



**ANALISA SENTIMEN MENGENAI SEPAK BOLA DI INDONESIA PADA
TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN SUPPORT
VECTOR MACHINE**

TUGAS AKHIR

Juliano Evanda
41518010090

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



**ANALISA SENTIMEN MENGENAI SEPAK BOLA DI INDONESIA PADA
TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN SUPPORT
VECTOR MACHINE**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Juliano Eyanda
41518010090

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

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518010090

Nama : Juliano Evanda

Judul Tugas Akhir : Analisa Sentimen Mengenai Sepak Bola Di Indonesia Pada Twitter Menggunakan Metode Naïve Bayes Dan Support Vector Machine

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, 29 Juli 2022



Juliano Evanda

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Juliano Evanda
NIM : 41518010090
Judul Tugas Akhir : Analisa Sentimen Mengenai Sepak Bola Di Indonesia Pada Twitter Menggunakan Metode Naïve Bayes Dan Support Vector Machine

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, 29 Juli 2022



Juliano Evanda

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Juliano Evanda
NIM : 41518010090
Judul Tugas Akhir : Analisa Sentimen Mengenai Sepak Bola Di Indonesia Pada Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	
		Jurnal Nasional Terakreditasi	✓
		Jurnal International Tidak Bereputasi	
		Jurnal International Bereputasi	
	Disubmit/dipublikasikan di :	Nama Jurnal	Ilmu Teknik Dan Komputer
		ISSN	2621-1491
	Link Jurnal	http://publikasi.mercubuana.ac.id/index.php/jitkom	
	Link File Jurnal Jika Sudah di Publish		

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

Jakarta, 29 Juli 2022


Raka Yusuf, ST, MTI


Juliano Evanda

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010090
Nama : Juliano Evanda
Judul Tugas Akhir : Analisa Sentimen Mengenai Sepak Bola Di Indonesia Pada Twitter Menggunakan Metode Naïve Bayes Dan Support Vector Machine

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, 29 Juli 2022

Menyetujui,

(Misbahul Fajri, M.TI)
Dosen Penguji 1
UNIVERSITAS
MERCU BUANA


LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010090
Nama : Juliano Evanda
Judul Tugas Akhir : Analisa Sentimen Mengenai Sepak Bola Di
Indonesia Pada Twitter Menggunakan Metode Naïve
Bayes Dan Support Vector Machine

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, 29 July 2022

Menyetujui,


(Ir. Emil K. Kaburuan, Ph.D., IPM.)
Penguji 2

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010090
Nama : Juliano Evanda
Judul Tugas Akhir : Analisa Sentimen Mengenai Sepak Bola Di Indonesia Pada Twitter Menggunakan Metode Naïve Bayes Dan Support Vector Machine

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, 29 July 2022

☒

Menyetujui.

(Runi)

(Saruni Dwiasnati, ST, MM, M.Kom)

Penguji 3
MERCU BUANA

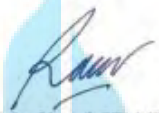
LEMBAR PENGESAHAN

NIM : 41518010090
Nama : Juliano Evanda
Judul Tugas Akhir : Analisa Sentimen Mengenai Sepak Bola Di Indonesia Pada
Twitter Menggunakan Metode Naïve Bayes Dan Support
Vector Machine

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, 7 Juli 2022

Menyetujui,

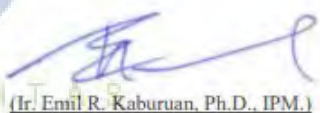


(Raka Yusuf, ST, MTI)
Dosen Pembimbing

Mengetahui,



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



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

MERCU BUANA

KATA PENGANTAR

Terimakasih dan rasa syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, dikarenakan Tugas Akhir yang berjudul “Analisa Sentimen Mengenai Sepak Bola Di Indonesia Pada Twitter Menggunakan Metode Naïve Bayes Dan Support Vector Machine” dapat diselesaikan dalam jangka waktu yang sudah ditentukan. Laporan Tugas Akhir ini dibuat sebagai syarat untuk LULUS sebagai sarjana Ilmu Komputer dari Universitas Mercu Buana.

Penulis menyadari bahwa pembuatan Tugas Akhir ini tidak terlepas dari bantuan dan bimbingan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kedua orang tua. Ayah dan Ibu, yang tak pernah lelah mendukung dan selalu percaya bahwa saya bisa menyelesaikan kuliah dengan baik, juga tak pernah luput mendoakan yang terbaik untuk proses meraih gelar sarjana bagi saya.
2. Bapak Achmad Kodar, Drs. MT selaku Dosen Pembimbing Akademik
3. Bapak Raka Yusuf, ST, MTI selaku Dosen Pembimbing Tugas Akhir yang telah memberikan masukan saat bimbingan dan meluangkan waktu sebagian besarnya untuk melakukan bimbingan dalam penyusunan tugas akhir ini hingga selesai.
4. Seluruh Dosen Program Studi Teknik Informatika yang sudah memberikan ilmu yang bermanfaat selama kuliah berlangsung. Memberi kesempatan untuk belajar, berkarya dan juga berkembang.
5. Sahabat dan kerabat, yang telah percaya bahwa saya bisa melewati dan menyelesaikan Tugas Akhir ini dan mendapatkan gelar sarjana dengan baik, juga tidak pernah bosan memberi dukungan dan doa.
6. Seluruh Staff Administrasi dan Tata Usaha yang telah banyak membantu dan memberikan kemudahan, terima kasih atas semua pelayanan dan arahannya.
7. Senior atas berbagai saran yang diberikan dan junior yang dengan semangat mendukung dikala bertemu.

8. Semua pihak dan personal yang tidak dapat disebutkan satu per satu yang terlibat dalam pembuatan Tugas Akhir ini sehingga dapat selesai dengan baik.

Akhir kata, hasil Tugas Akhir ini masih jauh dari sempurna. Masih terdapat kekurangan dalam eksperimen, cara penjelasan maupun kekeliruan penulisan. Untuk itu, kritik dan saran pembaca sangat dihargai dan diharapkan. Semoga Tugas Akhir ini dapat memberikan manfaat bagi para pembaca.

Jakarta, 7 Juli 2022



Juliano Evanda



DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	i
LEMBAR PERSETUJUAN PENGUJI.....	v
ABSTRAK	ix
ABSTRACT	x
KATA PENGANTAR.....	xi
NASKAH JURNAL	1
KERTAS KERJA	7
BAB 1. LITERATUR REVIEW	8
BAB 2. ANALISIS DAN PERANCANGAN.....	20
BAB 3. SOURCE CODE	22
BAB 4. DATASET.....	42
BAB 5. TAHAPAN EKSPERIMEN	44
BAB 6. HASIL SEMUA EKSPERIMEN.....	49
DAFTAR PUSTAKA	55
LAMPIRAN DOKUMEN HAKI.....	57
LAMPIRAN KORESPONDENSI	59

NASKAH JURNAL

Jurnal Ilmu Teknik dan Komputer
Vol. X No. X Jan 2017

ANALISA SENTIMEN MENGENAI SEPAK BOLA DI INDONESIA PADA TWITTER MENGUNAKAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

Juliano Evanda, Raka Yusuf

Teknologi Komputer, Fakultas Teknik dan Informatika,
Universitas Mercu Buana
julianoevanda@gmail.com
rksa@mercubuana.ac.id

ABSTRACT

Di masa pandemik ini sepak bola di indonesia sudah di mulai pertandingan nasional dan internasional tetapi masih dalam tahap pemantauan satgas covid-19. Hampir semua pemain di indonesia sudah melakukan proses sesuai aturan pemerintah, seperti contoh nya adalah vaksinasi dan melakukan swab antigen. Bahkan penonton saja belum di perbolehkan untuk menonton secara langsung di stadion. Dengan adanya polemik yang terjadi dalam beberapa bulan terakhir ini banyak menimbulkan opini maupun pendapat dari para pengguna media sosial salah satunya adalah twitter. Pada penelitian ini berfokus untuk mengklasifikasi terhadap opini terhadap dari cuitan di twitter yang mengandung nilai positif, negative mengenai Pesepak Bola. Pada penelitian ini melakukan eksperimen pada studi kasus Sepak Bola di Indonesia. Penelitian ini menggunakan dua algoritma yaitu Support Vector Machine dan Naive Bayes pada label otomatis dan manual. Pada pengujian label otomatis di dapat akurasi terbaik yaitu algoritma SVM dengan akurasi terbaik yaitu 82% pada percobaan 90:10. Kemudian pada label manual di dapat akurasi terbaik yaitu algoritma SVM dengan akurasi terbaik yaitu 83% pada percobaan 90:10. Pada penelitian ini pada pengujian algoritma terbaik adalah algoritma SVM pada label otomatis maupun label manual.

Kata Kunci: Analisis Sentimen, Sepak Bola Indonesia, Naive Bayes, Support Vector Machine.

PENDAHULUAN

Di masa pandemic ini sepak bola di indonesia sudah di mulai pertandingan nasional dan internasional tetapi masih dalam tahap pemantauan satgas covid-19. Hampir semua pemain di indonesia sudah melakukan proses sesuai aturan pemerintah, seperti contoh nya adalah vaksinasi dan melakukan swab antigen. Bahkan penonton saja belum di perbolehkan untuk menonton secara langsung di stadion. Dengan adanya polemik yang terjadi dalam beberapa bulan terakhir ini banyak menimbulkan opini maupun pendapat dari para pengguna media sosial salah satunya adalah twitter. Twitter merupakan layanan media sosial yang berkategori microblogging yang paling populer, pengguna dapat membaca dan berbagi pesan singkat dengan jumlah maksimal 280 karakter (Septian et al., 2019). Melalui beberapa penelitian, Dengan adanya

masalah yang terjadi dalam beberapa bulan terakhir ini banyak menimbulkan opini maupun pendapat dari para pengguna media sosial salah satunya adalah twitter tentang Sepak Bola Indonesia. Sepak Bola sangat berkaitan dengan olahraga, hiburan, social, ekonomi, Pendidikan dan lain-lain, oleh karena itu sepak bola berpengaruh untuk saat ini karena berkaitan di atas untuk membangkitkan Kembali kegiatan yang terhenti dengan berkaitan dari beberapa aspek. Pada penelitian sebelumnya yang berjudul "analisis sentiment tentang opini terhadap performa timnas sepak bola indonesia pada twitter" Pada pengujian penelitian ini dilakukan tiga kali pengujian dan setiap pengujian presentase data training nya berbeda karena jumlah data training juga berpengaruh terhadap peningkatan nilai akurasi. Dari tiga kali pengujian didapatkan akurasi tertinggi yaitu 87%, precision positif tertinggi yaitu 93%,

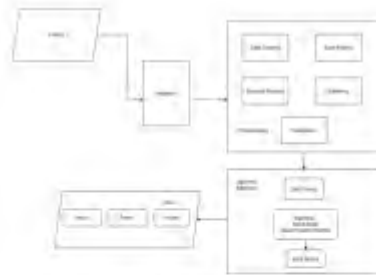
precision negatif tertinggi yaitu 85%, recall positif tertinggi yaitu 75%, dan recall negatif tertinggi yaitu 98% pada presentase 90% data training dan 10% data testing (Astiningrum et al., 2020). Berdasarkan pemaparan diatas menggunakan algoritma naïve bayes mendapatkan akurasi 87% ini belum maksimal karena tidak adanya perbandingan algoritma sehingga tidak adanya acuan terhadap akurasi dari algoritma lain. Pada penelitian ini berfokus untuk mengklasifikasi terhadap opini terhadap dari cuitan di twitter yang mengandung nilai positif, negative mengenai Pesepak Bola Indonesia menggunakan Algoritma Naïve bayes dan Support Vector Machine(SVM).

STUDI LITERATUR

Pada penelitian sebelumnya mengenai, analisis sentimen aplikasi ruang guru di twitter menggunakan algoritma klasifikasi. Hasil pengujian mendapatkan hasil bahwa aplikasi optimasi terbaik dalam model ini adalah algoritma PSO berbasis SVM dengan nilai akurasi sebesar 78,55% dan AUC sebesar 0,853 (Giovani et al., 2020). Penelitian selanjutnya Analisis Sentimen Opini Terhadap Vaksin Covid-19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes, hasil penelitian diperoleh bahwa algoritma SVM memiliki performa lebih baik pada bagian akurasi, presisi dan recall dengan nilai 80,47%, 80,23%, 80,78% dan performansi pada algoritma Naive Bayes adalah 78,64%, 77,32%, 88,13%, dengan selisih akurasi 1,83%, presisi 2,91% dan recall 2,65% (Fitriana et al., 2021). Penelitian selanjutnya Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naive Bayes, Metode untuk melakukan analisis sentimen menggunakan algoritma pengklasifikasi Naive Bayes, yang terdiri dari dua tahap yang dilakukan, yaitu tahap training dan tahap testing. Untuk menentukan kinerja dari mesin pengklasifikasi Naive Bayes dalam proses klasifikasi dilakukan eksperimen menggunakan 10 kali validasi silang (10-fold cross validation). Dari hasil eksperimen tersebut didapatkan rata-rata akurasi sebesar 93,33% (Artanti et al., 2018). Penelitian selanjutnya Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes, Berdasarkan hasil pengujian confusion matrix dengan teknik split validasi, penggunaan metode klasifikasi naïve bayes terhadap dataset yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori Good. Sementara nilai Precision sebesar 92% dan Recall sebesar 86%. algoritma Naive Bayes Classifier untuk melakukan analisis sentimen ini sudah sangat baik ditunjukkan dengan hasil akurasi yang tinggi sebesar 93% (Annur, 2018). Penelitian selanjutnya Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naive Bayes, Neural Network, KNN dan SVM, Hasil penelitian ini dapat dilihat bahwa sentimen positif muncul lebih pada Mudik 2019 sedangkan untuk sentimen negatif muncul lebih banyak hari-hari berikutnya Mudik 2019. Algoritma K-NN memiliki akurasi yang lebih tinggi, akurasi-90,76% dan AUC- 0,939; akurasi yang dihasilkan SVM-89,03% dan AUC- 0,5; akurasi Naive Bayes- 78,16% dan AUC- 0,567 kemudian yang dihasilkan Neural Network- 52,73% dan AUC-0,0 (Pertwi, 2019). Penelitian selanjutnya Analisis Sentimen Registrasi Ulang Kartu SIM Pada Twitter Menggunakan Algoritma SVM dan K-NN, hasil dari penelitian ini adalah analisis sentiment terhadap registrasi ulang kartu SIM, Metode SVM dianggap lebih baik dari pada metode k-NN dalam

penelitian ini karena menghasilkan Accuracy 78.97 % dan AUC 0.851 (Kusmira, 2019). Penelitian selanjutnya Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes, Hasil dari penelitian ini adalah klasifikasi Naive Bayes menemukan bahwa dataset tweet pemilu 2019 memiliki pola persepsi negatif sebesar 52% jauh lebih besar dari persepsi positif yaitu 18% dan Persepsi netral memiliki nilai 31% lebih tinggi dari persepsi positif. Tingkat akurasi klasifikasi Naive Bayes terhadap dataset training bernilai 81% dan dataset testing 76%, rata-rata nilai precision untuk sentimen positif 86,65%, sentimen negatif 77,15%, dan sentimen netral bernilai 80,95% sedangkan rata-rata nilai recall pada sentimen positif 36,8%, sentimen negatif 93,2% dan sentimen netral 86,8% (Juanita, 2020). Penelitian selanjutnya Perbandingan Naive Bayes, SVM, dan K-NN untuk Analisis Sentimen Gadget Berbasis Aspek. Penelitian ini bertujuan untuk mengetahui sentimen dari sebuah produk gadget. Berdasarkan hasil analisis dan pengujian yang telah dilakukan terhadap komentar Youtube gadget Samsung Galaxy Z Flip 3 dengan total 9.597 komentar, lebih banyak pengguna yang memberikan opini positif dalam aspek desain dan opini negatif pada aspek harga, spesifikasi dan citra merk. Rata-rata accuracy SVM sebesar 96.43% dilihat dari empat aspek, yaitu aspek desain sebesar 94.40%, aspek harga sebesar 97.44%, aspek spesifikasi sebesar 96.22%, dan aspek citra merk sebesar 97.63% (Iskandar & Nataliani, 2021). Penelitian selanjutnya Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes, Pada penelitian ini, data yang ada akan diklasifikasikan ke dalam 5 (lima) kelas, yaitu sangat negatif, negatif, netral, positif dan sangat positif. Data tersebut kemudian akan dievaluasi menggunakan pengujian confusion matrix dengan parameter akurasi, recall, dan precision. Hasil pengujian menunjukkan pada pengujian 3 kelas (negatif, netral dan positif) hasil terbaik didapatkan pada 90% data latih dan 10% data uji dengan nilai akurasi 77.78%, recall 93.33% dan precision 77.78% dan pada pengujian 5 kelas hasil terbaik didapatkan pada 90% data latih dan 10% data uji dengan nilai akurasi 59.33 %, recall 58.33 % dan precision 59.33 %. Hasil prediksi kelas data uji yang relevan dibandingkan antara kelas sentimen yang ditandai supervisor dan kelas sentimen yang dihasilkan oleh sistem analisis sentimen walaupun belum sepenuhnya akurat (Gunawan et al., 2018). Penelitian selanjutnya Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritma Support Vector Machine (SVM). Hasil dari penelitian ini menunjukkan parameter optimal dan pengaruh penggunaan Lexicon Based Features Dengan digunakan parameter C bernilai 10 dan learning rate bernilai 0,03 serta digunakan Lexicon Based Features dengan iterasi sebanyak 50 kali memberikan hasil accuracy sebesar 40%, precision 40%, 100% recall, dan f-measure sebesar 57,14% (Pravina et al., 2019). Berdasarkan pemaparan diatas pada analisis sentimen dapat di jadikan acuan untuk penelitian ini agar mendapatkan akurasi terbaik dari setiap algoritma dalam penelitian ini.

METHODOLOGI



Gambar 1 Tahapan Penelitian

1. Dataset

Dataset dilakukan dengan cara pengambilan data yang memanfaatkan API Twitter, dilakukan dengan mengintegrasikan API Twitter komentar netizen pada tweet- tweet mengenai Analisa Sentimen Mengenai Sepak Bola Di Indonesia Pada Twitter Menggunakan Metode Naïve Bayes Dan Support Vector Machine dan Rapid Miner yang dilakukan dengan cara scraping berdasarkan query atas term objek pada aplikasi pengelola data mining yang terhubung dengan API Twitter. Lalu proses pengumpulan data dimulai dari mengatur Twitter API di Rapid Miner Studio. Kemudian dari Twitter API dihubungkan ke internet untuk login Twitter. Langkah berikutnya adalah memasukkan query dan memilih bahasa Indonesia lalu dijalankan dan data akan tampil sesuai hashtag tersebut. Data dalam Rapid Miner Studio disimpan ke Microsoft Excel untuk dilakukakan proses selanjutnya. Tweet dan komentar akan sendirinya diperoleh berisi mengenai tingkat kepuasan masyarakat mengenai Analisa Sentimen Mengenai Sepak Bola Di Indonesia Pada Twitter Menggunakan Metode Naïve Bayes Dan Support Vector Machine.

2. Preprocessing

Preprocessing yaitu salah satu tahapan yang penting untuk data pada proses mining. Data yang digunakan dalam proses mining tidak selamanya dalam kondisi yang ideal untuk diproses. Seluruh proses melibatkan beberapa langkah: membersihkan teks online, penghapusan ruang spasi, memperluas singkatan, kata dasar (stemming), penghapusan kata henti (stopword removal), penanganan negasi dan terakhir seleksi fitur.

Berikut ini adalah penjelasan dari tahapan-tahapan preprocessing tersebut:

a. Data Cleansing adalah tahap untuk karakter dan tanda baca yang tidak diperlukan dihapus dari teks. Cleansing bertujuan untuk mengurangi gangguan/noise pada dataset. Tahap-tahap pada cleansing ini adalah Kata-kata yang tidak penting dihilangkan seperti URL, hashtag (#), username (@username), email, emoticon (:@, :-, :D), tanda baca seperti koma (,), titik (.) dan juga tanda baca lainnya (Iskak, n.d.).

b. Case folding adalah salah satu bentuk text preprocessing yang paling sederhana dan efektif meskipun sering diabaikan. Tujuan dari case folding untuk mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai 'z' yang diterima. Case folding dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital, oleh karena itu peran case folding dibutuhkan dalam mengonversi keseluruhan teks dalam

dokumen menjadi suatu bentuk standar (biasanya huruf kecil).

Contoh case folding, input kalimat : " kecewa dengan pelayanan dari CS Lazada kemarin saya diinfokan bisa loh utk aktivasi kembali akun", output kalimat : " kecewa dengan pelayanan dari cs lazada kemarin saya diinfokan bisa loh utk aktivasi kembali akun".

c. Tokenization merupakan proses memecah kalimat menjadi kata- kata menjadi lebih berarti dan bermakna. Tokenisasi memisah setiap kata yang menyusun kalimat. Contoh tokenizing, input kalimat: " kecewa dengan pelayanan dari cs lazada kemarin saya diinfokan bisa loh utk aktivasi kembali akun", output kalimat : " kecewa, dengan, pelayanan, dari, cs, lazada, kemarin, saya, diinfokan, bisa, loh, untuk, aktivasi, kembali, akun"

d. Stopword Removal merupakan proses menghilangkan kata-kata yang tidak mendeskripsikan sesuatu yang semestinya dihilangkan. Kata yang dihilangkan adalah kata yang ada dalam dictionary yang berisi daftar kata (stopword list). Contoh stop word removal, input kalimat : " kecewa, dengan, pelayanan, dari, cs, lazada, kemarin, saya, diinfokan, bias, loh, untuk, aktivasi, kembali, akun", output kalimat : " kecewa pelayanan cs lazada kemarin saya info bisa aktivasi kembali akun".

e. Stemming merupakan suatu proses untuk menemukan kata dasar dari sebuah kata. Dengan menghilangkan semua imbuhan (affixes) baik yang terdiri dari awalan (prefixes), sisipan (infixes), akhiran (suffixes) dan confixes (kombinasi dari awalan dan akhiran) pada kata turunan.

f. Remove Duplicate : Remove Duplicate dalam penelitian ini bertujuan untuk menghapus data yang memiliki duplikat.

g. Algoritma Klasifikasi adalah tempat berbagai algoritma pembelajaran mesin dilatih dan diimplementasikan. Alat yang digunakan untuk pemodelan sesuai dengan teknik tertentu, yaitu RapidMiner versi 9.8. Tahap klasifikasi adalah proses pengujian klasifikasi data tweet menggunakan algoritma klasifikasi yang digunakan. Hasil pengujian untuk masing-masing model adalah mengklasifikasikan tweet positif dan negatif untuk mendapatkan akurasi tertinggi untuk masing-masing algoritma.

3. Pelabelan

Tahap ini setelah data dicrawling. Jika data tweet telah di-crawl, data tersebut diberi label secara manual berupa sentimen positif dan negatif, data tersebut juga diubah menjadi vektor untuk dihitung menggunakan algoritma Naive Bayes dan Support Vector Machine untuk menghasilkan model klasifikasi sentimen.

Setelah melakukan tahap preprocessing, data yang telah dipreprocessing menjadi 1628 untuk data Sepak Bola Di Indonesia. Berikut merupakan hasil data setelah dilakukan tahapan preprocessing:



Gambar 2 Preprocessing Sepak Bola di Indonesia Dengan Labeling



Gambar 3 Preprocessing Sepak Bola di Indonesia Tanpa Labeling

Tabel 1 Nilai Klasifikasi

Sentimen	Positif	Negatif
Sepak Bola Indonesia	1140	499

Pada penelitian ini juga dilakukan eksperimen dengan menggunakan labeling manual terhadap kedua dataset tersebut. Label positif merupakan tweet yang memiliki pujian sedangkan tweet negative berupa kata-kata kasar dan sarkas.

Tabel 2 Klasifikasi Manual

Sentimen	Positif	Negatif
Sepak Bola Indonesia	1055	573

4. TF-IDF

TF-IDF adalah suatu proses untuk melakukan transformasi data dari data tekstual ke dalam data numerik untuk dilakukan pembobotan pada tiap kata atau fitur (Rian Hidayat, 2017).

HASIL DAN DISKUSI

Pada bagian ini akan menampilkan hasil semua eksperimen yang telah dilakukan. Pada penelitian ini melakukan eksperimen pada studi kasus Sepak Bola di Indonesia. Penelitian ini menggunakan dua algoritma yaitu Support Vector Machine dan Naïve Bayes. Terdapat tiga skenario yang menjadi pembagian persentase data dengan menggunakan split validation sebagai berikut :

Tabel 3 Split Validation

Eksperimen	Data Training	Data Testing
1	70%	30%
2	80%	20%
3	90%	10%

Hasil dari kinerja model dengan menggunakan metode labeling otomatis dan pembagian data atau percentage split 90% : 10% bahwa untuk studi kasus Sepakbola di Indonesia memiliki nilai akurasi 82% untuk algoritma SVM dan untuk algoritma Naïve Bayes sebesar 74%.

Tabel 4 SVM Split 90:10 Labeling Otomatis

	Precision	Recall	Accuracy
0	93%	48%	82%
1	79%	98%	82%

Tabel 5 Naïve Bayes Split 90:10 Labeling Otomatis

	Precision	Recall	Accuracy
0	68%	37%	74%
1	75%	92%	74%

Hasil dari kinerja model dengan menggunakan metode labeling otomatis dan pembagian data atau percentage split 80% : 20% bahwa untuk studi kasus Sepakbola di Indonesia memiliki nilai akurasi 80% untuk algoritma SVM dan untuk algoritma Naïve Bayes sebesar 73%.

Tabel 6 SVM Split 80:20 Labeling Otomatis

	Precision	Recall	Accuracy
0	88%	48%	80%
1	78%	97%	80%

Tabel 7 Naïve Bayes Split 80:20 Labeling Otomatis

	Precision	Recall	Accuracy
0	80%	26%	73%
1	72%	97%	73%

Hasil dari kinerja model dengan menggunakan metode labeling otomatis dan pembagian data atau percentage split 70% : 30% bahwa untuk studi kasus Sepakbola di Indonesia memiliki nilai akurasi 79% untuk algoritma SVM dan untuk algoritma Naïve Bayes sebesar 73%.

Tabel 8 SVM Split 70:30 Labeling Otomatis

	Precision	Recall	Accuracy
0	86%	41%	79%
1	77%	97%	79%

Tabel 9 Naïve Bayes Split 70:30 Labeling Otomatis

	Precision	Recall	Accuracy
0	61%	29%	73%
1	75%	92%	73%

Hasil dari kinerja model dengan menggunakan metode labeling manual dan pembagian data atau percentage split 90% : 10% bahwa untuk studi kasus Sepakbola di Indonesia memiliki nilai akurasi 83% untuk algoritma SVM dan untuk algoritma Naïve Bayes sebesar 73%.

Tabel 10 SVM 90:10 Labeling Manual

	Precision	Recall	Accuracy
0	78%	89%	83%
1	89%	78%	83%

Tabel 11 Naïve Bayes 90:10 Labeling Manual

	Precision	Recall	Accuracy
0	80%	60%	73%
1	68%	85%	73%

Hasil dari kinerja model dengan menggunakan metode labeling manual dan pembagian data atau percentage split 80% : 20% bahwa untuk studi kasus Sepakbola di Indonesia memiliki nilai akurasi 83% untuk algoritma SVM dan untuk algoritma Naïve Bayes sebesar 75%.

Tabel 12 SVM 80:20 Labeling Manual

	Precision	Recall	Accuracy
0	79%	87%	83%
1	86%	78%	83%

Tabel 13 Naïve Bayes 80:20 Labeling Manual

	Precision	Recall	Accuracy
0	79%	63%	75%
1	72%	85%	75%

Hasil dari kinerja model dengan menggunakan metode labeling manual dan pembagian data atau percentage split 70% : 30% bahwa untuk studi kasus Sepakbola di Indonesia memiliki nilai akurasi 80% untuk algoritma SVM dan untuk algoritma Naïve Bayes sebesar 73%.

Tabel 14 SVM 70:30 Labeling Manual

	Precision	Recall	Accuracy
0	75%	88%	80%
1	87%	74%	80%

Tabel 15 Naïve Bayes 70:30 Labeling Manual

	Precision	Recall	Accuracy
0	80%	56%	73%
1	69%	87%	73%

Hasil dari penelitian ini adalah eksperimen dengan menggunakan algoritma Support Vector Machine memiliki nilai akurasi paling tinggi pada tiga skenario eksperimen percentage split pada data yang diuji dengan menggunakan labeling otomatis dan manual. Hasil dari pembahasan terangkum dalam tabel berikut :

Tabel 16 Perbandingan Akurasi Labeling Otomatis

Percobaan	Algoritma	
	SVM	NB
90:10	82%	74%
80:20	80%	73%
70:30	79%	73%

Tabel 17 Perbandingan Akurasi Labeling Manual

Percobaan	Algoritma	
	SVM	NB
90:10	83%	73%
80:20	83%	75%
70:30	80%	73%

KESIMPULAN

Pada penelitian ini dalam menganalisa sentimen mengenai sepak bola menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM), terdapat tiga skenario eksperimen percentage split pada data yang diuji dengan menggunakan labeling otomatis dan manual. Pada

akurasi labeling otomatis algoritma SVM terdapat percobaan 90:10 didapati hasil akurasi sebesar 82%, percobaan 80:20 didapati hasil akurasi 80%, dan percobaan 70:30 didapati hasil akurasi sebesar 79%, sedangkan algoritma Naïve Bayes terdapat percobaan 90:10 didapati hasil akurasi sebesar 74%, percobaan 80:20 didapati hasil akurasi sebesar 73%, dan percobaan 90:10 didapati hasil akurasi sebesar 73%. Pada akurasi labeling manual algoritma SVM terdapat percobaan 90:10 didapati hasil akurasi sebesar 83%, percobaan 80:20 didapati hasil akurasi sebesar 83%, dan percobaan 70:30 didapati hasil akurasi sebesar 80%, sedangkan pada algoritma naïve hayes percobaan 90:10 didapati hasil akurasi sebesar 73%, percobaan 80:20 didapati hasil 75%, dan pada percobaan 70:30 didapati hasil akurasi sebesar 73%.

Penerbangan pada Dokumen Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2789–2797. <http://j-ptrik.ub.ac.id>

Rian Hidayat. (2017). *Uin Syarif Hidayatullah Jakarta Uin Syarif Hidayatullah Jakarta*. 95, 1–28. [http://repository.uinjkt.ac.id/dspace/bitstream/123456789/33026/1/NITA FITRIANI-FKIK.pdf](http://repository.uinjkt.ac.id/dspace/bitstream/123456789/33026/1/NITA%20FITRIANI-FKIK.pdf)

Septian, J. A., Fahrudin, T. M., & Nugroho, A. (2019). Journal of Intelligent Systems and Computation 43. *Journal of Intelligent Systems and Computation*, 43–49. https://doi.org/10.1007/978-1-4939-9100-5_1

DAFTAR PUSTAKA

- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160–165. <https://doi.org/10.33096/ilkom.v10i2.303.160-165>
- Artanti, D. P., Syukur, A., Prihandono, A., & Setiadi, D. R. I. M. (2018). *Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naive Bayes*. 8–9.
- Astiningrum, M., Hanih, M., & Pradana, Y. rahmat yoga. (2020). Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter. *Seminar Informatika Aplikatif Polinema (Siap)*, 35–39.
- Fitriana, F., Utami, E., & Al Fatta, H. (2021). Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes. *Jurnal Komitika (Komputasi Dan Informatika)*, 5(1), 19–25. <https://doi.org/10.31603/koitika.v5i1.5185>
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniarwati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 4(2), 143. <https://doi.org/10.26418/jp.v4i2.27526>
- Iskandar, J. W., & Nataliani, Y. (2021). Perbandingan Naive Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(6), 1120–1126. <https://doi.org/10.29207/resti.v5i6.3588>
- Juanita, S. (2020). Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes. *Jurnal Media Informatika Bulidarna*, 4(3), 552. <https://doi.org/10.30865/mih.v4i3.2140>
- Kusmira, M. (2019). Analisis Sentimen Registrasi Ulang Kartu SIM pada Twitter Menggunakan Algoritma SVM dan K-NN. *INTI Nusa Mandiri*, 14(1), 105–110.
- Pertiwi, M. W. (2019). Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naive Bayes, Neural Network, K-NN dan SVM. *Inti Nusa Mandiri*, 14(1), 27–32.
- Pravina, A. M., Cholissodin, I., & Adikara, P. P. (2019). Analisis Sentimen Tentang Opini Maskapai



KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal yang telah terlampir sebelumnya dengan judul “Analisa Sentimen Mengenai Sepak Bola Di Indonesia Pada Twitter Menggunakan Metode Naïve Bayes Dan Support Vector Machine”. Kertas kerja ini berisi semua material hasil penelitan Tugas Akhir. Di dalam kertas kerja ini disajikan beberapa bagian yang terdiri dari literature review, dataset yang digunakan, tahapan eksperimen, dan hasil eksperimen 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 latar belakang. Bagian III menjelaskan tentang source code yang digunakan pada penelitian ini. Bagian IV menjelaskan mengenai dataset yang digunakan. Bagian V memuat tahapan eksperimen yang disajikan dalam gambar beserta penjelasan dari tiap tahapan. Bagian VI merupakan bagian terakhir dari kertas kerja ini yang menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan, meliputi penjelasannya.

UNIVERSITAS
MERCU BUANA