



**ANALISA SENTIMEN MENGENAI PENDIDIKAN TATAP
MUKA PADA TWITTER MENGGUNAKAN METODE
NAÏVE BAYES DAN SUPPORT VECTOR MACHINE**

TUGAS AKHIR

UNIVERSITAS Fadhil Muhammad S
41518010044
MERCU BUANA

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



**ANALISA SENTIMEN MENGENAI PENDIDIKAN TATAP
MUKA PADA TWITTER MENGGUNAKAN METODE
NAÏVE BAYES DAN SUPPORT VECTOR MACHINE**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

UNIVERSITAS
MERCU BUANA

Oleh:
Fadhil Muhammad

41518010044

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA

2022

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518010044

Nama : Fadhl Muhammad

Judul Tugas Akhir : Analisa Sentimen Mengenai Pendidikan Tatap Muka Pada
Twitter Menggunakan Metode 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, 27 Juli 2022



Fadhl Muhammad

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Fadhl Muhammad
NIM : 41518010044
Judul Tugas Akhir : Analisa Sentimen Mengenai Pendidikan Tatap Muka Pada Twitter Menggunakan Metode 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, 27 Juli 2022



Fadhl Muhammad

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Fadhl Muhammad
 NIM : 41518010044
 Judul Tugas Akhir : Analisa Sentimen Mengenai Pendidikan Tatap Muka 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		Diajukan	v
		Jurnal Nasional Terakreditasi	v		
		Jurnal International Tidak Bereputasi		Diterima	
		Jurnal International Bereputasi			
Disubmit/dipublikasikan di :	Nama Jurnal	: Jurnal Informatika: Jurnal Pengembangan IT			
	ISSN	: 2477-5126			
	Link Jurnal	: https://ejournal.poltektegal.ac.id/index.php/informatika			
	Link File Jurnal Jika Sudah di Publish	:			

1. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
2. 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

Jakarta, 27 Juli 2022

Dosen Pembimbing TA

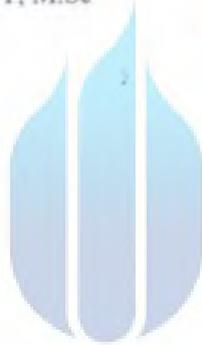
o.n.



Leonard Goeirianto, Dr., ST, M.Sc



Fadhl Muhammad



UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010044
Nama : Fadhl Muhammad
Judul Tugas Akhir : Analisa Sentimen Mengenai Pendidikan Tatap Muka Pada Twitter Menggunakan Metode 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, 27 Juli 2022



LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010044
Nama : Fadhl Muhammad
Judul Tugas Akhir : Analisa Sentimen Mengenai Pendidikan Tatap Muka Pada Twitter Menggunakan Metode 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, 27 Juli 2022



UNIVERSITAS
MERCUBUANA

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

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010044
Nama : Fadhl Muhammad
Judul Tugas Akhir : Analisa Sentimen Mengenai Pendidikan Tatap Muka Pada Twitter Menggunakan Metode 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, 27 Juli 2022



UNIVERSITAS
MERCU BUANA



(Anis Cherid, SE, MTI)

LEMBAR PENGESAHAN
LEMBAR PENGESAHAN

NIM : 41518010044
Nama : Fadhl Muhammad
Judul Tugas Akhir : Analisa Sentimen Mengenai Pendidikan Tatap Muka Pada
Twitter Menggunakan Metode 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, 27 Juli 2022

a.n. Menyetujui,

UNIVERSITAS
MERCU BUANA
(Leonard Goeirianto, Dr., ST, M.Sc)
Dosen Pembimbing

Mengetahui,



(Wawan Gunawan, S.Kom., MT)

Koord. Tugas Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph.D., IPM.)

Ka. Prodi Teknik Informatika

KATA PENGANTAR

Terimakasih dan rasa syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, dikarenakan Tugas Akhir yang berjudul “*Sentiment Analysis of Face-to-face education in Indonesia on Twitters*” 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 Dr. Harwikarya, MT selaku Dosen Pembimbing Akademik
3. Bapak Leonard Goermanto, Dr., ST, M.Sc 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, 6 Juli 2022

Fadhil Muhammad



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	vi
LEMBAR PENGESAHAN	ix
ABSTRAK	x
ABSTRACT	xi
KATA PENGANTAR	xii
DAFTAR ISI	xiv
NASKAH JURNAL	1
KERTAS KERJA	10
BAB 1. LITERATURE REVIEW	11
BAB 2. ANALISIS DAN PERANCANGAN	26
BAB 3. SOURCE CODE	30
BAB 4. DATASET	46
BAB 5. TAHAPAN EKSPERIMEN	48
BAB 6. HASIL SEMUA EKSPERIMEN	54
DAFTAR PUSTAKA	60
LAMPIRAN DOKUMEN HAKI	63
LAMPIRAN KORESPONDENSI	66

ANALISA SENTIMEN MENGENAI PENDIDIKAN TATAP MUKA PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

Fadhil Muhammad^{1*}, Leonard Goeirmanto²,

^{1,2}Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana, Jakarta

^{1,2}Jln. Raya, Meruya Sel., Kec. Kembangan, Jakarta, Daerah Khusus Ibukota Jakarta, 11650, Indonesia

email: ¹41518010044@student.mercubuana.ac.id, ²leonard@mercubuana.ac.id

Abstract – At the end of 2019, a new virus was found from Wuhan, China, namely the Corona virus (COVID-19). The virus has impacted all sectors of life, including the education sector. WHO or the World Health Organization explained that there was a decrease in the number of COVID-19 spread from January 3, 2020 to 18:37 CEST, June 21, 2022, there were 6,070,933 confirmed cases of COVID-19 with 156,700 deaths, as of June 14, 2022, a total of 417,522,347 vaccine doses had been administered. So that Indonesia stipulates Number 03 /KB/2021, Number 384 of 2021, Number HK.01.08 /MENKES/4242/2021, and Number 440-717 of 2021 concerning Guidelines for Implementing Learning During the Corona virus Disease 2019 (COVID-19) Pandemic, it is conveyed that learning in Higher Education starting from the odd semester of 2021/2022 is carried out face-to-face learning while still implementing health protocols or new normal. On the Twitter platform, there are often various kinds of public responses that they say about Face-to-Face Education in Indonesia that are negative and positive. In this study, an analysis of public opinion sentiment on Twitter was carried out on Face-to-Face Education using the Naïve Bayes Algorithm and Support Vector Machine using RapidMiner and Python tools. The experimental results showed that the Support Vector Machine algorithm was higher than the Naïve Bayes algorithm by providing the highest accuracy value at a Split percentage of 90:10, which was 93% with a K = 10 value of 90%.

Keywords: Education, Sentiment Analysis, Naïve Bayes, Support Vector Machine

Abstrak – Di penghujung tahun 2019, ditemukan virus baru dari Wuhan, China, yaitu virus Corona (COVID-19). Virus ini telah berdampak pada semua sektor kehidupan, termasuk sektor pendidikan. WHO atau Organisasi Kesehatan Dunia menjelaskan terjadi penurunan jumlah penyebaran COVID-19 dari 3 Januari 2020 menjadi 18:37 CEST, 21 Juni 2022, ada 6.070.933 kasus terkonfirmasi COVID-19 dengan 156.700 kematian, hingga 14 Juni 2022, total 417.522.347 dosis vaksin telah diberikan. Sehingga Indonesia menetapkan Nomor 03 /KB/2021, Nomor

384 Tahun 2021, Nomor HK.01.08 /MENKES/4242/2021, dan Nomor 440-717 Tahun 2021 tentang Pedoman Pelaksanaan Pembelajaran Pada Masa Pandemi Corona virus Disease 2019 (COVID-19), disampaikan bahwa pembelajaran di Perguruan Tinggi mulai semester ganjil 2021/2022 dilaksanakan pembelajaran tatap muka dengantetap melaksanakan protokol kesehatan atau newnormal. Pada platform Twitter, sering dijumpai berbagai macam tanggapan masyarakat yang mereka lontarkan mengenai Pendidikan Tatap Muka di Indonesia tersebut negatif maupun positif. Pada penelitian ini dilakukan analisa sentimen opini masyarakat di Twitter terhadap Pendidikan Tatap Muka menggunakan Algoritma Naïve Bayes dan Support Vector Machine dengan menggunakan tools RapidMiner dan Python. Hasil eksperimen menunjukkan bahwa algoritma Support Vector Machine lebih tinggi daripada algoritma Naïve Bayes dengan memberikan nilai akurasi paling tinggi pada percentage Split 90:10 yaitu 93% dengan nilai K=10 sebesar 90%.

Kata kunci:

Pendidikan, Sentimen, Naïve Bayes, Support Vector Machine

I. PENDAHULUAN

Di era sekarang Media Sosial salah satu sarana komunikasi yang dengan mudah berpartisipasi dan berbagi, media sosial juga salah satu sarana dan wadah untuk bertukar pikiran,beraspirasi dari internet ataupun weblog. Dalam hal ini, masyarakat mengkritik mengenai Pendidikan Tatap Muka. Analisis sentimen merupakan teknik mengekstrak data Text untuk mendapatkan informasi tentang sentimen bernilai positif maupun negatif. Analisis sentimen atau opinion mining adalah salah satu bagian dari text mining. 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.

Pada akhir tahun 2019 ditemukan sebuah virus baru yang berasal dari Wuhan, China, yaitu virus Corona

(COVID-19). Virus ini telah berdampak pada seluruh sektor kehidupan, tak terkecuali pada sektor Pendidikan. Penyebaran virus ini termasuk di Indonesia, membuat pemerintah Indonesia menekan penyebaran COVID-19 dengan menetapkan kebijakan PSBB atau Pembatasan Sosial Bersekala Besar sampai saat ini masa Transisi PPKM atau Pemberlakuan Pembatasan Kegiatan Masyarakat tingkat 3. Bidang yang terdampak dari kebijakan tersebut diantaranya adalah Pendidikan dan Perekonomian. WHO atau World Health Organization menjelaskan bahwa terjadi penurunan jumlah penyebaran COVID-19 dari 3 Januari 2020 tahun lalu hingga 21 Juni 2022, ditambah dengan total yang sudah divaksin berjumlah 417,522,347 manusia. Situasi saat ini terjadi di Indonesia, membuat pemerintahan Indonesia mengambil tindakan berdasarkan hasil keputusan bersama Menteri Pendidikan dan Kebudayaan, Menteri Agama, dan Menteri Kesehatan, dan Menteri Dalam Negeri Republik Indonesia Nomor 03/KB/2021, Nomor 384 Tahun 2021, Nomor HK.01.08/MENKES/4242/2021, dan Nomor 440-717 Tahun 2021 tentang Panduan Penyelenggaraan Pembelajaran di Masa Pandemi Corona virus Disease 2019 (COVID-19), dengan ini disampaikan bahwa pembelajaran di perguruan tinggi mulai semester gasal tahun akademik 2021/2022 diselenggarakan dengan pembelajaran tatap muka dilaksanakan melalui dua fase yaitu masa transisi atau terbatas dengan tetap menerapkan protokol kesehatan dan masa kebiasaan baru atau new normal. Dalam penyelenggaraan pembelajaran, perguruan tinggi harus tetap memprioritaskan kesehatan dan keselamatan warga kampus meliputi mahasiswa, dosen, tenaga kependidikan serta masyarakat sekitarnya.

Pada penelitian sebelumnya yang berjudul "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes" bahwa penelitian dilakukan dengan penambangan Text berbasis dokumen pada Twitter yang dianalisis menggunakan algoritma Naïve Bayes. Temuan menunjukkan bahwa pembelajaran daring memiliki 30% sentimen positif, 69% sentimen negatif, dan 1% netral pada periode tersebut. Tingginya sentimen Negatif dihasilkan karena ketidakpuasan masyarakat terhadap pembelajaran daring. Beberapa tweet menunjukkan kekecewaan dengan kata 'stres' dan 'malas' merupakan kata yang memiliki frekuensi tinggi dalam percakapan[1]. Pada penelitian lainnya yang berjudul "Analisis Sentimen Terhadap Pembelajaran Daring Di Indonesia Menggunakan Support Vector Machine (SVM)" melakukan analisis sentimen terhadap opini publik mengenai pembelajaran daring di Indonesia untuk memberikan informasi atau evaluasi terhadap opini publik pada media sosial twitter. Analisis sentimen dapat dilakukan dengan mengklasifikasi opini publik menjadi opini positif dan opini Negatif dengan metode Support Vector Machine (SVM). Dalam mengklasifikasikan data dapat dilakukan pelabelan data dan pembersihan data terlebih dahulu sebelum melalui proses Text

Preprocessing, kemudian data diberikan bobot setiap kata dengan Term Frequency-Invers Document Frequency (TF-IDF) yang akan dijadikan sebagai fitur setelah itu pembagian data menggunakan 10-Fold Cross Validation dan diklasifikasikan dengan metode Support Vector Machine (SVM). Hasil rata-rata evaluasi dengan confusion matrix yaitu accuracy sebesar 0,72[2].

Dari tabel berikut disimpulkan bahwa eksperimen dengan menggunakan Labeling manual lebih representatif dalam menentukan kelas Data. Model percentage split 90:10 memiliki nilai akurasi terbaik pada Data yang diuji Pendidikan Tatap Muka dengan nilai akurasi 93% pada eksperimen Labeling manual dan 87% pada eksperimen Labeling otomatis dan hasil perhitungan setelah validasi K-fold Cross Validation menunjukkan bahwa metode Support Vector Machine manual memiliki tingkat akurasi yang lebih tinggi sebesar 90% dengan rata-rata akurasi sebesar 90% yang paling baik pada penelitian ini.

Dari Permasalahan tersebut, maka dari itu penulis mengharapkan model yang telah dibuat untuk klasifikasi sentimen opini masyarakat di Twitter terhadap Pendidikan Tatap Muka agar dapat memperoleh hasil yang baik sehingga dapat diimplementasikan Pendidikan Tatap Muka di Indonesia.

Pada penelitian ini akan menggunakan algoritma Naïve Bayes dan Support Vector Machine, yang dimana algoritma ini adalah Naïve Bayes Classifier (NBC) merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan metode probabilitas dan statistic yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Menurut teorema Bayes, Naïve Bayes adalah Algoritma yang menggabungkan prior probability dengan probabilitas bersyarat pada rumus digunakan untuk menghitung kemungkinan tiap klasifikasi pada probabilitas. Naïve Bayes Classifier dalam melakukan klasifikasi terdapat dua proses penting yaitu learning (training) dan testing[3].

Sedangkan algoritma baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Penelitian ini berpusat pada klasifikasi biner yang terdiri dua kelas, kelas tersebut adalah Positif atau Negatif.

II. PENELITIAN YANG TERKAIT

A. M. Pravina, I. Cholissodin, and P. P. Adikara[4]. Pada penelitian ini dilakukan klasifikasi sentimen dengan fitur Lexicon Based yang dapat menerima opini berbahasa lain selain Bahasa Indonesia (dalam penelitian ini digunakan Bahasa Inggris) untuk melakukan analisis sentimen. Digunakan algoritma support vector machine untuk melakukan klasifikasi. 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%.

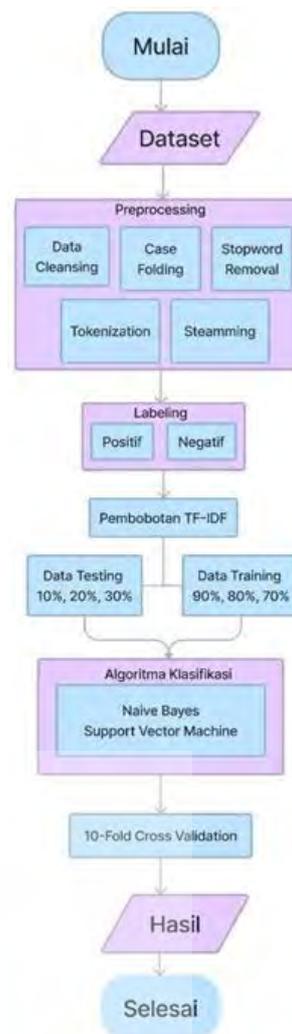
A. Hutapea and M. T. Furqon[5]. Pengecekan sistem terdiri dari pengecekan efek nilai K dan pengecekan efek nilai K Fold. Hasil pengujian pengaruh nilai K menunjukkan akurasi optimum sebesar 37,045% untuk K= 7 dan K-Fold = 10. Hasil pengujian pengaruh nilai K- Fold menunjukkan akurasi tertinggi sebesar 28,4462% Pada K-Fold = 5.

Rian Ardianto, Tri Rivanie, Yuris Alkhalifi, Fitra Septia Nugraha, Windu Gata[6]. Perbandingan tingkat akurasi, presisi dan recall antara Naïve Bayes dengan Support Vector Machine menggunakan metode SMOTE. Perbedaan akurasi antara Naïve Bayes dan Support Vector Machine dapat diamati sebesar 3,4%.

Ravinder Ahuja, Aakarsha Chug, Shruti Kohli, Shaurya Gupta, Pratyush Ahuja[7]. Dalam penelitian ini penulis menganalisis dampak dari dua fitur kata TF-IDF level dan, N-Gram pada dataset SS-Tweet dari analisis sentimen. Penulis menemukan bahwa dengan menggunakan tingkat kata TF-IDF (Term Frequency-Inverse Document Frequency) kinerja analisis sentimen adalah 3-4% lebih tinggi daripada menggunakan fitur N-gram, analisis dilakukan menggunakan enam algoritma klasifikasi (Decision Tree, Support vector Machine, K-Nearest Neighbour, Random Forest, Logistic Regression, Naive Bayes) dan mempertimbangkan parameter performa F-Score, Accuracy, Precision, dan Recall.

III. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif yang dimana Dataset yang telah terkumpul diolah melalui Twitter API menggunakan tools Rapid Miner. Dalam penelitian ini, Python dan r tools dibantu untuk melakukan Preprocessing dan Labeling pada Google Colab. Pada penelitian ini juga dilakukan eksperimen dengan membandingkan kinerja algoritma klasifikasi Naïve Bayes dan Support Vector Machine dan juga K-Fold Validation untuk mengetahui algoritma mana yang memiliki tingkat akurasi terbaik. Tahapan penelitian diilustrasikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Tahapan pertama yang dilakukan adalah mengumpulkan dataset. Selanjutnya pada tahapan berikutnya dilakukan tahapan Preprocessing untuk membersihkan dataset mentah. Setelah data selesai dilakukan Preprocessing, tahapan berikutnya adalah melakukan Labeling untuk mengetahui kelas positif dan negatif. Setelah itu dilakukan tahapan pembobotan TF-IDF untuk memberikan bobot pada setiap kata yang terdapat di dalam dokumen. Selanjutnya, dilakukan data testing dan data training untuk menghasilkan 3 split. Selanjutnya, dilakukan metode klasifikasi dengan menggunakan algoritma Naïve Bayes, Support Vector Machine dan juga K-Fold Cross Validation.

Sebelum tahap klasifikasi dilakukan, data terlebih dahulu dibagi menjadi data latih dan data uji. Data latih tersebut digunakan untuk proses pembelajaran pada algoritma Naïve Bayes dan Support Vector Machine. Langkah selanjutnya adalah menggunakan data uji untuk proses pengujian model klasifikasi Naïve Bayes dan Support Vector Machine. Pada penelitian ini digunakan algoritma Support Vector Machine karena cocok untuk klasifikasi Text dan cara kerja algoritma yang dapat mengatasi data outlier. Sedangkan algoritma Naïve Bayes cocok untuk data Text karena Naïve Bayes menggunakan metode probabilitas, peluang kata muncul.

A. Dataset

Data yang digunakan merupakan data yang didapat melalui media social Twitter. Total keseluruhan data yang digunakan sebanyak 1406 studi kasus Pendidikan Tatap Muka. Seluruh data diperoleh dengan dibantu menggunakan tools Rapid Miner. Proses pengambilan data dengan memasukkan kata kunci yang relevan dengan topik yang dibahas pada studi kasus, seperti yang ditunjukkan pada Tabel 1. Data yang dikumpulkan disimpan dalam format xlsx. Sumber Data diilustrasikan pada TABEL I.

TABEL I. Sumber Data

No	Query Twitter	Jumlah Data
1	Pendidikan Tatap Muka	9
2	Pendidikan Di Indonesia	174
3	Masuk Sekolah	359
4	Masuk Kuliah	562
5	Indonesia Belajar Tatap Muka	4
6	Belajar Offline	80
7	Bebas Pandemi	38
Total Data		1406

B. Preprocessing data

Preprocessing data bertujuan untuk membersihkan dataset dari data mentah menjadi data yang siap pakai agar memudahkan proses klasifikasi menjadi positif dan negatif. pre-processing data adalah teknik data mining yang melibatkan transformasi data mentah menjadi format yang mudah dimengerti. Pada penelitian ini digunakan lima langkah dalam tahapan Preprocessing. Berikut adalah penjelasan lima langkah Preprocessing:

- **Cleansing Data:** Cleansing dalam penelitian ini bertujuan untuk menghapus RT, username, hashtag, angka dan URL pada tweets.

Contoh tweet: RT @Kemenag_RI: Salam #SahabatReligi, Universitas Islam Internasional Indonesia (UIII) kembali membuka peluang bagi kamu yang ingin melanjutkan

Hasil Cleansing: Salam Universitas Islam Internasional Indonesia (UIII) kembalimembuka peluang bagi kamu yang ingin melanjutkan

- **Case Folding:** Case Folding dalam penelitian ini bertujuan untuk merubah Text dari penulisan huruf besar menjadi penulisan huruf kecil. Hasil Case Folding: salam universitas islam internasional indonesia (uiii) kembali membuka peluang bagi kamu yang ingin melanjutkan

- **Tokenization:** Tokenization dalam penelitian ini bertujuan untuk memisahkan kata kata dari dalam kalimat.

Hasil Tokenization: salam, universitas, islam, internasional, indonesia, (uiii), kembali, membuka, peluang, bagi, kamu, yang, ingin, melanjutkan,

- **Stopword Removal:** Bertujuan untuk menghapus kata sambung[8].

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[9].

Hasil Stopword Removal: salam, universitas, islam, internasional, indonesia, (uiii), kembali, membuka, peluang, bagi, kamu, yang, ingin, melanjutkan,

- **Stemming:** Stemming dalam penelitian ini bertujuan untuk membentuk kata kata dasar.

Hasil Stemming: salam, universitas, islam, internasional, indonesia, kembali, membuka, peluang, bagi, kamu, yang, ingin, melanjutkan,

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

Setelah melakukan tahap Preprocessing, data yang telah diPreprocessing menjadi 1375 untuk data Pendidikan Tatap Muka. Berikut merupakan hasil data setelah dilakukan tahapan Preprocessing. Preprocessing Pendidikan Tatap Muka dengan Labeling diilustrasikan pada Gambar 2 dan Preprocessing Pendidikan Tatap Muka tanpa Labeling diilustrasikan pada Gambar 3.

Klasifikasi	Tweet
868 positif	a whole comedy banget ma zel ambil nonton jam ...
1264 negatif	aaa gue ngonser udah bebas masker dongg
297 negatif	aaa males masuk sekolah
580 positif	aamiin nderr moga cape lelah bayar lolos sbm b...
292 positif	abis belanja siap anabul rumah beli pas selesa...
...	...
775 positif	yuk mari masuk kuliah paper streaming
346 positif	zaman sekolah minat lansung nak ajar maybe i w...
371 positif	zaman sekolah minat lansung nak ajar maybe i w...
827 positif	zaman suharto masuk kuliah ui gampang nyari ke...
688 negatif	zona nyaman nder explore semester awal tuh fas...

1375 rows x 2 columns

Gambar 2. Preprocessing Pendidikan Tatap

Muka dengan Labeling

	tweet_clean	Tweet
868	[a whole, comedy, banget, ma, zel, ambil, non...]	a whole comedy banget ma zel ambil nonton jam...
1264	[aaa, gue, ngonser, udah, bebas, masker, dongg]	aaa gue ngonser udah bebas masker dongg
297	[aaa, males, masuk, sekolah]	aaa males masuk sekolah
580	[aamiin, nderr, moga, cape, lelah, bayar, lolol...]	aamiin nderr moga cape lelah bayar lolol sbm b...
292	[abis, belanja, siap, anabul, rumah, beli, pas...]	abis belanja siap anabul rumah beli pas selesa...
...		
775	[yuk, mari, masuk, kuliah, paper, streaming]	yuk mari masuk kuliah paper streaming
346	[zaman, sekolah, minat, langsung, nak, ajar, ma...]	zaman sekolah minat langsung nak ajar maybe i w...
371	[zaman, sekolah, minat, langsung, nak, ajar, ma...]	zaman sekolah minat langsung nak ajar maybe i w...
827	[zaman, suharlo, masuk, kuliah, ui, gampang, n...]	zaman suharlo masuk kuliah ui gampang nyari ke...
688	[zona, nyaman, nder, explore, semester, awal...]	zona nyaman nder explore semester awal tuh fas...

Gambar 3. Preprocessing Pendidikan Tatap Muka tanpa Labeling

C. Labeling

Pada tahap ini, data yang dibersihkan selama Preprocessing disebut kelas data. Pengidentifikasi kelas data terbagi dalam dua kategori: kelas positif dan kelas negatif. Proses pelabelan dilakukan menggunakan bahasa R melalui alat Google Colab. Proses pelabelan dengan menggunakan bahasa R 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.

Proses pelabelan bahasa R membutuhkan kamus bahasa Indonesia yang terpasang di Google Drive. Kamus bahasa Indonesia yang digunakan terdiri dari dua kamus, kamus kata positif bahasa Indonesia dan kamus kata negatif bahasa Indonesia. Pada fase penandaan, setiap set diberi skor. Setiap kata yang dikenali diberi skor untuk menilai kelas emosional. Kata positif diberi nilai 1, kata negatif diberi nilai -1. Jika ada kata-kata yang tidak ada dalam kamus positif atau negatif, skor 0 diberikan. Skoring dilakukan dengan menghitung jumlah nilai setiap kata dalam kalimat. Jika nilainya ≥ 0 , ini ditandai sebagai emosi tweet positif. Jika tidak, nilai < 0 , yang ditandai sebagai sentimen tweet negatif. Jika skornya 0, itu akan ditandai sebagai tweet positif. Kelas data positif diberi nilai 1 dan kelas data negatif diberi nilai 0. Nilai Klasifikasi Otomatis dan Manual diilustrasikan pada TABEL II.

TABEL II. Nilai Klasifikasi Otomatis dan Manual

Klasifikasi	Positif	Negatif
Otomatis	1096	279
Manual	1214	161

Pada penelitian ini juga dilakukan eksperimen dengan menggunakan Labeling manual terhadap Dataset tersebut. Label positif merupakan tweet yang memiliki pujian sedangkan tweet negatif berupa kata-kata kasar dan sarkas.

D. Pembobotan TF-IDF

Mengikuti seberapa penting sebuah kata mewakili sebuah kalimat, dilakukan pembobotan atau perhitungan. Pemberian skor dalam TF-IDF berdasarkan frekuensi munculnya kata dalam dokumen[10].

E. Klasifikasi Algoritma

Text mining adalah salah satu subjek data mining, karena text juga adalah data. Namun karena cakupannya sangat luas sehingga penggunaan istilah text mining dan data testing dibedakan. Data mining menambang pengetahuan dari kumpulan data yang banyak dan biasanya terstruktur. Perbedaannya terdapat dari tugas data miningnya, seperti klasifikasi, klusterisasi, asosiasi, estimasi atau prediksi, bukan pada karakteristik data yang diolahnya. Text mining merupakan proses menambang data yang berupa Text dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen[11].

Analisis Sentimen adalah kegiatan melakukan analisis terhadap pendapat, opini, sikap atau emosi seseorang mengenai suatu produk, topik atau permasalahan tertentu sehingga bisa diketahui hal tersebut masuk kedalam sentimen positif, negatif[12].

Metode Klasifikasi adalah teknik data mining yang mengelompokkan data berdasarkan lampiran data pada data sampel. Metode klasifikasi juga merupakan proses mengklasifikasikan dokumen ke dalam satu atau lebih kategori tertentu atau kelas dokumen yang sama[13].

Metode Bayes merupakan pendekatan statistic untuk melakukan inferensi induksi pada persoalan klasifikasi. Pertama kali dibahas terlebih dahulu tentang konsep dasar dan definisi pada Teorema Bayes, kemudian menggunakan teorema ini untuk melakukan klasifikasi dalam Data mining[14]. Pengklasifikasi Naive Bayes dalam klasifikasi memiliki dua proses penting: pelatihan dan pengujian.[3].

Support Vector Machine (SVM) adalah metode yang relatif baru untuk membuat prediksi baik dari segi klasifikasi maupun regresi. Penelitian ini berfokus pada klasifikasi biner, yang terdiri dari dua kelas. Kelasnya positif atau negatif. Konsep klasifikasi menggunakan Support Vector Machine adalah mencari hyperplane optimal yang bertindak sebagai pembatas data pada kedua kelas tersebut. Algoritma mesin vektor dukungan bekerja dengan memisahkan data secara optimal dari titik terdekat menggunakan pengukuran jarak hyperplane daripada mencari titik maksimum, untuk memaksimalkan jarak antar label kelas berdasarkan pengukuran batas keanggotaan kelas[15].

K-Fold Cross Validation Menurut Rohani, Abbas., et al. (dalam Jiang, Ping., 2017) K-Fold Cross Validation dirancang untuk mengevaluasi kinerja proses metode algoritma dengan membagi sampel data secara acak dan mengelompokkan data K-Fold ke nilai K. Suatu jenis uji validasi silang. Kemudian salah satu kelompok

K-Fold digunakan sebagai data uji dan kelompok lainnya digunakan sebagai data latih.

F. Validasi

Validasi model dilakukan untuk memeriksa *overfitting* pada model klasifikasi. Metode validasi yang dipakai pada penelitian ini adalah *K-Fold Cross Validation*. Pada proses ini seluruh *Dataset* akan diacak (*shuffle*) lalu dibagi menjadi k bagian sama besar, di mana 1 bagian akan menjadi data uji, sedangkan sisanya akan menjadi data latih. Pada *10-Fold Cross Validation*, 1 bagian data akan menjadi data latih dan 9 bagian data lainnya akan menjadi data uji. Contohnya pada *Dataset* penelitian ini terdapat 1355 data akan dibagi ke dalam *10-fold*. *Fold* ke-1 akan menjadi data uji sedangkan sisanya menjadi data latih. Setelah *training* tahap pertama selesai selanjutnya *training* dilakukan dengan menjadikan *fold* ke-2 sebagai data uji dan 9 data bagian lainnya menjadi data latih. Seperti itu seterusnya sampai proses selesai [16].

IV. HASIL DAN PEMBAHASAN

Pada bagian ini akan menampilkan hasil semua

eksperimen yang telah dilakukan. Pada penelitian ini melakukan eksperimen pada studi kasus yaitu Pendidikan Tatap Muka. Eksperimen dilakukan dengan menggunakan dua algoritma yaitu Support Vector Machine dan Naïve Bayes. Terdapat tiga skenario yang menjadi pembagian persentase data dan juga 10-Fold Cross Validation. Pembagian presentasi data merupakan pemisahan data training dan testing berdasarkan persentasenya, contoh 90%:10% berarti 90% merupakan data training dan 10% merupakan data testing. Pada penelitian ini pembagian persentase dibagi menjadi tiga skenario eksperimen yaitu eksperimen pertama menggunakan pembagian data 90%:10%, eksperimen kedua menggunakan pembagian data 80%:20% dan eksperimen yang ketiga menggunakan pembagian data 70%:30%. Penelitian ini juga membandingkan nilai akurasi yang dilakukan dengan menggunakan metode Labeling manual dan metode Labeling otomatis.

Hasil dari kinerja model dengan menggunakan metode Labeling otomatis dan pembagian data atau percentage Split 90%:10% bahwa studi kasus Pendidikan Tatap Muka memiliki nilai akurasi 86% untuk algoritma Support Vector Machine pada TABEL III dan 86% untuk algoritma Naïve Bayes pada TABEL IV.

TABEL III. Support Vector Machine Pendidikan Tatap Muka Split 90:10

Labeling Otomatis			
	Precision	Recall	Accuracy
0	83%	22%	86%
1	86%	99%	86%

TABEL IV. Naïve Bayes Pendidikan Tatap Muka Split 90:10 Labeling Otomatis

	Precision	Recall	Accuracy
0	100%	17%	86%
1	86%	100%	86%

metode Labeling otomatis dan pembagian data atau percentage Split 80%:20% bahwa studi kasus Pendidikan Tatap Muka memiliki nilai akurasi 86% untuk algoritma Support Vector Machine pada TABEL V dan 81% untuk algoritma Naïve Bayes pada TABEL VI.

TABEL V. Support Vector Machine Pendidikan Tatap Muka Split 80:20 Labeling Otomatis

	Precision	Recall	Accuracy
0	93%	27%	86%
1	85%	100%	86%

TABEL VI. Naïve Bayes Pendidikan Tatap Muka Split 80:20 Labeling Otomatis

	Precision	Recall	Accuracy
0	50%	8%	81%
1	82%	98%	81%

Hasil dari kinerja model dengan menggunakan metode Labeling otomatis dan pembagian data atau percentage Split 70%:30% bahwa untuk studi kasus Indihome memiliki nilai akurasi 87% untuk algoritma Support Vector Machine pada TABEL VII dan 85% untuk algoritma Naïve Bayes pada TABEL VIII.

TABEL VII. Support Vector Machine Pendidikan Tatap Muka Split 70:30 Labeling Otomatis

	Precision	Recall	Accuracy
0	95%	26%	87%
1	87%	100%	87%

TABEL VIII. Naïve Bayes Pendidikan Tatap Muka Split 70:30 Labeling Otomatis

Hasil dari kinerja model dengan menggunakan

Otomatis

	Precision	Recall	Accuracy
0	82%	13%	85%
1	85%	99%	85%

Hasil dari kinerja model dengan menggunakan metode Labeling manual dan pembagian data atau percentage Split 90%:10% bahwa studi kasus Pendidikan Tatap Muka memiliki nilai akurasi 93% untuk algoritma Support Vector Machine pada TABEL IX dan 88% untuk algoritma Naïve Bayes pada TABEL X.

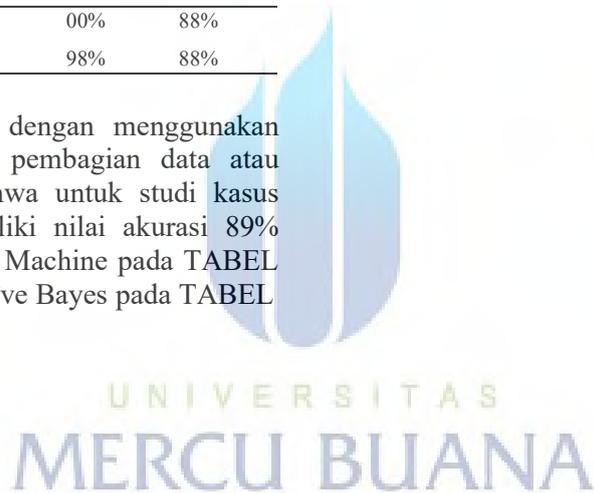
TABEL IX. Support Vector Machine Pendidikan Tatap Muka 90:10
Labeling Manual

	Precision	Recall	Accuracy
0	00%	00%	93%
1	93%	99%	93%

TABEL X. Naïve Bayes Pendidikan Tatap Muka 90:10 *Labeling Manual*

	Precision	Recall	Accuracy
0	00%	00%	88%
1	90%	98%	88%

Hasil dari kinerja model dengan menggunakan metode Labeling manual dan pembagian data atau percentage Split 80%:20% bahwa untuk studi kasus Pendidikan Tatap Muka memiliki nilai akurasi 89% untuk algoritma Support Vector Machine pada TABEL XI dan 88% untuk algoritma Naïve Bayes pada TABEL X.



XII.

TABEL XI. Support Vector Machine Pendidikan Tatap Muka 80:20

<i>Labeling Manual</i>			
	Precision	Recall	Accuracy
0	60%	10%	89%
1	90%	99%	89%

TABEL XII. Naïve Bayes Pendidikan Tatap Muka 80:20 *Labeling Manual*

	Precision	Recall	Accuracy
0	33%	3%	88%
1	88%	99%	88%

Hasil dari kinerja model dengan menggunakan metode Labeling manual dan pembagian data atau percentage Split 70%:30% bahwa untuk studi kasus Indihome memiliki nilai akurasi 89% untuk algoritma Support Vector Machine pada TABEL XIII dan 87% untuk algoritma Naïve Bayes pada TABEL XIV.

TABEL XIII. Support Vector Machine Pendidikan Tatap Muka 70:30 *Labeling Manual*

	Precision	Recall	Accuracy
0	20%	2%	89%
1	90%	99%	89%

TABEL XIV. Naïve Bayes Pendidikan Tatap Muka 70:30 *Labeling Manual*

	Precision	Recall	Accuracy
0	25%	4%	87%
1	88%	98%	87%

Hasil dari penelitian ini adalah eksperimen dengan menggunakan algoritma Support Vector Machine memiliki nilai akurasi paling tinggi pada tiga skenario eksperimen percentage Split pada model data yang diuji yaitu Pendidikan Tatap Muka dengan menggunakan Labeling Otomatis dan Labeling Manual. Hasil dari pembahasan terangkum dalam TABEL XV.

TABEL XV. Perbandingan Akurasi Labeling Otomatis dan Manual

	SVM		NB	
90:10	87%	83%	93%	89%
80:20	86%	81%	89%	88%
70:30	86%	86%	89%	87%

Hasil akurasi metode klasifikasi Naïve Bayes dengan validasi menggunakan metode 10-Fold Cross Validation otomatis pada TABEL XVI dan 10-Fold Cross Validation manual pada TABEL XVII.

TABEL XVI. Hasil perhitungan akurasi Naïve Bayes dengan *K-Fold Cross Validation*

Iterasi	Nilai Akurasi (dalam %)
1	82
2	82
3	87

4	80
5	80
6	80
7	78
8	82
9	82
10	82
Rata-rata	82.00

TABEL XVII. Hasil perhitungan akurasi Naïve Bayes manual dengan *K-Fold Cross Validation*

Iterasi	Nilai Akurasi (dalam %)
1	80
2	80
3	90
4	87
5	87
6	85
7	87
8	87
9	85
10	90
Rata-rata	86.00

Validasi pada model klasifikasi Naïve Bayes menggunakan dua model yaitu otomatis dan manual yang telah dibangun. Hasil validasi menunjukkan bahwa akurasi optimal berdasarkan metode K-Fold Cross Validation manual lebih tinggi dari akurasi otomatis, di mana model klasifikasi Naïve Bayes memiliki akurasi tertinggi pada iterasi ke-3 manual dengan nilai akurasi sebesar 90.47% dengan rata-rata akurasi sebesar 86,00%. Nilai akurasi metode klasifikasi Naïve Bayes menggunakan metode 10-Fold Cross Validation otomatis pada TABEL XVIII dan 10-Fold Cross Validation manual pada TABEL XIX.

TABEL XVIII. Hasil perhitungan akurasi Support Vector Machine dengan *K-Fold Cross Validation*

Iterasi	Nilai Akurasi (dalam %)
1	82
2	82
3	85
4	85
5	82
6	82
7	82
8	85
9	82
10	82
Rata-rata	84.00

TABEL XIX. Hasil perhitungan akurasi Support Vector Machine manual dengan *K-Fold Cross Validation*

Iterasi	Nilai Akurasi (dalam %)
1	88
2	88
3	88
4	90
5	90
6	90
7	90
8	90
9	90
10	90
Rata-rata	90.00

Validasi pada model klasifikasi Support Vector Machine menggunakan dua model yaitu otomatis dan manual yang telah dibangun. Hasil perhitungan setelah validasi menunjukkan bahwa metode Support Vector Machine manual memiliki tingkat akurasi yang lebih tinggi daripada metode Naïve Bayes otomatis, manual, Support Vector Machine otomatis dengan nilai akurasi Support Vector Machine manual sebesar 90.24% dengan rata-rata akurasi sebesar 90.00%. Sementara Naïve Bayes otomatis 82.00%, Naïve Bayes manual 86.00%, Support Vector Machine otomatis 84.00% setelah tervalidasi dengan Iterasi ke-4 s/d ke-10 sebagai akurasi tertinggi memiliki nilai sebesar 90.24% dengan rata-rata akurasi sebesar 90.00%.

V. KESIMPULAN

Dari hasil penelitian yang dilakukan dengan menggunakan algoritma Naïve Bayes, Support Vector Machine dengan pengujian dia labeling antara otomatis dan manual dapat disimpulkan bahwa eksperimen dengan menggunakan Labeling manual lebih representatif dalam menentukan kelas data. Sehingga menghasilkan model percentage split 90:10 memiliki nilai akurasi terbaik pada data yang diuji Pendidikan Tatap Muka dengan nilai akurasi 93% pada eksperimen Labeling manual dan 87% pada eksperimen Labeling otomatis dan hasil perhitungan setelah validasi K-fold Cross Validation dengan nilai K=10 menunjukkan bahwa metode Support Vector Machine manual memiliki tingkat akurasi yang lebih tinggi sebesar 90% dengan rata-rata akurasi sebesar 90% yang paling baik pada penelitian ini.

UCAPAN TERIMA KASIH

Penulis menyadari bahwa dalam penelitian ini tidak terlepas dari bantuan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada semua pihak yang telah membantu. Penulis berharap hasil dari penelitian ini dapat bermanfaat bagi pembaca, walaupun di dalam nya masih terdapat kekurangan dalam penulisan, maupun eksperimen.

DAFTAR PUSTAKA

- [1] W. A. Prabowo and C. Wiguna, "Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 1, p. 149, Jan. 2021, doi: 10.30865/mib.v5i1.2604.
- [2] A. N. Indrainsi, I. Ernawati, and A. Zaidah, "ANALISIS SENTIMEN TERHADAP PEMBELAJARAN DARING DI INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)."
- [3] M. Rasyida, "Naïve Bayes Classification untuk Penentuan Status Penduduk Miskin," *J. Inform. Kaputama(JIK)*, vol. 4, no. 2, 2020.
- [4] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," 2019. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [5] A. Hutapea and M. T. Furqon, "Penerapan Algoritme Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia," vol. 2, no. 10, pp. 3957–3961, 2018.
- [6] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. Septia Nugraha, and W. Gata, "SENTIMENT ANALYSIS ON E-SPORTS FOR EDUCATION CURRICULUM USING NAIVE BAYES AND SUPPORT VECTOR MACHINE," 2020.
- [7] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The impact of features extraction on the sentiment analysis," in *Procedia Computer Science*, 2019, vol. 152, pp. 341–348. doi: 10.1016/j.procs.2019.05.008.
- [8] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 3, p. 650, Jul. 2020, doi: 10.30865/mib.v4i3.2181.
- [9] A. Nugroho Sekolah Tinggi Teknologi Pelita Bangsa Jl Inspeksi Kalimantan, T. Danas, C. Pusat, and K. Bekasi, "Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstrasi Fitur N-Gram," 2018. [Online]. Available: <http://tunasbangsa.ac.id/ejurnal/index.php/jsakti>
- [10] A. Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes Muljono, D. Putri Artanti, A. Syukur, A. Prihandono, and D. I. Rosal Moses Setiadi, "Konferensi Nasional Sistem Informasi 2018 STMIK Atma Luhur Pangkalpinang," 2018. [Online]. Available: <http://twitter.com>

KERTAS KERJA

Ringkasan

Kertas kerja ini berisi tentang kelengkapan material dari artikel jurnal dengan judul “Analisa Sentimen Mengenai Pendidikan Tatap Muka Pada Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine”. Seluruh hasil penelitian Tugas Akhir yang tidak dimasukkan kedalam artikel jurnal. Pada kertas kerja ini disajikan: literatur *review*, analisis perancangan, *Source code*, *Dataset* yang digunakan, tahapan eksperimen dan seluruh hasil eksperimen.

- Bagian 1: *Literature review* menjabarkan mengenai beberapa jurnal yang terkait dengan penelitian.
- Bagian 2: Analisis dan perancangan menjabarkan analisis masalah dan analisis model.
- Bagian 3: *Source code* menjabarkan kumpulan kode di setiap prosesnya mulai dari Membaca data, *Preprocessing*, *Labeling*, Implementasi model, sampai pengujian dan evaluasi.
- Bagian 4: *Dataset* melampirkan pengambilan *dataset*.
- Bagian 5: Tahapan eksperimen berisi pengumpulan data, *Preprocessing*, *Labeling*, Implementasi model, dan evaluasi model.
- Bagian 6: Hasil implementasi eksperimen secara keseluruhan yang mencakup jawaban dari skenario percobaan pengujian yang sudah disebutkan pada bagian 5. Bagian ini berguna untuk menemukan model analisa sentimen terbaik yang dilihat dari beberapa pengujian berdasarkan opini masyarakat yang positif dan negatif. Kemudian, hasil yang didapat digunakan berupa akurasi. Pendidikan tatap muka di Indonesia untuk bahan evaluasi atas kebijakan yang diterapkan.