

**ANALISIS SENTIMEN PEMBERLAKUAN KARANTINA DI MASA
COVID-19 PADA MEDIA SOSIAL TWITTER DI INDONESIA
MENGUNAKAN ALGORITME *SUPPORT VECTOR MACHINE* (SVM)
DAN *LEXICON BASED***

(Skripsi)

Oleh

**MUHAMMAD ARSYI SOBIRIN
1817051066**



**JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2023**

ABSTRAK

ANALISIS SENTIMEN PEMBERLAKUAN KARANTINA DI MASA COVID-19 PADA MEDIA SOSIAL TWITTER DI INDONESIA MENGUNAKAN ALGORITME *SUPPORT VECTOR MACHINE* (SVM) DAN *LEXICON BASED*

Oleh

MUHAMMAD ARSYI SOBIRIN

Corona Virus Disease 2019 atau COVID-19 adalah virus yang sangat berbahaya karena tingkat penularannya yang tinggi. Virus ini dapat menyebar melalui kontak langsung atau berada di dekat orang yang terinfeksi dan barang yang terkontaminasi sehingga mudah menyebar ke berbagai dunia. Virus COVID-19 membuat sebagian manusia terganggu dalam melakukan aktivitasnya, bahkan musibah ini membuat manusia jatuh sakit dan sampai kehilangan nyawa. Pemberlakuan karantina diterapkan sebagai salah satu upaya untuk mengatasi penyebaran virus COVID-19. Upaya yang dilakukan menuai kritik dan saran dari berbagai kalangan masyarakat khususnya dalam penerapan karantina kesehatan. Dengan berkembangnya teknologi yang ada, masyarakat dapat mengeluarkan pendapatnya di berbagai jejaring media sosial salah satunya Twitter. Pada penelitian ini data yang didapat dari twitter *dipreprocessing* dan dilabeli kedalam kelas positif atau negatif secara manual dan dengan metode *lexicon based*, sedangkan untuk klasifikasinya menggunakan metode *Support Vector Machine* (SVM). Berdasarkan hasil penelitian yang dilakukan didapatkan hasil dari klasifikasi algoritme *Support Vector Machine* (SVM) dengan pelabelan secara manual menunjukkan nilai klasifikasi yang lebih baik dengan percobaan *oversampling* pembagian data *training* 90% dan *testing* 10% didapatkan hasil akurasi sebesar 97,36% dibandingkan dengan pelabelan *InSetLexicon* sebesar 97% .

Kata kunci: Analisis Sentimen, Karantina, *Support Vector Machine*, *Lexicon Based*.

ABSTRACT

SENTIMENT ANALYSIS OF QUARANTINE IMPLEMENTATION DURING COVID-19 ON TWITTER SOCIAL MEDIA IN INDONESIA USING SUPPORT VECTOR MACHINE (SVM) AND LEXICON BASED ALGORITHMS

By

MUHAMMAD ARSYI SOBIRIN

Corona Virus Disease 2019 or COVID-19 is a very dangerous virus because of its high transmission rate. This virus can be spread by direct contact or by being near infected people and contaminated items so that it can easily spread to various parts of the world. The COVID-19 virus has disturbed some humans in carrying out their activities, and this disaster has even made humans fall ill and even lose their lives. Quarantine implementation is implemented as an effort to overcome the spread of the COVID-19 virus. The efforts made have drawn criticism and suggestions from various groups of people, especially in the implementation of health quarantine. With the development of existing technology, people can express their opinions on various social media networks, one of which is Twitter. In this study, the data obtained from Twitter was preprocessed and labeled into positive or negative classes manually and using the lexicon-based method, while the Support Vector Machine (SVM) method was used for classification. Based on the results of the research conducted, the results obtained from the classification of the Support Vector Machine (SVM) algorithm with manual labeling show a better classification value. With the oversampling experiment, the distribution of 90% training data and 10% testing resulted in an accuracy of 97.36% compared to InSetLexicon labeling. accuracy of 97%.

Key words: Sentiment Analysis, Quarantine, Support Vector Machine, Lexicon Based.

**ANALISIS SENTIMEN PEMBERLAKUAN KARANTINA DI MASA
COVID-19 PADA MEDIA SOSIAL TWITTER DI INDONESIA
MENGUNAKAN ALGORITME *SUPPORT VECTOR MACHINE* (SVM)
DAN *LEXICON BASED***

Oleh

**MUHAMMAD ARSYI SOBIRIN
1817051066**

Skripsi

Sebagai salah satu syarat untuk memperoleh gelar
SARJANA KOMPUTER

Pada

Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung



**JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2023**

Judul Skripsi

: **ANALISIS SENTIMEN PEMBERLAKUAN KARANTINA
DI MASA COVID-19 PADA MEDIA SOSIAL TWITTER
DI INDONESIA MENGGUNAKAN ALGORITME
SUPPORT VECTOR MACHINE (SVM) DAN LEXICON
BASED**

Nama Mahasiswa

: **Muhammad Arsyi Sobirin**

Nomor Induk Mahasiswa : **1817051066**

Program Studi

: **Ilmu Komputer**


Fakultas

: **Matematika dan Ilmu Pengetahuan Alam**



Dr. Ir. Kurnia Muludi, M.S.Sc.
NIP. 19640616 198902 1 001

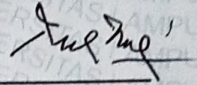
2. **Ketua Jurusan Ilmu Komputer**


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

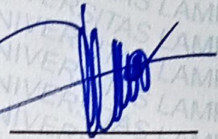
MENGESAHKAN

1. Tim Penguji

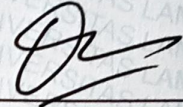
Ketua : Dr. Ir. Kurnia Muludi, M.S.Sc



**Penguji I
Penguji utama : Didik Kurniawan, S.Si., M.T.**



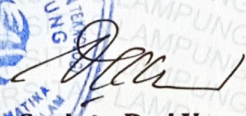
**Penguji II
Penguji Pembahas : Ossy Dwi Endah Wulansari, M.T.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



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



Tanggal Lulus Ujian Skripsi : 02 Februari 2023

PERNYATAAN

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi saya yang berjudul "**Analisi Sentimen Pemberlakuan Karantina Di Masa Covid-19 Pada Media Sosial Twitter Di Indonesia Menggunakan Algoritme *Support Vector Machine (SVM) Dan Lexicon Based***" merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti skripsi saya merupakan hasil salinan atau dibuat orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung, 10 Februari 2023



MUHAMMAD ARSYI SOBIRIN
NPM 1817051066

RIWAYAT HIDUP



Penulis dilahirkan di Bandar Lampung pada tanggal 09 April 2000, sebagai anak kedua dari Ayah Hapizulfikri, S.H., M.M. dan Ibu Melyanah. Penulis menyelesaikan pendidikan formal pertama kali di Pendidikan Sekolah Dasar (SD) di SD Negeri 2 Sukabumi Kota Bandar Lampung diselesaikan pada tahun 2012. Kemudian, pendidikan Sekolah Menengah Pertama (SMP) di MTs Negeri 2 Bandar Lampung diselesaikan pada tahun 2015, dan melanjutkan Sekolah Menengah Atas (SMA) di SMK SMTI Bandar Lampung diselesaikan pada tahun 2018.

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

1. Menjadi anggota Adapter Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2018/2019.
2. Menjadi anggota Magang Bidang Eksternal Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2018/2019.
3. Pada bulan Februari 2021 sampai dengan bulan Maret 2021 penulis melaksanakan Kerja Praktik di Kantor Kesyahbandaran dan Otoritas Pelabuhan Kelas I Panjang.
4. Pada bulan September 2021 penulis melaksanakan kegiatan Kuliah Kerja Nyata (KKN) di desa Sukarame II, kecamatan Teluk Betung Barat, Kota Bandar Lampung.

PERSEMBAHAN

Alhamdulillahirobbilalamin

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

Aku persembahkan karya ini kepada:

Kedua Orang Tuaku Tercinta

Sebagai tanda terimakasihku kepada bapak dan mamahku yang tercinta dan yang tersayang. Terima kasih telah mendidik dan membesarkanku dengan kasih sayang kalian. Terima kasih selalu mendukungku dan mendoakanku dalam segala pilihanku. Terima kasih atas semua pengorbanan dan perjuangan kalian yang tiada hentinya.

Kakakku, Aqmarani Adzani Fitriana, S.Pd.

Terima kasih telah memberikan semangat, dukungan, dan doa.

Seluruh Keluarga Besar Ilmu Komputer 2018, Sahabat, dan Teman-teman
yang selalu memberikan semangat dan dukungan.

Almamater Tercinta, Universitas Lampung

MOTTO

"Sesungguhnya Allah tidak akan mengubah keadaan suatu kaum sebelum mereka mengubah keadaan diri mereka sendiri."

(Q.S Ar-Ra'd: 11)

"Sesungguhnya bersama kesulitan itu ada kemudahan."

(Q.S Al Insyirah: 6)

"Apapun yang menjadi takdirmu, akan mencari jalannya menemukanmu."

(Abi bin Abi Thalib)

"Kesuksesan bukanlah kunci dari kebahagiaan. Sebaliknya kebahagiaan adalah kunci dari kesuksesan."

(Bob Dylan)

"Keep your eyes on the stars and your feet on the ground."

(Theodore Rosevelt)

SANWACANA

Puji syukur kita haturkan pada Allah SWT atas segala berkat-Nya penulis dapat menyelesaikan skripsi yang berjudul “Analisi Sentimen Pemberlakuan Karantina Di Masa Covid-19 Pada Media Sosial Twitter Di Indonesia Menggunakan Algoritme *Support Vector Machine* (SVM) Dan *Lexicon Based*” dengan baik.

Penulis menyadari selesainya skripsi ini tidak terlepas dari partisipasi bimbingan serta bantuan dari berbagai pihak baik secara langsung maupun tidak langsung. Maka kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Kedua Orang tua tercinta, bapak Hapizulfikri dan ibu Melyanah yang selalu memberi dukungan, memotivasi, dan menyemangati penulis selama proses perkuliahan sampai dengan penyusunan skripsi. Semoga Allah SWT selalu menyertai, memberkati, dan memberi kesehatan dan kebahagiaan yang berlimpah.
2. Bapak Dr. Ir. Kurnia Muludi, M.S.Sc. selaku dosen pembimbing utama yang telah meluangkan banyak waktu dan dengan sabar membimbing penulis, serta memberikan banyak dukungan, motivasi, dan dorongan untuk menyelesaikan skripsi ini. Terimakasih juga penulis ucapkan atas kritik dan saran yang membangun sehingga penulisan skripsi ini dapat diselesaikan.
3. Bapak Didik Kurniawan, S.Si., M.T. selaku dosen pembahas satu yang telah memberikan banyak masukan, serta ilmu dan pengetahuan baru yang bermanfaat dalam perbaikan skripsi ini dan selaku Ketua Jurusan Ilmu Komputer Universitas Lampung.
4. Ibu Ossy Dwi Endah Wulansari, M.T. selaku dosen pembahas kedua yang telah memberikan banyak masukan, serta memberikan ide, kritik dan saran yang membangun sehingga penulisan skripsi ini dapat diselesaikan.

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 Dr. rer. nat. Akmal Junaidi, M.Sc. selaku Sekretaris Jurusan Ilmu Komputer Universitas Lampung.
8. Bapak dan Ibu Dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung yang tidak dapat saya sebutkan satu per satu, atas bimbingan dan pengajarannya selama penulis menjadi mahasiswa FMIPA Universitas Lampung.
9. Seluruh Staf dan karyawan Fakultas MIPA Universitas Lampung: Ibu Ade Nora Maela, Bang Zainuddin, Mas Sam, Mas Ardi Novalia, dan lainnya yang tidak bisa penulis sebutkan satu per satu, yang telah membantu segala urusan administrasi penulis.
10. Aqmarani Adzani Fitriana, S.Pd. sebagai kakak yang telah mendoakan, menyemangati, dan mendukung selama ini.
11. Seluruh keluarga besar yang telah memberikan doa, semangat, dukungan dan motivasi yang membangun selama ini.
12. Amara Indah Pancarani yang telah menyemangati dan membantu selama masa perkuliahan dan proses pembuatan skripsi berlangsung.
13. Teman-teman terdekat, Ajul, Nabil, Lita, Rahma, dan Sasya yang telah memberikan banyak dukungan moril, segala bentuk bantuan, dan selalu menemani dari awal perkuliahan.
14. Teman masa SMA yang sampai saat ini saling menyemangati dan mendukung, Naufal, Raka, Rizky, Taufik, Niko, Imam.
15. Keluarga Besar Ilmu Komputer 2018 serta kakak tingkat dan adik tingkat yang tidak bisa penulis sebutkan satu per satu.
16. Almamater tercinta, Universitas Lampung yang sudah memberi banyak wawasan dan pengalaman berharga.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, semoga skripsi ini dapat berguna dan bermanfaat bagi agama, masyarakat, bangsa dan negara, para

mahasiswa, akademisi, serta pihak-pihak lain yang membutuhkan terutama penulis. Saran dan kritik yang bersifat membangun sangat diharapkan.

Akhir kata, penulis mengucapkan terima kasih. Semoga Allah SWT senantiasa memberikan perlindungan dan kebaikan bagi kita semua.

Bandar Lampung, 10 Februari 2023

Muhammad Arsyi Sobirin

NPM. 1817051066

DAFTAR ISI

	Halaman
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	x
DAFTAR KODE PROGRAM	xi
I. PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
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. <i>Text Mining</i>	7
2.2. Analisis Sentimen.....	9
2.3. Karantina	9
2.4. Twitter	11
2.5. Python.....	11
2.6. <i>Word2vec</i>	12
2.7. <i>Split Method</i>	13
2.8. <i>Support Vector Machine</i>	15
2.9. <i>Lexicon Based</i>	17
2.10. Teknik <i>Resampling</i>	18
2.11. Penelitian Terdahulu	18

III. METODELOGI PENELITIAN.....	23
3.1. Tempat dan Waktu Penelitian.....	23
3.2. Alat dan Bahan.....	23
3.3. Tahapan Penelitian.....	25
IV. HASIL DAN PEMBAHASAN	33
4.1. Pengumpulan data	33
4.2. <i>Labeling</i>	34
4.3. <i>Preprocessing</i> Data	39
4.4. Pembagian data	44
4.5. <i>Word embedding</i>	45
4.6. Klasifikasi Support Vector Machine (SVM)	46
4.7. Evaluasi hasil klasifikasi	47
V. SIMPULAN DAN SARAN.....	68
5.1. Simpulan.....	68
5.2. Saran.....	69
DAFTAR PUSTAKA	70

DAFTAR TABEL

Tabel	Halaman
1. Jenis <i>Kernel</i> SVM.....	17
2. Penelitian terdahulu.....	20
3. Proses <i>Casefolding</i>	28
4. Proses <i>Cleansing</i>	29
5. Proses <i>Normalization</i>	29
6. Proses <i>Stopwords</i>	29
7. Proses <i>Stemming</i>	30
8. Proses <i>Tokenizing</i>	30
9. <i>Confusion Matrix</i>	31
10. Contoh pelabelan manual.....	34
11. Hasil pelabelan manual.....	36
12. Contoh pelabelan <i>InSetLexicon</i>	37
13. Hasil pelabelan <i>InSet Lexicon</i>	38
14. Kinerja metode SVM (Pelabelan manual) <i>Imbalanced</i> data.....	50
15. <i>Confusion matrix</i> Pelabelan manual <i>Imbalanced</i> data.....	50
16. Kinerja metode SVM (Pelabelan <i>InSetLexicon</i>) <i>Imbalanced</i> data.....	51
17. <i>Confusion matrix</i> Pelabelan <i>InSetLexicon</i> <i>Imbalanced</i> data.....	51
18. Kinerja metode SVM (Pelabelan manual) <i>Undersampling</i>	53
19. <i>Confusion matrix</i> Pelabelan manual <i>Undersampling</i>	53
20. Kinerja metode SVM (Pelabelan <i>InSetLexicon</i>) <i>Undersampling</i>	54
21. <i>Confusion matrix</i> Pelabelan <i>InSetLexicon</i> <i>Undersampling</i>	54

22. Kinerja metode SVM (Pelabelan manual) <i>Oversampling</i>	56
23. <i>Confusion matrix</i> Pelabelan manual <i>Oversampling</i>	57
24. Kinerja metode SVM (Pelabelan <i>InSetLexicon</i>) <i>Oversampling</i>	57
25. <i>Confusion matrix</i> Pelabelan <i>InSetLexicon</i> <i>Oversampling</i>	58
26. Perbandingan kinerja klasifikasi <i>Support Vector Machine</i> (SVM) <i>Imbalanced data</i>	59
27. Perbandingan kinerja klasifikasi <i>Support Vector Machine</i> (SVM) <i>Undersampling</i>	60
28. Perbandingan kinerja klasifikasi <i>Support Vector Machine</i> (SVM) <i>Oversampling</i>	61

DAFTAR GAMBAR

Gambar	Halaman
1. Contoh Pembagian dataset metode <i>Split</i>	15
2. Metode SVM (Parapat, 2018).....	16
3. Tahapan Penelitian	26
4. Hasil <i>Crawling</i>	33
5. Diagram Pelabelan Manual	37
6. Diagram Pelabelan <i>InSetLexicon</i>	39
7. Diagram Dataset <i>Preprocessing</i>	40
8. Frekuensi Kemunculan Kata	43
9. Tampilan Kata Terpopuler dengan <i>Wordcloud</i>	43
10. Diagram <i>Imbalanced</i> pelabelan manual	49
11. Diagram <i>Imbalanced</i> pelabelan <i>InSetLexicon</i>	49
12. Diagram <i>Undersampling</i> pelabelan manual	52
13. Diagram <i>Undersampling</i> pelabelan <i>InSetLexicon</i>	52
14. Diagram <i>Oversampling</i> pelabelan manual	55
15. Diagram <i>Oversampling</i> pelabelan <i>InSetLexicon</i>	55
16. Hasil perbandingan Akurasi	64
17. Hasil perbandingan Presisi	65
18. Hasil perbandingan <i>Recall</i>	66
19. Hasil perbandingan <i>F1-score</i>	67

DAFTAR KODE PROGRAM

Kode Program	Halaman
1. Tahap <i>Casefolding</i>	40
2. Tahap <i>Cleansing</i>	41
3. Tahap <i>Normalization</i>	41
4. Tahap <i>Stopwords</i> dan <i>Stemming</i>	42
5. Tahap <i>Tokenizing</i>	42
6. Pembagian Data	44
7. <i>Vocab Size Word2vec</i>	45
8. <i>Total Words Word2vec</i>	45
9. Model Linear <i>Support Vector Machine</i>	46
10. <i>Confusion Matrix SVM</i>	47
11. Model RBF <i>Support Vector Machine</i>	47

I. PENDAHULUAN

1.1. Latar Belakang

Belakangan ini dunia sedang dilanda musibah yang sangat mengerikan sehingga membuat sebagian manusia terganggu dalam melakukan aktivitasnya, bahkan musibah ini membuat manusia jatuh sakit dan sampai kehilangan nyawa. Hal ini disebabkan karena adanya virus yang muncul dan menyerang manusia, setelah dilakukan penelitian virus ini disebut *Corona Virus Disease 2019* atau COVID-19.

COVID-19, yang lebih dikenal sebagai *Corona Virus Disease 2019*, pertama kali menyebar di kota Wuhan, China pada akhir tahun 2019. COVID-19 merupakan virus yang sangat berbahaya karena memiliki tingkat penularan yang tinggi. Virus ini dapat menyebar melalui kontak langsung dengan orang yang terinfeksi, barang yang terkontaminasi, atau berada di dekat orang yang terinfeksi, sehingga mudah menyebar ke berbagai tempat di seluruh dunia. Menurut laporan WHO, pada tahun 2020 sudah banyak negara yang melaporkan konfirmasi kasus infeksi virus corona atau COVID-19 (Felicia Watratan et al., 2020).

Setelah berita yang ada mengenai virus corona ini beredar di seluruh dunia, pada akhirnya virus ini mulai memasuki wilayah Indonesia. Virus ini pertama kali terdeteksi di Indonesia pada tanggal 2 Maret 2020 yang dimana 2 orang warga asal Depok, Jawa Barat yang terjangkit virus corona (Susilo et al., 2020). Sejak saat itu jumlah kasus masyarakat Indonesia yang terjangkit virus corona semakin banyak setiap harinya sampai tanggal 26 Oktober 2020 tercatat lebih dari 392 ribu kasus positif dan tingkat kematian sebanyak 13 ribu jiwa (Maharani, 2020). Dengan kondisi seperti ini memberikan dampak yang nyata kepada seluruh masyarakat di Indonesia dan menimbulkan masalah khususnya

pada bidang kesehatan, bidang ekonomi, pendidikan dan lainnya yang ikut terkena dampak dari penyebaran virus corona tersebut.

Dalam meminimalisir penularan virus ini, pemerintah mulai melakukan beberapa upaya yang ada. Upaya-upaya yang dilakukan oleh pemerintahan bertujuan untuk memperlambat tingkat penyebaran virus corona agar dampak negatif yang ditimbulkan dapat dikendalikan. Pemerintahan mengeluarkan kebijakan-kebijakan seperti menghimbau proses belajar di rumah atau daring, bekerja di rumah, kegiatan beribadah di rumah dan menghimbau agar masyarakat tidak berpergian keluar rumah apabila tidak dirasa penting atau lebih dikenal dengan istilah karantina.

Karantina kesehatan adalah tindakan yang efektif untuk mengurangi dampak dari bencana, seperti wabah yang menyebabkan kerugian atau dampak yang luas bagi negara. Dalam UU Tentang Keekarantinaan Kesehatan, diatur mengenai pembatasan keluar masuk suatu daerah yang telah terkena wabah, serta aturan tentang perintah untuk melakukan vaksinasi, isolasi, dan karantina wilayah untuk memutus mata rantai penyebaran wabah.

Namun, upaya yang dilakukan oleh pemerintah tersebut menuai kritik dan saran dari berbagai kalangan masyarakat khususnya dalam penerapan karantina kesehatan. Dengan berkembangnya teknologi yang ada, masyarakat dapat mengeluarkan pendapatnya di berbagai jejaring media sosial. Kebijakan pemerintah tentang karantina menimbulkan berbagai tanggapan dari masyarakat, dengan beberapa orang yang merasa bahwa hal ini penting untuk mengurangi penyebaran COVID-19, sementara yang lain merasa bahwa karantina merugikan karena membatasi gerak mereka. Berbagai tanggapan masyarakat yang beragam dapat ditemukan di berbagai akun Twitter milik masyarakat.

Twitter merupakan salah satu media sosial dengan jumlah pengguna terbesar ke-3 di Indonesia, dan dapat menjadi indikator untuk mengevaluasi kebijakan pemerintah berdasarkan tanggapan masyarakat yang diposting di sana (Corallo et al., 2015). Selain itu, Twitter memiliki keunggulan dalam interaksi antar

pengguna karena sifat terbukanya, yang memungkinkan pengguna mudah menuliskan pendapat mereka dan menandai (*mention*) siapapun tanpa harus membuat *tweet* di halaman fans atau akun pihak yang ditandai (Mahendrajaya et al., 2019).

Dalam kondisi tersebut, maka dilakukan penelitian ini yang bertujuan untuk menganalisis berbagai respon atau asumsi masyarakat terhadap pemberlakuan karantina dalam mengurangi penyebaran penularan virus COVID-19 khususnya di Indonesia dalam media sosial Twitter. Hal ini juga bertujuan untuk mengetahui berbagai respon positif dan negatif dari berbagai masyarakat, sehingga dapat diklasifikasikan dan mendapatkan hasil yang sesuai dengan data yang ada.

Penelitian ini menggunakan menggunakan metode *Support Vector Machine* dan *Lexicon Based*. Alasan dari penggunaan metode *Support Vector Machine* karena metode tersebut dapat menghasilkan akurasi yang tinggi dan banyak diusulkan oleh para peneliti (Keramati & Yousefi, 2011). Untuk penggunaan Metode *Lexicon Based* dapat digunakan untuk mengklasifikasikan sentimen dari setiap opini sehingga kalimat sentimen dapat dikelompokkan menjadi kategori positif atau negatif. Dengan menggunakan metode ini, data kalimat sentimen akan lebih mudah untuk diklasifikasikan (Mahendrajaya et al., 2019).

1.2. Rumusan Masalah

Rumusan dalam penelitian ini yaitu:

- a. Bagaimana pengklasifikasian sentimen positif dan sentimen negatif terhadap data yang didapatkan dari Twitter dengan menggunakan algoritme *Support Vector Machine* dan *Lexicon Based*?
- b. Bagaimana tingkat akurasi data pada analisis sentimen dengan menggunakan algoritme *Support Vector Machine* dan *Lexicon Based*?

1.3. Batasan Masalah

Penelitian ini dibatasi oleh beberapa hal, diantaranya:

- a. Menggunakan studi kasus mengenai karantina di masa pandemic COVID-19 di Indonesia.
- b. Proses pembuatan model menggunakan data *tweet* dan menggunakan Twitter API yang merupakan *library* untuk menarik *tweet*.
- c. Data yang digunakan data *tweet* pada 6 bulan, September 2021 sampai Februari 2022.
- d. Data *tweet* yang digunakan hanya *tweet* yang menyebutkan kata kunci yang berhubungan dengan karantina di masa pandemi COVID-19 di Indonesia.
- e. *Tweet* yang digunakan berbahasa Indonesia.
- f. Algoritme klasifikasi yang digunakan adalah *Support Vector Machine* dan *Lexicon Based*.

1.4. Tujuan Penelitian

Tujuan dari penelitian yang akan dicapai sebagai berikut:

- a. Mengimplementasi algoritme *Support Vector Machine* dan *Lexicon Based* untuk melakukan analisis sentimen data *tweet* tentang Pemberlakuan Karantina di masa pandemi COVID-19 di Indonesia.
- b. Mendapatkan hasil analisa dan informasi terkait sentimen masyarakat pengguna media Twitter Pemberlakuan Karantina di masa pandemi COVID-19 di Indonesia.
- c. Menguji dan membandingkan hasil akurasi, presisi, *recall*, dan *f1-score* algoritme *Support Vector Machine* dan *Lexicon Based* pada analisis sentimen.

1.5. Manfaat Penelitian

Manfaat yang ingin dicapai dalam penelitian ini yaitu:

- a. Untuk membantu pihak-pihak yang membutuhkan informasi mengenai pemberlakuan karantina di masa pandemi COVID-19 di Indonesia.

- b. Untuk mengetahui asumsi masyarakat terkait pemberlakuan karantina menggunakan algoritme klasifikasi *Support Vector Machine* dan *Lexicon Based*.

1.6. Sistematika Penulisan

Penulisan skripsi terbagi menjadi tiga bagian utama, yaitu bagian awal, bagian pokok, dan bagian akhir, berikut uraiannya:

- a. Bagian Awal

Pada bagian awal skripsi, terdapat beberapa hal yang disajikan, yaitu halaman judul, halaman pengesahan, motto dan persembahan, kata pengantar, abstrak, daftar isi, daftar tabel, dan daftar gambar.

- 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, metodologi penelitian dan sistematika penulisan.

BAB II Tinjauan Pustaka

Bab ini berisi mengenai landasan teori yang menjelaskan tentang teori umum yang berkaitan dengan topik, yaitu analisis sentimen. Dan menjelaskan teori program yang berkaitan dengan sistem yang digunakan, yaitu istilah-istilah khusus yang dipakai dalam analisis sentimen.

BAB III Metode Penelitian

Bab ini berisi tentang prosedur atau langkah-langkah yang dilakukan dalam penelitian ini meliputi tahapan penelitian.

BAB IV Hasil dan Pembahasan

Bab ini berisi hasil dari pengambilan data, pemrosesan data, dan evaluasi data dengan menggunakan metode yang telah

dipilih sebagai penentuan hasil akhir akurasi dari analisis data yang telah diperoleh dan diproses.

BAB V Penutup

Bab ini berisi kesimpulan mengenai hasil analisis yang telah didapatkan dan disertai saran yang diberikan oleh peneliti.

c. Bagian Akhir

Bagian akhir skripsi ini mencakup daftar pustaka yang berisi sumber-sumber yang digunakan dalam penulisan skripsi.

II. TINJAUAN PUSTAKA

2.1. *Text Mining*

Text mining didefinisikan sebagai proses pengetahuan adalah penemuan informasi baru yang sebelumnya tidak diketahui oleh komputer, dengan mengekstraksi informasi dari berbagai sumber tertulis (Hotho & Nürnberger, 2005). Sumber data berupa situs web, buku, email, ulasan, dan artikel. Mendapatkan informasi berkualitas tinggi dari teks adalah proses yang dikenal sebagai *text mining* atau penambangan data teks. Dengan membuat pola dan tren menggunakan teknik seperti pembelajaran pola statistik dan pengenalan pola, informasi berkualitas tinggi dapat dikumpulkan. Penambangan teks adalah proses penataan teks, yang meliputi penguraian, menambahkan beberapa fitur turunan linguistik, dan memasukkan data teks berikutnya ke dalam database. Ini juga melibatkan menemukan pola dalam data terstruktur, mengevaluasi dan menafsirkan hasil keluaran, dan mengidentifikasi pola dalam data yang tidak terstruktur. Pengelompokan teks, ekstraksi konsep/entitas, analisis sentimen, meringkas dokumen, dan pemodelan koneksi adalah beberapa tugas yang terlibat dalam penambangan teks (Hotho & Nürnberger, 2005).

Analisis teks meliputi beberapa langkah untuk mengekstrak informasi dari teks, seperti menganalisis makna kata, mengenali pola, memberikan penandaan/anotasi, menggunakan teknik penambangan data seperti analisis tautan dan asosiasi, menggunakan visualisasi, serta menggunakan analitik prediktif. Tujuan dari analisis teks ini adalah untuk mengubah teks menjadi data yang dapat dianalisis, baik dengan menggunakan *Natural Language Processing* (NLP) maupun berbagai jenis algoritme dan metode analitik.

Penafsiran informasi yang dikumpulkan merupakan tahap penting dari proses ini. (Feldman, 2007).

Adapun tahapan dari *Text Mining* seperti yang dijelaskan oleh (Feldman, 2007) yaitu:

1. *Text Preprocessing*

Tahap ini adalah tahap awal dalam melakukan text mining, hal yang dilakukan dalam tahap ini yaitu mempersiapkan data tekstual yang nantinya digunakan agar dapat diproses untuk tahap selanjutnya. Proses yang dilakukan dalam tahap ini, yaitu:

- a. *Case folding*
- b. *Tokenizing*
- c. *Stemming*
- d. *Tagging*

2. *Feature Selection*

Merupakan tahapan yang bertujuan untuk mereduksi dimensi data teks agar hasil dari proses *text mining* menjadi lebih baik. Tahap ini melakukan proses *stopword removal* yaitu menghilangkan kata-kata yang dianggap tidak penting atau tidak dapat memberikan informasi penting untuk memahami isi kalimat.

3. *Text Representation*

Ini adalah tahap di mana data teks diubah menjadi bentuk yang lebih dapat dimanipulasi. Pada tahap ini, sebuah kalimat direpresentasikan oleh sebuah objek, dan kata-kata yang terkandung dalam kalimat tersebut direpresentasikan oleh fitur. Data teks akan membentuk ruang dengan objek sebanyak kalimat dan fitur sebanyak kata yang berbeda.

4. *Application of Text Mining Techniques*

Ini adalah tahap utama dalam proses penambangan teks. Pada fase ini, teknik untuk mengekstraksi informasi dari data teks yang telah diproses sebelumnya diterapkan. Beberapa teknik yang dapat digunakan di antaranya adalah *classification*, *clustering*, *information extraction*, *trend analysis*, *distribution analysis*, dan *association*. Pemilihan teknik yang

tepat akan disesuaikan dengan jenis informasi yang ingin diambil dari data tekstual yang tersedia.

2.2. Analisis Sentimen

Menurut Liu, analisis sentimen merupakan studi yang mengkaji pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas seperti produk, layanan, organisasi, masalah, peristiwa, topik, dan atributnya (Liu, 2012). Analisis sentimen merupakan bidang penelitian baru di bidang *Natural Language Processing* (NLP) yang bertujuan untuk mendeteksi subjektivitas pada teks atau mengekstrak, mengklasifikasikan pendapat dan sentimen. Terdapat tiga strategi yang digunakan dalam metode klasifikasi sentimen: *Machine Learning*, *Lexicon Based*, dan *Hybrid Approach* (D'Andrea et al., 2015). Dalam penelitian analisis sentimen yang dilakukan saat ini dengan menggunakan pendekatan *machine learning* karena pendekatan ini digunakan untuk memprediksi polaritas sentimen berdasarkan data yang telah dilatih sesuai dengan data uji dan pendekatan *Lexicon Based* karena pendekatan ini menggunakan kata-kata yang dinilai berdasarkan polaritasnya untuk mengetahui tanggapan atau opini masyarakat, kata-kata ini diberikan skor 1 untuk kata positif dan -1 kata negatif.

Sesuai dengan penjelasan Bhonde & Prasad, proses *review* analisis sentimen terdiri dari beberapa tahap. Pertama adalah mengumpulkan data teks tidak terstruktur, yang berasal dari sumber-sumber seperti blog, *tweet*, forum, dan ulasan. Kedua adalah *Data Preprocessing*, yang meliputi proses *stemming*, penghapusan kata-kata yang tidak memiliki makna (*stopword*), dan tokenisasi. Tahap terakhir adalah mengidentifikasi dan mengklasifikasikan sentimen dari data teks yang telah diproses (Prasad, 2015).

2.3. Karantina

Pengertian mengenai karantina terdapat pada UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 6 TAHUN 2018 TENTANG KEKARANTINAAN KESEHATAN, yang menjelaskan Karantina kesehatan

adalah upaya pencegahan dan penanggulangan keluarnya atau masuknya penyakit dan/atau faktor risiko kesehatan masyarakat yang berpotensi menimbulkan peristiwa kesehatan masyarakat yang luar biasa, peristiwa yang mewakili risiko terhadap kesehatan masyarakat dan berpotensi menyebar secara regional atau internasional. Peristiwa ini dapat disebabkan oleh radiasi nuklir, polusi biologis, kontaminasi kimia, bioterorisme, atau makanan.

Karantina didefinisikan berdasarkan undang-undang sebagai tindakan membatasi aktivitas individu dan/atau pemisahan dari orang lain ketika individu tersebut dianggap telah terpapar penyakit menular, meskipun mereka belum menunjukkan gejala atau masih dalam fase inkubasi. Untuk menghentikan penularan penyakit kepada orang dan/atau komoditas terdekat, karantina juga dapat melibatkan pemisahan benda-benda seperti wadah, moda transportasi, atau barang lain yang dianggap telah terkontaminasi oleh manusia, penyakit itu sendiri, atau sumber kontaminasi lainnya. Untuk melindungi masyarakat dari penyakit dan/atau faktor risiko kesehatan masyarakat yang dapat mengakibatkan keadaan darurat kesehatan masyarakat tujuan dari kekarantinaan kesehatan dalam UU di atas, yaitu:

- a. Melindungi masyarakat;
- b. mencegah dan menangkal penyakit dan/atau faktor risiko kesehatan masyarakat;
- c. meningkatkan ketahanan nasional di bidang kesehatan masyarakat; dan
- d. menawarkan perlindungan dan kepastian hukum bagi masyarakat dan petugas kesehatan.

Melihat dari situasi saat ini, yaitu penyebaran virus COVID-19 sangat cepat menular melalui kontak langsung, barang yang terkontaminasi, atau berdekatan dengan orang yang terinfeksi. Karantina adalah salah satu upaya dalam meminimalisir penyebaran virus COVID-19.

2.4. Twitter

Twitter merupakan layanan mikroblog dan jejaring sosial yang didirikan oleh Jack Dorsey pada tahun 2009. Pengguna dapat mengunggah dan berinteraksi dengan pesan singkat yang disebut "*tweets*". Awalnya, *tweet* hanya dibatasi hingga 140 karakter, tetapi pada tanggal 7 November 2017, batas ini dinaikkan menjadi 280 karakter untuk semua bahasa kecuali Cina, Jepang, dan Korea. Pengguna terdaftar dapat mengunggah, menyukai, dan me-*retweet tweet*, tetapi pengguna yang tidak terdaftar hanya dapat membacanya. Twitter dapat diakses melalui situs webnya, melalui *Short Message Service* (SMS), atau melalui aplikasi perangkat seluler. Twitter, Inc. berbasis di San Francisco, California, dan memiliki lebih dari 25 kantor di seluruh dunia (Kelly, 2009).

Pada penelitian ini, Twitter digunakan sebagai media tempat pengunduhan data yang akan digunakan untuk proses sentimen analisis. Dalam pengambilan data menggunakan Twitter API yang merupakan *library* Twitter untuk pemrograman yang dapat mengambil data dari Twitter.

2.5. Python

Python adalah bahasa pemrograman yang dikembangkan oleh Guido van Rossum pada tahun 1990 di CWI, Amsterdam sebagai kelanjutan dari bahasa pemrograman ABC. Versi terakhir yang dikeluarkan oleh CWI adalah 1.2. Python adalah bahasa interpretatif yang dapat digunakan untuk berbagai keperluan. Desain filosofi Python didesain untuk memfokuskan pada keterbacaan kode, sehingga programmer dapat dengan mudah memahami kode yang ditulis. Selain itu, Python dikenal dengan sintaksnya yang jelas dan berbagai pustaka standar, dan dianggap sebagai bahasa pemrograman tingkat tinggi yang dirancang agar mudah digunakan oleh programmer, karena efisiensi, kemudahan pengembangan, dan kompatibilitasnya dengan sistem. Selain itu, Python dapat digunakan untuk membuat aplikasi mandiri atau untuk melakukan pemrograman scripting. (Saragih, 2016).

Menurut (Saragih, 2016) bahasa pemrograman Python memiliki beberapa keunggulan, diantaranya:

- a. Mudah digunakan.
- b. Kompabilitas dan Kemampuan Tinggi.
- c. Mendukung OOP.
- d. Platform Independent.
- e. *Open Source*.

2.6. *Word2vec*

Word2vec merupakan model yang dirancang oleh tim peneliti dari Google yang dipimpin oleh Tomas Mikolov pada tahun 2013. Algoritme ini dapat mengubah kata-kata dari sebuah dokumen atau kalimat menjadi vektor kata yang kemudian disimpan dalam sebuah ruang vektor berdimensi ratusan. Setiap kata unik dalam kumpulan teks akan diberikan vektor yang sesuai dalam ruang tersebut. *Word2vec* merupakan metode yang lebih efektif dibandingkan dengan algoritme sebelumnya seperti *latent semantic analysis*. Vektor kata diposisikan dalam ruang vektor sehingga kata-kata yang berbagi konteks umum dalam kumpulan tulisan terletak berdekatan satu sama lain dalam ruang tersebut (Mikolov et al., 2013).

Keunggulan dari *word2vec* adalah memiliki ekstensi untuk menyimpan kata-kata dari seluruh kalimat yang disebut dengan *paragraf2vec* atau *doc2vec*. *Word2vec* juga telah diimplementasikan dalam beberapa bahasa pemrograman yaitu C, Python, dan Java yang dapat mendukung inferensi penyimpanan dokumen pada dokumen tak terlihat (Mikolov & Com, 2014).

Word2vec dapat menggunakan salah satu dari dua model arsitektur untuk menghasilkan representasi kata-kata yang terdistribusi: bagan kata-kata kontinu (CBOW) atau *skip-gram* kontinu. Dalam arsitektur *bag-of-words* yang berkelanjutan, model memprediksi kata saat ini dari jendela kata konteks sekitarnya. Urutan kata konteks tidak memengaruhi prediksi (asumsi *bag-of-words*). Dalam arsitektur *skip-gram*, model menggunakan kata saat ini untuk memprediksi jendela kata konteks di sekitarnya. Arsitektur *skip-gram*

menimbang kata-kata konteks terdekat lebih berat daripada kata-kata konteks lebih jauh (Mikolov et al., 2013).

Pada penelitian ini, *Word2vec* digunakan untuk penghitungan kemunculan kata dan menyimpan kata yang muncul dari keseluruhan data. *Word2vec* hanya melakukan penghitungan kata yang muncul dari setiap data dan yang paling sering muncul. *Word2vec* dapat membantu dalam pengkategorian nilai sentimen dari data yang telah didapatkan.

2.7. *Split Method*

Split method dalam *data science* adalah teknik membagi data menjadi beberapa bagian untuk memudahkan dalam proses analisis dan pemodelan. Umumnya, data akan dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk mengevaluasi performa model. *Split method* juga dapat digunakan untuk membagi data menjadi bagian-bagian lebih kecil untuk mempermudah proses *cross validation* atau *data balancing*. Contoh dari *split method* adalah *Holdout method*, *K-Fold Cross Validation*, *Stratified K-Fold Cross Validation*, *Leave One Out Cross Validation*, *Stratified Leave One Out Cross Validation*.

Hold-out cross validation adalah teknik validasi yang paling sederhana, pada metode ini data dibagi secara acak menjadi dua bagian. Bagian ini disebut data pelatihan (*training*) dan data tes (*validation*), dengan rasio tipikal pembagian 60:40 atau 80:20. Teknik ini digunakan untuk melatih model menggunakan data pelatihan dan mengevaluasi model menggunakan data tes. Teknik evaluasi model yang digunakan pada data pelatihan dalam menghitung kesalahan bergantung dari jenis masalah pada penelitian, jika data pelatihan atau pengujian tidak dapat merepresentasikan data secara lengkap maka hasil data dapat menjadi tidak baik (Yadav & Shukla, 2016).

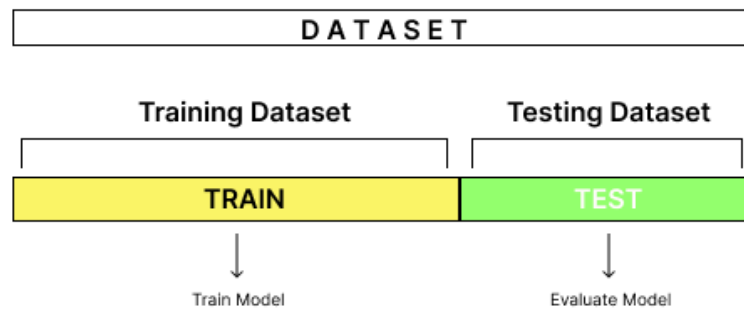
Pada kondisi terbatasnya data yang tersedia untuk proses training dan testing, diperlukan metode yang dapat membantu untuk mengevaluasi kinerja sebuah

model *machine learning*. Salah satu metode yang sering digunakan untuk mengatasi masalah ini adalah metode *holdout*. Metode *holdout* adalah metode validasi yang menyediakan sejumlah data untuk digunakan sebagai data *testing*, sementara sisanya digunakan sebagai data *training*. Dengan demikian, metode *holdout* membantu untuk mengetahui seberapa baik model dapat memprediksi data yang belum pernah dilihat sebelumnya (data *testing*). (Nugraha, 2013).

Ketika proses pembagian data menjadi data *training* dan data *testing* dilakukan dengan cara pengacakan data, maka ada kemungkinan terjadinya *overrepresented* pada salah satu atau beberapa klasifikasi. *Overrepresented* terjadi ketika salah satu klasifikasi dominan dibandingkan dengan klasifikasi lainnya, sehingga data *training* dan *testing* yang tercipta tidak representatif. Hal ini dapat mempengaruhi akurasi model yang dilatih dengan menggunakan data tersebut.

Untuk mengatasi masalah *overrepresented* ini, diperlukan prosedur *stratification holdout*. Prosedur *stratification holdout* memastikan bahwa setiap klasifikasi dapat terwakili pada data *training* dan data *testing* yang tercipta secara proporsional. Dengan demikian, distribusi klasifikasi pada data *training* dan data *testing* akan sesuai dengan distribusi klasifikasi pada data asli. Menurut Freitas 2002, proporsi klasifikasi yang terbagi dari hasil proses *holdout* harus selalu dipertahankan agar sesuai dengan proporsi klasifikasi asli. (Nugraha, 2013).

Repeated holdout adalah prosedur yang melakukan perulangan terhadap seluruh proses *training* dan *testing* beberapa kali dengan menggunakan data *training* dan data *testing* yang teracak. Setelah perulangan selesai, nilai rata-rata dari seluruh hasil evaluasi yang diperoleh akan diambil sebagai nilai akhir. Dengan cara ini, diharapkan dapat diperoleh nilai evaluasi yang lebih akurat dan stabil (Nugraha, 2013). Berikut contoh *holdout cross validation* dapat dilihat pada Gambar 1.



Gambar 1. Contoh Pembagian dataset metode *Split*

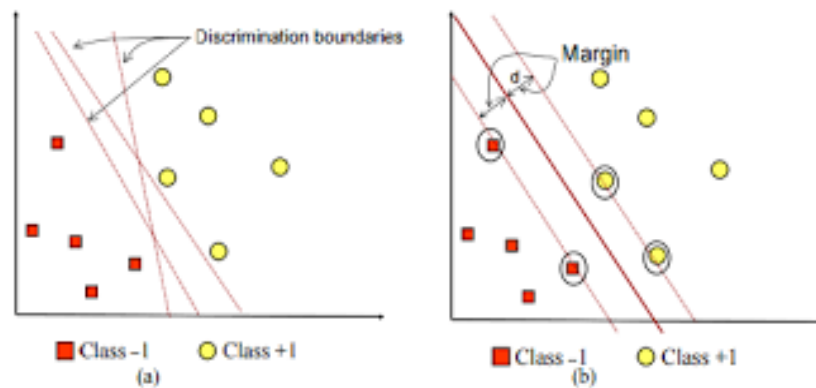
2.8. Support Vector Machine

Support Vector Machine (SVM) adalah sebuah metode pembelajaran yang digunakan untuk melakukan prediksi pada data, baik untuk klasifikasi maupun regresi (Santosa, 2015). Metode ini memerlukan tahap pelatihan, di mana model dibentuk menggunakan *sequentially training* SVM dan kemudian diuji pada tahap selanjutnya. SVM pertama kali diperkenalkan oleh Boser, Guyon, dan Vapnik pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. SVM merupakan sebuah sistem pembelajaran terbimbing yang menggunakan ruang hipotesis yang berupa fungsi-fungsi linear dalam sebuah ruang fitur (*feature space*) yang memiliki dimensi tinggi untuk melakukan pengklasifikasian.

Support Vector Machine (SVM) adalah teknik yang sangat efektif untuk melakukan klasifikasi pada data teks. Secara geometris, klasifikasi biner dapat dilihat sebagai *hyperplane* di ruang fitur yang memisahkan titik-titik yang mewakili kelas positif dari kelas negatif. Pada proses pelatihan, SVM memilih *hyperplane* yang unik yang dapat memisahkan instance positif yang diketahui dari instance negatif. SVM memiliki keunggulan dalam pendekatannya yang tidak mudah *overfitting* sehingga dapat bekerja dengan baik (Feldman, 2007).

Support Vector Machine adalah suatu teknik yang menggunakan dua titik (*vector*) untuk membentuk suatu garis pembatas yang disebut *hyperplane*. Jika dimensi yang digunakan lebih dari tiga, maka garis pembatas tersebut

akan berubah menjadi sisi pembatas. SVM digunakan untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi.



Gambar 2. Metode SVM (Parapat, 2018)

Pada Gambar 2, Dapat dilihat bahwa pada gambar A garis-garis (*hyperplane*) tersebut berhasil memisahkan 2 kelas tersebut tetapi dengan margin yang lebih kecil, sedangkan pada gambar B terdapat garis lurus atau garis linear yang memisahkan kedua kelas tersebut. Anggota yang terdekat pada *hyperplane* disebut dengan *support vector*. Dapat diamati bahwa ketika ada dua klasifikasi data, tugas SVM adalah membagi kedua kelompok ini untuk menemukan *hyperplane* terbaik, yang merupakan pembagian di mana garis batas dapat membagi dua kelompok dengan jarak terbesar antara titik terluar setiap kelompok (Kowalczyk, 2017).

Metode SVM dapat membagi data menjadi kelas yang berbeda dengan menggunakan *hyperplane* dalam ruang dimensi yang tinggi. *Hyperplane* yang terbaik adalah yang memiliki jarak terbesar ke titik data latih terdekat dari masing-masing kelas, karena margin yang lebih besar akan mengurangi kemungkinan kesalahan dalam klasifikasi (Hayati & Alifi, 2021).

Pendekatan SVM digunakan dalam pemodelan untuk mendapatkan akurasi terbaik. Dalam metode pemodelan ini, algoritma Python dijalankan menggunakan fungsi klasifikasi SVM dengan mengimpor *Support Vector Classification* (SVC). SVM adalah metode *supervised learning* yang digunakan untuk klasifikasi, regresi dan deteksi outlier. SVM digunakan karena menunjukkan akurasi analisis teks yang kuat. *Kernel* adalah kumpulan

operasi matematika yang dimiliki oleh SVM yang mengubah data input menjadi ruang dimensi tinggi (*Kernel Space*). *Kernel* yang populer dalam SVM diantaranya *Gaussian Radial basis Function (RBF)*, *linear kernel*, *polynomial function*, *sigmoid*, *Laplacian radial basis function*, dan *linear spline kernel* yang dapat dilihat pada Tabel 1 (Wati & Ernawati, 2021).

Tabel 1. Jenis *Kernel* SVM

Jenis <i>Kernel</i>	Definisi
<i>Linear</i>	$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i^t \vec{x}_j$
<i>Polynomial</i>	$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i, \vec{x}_j + 1)^p$
<i>Gaussian</i>	$K(\vec{x}_i, \vec{x}_j) = \exp \frac{-\ \vec{x}_i - \vec{x}_j\ ^2}{2\sigma^2}$
<i>Sigmoid</i>	$K(\vec{x}_i, \vec{x}_j) = \tanh(\alpha \vec{x}_i, \vec{x}_j + \beta)$

SVM merupakan salah satu metode yang digunakan untuk memisahkan dua kelas data di dalam ruang input yang tidak linear dengan menggunakan pendekatan *kernel* pada himpunan fitur data awal. Fungsi *kernel* membantu untuk memetakan dimensi data awal ke dimensi yang lebih tinggi. SVM mencari *hyperplane* pemisah terbaik yang terletak di tengah diantara dua kelas data. *Hyperplane* yang terbaik adalah yang memiliki jarak terbesar ke titik data terdekat dari kelas manapun, dan merupakan solusi terbaik untuk memisahkan kedua kelas data tersebut. Biasanya, kedua kelas data tidak dapat terpisah secara sempurna, sehingga *hyperplane* yang dicari adalah yang terbaik dari segi pemisahan kedua kelas tersebut. (Rizki et al., 2021).

2.9. *Lexicon Based*

Lexicon Based merupakan metode yang sederhana, layak, dan praktis untuk analisis sentimen. Situs media sosial seperti Twitter, Facebook, dan lainnya menyediakan data yang dapat dimanfaatkan untuk menentukan bagaimana perasaan orang tentang suatu barang atau jasa (Matulatuwa, Sedyono, and Iriani 2017). Tidak membutuhkan data berlabel dan prosedur pembelajaran merupakan keuntungan yang diperoleh dengan metode *Lexicon Based*

(Devika, Sunitha, and Ganesh 2016). *Lexicon Based* menggunakan kata-kata yang dievaluasi berdasarkan polaritasnya untuk menemukan reaksi atau sudut pandang orang. Kata-kata positif diberi skor 1, sedangkan kata-kata negatif diberi skor -1, dan kata-kata netral diberi skor 0.

2.10. Teknik *Resampling*

Teknik *resampling* dataset adalah metode yang digunakan untuk memperoleh sampel baru dari dataset yang sudah ada. Ada beberapa teknik *resampling* yang biasa digunakan, di antaranya:

- *Cross-validation*: proses membagi dataset menjadi beberapa fold dan melakukan pelatihan dan evaluasi model dengan menggunakan fold yang berbeda.
- *Bootstrapping*: proses mengambil sampel acak dengan replacement dari dataset asli.
- *Oversampling*: proses menambahkan sampel kelas minoritas ke dataset asli secara acak.
- *Undersampling*: proses mengurangi sampel kelas mayoritas dari dataset asli secara acak.
- *Stratified sampling*: proses membagi dataset menjadi beberapa stratum berdasarkan kelas, kemudian mengambil sampel secara acak dari masing-masing stratum.

Teknik *resampling* ini biasanya digunakan untuk mengevaluasi performa model atau estimator dengan menggunakan data yang sudah ada, memperbaiki masalah *imbalanced* dataset, atau untuk memperoleh sampel yang representatif dari dataset asli. Pada penelitian kali ini menggunakan *Oversampling* dan *Undersampling*.

2.11. Penelitian Terdahulu

Penelitian sebelumnya mengenai *text mining* untuk analisis sentimen telah banyak dilakukan dengan menggunakan berbagai metode yang berbeda.

Beberapa penelitian terdahulu yang berkaitan dengan analisis sentimen dapat dilihat pada Tabel 2.

Tabel 2. Penelitian terdahulu

Perbandingan Teori	Peneliti (Tahun Penelitian)	Tujuan	Langkah-langkah	Metode	Hasil
Perbandingan Metode <i>Lexicon based</i> dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter	(Najib et al., 2019)	Melakukan analisis sentimen pada Kampanye Pilpres Indonesia Tahun 2019 dengan pendekatan berbasis ontologi	<i>Crawling</i> data Pelabelan <i>tweet</i> Pembuatan ontologi Praproses dan klasifikasi Pembaharuan ontologi Pembaharuan klasifikasi <i>tweet</i> Seleksi ulang dataset Pembuatan dan pemilihan model Perhitungan skor sentiment Pemetaan hasil skor sentimen berdasarkan ontologi	Metode <i>Lexicon-based</i> dan SVM	Kekurangan: Metode <i>Lexicon</i> memiliki performa kurang baik, karena kurangnya penyempurnaan ontologi yang telah dibuat. Kelebihan: Sedangkan SVM memiliki performa yang lebih baik, faktor yang memengaruhi akurasi dari analisis sentimen SVM adalah pemilihan dataset. Hasil: Dari uji analisis ditemukan bahwa nilai akurasi untuk analisis sentimen menggunakan <i>Lexicon</i> sebesar 39,9% dan SVM memiliki akurasi sebesar 83,9%.
Penerapan <i>Lexicon Based</i> Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19	(Nooryuda Prasetya & Winarso, 2021)	Mengelompokkan opini masyarakat terhadap isu Covid-19 ke dalam kategori sentimen positif, negatif atau netral menggunakan <i>lexicon</i>	Pengumpulan data <i>tweet</i> <i>Text preprocessing</i> Klasifikasi menggunakan <i>Lexicon</i> Visualisasi	Metode <i>Lexicon-based</i>	Kekurangan: Kesalahan <i>preprocessing</i> teks akibat data yang tidak dinormlisasikan. Kelebihan: Sejumlah data besar digunakan yaitu 72.686 <i>tweet</i> dari Maret 2020 hingga Agustus 2020 & metode <i>lexicon-based</i> menghasilkan hasil yang baik. Hasil: Opini dengan sentimen positif sebesar 58.08% dari total <i>tweet</i> , opini dengan sentimen kategori negatif sebesar

Tabel 2 (lanjutan)

Perbandingan Teori	Peneliti (Tahun Penelitian)	Tujuan	Langkah-langkah	Metode	Hasil
					37.61%, dan sentimen kategori netral sebesar 4.31%. Dapat disimpulkan masyarakat pada twitter yang percaya bahwa COVID-19 adalah tinggi.
Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi <i>Support Vector Machine</i> (SVM)	(Tineges et al., 2020)	Mengetahui sentimen positif dan negatif tentang pengguna layanan Indihome, serta untuk mengetahui seberapa puas pengguna layanan Indihome berdasarkan Twitter	Pengumpulan data Data <i>Preprocessing</i> Pembagian data latih Pemodelan Data uji Evaluasi Opini positif & negatif	<i>Support Vector Machine</i> (SVM)	Kekurangan: Terdapat kesalahan pada hasil prediksi data uji dan data latih. Kelebihan: Metode SVM menghasilkan hasil <i>Confussion Matrix</i> yang baik. Hasil: Dari evaluasi <i>Confussion Matrix</i> , didapat akurasi sebesar 87% dengan (precision) sebesar 86%, (<i>recall</i>) sebesar 95%, (<i>error rate</i>) sebesar 13%, sedangkan (<i>f1-score</i>) adalah sebesar 90%. Dan berdasarkan data baru, dengan metode <i>Support Vector Machine</i> mendapatkan hasil nilai positif sebesar 18,4 % dan hasil nilai negatif sebesar 81,6%.

Pada Tabel 2, menjelaskan tentang penelitian (Najib et al., 2019), hasil dari metode SVM lebih baik dibandingkan metode *lexicon-based*. Faktor-faktor seperti penggunaan bahasa yang tidak baku, kurangnya kata-kata dalam kamus, dan kemunculan kata-kata positif lebih banyak dibandingkan kata bernilai negatif dapat mempengaruhi performa dari metode *Lexicon-based*. Selain itu, kurangnya proses ekstraksi dan pembuatan ontologi secara semi-otomatis juga dapat mempengaruhi performa metode tersebut. Namun demikian, faktor yang paling mempengaruhi akurasi dari analisis sentimen menggunakan metode SVM adalah pemilihan dataset yang digunakan.

Pada penelitian (Prasetya & Winarso, 2021), pada tahapan *text preprocessing convert word*, masih ada kata-kata tertentu yang tidak bisa di *convert* selama tahap *preprocessing* dikarenakan sulit dinormalisasi dan kesalahan dalam pengetikan atau kalimat yang salah eja dalam beberapa data *tweet*. Selain itu, dalam bentuk visualisasi *wordcloud* istilah "konspirasi," yang pada kenyataannya memiliki konotasi buruk atau negatif, memiliki konotasi yang baik atau positif. Ini mungkin terjadi karena sentimen positif sering menyertakan kata "konspirasi". Temuan ini menunjukkan bahwa, berbeda dengan mereka yang memegang teori konspirasi, persentase pengguna Twitter yang lebih tinggi memiliki keyakinan bahwa COVID-19 itu nyata.

Pada penelitian (Tineges et al., 2020), pada pemodelan dari data latih yang digunakan untuk data uji sebanyak 280 tweet, terdapat kesalahan hasil prediksi. Dari 194 data yang diprediksi sebagai negatif, terdapat 27 data yang seharusnya berada pada kelas positif. Sedangkan dari 86 data yang diprediksi sebagai positif, terdapat 9 data yang seharusnya berada pada kelas negatif. Berdasarkan data baru, nilai positif sebesar 18,4% dan nilai negatif sebesar 81,6%. Dapat disimpulkan bahwa tingkat kepuasan pengguna layanan Indihome cukup rendah.

III. METODELOGI PENELITIAN

3.1. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung. Penelitian dilaksanakan mulai bulan Januari 2022 sampai bulan November tahun 2022.

3.2. Alat dan Bahan

Penelitian ini dilakukan dengan menggunakan alat untuk mendukung dan menunjang pelaksanaan penelitian, sebagai berikut:

a. Perangkat Keras (*Hardware*)

Dalam penelitian ini, digunakan satu unit laptop sebagai perangkat keras yang mendukung pelaksanaan penelitian. Laptop ini memiliki spesifikasi:

- Processor : Intel(R) Core(TM) i3-5005U CPU @ 2.00GHz (4 CPUs), ~2.0GHz
- RAM : 8192 MB

b. Perangkat Lunak (*Software*)

Perangkat lunak yang digunakan dalam analisis sentimen ini adalah:

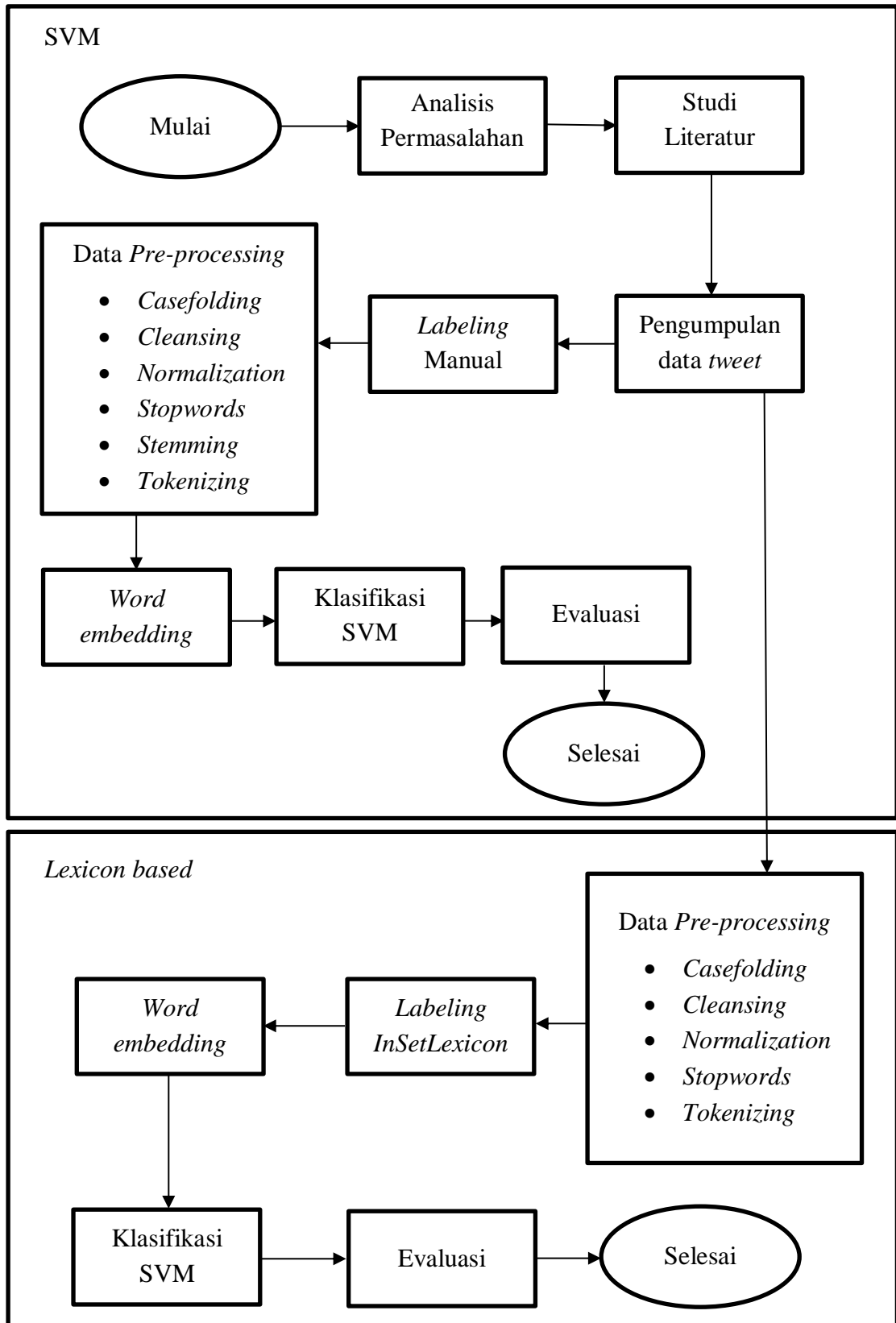
- Sistem operasi Windows 10
- Google Chrome yang digunakan untuk web browser.
- Python 3 yang digunakan sebagai bahasa pemrograman.
- Jupyter Notebook 6.4.5 yang digunakan untuk membuat *script* program.

c. *Library* yang digunakan:

- *YAML (Yet Another Markup Language)*: bahasa *markup* yang digunakan untuk menyimpan data dalam format yang mudah dibaca.
- *JSON (JavaScript object notation)*: format yang digunakan untuk menyimpan dan mentransfer data.
- *Searchtweets*: *library* Python untuk pengguna mengakses *premium search* API dari Twitter.
- *Pandas (Python for Data Analysis)*: *library* Python yang digunakan untuk mengelola dan menganalisis data secara efisien
- *NumPy (Numerical Python)*: *library* Python yang digunakan untuk melakukan komputasi numerik dan *scientific computing*
- *Sastrawi*: *library* bahasa Indonesia yang ditujukan untuk melakukan *stemming* (pengubahan kata ke dasar kata) pada teks bahasa Indonesia.
- *NLTK (Natural Language Toolkit)*: *library* Python yang menyediakan alat-alat yang dibutuhkan untuk memproses dan menganalisis bahasa alami, seperti mengekstrak informasi dari teks, menghitung frekuensi kata, dan mengidentifikasi tipe kata.
- *Matplotlib*: *library* Python yang digunakan untuk membuat plot atau grafik.
- *Wordcloud*: *library* Python digunakan untuk membuat visualisasi data dalam bentuk *wordcloud* (awan kata).
- *Gensim*: *library* Python untuk merepresentasikan dokumen sebagai vektor semantik.
- *SKLearn*: *library* yang membantu dalam melakukan pemodelan *machine learning*.
- *Imblearn*: *library* yang berguna dalam melakukan metode *sampling*.

3.3. Tahapan Penelitian

Penelitian ini melalui beberapa tahapan yang dapat dilihat pada gambar dihalaman selanjutnya:



Gambar 3. Tahapan Penelitian

Pada Gambar 3 menjelaskan tentang tahapan-tahapan penelitian yang dilakukan diantaranya:

a. Analisis Permasalahan

Pada tahap ini dilakukan analisis permasalahan yang ada yaitu mengenai metode yang dapat menghasilkan akurasi yang tinggi mengenai analisis sentimen. Topik dalam penelitian ini adalah analisis sentimen dengan pengklasifikasian sentimen positif dan negatif terhadap data yang didapatkan dari media sosial Twitter dengan menggunakan algoritme *Support Vector Machine (SVM)* dan *Lexicon Based* tentang Pemberlakuan Karantina di Masa COVID-19 untuk mengetahui persepsi masyarakat terhadap Pemberlakuan Karantina di Indonesia guna membantu pihak-pihak yang membutuhkan informasi mengenai pemberlakuan karantina di masa pandemi COVID -19 di Indonesia.

b. Studi Literatur

Pada tahap ini dilakukan pengumpulan informasi dan penelitian terdahulu yang berkaitan dengan topik permasalahan atau penelitian yang sama dan mempelajari berbagai metode penelitian. Untuk penelitian terdahulu mengenai analisis sentimen, terdapat tiga penelitian terdahulu dengan topik yang berbeda yang dijadikan sebagai studi literatur untuk penelitian ini. Algoritme yang digunakan dari ketiga penelitian terdahulu berbeda-beda.

c. Pengumpulan data

Pada tahap ini dilakukan pengumpulan data *tweet* dari media sosial Twitter menggunakan API key yang disediakan oleh Developer Twitter dengan menggunakan *library searchtweets*. Untuk mendapatkan API key diperlukan akun Twitter yang telah didaftarkan menjadi Developer Twitter. Data yang dikumpulkan merupakan data teks yang tidak memiliki nilai positif dan negatif. Kata kunci untuk pengumpulan data Twitter menggunakan 3 kata kunci yaitu “Karantina”, “Karantina Covid” dan “Karantina di tengah pandemi”. *Tweet* yang diambil *tweet* bulan September 2021 sampai Februari 2022 dikarenakan pada waktu tersebut tepatnya pada bulan September 2021 pemerintahan memperketat perjalanan Internasional

dari luar negeri dan Satgas Penanganan Covid-19 juga membuat perubahan yang menyoroti aturan baru tentang ketentuan bagi pelaku perjalanan Internasional yang terus diperbaharui menyesuaikan kondisi di Indonesia.

- d. *Labeling*, proses *Labeling* dilakukan dengan 2 cara yaitu *labeling* manual dan *labeling InSetLexicon*. *Labeling* secara manual dilakukan sebelum *preprocessing* dengan memberikan nilai sentimen 2 kelas yaitu positif dan negatif. Untuk menghindari subjektivitas dalam pelabelan dilakukan oleh 3 orang dan hasilnya disimpulkan dengan label mayoritas atau label terbanyak dari hasil pelabelan dalam satu *tweet*. Selanjutnya dilakukan juga pelabelan menggunakan *lexicon based* dengan cara menghitung skor sentimen menggunakan kamus *InSetLexicon* dari penelitian sebelumnya untuk mengklasifikasi sentimen terhadap twitter. Fitur "*Lexicon Based*" merupakan komponen yang mengidentifikasi apakah sebuah kata memiliki sentimen positif atau negatif berdasarkan pengklasifikasian yang terdapat dalam kamus atau *lexicon*. Jadi, dengan menggunakan fitur ini, kita bisa mengetahui apakah suatu kata memiliki konotasi positif atau negatif berdasarkan penilaian yang terdapat dalam kamus atau *lexicon*.

- e. Pengolahan dataset

Pada tahap ini data yang diperoleh dari Twitter dilanjutkan ke tahap *pre-processing* diantaranya:

- a) *Casefolding*, merupakan proses mengubah teks menjadi bentuk yang sama, terlepas dari apakah hurufnya besar atau kecil. Ini biasanya dilakukan untuk memastikan bahwa teks dapat dibandingkan dengan benar tanpa memperhatikan huruf besar atau kecil. Proses *casefolding* ditunjukkan pada Tabel 3.

Tabel 3. Proses *Casefolding*

<i>Input</i>	<i>Output</i>
Baru aja berita indo bisa masuk thai tanpa karantina eh lgsg dpt notif dri google, tau ajanih \ud83d\ude02\ud83e\udd23 https://t.co/C5LCMhr31Y	baru aja berita indo bisa masuk thai tanpa karantina eh lgsg dpt notif dri google, tau ajanih \ud83d\ude02\ud83e\udd23 https://t.co/C5LCMhr31Y

- b) *Cleansing*, merupakan proses penghapusan karakter non-alfabetis seperti simbol, tanda baca, atau karakter spesial yang tidak merupakan huruf atau angka. Penghapusan karakter non-alfabetis dilakukan untuk mengurangi noise atau gangguan yang dapat mempengaruhi proses analisis atau pemrosesan data. Karakter yang dihapus berupa tanda baca, simbol-simbol seperti tanda '@' untuk nama pengguna, hashtag (#), RT tweets, Duplicate tweet dan url dari situs web. Proses *cleansing* ditunjukkan pada Tabel 4.

Tabel 4. Proses *Cleansing*

<i>Input</i>	<i>Output</i>
baru aja berita indo bisa masuk thai tanpa karantina eh lgsg dpt notif dri google, tau ajanih \ud83d\ude02\ud83e\udd23 https://t.co/C5LCMhr31Y	baru aja berita indo bisa masuk thai tanpa karantina eh lgsg dpt notif dri google tau ajanih

- c) *Normalization*, merupakan proses memperbaiki ejaan yang disingkat atau *handling slangword*. Proses *Normalization* ditunjukkan pada Tabel 5.

Tabel 5. Proses *Normalization*

<i>Input</i>	<i>Output</i>
baru aja berita indo bisa masuk thai tanpa karantina eh lgsg dpt notif dri google tau ajanih	baru aja berita indo bisa masuk thai tanpa karantina eh langsung dapat notif dari google tau ajanih

- d) *Stopwords*, merupakan proses *filtering* yang bertujuan untuk memilih kata-kata penting dari hasil *tokenizing*. Pemilihan kata-kata penting ini dilakukan dengan mengecualikan *stopwords* dari daftar token yang dihasilkan, sehingga hanya tersisa kata-kata yang dianggap relevan atau memiliki makna dalam dokumen tersebut. Proses *Stopwords* ditunjukkan pada Tabel 6.

Tabel 6. Proses *Stopwords*

<i>Input</i>	<i>Output</i>
baru aja berita indo bisa masuk thai tanpa karantina eh langsung dapat notif dari google tau ajanih	baru aja berita indo masuk thai karantina eh langsung dapat notif dari google tau ajanih

<i>Input</i>	<i>Output</i>
iya tadi lihat daily case nya masih 8 9rb an ya ternyata plus baliknya harus karantina jadi mending ntar dulu deh tahan	iya tadi lihat daily case nya 8 9rb an ternyata plus baliknya karantina jadi mending ntar dulu deh tahan

- e) *Stemming*, proses mengubah kata-kata yang sudah disaring menjadi kata dasar atau kata dasar dari setiap kata yang memiliki imbuhan. Proses *Stemming* ditunjukkan pada Tabel 7.

Tabel 7. Proses *Stemming*

<i>Input</i>	<i>Output</i>
baru aja berita indo masuk thai karantina eh langsung dapet notif dari google tau ajanih	baru aja berita indo masuk thai karantina eh langsung dapet notif dari google tau ajanih
iya tadi lihat daily case nya 8 9rb an ternyata plus baliknya karantina jadi mending ntar dulu deh tahan	iya tadi lihat daily case nya 8 9rb an nyata plus balik karantina jadi mending ntar dulu deh tahan

- f) *Tokenizing*, merupakan proses memecah teks menjadi token-token yang lebih kecil. Token dapat diartikan sebagai unit terkecil dari teks yang masih memiliki makna. Dengan memecah teks menjadi token-token yang lebih kecil, kita bisa dengan mudah menghitung frekuensi kemunculan kata-kata dalam suatu *corpus* atau kumpulan teks. Proses *Tokenizing* ditunjukkan pada Tabel 8.

Tabel 8. Proses *Tokenizing*

<i>Input</i>	<i>Output</i>
baru aja berita indo masuk thai karantina eh langsung dapet notif dari google tau ajanih	baru, aja, berita, indo, masuk, thai, karantina, eh, langsung, dapet, notif, dari, google, tau, ajanih,

- f. *Word Embedding*

Word embedding merupakan proses yang digunakan untuk teknik pemrosesan teks yang bertujuan untuk mengubah representasi kata-kata dalam teks menjadi vektor-vektor numerik yang dapat diolah oleh model *machine learning*. Pada penelitian ini menggunakan *word2vec*.

g. Klasifikasi Sentimen

Pada tahap ini, akan dilakukan klasifikasi sentimen menggunakan metode *Support Vector Machine* (SVM). Sebelum itu, data *tweet* yang telah diproses dan diberi bobot pada setiap katanya harus dibagi menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk melatih algoritme atau metode, sedangkan data uji digunakan untuk menguji performa algoritme yang sudah dilatih terhadap data baru yang belum pernah ditemui sebelumnya.

h. Evaluasi

Pada tahap evaluasi hasil, dilakukan pengukuran kinerja sistem untuk mengetahui seberapa baik sistem dalam melakukan klasifikasi data. Salah satu metode yang sering digunakan untuk evaluasi hasil adalah *Confusion Matrix*. *Confusion Matrix* merupakan tabel yang menggambarkan hubungan antara label prediksi dengan label sebenarnya dari suatu sistem. Dari *Confusion Matrix* dapat dihitung beberapa metrik evaluasi seperti akurasi, presisi, *recall*, dan *f1-score* yang berguna untuk menilai seberapa baik sistem dalam melakukan klasifikasi data. Dengan demikian, proses evaluasi hasil sangat penting untuk menarik kesimpulan dari proses penelitian yang telah dilakukan dan mengetahui seberapa efektif sistem dalam melakukan klasifikasi data. Berikut tabel *Confusion Matrix* yang dapat dilihat pada Tabel 9.

Tabel 9. *Confusion Matrix*

Data Aktual	Data Prediksi	
	Negatif	Positif
Negatif	TN (<i>True Negative</i>)	FP (<i>False Positive</i>)
Positif	FN (<i>False Negative</i>)	TP (<i>True Positive</i>)

Keterangan:

TP (*True Positive*) = Jumlah data yang terprediksi sebagai positif dan sebenarnya positif.

TN (*True Negative*) = Jumlah data yang terprediksi sebagai negatif dan sebenarnya negatif.

FP (*False Positive*) = Jumlah data yang terprediksi sebagai positif tetapi sebenarnya negatif.

FN (*False Negative*) = Jumlah data yang terprediksi sebagai negatif tetapi sebenarnya positif.

Dari Tabel 9 Digunakan untuk mengukur nilai akurasi, presisi, *recall* dan *f1-score* sebagai berikut:

a) Akurasi merupakan rasio prediksi benar (positif, dan negatif) dengan keseluruhan data.

$$akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

b) Presisi merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

$$presisi = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

c) *Recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

$$recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

d) *F1-Score* merupakan perbandingan rata-rata presisi dan recall yang dibobotkan.

$$f1\ score = 2 \times \frac{(recall \times presisi)}{(recall + presisi)} \quad (4)$$

V. SIMPULAN DAN SARAN

5.1. Simpulan

Pada penelitian ini dapat disimpulkan bahwa:

1. Penelitian sudah berhasil mengimplementasikan algoritme *Support Vector Machine* (SVM) dan *Lexicon Based* dengan menggunakan *InSetLexicon*. Hasil dari pemberian sentimen secara manual terhadap data dari Twitter tentang Pemberlakuan karantina di masa pandemi Covid-19 menghasilkan sebanyak 2461 *tweet* data kelas positif dan 741 *tweet* data kelas negatif sedangkan pemberian sentimen menggunakan kamus *InSetLexicon* dihasilkan sebanyak 2419 *tweet* data kelas positif dan 783 *tweet* data kelas negatif. Hal tersebut berarti bahwa persepsi atau pandangan masyarakat terhadap pemberlakuan karantina covid-19 pada bulan September 2021 sampai Februari 2022 memiliki kecenderungan positif, dengan terdapat sebanyak 76,9% opini masyarakat termasuk kedalam sentimen positif dan 23,1% opini masyarakat termasuk kedalam sentimen negatif.
2. Perbandingan hasil kinerja dari klasifikasi algoritme *Support Vector Machine* (SVM) dengan pelabelan secara manual menunjukkan nilai klasifikasi yang lebih baik daripada pelabelan menggunakan *InSetLexicon*. Percobaan *Oversampling* lebih baik dibandingkan dengan percobaan *Imbalanced* data dan percobaan *Undersampling*. Algoritme *Support Vector Machine* (SVM) dengan pelabelan manual menghasilkan hasil tertinggi dengan percobaan *oversampling* menggunakan skenario ke-5 yaitu 90% data *training* & 10% data *testing* dengan hasil akurasi sebesar 97,36%, presisi 97%, *recall* 98%, dan *f1-score* 97% sedangkan hasil pelabelan menggunakan *InSetLexicon* menghasilkan hasil tertinggi dengan percobaan

oversampling menggunakan skenario ke-5 juga yaitu 90% data *training* & 10% data *testing* dengan hasil akurasi sebesar 97%, presisi 96%, *recall* 96%, dan *f1-score* 97%.

5.2. Saran

Berdasarkan hasil pengujian yang diperoleh, terdapat beberapa kekurangan yang perlu diperbaiki untuk meningkatkan performa dalam penelitian ini. Oleh karena itu, penulis memberikan beberapa saran untuk dapat dilakukan dalam pengembangan penelitian selanjutnya:

1. Dataset yang digunakan dapat diperbanyak jumlahnya, agar data latih yang digunakan dapat lebih banyak dan menyetarakan kelas data positif dan negatif sehingga sistem klasifikasi dapat memiliki nilai akurasi yang lebih baik.
2. Dapat menggunakan algoritma klasifikasi yang lain seperti Naïve Bayes atau mengkombinasikan dengan *Deep Learning* seperti *Reccurent Neural Network*, *Deep Neural Networks*, dan *Convolutional Neural Networks* sehingga dapat membandingkan hasil uji model yang dilakukan untuk mendapatkan hasil akurasi yang lebih baik.
3. Waktu pengambil *tweet* dengan kisaran yang lebih panjang agar mendapatkan opini yang lebih beragam untuk meningkatkan hasil akurasi.

DAFTAR PUSTAKA

- A. Kowalczyk. (2017). *Support Vector Machines Succinctly*. Syncfusion Inc.
<https://www.syncfusion.com/succinctly-free-ebooks/support-vector-machines-succinctly>
- Corallo, A., Fortunato, L., Matera, M., Alessi, M., Camillò, A., Chetta, V., Giangreco, E., & Storelli, D. (2015). Sentiment Analysis for Government: An Optimized Approach. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9166, 98–112. https://doi.org/10.1007/978-3-319-21024-7_7
- D’Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, Tools and Applications for Sentiment Analysis Implementation. *International Journal of Computer Applications*, 125(3), 26–33.
<https://doi.org/10.5120/ijca2015905866>
- Devika, M. D., C. Sunitha, and Amal Ganesh. 2016. “Sentiment Analysis: A Comparative Study on Different Approaches.” *Procedia Computer Science* 87: 44–49. <https://doi.org/10.1016/j.procs.2016.05.124>.
- Feldman, J. S. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*.
- Felicia Watratan, A., Puspita, A. B., Moeis, D., Informasi, S., & Profesional Makassar, S. (2020). Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. *Journal of Applied Computer Science and Technology*, 1(1), 7–14.
<https://doi.org/10.52158/JACOST.V1I1.9>
- Hayati, H., & Alifi, M. R. (2021). Analisis Sentimen Pada Tweet Terkait Vaksin Covid-19 Menggunakan Metode Support Vector Machine. *Jurnal Teknologi Terapan* /, 7(2), 110–119.
<https://jurnal.polindra.ac.id/index.php/jtt/article/view/349>
- Hotho, A., & Nürnberger, A G. P. (2005). A Brief Survey of Text Mining.

- Journal for Computational Linguistics and Language Technology*, 20(1), 19–62.
- Kelly, R. (2009). Twitter Study – August 2009 Introduction. *New York, 2010*(August), 1–17.
- Keramati, A., & Yousefi, N. (2011). A Proposed Classification of Data Mining Techniques in Credit Scoring. *Techniques*, September, 416–424.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.
- Maharani, T. (2020). UPDATE 26 Oktober: Tambah 112, Pasien Covid-19 Meninggal Jadi 13. *Kompas.Com*.
- Mahendrajaya, R., Buntoro, G. A., & Setyawan, M. B. (2019). Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine. *Komputek*, 3(2), 52. <https://doi.org/10.24269/jkt.v3i2.270>
- Matulatuwa, Febrilien Matresya, Eko Sedyono, and Ade Iriani. 2017. “Text Mining Dengan Metode Lexicon Based Untuk Sentiment Analysis Pelayanan PT. Pos Indonesia Melalui Media Sosial Twitter.” *Jurnal Masyarakat Informatika Indonesia* 2 (3): 52–65.
- Mikolov, Tomas & Chen, Kai & Corrado, G.s & Dean, Jeffrey. (2013). Efficient Estimation of Word Representations in Vector Space. *Proceedings of Workshop at ICLR*. 2013.
- Mikolov, T., & Com, T. G. (2014). *Distributed Representations of Sentences and Documents*. 32.
- Najib, A. C., Irsyad, A., Qandi, G. A., & Rakhmawati, N. A. (2019). Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter. *Fountain of Informatics Journal*, 4(2), 41. <https://doi.org/10.21111/fij.v4i2.3573>
- Nugraha, P. A. (2013). *Perbandingan Metode Probabilistik Naive Bayesian Classifier dan Jaringan Syaraf Tiruan Learning Vector Quantization dalam Kasus Klasifikasi Penyakit Kandungan*. 2(2), 20–33.
- Parapat, I. M. (2018). *Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak* (Doctoral dissertation, Universitas Brawijaya).

- Prasad, S. B. B. and J. R. (2015). Sentiment Analysis - Methods, Applications and Challenges,. *International Journal of Electronics Communication and Computer Engineering*, 6(6), 2249–71.
- Prasetya, Y., & Winarso, D. (2021). Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19. *Jurnal Fasilkom*, 11(2), 97–103.
- Rizki, M. F., Auliasari, K., & Primaswara Prasetya, R. (2021). Analisis Sentiment Cyberbullying Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(2), 548–556. <https://doi.org/10.36040/jati.v5i2.3808>
- Santosa, B. (2015). *Tutorial Support Vector Machine. [e-book]*.
- Saragih, R. R. (2016). *PEMROGRAMAN DAN BAHASA PEMROGRAMAN*.
- Susilo, A., Rumende, C. M., Pitoyo, C. W., Santoso, W. D., Yulianti, M., Herikurniawan, H., Sinto, R., Singh, G., Nainggolan, L., Nelwan, E. J., Chen, L. K., Widhani, A., Wijaya, E., Wicaksana, B., Maksum, M., Annisa, F., Jasirwan, C. O. M., & Yunihastuti, E. (2020). Coronavirus Disease 2019: Tinjauan Literatur Terkini. *Jurnal Penyakit Dalam Indonesia*, 7(1), 45–67. <https://doi.org/10.7454/JPDI.V7I1.415>
- Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). 4, 650–658. <https://doi.org/10.30865/mib.v4i3.2181>
- Wati, R., & Ernawati, S. (2021). Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 06, 240–247. <http://ejournal.ust.ac.id/index.php/JTIUST/article/view/1465>
- Yadav, S., & Shukla, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 78-83.