



**“PERBANDINGAN ALGORITMA NAÏVE BAYES DAN SUPPORT
VECTOR MACHINE PADA ANALISIS SENTIMEN MENGENAI
HEPATITIS AKUT PADA TWITTER”**
TUGAS AKHIR

Muhammad Adriansyah Budi Prasetyo
41518010123

MERCU BUANA

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022**



**“PERBANDINGAN ALGORITMA NAÏVE BAYES DAN SUPPORT
VECTOR MACHINE PADA ANALISIS SENTIMEN MENGENAI
HEPATITIS AKUT PADA TWITTER”**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Muhammad Adriansyah Budi Prasetyo
41518010123

UNIVERSITAS
MERCU BUANA
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022

LEMBAR PERNYATAAN ORISINALITAS

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518010123

Nama : Muhammad Adriansyah Budi Prasetyo

Judul Tugas Akhir : Perbandingan Algoritma Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Mengenai Hepatitis Akut Pada Twitter

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 07 Juli 2022



M. Adriansyah Budi P

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Muhammad Adriansyah Budi Prasetyo
NIM : 41518010123
Judul Tugas Akhir : Perbandingan Algoritma Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Mengenai Hepatitis Akut Pada Twitter

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 07 Juli 2022


M. Adriansyah Budi P.

SURAT PERNYATAAN LUARAN TUGAS AKHIR

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Muhammad Adriansyah Budi Prasetyo
 NIM : 41518010123
 Judul Tugas Akhir : Perbandingan Algoritma Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Mengenai Hepatitis Akut Pada Twitter

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan
		Jurnal Nasional Terakreditasi	✓
		Jurnal International Tidak Bereputasi	Diterima
		Jurnal International Bereputasi	✓
Disubmit/dipublikasikan di :	Nama Jurnal : Jurnal Ilmiah Kaputama		
	ISSN : 2548-9720		
	Link Jurnal : https://jurnal.kaputama.ac.id/index.php/JIKA		
	Link File Jurnal Jika Sudah di Publish : https://docplayer.info/229083193-Perbandingan-algoritma-naive-bayes-dan-support-vector-machine-pada-analisis-sentimen-mengenai-hepatitis-akut-pada-twitter.html		

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui
Dosen Pembimbing TA

Afiyati, SSi., MT.

Jakarta, 07 Juli 2022



M Adriansyah Budi P

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010123
Nama : Muhammad Adriansyah Budi Prasetyo
Judul Tugas Akhir : Perbandingan Algoritma Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Mengenai Hepatitis Akut Pada Twitter

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Juli 2022



(Anis Cherid, SE., MTI)

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010123
Nama : Muhammad Adriansyah Budi Prasetyo
Judul Tugas Akhir : Perbandingan Algoritma Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Mengenai Hepatitis Akut Pada Twitter

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Juli 2022



UNIVERSITAS
MERCU BUANA

(Muhammad Rifqi, S.Kom, M.Kom)

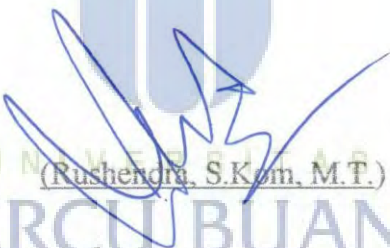
LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010123
Nama : Muhammad Adriansyah Budi Prasetyo
Judul Tugas Akhir : Perbandingan Algoritma Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Mengenai Hepatitis Akut Pada Twitter

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Juli 2022


UNIVERSITAS
(Rushendra, S.Kom, M.T.)
MERCU BUANA

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

NIM : 41518010123
Nama : Muhammad Adriansyah Budi Prasetyo
Judul Tugas Akhir : Perbandingan Algoritma Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Mengenai Hepatitis Akut Pada Twitter

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Juli 2022

Menyetujui,

UNIVERSITAS
MERCU BUANA

(Afiyati, SSi., MT.)
Dosen Pembimbing

Mengetahui,



(Wawan Gunawan, S.Kom, MT)
Koord. Tugas Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph.D., IPM.)
Ka. Prodi Teknik Informatika

KATA PENGANTAR

Puji syukur kita panjatkan kepada kehadiran Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, dikarenakan Tugas Akhir ini yang berjudul “Perbandingan Algoritma Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Mengenai Hepatitis Akut Pada Twitter” dapat diselesaikan. Tugas Akhir ini dibuat sebagai syarat untuk menyelesaikan program Sarjana (S1) Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Mercu Buana.

Penulis menyadari bahwa Tugas Akhir ini tidak mungkin terselesaikan tanpa adanya dukungan, bantuan, bimbingan, dan nasihat dari berbagai pihak selama penyusunan tugas akhir. Pada kesempatan ini penulis menyampaikan terimakasih setulus-tulusnya kepada :

1. Seluruh Dosen Program Studi Teknik Informatika yang sudah memberikan ilmu yang bermanfaat selama kuliah berlangsung. Memberi kesempatan untuk belajar, berkarya dan juga berkembang.
2. Afiyati, SSi., MT. Selaku dosen Pembimbing Tugas Akhir yang telah memberikan masukan, nasihat, dan motivasi dalam menyelesaikan tugas akhir dan juga meluangkan waktu sebagian besarnya untuk melakukan bimbingan hingga dapat menyelesaikan tugas akhir dengan baik.
3. Anis Cherid, SE, MTI selaku Dosen Pembimbing Akademik
4. Seluruh Staff Administrasi dan Tata Usaha yang telah banyak membantu dan memberikan kemudahan, terima kasih atas semua pelayanan dan arahannya.
5. Keluarga tercinta Ibu, Mba Puput dan Adik saya Aurel yang telah memberikan dukungan tiada habisnya. Semangat serta doa yang tiada hentinya dalam upaya meraih setiap impian penulis.
6. Keluarga besar tercinta Nenek Cia, Tante Ci, Om Syamsul, Om Haritsz, Om erik, Tante Ama, Om Eji, dan Om Din yang telah memberikan semangat kepada penulis untuk terus berusaha menyelesaikan tugas akhir dengan baik.

7. Sepupu tercinta Kak Ocen, Kak Nisa, Kak Zia, Kak Nurul, Kak Diah, Ica, Iki, dan Vio yang memberikan doa serta dukungan kepada penulis untuk tetap semangat dalam menyelesaikan tugas akhir.
8. Rekan seperjuangan Wahyu Saputra, Wamandaka Wardani, Ahmad Rezza Ferdian yang selalu membantu penulis ketika mendapatkan kesulitan dalam penulisan skripsi dan selalu memberikan semangat dan motivasi kepada penulis untuk selalu semangat. Yang selalu berjuang bersama penulis dalam menyelesaikan skripsi.
9. Kepada semua pihak yang tidak dapat disebutkan satu persatu yang dengan tulus Ikhlas memberikan Doa, Motivasi, dan Support sehingga dapat terselesaikan tugas akhir ini.

Dalam penelitian tugas akhir ini masih banyak kekurangan dan kesalahan karena itu segala kritik dan saran yang membangun dan menyempurnakan penelitian tugas akhir ini serta bermanfaat bagi penulis dan pembaca

Jakarta, 26 Juni 2022

Penulis

Muhammad Adriansyah Budi Prasetyo



UNIVERSITAS
MERCU BUANA

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR ...	iii
SURAT PERNYATAAN LUARAN TUGAS AKHIR	iv
LEMBAR PERSETUJUAN	v
LEMBAR PERSETUJUAN PENGUJI	vi
LEMBAR PENGESAHAN	ix
ABSTRAK	x
ABSTRACT	xi
KATA PENGANTAR	xii
DAFTAR ISI	xiii
NASKAH JURNAL	1
KERTAS KERJA	11
BAB 1. LITERATUR REVIEW	12
BAB 2. ANALISIS DAN PERANCANGAN	26
BAB 3. SOURCE CODE	27
BAB 4. DATASET	45
BAB 5. TAHAPAN EKSPERIMEN	47
BAB 6. HASIL SEMUA EKSPERIMEN	53
BAB 7. KESIMPULAN	55
DAFTAR PUSTAKA	56
LAMPIRAN DOKUMEN HAKI	58
LAMPIRAN KORESPONDENSI	59

NASKAH JURNAL

“Perbandingan Algoritma Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Mengenai Hepatitis Akut Pada Twitter”

M. Adriansyah Budi Prasetyo,
Universitas Mercu Buana

*Jl. Raya, RT.4/RW.1, Meruya Sel., Kec. Kembangan, Jakarta, Daerah Khusus
Ibukota Jakarta*

E-mail: 41518010123@student.mercubuana.ac.id

ABSTRACT

At this time of course it is not foreign to us to hear of various new viruses, one of which is acute hepatitis which attacks the liver which causes symptoms such as fever, nausea, vomiting, loss of appetite, acute diarrhea and so on. This study was made by researchers to find out the response of the Indonesian people regarding acute hepatitis by using social media in the form of twitter. Twitter is a utility that allows users to send messages worldwide with tweets consisting of 250 characters that allow people to give their own opinion, without restrictions, of course in this case it can be used to analyze public opinion about acute hepatitis. By using Naïve Bayes and Support Vector Machine which is an algorithm that aims to classify the data to be selected such as acute hepatitis which is taken on social media in the form of twitter. In this section will display the results of the experiments that have been carried out. In this research, we take a case study on Acute Hepatitis by using two algorithms, namely Nave Bayes and Support Vector Machine. data sharing is the separation of training and testing data based on their proportions, for example 70%: 30% means 70% is training data and 30% is data testing. Then the classification process is carried out using the Naïve Bayes algorithm and the Support Vector Machine. Naive Bayes multinomial accuracy value is 0.66. In my opinion, this accuracy is not high enough in predicting or classifying the sentiments of twitter netizens on the topic of acute hepatitis. And for the accuracy value generated by the support vector machine model of 0.81 which makes this model have the most precise and optimal classification/prediction results when used to classify twitter citizen sentiment when compared to the multinomial naive bayes model.

Keywords: Hepatitis Akut, Twitter, Naïve bayes, Support Vector Machine

ABSTRAK

Pada masa ini tentunya tidak asing bagi kita mendengar berbagai virus baru salah satunya yaitu Hepatitis akut yang menyerang liver yang menimbulkan gejala seperti demam, mual, muntah, hilang nafsu makan, diare akut dan sebagainya. Penelitian ini dibuat oleh peneliti untuk mengetahui tanggapan masyarakat Indonesia mengenai hepatitis akut ini dengan menggunakan medsos berupa twitter. Twitter ialah utilitas yang memungkinkan pengguna mengirimkan pesan pada semua global dengan tweet yang terdiri dari 250 karakter yang membuat masyarakat mampu memberikan pendapatnya masing-masing, tanpa Batasan, tentunya dalam hal ini bisa dimanfaatkan untuk menganalisis pendapat masyarakat mengenai hepatitis akut. Dengan menggunakan Naïve Bayes dan Support Vector Machine yang merupakan algoritma yang bertujuan untuk mengklasifikasikan data yang akan dipilih seperti hepatitis akut yang diambil pada medsos berupa twitter. Pada bagian ini akan menampilkan hasil dari eksperimen yang sudah dilakukan. Pada penelitian ini mengambil studi kasus mengenai Hepatitis Akut dengan menggunakan dua Algoritma yaitu Naïve Bayes dan Support Vector Machine. Pembagian presentasi data merupakan pemisahan data training dan testing berdasarkan persentasenya, contoh 70% : 30% berarti 70% merupakan data training dan 30% merupakan data testing. Kemudian dilakukan proses klasifikasi dengan menggunakan algoritma Naïve Bayes dan Support Vector Machine. Nilai akurasi multinomial naive bayes sebesar 0.66. Menurut saya akurasi tersebut tidak cukup tinggi dalam memprediksi atau mengklasifikasikan sentimen warganet twitter terhadap topik hepatitis akut. Dan untuk Nilai akurasi yang dihasilkan oleh model support vector machine sebesar 0.81 yang menjadikan model ini memiliki hasil klasifikasi/prediksi yang paling tepat dan optimal jika digunakan untuk mengklasifikasikan sentimen warganet twitter jika dibandingkan dengan model multinomial naive bayes.

Kata Kunci: Hepatitis Akut, Twitter, Naïve bayes, Support Vector Machine

I. PENDAHULUAN

Pada tahun 2022, ditemukan virus baru yang pertama kali muncul di inggris dan kemudian menyebar ke negara bagian eropa hingga asia. Dan meluasnya virus ini, termasuk di Indonesia yang membuat pemerintah Indonesia mengupayakan berbagai cara untuk mencegah penyebaran virus ini di Indonesia. Salah satu upaya yang dilakukan pemerintah Indonesia adalah dengan memberikan imbauan dan kesadaran akan adanya virus hepatitis akut yang menyerang

pada organ lever. Yang menyebabkan seluruh sektor pemerintahan, selain itu juga berdampak pada pola hidupan sehari-hari masyarakat yang berubah. Dengan fenomena tersebut, masyarakat banyak yang memberikan berbagai macam pendapat baik itu pujian atau keluh kesah yang diposting di berbagai media sosial, salah satunya adalah Twitter.

Pendapat masyarakat tersebut dibedakan menjadi pendapat positif dan negatif menggunakan analisis sentimen. Sehingga sentimen positif

dan negatif pada tweet maka dikerjakan secara manual, namun dari jumlah yang di analisis maka dibutuhkan waktu dan usaha yang tentunya banyak juga. Oleh sebab itu, Teknik manual sudah tidak efektif lagi sehingga menggunakan mesin yang dapat menganalisis yaitu Naïve Bayes dan Support Vector Machine. Dalam melakukan penelitian ini peneliti menyiapkan beberapa tahapan alur metode dalam penelitian ini, meliputi sebuah riset dan tahapan crawling data pada twitter menggunakan keyword yang sudah ditentukan dan banyak data sebanyak 900 lebih data yang berisi tanggal, nama user, dan isi tweet. Kemudian, dilakukan pre-processing data untuk dilanjutkan ke pengolahan data menggunakan algoritma Naïve Bayes dan Support Vector Machine untuk mendapatkan hasil sentimen dari data tweet yang digunakan.

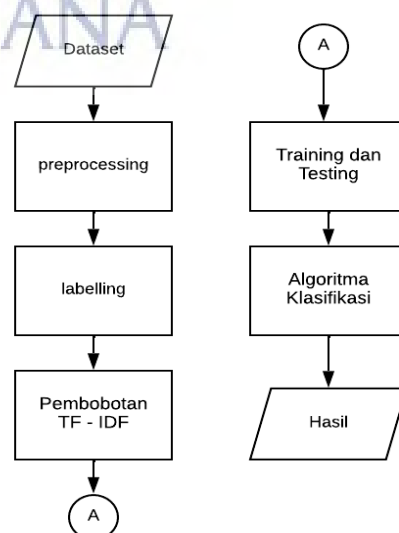
Proses pengklasifikasian teks yang terkandung dalam sebuah kata, kalimat atau dokumen dan kemudian menentukan pendapat yang diungkapkan dalam kalimat atau dokumen apakah itu positif atau negatif disebut analisis sentimen [1]

II. METHODS

Penelitian ini menggunakan metode kuantitatif dimana data yang telah terkumpul diolah dan didapatkan melalui Twitter API menggunakan python kemudian data yang diambil 1 minggu belakang dengan tema yang akan dipilih. Dalam penelitian ini, penelitian ini menggunakan python dibantu untuk melakukan preprocessing dan labeling pada Google Colab serta proses pengambilan data. Python adalah Bahasa pemrograman

interpretative yang dianggap mudah dipelajari serta berfokus pada keterbacaan kode. Dengan kata lain, python di klaim sebagai bahasa pemrograman yang memiliki kode-kode pemrograman yang sangat jelas, lengkap, dan mudah dipahami. Python secara umum berbentuk pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Python dapat digunakan untuk berbagai keperluan pembangun perangkat lunak dan dapat berjalan di berbagai platform sistem operasi [2]. Pada penelitian ini juga dilakukan eksperimen dengan membandingkan kinerja algoritma klasifikasi Naïve Bayes dan Support Vector Machine untuk mengetahui algoritma mana yang memiliki tingkat akurasi terbaik. Alur penelitian diilustrasikan dalam gambar 2.1 Tahapan Penelitian

Tahap awal penelitian ini adalah menyiapkan dataset Hepatitis Akut.



Gambar 2.1 Tahapan Penelitian

Pengumpulan dataset dilakukan menggunakan Visual Studio Kode

python yang diambil melalui Twitter API dengan memasukkan kata kunci yang relevan sesuai topik yang dipilih. Tahap selanjutnya adalah tahap preprocessing untuk membersihkan dataset yang telah terkumpul. Pada penelitian ini, tahapan preprocessing yang dilakukan adalah cleansing, case folding, tokenization, stopword removal, stemming dan remove duplikat. Setelah tahap preprocessing, tahap selanjutnya adalah tahap labeling untuk memberi label pada dataset yang telah dibersihkan dengan menggunakan TextLob. Dalam penelitian ini, proses pelabelan dibantu dengan menggunakan python.

Tahap selanjutnya adalah pembobotan TF-IDF. Pembobotan TF-IDF adalah proses pemberian bobot pada setiap kata yang terdapat dalam sebuah dokumen. Pada tahap pembobotan TF-IDF dilakukan untuk mengubah data berupa teks menjadi angka sehingga dapat diolah oleh komputer pada tahap klasifikasi [2].

Kemudian data yang sudah melewati tahap TF-IDF dilakukan training dan testing sebelum melakukan klasifikasi dengan tingkat presentase ukuran data sebesar 30% untuk data test dan 70% untuk data training. Data training merupakan sekumpulan data yang memiliki atribut label/kelas yang digunakan oleh mesin untuk mengenal karakteristik kumpulan data sehingga menghasilkan sebuah pola/model data. Sementara data testing adalah sekumpulan data yang juga memiliki label/kelas yang digunakan untuk menguji ketepatan pola/model dalam mengklasifikasikan data testing [3].

Tahap terakhir adalah proses klasifikasi menggunakan algoritma Naïve Bayes dan Support Vector Machine. Sebelum tahap klasifikasi dilakukan, data terlebih dahulu dibagi menjadi data Training dan Testing. Data Training tersebut digunakan untuk proses pembelajaran pada algoritma Naïve Bayes dan Support Vector Machine. Langkah selanjutnya adalah menggunakan data Testing untuk proses pengujian model klasifikasi Naïve Bayes dan Support Vector Machine.

A. Dataset

Pada tahapan ini, pengumpulan data didapatkan melalui Twitter API menggunakan Python Tweepy dan TextLob sebagai labelling. Proses yang dilakukan adalah memasukkan kata kunci yang relevan dengan topik yang dibahas. Penulis menggunakan beberapa kata kunci untuk mengumpulkan data terkait, antara lain “hepatitisakut”, “hepatitis akut”, “hepatitis”. Dari tweet yang telah dikumpulkan pada waktu saat pengambilan data pada tanggal 17 Juni 2022 sampai dengan 26 Juni 2022, data yang telah dikumpulkan terdapat pada gambar 2.2 Data Twitter.

	Username
0	Base Anak FK Open DM
1	Billion Store ID
2	Lyndsay
3	Lyndsay
4	Lyndsay
...	...
1714	Snow White
1715	Dindu Nuffin
1716	nella sari
1717	Duah Menaria
1718	sunday

1719 rows × 3 columns

Gambar 2.2 Data Twitter

B. Preprocessing Data

Preprocessing data bertujuan untuk membersihkan dataset yang masih mentah menjadi kata yang siap digunakan saat dilakukan proses klasifikasi menjadi netral, negatif, dan positif. Pra-pemrosesan data adalah teknik penambangan data yang melibatkan transformasi data mentah menjadi format yang mudah dipahami. Langkah pre-preprocessing data diperlukan untuk menyelesaikan beberapa jenis masalah antara lain data noise, redundansi data, nilai data yang hilang, dll [2]. Pada penelitian ini menggunakan 6 langkah dalam tahapan preprocessing

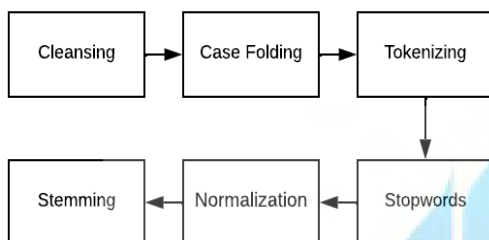
- **Cleansing Text** : Bertujuan untuk menghilangkan huruf-huruf yang tidak sesuai, simbol-simbol, dan emoticon. Didalam kolom “username” dan “Tweet” .
- **Case Folding** : Bertujuan untuk mengubah ukuran setiap

kalimat yang ada dikolom tweet dari ukuran kalimat yang besar menjadi huruf kecil.

- **Tokenizing** : Tokenisasi merupakan proses pemisahan suatu rangkaian karakter berdasarkan karakter spasi, dan mungkin pada waktu yang bersamaan [4]. Digunakan untuk memecah atau memotong susunan kalimat menjadi kata per kata dari kolom Tweet case folding
- **Stopwords** : Bertujuan untuk membuang kata-kata yang tidak deskriptif dalam melakukan pendekatan bag-of-words.
- **Normalization**: sebuah teknik untuk mengubah kata-kata "gaul" atau kata-kata yang tidak baku menjadi sebuah kata yang tepat. Normalisasi teks penting untuk teks berisik seperti pesan teks, ulasan produk, komentar media sosial, dan komentar ke posting blog di mana salah eja, singkatan, dan penggunaan kata-kata [5].

- Stemming : bertujuan untuk menambah kata imbuhan terhadap suatu kata yang tidak mengubah kelas kata tersebut. Kata-kata yang sudah diubah menjadi huruf kecil perlu dilakukan pengecekan [4]

C. Labelling



Gambar 2.3 Preprocessing

Pada tahap ini, data yang telah dibersihkan pada tahap preprocessing diklasifikasikan sebagai positif dan negatif menggunakan bahasa python dengan textblob melalui Google Colab. Pelabelan kata disesuaikan dengan kamus bahasa Indonesia dengan cara melakukan normalisasi yang telah terintegrasi dalam file Google Drive. Kamus bahasa Indonesia terdiri dari 2 file yang disimpan dalam file .csv dan .xlsx yang sudah terdiri dari sentimen positif dan negatif

Kata yang berlabel positif adalah kata yang terdeteksi menurut kamus positif, sedangkan kata yang berlabel negatif adalah kata yang terdeteksi menurut kamus negatif. Setiap kata yang terdeteksi diberi skor untuk menilai kelas sentimen. Untuk kata positif diberi nilai 1 sedangkan untuk kata negatif diberi nilai -1. Jika ada kata yang tidak ada dalam kamus

positif atau negatif, maka diberi nilai 0.

Penilaian dilakukan dengan menghitung jumlah poin pada setiap kata dalam satu kalimat. Jika nilainya ≥ 0 maka dilabeli sebagai sentimen tweet positif, sebaliknya jika nilainya < 0 maka dilabeli sebagai sentimen tweet negatif sedangkan $=0$ maka dilabeli sebagai sentimen netral. Tweet sentimen diberi nilai 1 untuk klasifikasi positif dan -1 untuk klasifikasi negatif. Berikut adalah data setelah klasifikasi sentimen diberikan:

Sentimen	Tweet
2	Positive 558
1	Neutral 243
0	Negative 98

Gambar 2.4 Labelling

D. TF-IDF Weighting

Pembobotan kata adalah proses menetapkan bobot untuk setiap kata yang terkandung dalam dokumen. Dalam mencari informasi pemeringkatan berdasarkan frekuensi kata, salah satu metode yang paling populer adalah metode TF IDF (Term Frequency Inversed Document Frequency) [4].

Untuk mengetahui seberapa penting sebuah kata mewakili sebuah kalimat, pembobotan atau perhitungan dilakukan. Penilaian dalam TF-IDF berdasarkan frekuensi terjadinya kata-kata dalam dokumen (Arsya Monica Pravina, Imam Cholissodin, 2019). TF-IDF menyajikan skor frekuensi kata,

terutama untuk kata-kata yang menarik, seperti kata-kata yang sering muncul dalam satu dokumen tetapi tidak untuk keseluruhan dokumen (Ahuja et al., 2019).

Kritik kedua datang dari itu dimensi (ukuran set fitur) di TFIDF untuk data tekstual adalah ukuran kosakata di seluruh himpunan data, menghasilkan bahwa itu membawa perhitungan besar pada pembobotan semua ini istilah [5].

TF-IDF adalah model statistik untuk mengevaluasi signifikansi kata-kata dalam kumpulan dokumen. Mungkin saja dihitung dengan mengalikan dua metrik: TF-matrix yang merupakan matriks dua dimensi dan IDF yang merupakan matriks satu dimensi. Berbagai penyempurnaan telah diusulkan oleh beberapa peneliti untuk ditingkatkan model TF-IDF klasik [6].

E. Algoritma Klasifikasi

Pada penelitian ini menggunakan dua algoritma yaitu Naïve Bayes dan Support Vector Machine Naive Bayes Classifier adalah sejenis mekanisme klasifikasi probabilistik yang berakar dari Teorema Bayes yang merupakan teori anumerta Thomas Bayes [7][8]. Naive Bayes merupakan cabang matematika yang dikenal sebagai teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan melihat frekuensi masing-masing klasifikasi dalam data pelatihan [9]. Klasifikasi Naive Bayes merupakan salah satu klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas, klasifikasi ini didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi yang mirip

dengan pohon keputusan dan jaringan saraf yang terbukti memiliki akurasi dan kecepatan tinggi ketika diimplementasikan ke dalam database dengan dataset yang besar [10].

Metode SVM merupakan suatu metode yang relatif baru untuk melakukan prediksi pada kasus regresi atau klasifikasi. SVM adalah model supervised learning yang implementasinya membutuhkan tahap pelatihan menggunakan sequential training SVM dan diikuti proses pengujian [11]. Konsep kerja Support Vector Machine yaitu dengan mencari hyperplane atau garis pembatas paling optimal yang berfungsi untuk memisahkan dua kelas [12][9].

III. HASIL DAN PEMBAHASAN

Pada bagian ini akan menampilkan hasil dari eksperimen yang sudah dilakukan. Pada penelitian ini mengambil studi kasus mengenai Hepatitis Akut dengan menggunakan dua Algoritma yaitu Naïve Bayes dan Support Vector Machine. Pembagian presentasi data merupakan pemisahan data training dan testing berdasarkan persentasenya, contoh 30% : 70% berarti 30% merupakan data training dan 70% merupakan data testing. Kemudian dilakukan proses klasifikasi dengan menggunakan algoritma Naïve Bayes dan Support Vector Machine.

1. Naïve Bayes

Nilai akurasi multinomial naive bayes sebesar 0.66. Menurut saya akurasi tersebut tidak cukup tinggi dalam

memprediksi atau mengklasifikasikan sentimen warganet twitter terhadap topik hepatitis akut. Maka dari itu model multinomial naive bayes tidak disarankan untuk digunakan kembali dalam mengklasifikasikan sentimen warganet twitter terhadap topik hepatitis akut karena hasil klasifikasinya belum cukup optimal.

Negative	0.42	0.19	0.26	26
Neutral	0.64	0.23	0.34	79
Positive	0.68	0.95	0.79	165
accuracy			0.66	270
macro avg	0.58	0.46	0.46	270
weighted avg	0.64	0.66	0.61	270

Gambar 3.1 Hasil Algoritma Naïve Bayes

2. Support Vector Machine

Nilai akurasi yang dihasilkan oleh model support vector machine sebesar 0,81 yang menjadikan model ini memiliki hasil klasifikasi/prediksi yang paling tepat dan optimal jika digunakan untuk mengklasifikasikan sentimen warganet twitter jika dibandingkan dengan model multinomial naïve bayes

Negative	0.47	0.31	0.37	26
Neutral	0.71	0.85	0.77	79
Positive	0.90	0.87	0.88	165
accuracy			0.81	270
macro avg	0.69	0.67	0.68	270
weighted avg	0.80	0.81	0.80	270

Gambar 3.2 Hasil Algoritma SVM

IV. KESIMPULAN

Hasil penelitian ini adalah eksperimen menggunakan algoritma Support Vector Machine memiliki nilai akurasi tertinggi pada penelitian ini dengan menggunakan hasil testing sebesar 30% dan training 70%, yaitu. Hasil diskusi dirangkum dalam grafik berikut:

Dapat dilihat perbandingan




Gambar 4 Perbandingan NB dan SVM dalam Grafik

antara kedua algoritma Naïve Bayes dengan warna merah dan Support Vector Machine dengan warna biru. Dapat disimpulkan pada penelitian ini dengan hasil testing 30% dan training 70% tingkat akurasi tertinggi ada pada algoritma Support Vector Machine dengan akurasi sebesar 0.81%

REFERENSI

- [1]. Putra Nuansa, “Analisis Sentimen Pengguna Twitter Terhadap Pemilihan Gubernur

- Dki Jakarta Dengan Metode Naïve Bayesian Classification Dan Support Vector Machine,” Inst. Teknol. Sepuluh Nop. Surabaya, pp. 1–101, 2017.
- [2]. F. Syadid, “Analisis Sentimen Komentar Netizen Terhadap Calon Presiden Indonesia 2019 Dari Twitter Menggunakan Algoritma Term Frequency-Invers Document Frequency (Tf-Idf) Dan Metode Multi Layer Perceptron (Mlp) Neural Network,” Skripsi Univ. Islam Negeri Syarif Hidayatullah Jakarta, p. 72, 2019.
- [3]. W. A. Prabowo and C. Wiguna, “Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM,” *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021.
- [4]. B. Gunawan, H. S. Pratiwi, and E. E. Pratama, “Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes,” *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018.
- [5]. W. Zhang, T. Yoshida, and X. Tang, “A comparative study of TF*IDF, LSI and multi-words for text classification,” *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2758–2765, 2011.
- [6]. M. Das, S. Kamalanathan, and P. Alphonse, “A Comparative Study on TF-IDF feature weighting method and its analysis using unstructured dataset,” *CEUR Workshop Proc.*, vol. 2870, pp. 98–107, 2021.
- [7]. F. J. Yang, “An implementation of naive bayes classifier,” *Proc. - 2018 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2018*, pp. 301–306, 2018.
- [8]. M. Lestandy, A. Abdurrahim, and L. Syafa’ah, “Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naive Bayes,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 802–808, 2021.
- [9]. A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter

- Menggunakan Algoritma Klasifikasi,” J. Teknoinfo, vol. 14, no. 2, p. 115, 2020.
- [10]. H. Annur, “Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes,” Ilk. J. Ilm., vol. 10, no. 2, pp. 160–165, 2018.
- [11]. V. K. S. Que, A. Iriani, and H. D. Purnomo, “Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization,” J. Nas. Tek. Elektro dan Teknol. Inf., vol. 9, no. 2, pp. 162–170, 2020.
- [12]. H. C. Husada and A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” Teknika, vol. 10, no. 1, pp. 18–26, 2021.
- 

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal yang telah terlampir sebelumnya dengan judul “Perbandingan Algoritma Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Mengenai Hepatitis Akut Pada Twitter”. Kertas kerja ini berisi semua material hasil penelitian Tugas Akhir. Di dalam kertas kerja ini disajikan beberapa bagian yang terdiri dari literature review, dataset yang digunakan, hasil dan pembahasan secara keseluruhan.

Bagian I membahas mengenai literature review yang berisi artikel jurnal sebelumnya yang menjadi dasar atau landasan dalam penelitian ini. Bagian II menjelaskan tentang analisis sistem dan perancangan yang digunakan pada penelitian ini. Bagian III menjelaskan tentang source code yang digunakan pada penelitian ini. Bagian IV menjelaskan mengenai dataset yang digunakan. Bagian V merupakan Tahapan Eksperimen pada penelitian ini. Bagian VI merupakan hasil dari berbagai metode yang di gunakan pada penelitian ini. Bagian VII bagian terakhir dari kertas kerja ini yang menjelaskan hasil keseluruhan yang telah dibuat, meliputi penjelasannya.

UNIVERSITAS
MERCU BUANA