ANALISIS SENTIMEN APLIKASI PEDULILINDUNGI PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER

(Skripsi)

Oleh

SASYA SALSABILA JANERDI

1867051001

S1 ILMU KOMPUTER
JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2022

ANALISIS SENTIMEN APLIKASI PEDULILINDUNGI PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER

Oleh

SASYA SALSABILA JANERDI

Skripsi

Sebagai salah satu syarat untuk memperoleh gelar

SARJANA KOMPUTER

Pada

Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

S1 ILMU KOMPUTER
JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2022

ABSTRAK

ANALISIS SENTIMEN APLIKASI PEDULILINDUNGI PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER

Oleh

SASYA SALSABILA JANERDI

Analisis sentimen merupakan pengolahan data tekstual yang bertujuan untuk klasifikasi polaritas dari teks pada kalimat atau opini . Klasifikasi tersebut bertujuan untuk melihat polaritas dari suatu kalimat atau opini apakah bersifat positif, negatif, atau netral. Masyarakat memiliki berbagai opini yang diekspresikan di berbagai media, diantaranya media sosial twitter. Besarnya komunitas pengguna Twitter di Indonesia tentunya berpengaruh pada ragam opini mengenai aplikasi PeduliLindungi. Aplikasi PeduliLindungi sebagai upaya dalam melakukan *tracing*, *tracking*, dan pemberi peringatan selama pandemi COVID-19. PeduliLindungi perlu terus dikembangkan untuk meningkatkan rasa nyaman bagi penggunanya sehingga menjamin penggunaan yang berkelanjutan. Penelitian ini menggunakan metode Naive Bayes Classifier dalam analisis sentimen menggunakan data Aplikasi PeduliLindungi dari Twitter. Berdasarkan hasil penelitian , menghasilkan nilai akurai sebesar 0,85 atau 85% dan akurasi meningkat sebesar 0.89 atau 89% dengan pengujian menggunakan k-fold cross validation.

Kata Kunci: Analisis Sentimen, PeduliLindungi, Naïve Bayes Classifier

ABSTRACT

SENTIMENT ANALYSIS OF PEDULINDUNGI APPLICATIONS ON TWITTER SOCIAL MEDIA USING NAIVE BAYES CLASSIFIER METHOD

By

SASYA SALSABILA JANERDI

Sentiment analysis is textual data processing that aims to classify text polarity in sentences or opinions. Classification aims to see the polarity of a sentence or opinion whether positive, negative or neutral. The public has various opinions which are expressed in various media, including social media Twitter. The size of the Twitter user community in Indonesia certainly influences the diversity of opinions regarding the PeduliLindungi application. The PeduliLindungi application is a tracing, tracking, and alerting effort during the COVID-19 pandemic. PeduliLindungi needs to be continuously developed to increase comfort for its users so they can continue to use it. This study uses the Naive Bayes Classifier method in sentiment analysis using data from the PeduliLindungi application from Twitter. Based on the research results, it produces an accuracy value of 0.85 or 85% and accuracy increases by 0.89 or 89% by testing using k-fold cross validation.

Keywords: Sentiment Analysis, PeduliLindungi, Naive Bayes Classifier

Judul Skripsi	: **ANALISIS SENTIMEN APLIKASI PEDULILINDUNGI PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER**
Nama Mahasiswa	: Sasya Salsabila Janerdi
Nomor Pokok Mahasiswa	: **1867051001**
Jurusan	: **Ilmu Komputer**
Fakultas	: **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Anie Rose Irawati, S.T. M.Cs.
NIP. 196701031992031003

2. Ketua Jurusan Ilmu Komputer

Didik Kurniawan, S.Si., M.T.
NIP. 198004192005011004

MENGESAHKAN

1. Tim Penguji

   Ketua : **Anie Rose Irawati, S.T. M.Cs** .....

   Penguji I
   Bukan Pembimbing : **Aristoteles, S.Si., M.Si.** .....

   Penguji II
   Bukan Pembimbing : **Tristiyanto, S.Kom., M.I.S., Ph.D** .....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T.
NIP 197407052000031001

Tanggal Lulus Ujian Skripsi: **20 Oktober 2022**

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Sasya Salsabila Janerdi

NPM : 1867051001

Dengan ini menyatakan bahwa skripsi saya yang berjudul **"Analisis Sentimen Aplikasi PeduliLindungi Pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier"** 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 dikemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung, 13 Desember 2022

SASYA SALSABILA JANERDI

NPM. 1867051001

RIWAYAT HIDUP

Penulis dilahirkan di Bandar Lampung pada tanggal 08 Januari 2000. Sebagai anak kedua dari dua bersaudara dari pasangan Bapak Riadi Burniat, S.Sos dan Ibu Erwanalita. Penulis menyelesaikan pendidikan formal pertamanya di Taman Kanak Kanak (TK) Dharma Wanita pada tahun 2005, setelah itu melanjutkan pendidikan Sekolah Dasar (SD) di SD N 2 Rajabasa dan selesai pada tahun 2012. Kemudian pendidikan menengah pertama di SMP N 28 Bandar Lampung yang diselesaikan pada tahun 2015, lalu melanjutkan ke pendidikan menengah atas di SMA N 1 Natar 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 prestasi khusus. Selama menjadi mahasiswa, penulis aktif melakukan beberapa kegiatan, antara lain:

1. Menjadi anggota Adapter Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2018/2019.
2. Menjadi anggota bidang Internal Himpunan Mahasiswa Jurusan Ilmu Komputer periode 2018/2019.

3. Menjadi Bendahara Pelaksana Acara Kompetisi Olahraga Jurusan Ilmu Komputer (Kokom) 2019/2020
4. Menjadi Sekretaris Pelaksana Acara PRJ VIII Jurusan Ilmu Komputer 2019/2020.
5. Menjadi Sekretaris Divisi Komisi Disiplin (Komdis) 2019/2020
6. Melaksanakan kerja praktik pada bulan Februari periode 2020/2021 di Jasaraharja Putera Cabang Lampung
7. Melaksanakan KKN di Desa Susunan Baru, Kecamatan Tanjung Karang Barat, Kota Bandar Lampung, Lampung pada tahun 2020/2021.

MOTTO

“Apa yang melewatkanku tidak akan pernah menjadi takdirku, dan apa yang ditakdirkan untukku tidak akan pernah melewatkanku”

(Umar bin Khattab)

“Sesungguhnya bersama kesulitan itu ada kemudahan”

(Al – Insyirah : 6)

“Jika sesuatu yang kau senangi tidak terjadi, maka senangilah apa yang terjadi”

(Ali bin abi thalib)

“If you don’t believe in yourself, no one will do it for you ”

(Kobe Bryant)

PERSEMBAHAN

Alhamdulillahirobbilalamin

Puji dan syukur tercurahkan kepada Allah Subhanahu Wa Ta'alaa atas segala Rahmat dan Karunia-Nya sehingga saya dapat menyelesaikan skripsi ini. Shalawat serta salam selalu tercurahkan kepada Nabi Muhammad SAW.

Kupersembahkan karya ini kepada:

Kedua Orang Tuaku Tercinta

Yang senantiasa memberikan yang terbaik, dan melantukan do'a yang selalu menyertaiku. Kuucapkan pula terima kasih sebesar-besarnya karena telah mendidik dan membesarkanku dengan cara yang dipenuhi kasih sayang, dukungan, dan pengorbanan yang belum bisa terbalaskan.

Kakakku

Terima kasih telah memberikan semangat, dukungan, dan doa.

Almamater Tercinta, Universitas Lampung dan Jurusan Ilmu Komputer

Tempat bernanung mengemban semua ilmu untuk menjadi bekal hidup.

SANWACANA

Puji syukur kehadirat Allah SWT atas berkah, rahmat dan hidayat-Nya, serta petunjuk dan pedoman dari Rasulullah Nabi Muhammad Sholallahu Alaihi Wasallam penulis dapat menyelesaikan skripsi yang berjudul "Analisis Sentimen Aplikasi PeduliLindungi Pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier" dengan baik dan lancar.

Terima kasih penulis ucapkan kepada semua pihak yang telah membantu dan berperan besar dalam menyusun skripsi ini, antara lain.

1. Mami, Papi, Kakak serta Keluarga tercinta yang selalu memberi dukungan, do'a, semangat, motivasi, dan kasih sayang yang luar biasa tak terhingga. Semoga Allah SWT selalu memberikan kebahagiaan dan keberkahan dalam kehidupan kalian di dunia dan akhirat
2. Ibu Anie Rose Irawati, ST, M.Cs. sebagai pembimbing utama yang telah memberikan arahan, ide, kritik serta saran kepada penulis sehingga dapat menyelesaikan skripsi ini dengan baik.
3. Bapak Aristoteles, S.Si., M.Si. sebagai pembahas yang telah memberikan masukan yang bermanfaat dalam perbaikan skripsi ini.
4. Bapak Tristiyanto, S.Kom., M.I.S., Ph.D sebagai pembahas yang telah memberikan masukan yang bermanfaat dalam perbaikan skripsi ini.

5. Ibu Yunda Heningtyas, M. Kom. selaku pembimbing akademik penulis yang selalu mendukung peningkatan akademik penulis.
6. Bapak Didik Kurniawan, S.Si., M.T. selaku Ketua Jurusan Ilmu Komputer FMIPA Universitas Lampung.
7. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc. selaku Sekretaris Jurusan Ilmu Komputer FMIPA Universitas Lampung.
8. Ibu Ade Nora Maela yang telah membantu segala urusan administrasi penulis di Jurusan Ilmu Komputer.
9. Bapak dan Ibu Dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung yang telah memberikan ilmu dan pengalaman dalam hidup untuk menjadi lebih baik.
10. Rahmayanti Kurniasih, Lita Amelia, Amara Indah Pancarani, Ahmad Julio Rizki, M Arsyi Sobirin dan Aulia Ahmad Nabil selaku teman seperjuangan penulis yang telah mendukung memberi semangat, segala bentuk bantuan dan do'a dalam proses skripsi.
11. Aisha Rahmayanti, Athiyyah Berlianda, Kurniati dan Yolanda Olivia selaku sahabat penulis yang telah mendukung memberi semangat dan doa kepada penulis dalam menyelesaikan skripsi ini.
12. Teman-teman Ilmu Komputer 2018 yang menjadi keluarga satu angkatan selama menjalankan proses skripsi.
13. Teman-teman Himakom yang sudah mengajarkan banyak hal dalam berorganisasi, memberikan banyak pengalaman, berjuang bersama memajukan organisasi dengan membawa nama baik Jurusan Ilmu Komputer.

14. Last but not least, I wanna thank me, I wanna thank me for believing in me, I wanna thank me for doing all this hard work, I wanna thank me having no days off, I wanna thank me for never quitting, for just being me at all times

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, semoga skripsi ini membawa manfaat dan keberkahan bagi semua civitas Ilmu Komputer Universitas Lampung.

Bandar Lampung, 13 Desember 2022

Sasya Salsabila Janerdi
NPM. 1867051001

DAFTAR ISI

DAFTAR TABEL

Tabel Halaman

DAFTAR GAMBAR

Gambar Halaman

I. PENDAHULUAN

1.1. Latar Belakang

Coronavirus Disease 2019 (Covid-19) adalah wabah penyakit pernafasan yang mengancam bagi kesehatan global. Covid-19 yang diidentifikasi pada Desember 2019 dengan cepat terbukti disebabkan oleh virus corona baru yang secara struktural terkait dengan virus penyebab sindrom pernafasan akut parah (SARS). Kota Wuhan di provinsi Hubei, Cina menjadi pusat wabah covid-19 pertama kali. Penyebaran Covid-19 telah menimbulkan tantangan kritis bagi kesehatan masyarakat, penelitian, dan komunitas medis. Indonesia adalah negara terpadat keempat di dunia dan berdasarkan data pada Februari sudah 4.901.328 ribu jiwa yang terpengaruh secara signifikan oleh Covid-19 dalam periode yang lebih lama (Noviriandini, A. et al., 2022) . Pandemi ini telah menghambat berbagai aktivitas kehidupan di seluruh dunia termasuk Indonesia, hingga berbagai upaya telah dilakukan pemerintah untuk mencegah penyebaran virus dengan tujuan untuk mengendalikan penyebaran virus corona.

Beberapa bentuk upaya yang telah dilakukan pemerintah diantaranya yaitu diberlakukannya Pembatasan Sosial Berskala Besar (PSBB), himbauan untuk melakukan *social distancing*, pemakaian masker, rajin untuk cuci tangan dan mengembangkan vaksin. Vaksin tidak hanya melindungi mereka yang divaksinasi tetapi juga masyarakat luas dengan mengurangi penyebaran penyakit dalam populasi (Sari & Sriwidodo, 2020). Selain itu, pemerintah telah mengembangkan aplikasi bernama PeduliLindungi sebagai upaya dalam melakukan *tracing*, *tracking*, dan pemberi peringatan selama pandemi COVID-19 bagi masyarakat Indonesia (Ainiyah, 2022).

Pada Februari 2022 terdapat 40 juta lebih yang mengunduh aplikasi PeduliLindungi. Selain itu aplikasi tersebut memiliki penilaian dengan total

rata – rata 3,5 yang diberikan oleh pengguna. Oleh karena itu dengan menimbang pentingnya penggunaan aplikasi PeduliLindungi untuk mengendalikan Covid-19 maka aplikasi PeduliLindungi perlu terus dikembangkan untuk memperbaiki kualitas layanan yang salah satunya dapat menggunakan opini masyarakat. Aplikasi yang lebih stabil meningkatkan rasa nyaman bagi penggunanya sehingga menjamin penggunaan yang berkelanjutan.

Masyarakat Indonesia memiliki berbagai opini terhadap kebijakan penggunaan aplikasi PeduliLindungi, baik opini yang bersifat negatif maupun positif. Opini dari masyarakat yang diekspresikan dengan berbagai media , salah satu nya media sosial Twitter dengan popularitas tertinggi di Indonesia. Menurut (Statista, 2022) tercatat ada sebanyak 17.55 juta pengguna Twitter di Indonesia, menempatkan negara keenam terbesar pengguna Twitter setelah Amerika Serikat, Jepang, India, Brazil, dan Inggris per Oktober 2021. Besarnya komunitas pengguna Twitter di Indonesia tentunya berpengaruh pada ragam opini mengenai aplikasi PeduliLindungi. Ini menjadi peluang sumber data yang sangat besar dan dapat dimanfaatkan untuk analisis sentimen terhadap banyak permasalahan, terutama masalah penggunaan aplikasi PeduliLindungi di Indonesia.

Analisis sentimen merupakan pengolahan data tekstual yang bertujuan untuk klasifikasi polaritas dari teks. Klasifikasi tersebut bertujuan untuk melihat polaritas dari suatu tweet apakah bersifat positif, negatif, atau netral (Fanissa et al., 2018). Berdasarkan penelitian yang telah dilakukan, algoritma yang tepat untuk klasifikasi sentimen adalah algoritma *Naive Bayes*, karena algoritma ini mudah untuk dipahami, lebih cepat dalam hal perhitungan dan hanya memerlukan sedikit data training (Rahman et al., 2020) . Rahman et al dengan penelitian berjudul "Metode *Naive Bayes* Untuk Menganalisis Akurasi Sentimen Komentar di Youtube" mempelajari sentimen yang diungkapkan dalam komentar menunjukkan apakah *Word War III* (WW3) mendapatkan umpan balik positif atau negatif, dengan hasil pengujian sebesar 1500 data

menghasilkan sentimen positif sebanyak 30.3% dan sentimen negatif sebanyak 60.6% dengan akurasi sebesar 78.17%. Selanjutnya (Buntoro, 2017) dengan penelitian berjudul "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter". Untuk proses klasifikasinya menggunakan metode Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM). Akurasi tertinggi didapat saat menggunakan metode klasifikasi Naïve Bayes Classifier (NBC), dengan nilai rata-rata akurasi mencapai 95%, nilai presisi 95%, nilai *recall* 95% nilai TP rate 96,8% dan nilai TN rate 84,6%. Dalam penelitian ini juga dapat diketahui metode klasifikasi *Naïve Bayes Classifier* (NBC) lebih tinggi akurasinya untuk klasifikasi sentimen Tweet Bahasa Indonesia dibandingkan dengan metode klasifikasi Support Vector Machine (SVM). Selanjutnya penelitian dilakukan oleh (Simanjutak, 2018) berjudul "Analisis Sentimen Pada Layanan Gojek Indonesia". Dengan menggunakan metode *Naive Bayes* menghasilkan akurasi sebesar 92,30% .

Berdasarkan penjelasan tersebut, algoritma yang digunakan dalam penelitian ini adalah algoritma *Naive Bayes Classifier* untuk klasifikasi opini. Opini yang digunakan terdapat 2 kelas yaitu positif dan negatif. Penelitian ini akan menganalisis sentimen yang mengkaji tentang Aplikasi PeduliLindungi melalui cara pengumpulan data *tweet* di Twitter dan menganalisis data tersebut dengan *tools* tertentu yang berjudul Analisis Sentimen Aplikasi PeduliLindungi Pada Media Sosial Twitter Menggunakan Metode *Naive Bayes Classifier*.

1.2. Rumusan Masalah

Berdasarkan uraian dari latar belakang , rumusan masalah yang menjadi fokus dari penelitian ini yaitu Bagaimana memberikan informasi sentimen pengguna aplikasi PeduliLindungi berdasarkan opini pengguna di twitter.

1.3. Batasan Masalah

Dalam penelitian ini penulis memberikan batasan masalah , yaitu :

1.3.1. Data tweet yang digunakan hanya yang menyebutkan kata kunci yang terkait dengan aplikasi PeduliLindungi

1.3.2. Tweet yang digunakan hanya yang berbahasa Indonesia

1.3.3. Tweet yang digunakan yaitu tweet pada September 2021 – Februari 2022

1.3.4. Algoritme yang digunakan adalah *Naive Bayes Classifier*

1.4. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk melakukan analisis sentimen berdasarkan tweet terhadap data yang didapatkan dari Twitter dengan menggunakan algortima *Naive Bayes*.

1.5. Manfaat Penelitian

Manfaat yang didapat dalam melakukan penelitian ini yaitu mendapatkan informasi mengenai kualitas aplikasi PeduliLindungi untuk pengembangan aplikasi oleh pihak terkait.

II. LANDASAN TEORI

2.1. Penelitian Terdahulu

Penelitian terdahulu dalam penelitian ini telah dilakukan oleh orang lain yang sejenis bahkan menjadi acuan penelitian ini. Terdapat beberapa penelitian yang memiliki topik yang sama terkait dengan analisis sentimen dapat dilihat pada Tabel 1.

Tabel 1. Penelitian Terdahulu

No.	Penelitian	Data	Metode	Hasil
1.	Analisis Sentiment Berdasarkan Ulasan Komentar Terhadap Aplikasi PeduliLindungi Menggunakan Metode Naive Bayes (Zefanya et al., 2020)	Jumlah : 496 Tweet(kalimat)	Metode Klasifikasi: *Naive Bayes*	Akurasi: 52%
2.	Analisis Sentimen Keputusan Pemindahan Ibukota Negara Menggunakan Klasifikasi Naive Bayes (Amar P, 2019)	Jumlah : 200 Tweet(kalimat) Positif : 100 Tweet(kalimat) Negatif : 100 Tweet(kalimat)	Metode Klasifikasi: *Naive Bayes* Fitur seleksi: *Term Frequency*	Akurasi : 89,86%

3.	Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier (Astari et al., 2020)	Jumlah : 796 Tweet(kalimat) Positif : 290 Tweet(kalimat) Negatif : 506 Tweet(kalimat)	Metode Klasifikasi: *Naive Bayes*	Akurasi : 73%
4.	Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification (Naraswati et al., 2021)	Jumlah: 2330 Tweet(kalimat) Positif : 646 Tweet(kalimat) Negatif :1684 Tweet(kalimat)	Metode Klasifikasi: *Naive Bayes*	Akurasi : 87,34% Sensitivitas : 93,43% spesifisitas : 71,76%
5.	Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier (Sudiantoro & Zuliarso, 2018)	Jumlah : 100 Tweet(kalimat) Positif : 32 Tweet(kalimat) Negatif : 68 Tweet(kalimat)	Metode Klasifikasi: *Naive Bayes*	Akurasi : 84%
6.	Analisis Sentimen Twitter dengan Klasifikasi Naive Bayes Menggunakan Seleksi Fitur Mutual	Jumlah: 3195 Tweet(kalimat) Positif : 1065 Tweet(kalimat)	Metode Klasifikasi : *Naive Bayes* *Mutual Information* (MI) dan	MI: 71.89% IDF: 60.67%

Information Dan Inverse Document Frequency (Putra, 2017)	Negatif: 1065 Tweet(kalimat) Netral : 1065 Tweet(kalimat)	*Inverse Document Frequency* (IDF)

Berdasarkan Tabel 1 diketahui bahwa ada beberapa penelitian yang melakukan sentimen analisis menggunakan data tweet. Dari penelitian tersebut diketahui bahwa penggunaan Algoritma Naive Bayes menghasikan akurasi yang baik dalam rentang minimal 52% sampai dengan 89,86% .

2.2. Text Mining

Tujuan penting dari *text mining* adalah untuk mendapatkan informasi berkualitas tinggi dari teks. Hal ini biasanya dilakukan untuk mengekstraksi pengetahuan dan informasi dari pola dalam teks dokumen (Hermanto et al., 2020). Hal ini juga dilakukan dengan menemukan pola dan tren dengan cara seperti pembelajaran pola statistik, topik, dan pemodelan bahasa statistik. Penambangan teks umumnya membutuhkan penataan teks input (misalnya parsing, bersama dengan menambahkan beberapa fitur dan penghapusan linguistik turunan lainnya, dan penyisipan berikutnya ke dalam database). Ini diikuti oleh menurunkan pola dalam data terstruktur, dan evaluasi dan interpretasi keluaran. "Kualitas tinggi" dalam teks pertambangan biasanya mengacu pada kombinasi relevansi, minat, dan sesuatu yang baru.

2.3. Analisis Sentimen

Analisis sentimen adalah proses menentukan sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat sehingga kategori dapat ditentukan sebagai sentimen positif , negatif atau netral

(Ramadhan & Erwin, 2019). Saat ini, peneliti secara luas menggunakan analisis sentimen sebagai salah satu cabang penelitian dalam ilmu komputer. Jejaring sosial, seperti Twitter, umumnya digunakan dalam analisis sentimen untuk menentukan persepsi publik (Rifan et al., 2019).

Dalam analisis sentimen, data mining dilakukan untuk menganalisis, memproses dan mengambil data teks dalam suatu entitas seperti layanan, produk, orang, fenomena, atau topik tertentu. Proses analisis dapat mencakup teks ulasan, forum, tweet, atau blog, dengan data preprocessing mencakup proses tokenization, stopword, penghapusan, stemming, identifikasi sentimen, dan klasifikasi sentimen (Watrianthos et al., 2019).

2.4. Twitter

Twitter adalah layanan jejaring sosial online dimana pengguna terdaftar dapat memposting pesan, yang dikenal sebagai "tweet". Pesan-pesan ini awalnya hanya dibatasi hingga 140 karakter, tetapi pada 7 November 2017, batasnya digandakan menjadi 280 karakter untuk semua bahasa kecuali Jepang, Korea, dan Cina. Pengguna terdaftar dapat memposting tweet, tetapi bagi mereka yang tidak terdaftar hanya dapat membacanya saja. Pengguna terdaftar dapat mengunggah, menyukai, dan me-retweet tweet, tetapi pengguna yang tidak terdaftar hanya dapat membacanya (Brian, 2018) .

Twitter ini juga bisa menjadi salah satu media sosial untuk bisa berkomunikasi dengan orang-orang yang dikenal maupun yang tidak dikenal. Selain itu , jejaring media sosial twitter juga seringkali digunakan sebagai tempat menyampaikan tanggapan atau pendapat mengenai sesuatu hal berupa cuitan atau tweet yang dapat berupa tanggapan secara positif maupun negatif. Twitter menjadi wadah bagi para penggunanya untuk menyalurkan apa yang mereka rasakan, alami, dan juga bisa memberikan respon terhadap apa yang dunia sedang alami melalui suatu fenomena ataupun kejadian yang sedang terjadi (Wira et al., 2019).

2.5. Klasifikasi

Klasifikasi merupakan pengelompokkan fakta yang memenuhi kriteria tertentu. Klasifikasi dibagi menjadi dua kelompok, yaitu klasifikasi sederhana dan klasifikasi kompleks. Klasifikasi sederhana adalah klasifikasi yang mengelompokkan objek ke dalam dua kategori atau kelas. Namun, klasifikasi kompleks akan mengelompokkan objek ke dalam tiga kategori atau lebih. Proses Klasifikasi adalah proses pengelompokan objek menurut kelas yang ada. Pengklasifikasian suatu data perlu melalui 2 proses terlebih dahulu. Proses awal yang perlu dilakukan adalah pelatihan atau training yang dilakukan untuk menganalisis data latih untuk menjadi model prediksi. Setelah proses pelatihan terpenuhi, baru dijalankan proses klasifikasi. Proses klasifikasi dilakukan untuk mengestimasi akurasi data yang didapat dari hasil model prediksi yang diuji dengan data test atau uji. Jika akurasi yang didapat sesuai, maka model tersebut dapat digunakan untuk prediksi kelas atau kategori data yang belum diketahui (Wanda Athira, 2018).

2.6. Naive Bayes Classifier

Naive Bayes Classifier merupakan metode klasifikasi yang berdasar pada teorema Bayes. Metode klasifikasi ini cocok digunakan ketika jumlah masukan yang sangat besar. Klasifikasi ini lebih disukai karena kecepatan dan kesederhanaannya (Goel et al., 2016). Meskipun klasifikasi ini bisa dibilang klasifikasi yang sederhana, namun hasil yang diperoleh dari klasifikasi ini sering mencapai performa yang sebanding dengan algoritme lain seperti *Decision tree* dan *Neural Network classifier*. Keuntungan menggunakan algoritma ini adalah hanya membutuhkan sedikit data latih (training data) untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi. Karena diasumsikan sebagai variabel bebas , maka hanya membutuhkan varian dari suatu variable dalam kelas untuk menentukan klasifikasi , bukan seluruh matriks kovarians (Muslehatin et al., 2017). Untuk menyelesaikan metode Naive Bayes, dapat menggunakan persamaan - persamaan sebagai berikut :

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

Keterangan :

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

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

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

P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) = Probabilitas dari X

Penjabaran lebih lanjut rumus Naïve Bayes tersebut dilakukan dengan menjabarkan secara terperinci (C|X1…,Xn) menggunakan aturan perkalian sebagai berikut :

P(C|x1,.....,xn = P C P x ,....., x | C ( ) (1 n)

= P(C)P(X1|C)P(X2,….,Xn|C,X1)

= (C)P(X1|C)P(X2|C,X1)P(X3,...Xn|C,X1,X2(C)P(X1|C)

P(X2|C,X1)P(X3|C,X1,X2)P(X4,....,Xn|C,X1,X2,X3)P(C)

= P(X1|C)P(X2|C,X1)P(X3|C,X1,X2)....

P(Xn|C,X1,X2,X3,.....Xn- 1..... (2)

Jika semakin banyak faktor-faktor yang semakin kompleks yang berpengaruh terhadap nilai probabilitas, maka semakin tidak mungkin untuk menghitung nilai tersebut satu persatu. Proses perhitungan akan semakin susah untuk dilakukan, maka disinilah digunakan asumsi independensi yang sangat tinggi, bahwa masing masing atribut dapat saling bebas. Dengan asumsi tersebut, diperlukan persamaan 3 :

$$P(X_i|X_j) = \frac{P(X_i|H) * P(H_j)}{P(X_j)} = \frac{P(X_i \cap H_j)}{P(X_j)} = P(X_i)$$

Untuk i≠j, sehingga

$$P(X_i/C, X_j = P(X_i/C) \quad (3)$$

Dari persamaan diatas dapat di ambil kesimpulan bahwa asumsi independensi membuat syarat perhitungan menjadi lebih sederhana. Selanjutnya penjabaran (P(C|X1,…..,Xn) dapat disederhanakan menjadi persamaan 4 :

$$P(X_2/C)P(X_3/C)...P(C/X_1,...X_n) = P(X_1/C) = \prod_{i=1}^{n} P(Xi\,|C) \quad (4)$$

Keterangan :

$\prod_{i=1}^{n} P(Xi\,|C)$ = perkalian ranting antar atribut.

Persamaan 4 merupakan teorema bayes yang kemudian akan digunakan untuk melakukan perhitungan klasifikasi. Untuk klasifikasi dengan data continue atau data angka menggunakan rumus distribusi Gaussian dengan 2 parameter : mean μdan varian σ :

$$P(X_{\text{i}} = X_{\text{i}} \mid C = c_j) = \frac{1}{\sqrt{2\pi\sigma ij}} \exp \frac{(xi - \mu ij)}{2\sigma 2\text{ij}} \quad (5)$$

Keterangan :

P : Peluang

Xi : Atribut ke i

Xj : Nilai atribut ke i

C : Kelas yang dicari

Ci : Sub kelas Y yang dicari

μ : Menyatakan rata–rata dari seluruh atribut

σ : Deviasi standar, menyatakan varian dari seluruh atribut.

Dalam metode Naïve Bayes diperlukan data latih dan data uji yang ingin diklasifikasikan. Semakin banyak data latih yang yang dilibatkan, semakin baik hasil yang prediksi yang diberikan. Menghitung P(Ci) yang merupakan

probabilitas prior untuk setiap sub kelas C yang akan dihasilkan menggunakan persamaan 6 :

$$P(Ci) = \frac{Si}{S} \quad (6)$$

Si adalah jumlah data training dari kategori Ci, dan s adalah jumlah total data training. Menghitung P(Xi|Ci) yang merupakan probabilitas posterior Xi dengan syarat C menggunakan persamaan 4.

2.7. Bahasa Pemrograman Python

Python adalah bahasa pemrograman interpretatif multiguna yang filosofi desainnya mengacu pada sejauh mana kode terbuka. Python dibuat oleh Guido van Rossum dan pertama kali dirilis pada tahun 1991. Filosofi desain Python menekankan pembacaan kode dengan penggunaan spasi yang signifikan. Bahasa dan pendekatan berorientasi objek Python dirancang untuk membantu programmer menulis kode yang dapat dimengerti dan logis untuk proyek kecil dan besar (John V, 2016).

Python merupakan *programming languange* yang sukses dalam peningkatan pencarian yang sangat baik dalam beberapa tahun terakhir. Selain dari perubahan *library* yang semakin bagus dan baik, kontibutor dari Python yang banyak akhirnya menciptakan Python menjadi salah satu *programming languange* yang solid dan berkembang cepat. Bahasa pemrograman Python memiliki beberapa *library* dan *framework* yang digunakan untuk melakukan analisis data. Beberapa *library* yang digunakan dalam penelitian ini sebagai berikut :

1. Pandas

   Library ini bersifat *open source* ini menyediakan struktur data tingkat tinggi yang fleksibel serta berbagai alat analisis. Pandas digunakan untuk memproses data yang meliputi analisis data, manipulasi data, dan pembersihan data.
2. *Natural Languange ToolKit (*NLTK)

Natural Language Toolkit (NLTK) adalah sebuah platform yang digunakan untuk mempermudah proses data teks. Platform ini awalnya dirilis oleh Steven Bird dan Edward Loper pada tahun 2001.

3. Sastrawi

   Sastrawi salah satu *library* yang digunakan dalam melakukan proses *stemming* bahasa Indonesia menjadi bentuk dasarnya. Sastrawi merupakan pengembangan dari proyek PHP Sastrawi.

4. Scikit Learn

   Scikit-learn adalah *library* python terkenal yang digunakan untuk data kompleks. Perpustakaan *open source* ini mendukung *machine learning* dengan mendukung berbagai algoritma yang diawasi dan tidak diawasi seperti regresi linier, klasifikasi, pengelompokan, dan lain sebagainya.

5. NumPy (Numerical Python)

   Numpy adalah *library python* yang digunakan untuk bekerja dengan *array* dan juga memiliki fungsi yang bekerja dalam domain aljabar linier, transformasi fourier, dan matriks. *Library* yang dibuat pada 2005 oleh Travis Oliphant ini merupakan proyek *open source* sehingga dapat digunakan secara bebas.

6. TfidfVectorizer

   Merupakan *Machine Learning* berdasarkan TF-IDF yang khusus mengolah kata - kata dari sebuah dokumen.

7. *Countvectorizer*

   Countvectorizer berfungsi untuk menghitung frekuensi kata dalam dokumen. *Countvectorizer* dapat mengubah fitur teks menjadi sebuah representasi vector.

2.8. Aplikasi PeduliLindungi

PeduliLindungi merupakan Aplikasi yang dipergunakan dalam pelaksanaan surveilans kesehatan dalam menangani penyebaran covid-19, dengan menyelenggarakan *Tracing* yaitu, melakukan pelacakan terhadap orang-orang

yang berkontak dengan orang yang diduga mengidap covid-19, selain itu juga *Tracking* yaitu melacak persebaran virus corona dengan melihat siapa saja yang telah bertemu dengan penderita virus corona dan menyelenggaraakan *Warning and Fencing* yaitu adanya peringatan dan pengawasan dengan membatasi pergerakan seseorang yang sedang dalam karantina atau isolasi (Nurhidayati et al., 2021). Aplikasi PeduliLindungi digunakan pada masa darurat covid-19. Agar bisa melacak riwayat kontak dengan pasien Covid-19, peran serta masyarakat sangat diperlukan, dengan saling berbagi data lokasi saat bepergian.

Aplikasi PeduliLindungi akan merekam data pergerakan pasien selama 14 hari terakhir, aplikasi terhubung dengan telepon seluler untuk menghasilkan visualisasi pergerakan, sistem aplikasi akan memberikan peringatan melalui ponsel orang orang disekitar pasien yang terdeteksi agar menjalankan protokol ODP (orang dalam pemantauan) (Kompas, 2020).

2.9. *Term Frequency - Inverse Document Frequency* (TF-IDF)

Data yang telah melalui tahap preprocessing harus berbentuk numerik. Untuk mengubah data tersebut menjadi numerik yaitu menggunakan metode pembobotan TF-IDF. Metode *Term Frequency Invers Document Frequency* (TF-IDF) merupakan metode yang digunakan menentukan seberapa jauh keterhubungan kata (term) terhadap dokumen dengan memberikan bobot setiap kata. Perhitungan ini perlu dilakukan untuk menentukan seberapa relevan sebuah kata di dalam dokumen (Nurjannah et al., 2016) . Metode TF-IDF ini menggabungkan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut (Herwijayanti et al., 2018). Dalam perhitungan bobot menggunakan TF-IDF, dihitung terlebih dahulu nilai TF tiap kemunculan kata diberikan bobot 1. Sedangkan nilai IDF diformulasikan pada Persamaan (1)

$$\text{IDF (Word)} = log\ \frac{td}{df} \quad (1)$$

DF adalah nilai IDF dari setiap kata yang akan dicari, *td* adalah jumlah keseluruhan dokumen yang ada, sedangkan *df* adalah jumlah kemunculan kata pada semua dokumen.

2.10. K-fold Cross Validation

Cross-validation atau bisa dianggap perkiraan pemutar adalah suatu cara validasi untuk penilaian hasil analisis statistik yang akan menormalkan kombinasi data independen. *Cross validation* mampu bekerja dengan cepat dengan pengambilan sampel yang lebih struktur, jadi dalam jumlah pengujian beberapa pun set data latih dan set data uji akan diambil dengan data yang berbeda dengan percobaan atau literasi sebelumnya. Cara ini dipakai untuk prediksi contoh & memprediksi beberapa contoh prediktif saat digunakan pada penerapannya. Satu caranya adalah validasi silang merupakan *k-fold cross validation*, maksudnya adalah memecah data sebagai k bagian dari kumpulan data menggunakan ukuran yang sama. Penggunaan *k-fold cross validation* untuk menghilangkan penyimpangan dalam data. Data latih dan data uji dilakukan sebanyak k kali. Pada percobaan pertama, subset S1 diperlakukan sebagai data uji dan subset lainnya diperlakukan sebagai data latih, dalam percobaan ke 2 subset S1, S3,…Sk sebagai data latih dan S2 sebagai data uji, dan setererusnya (Tempola et al., 2018) .

Dalam penelitian ini k yang digunakan adalah 10-*fold* yang berarti akan dilakukan 10 kali dengan posisi data tes berbeda di setiap iterasi nya pada seluruh isi dokumen secara acak. 10 *fold cross validation* adalah salah satu k-*fold cross validation* yang direkomendasikan untuk pemilihan model terbaik karena cenderung memberikan estimasi akurasi yang kurang bias dibandingkan dengan *cross validation* biasa, *leave-one-out cross validation* dan bootstrap (Yunitasari & Putera, 2021).

2.11. Lexicon Based

Lexicon Based merupakan metode yang sederhana, layak, dan praktis untuk analisis sentimen. Data yang bisa digunakan berasal dari media sosial seperti

Twitter, Facebook, dan media sosial lain mengenai opini suatu produk atau layanan jasa (Matulatuwa et al., 2017). Keuntungan yang diperoleh dengan metode Lexicon Based adalah tidak membutuhkan data berlabel dan prosedur pembelajaran (Devika et al., 2016). *Lexicon Based* menggunakan kata-kata yang dinilai berdasarkan polaritasnya untuk mengetahui opini masyarakat.

III. METODOLOGI PENELITIAN

3.1. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung yang berlokasi di Jalan Soemantri Brojonegoro No.1 Gedong Meneng, Bandar Lampung. Penelitian ini dilakukan mulai pada bulan Januari 2022 hingga sampai bulan Agustus 2022.

3.2. Perangkat Penelitian

3.2.1. Perangkat Keras (Hardware)

Perangkat keras yang digunakan dalam penelitian ini adalah sebagai berikut :

- *Processor* : *Intell® Core™ i3-5005U*
- *Installed RAM* : *8.00 GB*
- *System Type* : *64-bit operating system, x-64-based processor.*

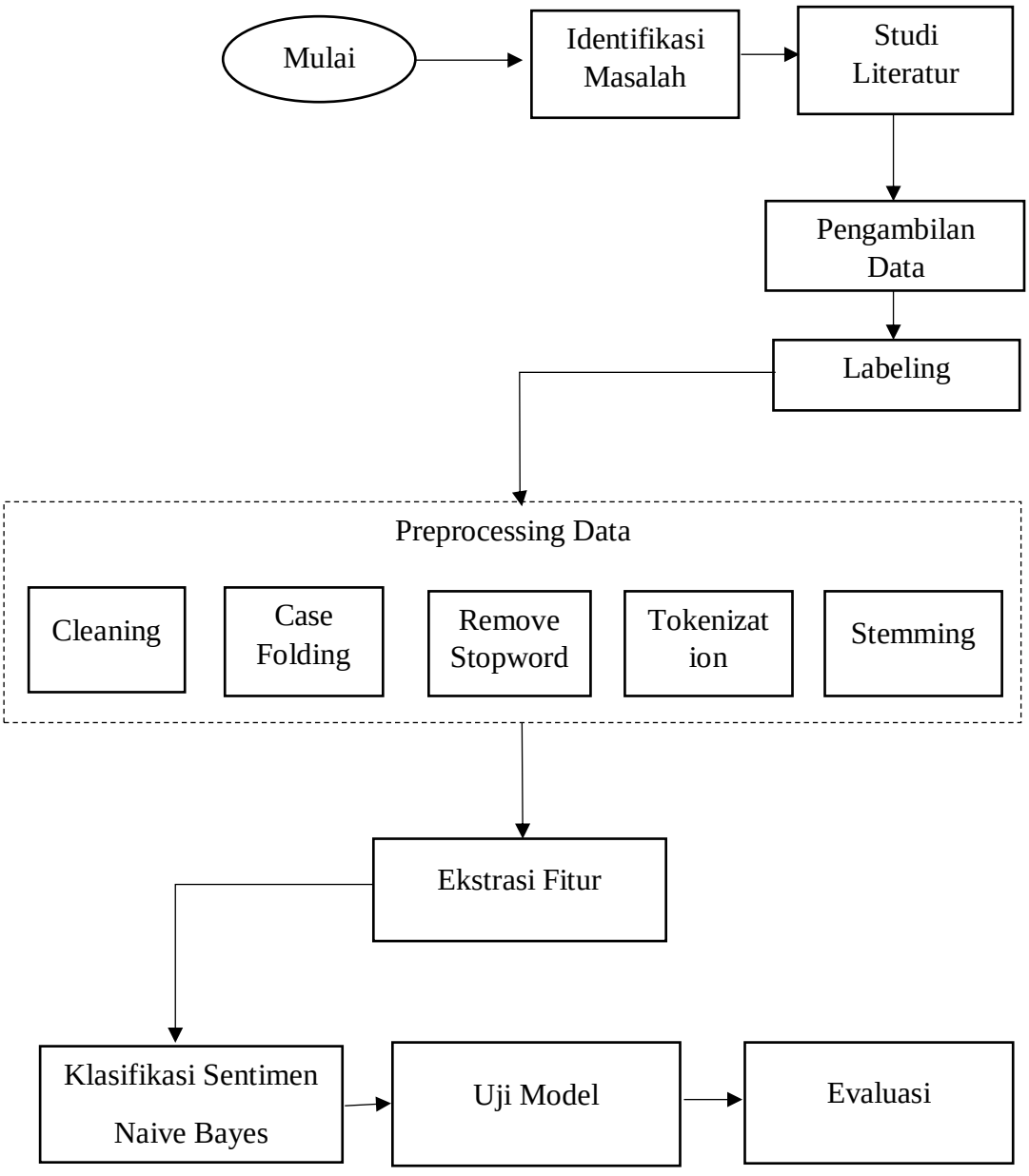
3.2.2. Perangkat Lunak (Software)

Perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut :

- Sistem Operasi : Windows 10
- *Web Browser*: *Google chrome*, digunakan untuk pengunduhan data
- Python 3, digunakan sebagai bahasa pemrograman
- Jupyter Notebook, digunakan untuk membuat script program.

3.3. Tahapan Penelitian

Alur penelitian ini melalui beberapa tahapan yang diilustrasikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

3.3.1. Identifikasi Masalah

Tahapan awal dalam penelitian ini perlu dilakukan penentuan permasalahan yang akan dianalisis secara jelas dan sederhana. Identifikasi masalah ini bertujuan untuk mentransformasi topik sehingga dapat disesuaikan dengan kemampuan dan keterbatasan sumber daya yang ada. Setelah menentukan berbagai aspek masalah yang diperoleh, selanjutnya disusun informasi tentang masalah tersebut untuk dijawab menjadi suatu identifikasi masalah. Seperti yang diketahui dunia termasuk indonesia tengah mengalami penyebaran *Corona Virus Disease* 2019 (COVID-19). Kondisi tersebut menuntut pemerintah harus bergerak cepat dalam memantau penyebaran virus salah satunya dengan membuat aplikasi PeduliLindungi.

3.3.2. Studi Literatur

Pada tahapan ini dilakukan mencari referensi dan penelitian terdahulu yang berkaitan dengan penelitian yang sama. Tujuannya adalah untuk mendapatkan dasar yang cukup dari masalah yang akan dianalisis. Adapun penelitian sebelumnya tentang analisis sentimen, terdapat enam penelitian terdahulu dengan topik yang berbeda dijadikan sebagai studi literatur untuk penelitian ini.

3.3.3. Pengambilan Data

Proses pengambilan data ini yang terlebih dahulu mempunyai akun twitter yang telah didaftarkan menjadi akun developer twitter. Setelah itu , pengumpulan data tweet menggunakan API key yang disediakan oleh Developper Twitter dengan menggunakan *library search tweet.* Data yang dikumpulkan dari Twitter merupakan data teks yang tidak memiliki nilai sentimen positif ataupun negatif. Kata kunci untuk pengambilan data dari Twitter menggunakan kata “Peduli Lindungi”.

3.3.4. Labeling

Proses pelabelan pada setiap data dilakukan secara manual kemudian dilakukan juga untuk memvalidasi hasil labeling menggunakan algoritma lexicon based untuk menentukan apakah data tersebut termasuk sentimen positif atau negatif. Tahapan pelabelan data ini dilakukan karena pada saat pengumpulan data, data-data yang didapatkan belum memiliki nilai sentimen positif ataupun negatif.

3.3.5. Preprocessing Data

Tahap ini mengolah data yang telah diperoleh dari Twitter dengan melakukan preprocessing data. Preprocessing ini dilakukan bertujuan untuk menghilangkan karakter yang tidak relevan dan meningkatkan kualitas data latih yang akan digunakan. Tahapan yang dilakukan saat preprocessing data yaitu *cleaning , case folding,* penghapusan *stopword, tokenization* dan *stemming*. Berikut penjelasan dari tahapan preprocessing data yang dilakukan.

a. *Cleaning* (Pembersihan Data)

   Pembersihan data adalah proses pembersihan kata-kata dengan menghapus koma pembatas (,), titik (.), dan tanda baca lainnya. Atribut yang tidak berpengaruh tersebut akan dihilangkan dari dokumen kemudian akan digantikan dengan karakter spasi. Pembersihan kata yang bertujuan untuk mengurangi *noise* pada data. Terkadang tanda baca, simbol, dan angka dalam komentar pengguna situs membuat data menjadi tidak efektif dan tidak punya arti. Contoh *tweet* dapat dilihat pada Tabel 2.

   Tabel 2. Contoh Proses Penerapan Pembersihan Data

Input	*Output*
Pasien Covid-19 akan mendapati warna hitam pada aplikasi	Pasien Covid akan mendapati warna hitam pada

PeduliLindungi mereka. Artinya, mereka tidak boleh masuk ke fasilitas publik yang mengharuskan memindai aplikasi tersebut. https://t.co/WGByQ49rI5	aplikasi PeduliLindungi mereka Artinya mereka tidak boleh masuk ke fasilitas publik yang mengharuskan memindai aplikasi tersebut

b. Case Folding

Case folding berfungsi untuk menormalkan kata atau kalimat dalam suatu data. *Case folding* tujuannya adalah untuk memperbaiki struktur kata dan kosa kata dalam kalimat. dalam penelitian ini *case folding* akan dilakukan pada semua data teks penelitian yang digunakan. Contoh *tweet* dapat dilihat pada Tabel 3.

Tabel 3. Contoh Proses Penerapan *Case Folding*

Input	*Output*
Pasien Covid akan mendapati warna hitam pada aplikasi PeduliLindungi mereka Artinya mereka tidak boleh masuk ke fasilitas publik yang mengharuskan memindai aplikasi tersebut	pasien covid akan mendapati warna hitam pada aplikasi pedulilindungi mereka artinya mereka tidak boleh masuk ke fasilitas publik yang mengharuskan memindai aplikasi tersebut

c. Remove Stopword

Stopword adalah kata yang sering muncul tetapi jika dihapus tidak mengubah makna dari *tweet* tersebut. Pembuangan *stopword* dimaksudkan untuk pembuangan kata dasar yang tidak memiliki arti atau tidak relevan. Pada tahapan ini menggunakan bantuan

*library Natural Languange Toolkit (*NLTK). Contoh *tweet* dapat dilihat pada Tabel 4.

Tabel 4. Contoh Proses Penerapan *Remove Stopword*

Input	*Output*
pasien covid akan mendapati warna hitam pada aplikasi pedulilindungi mereka artinya mereka tidak boleh masuk ke fasilitas publik yang mengharuskan memindai aplikasi tersebut	pasien covid mendapati warna hitam aplikasi pedulilindungi arti masuk fasilitas publik mengharuskan memindai aplikasi

d. Tokenization

Pada proses tokenisasi akan dilakukan proses pemotongan string input berdasarkan tiap kata yang menyusunnya. Tokenisasi secara garis besar memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata, bagaimana membedakan karakter - karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan. Pada umumnya setiap kata terindentifikasi atau terpisahkan dengan kata yang lain oleh karakter spasi, sehingga proses tokenisasi mengandalkan karakter spasi pada dokumen untuk melakukan pemisahan kata. *Tokenization* dengan bantuan *library* pandas yang ada pada bahasa pemrograman python. Contoh *tweet* dapat dilihat pada Tabel 5.

Tabel 5. Contoh Proses Penerapan *Tokenization*

Input	*Output*
pasien covid mendapati warna hitam aplikasi pedulilindungi arti masuk fasilitas publik mengharuskan memindai aplikasi	[‘pasien’, ‘covid’, ‘mendapati’, ‘warna’, ‘hitam’, ‘aplikasi’, ‘pedulilindungi’, ‘arti’, ‘masuk’, ‘fasilitas’, ‘publik’, ‘mengharuskan’, ‘memindai’, ‘aplikasi’]

e. Stemming

Stemming merupakan proses mengubah kata menjadi kata dasar atau imbuhan di depan ada imbuhan di belakang kata dihapuskan dengan menghilangkan imbuhan pada kata dalam dokumen.Pada tahap ini menggunakan bantuan *library* sastrawi. Sebelum melakukan proses *stemming, library* sastrawi harus di i*nstall* terlebih dahulu dengan perintah pip sastrawi. *Library* sastrawi berfungsi mengubah kata-kata menjadi kata dalam bentuk dasarnya. Contoh *tweet* dapat dilihat pada Tabel 6.

Tabel 6. Contoh Proses Penerapan *Stemming*

Input	*Output*
[‘pasien’, ‘covid’, ‘mendapati’, ‘warna’, ‘hitam’, ‘aplikasi’, ‘pedulilindungi’, ‘arti’, ‘masuk’, ‘fasilitas’, ‘publik’, ‘mengharuskan’, ‘memindai’, ‘aplikasi’]	pasien covid dapat warna hitam aplikasi pedulilindungi arti masuk fasilitas publik harus pindai aplikasi

3.3.6. Ekstraksi Fitur

Setelah melakukan semua tahapan preprocessing, langkah selanjutnya adalah membuat fitur untuk mempermudah proses klasifikasi. Pada tahapan pembuatan fitur ini dilakukan dengan menggunakan *Tf-Idf* untuk pembobotan kata. Membuat kalimat yang sudah menjadi array menjadi matriks. Setiap baris matriks mewakili satu baris dokumen dan kolom dalam matriks mewakili semua kata dalam teks yang ada. Tujuan *Tf-idf* untuk memberikan bobot setiap kalimat atau kata dalam dokumen. Dataset yang sudah siap untuk digunakan pada training menggunakan perhitungan *naive bayes.*

3.3.7. Klasifikasi Sentimen

Proses klasifikasi dalam penelitian ini menggunakan algoritma Naive Bayes. Data yang akan diklasifikasi adalah data tweet yang telah melewati tahap pengolahan data. Untuk mengklasifikasi sentimen, menggunakan data dari praproses hingga pembobotan kata dengan *Tf-Idf*. Setelah berhasil memproses data latih, data uji digunakan untuk memeriksa keakuratan klasifikasi yang dilakukan. Hasil akhir yang diperoleh dari klasifikasi akan disajikan sebagai prediksi sentimen positif dan negatif dalam bentuk *confusion matrix* dan *classification report*, yang isinya adalah akurasi, presisi, *recall*, dan *f1-score* dari hasil klasifikasi.

3.3.8. Uji Model

Proses pengujian model ini dilakukan setelah proses pelatihan data selesai. Pengujian model dilakukan untuk mengetahui kinerja dari model. Jumlah data yang digunakan untk menguji model diambil dari 0,2% data latih. Data diambil secara acak dengan bantuan library python. Setelah menjalankan uji model, maka akan terlihat seberapa akurat metode yang digunakan.

3.3.9. Evaluasi

Evaluasi yang dilakukan adalah melakukan perbandingan nilai sentimen dari pelabelan manual dan analisis sentimen menggunakan algoritma Naive Bayes. Perhitungan yang dilakukan pada tahap ini adalah melihat hasil akurasi dari setiap tahap yang dilakukan penelitian. Hasil klasifikasi yang didapatkan adalah *confusion matrix* dan *classification report* berupa akurasi, presisi, *recall* dan *f1-score.*

Pada *confusion matrix* sendiri berisi informasi prediksi klasifikasi dari data aktual yang dilakukan oleh sistem klasifikasi. Kinerja sistem klasifikasi umumnya dihitung menggunakan data dalam tabel *confusion matrix* pada Tabel 7.

Tabel 7. *Confusion Matrix*

Fakta	**Prediksi**	
	Negatif	**Positif**
Negatif	TP (*True Positif*)	FP (*False Positif*)
Positif	FN (*False Negatif)*	TN (*True Negatif)*

Nilai *True Positive* (TP) merupakan data positif yang terdeteksi benar, sedangkan *False Positive* (FP) merupakan data negatif namun terdeteksi positif, *False Negative* (FN) merupakan data positif yang terdeteksi negatif, sedangkan *True Negative* (TN) merupakan jumlah data negatif yang terdeteksi dengan benar. Tabel *confusion matrix* digunakan untuk mengukur kinerja suatu metode klasifikasi dengan menghitung nilai akurasi, presisi, *recall*, dan *f1-score*.

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

$$Accuracy = \frac{TP + TN}{TP+TN+FP+FN} x\ 100\% \quad (7)$$

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

$$Precision = \frac{TP}{TP+ FP} x\ 100\% \quad (8)$$

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

$$Recall = \frac{TP}{TP+ FN} x\ 100\% \quad (9)$$

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

$$f1\text{-}score = 2\ x\ \frac{(recall\ x\ precission)}{(recall + precision)} \quad (10)$$

V. SIMPULAN DAN SARAN

5.1. *Simpulan*

Berdasarkan penelitian yang telah dilakukan, maka didapatkan hasil sebagai berikut :

1. Hasil penelitian yang sudah dilakukan sentimen aplikasi didapatkan sentimen positif sebesar 969 ulasan dan sentimen negatif sebesar 964 ulasan. Dimana dalam frekuensi kata terdapat 10 besar kata yang sering muncul diantaranya adalah ‘pedulilindungi’, ‘aplikasi’, ‘vaksin’, ‘status’, ‘covid’, ‘tes’, ‘hasil’, ‘pcr’, ‘positif’, ‘hitam’ dengan frekuensi yang tinggi dimana pengguna banyak mengaitkan aplikasi PeduliLindungi dngan status vaksinasi , status kesehatan, dan status hasil tes covid-19.
2. Hasil klasifikasi dari algoritma *Naive Bayes Clasifier* menunjukkan nilai klasifikasi yang cukup baik dengan akurasi 0.85 atau 85% .
3. Tingkat akurasi yang lebih baik dihasilkan dengan melakukan pengujian menggunakan *K-fold cross Validation* dengan k sebesar 10 yang hasil nilai akurasinya 0.89 atau 89%

5.2. Saran

Berdasarkan hasil uji coba yang diperoleh, penulis menyadari bahwa dalam penelitian ini masih memerlukan beberapa perbaikan untuk meningkatkan performa sistem. Oleh karena itu, penulis menyarankan beberapa hal berikut untuk dilakukan bagi pengembangan penelitian selanjutnya:

1. Dataset yang digunakan untuk menganalisa aplikasi PeduliLindungi bisa didapatkan dari media *social* atau aplikasi lain untuk meningkatkan hasil akurasi.

2. Mengambil tweet dengan kisaran waktu yang lebih panjang agar mendapatkan opini yang lebih beragam untuk meningkatkan hasil akurasi.
3. Dapat menggunakan algoritma klasifikasi yang lain sehingga dapat membandingkan hasil uji model yang dilakukan untuk mendapatkan hasil analisis sentimen yang lebih baik.
4. Melakukan perhitungan menggunakan 2 *class* data (positif dan negatif) untuk melihat perbedaan hasil sentimen analisis.

DAFTAR PUSTAKA

Ainiyah, K. 2022. Analisis sentimen terhadap aplikasi pedulilindungi menggunakan seleksi fitur query expansion ranking dengan metode support vector machine. *Doctoral Dissertation, Universitas Islam Negeri Maulana Malik Ibrahim*.

Amar P, N. 2019. Analisis Sentimen Keputusan Pemindahan Ibukota Negara Menggunakan Klasifikasi Naive Bayes. *In SENSITIf: Seminar Nasional Sistem Informasi Dan Teknologi Informasi*, 47–53.

Astari, N. M. A. J., Divayana, D. G. H., & Indrawan, G. 2020. Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Dan Informatika (JSI)*, *15*(1), 27–29.

Brian, A. 2018. Analisis Sentimen Konten Radikal Melalui Dokumen Twitter Menggunakan Metode Backpropagation. *Doctoral Dissertation, Universitas Brawijaya*.

Buntoro, G. A. 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, *2*(1).

Devika, M. D., Sunitha, C., & Ganesh, A. 2016. Sentiment analysis: a comparative study on different approaches. *Procedia Computer Science*, *87*, 44–49.

Fanissa, S., Fauzi, M. A., & Adinugroho, S. 2018. Analisis Sentimen Pariwisata di Kota Malang Menggunakan Algoritma Naive Bayes dan Seleksi Fitur Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*.

Goel, A., Jyoti, G., & Sitesh, K. 2016. Real time sentiment analysis of tweets using Naive Bayes. *In 2016 2nd International Conference on Next*

Generation Computing Technologies (NGCT), 257–261.

Hermanto, H., Mustopa, A., & Yadi Kuntoro, A. 2020. Algoritma Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Layanan Komplain Mahasiswa. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *5*(2), 211–220.

Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. 2018. Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*, 306–312.

John V, G. 2016. *Introduction to computation and programming using Python With application to understanding data*. MIT Press.

Kompas, J. 2020. Vakum Regulasi Aplikasi Pelacak Covid-19 di Indonesia. Kompas.Com. In *https://kompas.id/baca/humaniora/kesehatan/2020/03/29/vakum-regulasi-aplikasi-pelacak-covid-19-di-indonesia/?_t=sIFkSkDbBPvEui33hzbwsZLC4Mjqyl7UrQE2O0lf1ZTtqOLt0C2CGthSFtg.*

Kundi, F. M., Khan, A., Ahmad, S., & Asghar, M. Z. 2018. Lexicon-based sentiment analysis in the social web. *Journal of Basic and Applied Scientific Research, 4(6)*, 238-48.

Matulatuwa, F. M., Sediyono, E., & Iriani, A. 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.

Muslehatin, W., Ibnu, M., & Mustakim, M. 2017. Penerapan Naïve Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau. *In Seminar Nasional Teknologi Informasi Komunikasi Dan Industri*, 250–256.

Naraswati, N. P. G., Nooraeni, R., Rosmilda, D. C., Desinta, D., Khairi, F., & Damaiyanti, R. 2021. Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification. *Sistemasi: Jurnal Sistem Informasi*, *10*(1), 222–238.

Noviriandini, A., Hermanto, H., & Yudhistira, Y. 2022. KLASIFIKASI SUPPORT VECTOR MACHINE BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK ANALISA SENTIMEN PENGGUNA APLIKASI PEDULILINDUNGI. *JIKA (Jurnal Informatika)*, *6(1)*, 50–56.

Nurhidayati, N., Sugiyah, S., & Yuliantari, K. 2021. Pengaturan Perlindungan Data Pribadi Dalam Penggunaan Aplikasi Pedulilindungi. *Widya Cipta: Jurnal Sekretari Dan Manajemen*, *5*(1), 39–45.

Nurjannah, M., Hamdani, H., & Astuti, I. F. 2016. Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) untuk Text Mining. *Jurnal Ilmiah Ilmu Komputer*, *8(3)*, 110–113.

Putra, R. S. 2017. *Analisis Sentimen Twitter dengan Klasifikasi Naïve Bayes menggunakan Seleksi Fitur Mutual Information dan Inverse Document Frequency*.

Rahman, A., Rahmat, F., Fariqi, M. Y., & Adi, S. 2020. metode naive bayes untuk menganalisis akurasi sentimen komentar di youtube. *Jurnal EECCIS (Electrics, Electronics, Communications, Controls, Informatics, Systems)*, *14*(1), 31–34.

Ramadhan, D. A., & Erwin, B. S. 2019. Analisis Sentimen Program Acara di SCTV pada Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine. *EProceedings of Engineering*, *6*(2).

Rifan, F., Kusrini, K., & Wibowo, F. W. 2019. Analisis Sentimen Wisata Jawa Tengah Menggunakan Naïve Bayes. *Jurnal Informa: Jurnal Penelitian Dan Pengabdian Masyarakat*, *5*(3), 55–60.

Sari, I. P., & Sriwidodo, S. 2020. Perkembangan Teknologi Terkini dalam Mempercepat Produksi Vaksin COVID-19. *Majalah Farmasetika*, *5(5)*, 204–217.

Simanjutak, R. A. 2018. *ANALISIS SENTIMEN PADA LAYANAN GOJEK INDONESIA*.

Statista. 2022. *Leading countries based on number of Twitter users as of October*. Https://Www.Statista.Com/Statistics/242606/Number-of-Active-Twitter-Users-in-Selected-Countries/.

Sudiantoro, A. V., & Zuliarso, E. 2018. Analisis sentimen twitter menggunakan text mining dengan algoritma Naïve Bayes Classifier. *Jurnal Dinamika Informatika*, *10*(2), 69–73.

Tempola, F., Muhammad, M., & Khairan, A. 2018. Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *5 (5)*, 577–584.

Wanda Athira, L. 2018. Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Doctoral Dissertation, Universitas Brawijaya*.

Watrianthos, R., Suryadi, S., Irmayani, D., Nasution, M., & Simanjorang, E. F. S. 2019. Sentiment analysis of traveloka app using naïve bayes classifier method. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, *8*(7).

Wira, B., Riski, I., Dwi, K., Nooraeni, R., Siahaan, T., & Dhea, Y. 2019. Analisis Text Mining dari Cuitan Twitter Mengenai Infrastruktur di Indonesia dengan Metode Klasifikasi Naïve Bayes. *Eigen Mathematics Journal*, *1*(2), 92–101.

Yunitasari, Y., & Putera, A. R. 2021. Analisis Sentimen Masyarakat di Twitter

Terkait Pandemi Covid-19. *SMATIKA JURNAL*, *11(01)*, 22–26.

Zefanya, C. F., Akbar, D. R., Titirloloby, Y., Ramadhan, W. R., & Situmorang, F. F. 2020. Analisis Sentiment Berdasarkan Ulasan Komentar Terhadap Aplikasi PeduliLindungi Menggunakan Metode Naive Bayes. *Jurnal Techno Nusa Mandiri*.