



**Perbandingan Klasifikasi Analisis Sentimen Media Sosial Mengenai
Larangan Mudik Di Masa Pandemi 2020**

TUGAS AKHIR

Bekti Pangastuti
41517010026

PROGRAM STUDI TEKNIK INFORMATIKA
U FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2021



**Perbandingan Klasifikasi Analisis Sentimen Media Sosial Mengenai
Larangan Mudik Di Masa Pandemi 2020**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Bekti Pangastuti
41517010026

UNIVERSITAS
MERCU BUANA

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2021

LEMBAR PERNYATAAN ORISINALITAS

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41517010026

Nama : Bekti Pangastuti

Judul Tugas Akhir : Perbandingan Klasifikasi Analisis Sentimen Media Sosial
Mengenai Larangan Mudik Di Masa Pandemi 2020

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, 17 Februari 2021



bekti pangastuti



UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAAN PUBLIKASI TUGAS AKHIR

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Bekti Pangastuti
NIM : 41517010026
Judul Tugas Akhir : Perbandingan Klasifikasi Analisis Sentimen Media Sosial Mengenai Larangan Mudik Dimasa Pandemi 2020

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 pemyataan ini saya buat dengan sebenarnya.

Jakarta, 17 Februari 2021



UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Bekti Pangastuti
 NIM : 41517010026
 Judul Tugas Akhir : Perbandingan Klasifikasi Analisis Sentimen Media Sosial Mengenai Larangan Mudik Di Masa Pandemi 2020

Menyatakan bahwa :

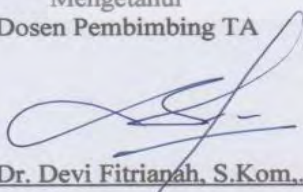
1. Luaran Tugas Akhir saya adalah sebagai berikut :

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 Teknik Informatika dan Sistem Informasi (JATISI)			
	ISSN	:			
	Link Jurnal	:			
	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


Dr. Devi Fitriani, S.Kom., M.TI

Jakarta, 17 Februari 2021



Bekti Pangastuti

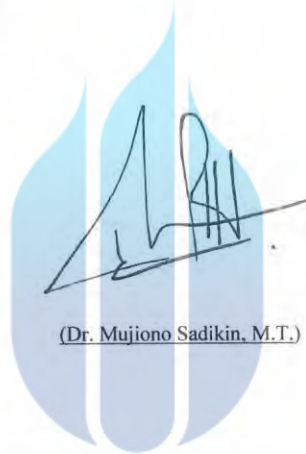
LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41517010026
Nama : Bekti Pangastuti
Judul Tugas : Perbandingan Klasifikasi Analisis Sentimen Media Sosial Mengenai Larangan Mudik Di Masa Pandemi 2020
Akhir

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, 8 Februari 2021



(Dr. Mujiono Sadikin, M.T.)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41517010026
Nama : Beki Pangastuti
Judul Tugas : Perbandingan Klasifikasi Analisis Sentimen Media Sosial Mengenai Larangan Mudik Di Masa Pandemi 2020
Akhir

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, 8 Februari 2021



(Sabar Rudiarto, S.Kom., M.Kom)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41517010026
Nama : Bekti Pangastuti
Judul Tugas Akhir : Perbandingan Klasifikasi Analisis Sentimen Media Sosial Mengenai Larangan Mudik Di Masa Pandemi 2020

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, 8 Februari 2021


(Hery Derajad Wijaya, S.Kom., M.M)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

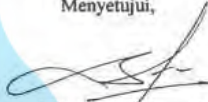
LEMBAR PENGESAHAN

NIM : 41517010026
Nama : Bekti Pangastuti
Judul Tugas Akhir : Perbandingan Klasifikasi Analisis Sentimen Media Sosial
Mengenai Larangan Mudik Di Masa Pandemi 2020

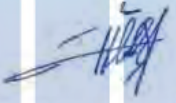
Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.


Jakarta, 8 Februari 2021

Menyetujui,


(Dr. Devi Fitriana, S.Kom., M.TI)
Dosen Pembimbing

Mengetahui,


(Diky Firdaus, S.Kom, MM)
Koord. Tugas Akhir Teknik Informatika


(Desi Ramayanti, S.Kom, MT)
Ka. Prodi Teknik Informatika

UNIVERSITAS
MERCU BUANA

ABSTRAK

Nama : Bekti Pangastuti
NIM : 41517010026
Pembimbing TA : Dr. Devi Fitrianah, S.Kom., M.TI
Judul : Perbandingan Klasifikasi Analisis Sentimen Media Sosial Mengenai Larangan Mudik Di Masa Pandemi 2020

Mudik sangat identik di Indonesia, kebiasaan tersebut rutin dilakukan pada tiap tahun menjelang hari raya umat beragama. Analisis sentimen berguna untuk mengolah atau mengekstrak data komentar publik. Penelitian ini bertujuan untuk membandingkan tiga algoritma yaitu Support Vector Machine, K-Nearest Neighbor, dan Decision Tree dengan menggunakan data yang berasal dari komentar di media sosial twitter dan youtube. Agar dapat mengetahui komentar masyarakat mengenai peraturan dilarang mudik pada masa pandemi yang dikeluarkan oleh pemerintah. Tools yang digunakan peneliti ialah python, dan untuk proses terdiri dari beberapa tahapan yaitu labelling data, pre-processing, split training dan testing, fitur extraction, mencari parameter terbaik, model, dan yang terakhir evaluasi. Hasil akurasi menyimpulkan bahwa algoritma SVM lebih unggul yaitu sebesar 91% dibandingkan algoritma KNN yang mendapatkan hasil 89% dan DT mendapatkan hasil 86%.

Kata kunci:

Analisis sentimen, Support vector machine, K-Nearest Neighbor, Decision Tree, Klasifikasi teks.



UNIVERSITAS
MERCU BUANA

ABSTRACT

Name : Bekti Pangastuti
Student Number : 41517010026
Counsellor : Dr. Devi Fitrianah, S.Kom., M.TI
Title : Perbandingan Klasifikasi Analisis Sentimen Media Sosial Mengenai Larangan Mudik Di Masa Pandemi 2020

Homecoming is very synonymous in Indonesia, this habit is routinely carried out every year before the religious feast. Sentiment analysis is useful for processing or extracting public comment data. This study aims to compare three algorithms, namely Support Vector Machine, K-Nearest Neighbor, and Decision Tree using data derived from comments on social media twitter and youtube. In order to find out the comments of the public regarding the regulations on the prohibition of going home during the pandemic period issued by the government. The tools used by researchers are python, and the process consists of several stages, namely data labeling, pre-processing, split training and testing, feature extraction, finding the best parameters, models, and finally evaluation. The accuracy results conclude that the SVM algorithm is superior at 91% compared to the KNN algorithm which gets 89% and DT gets 86%.

Key words:

Sentiment analysis, Support vector machine, K-Nearest Neighbor, Decision Tree, Classification teks.



UNIVERSITAS
MERCU BUANA

KATA PENGANTAR

Puji syukur kita panjatkan kepada Tuhan Yang Maha Esa atas berkat rahmat dan karunia Nya, sehingga penulis dapat menyelesaikan penulisan tugas akhir yang berjudul “Perbandingan Klasifikasi Analisis Sentimen Media Sosial Mengenai Larangan Mudik Di Masa Pandemi 2020” tepat pada waktunya. Tugas akhir ini disusun untuk memenuhi salah satu syarat memperoleh gelar sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, penulisan dan penelitian ini tidak akan terlaksana dengan baik. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Ibu Dr. Devi Fitriana, S.Kom., MTI, selaku dosen pembimbing Tugas Akhir, yang telah meluangkan waktunya, memberi motivasi, bimbingan serta arahan dalam penyusunan tugas akhir ini hingga selesai.
2. Ibu Desi Ramayanti, S. Kom, M.T selaku Kepala Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Mercu Buana.
3. Bapak Diky Firdaus, S. Kom, M.M selaku Koordinator Tugas Akhir Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Mercu Buana.
4. Kedua orang tua dan keluarga yang telah memberikan semangat, doa serta dukungan selama penulis menyelesaikan penelitian.
5. Teman-teman seperjuangan yang telah memberikan bantuan, mendoakan serta memberikan dukungan sehingga penulisan ini dapat di selesaikan.
6. Semua pihak yang telah membantu dalam penelitian dan penyusunan tugas akhir ini.

Akhir kata, penulis berharap penelitian ini dapat memberikan manfaat bagi semua pihak.

Jakarta, Januari 2021
Penulis

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.....	vii
LEMBAR PENGESAHAN	viii
ABSTRAK	ix
ABSTRACT.....	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xii
NASKAH JURNAL	1
KERTAS KERJA.....	13
BAB 1. LITERATUR REVIEW	14
BAB 2. SOURCE CODE	24
BAB 3. DATASET.....	36
BAB 4. TAHAPAN EKSPERIMEN	39
BAB 5. HASIL SEMUA EKSPERIMEN.....	42
DAFTAR PUSTAKA	46
LAMPIRAN DOKUMEN HAKI.....	48
LAMPIRAN KORESPONDENSI	50

NASKAH JURNAL

Perbandingan Klasifikasi Analisis Sentimen Media Sosial Mengenai Larangan Mudik Di Masa Pandemi 2020

Bekti Pangastuti¹, Devi Fitriyah²

^{1,2}Universitas Mercu Buana;

Jl. Meruya Selatan, Jakarta 11650, Indonesia

Jurusan Teknik Informatika

e-mail: ¹41517010026@student.mercubuana.ac.id, ²devi.fitriyah1@mercubuana.ac.id

Abstrak

Mudik sangat identik di Indonesia, kebiasaan tersebut rutin dilakukan pada tiap tahun menjelang hari raya umat beragama. Analisis sentimen berguna untuk mengolah atau mengekstrak data komentar publik. Penelitian ini bertujuan untuk membandingkan tiga algoritma yaitu Support Vector Machine, K-Nearest Neighbor, dan Decision Tree dengan menggunakan data yang berasal dari komentar di media sosial twitter dan youtube. Agar mengetahui komentar masyarakat mengenai peraturan dilarang mudik pada masa pandemi yang dikeluarkan oleh pemerintah. Tools yang digunakan peneliti ialah python, dan untuk proses terdiri dari beberapa tahapan yaitu labelling data, pre-processing, split training dan testing, fitur extraction, mencari parameter terbaik, model, dan yang terakhir evaluasi. Hasil akurasi menyimpulkan bahwa algoritma SVM lebih unggul yaitu sebesar 91% dibandingkan algoritma KNN yang mendapatkan hasil 88% dan DT mendapatkan hasil 86%.

Kata kunci— Analisis sentimen, Support vector machine, K-Nearest Neighbor, Decision Tree, Klasifikasi teks.

Abstract

Homecoming is very synonymous in Indonesia, this habit is routinely carried out every year before the religious feast. Sentiment analysis is useful for processing or extracting public comment data. This study aims to compare three algorithms, namely Support Vector Machine, K-Nearest Neighbor, and Decision Tree using data derived from comments on social media twitter and youtube. In order to find out the comments of the public regarding the regulations on the prohibition of going home during the pandemic period issued by the government. The tools used by researchers are python, and the process consists of several stages, namely data labeling, pre-processing, split training and testing, feature extraction, finding the best parameters, models, and finally evaluation. The accuracy results conclude that the SVM algorithm is superior at 91% compared to the KNN algorithm which gets 88% and DT gets 86%.

Keywords— Sentiment analysis, Support vector machine, K-Nearest Neighbor, Decision Tree, Text classification.

1. PENDAHULUAN

Penyakit Coronavirus 2019, disebabkan oleh virus baru coronavirus dengan reproduktifitas lebih tinggi daripada SARS, Negara Republik Rakyat Tiongkok yang

menjadi asal dari penyebaran virus tersebut, dan peristiwa tersebut terjadi pada Desember 2019 [1]. Coronavirus 2 (SARS-CoV-2) mengakibatkan terjadinya sindrom pernafasan akut parah pada penderita, Virus tersebut telah menyebabkan menurunnya kesehatan masyarakat global, dan terus menyebar ke seluruh dunia [2]. Karena adanya penyakit baru ini pemerintah mengeluarkan kebijakan larangan mudik, mudik ialah kegiatan para perantau untuk pulang ke kampung halaman agar dapat bertemu orangtua dan sanak saudara yang berada dikampung halaman. Terjadinya peningkatan jumlah pemudik pada tiap tahunnya mengakibatkan kondisi stasiun, terminal, dan bandara selalu ramai di datangi oleh para pemudik yang menggunakan jasa transportasi umum tersebut. Berdasarkan data pada tahun 2013 diperkirakan sebanyak 24 juta pemudik yang bergerak menuju kampung halaman. Sebagian besar jumlah pemudik tersebut berasal dari Jakarta menuju ke Jawa Tengah. Dan jumlah pemudik akan diperkirakan terus melonjak pada tiap tahunnya [3].

Studi komputasional dari opini pengguna internet, sentimen dan emosi melalui entitas dan atribut yang dimiliki dan diekspresikan dalam bentuk teks disebut dengan analisis sentimen atau opinion mining [4]. Keluhan masyarakat yang diterima melalui media sosial memiliki tingkat tatanan bahasa yang berbeda-beda, oleh karena itu perlu adanya sistem filterasi keluhan dengan cara menggunakan analisis sentimen pada setiap keluhan yang ditujukan pada pemerintah. Proses mengelompokkan teks yang terdapat dalam sebuah kata, kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif atau negatif disebut pengertian tersebut disebut dengan analisis sentimen [5]. Analisis sentimen atau opinion mining sendiri dapat diartikan sebagai proses memahami, mengekstrak dan melakukan olah data teks secara otomatis agar mendapatkan informasi yang terkandung dalam sebuah kata atau suatu kalimat pada opini tersebut. Terdapat sekitar 20-30 perusahaan di Amerika yang memfokuskan pada layanan analisis sentimen, karena mengingat pengaruh serta manfaat dari analisis sentimen yang menyebabkan penelitian analisis sentimen dapat berkembang dengan pesat [6]. Penelitian lain yang juga menggunakan analisis sentimen terdapat dalam [7]-[8]. Dan pada penelitian [9] juga melakukan analisis sentimen menggunakan pembobotan tekstual jumlah retweet dapat meningkatkan akurasi yaitu sebesar 82,50% dibandingkan dengan menggunakan pembobotan non-tekstual yaitu sebesar 60%.

Suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia disebut dengan klasifikasi [10]. Penelitian yang dilakukan oleh [11] membuat penelitian tentang klasifikasi dengan menggunakan algoritma Naïve Bayes dan Support Vector Machine mendapatkan hasil yang diperoleh yaitu akurasi bervariasi antara 0,85 dan 0,89, recall antara 0,34 dan 0,47, dan presisi antara 0,33 dan 0,52. Penelitian lain yang juga melakukan klasifikasi [12] melakukan klasifikasi tiga algoritma dan mendapatkan hasil percobaan menunjukkan bahwa metode kombinasi, yang menggabungkan berbagai set fitur dan klasifikasi algoritma, mampu mencapai hasil terbaik dengan nilai F-measure 94,48%.

Pada penelitian ini, mengklasifikasi komentar pada media sosial mengenai larangan mudik pada masa pandemi untuk mengelompokkan komentar tersebut berupa positif atau negatif, dengan menggunakan algoritma Support Vector Machine, K-Nearest Neighbor dan Decision Tree. Dan hasil dari masing-masing algoritma dilakukan perbandingan untuk mengetahui algoritma mana yang mendapatkan hasil terbaik.

2. TINJAUAN PUSTAKA

2.1 Support Vector Machine

Salah satu metode seleksi untuk membandingkan parameter standar seperangkat nilai diskrit yang disebut kandidat set, dan salah satu metode yang memiliki akurasi klasifikasi terbaik disebut dengan *Support Vector Machine* (SVM) [13]. *Support Vector Machine* dapat dikategorikan ke dalam kelas masuk supervised learning, dalam implementasinya perlu adanya tahap pelatihan dan disusul tahap pengujian [14].

2.2 K- Nearest Neighbor

Metode k-NN adalah model pembelajaran supervised learning dimana database populasi yang ingin digunakan untuk dilakukan identifikasi terlebih dahulu ke dalam kelas tertentu. Metode ini pertama kali dikemukakan pada awal tahun 1960 dan menjadi salah satu metode yang sering digunakan. Cara kerja metode ini yaitu melakukan pekerjaan secara intensif ketika diberikan data training yang besar [15].

Metode yang dilakukan untuk mengklasifikasi objek menggunakan data training dengan cara menggunakan jarak terdekat pada objek pernyataan tersebut adalah pengertian dari Algoritma *K-Nearest Neighbor*. Cara kerja dari Metode ini yaitu dengan menggunakan jarak terpendek dari *query distance* ke data training sample untuk menentukan *K-Nearest Neighbor*[16]

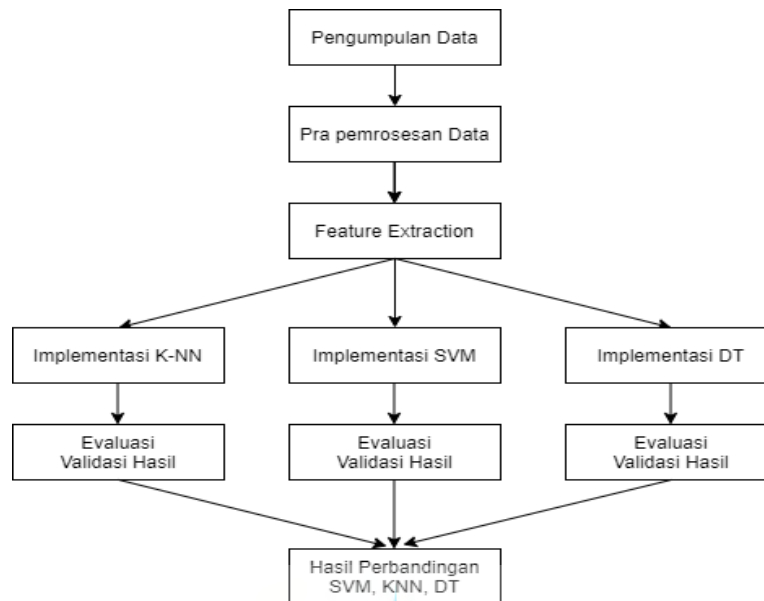
2.3 Decision Tree

Decision tree merupakan salah satu metode untuk mengklasifikasikan data [17]. Decision tree adalah pohon keputusan dimana pada tiap cabang menunjukkan pilihan alternative lainnya yang ada. Dan pada setiap daunnya untuk menunjukkan keputusan yang akan diambil. Decision tree berguna untuk mendapatkan suatu informasi untuk tujuan dalam mengambil suatu keputusan. Kelebihan dari metode ini adalah cara pengambilan keputusan yang sebelumnya rumit dan sangat kompleks, menjadi lebih spesifik dan simpel.

Cara kerja decision tree yaitu dimulai dari sebuah titik awal (root node) yang dipakai peneliti untuk mengambil tindakan. Lalu setelah itu peneliti memecahnya sesuai dengan algoritma decision tree. Hasil akhirnya berupa sebuah decision tree yang tiap cabangnya menunjukkan kemungkinan keputusan yang akan diambil beserta dengan hasil [18]

MERCU BUANA

3. METODE PENELITIAN



Gambar 1. Tahapan penelitian

Tahap pertama yang dilakukan adalah proses pengumpulan data, untuk penelitian ini dataset yang digunakan diambil dari data twitter dan youtube menggunakan hashtag dan link dari media sosial youtube, Untuk tools yang digunakan peneliti yaitu *python* dan *rapidminer* digunakan hanya untuk awal pengambilan data pada twitter. Hashtag yang digunakan pada twitter ialah #dilarangmudik dan untuk youtube diambil dari komentar video mengenai larangan mudik yang terdapat pada kanal CNN dan Kompas TV. Total keseluruhan data yang didapatkan data dari hasil pengambilan dataset sebanyak 2039 komentar. sebanyak 409 komentar berasal dari media sosial twitter dan sebanyak 1.649 komentar didapatkan dari akun youtube. Berikut tabel penjelasan mengenai pengumpulan data :

Tabel 1. Sumber Data

Media sosial	Link/hashtag	Tanggal
Twitter	#dilarangmudik	29 agustus 2020
Youtube (Kanal CNN dan Kompas TV)	https://www.youtube.com/watch?v=Lh1vCIEuMhE https://www.youtube.com/watch?v=71_6SQ9mepQ https://www.youtube.com/watch?v=YImkZpjh720 https://www.youtube.com/watch?v=6oKoQWkyYK	23 september 2020

Tahap kedua adalah preprocessing yang bertujuan agar dataset yang sudah didapatkan tersebut menjadi lebih rapih dan tersusun sehingga mudah dalam proses klasifikasi. prapemrosesan teks terdiri dari empat langkah *case folding*, *tokenizing*, *filtering (stopword removal)*, dan *stemming*. Berikut adalah penjelasan dari empat langkah pemrosesan teks yang dilakukan :

- Case Folding : berfungsi untuk mengubah huruf capital menjadi huruf kecil
- Tokenezing : memecah kalimat menjadi kumpulan kata, mengidentifikasi tanda baca yang tidak dibutuhkan

- Filtering : berguna untuk membuang kata-kata yang kurang penting yang tidak berpengaruh terhadap penelitian yang dilakukan.
- Stemming : mentransformasi kata menjadi sebuah kata dasar dengan cara membuang kata yang berimbuhan pada awalan dan akhiran, contoh yaitu membela : bela, dikatakan : kata

Terdapat proses penambahan beberapa langkah pada tahapan preprocessing yaitu:

- a. *Remove Unnecesarry Characters* bertujuan untuk menghapus bagian yang tidak diperlukan pada data seperti emoji, tanda baca, link, dan karakter yang berulang-ulang. Sehingga dapat menghasilkan kata atau kalimat yang sudah sesuai dengan kaidah kamus besar bahasa Indonesia.
- b. Normalisasi merupakan suatu proses yang digunakan untuk mengubah kata atau kalimat yang tidak baku menjadi sebuah kata atau kalimat baku yang telah sesuai dengan kaidah kamus besar bahasa Indonesia (KBBI). Perbaikan kata atau kalimat yang tidak baku dapat dilihat pada apakah sudah sesuai dan benar dalam cara penulisannya, seperti contoh dengan menggunakan bahasa modern (gue = saya), penulisan dengan kaidah kamus besar bahasa Indonesia (kwalitas = kualitas), penulisan dengan menggunakan kata singkat (adlh = adalah), penulisan dengan menggunakan angka (n41k = naik), dan penulisan dengan menggunakan kata salah eja (jelsk = jelek) [19]
- c. Metode n-gram adalah metode yang sering digunakan dan metode yang paling sederhana [20]. n-gram adalah potongan dari sejumlah n karakter dari sebuah tipe data string. Metode ini sering digunakan untuk mengambil potongan-potongan karakter dari kata atau kalimat sebanyak jumlah karakter pada kata atau kalimat tersebut. Keunggulan dari metode ini ialah tidak sensitif terhadap kesalahan penulisan kata. Beberapa karakteristik dari N-gram yaitu sebagai berikut : 1. Hanya membutuhkan penyimpanan yang sederhana dan dapat berjalan dengan efisien 2. Tetap berfungsi dengan baik walaupun terdapat kesalahan tekstual 3. Dapat bekerja dengan waktu yang relatif cepat [21]
- d. Memberikan label pada data yang dilakukan secara manual, setiap kalimat diberikan label masing- masing apakah kalimat tersebut masuk ke dalam kalimat positif (kalimat atau kata yang tidak menjelekan seseorang atau menyudutkan suatu pihak) atau kalimat negatif (kalimat atau kata yang berisi ejekan untuk menyudutkan seseorang atau pihak tertentu)

Tahap ketiga adalah fitur ekstraksi, proses pemberian bobot untuk setiap kata yang terdapat dalam sebuah dokumen disebut dengan pembobotan kata. metode TFIDF (Term Frequency - Inversed Document Frequency) adalah metode dalam pencarian informasi peringkat berdasarkan frekuensi kata, dan menjadi salah satu metode yang paling sering digunakan [22]. Dalam metode TF-IDF, Term Frequency lebih kepada istilah semakin besar jumlah term dalam dokumen, maka akan semakin

besar bobotnya, sedangkan Inverse Document Frequency semakin sedikit dokumen yang mengandung term maka nilai IDF nya akan semakin besar.

Tahap keempat adalah proses klasifikasi dengan menggunakan *python*. Metode Klasifikasi yang digunakan oleh peneliti yaitu *Support Vector Machine*, *K-Nearest Neighbor*, dan *Decision Tree*. Model klasifikasi data diuji dengan menggunakan ialah metode *10 fold cross validation*. Cara kerja dari metode ini yaitu dengan membagi dua dataset, yaitu 10 bagian dengan menggunakan 9/10 untuk proses training dan 1/10 digunakan untuk proses testing. Iterasi terjadi 10 kali dengan menggunakan variasi data training dan testing menggunakan kombinasi 10 bagian data [23]

Tahap kelima adalah evaluasi model menghasilkan prediksi dari data uji, untuk dapat mengetahui hasil akurasi dari algoritma dengan menggunakan *Confusion Matrix*.

Tabel 2. Model klasifikasi menggunakan *confusion matrix*

Kelas Prediksi	Kelas aktual	
	True Negatif	True Positif
Pediksi Positif	FP	TP
Prediksi Negatif	TN	FN

Untuk mencari akurasi, digunakan alat ukur *confusion matrix precision* dan *recall* dapat di gunakan rumus sebagai berikut [24] :

$$\text{Accuracy} = \frac{\square\square + \square\square}{\square\square + \square\square + \square\square + \square\square}$$

$$\text{Precision} = \frac{\square\square}{\square\square + \square\square}$$

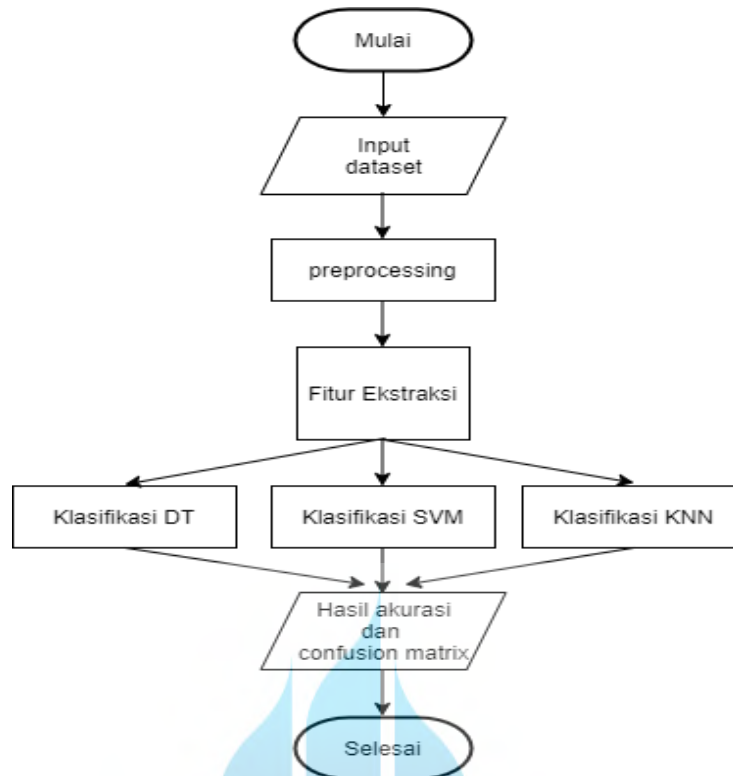
$$\text{Recall} = \frac{\square\square}{\square\square + \square\square}$$

Keterangan : UNIVERSITAS

- TP (*True Positif*) kelas positif yang benar dilabeli oleh klasifikasi.
- TN (*True Negatif*) kelas negatif yang benar dilabeli oleh klasifikasi
- FP (*False Positif*) kelas negatif salah dilabeli oleh klasifikasi yang seharusnya positif.
- FN (*False Negatif*) kelas positif yang salah dilabeli oleh klasifikasi yang seharusnya negatif

Metode yang digunakan peneliti yaitu klasifikasi dengan membandingkan tiga algoritma Support Vector Machine, K-Nearest Neighbor dan Decision Tree. Ketiga algoritma tersebut dapat menghasilkan nilai yang baik pada pengklasifikasi data analisis sentimen peraturan dilarang mudik di masa pandemi yang didapatkan dari media sosial. Dataset yang digunakan berasal dari media sosial twitter dan juga youtube. Hasil dari penelitian ini adalah dapat melihat hasil respon dari masyarakat terhadap peraturan pelarangan mudik pada masa pandemi. Dibawah ini merupakan skenario pohon keputusan dari masing-masing algoritma.

Gambar 2. Implementasi tiga metode klasifikasi



4. HASIL DAN PEMBAHASAN

Dibawah ini merupakan beberapa contoh komentar yang berhasil di crawling lalu setelah itu komentar diberikan label “positif” atau “negatif” satu persatu secara manual sebanyak dataset yang digunakan pada penelitian. Komentar dapat dikategorikan ke dalam sentimen positif memiliki kriteria apabila : berisi pujian, memberikan semangat, mendukung program pemerintah. Sebaliknya komentar dapat dikategorikan dalam sentimen negatif apabila : sumpah serapah, menyudutkan atau menjelekkan suatu pihak atau perusahaan.

Tabel 3. Contoh data yang telah diberi label

Text	Sentimen
Dishub Banda Aceh : Dilarang Mudik Sayangi Keluarga https://t.co/4jYuumV0oL	Positif
Fuck covid ! Brp kali bikin gagal mudik !!!	Negatif
Agustus 2020 sudah di pertengahan. Gak terasa kurang dari 4 bulan lagi akan meninggalkan 2020. Semoga pandemi #Covid19 ini segera berakhir.	Positif

Pada penelitian ini menggunakan pembagian data sebesar 80% data pelatihan dan 20% data pengujian, serta menggunakan fitur ekstraksi untuk peningkatan performa model menjadi lebih baik. dan GridSearchCV untuk mendapatkan nilai parameter yang tertinggi pada klasifikasi yang digunakan dan melakukan validasi. Berikut adalah tabel hasil eksperimen menggunakan fitur ekstraksi yang menyatakan “UNI” berarti fitur

unigram, “UNI+BI” berarti fitur unigram dan bigram, dan “UNI+TRI” berarti fitur unigram dan trigram.

Tabel 4. Eksperimen fitur ekstraksi

Klasifikasi model	Pembagian Data	TF-IDF Vectorizer Acc Score (%)		
		UNI	UNI+BI	UNI+TRI
SVM	80:20	91	90	90
KNN	80:20	88	87	87
DT	80:20	86	85	84

Hasil dari melakukan lima eksperimen dengan menggunakan tiga algoritma, *Support Vector Machine*, *K-Nearest Neighbor* dan *Decision Tree*. Terdapat lima skenario yang menjadi pembagian presentase data. Ekperimen pertama data yang digunakan ialah pembagian data dengan proporsi 50% : 50% , eksperimen kedua menggunakan pembagian data dengan proporsi 60% : 40%, eksperimen ketiga pembagian data dengan proporsi 70% : 30%, eksperimen keempat pembagian data dengan proporsi 80% : 20%, dan eksperimen yang terakhir data yang menggunakan pembagian data dengan proporsi 90% : 10% data pengujian. Dibawah ini tabel 5. adalah parameter yang dihasilkan pada percobaan *Support Vector Machine*, tabel 6 . percobaan menggunakan *Decision Tree*, dan yang terakhir tabel 7. menunjukkan hasil *K-Nearest Neighbor*. Jika dilihat dari rata-rata akurasi yang dihasilkan dapat diketahui bahwa skenario eksperimen pertama dan kedua terbaik untuk algoritma *Decision Tree* dan eksperimen kedua terbaik untuk *K-Nearest Neighbor* dan *Support Vector Machine*. Hasil akurasi *Support Vector Machine* pada skenario kedua yaitu 91%, dan *K-Nearest Neighbor* yaitu 88%, untuk *Decision Tree* yaitu 86% pada skenario pertama dan kedua

Tabel 5. percobaan Support Vector Machine

Percobaan	Akurasi	Precision	Recall
90:10	90	90	89,5
80:20	91	92	89,5
70:30	83	83,5	81
60:40	82	82	80
50:50	85	86	81,5

Tabel 6. Percobaan Decision Tree

Percobaan	Akurasi	Precision	Recall
90:10	86	86,5	85,5
80:20	86	85	84
70:30	83	83	82
60:40	84	83,5	83
50:50	80	79,5	78

Tabel 7. percobaan K-Nearest Neighbor

Percobaan	Akurasi	Precision	Recall
90:10	85	85,5	83,5
80:20	88	88,5	86
70:30	86	85,5	85
60:40	84	84	83
50:50	82	83	80

Hasil prediksi pada tabel 8 menyatakan bahwa pada penelitian ini didapatkan hasil *accuracy* dari *Support Vector Machine* sebesar 91%, *K-Nearest Neighbor* sebesar 88%, dan *Decision Tree* sebesar 86%. Metode klasifikasi *Support Vector Machine* dapat menghasilkan nilai akurasi yang paling tinggi dibandingkan dengan dua metode lainnya dalam analisis sentimen mengenai larangan mudik. Dan hasil dari prediksi yang baik didapatkan dari data yang berkualitas

Tabel 8. Hasil Perbandingan akurasi

Algoritma	Accuracy	Precision	Recall
SVM	91%	92%	89,5%
KNN	88%	88%	87%
DT	86%	84,5%	83,5%

Langkah berikutnya setelah memvalidasi hasil yaitu melakukan validasi silang menggunakan *cross validation*. Dengan menggunakan jumlah lipatan *k-fold* yang berbeda yaitu $k = 5$, $k = 10$, dan $k = 20$ lalu diterapkan pada *Support Vector Machine*, *K-Nearest Neighbor*, dan *Decision Tree*. Hasil eksperimen dapat dilihat pada tabel 9. Yang berarti bahwa dalam penelitian ini, *Support Vector Machine* menjadi algoritma yang memiliki kinerja terbaik dibandingkan dengan algoritma lainnya. Dan untuk penelitian ini sendiri *k-fold* yang digunakan peneliti ialah $k = 10$ karena hasil dari rata-rata akurasi keseluruhan yang dapat dikatakan baik.

Tabel 9. Hasil perbandingan Cross Validation

Algoritma	5 Fold	10 Fold	20 Fold
SVM	85,41%	86,51%	86,94%
KNN	84,86%	85,90%	86,08%
DT	82,40%	84,49%	83,82%

Berikut ini merupakan hasil dari tabel *Confusion Matrix* yang menggunakan perbandingan tiga metode algoritma yaitu *Support Vector Machine*, *K-Nearest Neighbor*, dan *Decision Tree* terdapat pada tabel dibawah ini.

Tabel 5. Hasil *Confusion Matrix*

Algoritma	TP	TN	FP	FN
SVM	128	244	7	29

KNN	123	238	18	29
DT	108	241	26	33

4. KESIMPULAN

Pada sistem analisis sentimen tentang larangan mudik pada media sosial menggunakan algoritma klasifikasi *Support Vector Machine*, *K-Nearest Neighbour* dan *Decision Tree* dengan menggunakan metode *cross validation*, dan data 1212 review positif dan 827 review negatif serta. Hasil yang diperoleh pada metode *Support Vector Machine* berupa akurasi sebesar 91%, *precision* 92%, dan *recall* 89,5%. Untuk metode *K-Nearest Neighbor* memperoleh hasil akurasi sebesar 88%, *precision* 88%, dan *recall* 87%. Dan yang terakhir yaitu *Decision Tree* mendapatkan hasil akurasi 86%, *precision* 84,5%, dan *recall* 83,5%. Dari hasil perbandingan metode dapat disimpulkan bahwa *Support Vector Machine* yang memperoleh hasil yang terbaik dari algoritma klasifikasi lainnya. Penelitian ini menunjukkan bahwa menggunakan algoritma *Support Vector Machine*, *K-Nearest Neighbor* dan *Decision Tree* dapat menghasilkan klasifikasi terhadap data larangan mudik yang berasal dari media sosial.

Saran untuk penelitian selanjutnya, untuk pengembangan studi kedepannya diharapkan dapat dilakukan penambahan data agar tingkat akurasi lebih meningkat. Selain itu pada bagian pra pemrosesan data, proses stemming bisa menggunakan pendekatan lain selain algoritma sastrawi yang digunakan.

DAFTAR PUSTAKA

- [1] T. Wang, K. Lu, K. P. Chow, and Q. Zhu, "COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model," *IEEE Access*, vol. 8, pp. 138162–138169, 2020, doi: 10.1109/ACCESS.2020.3012595.
- [2] Z. Wang, Q. Liu, and Q. Dou, "Contrastive Cross-Site Learning with Redesigned Net for COVID-19 CT Classification," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 10, pp. 2806–2813, 2020, doi: 10.1109/JBHI.2020.3023246.
- [3] A. H. Arribathi and Q. Aini, "Mudik Dalam Perspektif Budaya Dan Agama (Kajian Realistis Perilaku Sumber Daya Manusia)," *Cices*, vol. 4, no. 1, pp. 45–52, 2018, doi: 10.33050/cices.v4i1.475.
- [4] Rudianto and Sopyan, "(Journal of Informatics and Telecommunication Engineering)," vol. 2, no. 2, pp. 62–68, 2019.
- [5] E. 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.
- [6] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *Integer J. Maret*, vol. 1, no. 1, pp. 32–41, 2017.
- [7] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel," *J. Khatulistiwa Inform.*, vol. VI, no. 1, pp. 64–69, 2018.
- [8] D. J. Haryanto, L. Muflikhah, and M. A. Fauzi, "Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2909–2916, 2018.
- [9] W. E. Nurjanah, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Terhadap

- Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1750–1757, 2017, doi: 10.1074/jbc.M209498200.
- [10] 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*, pp. 1–89, 2019.
- [11] A. P. Kirilenko, S. O. Stepchenkova, H. Kim, and X. (Robert) Li, “Automated Sentiment Analysis in Tourism: Comparison of Approaches,” *J. Travel Res.*, vol. 57, no. 8, pp. 1012–1025, 2018, doi: 10.1177/0047287517729757.
- [12] A. Al-Saffar, S. Awang, H. Tao, N. Omar, W. Al-Saiagh, and M. Al-bared, “Malay sentiment analysis based on combined classification approaches and Senti-lexicon algorithm,” *PLoS One*, vol. 13, no. 4, pp. 1–18, 2018, doi: 10.1371/journal.pone.0194852.
- [13] Suhardjono, G. Wijaya, and H. Abdul, “Prediksi Waktu Kelulusan Mahasiswa Menggunakan Svm Berbasis Pso,” *Bianglala Inform.*, vol. 7, no. 2, pp. 97–101, 2019.
- [14] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1725–1732, 2017.
- [15] R. Sari, “Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn),” *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 10–17, 2020, doi: 10.31294/evolusi.v8i1.7371.
- [16] M. Riefky and W. Pramesti, “Sentiment Analysis of Southeast Asian Games (SEA Games) in Philippines 2019 Based on Opinion of Internet User of Social Media Twitter with K-Nearest Neighbor and Support Vector Machine,” *J. Mat. Stat. dan Komputasi*, vol. 17, no. 1, pp. 26–41, 2020, doi: 10.20956/jmsk.v17i1.9947.
- [17] N. T. Romadloni, I. Santoso, and S. Budilaksono, “Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl,” *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [18] D. Setiawati, I. Taufik, J. Jumadi, and W. B. Zulfikar, “Klasifikasi Terjemahan Ayat Al-Quran Tentang Ilmu Sains Menggunakan Algoritma Decision Tree Berbasis Mobile,” *J. Online Inform.*, vol. 1, no. 1, p. 24, 2016, doi: 10.15575/join.v1i1.7.
- [19] P. Antinasari, R. S. Perdana, and M. A. Fauzi, “Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1733–1741, 2017.
- [20] E. Indrayuni, “Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes,” *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 29–36, 2019, doi: 10.31294/jki.v7i1.1.
- [21] A. Nugroho, “Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstraksi Fitur N-Gram,” *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 2, no. 2, p. 200, 2018, doi: 10.30645/j-sakti.v2i2.83.
- [22] 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, doi: 10.26418/jp.v4i2.27526.
- [23] G. A. Buntoro, “Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine,” *J. Din. Inform.*, vol. 5, no. September, pp. 11–40, 2016.
- [24] andi nurul Hidayat, “Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine Dan Naive Bayes,” *J. Elektron. Sistim Inf. Dan Komput.*, vol. 1, no. 1, pp. 1–7, 2015.



KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal yang telah terlampir sebelumnya dengan judul “Perbandingan Klasifikasi Analisis Sentimen Media Sosial Mengenai Larangan Mudik Di Masa Pandemi 2020”. 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, 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 source code yang digunakan pada penelitian ini. Bagian III menjelaskan mengenai dataset yang digunakan, meliputi : cara memperoleh data, sumber data yang digunakan, pelabelan data, contoh beberapa sampel yang berasal dari dataset, lalu data dapat diolah. Bagian IV memuat tahapan eksperimen yang disajikan dalam gambar beserta penjelasan dari tiap tahapan. Bagian V merupakan bagian terakhir dari kertas kerja ini yang menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan, meliputi penjelasannya.

UNIVERSITAS
MERCU BUANA