

**PERBANDINGAN AKURASI DAN WAKTU PROSES ALGORITMA
NAÏVE BAYES DENGAN ALGORITMA SUPPORT VECTOR MACHINE
DALAM ANALISIS SENTIMEN TWITTER APLIKASI PEDULI
LINDUNGI
SKRIPSI**



disusun oleh
Maulana Rizky Hidayat
18.11.1816

**PROGRAM SARJANA
PROGRAM STUDI SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**PERBANDINGAN AKURASI DAN WAKTU PROSES ALGORITMA
NAÏVE BAIYES DENGAN ALGORITMA SUPPORT VECTOR
MACHINE DALAM ANALISIS SENTIMEN TWITTER APLIKASI
PEDULI LINDUNGI**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Sistem Informasi



disusun oleh

Maulana Rizky Hidayat

18.11.1816

**PROGRAM SARJANA
PROGRAM STUDI SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

PERBANDINGAN AKURASI DAN WAKTU PROSES ALGORITMA NAÏVE BAYES DENGAN ALGORITMA SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN TWITTER APLIKASI PEDULI LINDUNGI

yang dipersiapkan dan disusun oleh

Maulana Rizky Hidayat

18.11.1816

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 14 Mei 2016

Dosen Pembimbing,

Mulia Sulistivono, M.Kom
NIK. 190302248

PENGESAHAN
SKRIPSI
PERBANDINGAN AKURASI DAN WAKTU PROSES ALGORITMA
NAÏVE BAYES DENGAN ALGORITMA SUPPORT VECTOR MACHINE
DALAM ANALISIS SENTIMEN TWITTER APLIKASI PEDULI
LINDUNGI

yang dipersiapkan dan disusun oleh

Maulana Rizky Hidayat

18.11.1816

telah dipertahankan di depan Dewan Penguji
pada tanggal 10 Oktober 2016

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Nama dan Gelar Penguji 1

NIK. 190302xxx

Nama dan Gelar Penguji 2

NIK. 190302xxx

Nama dan Gelar Penguji 3

NIK. 190302xxx

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 10 November 2016

KETUA STMIK AMIKOM YOGYAKARTA

Prof. Dr. M. Suyanto, M.M.

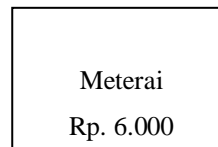
NIK. 190302001

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta,



Maulana Rizky Hidayat

NIM. 18.11.1816

MOTTO

KATA PENGANTAR

DAFTAR ISI

JUDUL.....	I
PERSETUJUAN	II
PENGESAHAN	III
PERNYATAAN.....	IV
MOTTO.....	VI
KATA PENGANTAR.....	VII
DAFTAR ISI.....	VIII
DAFTAR TABEL	X
ABSTRACT	XII
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH.....	1
1.3 BATASAN MASALAH.....	4
1.4 MAKSUD DAN TUJUAN PENELITIAN	4
1.5 MANFAAT PENELITIAN	4
1.6 METODE PENELITIAN	4
1.7 SISTEMATIKA PENULISAN	5
BAB II LANDASAN TEORI.....	7
2.1 KAJIAN PUSTAKA	7
2.2 DASAR TEORI.....	11
2.2.1 <i>Data Mining</i>	11
2.2.2 <i>Sentiment Analysis</i>	11
2.2.3 <i>Twitter</i>	11
2.3 METODE ANALISIS.....	12
2.3.1 <i>Text Mining</i>	12
2.3.2 <i>Preprocessing</i>	12
2.3.3 <i>Naive Bayes Classifier</i>	12
2.3.4 <i>Support Vector Machine</i>	13
2.3.5 <i>Evaluasi</i>	14
2.3.6 <i>Validasi</i>	16

BAB III METODE PENELITIAN	17
3.1 ALAT DAN BAHAN PENELITIAN	17
3.2 ALUR PENELITIAN	17
3.2.1 Pengumpulan Data	19
3.2.2 Text Preprocessing.....	19
3.2.3 Pembobotan Term.....	20
3.2.4 Klasifikasi.....	20
3.2.5 Validasi dengan K-FOLD CROSS Validation	23
3.2.6 Evaluasi.....	23
BAB IV HASIL DAN PEMBAHASAN	24
4.1 IMPLEMENTASI PENGAMBILAN DATA & TEXT PREPROCESSING	24
4.2 IMPLEMENTASI ALGORITMA KLASIFIKASI SERTA PENGUJIAN.	26
4.3 PERBANDINGAN ALGORITMA KLASIFIKASI.....	31
BAB V PENUTUP	33
5.1 KESIMPULAN	33
5.2 SARAN	33
DAFTAR PUSTAKA.....	35

DAFTAR TABEL

Tabel 2. 1 Matrik Literatur Review dan Posisi Penelitian.....	9
Tabel 4. 1 Confusion Matrix Klasifikasi Naive Bayes Dengan Split 8020.....	26
Tabel 4. 2 Perhitungan Klasifikasi Naive Bayes Dengan Split 8020.....	26
Tabel 4. 3 Hasil Pengujian Naïve Bayes	27
Tabel 4. 4 Hasil Pengujian K Fold Cross Validation	28
Tabel 4. 5 Waktu Proses Naïve Bayes	28
Tabel 4. 6 Hasil Klasifikasi SVM Dengan Split 8020	29
Tabel 4. 7 Perhitungan Klasifikasi SVM Dengan Split 8020.....	29
Tabel 4. 8 Hasil Pengujian Naïve Bayes	30
Tabel 4. 9 Hasil Pengujian K Fold Croos.....	31
Tabel 4. 10 Waktu Proses SVM.....	31
Tabel 4. 11 Hasil Perbandingan Algoritma SVM	31

DAFTAR GAMBAR

Gambar 2 . 1 Ilustrasi SVM [2].....	14
Gambar 2 . 2 Confusion Matrix (sumber : https://ksnugroho.medium.com/)	15
Gambar 3 . 1 WorkFlow Penelitian.....	19
Gambar 3 .2 Diagram alir tahapan Naïve Bayes.	21
Gambar 3 . 3 Diagram Alir Support Vector Machine	23
Gambar 4. 1 Dataset Twitter.....	24
Gambar 4. 2 Tahap Text Preprocessing.....	25

INTISARI

Salah satu cabang ilmu computer yaitu mesin learning (machine learning) menjadi tren dalam beberapa waktu terakhir. Pembelajaran mesin bekerja dengan memanfaatkan data dan algoritma untuk membuat model dengan pola kumpulan data tersebut. Selain itu, pembelajaran machine learning juga mempelajari bagaimana model yang telah dibuat memprediksi keluaran berdasarkan pola yang ada.

Terdapat dua jenis metode pembelajaran mesin learning yang dapat digunakan untuk analisis sentimen: yaitu supervised learning dan unsupervised learning.

Penelitian ini akan membandingkan dua algoritma klasifikasi yang termasuk dari supervised learning yang akan diimplementasikan kedalam analisis sentimen Aplikasi Peduli Lindungi algoritma tersebut ialah Algoritma Naïve Bayes Classification serta Algoritma Support Vector Machine, dengan cara membuat model dengan masing masing algoritma mana yang lebih baik dalam segi akurasi dan waktu proses. Hasil pada perhitungan akurasi menunjukkan bahwa metode Naïve Bayes lebih unggul dengan 80% tanpa menggunakan K-Fold Cross Validation dan 83% dengan menggunakan K-Fold Cross Validation pada fold - 4. Sedangkan perhitungan waktu proses Naïve Bayes juga unggul dengan mendapatkan waktu 0.009365 detik.

Kata Kunci: Machine Learning, Naïve Bayes, Support Vector Machine, Validasi

ABSTRACT

One branch of computer science, namely machine learning (machine learning) has become a trend in recent times. Machine learning works by utilizing data and yahoo to create a model with the pattern of the data set. In addition, machine learning also learns how the model that has been created predicts output based on existing patterns. There are two types of machine learning methods that can be used to analyze sentiment: supervised learning and unsupervised learning. This study will compare two algorithms that are included in supervision learning which will be implemented in sentiment analysis. The Peduli Protect application includes the Naïve Bayes Classification Algorithm and the Support Vector Machine Algorithm, by making a model with each algorithm which is better in terms of accuracy. and processing time. The calculation results show that the Naïve Bayes method is superior by 80% without using K-Fold Cross Validation and 83% using K-Fold Cross Validation at fold - 4. Meanwhile the Naïve Bayes time calculation is also superior by getting a time of 0.009365 seconds.

Keyword: *Machine Learning, Nave Bayes, Support Vector Machine, Validation*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seperti yang ditunjukkan oleh Organisasi Kesehatan Dunia (WHO) di situs otoritasnya, Indonesia adalah negara dengan 4.232.099 kasus positif dan 142.828 kasus meninggal angka dari informasi tersebut adalah angka yang besar. Angka ini secara umum akan terus bertambah jika tidak dilakukan pengawasan oleh pemerintah. Lockdown PSBB (Pembatasan Sosial Bersekala Besar), PPKM level 1-4, serta imunisasi di berbagai daerah telah dilakukan di Indonesia untuk mengurangi peningkatan COVID-19, dengan penanganan yang telah berjalan. Kasus COVID-19 di bulan Oktober, serta untuk mengendalikan COVID-19, Pemerintah Indonesia juga telah membuat aplikasi bernama Peduli Lindungi yang dibuat untuk memantau aktifitas serta diharapkan untuk mengurangi lonjakan penyebaran Covid – 19 di Indonesia [1]

Dengan adanya Aplikasi Peduli Lindungi ini menimbulkan berbagai tanggapan masyarakat terhadap aplikasi Peduli Lindungi sehingga mengundang banyak kalangan untuk memberikan pendapat opini di berbagai kalangan melalui media sosial, media sosial yang sering digunakan masyarakat untuk menyuarakan pendapat dan opini adalah Twitter.

Twitter telah menjadi media berbasis web dan smartphone yang paling populer dengan lebih dari 200 juta pengguna dan lebih dari 10,6 miliar tweet yang telah dihasilkan. Melalui posting Twitter, dapat berbagi dan mendapatkan data tentang masalah apapun dengan menggunakan informasi dari Twitter, dapat digunakan untuk menganalisis pendapat dan opini masyarakat terhadap Aplikasi

Peduli Lindungi melalui analisis sentimen dengan menggunakan klasifikasi pendapat ke dalam 2 kelas yaitu “Positif” dan “Negatif”. [2]

Analisis Sentimen merupakan suatu peristiwa dari penelitian yang bergantung pada sikap individu terhadap suatu hal. Analisis Sentimen biasanya dilakukan untuk mengumpulkan dan menemukan opini public dalam postingan Blog, Twitter, Facebook dan lainnya. Analisis sentimen diperlukan untuk mengetahui penilaian umum pada suatu objek. Dengan menggunakan informasi dari media online Twitter, dapat dilakukan Analisis Sentimen terhadap Aplikasi Peduli Lindungi. [2]

Support Vector Machine merupakan salah satu algoritma klasifikasi yang baik. Kemampuan Algoritma Support Vector Machine mendefinisikan hyperplane yang baik menjadikan algoritma SVM memiliki tingkat akurasi yang baik dibandingkan algoritma lainnya, Sedangkan algoritma Naïve Bayes merupakan algoritma klasifikasi machine learning dengan penalaran probabilitas yang tak kalah baik dengan algoritma lainnya. Algoritma Naïve Bayes juga memiliki kelebihan yaitu dengan menggunakan data latih yang tidak terlalu banyak dan memiliki tingkat akurasi yang baik. [3]

Terdapat beberapa penelitian analisis sentimen menggunakan algoritma klasifikasi yang pernah dilakukan oleh Henni Tuhuteru [4] dengan penelitian yang berjudul Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier pada penelitian ini menggunakan data set sebanyak 2378 dengan menggunakan algoritma Naïve Bayes dan SVM dengan pengujian validasi menggunakan K Fold Cross Validation dengan hasil SVM lebih baik dari pada Naïve Bayes

dengan tingkat akurasi sebesar 81,61% sedangkan Naïve Bayes memiliki tingkat akurasi 67.29%

Selanjutnya penelitian yang pernah dilakukan oleh Friska Fitriana [3] mengenai Analisis Sentimen Opini Terhadap Vaksin Covid-19 pada media sosial media twitter menggunakan Support Vector Machine serta Naïve Bayes dengan menggunakan data set yang diambil dari twitter sebanyak 1000 dengan keyword “Vaksin Covid-19” berdasarkan penelitian tersebut metode SVM lebih baik sedikit dibandingkan metode Naive Bayes dengan tingkat akurasi 90,47% untuk algoritma SVM sedangkan Naïve Bayes 88,64%

Kemudian penelitian dilakukan oleh Arifin Oki mengenai Analisa Perbandingan Tingkat Performansi Metode Support Vector Machine dan Naïve Bayes untuk Klasifikasi Jalur Minat Bakat SMA. Data set diambil dari dua buah sma dengan data sejumlah 288 siswa, dalam penelitian ini digunakan pengujian menggunakan K Fold Cross Validation, serta menggunakan algoritma Naïve Bayes dan serta SVM dari penelitian ini didapatkan hasil Algoritma SVM lebih baik dari pada Algoritma Naïve Bayes. [5]

Berdasarkan perbandingan ketiga jurnal diatas algoritma SVM dan Naïve Bayes memiliki beberapa kelebihan serta kekurangan masing masing. Pada penelitian ini akan membahas proses analisis sentimen dengan melakukan perbandingan dua metode yaitu Naïve Bayes Classifier dan Support Vector Machine untuk mengetahui perbandingan tingkat akurasi serta waktu pemrosesan yang diimplementasikan pada analisis sentimen twitter terhadap aplikasi Peduli Lindungi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, maka perlu dirumuskan suatu masalah penelitian yaitu perbandingan Akurasi serta Waktu Pemrosesan algoritma Naïve Bayes dengan algoritma Support Vector Machine dalam analisis sentimen Aplikasi Peduli Lindungi.

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian adalah:

- Penelitian ini mengambil objek berupa opini masyarakat terhadap aplikasi Peduli Lindungi.
- Algoritma yang digunakan untuk mengolah data ialah metode Naïve Bayes serta Support Vector Machine.
- Dataset yang digunakan diambil dari twitter dengan menggunakan tools twint dengan keyword “PLindungi”, “Peduli Lindungi”
- Tweet yang digunakan berupa text dan tidak mengandung gambar.

1.4 Maksud dan Tujuan Penelitian

Berdasarkan rumusan masalah yang terdapat diatas maksud dan tujuan yang dikedepankan dalam penelitian ini adalah membandingkan akurasi serta waktu proses algoritma Naïve Bayes dengan algoritma Support Vector Machine dalam opini pengguna Twitter terhadap Aplikasi Peduli Lindungi.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari adanya penelitian ini yaitu:

1. Mengetahui seberapa besar tingkat akurasi dan performansi algoritma Support Vector Machine dengan Naïve Bayes
2. Mengetahui algoritma yang sesuai untuk klasifikasi analisis sentimen Aplikasi Peduli Lindungi
3. Sebagai referensi untuk penelitian selanjutnya dan mengembangkan penelitian ini.

1.6 Metode Penelitian

1. Metode Pengumpulan Data

Dalam metode ini penulis mencoba mendapatkan data dari twitter dengan cara crawling data menggunakan tools Twint dari akun akun acak berdasarkan keyword “Peduli Lindungi”, “Plindungi,” “Aplikasi Peduli Lindungi”

2. Metode Pengolahan data

Pada pengolahan data ini terdiri dari beberapa tahapan yang memiliki masing masing metode untuk Analysis Sentiment. Hal yang pertama kali dilakukan adalah mengcrawling data twitter menggunakan tools Twint selanjutnya data tersebut disimpan dalam bentuk file document, lalu data yang telah diproses dibagi menjadi dua tipe data yaitu data training dan data testing, berikut tahapan selanjutnya dalam pengolahan data training dan data testing.

1. Pelabelan Manual
2. Cleaning
3. Tweet tokenize
4. Stopwords
5. Normalization
6. Text Stemming

3. Metode Analisis

Penelitian ini menganalisis setiap data yang di peroleh melalui cara kerja algoritma Naïve Bayes dan Support Vector Machine dalam mengklasifikasikan setiap data tweet yang diperoleh dari crawling data itu sendiri.

1.7 Sistematika Penulisan

Adapun sistematika penulisan Skripsi ini adalah sebagai berikut:

1. BAB 1 Pendahuluan

Pada Bab 1 diuraikan isi dan rencana pengerjaan Skripsi secara keseluruhan yang meliputi latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat, metode penelitian yang digunakan.

2. BAB 2 Landasan Teori

Bab 2 memaparkan dasar dasar teori yang berkaitan dengan Analisis Sentiment, Metode Naïve Bayes, Metode Support Vector Machine.

3. BAB 3 Perancangan dan Implementasi

Perancangan sistem dan contoh implementasi dari sistem yang dibangun dan dipaparkan pada bab ini. Selanjutnya dilakukan proses implementasi.

4. BAB 4 Pengujian dan Analisis

Pada bab ini dibahas skenario dan hasil dari pengujian yang dilakukan pada hasil implementasi system.

5. BAB 5 Kesimpulan dan Saran

Bab ini berisi kesimpulan dan saran yang didapatkan dari hasil implementasi sistem secara keseluruhan.

BAB II

LANDASAN TEORI

2.1 Kajian Pustaka

Pada penelitian yang dilakukan oleh Siti Mujilahwati [6] dengan Judul Preprocessing Text Mining Pada Informasi Twitter disimpulkan jika ujicoba klaifikasi dengan menggunakan metode Naïve Bayes dapat menghasilkan akurasi sebesar 93,11% dari hasil preprocessing yang sudah dijalani. Klasifikasi di atas dianggap masih kurang baik sebab masih terdapat mention terhadap customer support.

Penelitian tentang analisis sentimen juga pernah dilakukan oleh G.Buntoro [7] yang berjudul. Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine, penelitian ini bertujuan untuk mengukur analisis sentimen hatespeech dari data yang didapatkan pada twitter yang berjumlah 522 kemudian data tersebut diolah dengan menggunakan Algoritma Naïve Bayes dan Support Vector Machine, pada penelitian yang telah dilakukan, hasil akurasi tertinggi didapatkan oleh Algoritma Support Vector Machine dengan akurasi 66.7% namun penelitian ini belum maksimal dan dapat dikembangkan menggunakan stopword list dan stemmer Bahasa Indonesia yang mampu meningkatkan akurasi dalam analisis sentiment.

Kemudian penelitian yang dilakukan oleh Frizka Fitriana dan Hanif Al Fatta [3] dengan judul Analisis Sentimen Opini Terhadap Vaksin Covid-19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naïve Bayes, dataset yang digunakan berjumlah 1000 dengan keyword Vaksin Covid-19 setelah data didapatkan kemudian data tersebut masuk ke tahap text preprocessing melakukan sorter ke dalam proses case folding tokenise, stopword

removal, stemming, translation, normalization kemudian data siap diimplementasikan ke dalam algoritma Naïve Bayes dan Support Vector Machine dengan menggunakan bahasa pemrograman PHP 7.4 dan Javascript serta database dengan MySQL kemudian data dilakukan tahap validasi menggunakan K-Fold Cross Validation. Dari penelitian tersebut didapatkan nilai akurasi Support Vector Machine lebih baik dari pada algoritma Naïve Bayes dengan tingkat akurasi 90,44% sedangkan Naïve Bayes 88,64%

Tabel 2. 1 Matrik Literatur Review dan Posisi Penelitian

No	Judul	Peneliti	Tahun	Pokok Penelitian	Hasil	Kelebihan	Kekurangan	Perbandingan dengan yang akan diteliti
1	Pre-Processing Text Mining Pada Data Twitter	Siti Mujilawati	2016	Teknik preprocessing penanganan preprocessing data komentar twitter	Pada penelitian ini memiliki hasil akurasi sebesar 93.11%	Tanpa proses steaming hasil yang didapatkan sudah sangat baik,	Pada penelitian ini tidak menggunakan proses steaming sehingga masih terdapat data mention terhadap customer support dan remove hastag	Tingkat akurasi metode Naïve Bayes
2	Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes	Fitriana, Frizka Utami, Ema Al Fatta, Hanif	2021	Membandingkan dua algoritma SVM dan NBC yang dimplementasikan pada analisis sentimen tentang vaksin covid 19	Algoritma SVM memiliki tingkat akurasi yang baik sedangkan Algoritma naïve bayes memiliki waktu pemrosesan yang lebih sedikit	Algoritma SVM memiliki performa lebih baik pada bagian akurasi sebesar 90,47% sedangkan algoritma Naïve Bayes unggul dalam segi	Performa akurasi dapat ditingkatkan melalui optimasi kedua algoritma SVM dan NBC seperti menggunakan particle swarm optimization	Membandingkan akurasi dan optimasi algoritma Naïve Bayes dan Support Vector Machine.

						waktu traning		
3	Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine	Ghulam Asrofi Buntoro	2016	Sentimen analisis twitter berbahasa Indonesia dengan taggar Hatespeech	Algoritma SVM dengan melakukan preprocessing (tokenisasi, cleansing dan filtering.)memiliki tingkat akurasi sebesar 66%	Algoritma SVM memiliki akurasi lebih baik dari pada metode lain	Dalam penelitian ini masih dapat dikembangkan dengan menambahkan stopword serta data yang digunakan masih sedikit	Membandingkan akurasi dan optimasi agoritma Naïve Bayes dan Support Vector Machine

2.2 Dasar Teori

2.2.1 Data Mining

Data Mining merupakan proses penting dimana metode digunakan untuk mengekstrak pola yang valid. Data mining tersebut termasuk klasifikasi, regresi, pengelompokan, atau asosiasi. Sebagian besar Data mining dikenal dan digunakan dalam banyak aplikasi bagian ini memberikan ulasan singkat tentang teknik terpilih yang dianggap paling penting pentingnya dan sering terjadi pada Data Mining.[8]

2.2.2 Sentiment Analysis

Sentiment Analysis didefinisikan sebagai tugas mencari pendapat penulisan tentang entitas tertentu. Proses pengambilan keputusan dari orang dipengaruhi oleh opini yang dibentuk, oleh masyarakat umum, pemikiran oleh pemimpin. Analisis sentimen dapat didapatkan dari media sosial, Twitter, Facebook, Blog dan forum pengguna lainnya. Sentimen Analisis biasanya digunakan untuk mengambil feedback dari sebuah produk atau kinerja sebuah organisasi [9]

2.2.3 Twitter

Twitter merupakan salah satu media sosial yang bertipe microblogging sebagai layanan interaksi. Twitter sebagai salah satu layanan media sosial yang sangat populer didunia dengan 200 juta pengguna aktif serta lebih dari 10.6 milyar tweet yang sudah dihasilkan. [10]

Twitter diciptakan sebagai platform bertukar pengalaman dan sharing apapun antar sesama penggunanya tanpa adanya sekat penghalang. Dengan menggunakan twitter pengguna akan lebih mudah mengikuti trend serta

informasi dan berita dari seluruh penjuru dunia. Selain itu twitter juga membantu penggunaanya untuk selalu terhubung dengan orang-orang terdekat.[11]

2.3 Metode Analisis

2.3.1 Text Mining

Text mining merupakan bagian dari data mining dimana proses yang dilakukan utamanya adalah melakukan ekstraksi pengetahuan dan informasi dari pola-pola yang terdapat dalam sekumpulan document, teks menggunakan alat analisis tertentu[9] Text mining bertujuan untuk mencari kata-kata yang dapat mewakili apa yang ada didalam document sehingga dapat dilakukan analisa keterhubungan antar document. Text mining memiliki tahapan.

2.3.2 Preprocessing

Penerapan suatu algoritma data mining pada suatu data set dalam sentiment analysis memerlukan persiapan data yang akan diolah terlebih dahulu proses persiapan tersebut dinamakan preprocessing. Proses ini berfungsi untuk sebagai tahapan pengkondisian suatu data agar dapat melakukan pengolahan secara cepat dan efektif. Preprocessing juga bertujuan untuk membersihkan data yang muncul berulang dan membakukan kalimat yang tidak baku[11].

2.3.3 Naïve Bayes Classifier

Naïve Bayes Classifier (NBC) merupakan salah satu algoritma klasifikasi yang berakar pada teorema bayes. Naïve Bayes Classifier bekerja sangat baik dibandingkan dengan model classifier lainnya seperti Decision Tree ataupun Neural Network. Keuntungan pengguna metode ini adalah metode ini hanya membutuhkan data training yang kecil untuk menentukan parameter yang diperlukan didalam proses pengklasifikasian. Adapun Naïve Bayes Classifier memiliki bentuk umum seperti berikut: [1]

$$P(c_i|x) = \frac{P(x|c_i) \cdot P(c_i)}{P(x)} \dots\dots\dots(1)$$

Keterangan :

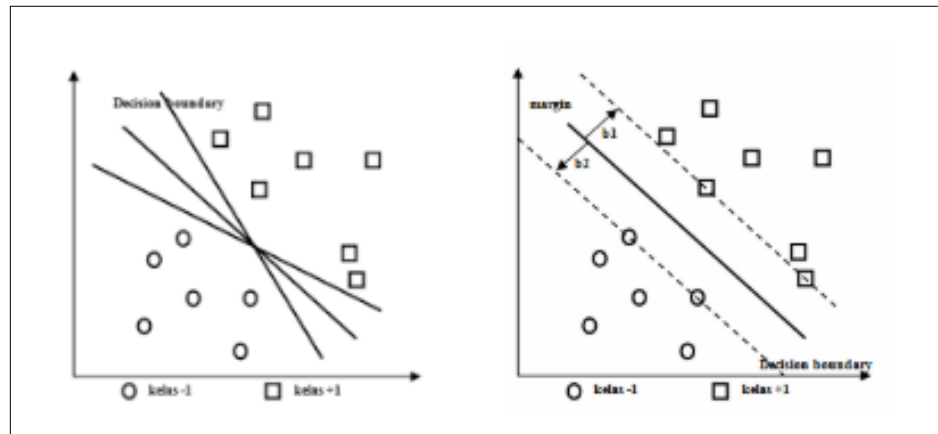
- X : Hipotesis data Ci adalah kelas khusus
 Ci : Data dengan kelas tidak dikenal
 P(Ci|X) : Probabilitas hipotesis X didasarkan pada kondisi Ci
 P(X|Ci) : Probabilitas hipotesis Ci didasarkan pada kondisi X
 P(Ci) : Probabilitas hipotesis X
 P(X) : Probabilitas hipotesis Ci

2.3.4 Support Vector Machine

Support Vector Machine (SVM) adalah metode pembelajaran terbimbing yang menganalisis data dan mengenali pola, digunakan untuk klasifikasi dan analisis regresi. Metode ini dikembangkan oleh Boser Guyon Vapnik dan pertama kali dipresentasikan pada tahun 1992 di *Annual Workshop on Computational Learning Theory* [11]

SVM juga termasuk klasifikasi yang memberi aturan untuk mengubah document text menjadi vector multidimensional tf-idf, vector dalam penelitian ini memiliki dua komponen yaitu dimensi (word id) dan bobot.

Prinsip kerja dari metode ini adalah mencari ruang pemisah yang paling optimal dari suatu dataset dalam kelas yang berbeda. Klasifikasi ini dilakukan dengan mencari hyperplane atau garis pembatas (*decision boundary*) yang memisahkan antara satu kelas dengan kelas lainnya [11]



Gambar 2 . 1 Ilustrasi SVM [2]

Gambar di atas menjelaskan konsep klasifikasi SVM. Pada gambar (a) dijelaskan terdapat sejumlah data dengan lingkaran sebagai kelas -1 dan kotak sebagai kelas +1. Pada gambar tersebut juga terdapat sejumlah hyperplane yang mungkin untuk set data. Gambar (b) adalah hyperplane yang paling maksimal. Perhitungan hyperplane dilakukan dengan cara menghitung jarak margin dengan data terdekat dari masing masing kelas. Data terdekat ini disebut sebagai Support Vector Machine. Inti dari metode ini adalah pencarian hyperplane terbaik dari setiap kemungkinan

Persamaan hyperplane yang terletak pada support vector adalah

$$w, x + b = -1 \dots\dots\dots(2)$$

$$w, x + b = +1$$

2.3.5 Evaluasi

Evaluasi berfungsi untuk mengetahui akurasi dari model algoritma yang dibuat. Kriteria evaluasi yang dipertimbangkan adalah akurasi, standart deviation, f1 score, recall, presicion dan specificity

Confusion Matrix biasanya digunakan dalam machine learning untuk mengevaluasi atau memvisualisasikan perilaku model dalam konteks klasifikasi terawasi.[12]

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive) <i>Type I Error</i>
	0 (Negative)	FN (False Negative) <i>Type II Error</i>	TN (True Negative)

Gambar 2 . 2 Confusion Matrix (sumber: <https://ksnugroho.medium.com/>)

Nilai true positive (TP) dan true negative (TN) adalah hasil klasifikasi yang benar. Nilai false positif (FP) adalah nilai dimana hasilnya diprediksi sebagai kelas-2 namun sebenarnya merupakan kelas-1 sedangkan false negative (FN) adalah nilai dimana produksi pengklasifikasian sebagai kelas 1 namun faktanya termasuk dalam klasifikasi kelas 2.

Akurasi adalah proporsi dari total prediksi true dari semua data. Rumus akurasi adalah:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(3)$$

Precision adalah ukuran ketepatan dari hasil suatu model. Persamaannya adalah perbandingan antara true positive dengan total data dengan label positive:

$$Presicion = \frac{TP}{TP+FP} \dots\dots\dots(4)$$

Recall adalah ukuran kelengkapan dari sebuah model. Persamaan recall perbandingan antara true positive terhadap total contoh yang benar benar positive:

$$Recall = \frac{TP}{TP+FN} \dots \dots \dots (5)$$

F-measure merupakan harmonic mean dari presicion dan recall dimana persamaannya adalah[13]

$$F1\ Score = \frac{2*Presicion*Recall}{Presicion+Recall} \dots \dots \dots (6)$$

2.3.6 Validasi

K-Fold Cross Validation merupakan salah satu metode yang dapat memeriksa overfitting pada suatu model. Data yang dibagi menjadi K bagian membolehkan setiap bagian data berhenti memprediksi data lebih cepat ketimbang tidak dibagi terlebih dahulu [14]

Pada K-fold Cross Validation model yang telah dibuat dibagi menjadi k bagian yang sama atau mendekati ukuran. Akurasi model akan diuji menggunakan data uji pada setiap fold, dan berlanjut ke fold selanjutnya sampai selesai. Akurasi akan ditotal dan dibagi dengan banyaknya [15]

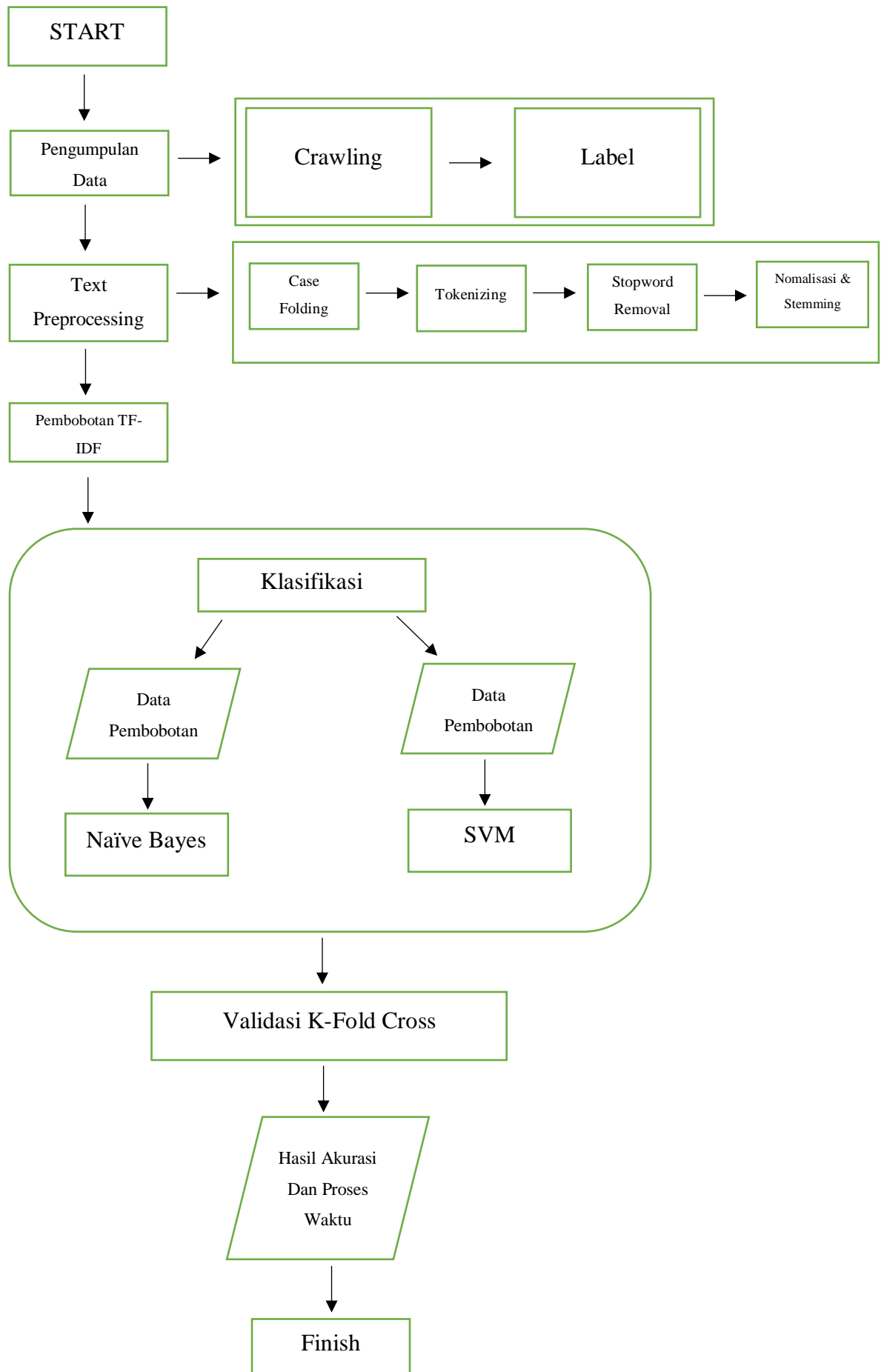
BAB III METODE PENELITIAN

3.1 Alat dan Bahan Penelitian

1. Google Collaboratory
2. Python
3. Numpy
4. Pandas
5. NLTK
6. Twint
7. Flask & Heroku
8. Twitter (“PLindungi”, “Pedulilindungi”)

3.2 Alur Penelitian

Dalam penelitian ini menggunakan acuan, pedoman dan tahapan yang akan diterapkan pada sebuah penelitian ini untuk mencapai tujuan penelitian tersebut yang didapatkan sesuai dengan yang diharapkan oleh karena itu diperlukan tahapan yang tersusun. Gambar 3.1 berikut ini merupakan tahapan yang dilakukan pada penyusunan skripsi / Tugas Akhir ini:



Gambar 3 . 1 WorkFlow Penelitian.

3.2.1 Pengumpulan Data

Pengumpulan Data merupakan tahap awal dalam metodologi penelitian. Pada tahap pengumpulan data tersebut menggunakan proses crawling, crawling yaitu proses dimana kita mengambil data dari media sosial twitter untuk nantinya digunakan dalam penelitian. Pada pengambilan data ini digunakan tools twint untuk mengambil data dari twitter tersebut kemudian masuk pada tahap labeling data,

Labeling merupakan proses pelabelan data yang sudah diambil dari teknik crawling, dengan label positif dan negatif yang dilabeli dengan cara pelabelan manual, Pada proses pelabelan data twitter biasanya terdapat bahasa yang tidak baku maka pelabelan dilakukan berdasarkan kalimat yang mengandung makna positif ataupun negatif sebagai contoh “Asyik dan bermanfaat banget sih aplikasi Pedulilindungi ini” & “Aplikasi buuuuuurrrriiikkk, bikin jengkel saja” .

3.2.2 Text Preprocessing

Pada tahap preprocessing ini terdapat empat proses yang dilakukan dimana pada tahap ini akan memproses atau mengolah data yang telah dilabeli. tahap ini dilakukan untuk membersihkan data dan memaksimalkan hasil dan konsisten. Proses pertama yaitu Case Folding. Pada proses ini akan dilakukan pengubahan semua huruf yang terdapat pada data tersebut menjadi huruf kecil atau lowercase dan menghilangkan karakter selain huruf. Proses ini dilakukan untuk mendapatkan data yang redundant. Selanjutnya proses tokenizing dimana pada proses ini akan dilakukan pemotongan atau pemisahan setiap kata yang terdapat daalm document yang kemudian diesbut dengan token. Setelah proses

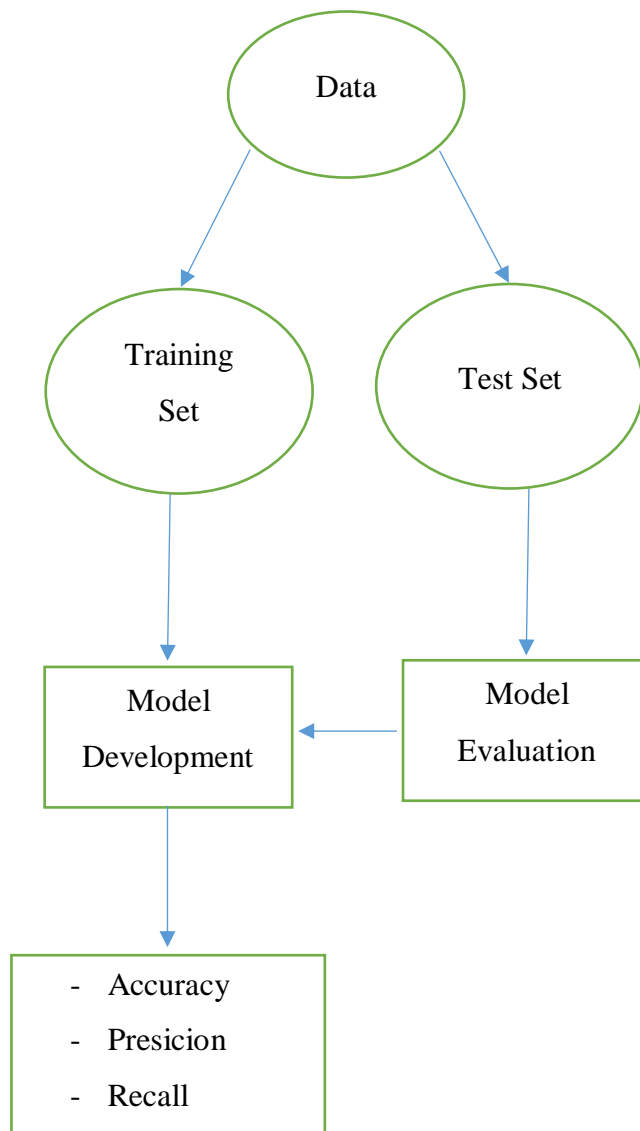
tokenizing proses selanjutnya yang dilakukan yaitu stopwords removal. Pada proses ini semua kosakata yang tidak memiliki makna seperti kata ‘dan’, ‘di’, ‘oleh’ akan dihilangkan dari dalam document sehingga menyisakan kata yang bermakna saja didalam document selanjutnya proses terakhir dalam tahapan text preprocessing yaitu proses stemming semua kata yang terdapat dalam document akan diubah menjadi bentuk kata dasar dengan menghapus atau menghilangkan imbuhan yang terdapat pada kata tersebut. Stemming dilakukan untuk mengatasi adanya kata yang tidak biasa memperkecil jumlah indeks yang berbeda dari suatu document.

3.2.3 Pembobotan Term

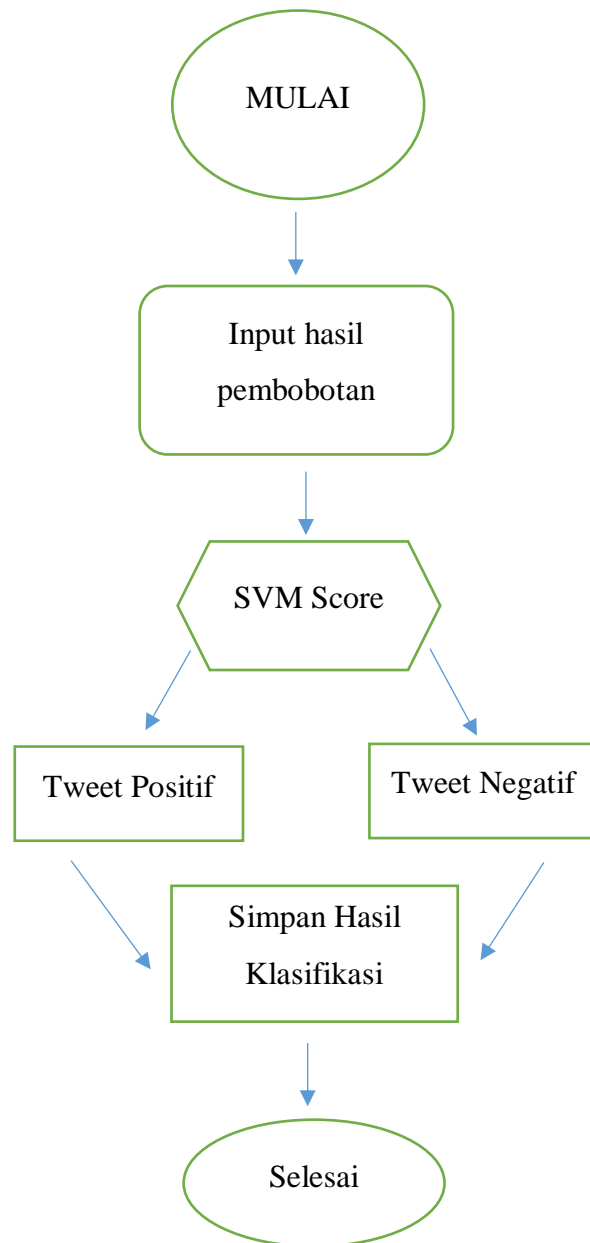
Tahap selanjutnya yang dilakukan yaitu pembobotan term. Hal ini dilakukan dengan tujuan untuk menghitung seberapa banyak kemunculan kata pada dokumen. Dalam penelitian yang akan dilakukan ini metode yang digunakan yaitu TF-IDF sebagai proses pembobotan.

3.2.4 Klasifikasi

Setelah melalui tahap pengumpulan data, text preprocessing dan pembobotan term tahap selanjutnya yaitu tahap klasifikasi. Tahap preprocessing serta sudah dilakukan pembobotan dengan metode TF-IDF, akan diproses dengan algoritma yang telah dipilih yaitu algoritma Naïve Bayes dan Support Vector Machine. Dalam tahapan ini mesin akan diajari untuk mengenal pola data atau dokumen yang ada untuk kemudian dapat mengklasifikasi sebuah data ke dalam dua kelas yaitu , positif dan negatif.



Gambar 3 .2 Diagram alir tahapan Naïve Bayes.



Gambar 3 . 3 Diagram Alir Support Vector Machine

3.2.5 Validasi dengan K-FOLD CROSS Validation

Tahap validasi ini menggunakan K-Fold Cross Validation dokumen akan dibagi menjadi 10 bagian. Akan dilakukan percobaan sebanyak 10 kali percobaan, klasifikasi dokumen dan setiap percobaan data akan diacak urutannya terlebih dahulu sebelum akhirnya dimasukkan dalam fold. Tujuan hal ini dilakukan guna menghindari pengelompokan dokumen dari satu kategori tertentu pada suatu fold pada pengujian data ini menggunakan data training sebesar 80% serta data testing sebanyak 20%. Untuk mendapatkan tingkat akurasi yang maksimal.

3.2.6 Evaluasi

Pada tahap evaluasi ini akan dilakukan perbandingan akurasi dan hasil pemrosesan waktu yang diperoleh dari algoritma Naive bayes serta SVM. Serta mengambil hasil dan kesimpulan pada penelitian ini.

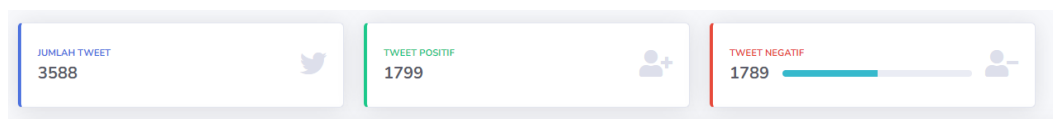
BAB IV

Hasil dan Pembahasan

Berdasarkan langkah langkah yang telah dipaparkan dalam BAB III, penelitian ini bertujuan untuk membahas tentang perbandingan tingkat akurasi serta perbandingan waktu proses menggunakan algoritma Naïve Bayes serta Algoritma SVM serta dilakukan pengujian dengan menggunakan KFold Cross Validation. Berikut ini pembahasan hasil dari percobaan yang dilakukan

4.1 Implementasi Pengambilan Data & Text Preprocessing

Tahap Implementasi Pengambilan data menggunakan program sederhana python dengan menggunakan Twint, dilakukan dengan menggunakan browser menggunakan google collabs untuk mengambil data twitter dengan kata kunci “Plindungi” serta search “Peduli lindungi” yang digunakan untuk memfilter data yang diambil dari twitter serta melimitasi data yang diambil sebanyak 3.588 data set. Kemudian setelah data set diambil selanjutnya dilakukan proses pelabelan data secara manual sebanyak data yang didapatkan.



Gambar 4.1 Dataset Twitter

Setelah dilakukan tahap pelabelan, dataset yang telah dilabeli tersebut kemudian akan dilanjutkan ke proses text preprocessing.

No	Value	Tweet	Cleaning	Tweet Tokenize	Stopwords	Normalization	Text Stemming
1	Negatif	yuk kenali setiap warna yang ada dari pedulilindungi #pentingnyavaksinasi #ikhtiaratasipandemi https://t.co/jenivcyafk	yuk kenali setiap warna yang ada dari pedulilindungi	['yuk', 'kenali', 'setiap', 'warna', 'yang', 'ada', 'dari', 'pedulilindungi']	['kenali', 'warna', 'pedulilindungi']	kenali warna pedulilindungi	kenal warna pedulilindungi
2	Negatif	yth @plindungi @kemenkesri status saya di aplikasi peduli lindungi tiba2 berubah menjadi hitam padahal saya tidak pernah melakukan test pcr atau antigen dengan hasil positif dan saya juga tidak ada kontak erat dengan pasein covid	yth status saya di aplikasi peduli lindungi tiba berubah menjadi hitam padahal saya tidak pernah melakukan test pcr atau antigen dengan hasil positif dan saya juga tidak ada kontak erat dengan pasein covid	['yth', 'status', 'saya', 'di', 'aplikasi', 'peduli', 'lindungi', 'tiba', 'berubah', 'menjadi', 'hitam', 'padahal', 'saya', 'tidak', 'pernah', 'melakukan', 'test', 'pcr', 'atau', 'antigen', 'dengan', 'hasil', 'positif', 'dan', 'saya', 'juga', 'tidak', 'ada', 'kontak', 'erat', 'dengan', 'pasein', 'covid']	['status', 'aplikasi', 'peduli', 'lindungi', 'berubah', 'hitam', 'test', 'pcr', 'antigen', 'hasil', 'positif', 'kontak', 'erat', 'pasein', 'covid']	status aplikasi peduli lindungi berubah hitam test pcr antigen hasil positif kontak erat pasein covid	status aplikasi peduli lindungi ubah hitam test pcr antigen hasil positif kontak erat pasein covid

Gambar 4. 2 Tahap Text Preprocessing

Pada tahap preprocessing ini dataset pertama – tama akan melewati proses case folding terlebih dahulu untuk konversi text menjadi suatu bentuk standar, pada tahap ini biasanya dipilih lowercase untuk membuat huruf capital menjadi huruf kecil. Kemudian dilanjut dengan melakukan text cleansing dengan melakukan penghapusan symbol dan karakter special yang diperoleh pada data tersebut.

Selanjutnya setelah proses cleansing dataset terebut melalui proses tokenizing ditahap tokenizing ini data dilakukan pemisahan yang disebut sebagai token yang kemudian digunakan untuk analisa lebih lanjut. Setelah proses tokenizing data kemudian dilanjutkan pada proses stopwords removal pada tahap ini dilakukan proses penghapusan kalimat yang tidak memiliki makna atau tidak berpengaruh dalam penelitian seperti “yg”, “dg”, “dgn”, “ny”, “ya” dan masih banyak lagi. Setelah memasuki stopwords kemudian data tersebut dilakukan proses normalisasi. Dalam proses normalisasi ini berfungsi untuk mengubah kata kata yang tidak baku menjadi kata kata baku. Kemudian setelah proses normalisasi dilanjutkan dengan proses stemming yaitu menjadikan kata kata yang sudah dinormalisasi tadi dijadikan kata dasar dengan menggunakan python library.

4.2 Implementasi Algoritma Klasifikasi Serta Pengujian

Untuk melihat perbandingan akurasi serta waktu proses pada Algoritma Naïve bayes serta Algoritma Support Vector Machine dengan menggunakan data set yang telah melalui proses text preprocessing, maka diimplementasikan algoritma Klasifikasi yang telah sekilas dibahas pada BAB III dalam penelitian perbandingan akurasi algoritma Naïve Bayes serta algoritma Support Vector Machine yang diimplementasikan pada analisis sentimen peduli lingkungan.

4.2.1 Implementasi Algoritma Naïve Bayes

Implementasi algoritma klasifikasi Naïve Bayes pada dataset Peduli Lingkungan yang telah melalui tahap text preprocessing.

Tabel 4. 1 Confusion Matrix Klasifikasi Naive Bayes Dengan Split 8020

Actual Classification	Prediction Classification	
	Positive (Kelas Negative)	Negative (Kelas Positive)
Positive (Kelas Negative)	284	66
Negative (Kelas Positive)	76	292

Tabel 4. 2 Perhitungan Klasifikasi Naive Bayes Dengan Split 8020

KLASIFIKASI NAÏVE BAYES DENGAN MENGGUNAKAN DATASET PEDULI LINDUNGI	
Akurasi	$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$ $= \frac{284 + 292}{284+292+66+76} \times 100\%$ $= \frac{574}{718} \times 100\%$

	$= 80\%$
Presisi	$\text{Presisi} = \frac{TP}{TP+FP} \times 100\%$ $= \frac{284}{284+66} \times 100\%$ $= \frac{284}{350} \times 100\%$ $= 81\%$
Recall	$\text{Recall} = \frac{TP}{TP+FN} \times 100\%$ $= \frac{284}{284+76} \times 100\%$ $= \frac{284}{360} \times 100\%$ $= 79\%$
Skor F1	$\text{Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$ $= \frac{284}{284 + \frac{1}{2}(66 + 76)}$ $= \frac{284}{355}$ $= 80\%$

Tabel 4. 3 Hasil Pengujian Naïve Bayes

KLASIFIKASI NAÏVE BAYES DENGAN MENGGUNAKAN DATASET PEDULI LINDUNGI	
Akurasi	80%
Presisi	81%
Recall	79%
Skor F1	80%

Pada penelitian di atas dapat dilihat pada Tabel 4.2 bahwa Akurasi yang didapatkan dalam algoritma Naïve Bayes sebanyak 80%, serta Presisi sebanyak 81% , Recall 79%, serta Score F1 sebanyak 80%

4.3.1 Pengujian menggunakan K Fold Cross Validation

Pada tahap pengujian dilakukan dengan menggunakan Cross Validation dengan CV 10 pada kombinasi ini dataset 80% data latih 20% data uji .

Tabel 4. 4 Hasil Pengujian K Fold Cross Validation

Nilai K-Fold	Accuracy
1	0.75
2	0.80
3	0.85
4	0.83
5	0.76
6	0.69
7	0.70
8	0.82
9	0.82
10	0.75

Dari Tabel 4.3 diperoleh akurasi tertinggi berada pada iterasi ke-3 dengan nilai akurasi sebesar 85% dengan akurasi rata rata keseluruhan 78%

4.3.2 Waktu Proses Algoritma Naïve Bayes

Selanjutnya waktu proses testing serta waktu proses prediksi yang digunakan dlm penelitian ini.

Tabel 4. 5 Waktu Proses Naïve Bayes

Training Time	Prediction Time
0.052144 s	0.009365 s

Pada table diatas waktu yang digunakan untuk melatih data pada algoritma Naïve Bayes dalam penelitian ini mendapatkan waktu 0.05 detik, serta waktu Prediksi mendapatkan waktu 0.0093 detik.

4.2.2 Implementasi Algoritma Support Vector Machine

Implementasi algoritman klasifikasi Support Vector Machine pada dataset Peduli Lindungi yang telah melalui tahap text preprocessing.

Tabel 4. 6 Hasil Klasifikasi SVM Dengan Split 8020

Actual Classification	Prediction Classification	
	Positive (Kelas Negative)	Negative (Kelas Positive)
Positive (Kelas Negative)	278	72
Negative (Kelas Positive)	73	295

Tabel 4. 7 Perhitungan Klasifikasi SVM Dengan Split 8020

KLASIFIKASI SUPPORT VECTOR MACHINE DENGAN MENGGUNAKAN DATASET PEDULI LINDUNGI	
Akurasi	$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$ $= \frac{278+295}{278+295+72+73} \times 100\%$ $= \frac{573}{718} \times 100\%$ $= \mathbf{79\%}$
Presisi	$\text{Presisi} = \frac{TP}{TP+FP} \times 100\%$ $= \frac{278}{278 + 72} \times 100\%$ $= \frac{278}{350} \times 100\%$ $= \mathbf{80\%}$

Recall	$\text{Recall} = \frac{TP}{TP+FN} \times 100\%$ $= \frac{278}{278+73} \times 100\%$ $= \frac{278}{351} \times 100\%$ $= \mathbf{80\%}$
Skor F1	$\text{Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$ $= \frac{278}{278 + \frac{1}{2}(72 + 73)}$ $= \frac{278}{351}$ $= \mathbf{80\%}$

Tabel 4. 8 Hasil Pengujian Naïve Bayes

KLASIFIKASI SUPPORT VECTOR MACHINE DENGAN MENGGUNAKAN DATASET PEDULI LINDUNGI	
Akurasi	79%
Presisi	80%
Recall	80%
Skor F1	80%

Pada penelitian di atas dapat dilihat pada Tabel 4.6 bahwa Akurasi yang didapatkan dalam algoritma Support Vector Machine sebanyak 79%, serta Presisi sebanyak 80% , Recall 80%, serta Score F1 sebanyak 80%

4.3.3 Pengujian Menggunakan K Fold Cross Validation

Pada tahap pengujian dilakukan dengan menggunakan Cross Validation dengan CV 10 pada kombinasi ini dataset 80% data latih 20% data uji .

Tabel 4. 9 Hasil Pengujian K Fold Croos

Nilai K-Fold	Accuracy
1	0.73
2	0.77
3	0.86
4	0.83
5	0.75
6	0.69
7	0.66
8	0.82
9	0.85
10	0.72

Dari Tabel 4.7 diperoleh akurasi tertinggi berada di nilai K-Fold 3 dengan nilai akurasi sebesar 86% dengan rata rata keseluruhan akurasi sebesar 77%

4.3.4 Waktu Proses Algoritma SVM

Tabel 4. 10 Waktu Proses SVM

Training Time	Prediction Time
20.120 s	84.06 s

Pada table diatas waktu yang digunakan untuk melatih data pada algoritma Support Vector Machine dalam penelitian ini mendapatkan waktu 23 detik, serta waktu Prediksi mendapatkan waktu 8.4 detik.

4.3 Perbandingan Algoritma klasifikasi

Tabel 4. 11 Hasil Perbandingan Algoritma SVM

NO	Naïve Bayes			SVM		
	10-Fold	Split	Waktu Proses	10-Fold	Split	Waktu Proses
1	85 %	80%	0.009365s	86%	79%	8.4 s

Berdasarkan hasil perbandingan algoritma Naïve Bayes serta SVM pada table di 4.9 diatas. Pada perhitungan waktu algoritma SVM memiliki waktu proses yang lebih lambat dibandingkan algoritma Naïve Bayes dengan waktu proses 84.06s dengan cara kerja algoritma SVM yang lebih kompleks dengan menggunakan kernel yang bertujuan untuk mencari hyperplane sehingga menyebabkan waktu proses cenderung lebih lama. Berbeda dengan algoritma Naïve Bayes dengan cara kerja menghitung peluang dari satu kelas dari masing masing kelompok atribut yang ada dan menentukan probabilitas yang paling optimal, serta menggunakan data latih yang cenderung lebih sedikit, sehingga menghasilkan waktu proses yang singkat yaitu 0.009365s.

Selanjutnya tingkat akurasi dengan pengujian K-Fold Cross Validation, Algoritma Naïve Bayes mendapatkan akurasi 85% pada iterasi ke 3 sedangkan Algoritma Support Vector Machine mendapatkan akurasi lebih baik dengan tingkat akurasi 86%, pada iterasi ke 3. Kemudian dilakukan pengujian menggunakan teknik split 8020 algoritma Naïve Bayes memiliki tingkat akurasi 80% lebih baik sedangkan algoritma SVM memiliki akurasi 79%

BAB V

PENUTUP

5.1 Kesimpulan

Penelitian ini bertujuan untuk mengetahui waktu proses serta akurasi terbaik algoritma klasifikasi. Pengujian dalam penelitian ini membandingkan algoritma Naïve Bayes serta algoritma Support Vector machine dengan menggunakan dataset sebanyak 3.588 dataset kemudian dengan pengujian menggunakan teknik split 8020 serta K-Fold Cross, serta membandingkan waktu proses latih serta waktu proses prediksi. Berdasarkan hasil penelitian ini dapat ditarik kesimpulan:

1. Pada penelitian ini nilai akurasi serta waktu proses yang telah dilakukan dalam analisis sentimen Aplikasi peduli lindungi dengan menggunakan algoritma Naive Bayes serta algoritma Support Vector Machine. algoritma Support Vector Machine unggul dalam pengujian dengan K-Fold Cross dengan tingkat akurasi tertingginya 86% pada iterasi ke - 3, sedangkan pada pengujian split 8020 algoritma Naive Bayes lebih unggul sedikit dibandingkan algoritma SVM dengan tingkat akurasi 80%. serta waktu proses algoritma Naive Bayes lebih cepat dibandingkan algoritma SVM dengan waktu proses 0.009365s.
2. Dalam penelitian ini algoritma Naïve Bayes memiliki tingkat akurasi yang sama baiknya dengan algoritma Support Vector machine, yang membedakan hanya saja dalam waktu proses algoritma Naïve Bayes memiliki waktu proses yang lebih baik dari pada Support Vector Machine.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa saran untuk peneliti selanjutnya yakni:

1. Penelitian kedepannya dapat membandingkan beberapa algoritma klasifikasi lain untuk dilihat performanya serta waktu prosesnya
2. Penelitian kedepannya dapat menggunakan lebih banyak kelas serta dataset dengan ukuran yang lebih besar.
3. Penelitian kedepannya dapat memaksimalkan perfromansi algoritma klasifikasi dengan Xgboost dll

DAFTAR PUSTAKA

- [1] F. A. Nasution, F. P. Kurniyanto, I. Nindra, K. Forestyanto, K. A. Wijaya, and P. F. Supriyadi, "Tugas Besar Wawasan Global TIK Analisis Sentimen Kebijakan Pemerintah Mengenai Vaksin COVID-19 Di Indonesia Menggunakan Metode Naive Bayes Classifier Pada Sosial Media Twitter FAKULTAS INFORMATIKA," no. 1301194024, pp. 1–21, 2020.
- [2] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.
- [3] F. Fitriana, E. Utami, and H. Al Fatta, "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes," *J. Komtika (Komputasi dan Inform.*, vol. 5, no. 1, pp. 19–25, 2021, doi: 10.31603/komtika.v5i1.5185.
- [4] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *J. Inform. J. Pengemb. IT*, vol. 3, no. 3, pp. 394–401, 2018, doi: 10.30591/jpit.v3i3.977.
- [5] O. Arifin and T. B. Sasongko, "ANALISA PERBANDINGAN TINGKAT PERFORMANSI METODE SUPPORT VECTOR MACHINE DAN NAÏVE BAYES CLASSIFIER," no. 2016, pp. 67–72, 2018.
- [6] S. Mujilahwati, "Pre-Processing Text Mining Pada Data Twitter," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. Sentika, pp. 2089–9815, 2016.
- [7] G. A. Buntoro, "Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine," *J. Din. Inform.*, vol. 5, no. 2, pp. 1–13, 2016.
- [8] S. Garcia, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining. Intelligent Systems Reference Library. 2015*, vol. 10. 2015.
- [9] R. Feldman, "Techniques and applications for sentiment analysis," *Commun. ACM*, vol. 56, no. 4, pp. 82–89, 2013, doi: 10.1145/2436256.2436274.

- [10] M. Ibrahim, O. Abdillah, A. F. Wicaksono, and M. Adriani, "Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in a Twitter Nation," *Proc. - 15th IEEE Int. Conf. Data Min. Work. ICDMW 2015*, pp. 1348–1353, 2016, doi: 10.1109/ICDMW.2015.113.
- [11] M. S. Hadna, P. I. Santosa, and W. W. Winarno, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. Sentika, pp. 57–64, 2016, [Online]. Available: <https://fti.uajy.ac.id/sentika/publikasi/makalah/2016/95.pdf>.
- [12] O. Caelen, "A Bayesian interpretation of the confusion matrix," *Ann. Math. Artif. Intell.*, vol. 81, no. 3–4, pp. 429–450, 2017, doi: 10.1007/s10472-017-9564-8.
- [13] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach," *Expert Syst. Appl.*, vol. 57, pp. 117–126, 2016, doi: 10.1016/j.eswa.2016.03.028.
- [14] D. K. Barrow and S. F. Crone, "Crogging (Cross-Validation Aggregation) for Forecasting – a novel algorithm of Neural Network Ensembles on Time Series Subsamples."
- [15] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," *Proc. - 6th Int. Adv. Comput. Conf. IACC 2016*, no. Cv, pp. 78–83, 2016, doi: 10.1109/IACC.2016.25.

