Telkom Menggunakan Naïve Bayes Classifier Dan Support Vector Machine Jurnal Ilmiah." [Http://Repository.Unimus.Ac.Id](http://Repository.Unimus.Ac.Id).
- Kristiyanti, Dinar Ajeng Et Al. 2019. "Comparison Of SVM Naïve Bayes Algorithm For Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based On Public Opinion On Twitter." *2018 6th International Conference On Cyber And IT Service Management, CITSM 2018* (Citsm 2018): 1–6.
- Mardihartono, Agus, And Yuli Evadianti. 2019. "New Media Sebagai Sarana Promosi Pariwisata Lampung." *Seminar Nasional Fisip Unila (Sefila) - 3 Tahun 2019*.
- Nasution, Muhammad Rangga Aziz, And Mardhiya Hayaty. 2019. "Perbandingan Akurasi Dan Waktu Proses Algoritma K-NN Dan SVM Dalam Analisis Sentimen Twitter." *Jurnal Informatika* 6(2): 226–35.
- Nofiyanti, Endah, And Erry Maricha Oki Nur Haryanto. 2021. "Analisis Sentimen Terhadap Penanggulangan Bencana Di Indonesia." *Jurnal Ilmiah SINUS* 19(2): 17.
- Peryanto, Ari, Anton Yudhana, And Rusydi Umar. 2020. "Klasifikasi Citra Menggunakan Convolutional Neural Network Dan K Fold Cross Validation." *Journal Of Applied Informatics And Computing* 4(1): 45–51.
- Putri, Nadya Berliana, And Nada Arina Romli. 2021. "Analisis Dampak Adiksi Internet Pada Media Sosial Twitter Di Indonesia Dengan Pendekatan Teori Komunikasi." *Jurnal Komunikasi Universitas Garut: Hasil Pemikiran Dan Penelitian* 7(1): 582.

- Que, Valentino Kevin Sitanayah, Ade Iriani, And Hindriyanto Dwi Purnomo. 2020. "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization." *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi* 9(2): 162–70.
- Santoso, Imam, Windu Gata, And Atik Budi Paryanti. 2019. "Penggunaan Feature Selection Di Algoritma Support Vector Machine Untuk Sentimen Analisis Komisi Pemilihan Umum." *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)* 3(3): 364–70.
- Santoso, Valonia Inge, Gloria Virginia, And Yuan Lukito. 2017. "Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine." *Jurnal Transformatika* 14(2): 72.
- Setiawan, Hendrik, Ema Utami, And Sudarmawan. 2021. "Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine Dan Naive Bayes." *Jurnal Komtika (Komputasi Dan Informatika)* 5(1): 43–51.
- Sodik, Fajar, And Iqbal Kharisudin. 2021. "Analisis Sentimen Dengan SVM , NAIVE BAYES Dan KNN Untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 Pada Media Sosial Twitter." *Prisma* 4: 628–34.
- Sumastuti, Efriyani, Heri Prabowo, And Kristin Violinda. 2021. "Pengembangan Wisata Kota Semarang." *Khasanah Ilmu - Jurnal Pariwisata Dan Budaya* 12(1): 30–38.
- Susanti, Novita Dewi, Eko Sedyono, And Irwan Sembiring. 2016. "Uji Perbandingan Akurasi Analisis Sentimen Pariwisata Menggunakan Algoritma Support Vektor Machine Dan Naive Bayes." *Nusantara Of Engineering* 3(2): 26–33.
- Tineges, Rian, Agung Triayudi, And Ira Diana Sholihati. 2020. "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)." *Jurnal Media Informatika Budidarma* 4(3): 650.

- Tuhuteru, Hennie, And Ade Iriani. 2018. “Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine Dan Naive Bayes Classifier.” *Jurnal Informatika: Jurnal Pengembangan IT* 3(3): 394–401.
- Utami, Dian Siti, And Adhitia Erfina. 2021. “Analisis Sentimen Pinjaman Online Di Twitter Menggunakan Algoritma Support Vector Machine (SVM).” *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)* 1(1): 299–305.
- Widowati, Tanthy Tawaqalia, And Mujiono Sadikin. 2021. “Analisis Sentimen Twitter Terhadap Tokoh Publik Dengan Algoritma Naive Bayes Dan Support Vector Machine.” *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer* 11(2): 626–36.