



**ANALISIS SENTIMEN TERHADAP PENERIMAAN BANTUAN SOSIAL
TUNAI PADA TWITTER MENGGUNAKAN ALGORITMA NAIVE
BAYES DAN SUPPORT VECTOR MACHINE**

TUGAS AKHIR

Astiana
41518010036

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



**ANALISIS SENTIMEN TERHADAP PENERIMAAN BANTUAN SOSIAL
PADA TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES DAN
SUPPORT VECTOR MACHINE**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Astiana

41518010036

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

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 4151810036

Nama : Astiana

Judul Tugas Akhir : Analisa Sentimen Terhadap Penerimaan Bantuan Sosial
Tunai Pada Twitter Menggunakan Algoritma Naive 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, 4 Agustus 2022



Astiana



UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Astiana
NIM : 41518010036
Judul Tugas Akhir : Analisa Sentimen Terhadap Bantuan Sosial Tunai
Pada Twitter Menggunakan Algoritma Naive 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, 4 Agustus 2022

UNIVERSITAS
MERCU BUANA



Astiana

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Astiana
NIM : 41518010036
Judul Tugas Akhir : Analisa Sentimen Terhadap Bantuan Sosial Tunai Pada Twitter Menggunakan Algoritma 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	Diajukan ✓
		Jurnal Nasional Terakreditasi	
		Jurnal International Tidak Bereputasi	Diterima
		Jurnal International Bereputasi	
Disubmit/dipublikasikan di :	Nama Jurnal	: Jurnal Informatika: Jurnal Pengembangan IT	
	ISSN	: ISSN 2477-5126 (print), 2548-9356 (online)	
	Link Jurnal	: https://ejournal.poltektegal.ac.id/index.php/informatika/	
	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.

Jakarta, 4 Agustus 2022



LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010036
Nama : Astiana
Judul Tugas Akhir : Analisa Sentimen Terhadap Bantuan Sosial Tunai
Pada Twitter Menggunakan Algoritma Naive 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, 4 Agustus 2022



(Sabar Rudiarto, M.Kom)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010036
Nama : Astiana
Judul Tugas Akhir : Analisa Sentimen Terhadap Bantuan Sosial Tunai
Pada Twitter Menggunakan Algoritma Naive 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, 4 Agustus 2022



(Vina Ayumi, S.Kom., M.Kom)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010036
Nama : Astiana
Judul Tugas Akhir : Analisa Sentimen Terhadap Bantuan Sosial Tunai
Pada Twitter Menggunakan Algoritma Naive 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, 4 Agustus 2022



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

UNIVERSITAS
MERCU BUANA


LEMBAR PENGESAHAN

NIM : 41518010036
Nama : Astiana
Judul Tugas Akhir : Analisa Sentimen Terhadap Bantuan Sosial Tunai Pada Twitter Menggunakan Algoritma Naive 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, 4 Agustus 2022

Menyetujui,


(Umniy Salamah, ST, MMSI)
Dosen Pembimbing

Mengetahui,


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


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

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Allah SWT atas berkat, rahmat, dan hidayah-Nya penulis dapat menyelesaikan penulisan tugas akhir yang merupakan salah satu syarat kelulusan untuk mendapatkan gelar Sarjana Strata Satu (S-1) di program Studi Teknik Informatika Universitas Mercu Buana Jakarta dengan baik dan tepat waktu serta sesuai dengan ketentuan yang berlaku.

Tugas akhir ini berjudul “Analisa Sentimen Terhadap Bantuan Sosial Tunai Pada Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine”. Tugas akhir ini membahas tentang tanggapan masyarakat pada twitter mengenai bantuan sosial Tunai program pemerintah menggunakan algoritma naive bayes dan support vector machine.

Tugas akhir ini tidak dapat berjalan dengan baik tanpa bantuan, dukungan, saran dan bimbingan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis ingin menyampaikan terima kasih kepada:

1. Allah SWT atas segala berkat, rahmat, dan hidayah-Nya yang diberikan kepada penulis sehingga dapat menyelesaikan laporan tugas akhir dengan baik.
2. Kedua orang tua dan kaka, untuk bapak Nasrun ibu Sanis dan mas Eko yang tidak pernah lelah memberikan doa, dorongan, dan semangat selama penyusunan laporan tugas akhir.
3. Ibu Umniy Salamah, ST., MMSI selaku dosen pembimbing tugas akhir, atas bimbingan dan dukungan yang beliau berikan kepada penulis.
4. Bapak Ir. Emil R. Kaburuan, Ph.D., IPM. Selaku Ketua Program Studi Teknik Informatika Universitas Mercu Buana.
5. Bapak Wawan Gunawan, S.Kom, MT selaku Koordinator Tugas Akhir Prodi Informatika yang telah memberikannya arahan selama periode tugas akhir.
6. Bapak Dr. Harwkarya, MT selaku dosen pembimbing akademik saya yang telah membimbing selama perkuliahan hingga sampai pada tugas akhir.

7. Seluruh dosen Fakultas Ilmu Komputer, jurusan Teknik Informatika Universitas Mercu Buana yang telah membimbing saya dari semester awal sampai akhir.
8. Mas Doni Dwianto, Qonita Azillatin, Cemong dan Bocil (anabul) yang selaku sabar dan selalu memberikan dukungan, semangat pada pengerjaan tugas akhir ini.
9. Para Staff karyawan unit Pengabdian Masyarakat Universitas Mercu Buana tempat dimana penulis bekerja yang selalu membantu dan memberikan dukungan kepada penulis.
10. Terima kasih kepada seluruh teman angkatan 2018 dan juga khususnya untuk teman tercinta saya Yuza Giravanesa, Eka Pramudiansyah, Muhammad Iqbal, Ramzi, Fadhl Muhamad, Abdul RahanAlpat, Harvan Alans Handiva, Nabilah Mumtaz, Ismawati M, Hilwa, Citra Dewi, Willi, Yogi yang selalu memberikan saya semangat selama penulisan tugas akhir ini.
11. *Last but not least, I wanna thank me I wanna thank me for believing in me, I wanna thank me for doing all this hard work, I wanna thank me for having no days off, I wanna thank me for, for never quitting, I wanna thank me for always being a giver and tryna give more than I receive, I wanna thank me for tryna do more right than wrong, I wanna thank me for just being me at all times.*

Penulis menyadari bahwa penelitian ini jauh dari kesempurnaan. Oleh sebab itu, kritik maupun saran selalu penulis haapkan demi menghasilkan hasil terbaik dari penelitian ini. Besar harapan penulis, semoga penelitan ini dapat memberikan manfaat sekaligus menambah pengetahuan bagi berbagai pihak. Aamiin.

Jakarta, Juli 2022



Astiana

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 PENGUJI.....	v
LEMBAR PENGESAHAN.....	viii
ABSTRAK.....	ix
ABSTRACT.....	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
NASKAH JURNAL.....	1
KERTAS KERJA.....	15
BAB 1. LITERATUR REVIEW.....	16
BAB 2. SOURCE CODE.....	23
BAB 3. DATASET.....	43
BAB 4. TAHAPAN EKSPERIMEN.....	45
BAB 5. HASIL SEMUA EKSPERIMEN.....	51
DAFTAR PUSTAKA.....	63
LAMPIRAN DOKUMEN HAKI.....	66
LAMPIRAN KORESPONDENSI.....	69

NASKAH JURNAL

Analisis Sentimen Terhadap Penerimaan Bantuan Sosial Tunai Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine

Astiana^{*}), Umniy Salamah, ST., MMSI,

¹Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana, Jakarta

¹Jln. Raya, Meruya Sel., Kec. Kembangan, Jakarta, Daerah Khusus Ibukota Jakarta, 11650, Indonesia

email: ¹astianaa08@gmail.com, ²umniy.salamah@mercubuana.ac.id

Abstrak – Penelitian ini berdasarkan dari dampak covid – 19 yang merupakan virus baru yang baru muncul bertransmisi ke permukaan selama dua tahun terakhir yang ditetapkan oleh badan kesehatan internasional (WHO), hal tersebut membuat pemerintah memberikan tindakan dengan meningkatkan protokol kesehatan dan memberikan bantuan ekonomi. Bantuan Sosial Tunai (BST) merupakan salah satu bantuan dari pemerintah berupa uang tunai kepada masyarakat miskin yang terkena dampak dari covid – 19, yang merupakan bagian dari program Bantuan Sosial. Hal ini menimbulkan respon positif dan negatif dari masyarakat yang dituangkan pada media sosial Twitter. Pada penelitian ini dilakukan analisa sentimen untuk mengetahui opini masyarakat terhadap penerimaan bantuan. Opini tersebut didapatkan melalui platform Twitter dengan *keyword* adalah

“BST” “BLT” “Bantuan Sosial Tunai” “Bantuan Sosial” “Bansos” “BANSOS”. Disini metode yang digunakan yaitu algoritma klasifikasi Naive Bayes dan Support Vector Machine. Proses pelabelan teks dilakukan secara otomatis menggunakan *textblob*. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes memiliki performa yang lebih baik jika dibandingkan dengan algoritma SVM dengan nilai sebesar 87%.

Kata kunci: *BST, Naive Bayes, SVM, TextBlob*

Abstract – This research is based on the impact of covid-19, which is a new virus that has just emerged transmitted to the surface during the last two years defined by the international health agency (WHO), this has prompted the government to take action by improving health protocols and providing economic

assistance. Cash Social Assistance (BST) is one of the assistance from the government in the form of cash to the poor who have been affected by the COVID-19 impact. This has led to various responses from the public regarding improper distribution, the benefits of the assistance provided to suggestions submitted by the community regarding the effectiveness of the assistance provided. In this study, sentiment analysis was conducted to determine public opinion on the acceptance of aid. The opinion was obtained through the Twitter platform with the keywords being "BST" "BLT" "Cash Social Assistance" "Social Assistance" "Bansos" "BANSOS". Here the method used is the Naive Bayes classification algorithm and the Support Vector Machine. The text labeling process is done automatically using textblob. The results showed that the Naive Bayes algorithm had a better performance when compared to the SVM algorithm with a value of 87 percent.

Key words: *BST, Naive Bayes, SVM, TextBlob.*

PENDAHULUAN

Media sosial merupakan wadah bagi masyarakat untuk mencari informasi, menerima informasi dan menyampaikan informasi yang sedang menjadi bahan pembicaraan di lingkup masyarakat. Opini masyarakat banyak disampaikan melalui media sosial twitter, sebagai hal opini yang disampaikan masyarakat dari yang negatif dan positif. Salah satu pembahasan yang banyak di berikan pendapat oleh masyarakat twitter yaitu mengenai bantuan sosial tunai yang dimana merupakan program bantuan

pemerintah untuk memberikan bantuan tunai sebesar Rp.300.000 kepada setiap kepala keluarga yang belum mendapatkan program bantuan keluarga harapan selain tidak mendapatkan bantuan tersebut juga yang terdampak covid-19.

Analisa sentimen merupakan salah satu cabang ilmu dari textmining natural language program dan artificial intelligence. Bidang ini melakukan studi mengenai opini orang-orang, sentimen, evaluasi, tingkah laku dan emosi terhadap suatu entitas seperti produk, layanan, organisasi, individu, permasalahan, topik, acara dan atribut-atributnya.

Bantuan Sosial Tunai (BST) mendapatkan banyak tanggapan oleh masyarakat di media sosial twitter dari yang negatif sampai positif. Permasalahan yang ada yaitu banyak warga yang kurang mampu yang tidak bisa menerima BST. Hal ini dikarenakan data tidak di update, penyeleksian data masih secara manual, penyebaran informasi yang tidak merata dan penentuan keputusan yang kurang akurat sehingga peserta penerima BST tidak tepat sasaran. Tentu hal ini merugikan berbagai pihak terutama kalangan menengah kebawah yang seharusnya bisa merasakan bantuan BST ini.

Dari permasalahan tersebut, penulis mencoba untuk melakukan penelitian mengenai sentimen analisa penerima bantuan sosial tunai dari pemerintah dengan membandingkan kinerja 2 algoritma yaitu Naive Bayes dan Support Vector Machine.

*) **penulis korespondensi:** Astiana Email: astiana08@gmail.com

PENELITIAN YANG TERKAIT

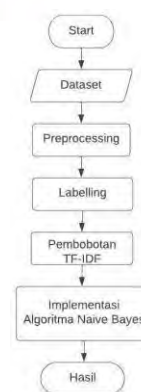
Pemilihan penggunaan metode algoritma serta data yang digunakan pada penelitian ini adalah didasarkan dari penelitian-penelitian sebelumnya yang membahas penelitian dengan topik yang sama, yaitu penelitian pertama yang membahas mengenai analisis perbandingan Detection Traffic Anomaly menggunakan metode Naive Bayes dan Support Vector Machine dengan hasil Naive Bayes memiliki nilai probabilitas 0.1 dan nilai probabilitas paling tinggi yaitu 0.8[1]. Penelitian kedua yaitu analisa sentimen terhadap aplikasi ruangguru dengan metode Naive Bayes, Random Forest dan SVM yang menghasilkan nilai akurasi tertinggi 97,16% oleh algoritma Random Fore, SVM 96.01% dan NB 94.16%[2]. Penelitian selanjutnya analisa sentimen aplikasi Halodoc dengan menggunakan algoritma Naive Bayes, Support Vector Machine dan KNN dengan hasil akurasi Naive Bayes 92.50%, SVM 93.00%, dan 95.00% KNN dimana KNN memiliki performa paling baik dan akurat dalam klasifikasi sentimen berbahasa Inggris[3]. Sentimen Analysis On E-Sports For Education Curriculum Using memiliki hasil Naive bayes memiliki akurasi yang lebih tinggi dibandingkan dengan algoritma SVM, dimana perbedaan akurasi yaitu 3,4 %. Dengan itu algoritma Naive Bayes dapat memprediksi capaian prestasi untuk kurikulum pendidikan dengan lebih baik [4]. Dari jurnal-jurnal referensi diatas penulis tertarik untuk menggunakan Metode pengujian Algoritma Naive Bayes dan Support Vector Machine dengan pelabellan menggunakan TextBlob dan pembobotan TF-IDF.

METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif karena penulis akan melakukan eksperimen dengan melakukan analisis perbandingan dengan melihat dari akurasi terbaik dari dua algoritma, adapun tahapan penelitian ini sebagai berikut:

A. Tahapan Penelitian

Tahapan pertama yang dilakukan adalah mengumpulkan dataset. Kemudian pada tahapan berikutnya dilakukan tahapan preprocessing untuk membersihkan dataset mentah. Setelah data selesai dilakukan preprocessing, tahapan berikutnya adalah dilakukan labelling untuk mengetahui kelas positif dan negatif. Setelah itu dilakukan tahapan pembobotan TF-IDF untuk memberikan bobot pada setiap kata yang terdapat di dalam dokumen. Dari sini dilakukan train test split yaitu melakukan pembagian data menjadi dua bagian untuk (data train & data test) menggunakan *train test split*. Selanjutnya, dilakukan metode klasifikasi dengan menggunakan algoritma Naive Bayes dan Support Vector Machine.



Gbr. 1 Tahapan penelitian

B. Rincian Tahapan Penelitian

Alur kerja penelitian ini ditunjukkan pada Gambar 1,

penjelasan setiap tahapannya sebagai berikut:

1. Dataset

Username	Tweet	Waktu Tweet
0 Toje Rendana	@striked @manki @anisobaweden @mika...	2022-05-21 05:57:02
1 HILIK	@widi_ross13 @sharrah23k heran kasih...!	2022-05-21 05:51:46
2 emma strenging has week	@ChRzali25 oke ig basos ke	2022-05-21 05:44:43
3 Muhammad Zahrul	@AndiKunigga Korupsi basos diakhir nasy...	2022-05-21 05:42:43
4 sangkaka	@pntkro2 jarta ming basos. nyara koru...	2022-05-21 05:40:26
...
2495 setyo prihandono	@getosaco Patah ming basos dan sarang koru...	2022-05-13 14:29:15
2496 Hmawati	@mwanaf @Adhiz21 @anisobaweden @benjam...	2022-05-13 14:27:26
2497 Inayah Inou	@Adhiz21 @Htm4_WA1 @anisobaweden @benjam...	2022-05-13 14:26:27
2498 Yayan Mulyati	SEMOGA TERKAJIKORVESTASI NEWAJONT NTYAME	2022-05-13 14:23:55
2499 Inayah Inou	@Htm4_WA1 @Adhiz21 @anisobaweden @benjam...	2022-05-13 14:23:04

Gbr. 2 Dataset

Data yang digunakan merupakan data yang didapat melalui media social Twitter. Total keseluruhan data yang digunakan sebanyak 2.500 data. Seluruh data diperoleh dengan dibantu menggunakan tools Google Collabs. Proses pengambilan data dengan memasukkan kata kunci yang relevan dengan topik yang dibahas pada studi kasus.

2. Preprocessing

Preprocessing data merupakan tahapan selanjutnya yang bertujuan untuk membersihkan dataset dari tweet atau atribut yang tidak berpengaruh atau tidak memiliki makna pada proses pengolahan data yang akan menjadikan data siap pakai agar memudahkan proses klasifikasi menjadi positif dan negatif. Data pre-processing adalah teknik data mining yang melibatkan transformasi data mentah menjadi format yang mudah dimengerti. Adapun beberapa langkah dari proses preprocessing sebagai berikut:

- **Cleaning Data**
Cleansing dalam penelitian ini bertujuan untuk menghapus Rt, username, hashtag, angka, emoticon, dan URL pada tweets.
- **Case Folding**
Case Folding dalam penelitian ini bertujuan untuk merubah isi teks dari penulisan huruf

besar menjadi penulisan huruf kecil/ lowercase.

Username	Tweet	Waktu Tweet
0 Toje Rendana	Mia mola dengan anar luan dan madam basos...	2022-05-21 05:57:03
1 HILIK	heran haan maan para kabrot sekarang sebh...	2022-05-21 05:51:46
2 emma strenging has week	oke ig basos	2022-05-21 05:44:43
3 Muhammad Zahrul	korupsi basos diakhir surya persoran ker...	2022-05-21 05:42:43
4 sangkaka	patah ming basos para koruptor kpi pap...	2022-05-21 05:40:26

Gbr. 3 Hasil Case Folding

- **Tokenization**
Tokenization dalam penelitian ini merupakan tahapan awal dalam txt-preprocessing bertujuan untuk memisahkan kata-kata dari tiap kalimat, paragraph menjadi perkata atau token.

Username	Tweet	Waktu Tweet
0 Toje Rendana	Mia mola dengan anar luan dan madam basos...	2022-05-21 05:57:03
1 HILIK	heran haan maan para kabrot sekarang sebh...	2022-05-21 05:51:46
2 emma strenging has week	oke ig basos	2022-05-21 05:44:43
3 Muhammad Zahrul	korupsi basos diakhir surya persoran ker...	2022-05-21 05:42:43
4 sangkaka	patah ming basos para koruptor kpi pap...	2022-05-21 05:40:26

Gbr. 4 Hasil Tokenization

- **Normalisasi**
Melakukan perubahan kalimat dasar dari bahasa selain bahasa Indonesia, perubahan bahasa 'gaul' menjadi kalimat yang tepat.

Username	Tweet	Waktu Tweet
0 Toje Rendana	Mia mola dengan anar luan dan madam basos...	2022-05-21 05:57:03
1 HILIK	heran haan maan para kabrot sekarang sebh...	2022-05-21 05:51:46
2 emma strenging has week	oke ig basos	2022-05-21 05:44:43
3 Muhammad Zahrul	korupsi basos diakhir surya persoran ker...	2022-05-21 05:42:43
4 sangkaka	patah ming basos para koruptor kpi pap...	2022-05-21 05:40:26

Gbr. 5 Hasil Normalisasi

- **Stopword**
Stopwords merupakan sebuah tahapan filtering yang bekerja dengan membuang kata - kata yang tidak deskriptif dalam melakukan pendekatan bag-of-words. Contoh dari Stopword Removal bahasa Indonesia adalah sebagai berikut : yang, juga, dari, dia, kami, kamu, aku, saya, ini, itu, atau, dan, pada, dengan, adalah, yaitu, ke, tak, tidak, di, pada, jika, maka, ada, pun, lain, saja, hanya, namun, seperti, kemudian, dll [5].

id	text	sentiment	score
1	positif	0.02
2	negatif	-0.02
3	netral	0.00

Gbr. 6 Hasil Stopword

- **Stemming**
text preprocessing untuk memperkecil jumlah indeks yang berbeda dari suatu data dan juga untuk melakukan pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk yang berbeda karena mendapatkan imbuhan yang berbeda pula.

id	text	stemming
1
2
3

Gbr.7 Hasil Stemming

- **Remove Duplicate** :
Remove Duplicate dalam penelitian ini bertujuan untuk menghapus data yang memiliki duplikat. Setelah melakukan tahap preprocessing, data yang telah dipreprocessing menjadi 1644 data.

3. Labelling

Pada tahapan ini, data yang sudah dibersihkan pada saat preprocessing akan dilakukan pelabelan kelas data. Pelabelan kelas data dibagi menjadi dua, kelas positif dan kelas negatif. Proses pelabelan dilakukan dengan menggunakan bahasa Python melalui tools Google Colab. Proses pelabelan dengan menggunakan bahasa Python diperlukan kamus bahasa Indonesia yang terintegrasi dengan Google Drive. Kamus berbahasa Indonesia yang digunakan terdiri

dari dua kamus, yaitu kamus berbahasa Indonesia dengan kata positif dan kamus berbahasa Indonesia dengan kata Negatif.



Gbr.8 Total Positif dan Negatif

Pada tahapan labeling, dilakukan skoring pada tiap kalimat. Setiap kata yang terdeteksi diberikan skoring untuk menilai kelas sentimen. Untuk kata positif, diberikan nilai 2 sedangkan untuk kata negatif diberikan nilai -2. Jika terdapat kata yang tidak ada dalam kamus positif ataupun negatif, diberikan nilai 0. Pemberian skoring dilakukan dengan cara menghitung jumlah nilai pada tiap kata dalam satu kalimat. Jika nilai ≥ 0.02 maka dilabeli sebagai sentimen tweet positif, sebaliknya jika nilai ≤ -0.02 maka dilabeli sebagai tweet sentimen negatif. Jika hasil skoring bernilai 0.00 maka dilabelling sebagai tweet positif.

Untuk tahapan labelling menggunakan metode TextBlob yang merupakan salah satu library pada python yang memiliki dua fungsi yaitu getSubjectivity dan getPolarity. Dimana Subjektivitas mengukur jumlah opini dan informasi faktual yang terpadat pada teks, sedangkan Polaritas mendefinisikan sentimen negatif dan sentimen positif dengan nilai $[-0.02, 0.02]$, -0.02 menetapkan sentimen negatif dan 0.02 menetapkan sentimen positif.

4. TF-IDF

Untuk mengukur seberapa penting hubungan sebuah kata mewakili suatu kalimat, dilakukan pembobotan atau perhitungan.

Dimana Term Frequency - Inverse Document Frequency ini menyatukan dua konsep yang terdiri dari frekuensi kemunculan kata di dalam suatu kalimat dan invers frekuensi kalimat[6].

5. Implementasi Algoritma

a. Naive Bayes

Naive Bayes merupakan pengklasifikasian dengan menggunakan probabilitas dan statistik yaitu memprediksi peluang di masa sebelumnya sehingga dikenal dengan Teorema Bayes (Wicaksono 2018). Dalam algoritma naive bayes sendiri bekerja dengan baik dan cepat dalam memprediksi class pada dataset.

Algoritma Naive Bayes merupakan salah satu algoritma yang terdapat dalam teknik klasifikasi[7]. Menurut teorema Bayes, Naive Bayes adalah Algoritma yang menggabungkan prior probability dengan probabilitas bersyarat pada rumus digunakan untuk menghitung kemungkinan tiap klasifikasi pada probabilitas[8].

Perhitungan pada algoritma naive bayes adalah sebagai berikut :

$$P(y) = \frac{P(x)P(x)}{P(y)}$$

Keterangan:

y = data dengan kelas yang belum diketahui.

x = hipotesis data y merupakan suatu kelas spesifik.

$P(x|y)$ = probabilitas hipotesis x berdasarkan kondisi y.

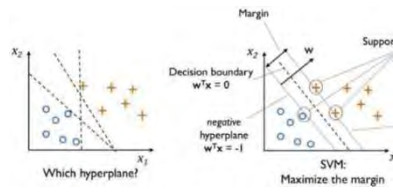
$P(x)$ = probabilitas hipotesis x.

$P(y|x)$ = probabilitas y berdasarkan kondisi pada hipotesis x.

$P(y)$ = probabilitas dari y..

b. Support Vector Machine

SVM merupakan suatu model klasifikasi berdasarkan atas kecerdasan buatan yang telah mempengaruhi berbagai macam proses pengambilan keputusan diberbagai bidang termasuk pada bidang kedokteran[9]. Model ini dapat digunakan untuk mengklasifikasikan dan memprediksi kejadian baru dan tidak terlihat berdasarkan pembelajaran dari data pelatihannya. Diberikan satu set titik data milik salah satu dari dari dua kelas, itu mewakili contoh data ke dalam ruang[10].SVM dipergunakan untuk mencari hyperplane terbaik menggunakan memaksimalkan jarak antar kelas. Hyperplane merupakan sebuah fungsi yang bisa dipergunakan buat pemisah antar kelas. dalam dua-D fungsi yang dipergunakan buat pembagian terstruktur mengenai antar kelas diklaim menjadi line whereas, fungsi yang dipergunakan buat penjabaran antar kelas pada tiga-D disebut plane similarly, sedangkan fungsi yg digunakan buat penjabaran pada pada ruang kelas dimensi yg lebih tinggi pada sebut hyperplane[11].



Gbr.9 Hyperplane yang memisahkan dua kelas positif (+1) dan negatif(-1) [11]

Adapun proses perhitungan pada penelitian ini menggunakan kernel linier sebagai berikut.

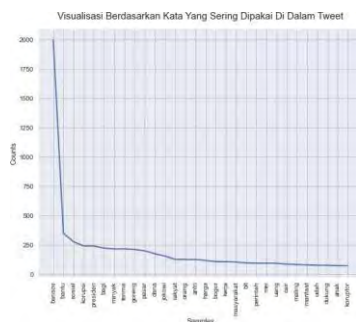
$$K(x,y)=x.y$$

Yang dimana kernel linier ini diterapkan pada saat data yang akan diklasifikasikan dapat terpisah dengan sebuah garis /hyperplane.

HASIL DAN PEMBAHASAN

Pada penelitian ini melakukan eksperimen pada studi kasus Bantuan Sosial Tunai. Eksperimen dilakukan dengan menggunakan dua algoritma yaitu Support Vector Machine dan Naïve Bayes dengan menggunakan Split Validation dan Cross Validation. Pembagian persentase split validation pada pembagian data yaitu 90%:10%, eksperimen kedua menggunakan pembagian data 80%:20% dan eksperimen yang ketiga menggunakan pembagian data 70%:30% dengan menggunakan cross validation 10 fold. Penelitian ini juga menggunakan confusion matrix pada tahap evaluasi data.

Adapun Visualisasi berdasarkan kata yang sering di pakai dalam Tweet.



Gbr.10 Visualisasi kata yang sering dipakai



Gbr.11 Menampilkan semua isi Tweet

Eksperimen pertama menggunakan pembagian data 90:10 diterapkan pada algoritma Naive Bayes dan Support Vector Machine. Hasil pengujian Naïve Bayes disajikan pada Tabel 1 dan Tabel 2.

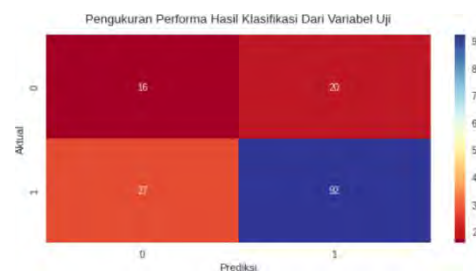
Tbl.1 Hasil Pengujian Data Test Naïve Bayes

Label	Accuracy	Precision	Recall
Positif	70%	37%	44%
Negatif	70%	82%	80%

Tbl.2 Hasil Pengujian Data Train Naïve Bayes

Label	Accuracy	Precision	Recall
Positif	88%	71%	66%
Negatif	88%	92%	93%

Hasil pengujian selanjutnya menggunakan confusion matrix pada presentase data 90:10 pada data test disajikan pada Gambar 12 dan pada data train disajikan pada Gambar 13.



Gbr 12 Confusion Matrix Data Test

Berikut adalah penjelasan dari grafik confusion matrix pada Gambar 12 atau nilai performa dari algoritma Naïve Bayes:

1. True Positive (TP) mendapatkan nilai sebesar 16.
2. True Negative (TN) mendapatkan nilai sebesar 92.
3. False Positive (FP) mendapatkan nilai sebesar 20.
4. False Negative (FN) mendapatkan nilai sebesar 27.



Gbr.13 Confusion Matrix Train

Berikut adalah penjelasan dari grafik *confusion matrix* pada Gambar 13 atau nilai performa dari algoritma Naïve Bayes:

1. True Positive (TP) mendapatkan nilai sebesar 187.
2. True Negative (TN) mendapatkan nilai sebesar 1029.
3. False Positive (FP) mendapatkan nilai sebesar 95.
4. False Negative (FN) mendapatkan nilai sebesar 76.

Hasil Pengujian menggunakan algoritma SVM pada presentase data 90:10 disajikan pada Tabel 3 dan 4.

Tbl.3 Hasil Pengujian Data Test Support SVM

Label	Accuracy	Precision	Recall
Positif	82%	83%	97%
Negatif	82%	75%	33%

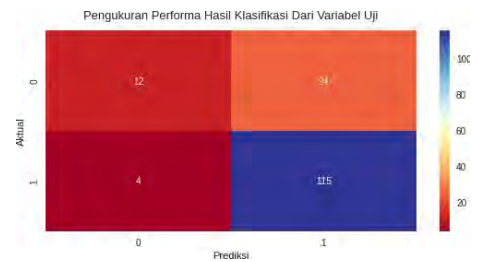
Tabel 4 Hasil Pengujian Data Train SVM

Label	Accuracy	Precision	Recall
Positif	98%	98%	100%
Negatif	98%	98%	91%

Negatif	98%	98%	91%
---------	-----	-----	-----

Hasil pengujian selanjutnya

menggunakan confusion matrix pada presentase data 90:10 pada data test disajikan pada Gambar 14 dan pada data train disajikan pada Gambar 15.



Gbr. 14 Confusion Matrix Data Test

Berikut adalah penjelasan dari grafik *confusion matrix* pada Gambar 14 atau nilai performa dari algoritma SVM:

1. True Positive (TP) mendapatkan nilai sebesar 12.
2. True Negative (TN) mendapatkan nilai sebesar 115.
3. False Positive (FP) mendapatkan nilai sebesar 24.
4. False Negative (FN) mendapatkan nilai sebesar 27.



Gbr.15 Confusion Matrix Train

Berikut adalah penjelasan dari grafik *confusion matrix* pada Gambar 15 atau nilai performa dari algoritma SVM:

1. True Positive (TP) mendapatkan nilai sebesar 258.

2. *True Negative* (TN) mendapatkan nilai sebesar 1100.
3. *False Positive* (FP) mendapatkan nilai sebesar 25.
4. *False Negative* (FN) mendapatkan nilai sebesar 5.

Eksperimen kedua menggunakan pembagian data 80:20 diterapkan pada algoritma Naive Bayes dan Support Vector Machine. Hasil pengujian Naive Bayes disajikan pada Tabel 5 dan Tabel 6.

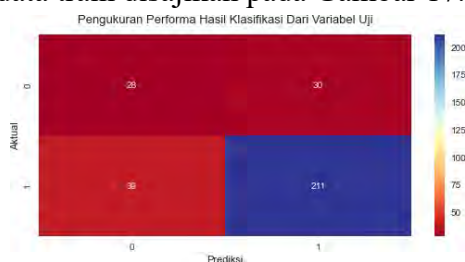
Tbl.5 Hasil Pengujian Data Test Naive Bayes

Label	Accuracy	Precision	Recall
Positif	78%	88%	84%
Negatif	78%	42%	48%

Tbl.6 Hasil Pengujian Data Train Naive Bayes

Label	Accuracy	Precision	Recall
Positif	87%	90%	92%
Negatif	87%	74%	60%

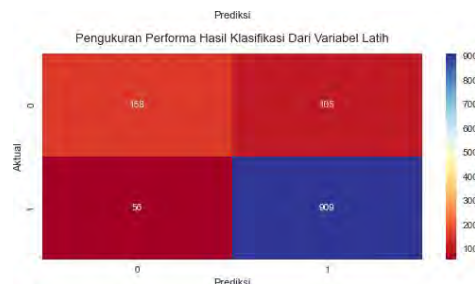
Hasil pengujian selanjutnya menggunakan confusion matrix pada presentase data 80:20 pada data test disajikan pada Gambar 16 dan pada data train disajikan pada Gambar 17.



Gbr.16 Confusion Matrix Data Test

Berikut adalah penjelasan dari grafik confusion matrix pada Gambar 16 atau nilai performa dari algoritma Naive Bayes:

1. *True Positive* (TP) mendapatkan nilai sebesar 28.
2. *True Negative* (TN) mendapatkan nilai sebesar 211.
3. *False Positive* (FP) mendapatkan nilai sebesar 30.
4. *False Negative* (FN) mendapatkan nilai sebesar 30.



Gbr.17 Confusion Matrix Train

Berikut adalah penjelasan dari grafik confusion matrix pada Gambar 17 atau nilai performa dari algoritma Naive Bayes:

1. *True Positive* (TP) mendapatkan nilai sebesar 158.
2. *True Negative* (TN) mendapatkan nilai sebesar 909.
3. *False Positive* (FP) mendapatkan nilai sebesar 105.
4. *False Negative* (FN) mendapatkan nilai sebesar 56.

Hasil Pengujian menggunakan algoritma SVM pada presentase data 80:20 disajikan pada Tabel 7 dan Table 8.

Tbl. 7 Hasil Pengujian Data Test Support SVM

Label	Accuracy	Precision	Recall
Positif	87%	87%	98%
Negatif	87%	84%	36%

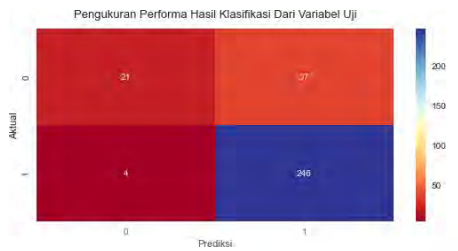
Tbl.8 Hasil Pengujian Data Train SVM

Label	Accuracy	Precision	Recall

<i>Positif</i>	98%	98%	100%
<i>Negatif</i>	98%	98%	91%

Hasil pengujian selanjutnya

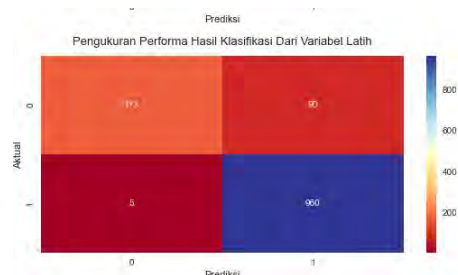
menggunakan confusion matrix pada presentase data 80:20 pada data test disajikan pada Gambar 18 dan pada data train disajikan pada Gambar 19.



Gbr.18 Confusion Matrix Data Test

Berikut adalah penjelasan dari grafik *confusion matrix* pada Gambar 18 atau nilai performa dari algoritma SVM:

1. *True Positive* (TP) mendapatkan nilai sebesar 21.
2. *True Negative* (TN) mendapatkan nilai sebesar 246.
3. *False Positive* (FP) mendapatkan nilai sebesar 37.
4. *False Negative* (FN) mendapatkan nilai sebesar 4.



Gbr.19 Confusion Matrix Train

Berikut adalah penjelasan dari grafik *confusion matrix* pada Gambar 19 atau nilai performa dari algoritma SVM:

1. *True Positive* (TP) mendapatkan nilai sebesar

2. *True Negative* (TN) mendapatkan nilai sebesar 960.
3. *False Positive* (FP) mendapatkan nilai sebesar 90.
4. *False Negative* (FN) mendapatkan nilai sebesar 5.

Eksperimen ketiga menggunakan pembagian data 70:30 diterapkan pada algoritma Naive Bayes dan Support Vector Machine. Hasil pengujian Naive Bayes disajikan pada Tabel 9 dan Tabel 10.

Tbl.9 Hasil Pengujian Data Test Naive Bayes

Label	Accuracy	Precision	Recall
<i>Positif</i>	75%	85%	83%
<i>Negatif</i>	75%	38%	42%

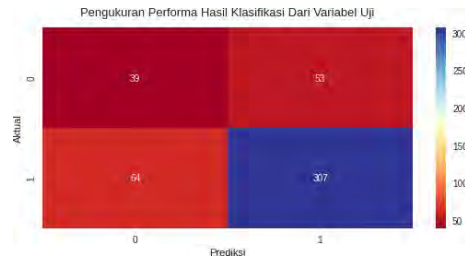
Tbl.10 Hasil Pengujian Data Train Naive Bayes

Label	Accuracy	Precision	Recall
<i>Positif</i>	87%	95%	94%
<i>Negatif</i>	87%	71%	61%

Hasil pengujian selanjutnya menggunakan *confusion matrix* pada presentase data 70:30 pada data test disajikan pada Gambar 20 dan pada data train disajikan pada Gambar 21.

Gbr.20 Confusion Matrix Data Test

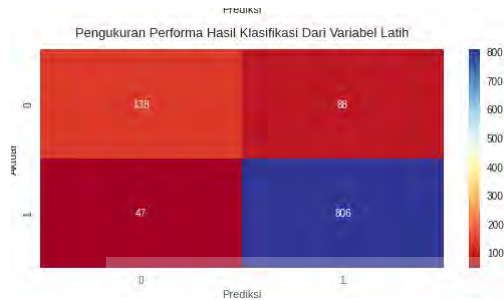
Berikut adalah penjelasan dari grafik *confusion matrix* pada Gambar 20 atau nilai performa dari algoritma Naïve Bayes:



1. *True Positive* (TP) mendapatkan nilai sebesar 39.
2. *True Negative* (TN) mendapatkan nilai sebesar 307.
3. *False Positive* (FP) mendapatkan nilai sebesar 53.
4. *False Negative* (FN) mendapatkan nilai sebesar 64.

<i>Negatif</i>	98%	98%	96%
----------------	-----	-----	-----

Hasil pengujian selanjutnya menggunakan *confusion matrix* pada presentase data 70:20 pada data test disajikan pada Gambar 22 dan pada data train disajikan pada Gambar 23.



Gbr.21 *Confusion Matrix Train*

Berikut adalah penjelasan dari grafik *confusion matrix* pada Gambar 21 atau nilai performa dari algoritma Naïve Bayes:

1. *True Positive* (TP) mendapatkan nilai sebesar 138.
2. *True Negative* (TN) mendapatkan nilai sebesar 806.
3. *False Positive* (FP) mendapatkan nilai sebesar 88.
4. *False Negative* (FN) mendapatkan nilai sebesar 47

Hasil Pengujian menggunakan algoritma SVM pada presentase data 70:30 disajikan pada Tabel 11 dan 12.

Tbl.11 Hasil Pengujian Data Test Support SVM

Label	Accuracy	Precision	Recall
<i>Positif</i>	85%	86%	98%
<i>Negatif</i>	85%	79%	36%

Tbl.12 Hasil Pengujian Data Train SVM

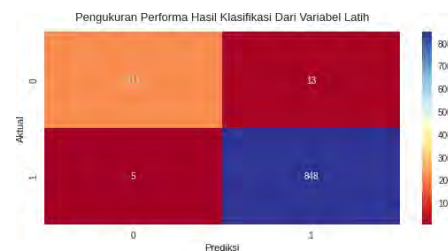
Label	Accuracy	Precision	Recall
<i>Positif</i>	98%	98%	99%



Gbr.22 *Confusion Matrix Data Test*

Berikut adalah penjelasan dari grafik *confusion matrix* pada Gambar 22 atau nilai performa dari algoritma SVM:

1. *True Positive* (TP) mendapatkan nilai sebesar 33.
2. *True Negative* (TN) mendapatkan nilai sebesar 362.
3. *False Positive* (FP) mendapatkan nilai sebesar 59.
4. *False Negative* (FN) mendapatkan nilai sebesar 9.



Gbr.23 *Confusion Matrix Data Test*

Berikut adalah penjelasan dari grafik *confusion matrix* pada Gambar 23 atau nilai performa dari algoritma SVM:

1. *True Positive* (TP) mendapatkan nilai sebesar 213.

2. *True Negative* (TN) mendapatkan nilai sebesar 848.
3. *False Positive* (FP) mendapatkan nilai sebesar 13.
4. *False Negative* (FN) mendapatkan nilai sebesar 5.

Hasil dari penelitian ini adalah eksperimen dengan menggunakan algoritma Naive Bayes memiliki nilai akurasi paling tinggi pada tiga skenario eksperimen percentage split. Hasil dari pembahasan terangkum dalam berikut:

Percobaan	Bantuan Sosial			
	SVM		NB	
	test	train	test	train
90:10	82%	98%	70%	88%
80:20	78%	87%	87%	92%
70:30	75%	87%	85%	98%

KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan menggunakan dua metode algoritma yaitu, algoritma Naive Bayes dan algoritma Support Vector Machine pada pengujian Analisa sentimen bantuan sosial tunai (BST) pada 3 model pengujian. Hasil pengujian analisa sentiment BST menunjukkan algoritma SVM lebih unggul dari pada algoritma Naive Bayes pada presentase data 90:10 menghasilkan akurasi sebesar 82% pada data test dan 98% train sedangkan Naive Bayes menghasilkan akurasi sebesar 70% pada data test dan 88% pada data train. Pada presentase 80:20 menunjukkan Naive Bayes lebih unggul dengan akurasi sebesar 87%

pada data test dan 92% pada data train sedangkan SVM menghasilkan akurasi sebesar 78% pada data test dan 87% pada data train. Presentase 70:30 menunjukkan Naive Bayes lebih baik dengan nilai akurasi sebesar 85% pada data test dan 98% pada data train sedangkan SVM menghasilkan akurasi sebesar 75% pada data test dan 87% pada data test. Sehingga dapat disimpulkan bahwa dalam analisa sentimen dalam BST menunjukkan Naive Bayes lebih baik dari pada SVM yang mana tanggapan positif lebih banyak dibandingkan negatif. Dari hasil ini diharapkan dapat menjadi masukan bagi pemerintah dengan mendapatkan lebih banyak tanggapan positif terhadap program BST.

DAFTAR PUSTAKA

- [1] I. Riadi, R. Umar, and F. D. Aini, "Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes Dan Support Vector Machine (Svm)," *Ilk. J. Ilm.*, vol. 11, no. 1, pp. 17–24, 2019, doi: 10.33096/ilkom.v11i1.361.17-24.
- [2] E. Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *J. Transform.*, vol. 18, no. 1, p. 71, 2020, doi: 10.26623/transformatika.v18i1.2317.
- [3] E. Indrayuni, A. Nurhadi, and D. A. Kristiyanti, "Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors untuk Analisa

- Sentimen Aplikasi Halodoc,” *Fakt. Exacta*, vol. 14, no. 2, p. 64, 2021, doi: 10.30998/faktorexacta.v14i2.9697.
- [4] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata, “Sentiment Analysis on E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine,” *J. Ilmu Komput. dan Inf.*, vol. 13, no. 2, pp.109–122, 2020, doi: 10.21609/jiki.v13i2.885.
- [5] 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.
- [6] N. I. Widiastuti, E. Rainarli, and K. E. Dewi, “Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen,” *J. Infotel*, vol. 9, no. 4, p. 416, 2017, doi: 10.20895/infotel.v9i4.312.
- [7] S. Dwiasnati and Y. Devianto, “Utilization of Prediction Data for Prospective Decision Customers Insurance Using the Classification Method of C.45 and Naive Bayes Algorithms,” *J. Phys. Conf. Ser.*, vol. 1179, no. 1, 2019, doi: 10.1088/1742-6596/1179/1/012023.
- [8] C. A. Sugianto and F. R. Maulana, “Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama),” *Techno.Com*, vol. 18, no. 4, pp. 321–331, 2019, doi: 10.33633/tc.v18i4.2587.
- [9] D. Fitriannah, W. Gunawan, and A. P. Sari, “Studi Komparasi Algoritma Klasifikasi C5.0, SVM dan Naive Bayes dengan Studi Kasus Prediksi Banjir,” *Techno.Com*, vol. 21, no. 1, pp. 1–11, 2022, doi: 10.33633/tc.v21i1.5348.
- [10] V. Ayumi and M. I. Fanany, “A comparison of SVM and RVM for human action recognition,” *Internetworking Indones. J.*, vol. 8, no. 1, pp. 29–33, 2016, doi: 10.13140/RG.2.1.3986.0560.
- [11] Samsudiney, “Penjelasan Sederhana tentang Apa Itu SVM?,” 2019.

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul “Analisa Sentimen Terhadap Bantuan Sosial Tunai Pada Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine”. Kertas kerja berisi semua material hasil penelitian Tugas Akhir yang tidak dimuat/ atau disertakan di artikel jurnal. Di dalam kertas ini disajikan, literatur review berisi artikel jurnal yang merupakan landasan teori pendukung dalam penelitian ini, dataset yang digunakan, source code, dan hasil eksperimen secara keseluruhan.

