

**STUDI KOMPARASI ALGORITMA *NAIVE BAYES CLASSIFIER* DAN
SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN TWEET**

(Studi Kasus : Data Tweet Dengan Kata Kunci ‘psbb’ dan ‘pandemi’)

SKRIPSI

Untuk Memenuhi Sebagian Persyaratan Mencapai Derajat Sarjana S-1

Program Studi Teknik Informatika



Disusun oleh:

Ahmad Putra Awwalu Raafi’u

17106050023

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2021

HALAMAN PENGESAHAN



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1206/Un.02/DST/PP.00.9/07/2021

Tugas Akhir dengan judul : STUDI KOMPARASI ALGORITMA NAIVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN TWEET

yang dipersiapkan dan disusun oleh:

Nama : AHMAD PUTRA AWWALU RAAFTU
Nomor Induk Mahasiswa : 17106050023
Telah diujikan pada : Jumat, 16 Juli 2021
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Muhammad Didik Rohmad Wahyudi, S.T., MT.
SIGNED

Valid ID: 61077b17c59dc



Penguji I

Ir. Muhammad Taufiq Nurozzaman, S.T.
M.Eng., Ph.D.
SIGNED

Valid ID: 61071c87770d9



Penguji II

Dr. Agung Fatwanto, S.Si., M.Kom.
SIGNED

Valid ID: 61034dc3efc4



Yogyakarta, 16 Juli 2021

UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6108ced5a81b1

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR



Universitas Islam Negeri Sunan Kalijaga



FM-UINSK-BM-05-03/R0

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi
Lamp :

Kepada
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
di Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Ahmad Putra Awwalu Raafi'u
NIM : 17106050023
Judul Skripsi : Studi Komparasi Algoritma *Naive Bayes Classifier* Dan *Support*

Vector Machine Dalam Analisis Sentimen Tweet

sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika

Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu 'alaikum wr. wb.

Yogyakarta, 06 Juli 2021
Pembimbing

Muhammad Didik Rohmad Wahyudi, S.T., MT.
NIP. 19760812 200901 1 015

PERNYATAAN KEASLIAN SKRIPSI

PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini:

Nama : Ahmad Putra Awwalu Raafi'u
NIM : 17106050023
Program Studi : Teknik Informatika
Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi saya yang berjudul “Studi Komparasi Algoritma Naive Bayes Classifier Dan Support Vector Machine Dalam Analisis Sentimen Tweet” merupakan hasil penelitian saya sendiri, tidak terdapat pada karya yang pernah di ajukan untuk memperoleh gelar kesarjana di suatu perguruan tinggi, dan bukan plagiasi karya orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA
Yogyakarta, 05 Juli 2021



Ahmad Putra Awwalu Raafi'u
NIM. 17106050023

KATA PENGANTAR

Segala puji dan syukur kehadiran Allah SWT atas berkah, rahmat dan hidayah-Nya yang senantiasa dilimpahkan kepada penulis, sehingga penulisan skripsi dengan judul “Studi Komparasi Algoritma *Naive Bayes Classifier* Dan *Support Vector Machine* Dalam Analisis Sentimen Tweet (Studi Kasus : Data Tweet Dengan Kata Kunci ‘psbb’ dan ‘pandemi’)” dapat terselesaikan dengan baik. Untuk itu, dengan segala kerendahan hati pada kesempatan ini penulis mengucapkan banyak terimakasih kepada:

1. Bapak Prof. Dr.Phil. Al Makin, S.Ag., M.A., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Ibu Ir.Maria Ulfa Siregar, S.Kom., MIT., Ph.D., selaku Ketua Program Studi S1 Teknik Informatika dan Dosen Pembimbing Akademik selama masa perkuliahan.
4. Bapak Muhammad Didik Rohmad Wahyudi, S.T., MT., selaku Dosen Pembimbing yang telah sabar membimbing, mengarahkan, dan memberi nasihat selama penyusunan skripsi.
5. Bapak-Ibu Dosen Teknik Informatika UIN Sunan Kalijaga yang telah memberikan banyak ilmu bagi penulis.
6. Seluruh staf dan karyawan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.

7. Staf dan karyawan Perpustakaan UIN Sunan Kalijaga Yogyakarta yang telah membantu dalam pencarian referensi penulisan skripsi.
8. Kedua Orang tua dan keluarga yang telah memberikan doa dan dukungan selama proses pembuatan skripsi.
9. Teman-teman seperjuangan Teknik Informatika 2017 UIN Sunan Kalijaga Yogyakarta.
10. Teman-teman KKN Mandiri 102 Kelompok 196 Plosorejo.
11. Teman-teman TSWF, dan King\$man.
12. Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu memberikan dukungan.

Penulis menyadari masih banyak kekurangan dalam skripsi ini, sehingga kritik dan saran yang membangun dari pembaca sangat penulis harapkan .Semoga skripsi ini dapat memberikan manfaat untuk mendorong penelitian penelitian selanjutnya.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Yogyakarta, Juli 2021

Penyusun

Ahmad Putra Awwalu Raafi'u

HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan untuk almamater, kedua orang tua, keluarga, dan teman-teman seperjuangan yang selalu memberikan support dan semangat kepada penulis serta semua pihak yang turut andil dalam proses penulisan ini yang tidak dapat disebutkan satu persatu.

Terimakasih Semuanya.



HALAMAN MOTTO

“Yang penting semangat dulu aja”

-Jerbasuki Mowo Bea-

“Setiap keberhasilan ada harga yang dibayarkan. Untuk meraih kesuksesan dibutuhkan sebuah perjuangan dan pengorbanan. Tidak mungkin tanpa ada usaha yang keras tiba-tiba mendapatkan kemenangan.”



DAFTAR ISI

HALAMAN PENGESAHAN	ii
SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
KATA PENGANTAR.....	v
HALAMAN PERSEMBAHAN	vii
HALAMAN MOTTO	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
INTISARI	xv
ABSTRACT	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan.....	5
1.5 Manfaat Penelitian.....	5
1.6 Keaslian Penelitian	6
1.7 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	8
2.1 Tinjauan Pustaka	8
2.2 Landasan Teori	15

2.2.1	Machine Learning.....	15
2.2.2	Supervised Learning.....	16
2.2.3	Klasifikasi.....	17
2.2.4	Analisis Sentimen.....	17
2.2.5	Pembobotan TF-IDF.....	18
2.2.6	Naive Bayes Classifier	19
2.2.7	Support Vector Machine.....	22
2.2.8	K-Fold Cross Validation.....	25
2.2.9	Twitter	26
2.2.10	Python.....	28
BAB III METODE PENELITIAN		30
3.1	Metode Penelitian.....	30
3.2	Tahapan-Tahapan Penelitian	30
3.2.1	Studi Pustaka	31
3.2.2	Pengumpulan Data.....	31
3.2.3	Seleksi Data dan Pelabelan Data	31
3.2.4	Preprocessing Data	32
3.2.5	Analisa dan Validasi.....	32
3.2.6	Implementasi	32
BAB IV HASIL DAN PEMBAHASAN		33
4.1	Pengumpulan Data	33
4.2	Seleksi Data dan Pelabelan Data	35
4.2.1	Seleksi Data.....	35

4.2.2	Pelabelan Data	36
4.3	Preprocessing.....	38
4.3.1	Cleansing	38
4.3.2	Casefolding.....	39
4.3.3	Tokeninasi	41
4.3.4	Convert Slangword.....	42
4.3.5	Stopword Removal	45
4.3.6	Stemming.....	47
4.4	Analisa dan Evaluasi	48
4.4.1	Analisa.....	48
4.4.2	Evaluasi Model Klasifikasi.....	71
4.5	Implementasi	78
BAB V KESIMPULAN DAN SARAN		84
5.1	Kesimpulan.....	84
5.2	Saran.....	85
DAFTAR PUSTAKA		87
LAMPIRAN.....		91
CURRICULUM VITAE.....		92

DAFTAR GAMBAR

Gambar 2. 1 Hyperlane Support Vector Machine.....	23
Gambar 2. 2 Model 3-fold cross validation (Tempola et al., 2018).....	26
Gambar 3. 1 Tahapan Penelitian	30
Gambar 4. 1 Proses Analisis Sentimen	33
Gambar 4. 2 Grafik Nilai Akurasi Tiap K.....	73
Gambar 4. 3 Grafik 16-fold cross validation Naive Bayes	74
Gambar 4. 4 Grafik 16-fold cross validation SVM kernel Linear	76
Gambar 4. 5 Grafik 16-fold cross validation SVM kernel RBF	77
Gambar 4. 6 Hasil Implementasi Model Klasifikasi Pada Data Uji	79
Gambar 4. 7 Hasil Implementasi Model Klasifikasi Pada Data Uji	80
Gambar 4. 8 WordCloud Sentimen positif Klasifikasi Naive Bayes	81
Gambar 4. 9 WordCloud Sentimen negatif Klasifikasi Naive Bayes	81
Gambar 4. 10 WordCloud Sentimen netral Klasifikasi Naive Bayes	82
Gambar 4. 11 WordCloud Sentimen positif Klasifikasi SVM.....	82
Gambar 4. 12 WordCloud Sentimen negatif Klasifikasi SVM.....	83
Gambar 4. 13 WordCloud Sentimen netral Klasifikasi SVM.....	83

DAFTAR TABEL

Tabel 2. 1 Tinjauan Pustaka	12
Tabel 4. 1 Contoh data tweet	34
Tabel 4. 2 Contoh data tweet yang telah diberi label	36
Tabel 4. 3 Contoh cleansing pada data tweet	38
Tabel 4. 4 Contoh casefolding pada data tweet	40
Tabel 4. 5 Contoh tokenisasi pada data tweet	41
Tabel 4. 6 Contoh kamus slangword.....	43
Tabel 4. 7 Contoh convert slangword pada data tweet	43
Tabel 4. 8 Contoh stopwords removal pada data tweet	46
Tabel 4. 9 Contoh stemming pada data tweet	47
Tabel 4. 10 Contoh data tweet	48
Tabel 4. 11 Nilai TF dan IDF.....	50
Tabel 4. 12 Nilai TF-IDF	52
Tabel 4. 13 Hasil perhitungan probabilitas term positif.....	55
Tabel 4. 14 Hasil perhitungan probabilitas term negatif.....	56
Tabel 4. 15 Hasil perhitungan probabilitas term netral.....	58
Tabel 4. 16 Contoh Klasifikasi 3 Kelas Dengan OVR.....	61
Tabel 4. 17 Contoh Data Latih One Versus Rest.....	63
Tabel 4. 18 Contoh Dari Nilai X.....	63
Tabel 4. 19 Contoh Dari Nilai Y	64
Tabel 4. 20 Nilai X Setiap Dokumen	66
Tabel 4. 21 Contoh Nilai Y Setiap Dokumen	67

Tabel 4. 22 Nilai Support Vector Setiap Dokumen	67
Tabel 4. 23 Nilai Support Vector Bias	68
Tabel 4. 24 Jumlah Data Tiap Kelas Sentimen	71
Tabel 4. 25 Hasil K-Fold Cross Validation Tiap K	72
Tabel 4. 26 Hasil akurasi SVM dengan Kernel Linear	74
Tabel 4. 27 Hasil Akurasi SVM Dengan Kernel RBF	75
Tabel 4. 28 Perbandingan nilai Akurasi kedua metode.....	77
Tabel 5. 1 Perbandingan Hasil Akurasi	84



STUDI KOMPARASI ALGORITMA *NAIVE BAYES CLASSIFIER* DAN *SUPPORT VECTOR MACHINE* DALAM ANALISIS SENTIMEN TWEET

(Studi Kasus : Data Tweet Dengan Kata Kunci ‘psbb’ dan ‘pandemi’)

Ahmad Putra Awwalu Raafi’u

17106050023

INTISARI

Mengawali tahun 2020, dunia digemparkan dengan penemuan virus baru yang berasal dari Wuhan, Kota bagian tengah di China. Virus baru ini termasuk jenis coronavirus kemudian diberi nama *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-COV2) dan menyebabkan penyakit *Corona Virus Disease 2019* (COVID-19). Pemerintah Indonesia memberlakukan berbagai macam kebijakan publik untuk menekan tingkat penyebaran virus tersebut salah satunya dengan upaya memberlakukan Pembatasan Sosial Berskala Besar (PSBB). Setelah setahun pandemi di Indonesia dan pemberlakuan kebijakan PSBB tersebut, banyak pro dan kontra mengenai kebijakan PSBB itu dikalangan masyarakat.

Penelitian ini bertujuan untuk menganalisis terhadap sentimen atau opini yang berkembang di masyarakat terhadap kebijakan PSBB oleh pemerintah saat pandemi COVID-19 menggunakan metode *Naive Bayes Classifier* dan *Support Vector Machine*. Data yang digunakan sebanyak 4040 data tweet sebagai data latih dengan label dan 6217 data tweet sebagai data uji tanpa label. Proses pembelajaran terhadap data latih menghasilkan nilai akurasi sebesar 62,02% pada model *Naive Bayes Classifier* dan 66,68% pada model *Support Vector Machine*.

Hasil implementasi pada data uji sebanyak 6217 data tweet menggunakan model *Naive Bayes Classifier* menghasilkan klasifikasi kelas sentimen positif sebanyak 40,8% dengan 2535 data, kelas sentimen netral sebanyak 54,8% dengan 3407 data, dan kelas sentimen negatif sebanyak 4,4% dengan 275 data sedangkan menggunakan model *Support Vector Machine* menghasilkan klasifikasi kelas sentimen positif sebanyak 38,0% dengan 2363 data, kelas sentimen netral sebanyak 46,8% dengan 2910 data, dan kelas sentimen negatif sebanyak 15,2% dengan 944 data.

Kata kunci : Covid-19, Pandemi, PSBB, Analisis Sentimen, *Naive Bayes Classifier*, *Support Vector Machine*

COMPARATIVE STUDY OF NAIVE BAYES CLASSIFIER AND SUPPORT VECTOR MACHINE ALGORITHM IN TWEET SENTIMENT ANALYSIS

(Case Study: Tweet Data With Keywords 'psbb' and 'pandemic')

Ahmad Putra Awwalu Raafi'u

17106050023

ABSTRACT

Starting 2020, the world was shocked by the discovery of a new virus originating from Wuhan, the central city in China. This new virus, including a type of coronavirus, was later named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV2) and caused Corona Virus Disease 2019 (COVID-19). The Indonesian government has implemented various public policies to suppress the spread of the virus, one of which is by imposing *Pembatasan Sosial Berskala Besar* (PSBB). After a year of the pandemic in Indonesia and the implementation of the PSBB policy, there are many pros and cons regarding the PSBB policy among the public.

This study aims to analyze the sentiment or opinion that developed in the community towards the PSBB policy by the government during the COVID-19 pandemic using the *Naive Bayes Classifier* and *Support Vector Machine* methods. The data used are 4040 tweets as training data with labels and 6217 tweets as test data without labels. The learning process on the training data resulted in an accuracy value of 62.02% on the *Naive Bayes Classifier* model and 66.68% on the *Support Vector Machine* model.

The results of the implementation on test data of 6217 tweets using the *Naive Bayes Classifier* model resulted in a positive sentiment class classification of 40.8% with 2535 data, a neutral sentiment class of 54.8% with 3407 data, and a negative sentiment class of 4.4% with 275 data while using the *Support Vector Machine* model resulted in a positive sentiment class classification of 38.0% with 2363 data, a neutral sentiment class of 46.8% with 2910 data, and a negative sentiment class of 15.2% with 944 data.

Keywords: Covid-19, Pandemic, PSBB, Sentiment Analysis, Naive Bayes Classifier, Support Vector Machine

BAB I

PENDAHULUAN

1.1 Latar Belakang

Mengawali tahun 2020, dunia digemparkan dengan penemuan virus baru yang berasal dari Wuhan, Kota bagian tengah di China. Virus baru ini termasuk jenis coronavirus kemudian diberi nama *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-COV2) dan menyebabkan penyakit *Corona Virus Disease 2019* (COVID-19). Penyakit ini menginfeksi saluran pernafasan dengan gejala umum batuk, demam dan sesak nafas. Di Indonesia COVID-19 kasus positif pertama diumumkan pada tanggal 2 Maret 2020. Kasus tersebut menyerang ibu dan anak di Depok yang sebelumnya berkontak erat dengan warga negara asing yang terjangkit COVID-19.

Setelah menjangkit berbagai negara termasuk Indonesia dengan tingginya penyebaran dan keparahan yang diakibatkan serta diluar kontrolnya penularan virus tersebut, maka organisasi kesehatan dunia (WHO) menetapkan virus COVID-19 ini sebagai pandemic. Setelah ditetapkannya sebagai pandemi, semua negara harus melakukan penanganan khusus untuk mengendalikan pandemi ini.

Pandemi sendiri merupakan sebuah epidemi yang telah menyebar ke berbagai benua dan negara, umumnya menyerang banyak orang. Sementara epidemi sendiri adalah sebuah istilah yang telah digunakan untuk mengetahui peningkatan jumlah kasus penyakit secara tiba-tiba pada suatu populasi area

tertentu. Dengan dijadikannya COVID-19 ini sebagai pandemi, pemerintah Indonesia memberlakukan berbagai macam kebijakan publik untuk menekan tingkat penyebaran virus tersebut salah satunya dengan upaya memberlakukan Pembatasan Sosial Berskala Besar (PSBB).

Pemerintah menerapkan Pembatasan Sosial Berskala Besar (PSBB) demi memutus mata rantai penyebaran virus corona. Meskipun banyak fasilitas umum yang ditutup, namun beberapa sektor vital seperti fasilitas kesehatan, pasar atau minimarket tetap buka selama PSBB (Nasruddin & Haq, 2020). Pemberlakuan PSBB adalah strategi yang dianggap pemerintah efektif untuk menurunkan penyebaran COVID-19.

Setelah setahun pandemi di Indonesia dan pemberlakuan kebijakan PSBB tersebut, banyak pro dan kontra mengenai kebijakan PSBB itu di kalangan masyarakat. Di dalam penerapannya, banyak kontra di masyarakat yang menyebutkan bahwa PSBB membuat guncangan disektor ekonomi kelas menengah kebawah. Lalu banyak masyarakat yang mengabaikan anjuran dari pemerintah untuk menjalankan roda ekonominya. Tetapi tidak sedikit juga yang pro terhadap kebijakan tersebut agar pandemi ini cepat teratasi. Dalam menanggapi kebijakan itu, masyarakat menggunakan media sosial untuk mengungkapkan opini yang meraka rasakan.

Dalam mengungkapkan opininya, masyarakat menggunakan sosial media salah satunya dengan media twitter. Berdasarkan data yang dirilis oleh Twitter Indonesia, bahwa Indonesia merupakan salah satu dari 5 negara terbesar

dalam penggunaan Twitter secara aktif, yakni 79% dari pengguna Twitter Indonesia merupakan pengguna aktif. Reaksi masyarakat Indonesia baik yang berbentuk opini maupun fakta terhadap berbagai hal yang terjadi, terkumpul di dalam media sosial Twitter dan muncul sebagai tweets. Isi Tweet ini menjadi sumber data apabila diolah dengan benar dan tepat dapat menghasilkan berbagai informasi yang berguna, misalnya dapat digunakan sebagai pengambilan keputusan atau digunakan untuk meningkatkan persepsi seseorang terhadap suatu hal yang baru diketahui. Bentuk analisis pendapat atau review dapat dilakukan menggunakan analisis sentimen (Sari et al., 2020).

Analisis sentimen atau dapat disebut *opinion mining* adalah suatu proses dengan melakukan pemahaman, pengekstrakan dan pengolahan data tekstual secara otomatis agar didapatkan informasi. Pertumbuhan data yang terjadi secara terus menerus menjadi tantangan sendiri untuk proses analisis sentimen.

Oleh karena itu, penelitian ini akan melakukan analisis terhadap sentimen atau opini yang berkembang di masyarakat terhadap kebijakan PSBB oleh pemerintah saat pandemi COVID-19. Data yang akan dianalisis adalah data tweet berbahasa Indonesia dengan kata kunci “PSBB” dan “Pandemi”. Data tersebut kemudian diklasifikasikan menjadi tiga kelas sentimen, yaitu positif, negatif, dan netral. Metode yang akan digunakan dalam penelitian ini adalah *Naive bayes Classifier* dan *Support Vector Machine*.

Dengan adanya penelitian ini, diharapkan dapat membantu pihak-pihak terkait yang membutuhkan analisis informasi mengenai kebijakan PSBB oleh

pemerintah dan dapat melakukan tindakan atau kebijakan kedepannya yang sesuai serta menjangkau ke seluruh elemen masyarakat.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas, maka dapat diambil rumusan masalah yang akan menjadi pembahasan di penelitian ini, yaitu : “ Bagaimana melakukan analisis sentimen mengenai kebijakan PSBB selama pandemi di indonesia pada sosial media twitter menggunakan algoritma *Naive Bayes Classifier* dan *Support Vector Machine*”.

1.3 Batasan Masalah

Adapun yang menjadi batasan masalah masalah dalam penelitian ini adalah :

1. Data yang digunakan adalah data tweet berbahasa indonesia dengan kata kunci “PSBB” dan “Pandemi”.
2. Data yang diambil adalah data random bulan maret – april 2021.
3. Menggunakan bahasa pemrograman python.
4. Algoritma yang digunakan dalam pengklasifikasian ini adalah *Naive Bayes Classifier* dan *Support Vector Machine*.
5. Data diklasifikasikan kedalam kedalam tiga kelas yaitu positif, negatif, dan netral.

1.4 Tujuan

Adapun tujuan yang ingin dicapai oleh penulis dari penelitian dan penyusunan tugas akhir ini adalah :

1. Melihat sejauh mana tingkat akurasi analisis sentimen menggunakan metode *Naive Bayes Classifier* dan *Support Vector Machine*
2. Mengklasifikasikan sentimen mengenai kebijakan PSBB selama pandemi di Indonesia.

1.5 Manfaat Penelitian

Dengan tercapainya tujuan dari penelitian ini, diharapkan akan bermanfaat bagi semua pihak yang terkait. Manfaat dari penelitian ini adalah :

1. Dengan algoritma *Naive Bayes Classifier* dan *Support Vector Machine* dapat melakukan klasifikasi sentimen terhadap data tweet.
2. Dapat mengetahui algoritma mana yang memiliki akurasi lebih tinggi dalam mengklasifikasikan sentimen terhadap data tweet.
3. Bagi UIN Sunan Kalijaga, dapat digunakan sebagai tambahan referensi terkait dengan peningkatan penggunaan sosial media seiring kemajuan teknologi.
4. Bagi penulis, dapat memberikan pengetahuan dan wawasan yang baru, bahwa data tweet dapat dijadikan sebagai sumber data yang bisa digunakan untuk menganalisis opini terhadap suatu topik.
5. Bagi pembaca, dengan adanya penelitian ini dapat digunakan sebagai sumber pembelajaran untuk penelitian selanjutnya dan menambah wawasan baru.

1.6 Keaslian Penelitian

Penelitian mengenai analisis sentimen dan klasifikasi pada sosial media twitter menggunakan metode *Naive Bayes Classifier* dan *Support Vector*

Machine, hingga saat ini sudah banyak dilakukan oleh peneliti sebelumnya. Namun, berdasarkan referensi dan tinjauan pustaka, penelitian yang diajukan sebagai Tugas Akhir S1 pada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga mengenai analisis sentimen terhadap data tweet dengan metode Naive Bayes Classifier dan Support Vector Machine dengan kata kunci “PSBB” dan “Pandemi” belum pernah dilakukan di UIN Sunan Kalijaga Yogyakarta.

1.7 Sistematika Penulisan

Berikut ini adalah sistematika penulisan yang dilakukan dalam penelitian ini :

BAB I PENDAHULUAN

Bab ini berisikan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, keaslian penelitian dan sistematika penulisan skripsi.

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Menjelaskan penelitian yang telah dilakukan sebelumnya serta landasan teori yang berhubungan dengan topik yang akan dibahas dalam penelitian ini.

BAB III METODE PENELITIAN

Membahas metode yang dilakukan dalam penelitian serta langkah-langkah yang harus dilakukan untuk mencapai tujuan dalam menyelesaikan penelitian tugas akhir.

BAB IV HASIL DAN PEMBAHASAN

Bagian ini menjelaskan proses yang dilakukan dalam penelitian mulai dari pengumpulan data hingga hasil yang didapatkan dari penelitian.

BAB V KESIMPULAN DAN SARAN

Menjelaskan kesimpulan dari seluruh tahapan penelitian serta saran, sehingga dapat menjadi bahan pertimbangan bagi pihak-pihak yang berkepentingan serta pengembangan selanjutnya.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Data yang digunakan adalah data yang diambil dari twitter yang diambil dengan cara *crawling* menggunakan *API twitter* dengan kata kunci “psbb” dan “pandemi”. Data yang didapatkan sebanyak 11104 data tweet, kemudian dilakukan seleksi data sehingga mendapatkan 4040 data latih dan 6217 data uji. Data diolah menggunakan metode *Naive Bayes Classifier* dan *Support Vector Machine* dengan pembobotan TF-IDF serta dicari akurasi menggunakan *k-fold cross validation*.

Berdasarkan penelitian yang telah dilakukan terhadap 4040 data latih dengan pembobotan TF-IDF menggunakan metode *Naive Bayes Classifier* didapatkan nilai rata-rata akurasi saat menggunakan *16-fold cross validation* sebesar 62,02%. Sedangkan nilai rata-rata akurasi saat menggunakan *16-fold cross validation* pada metode *Support Vector Machine* kernel *Linear* adalah sebesar 66,04% dan kernel RBF sebesar 66,68%. Berikut merupakan perbandingan nilai akurasi dari kedua metode tersebut pada Tabel 5.1 berikut :

Tabel 5. 1 Perbandingan Hasil Akurasi

	TF-IDF		
	16-fold cross validation		
	NBC	SVM Kernel Linear	SVM Kernel RBF
Akurasi	62,02%	66,04%	66,68%

Dari perbandingan nilai akurasi diatas didapatkan kesimpulan bahwa kombinasi penggunaan metode untuk mencari model klasifikasi menggunakan metode *Naive Bayes Classifier* dan *Support Vector Machine* kernel RBF dikarenakan hasil akurasi menggunakan kernel RBF lebih tinggi dari pada menggunakan kernel *Linear*.

Selanjutnya, hasil implementasi pada 6217 data uji yang belum memiliki label dengan menggunakan model *Naive Bayes Classifier* menghasilkan kelas sentimen positif sebanyak 40,8% dengan 2535 data, kelas sentimen netral sebanyak 54,8% dengan 3407 data, dan kelas sentimen negatif sebanyak 4,4% dengan 275 data. Sedangkan menggunakan model *Support Vector Machine* kernel RBF dengan parameter $C = 2$, $\gamma = 0,5$ menghasilkan kelas sentimen positif sebanyak 38,0% dengan 2363 data, kelas sentimen netral sebanyak 46,8% dengan 2910 data, dan kelas sentimen negatif sebanyak 15,2% dengan 944 data.

Dari kedua model tersebut sama sama menghasilkan nilai tertinggi dalam mengklasifikasikan data kelas sentimen netral dan nilai terendah dalam mengklasifikasikan data kelas sentimen negatif.

5.2 Saran

Pada penelitian ini tentunya tidak terlepas dari kekurangan dan kelemahan. Oleh karena itu peneliti menyarankan adanya perbaikan-perbaikan yang dapat dijadikan masukan untuk peneliti selanjutnya, adapun sebagai berikut :

1. Penelitian selanjutnya diharapkan menggunakan pengembangan dari algoritma *Naive Bayes Classifier*.

2. Penelitian selanjutnya diharapkan menggunakan data latih dengan jumlah yang lebih banyak agar mendapatkan model yang lebih akurat.
3. Penelitian selanjutnya diharapkan dapat melakukan proses pelabelan data latih dibawah pengawasan pakar bidang terkait sehingga validitas dapat dimaksimalkan dan tidak berdasar penilaian subjektif saja.
4. Penelitian selanjutnya diharapkan bisa menggunakan tahapan preprocessing yang berbeda



DAFTAR PUSTAKA

- Agustina, D. A., Subanti, S., & Zukhronah, E. (2021). Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine. *Indonesian Journal of Applied Statistics*, 3(2), 109. <https://doi.org/10.13057/ijas.v3i2.44337>
- Ahmad, A. (2017). ANALISIS BIG DATA PENGGUNAAN INTERNET PADA PENGGUNA LAYANAN UBIQU. *Jurnal Teknologi Indonesia*.
- Alfikri, M. Z. (2020). *Analisis Sentimen Twitter terhadap Kartu Prakerja di tengah Pandemi COVID-19 menggunakan Algoritma Pencocokan String dan library TextBlob*.
- Chalida, M., & R.Wahyudi, M. D. (2019). Analisis Sentimen Ujaran Kebencian Pemilihan Presiden 2019 Menggunakan Algoritma Naïve Bayes (Studi Kasus: Tweet #Pilpres2019 Di Kota Jakarta, Bandung, Semarang, Surabaya Dan Yogyakarta). *Jnanaloka (Jurnal Open Access Yayasan Lentera Dua Indonesia)*, 2001, 5–10.
- Februariyanti, H., & Zuliarso, E. (2012). Klasifikasi Dokumen Berita Teks Bahasa Indonesia menggunakan Ontologi. *Teknologi Informasi DINAMIK*, 17(1), 14–23. <http://www.unisbank.ac.id/ojs/index.php/fti1/article/view/1612/594>
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *Smatika Jurnal*, 10(02), 71–76. <https://doi.org/10.32664/smatika.v10i02.455>
- Handayani, F., & Pribadi, S. (2015). Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110. *Jurnal Teknik Elektro*, 7(1), 19–24. <https://doi.org/10.15294/jte.v7i1.8585>
- Lestari, A. R. T., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang

- Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Näive Bayes dan Pembobotan Emoji. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(12), 1718–1724. <http://j-ptiik.ub.ac.id>
- Lusiana, L. (2018). *ANALISIS SENTIMEN TWITTER DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) (Studi Kasus: 3556 Data Tweets dengan Kata Kunci Cadar dan Hijab)*.
- Nasruddin, R., & Haq, I. (2020). Pembatasan Sosial Berskala Besar (PSBB) dan Masyarakat Berpenghasilan Rendah. *SALAM: Jurnal Sosial Dan Budaya Syar-I*, 7(7). <https://doi.org/10.15408/sjsbs.v7i7.15569>
- Nugraha, A. S., & Purnamasari, K. K. (2019). *Penerapan Metode Support Vector Machine Pada Part of Speech Tag Bahasa Indonesia*. 112.
- Nugroho, G. A. (2020). *Analisis big data penggunaan internet pada pengguna layanan ubiqu*.
- Permadi, V. A. (2020). Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura. *Jurnal Buana Informatika*, 11(2), 140. <https://doi.org/10.24002/jbi.v11i2.3769>
- Pitria, P. (2019). Pengguna Twitter Pada Akun Resmi Samsung Indonesia Dengan Menggunakan Näive Bayes. *Undergraduate Theses from JBPTUNIKOMPP*.
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, 3(1), 50. <https://doi.org/10.35314/isi.v3i1.335>
- Rauhan, A. (2019). Pengolahan Data Menggunakan Machine Learning. *Student Paper Pertamina University*, 021, 1–4. [https://library.universitaspertamina.ac.id/xmlui/bitstream/handle/123456789/162/Jurnal Ilmiah.pdf?sequence=1&isAllowed=y](https://library.universitaspertamina.ac.id/xmlui/bitstream/handle/123456789/162/Jurnal%20Ilmiah.pdf?sequence=1&isAllowed=y)

- Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156. <https://doi.org/10.51211/biict.v7i2.1422>
- Rozi, I., Pramono, S., & Dahlan, E. (2012). Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi. *Jurnal EECCIS*, 6(1), 37–43.
- Sari, D. I., Wati, Y. F., & Widiastuti. (2020). Analisis Sentimen Dan Klasifikasi Tweets Berbahasa Indonesia Terhadap Transportasi Umum Mrt Jakarta Menggunakan Naïve Bayes Classifier. *Jurnal Ilmiah Informatika Komputer*, 25(1), 64–75. <https://doi.org/10.35760/ik.2020.v25i1.2427>
- Septian, J. A., Fahrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor. *Journal of Intelligent Systems and Computation*, September, 43–49.
- Sevsa, B. A., & R Wahyudi, M. D. (2019). Analisis Sentimen pada Indeks Kinerja Dosen Fakultas SAINTEK UIN Sunan Kalijaga Menggunakan Naive Bayes Classifier. *Jurnal Buana Informatika*, 10(2), 112. <https://doi.org/10.24002/jbi.v10i2.2250>
- Surohman, S., Aji, S., Rousyati, R., & Wati, F. F. (2020). Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K-Nearest Neighbor. *EVOLUSI : Jurnal Sains Dan Manajemen*, 8(1), 93–105. <https://doi.org/10.31294/evolusi.v8i1.7535>
- Tempola, F., Muhammad, M., & Khairan, A. (2018). Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(5), 577. <https://doi.org/10.25126/jtiik.201855983>
- Tuhuteru, H., & Iriani, A. (2018). Analisis Sentimen Perusahaan Listrik Negara

Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier. *Jurnal Informatika: Jurnal Pengembangan IT*, 3(3), 394–401. <https://doi.org/10.30591/jpit.v3i3.977>



LAMPIRAN

Source code dan data dapat diakses melalui link berikut :

<https://github.com/ahmad-putra/Analisis-Sentimen>

