Studi Perbandingan Metode Analisis *Naive Bayes Classifier* dengan *Support Vector Machine* untuk Analisis Sentimen (Studi Kasus: Tweet Berbahasa Indonesia Tentang Covid-19)

Skripsi

untuk memenuhi sebagai persyaratan

mencapai derajat Sarjana S-1

Program Studi Teknik Informatika

Disusun oleh:

Ahmad Nur Fauzi

16650029

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2020

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-1374/Un.02/DST/PP.00.9/06/2020

Tugas Akhir dengan judul : STUDI PERBANDINGAN METODE ANALISIS NAIVE BAYES CLASSIFIER DENGAN SUPPORT MACHINE UNTUK ANALISIS SENTIMEN (STUDI KASUS : TWEET BERBAHASA INDONESIA TENTANG COVID-19)

yang dipersiapkan dan disusun oleh:

Nama : AHMAD NUR FAUZI
Nomor Induk Mahasiswa : 16650029
Telah diujikan pada : Kamis, 25 Juni 2020
Nilai ujian Tugas Akhir : A

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

TIM UJIAN TUGAS AKHIR

Ketua Sidang/Penguji I

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

Valid ID: 5f0baf0ed3174

Penguji II

Agus Mulyanto, S.Si., M.Kom.
SIGNED

Valid ID: 5f0bc747e81ed

Penguji III

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

Valid ID: 5f0ba859171ff

Yogyakarta, 25 Juni 2020
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Dr. Murtono, M.Si.
SIGNED

Valid ID: 5f0bd591dd1b8

SURAT PERSETUJUAN SKRIPSI

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 Nur Fauzi
NIM : 16650029
Judul Skripsi : "Studi Perbandingan Metode Analisis Naive Bayes Classifier dengan Support Vector Machine untuk Analisis Sentimen (Studi Kasus: Tweet Berbahasa Indonesia Tentang Covid-19 Dari Tgl. 27 Maret 2020 S.D. 23 April 2020)"

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 mengharap agar skripsi tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 16 Juni 2020
Pembimbing

Agung Fatwanto, Ph.D.
NIP: 19770103 200501 1 003

PERNYATAAN KEASLIAN SKRIPSI

PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini :

Nama : Ahmad Nur Fauzi

NIM : 16650029

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi saya yang berjudul "**Studi Perbandingan Metode Analisis Naive Bayes Classifier dengan Support Vector Machine untuk Analisis Sentimen (Studi Kasus: Tweet Berbahasa Indonesia Tentang Covid-19 Dari Tgl. 27 Maret 2020 S.D. 23 April 2020)**" 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.

Yogyakarta, 16 Juni 2020

Yang menyatakan

Ahmad Nur Fauzi

NIM.16650029

KATA PENGANTAR

Alhamdulillahirobbil'alamin. Puji syukur kehadirat Allah SWT karena rahmat dan izin-Nya sehingga penulis dapat menyelesaikan penelitian yang berjudul **Studi Perbandingan metode analisis *Naive Bayes classifier* dengan *Support Vector Machine* dalam sentimen *twitter*** sebagai salah satu syarat untuk mencapai gelar sarjana program studi Teknik Informatika Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Sholawat serta salam selalu tercurahkan kepada junjungan kita Nabi Muhammad SAW. beserta seluruh keluarga dan sahabat beliau.

Penulis menyadari bahwa apa yang dilakukan dalam penyusunan laporan penelitian ini masih terlalu jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran yang berguna dalam penyempurnaan analisis ini di masa yang akan datang. Semoga apa yang telah penulis lakukan dapat bermanfaat bagi pembaca.

Selanjutnya, penulis ingin mengucapkan terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian skripsi ini, baik secara langsung maupun tidak langsung. Ucapan terimakasih penulis sampaikan kepada:

1. Bapak Prof. Dr. Phil. Al Makin, S.Ag., M.A. selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Bapak Dr. Murtono, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Bapak Sumarsono, S.T., M.Kom., selaku Ketua Program Studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta.

4. Bapak Agung Fatwanto, Ph.D selaku dosen pembimbing skripsi yang telah sabar dan meluangkan waktunya untuk memberikan koreksi dan kritik saran kepada penulis sehingga skripsi ini dapat diselesaikan sekaligus selaku Dosen Pembimbing Akademik.
5. Agus Mulyanto, S.Si., M.Kom., M. Didik Rohmad Wahyudi, S.T., MT, Aulia Faqih Rifa'i, M.Kom., M. Taufiq Nuruzzaman, S.T., Maria Ulfah Siregar, S.Kom. MIT., Ph.D., Nurochman, S.Kom., M.Kom., Rahmat Hidayat, S.Kom., M.Cs., Dr. Shofwatul 'Uyun, S.T., M.Kom., selaku dosen pengampu mata kuliah program studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta yang telah banyak membantu sehingga penulis dapat menyusun tugas akhir.
6. Pak Wahdan, dan seluruh staf karyawan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
7. Manusia terbaik, Ibu Sri Lestari dan Bapak Muhyidin. Segala yang baik baik berawal dan berasal dari mereka yang selalu mensupport sepenuhnya.
8. Seluruh Keluarga Besar Bapak Nardisihono dan Ibu Mursini yang senantiasa mendoakan dan memberikan dukungan kepada penulis.
9. Partner yang selalu support Bela Kurnia Davis dalam setiap sambat, serta Maria Josephine, Rafif Naufal, Ken ratri, Seno Adi dan Wening yang senantiasa menghibur dan membantu.
10. Teman-teman diskusi dalam banyak hal sdr Bayu Irfan, Syafaat Adi, Sholehudin Nur, Yakin Arif, Ulfa Mulya, Ari Lukman, Azis Alvri.
11. Teman sambat dan misuh Aditya, Agus, Nando, Irfan, Rahman.

12. Teman-teman yang pernah menjadi satu kelompok selama kuliah dan seluruh teman-teman Teknik Informatika 2016 yang tidak dapat penulis sebutkan satu per satu.
13. Keluarga Kreasi Kode yang senantiasa menjadi tempat mencari ilmu dan pengalaman.
14. Teman-teman KKN Kelompok 204 Angkatan 99 yang ikut menemani dalam perjalanan penyelesaian skripsi ini.
15. Serta semua pihak yang memberi dukungan *lillah*, sehingga penelitian ini dapat terselesaikan.

Yogyakarta, 16 Juni 2020

Penulis

HALAMAN PERSEMBAHAN

Alhamdulillahi rabbil 'alamin, segala puji syukur hanya bagi Allah SWT. Terima kasih kepada semua pihak yang telah banyak membantu penulis sampai saat ini. Oleh karena itu penulis ingin mempersembahkan hasil tulisan ini kepada semua pihak yang telah banyak membantu, mendukung, dan menginspirasi penulis.

Skripsi ini penulis persembahkan kepada:

1. Orang tuaku tercinta, Bapak Muhyidin dan Ibu Sri Lestari yang senantiasa memberikan nasehat dan motivasi serta tak pernah lelah mendoakan penulis.
2. Adikku, Hayfa Isnaini Al-Husna yang selalu menghibur dan memberi semangat kepada penulis.
3. Bapak Agung Fatwanto, Ph.D, yang telah memberikan arahan dan bimbingan dalam menyusun skripsi ini.
4. Partner yang selalu support Bela Kurnia Davis dalam setiap sambat, serta Maria Josephine, Rafif Naufal, Ken ratri, Seno Adi dan Wening yang senantiasa menghibur dan membantu.
5. Teman-teman KKN Kelompok 204 Angkatan 99
6. Semua pihak yang telah membantu dan mensupport penulis, yang tidak bisa penulis sebutkan satu persatu. Semoga Allah membalas amal kebaikan dan ibadah kalian semua

HALAMAN MOTTO

"*Beri nilai dari usahanya jangan dari hasilnya. Baru kita bisa mengerti kehidupan*"

- Albert Einstein.

"*Usaha Tidak Akan Mengkhianati Hasil*"

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

Studi Perbandingan Metode Analisis Naive Bayes Classifier dengan Support Vector Machine untuk Analisis Sentimen

(Studi Kasus: Tweet Berbahasa Indonesia Tentang Covid-19 Dari Tgl. 27 Maret 2020 S.D. 23 April 2020)

Ahmad Nur Fauzi

16650029

INTISARI

Analisis sentimen adalah sebuah teknik untuk mendeteksi opini terhadap suatu subyek (misalnya individu, organisasi ataupun produk) dalam sebuah kumpulan data. Kehidupan masyarakat saat ini mendukung munculnya interaksi sosial melalui media sosial yang menjadi saran dalam menyampaikan opini yang bersifat positif maupun negatif. Dalam penelitian ini menggunakan 2 metode yaitu *Naïve Bayes Classifiers* (NBC) dan *Support Vector Machine* (SVM) untuk melakukan klasifikasi sentimen.

Sebanyak 9015 data digunakan dalam proses analisa dengan teknik pembobotan TF-IDF dan *Term Frequency*. Evaluasi dilakukan menggunakan 2 cara yaitu *confusion matrix* dan *K-fold cross validation*. Dengan *confusion matrix* SVM memiliki nilai akurasi 83%, nilai presisi 83%, nilai *recall* 98,8%, dan nilai *specificity* 9%. Sedangkan NBC memiliki nilai akurasi 82,3%, nilai presisi 82,3%, nilai *recall* 99,7%, dan nilai *specificity* 3,4% dengan TF-IDF. Lalu, SVM memiliki nilai akurasi 82,8%, nilai presisi 83%, nilai *recall* 99%, dan *specificity* 8%. Sedangkan NBC memiliki nilai akurasi 80,1%, nilai presisi 85,7%, nilai *recall* 90%, dan *specificity* 29% dengan TF.

K-fold cross Validation dengan pembobotan TF-IDF, SVM nilai akurasi 82,6%, nilai presisi 82,9%, tingkat *recall* 98,8%, dan *specificity* 10,9%. NBC nilai akurasi 82%, nilai presisi 82%, tingkat *recall* 99,6%, dan *specificity* 4,8%. *Term Frequency* SVM nilai akurasi 82,5%, nilai presisi 82,6%, tingkat *recall* 99,1%, dan *specificity* 7%. NBC nilai akurasi 79,8%, nilai presisi 85,9%, tingkat *recall* 89,2%, dan *specificity* 32%. NBC waktu proses olah data lebih cepat dari SVM dengan NBC dibawah 5 detik sedangkan SVM diatas 500 detik.

Kata kunci: *analisis sentimen, TF-IDF, NBC, SVM, Twitter, confusion matrix, K-fold cross validation,* akurasi, presisi, *recall, specificity.*

Comparative Study of Naive Bayes Classifier Analysis Method with Support Vector Machine for Sentiment Analysis
(Case Study: Indonesian Language Tweet About Covid-19 From 27 March 2020 S.D. 23 April 2020)

Ahmad Nur Fauzi
16650029

ABSTRACT

Sentimen analysis is a technique for detecting opinions about a subject (for example individuals, organizations or products) in a data set. Community life today supports the emergence of social interaction through social media which is a suggestion in conveying positive and negatif opinions. In this study using 2 methods, namely Naïve Bayes Classifiers (NBC) and Support Vector Machine (SVM) to classify sentimens.

A total of 9015 data were used in the analysis process with TF-IDF weighting techniques and Term Frequency. Evaluation uses 2 ways, that is *confusion matrix* and K-fold cross validation. With the *confusion matrix*, SVM has an accuracy value of 83%, a precision value of 83%, a *recall* value of 98.8%, and a *specificity* value of 9%. While NBC has an accuracy value of 82.3%, a precision value of 82.3%, a *recall* value of 99.7%, and a *specificity* value of 3.4% with TF-IDF. Then, SVM has an accuracy value of 82.8%, a precision value of 83%, a *recall* value of 99%, and a *specificity* of 8%. While NBC has an accuracy value of 80.1%, a precision value of 85.7%, a *recall* value of 90%, and a *specificity* of 29% with TF.

K-fold cross Validation with TF-IDF weighting, SVM accuracy value of 82.6%, precision value of 82.9%, *recall* rate of 98.8%, and *specificity* of 10.9%. NBC accuracy is 82%, precision value is 82%, *recall* rate is 99.6%, and *specificity* is 4.8%. SVM Term Frequency accuracy value is 82.5%, precision value is 82.6%, 99.1% *recall* rate, and *specificity* is 7%. NBC accuracy value is 79.8%, precision value is 85.9%, *recall* rate is 89.2%, and *specificity* is 32%. NBC data processing time is faster than SVM with NBC under 5 seconds while SVM is above 500 seconds.

Keywords: *sentimen analysis, TF-IDF, NBC, SVM, Twitter, confusion matrix, K-fold cross validation, accuracy, precision, recall, specificity.*

BAB I
PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi informasi dan komunikasi jelas memberi dampak pada perubahan gaya hidup masyarakat dunia. Situs internet telah menjadi lautan informasi bagi siapapun untuk mendapatkan informasi mengenai hal apapun. Jumlah informasi yang banyak tersebut membuka celah untuk dilakukan berbagai kajian untuk mendapatkan manfaat yang baru sehingga dapat digunakan sebagai pondasi perkembangan teknologi untuk kedepannya. Informasi berbentuk teks saat ini banyak terdapat di internet dalam format forum, blog, media sosial, serta situs berisi review. Penggunaan media sosial juga menjadi wadah bagi komunitas yang memiliki bakat dan minat yang sama untuk saling mengenal. Salah satu media sosial yang cukup banyak dipakai masyarakat adalah Twitter. Di Twitter setiap pengguna bebas memberikan pendapat atau pun komentar terhadap sesuatu. Perhatian kepada pemerintah juga tidak luput dari pengamatan para pengguna twitter.

Pada masa ini terjadi bencana wabah virus COVID-19, hal tersebut cukup membuat masyarakat menjadi panik dan menyebabkan kerusuhan. Peran pemerintah dalam mengatasi berbagai problema yang disebabkan oleh COVID-19 sangat diperlukan agar masyarakat dapat tenang dan merasa aman. Namun dalam pelaksanaan kebijakan yang telah dibuat, banyak sekali pro dan kontra yang terjadi di dalam masyarakat. Para pengguna twitter juga ikut dalam menanggapi

penanganan COVID-19 oleh pemerintah. Pandangan masyarakat pun ada yang mendukung pemerintah ada juga yang tidak mematuhi himbauan pemerintah. Dari hal tersebut sentiment masyarakat terhadap pemerintah bisa terbagi menjadi 2 antara mendukung pemerintah dan mengkritik pemerintah. Untuk mengetahui bagaimana tingkat kepercayaan masyarakat terhadap pemerintah dalam melakukan penanganan COVID-19 perlu adanya survey pendapat. Dengan bantuan sentimen analysis hal tersebut dapat dilakukan dengan data tweet. informasi yang yang tadinya tidak terstruktur tersebut dapat diubah menjadi data yang lebih terstruktur dengan melakukan pengelompokan.

Analisis sentimen adalah sebuah teknik untuk mendeteksi opini terhadap suatu subyek (misalnya individu, organisasi ataupun produk) dalam sebuah kumpulan data (Nasukawa & Yi, 2003). Menurut Cutlip & Center dalam Resphati (Resphati, 2010), opini adalah pengekspresian suatu sikap mengenai persoalan yang mengandung suatu sikap mengenai persoalan yang mengandung pertentangan. Opini di sini bisa berupa *mention* di media sosial maupun artikel di situs berita dan blog pribadi. Opini yang terbentuk dapat disimpulkan dan diklasifikasikan berdasarkan kriteria yang telah ditentukan. Banyak metode yang dapat digunakan untuk memudahkan proses klasifikasi sentiment data. Setiap metode memiliki karakteristik berbeda sehingga tingkat keefektifan dalam menganalisa data juga berbeda-beda.

Di dalam penelitian ini, akan dibahas tahapan yang dilalui untuk melakukan perbandingan metode yaitu *Naive Bayes Classifier* dan *Support Vector Machine* serta bagaimana mengukur kualitas hasil analisis menggunakan beberapa parameter

seperti akurasi, presisi, *recall*, dan *specificity*. Hasil Komparasi diharapkan dapat dipilih metode yang memiliki tingkat performansi lebih baik dari metode lainnya dengan kasus yang sama.

1.2 Rumusan Masalah

Berdasarkan Latar Belakang diatas maka dapat dirumuskan permasalahan pada penelitian adalah "Bagaimana perbandingan metode antara *Naïve Bayes* dan *Support Vector Machine* saat dipakai untuk melakukan klasifikasi dalam menganalisis sentimen".

1.3 Batasan Masalah

Agar penyusunan tugas akhir ini tidak keluar dari pokok permasalahan yang dirumuskan, maka ruang lingkup pembahasan dibatasi pada:

a. Algoritma yang digunakan dalam pengklasifikasian ini adalah *Naïve Bayes Classifier* dan *Support Vector Machine*
b. Data yang digunakan terdiri dari *Tweet* provider telekomunikasi berbahasa Indonesia dengan jumlah data yang digunakan berdasarkan perhitungan sample *Tweet* dengan kata kunci "lockdown, karantina, psbb" yang berhubungan dengan COVID 19 yg di crawling dari tanggal 27 Maret 2020 – 23 April 2020.
c. Proses *Stopword* dan *Stemming* hanya berlaku pada kata-kata berbahasa Indonesia saja.

d. Menggunakan metode TF-IDF untuk menghitung bobot kata.
e. Pada tahap proses *Text Mining* pada penelitian ini tidak dilakukan tahap tagging atau *Part of Speech Tagging*.
f. Menggunakan Bahasa Pemrograman Python dalam menyelesaiakan alat bantu Penelitian.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, adapun tujuan dari penelitian ini adalah melakukan perbandingan metode *Naïve Bayes Classifier* dan *Support Vector Machine* untuk melakukan analisis sentimen pada tweet berbahasa Indonesia dengan studi kasus pada tema COVID-19 yang dibuat tanggal 27 Maret 2020 s.d 23 April 2020.

1.5 Manfaat Penelitian

Berdasarkan latar belakang dan tujuan di atas, adapun manfaat dari penelitian ini adalah sebagai berikut:

a. Mengklasifikasikan sentimen pada Twitter dalam jumlah yang besar secara otomatis.
b. Menguji tingkat akurasi, presisi, *recall*, dan *specificity* antara dua metode yaitu NBC (*Naive Bayes* Classifier) dan SVM (*Support Vector Machine*).

1.6 Keaslian Penelitian

Penelitian mengenai analisis sentimen dan klasifikasi pada media sosial twitter menggunakan metode *Support Vector Machine* dan *Naive Bayes* hingga saat ini sudah banyak dilakukkan oleh peneliti sebelumnya. Namun berdasarkan referensi dan tinjauan Pustaka, penelitian yang diajukan sebagai Tugas Akhir S1 Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan

Kalijaga mengenai perbandingan metode *Naive Bayes* dan *Support Vector Machine* dengan studi kasus analisis sentimen pada Twitter.

1.7 Sistematika Penulisan

Sebagai gambaran dan kerangka yang jelas mengenai pokok bahasan setiap bab dalam penelitian ini, maka diperlukan sistematika penulisan. Penyusunan laporan tugas akhir ini memiliki sistematika penulisan yang diawali dari BAB I dan diakhiri BAB V. Berikut adalah penjelasan pada tiap-tiap bab dalam laporan penelitian ini:

BAB I PENDAHULUAN

Bab pendahuluan berisikan penjelasan mengenai latar belakang dilakukannya penelitian, rumusan masalah penelitian, batasan masalah, tujuan penelitian, manfaat penelitian, keaslian penelitian, dan sistematika penulisan penelitian.

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Bab tinjauan pustaka dan landasan teori berisikan mengenai penelitian terdahulu dan teori-teori dasar yang terkait dengan penelitian ini. Teori yang digunakan terdiri dari analisis sentimen, *Term Frequency, TF-IDF*, metode *Naïve Bayes Classifier* dan *Support Vector Machine*, *text preprocessing*, *Feature Extraxtion*, *Confusion matrix* dan *Python*.

BAB III METODE PENELITIAN

Bab metode penelitian berisi tentang penjelasan mengenai metode ataupun algoritma yang digunakan serta tahapan-tahapan

yang dilakukan untuk mencapai tujuan dan kesimpulan tugas akhir.

BAB IV HASIL DAN PEMBAHASAN

Bab hasil dan pembahasan membahas analisis data dan hasil dari penelitian yang telah dilakukan.

BAB V PENUTUP

Bab penutup berisi tentang kesimpulan dari hasil penelitian yang telah dilakukan. Selanjutnya, kekurangan yang ada pada penelitian dituliskan pada saran untuk pengembangan penelitian di masa yang akan datang.

BAB V

PENUTUP

2.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan tentang Studi Perbandingan Metode Analisis *Naive Bayes Classifier* Dengan *Support Vector Machine* Dalam Sentimen Twitter, dengan menggunakan 9015 data tweet dengan kata kunci "lockdown, psbb, karantina" yang dipilih dari 180141 tweet berbahasa Indonesia dengan metode propotional random sampling beradasarkan lokasi yang telah diberikan pelabelan oleh 5 trainer dengan tingkat *margin of error* yang digunakan sebesar 1%, *confidence level* sebesar 95%, jumlah populasi tidak diketahui, lalu tingkat proporsi sampel sebesar 50%, dapat disimpulkan bahwa Klasifikasi yang dilakukan menggunakan dua algoritma yaitu NBC dan SVM menghasilkan nilai akurasi, presisi, dan *recall* yang berbeda. SVM memiliki hasil yang lebih tinggi dari pada NBC dalam dua kali evaluasi yaitu menggunakan *confusion matrix* dan *K-fold cross validation*. Pada evaluasi dengan *confusion matrix* dengan pembobotan TF-IDF, SVM memiliki nilai akurasi 83%, nilai presisi 83%, nilai *recall* 98,8%, dan nilai *specificity* 9%. Sedangkan NBC memiliki nilai akurasi 82,3%, nilai presisi 82,3%, nilai *recall* 99,7%, dan nilai *specificity* 3,4%. Pada pembobotan dengan *Term Frequency* SVM memiliki nilai akurasi 82,8%, nilai presisi 83%, nilai *recall* 99%, dan *specificity* 8%. Sedangkan NBC memiliki nilai akurasi 80,1%, nilai presisi 85,7%, nilai *recall* 90%, dan *specificity* 29%. Pada evaluasi dengan *K-fold cross Validation* dengan pembobotan TF-IDF, SVM memiliki nilai akurasi 82,6%, nilai presisi 82,9%, tingkat *recall* 98,8%, dan *specificity* 10,9%. Sedangkan NBC

memiliki nilai akurasi 82%, nilai presisi 82%, tingkat *recall* 99,6%, dan *specificity* 4,8%. Pada pembobotan dengan *Term Frequency* SVM memiliki nilai akurasi 82,5%, nilai presisi 82,6%, tingkat *recall* 99,1%, dan *specificity* 7%. Sedangkan NBC memiliki nilai akurasi 79,8%, nilai presisi 85,9%, tingkat *recall* 89,2%, dan *specificity* 32%. Dalam memproses data set dari setiap trainer, NBC memiliki waktu proses yang sangat cepat dibandingkan dengan SVM. Pada NBC dalam evaluasi dengan *confusion matrix* rata-rata waktu proses TF-IDF 1,43 detik dan TF 3 detik. Pada *K-fold cross validation* rata-rata waktu proses TF-IDF 0,8 detik dan TF 2.5 detik. Sedangkan SVM dalam evaluasi dengan *confusion matrix* rata-rata waktu proses TF-IDF 900 detik dan TF 880 detik. Pada *K-fold cross validation* rata-rata waktu proses TF-IDF 1097 detik dan TF 935 detik. Berikut adalah hasil dari Perbandingan NBC dan SVM pada 9015 data dari 5 trainer.:

Tabel 5.1 *Kesimpulan perbandingan NBC dan SVM*

		TF-IDF		TF	
		NBC	SVM	NBC	SVM
CM	waktu	1,434 ± 0,23	900,875 ± 72,9	2,922 ± 0,23	880,875 ± 41,8
	akurasi	0,823 ± 0,06	0,831 ± 0,056	0,801 ± 0,055	0,828 ± 0,058
	presisi	0,823 ± 0,06	0,833 ± 0,055	0,857 ± 0,043	0,829 ± 0,058
	recall	0,997 ± 0,002	0,989 ± 0,011	0,900 ± 0,04	0,991 ± 0,099
	specificity	0,034 ± 0,04	0,096 ± 0,09	0,29 ± 0,16	0,08 ± 0,08
K-fold	waktu	0,801 ± 0,02	1097,431 ± 51	2,581 ± 0,04	935,440 ± 57
	akurasi	0,821 ± 0,062	0,827 ± 0,058	0,798 ± 0,058	0,825 ± 0,059
	presisi	0,821 ± 0,063	0,829 ± 0,057	0,859 ± 0,048	0,827 ± 0,059
	recall	0,996 ± 0,003	0,988 ± 0,012	0,892 ± 0,043	0,991 ± 0,011
	specificity	0,049 ± 0,049	0,109 ± 0,08	0,322 ± 0,17	0,077 ± 0,086

2.2 Saran

Pada penelitian ini masih banyak sekali kekurangan. Maka dari itu penulis menyarankan beberapa hal untuk penelitian selanjutnya, diantaranya:

1. Diharapkan penelitian selanjutnya pemberian labeling atau sentimen secara manual dapat dilakukan di bawah pengawasan pakar bidang terkait sehingga validitas dapat lebih dimaksimalkan dan tidak hanya berdasar penilaian subjektif.
2. Penelitian selanjutnya dapat melakukan percobaan dengan menggunakan 3 kelompok sentimen atau lebih sehingga klasifikasi dapat lebih relevan dan lebih baik.
3. Membandingkan kembali dengan metode pembelajaran mesin lainnya untuk menemukan model klasifikasi yang lebih baik.
4. Menggunakan data yang lebih banyak dan preprocessing yang berbeda sehingga menghasilkan model klasifikasi yang lebih akurat.

DAFTAR PUSTAKA

Alam, M., Sumy, S. A., & Parh, Y. A. (2015). Selection of the Samples with Probability Proportional to Size. *Science Journal of Applied Mathematics and Statistics*, 230-233.

Aliandu, P. (2013). Sentiment Analysis on Indonesian Tweet. *The Proceedings of The 7th ICTS*, 203-208.

Anita Novantirani, M. K. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine . *e-Proceeding of Engineering* , 1178.

Arief, R., & Imanuel, K. (2019). Analisis Topik Viral Desa Penari pada Media Social Twitter dengan Metode Lexicon Based. *Jurnal Ilmiah Matrix*, 242-250.

Bakliwal, A., Arora, P., Madhappan, S., Kapre, N., Singhand, M., & Varma, V. (2012). Mining Sentiments from Tweets. *Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, 11-18.

Baktikominfo. (2019, September 2). *Bahasa Pemrograman Python : pengertian, sejarah, kelebihan dan kekurangannya*. Retrieved from Baktikominfo: https://www.baktikominfo.id/en/informasi/pengetahuan/bahasa_pemrograman_python_pengertian_sejarah_kelebihan_dan_kekurangannya-954

Bintang, I. (2019, November 27). *Menyiapkan Data untuk Machine Learning*. Retrieved from Medium: https://medium.com/@nullphantom/menyiapkan-data-untuk-machine-learning-a2aa59f03256

Buntoro. (2016). Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine. *Sains Techno*, 12-25.

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information Credibility on Twitter. *WWW*, 675-684.

Cevikalp, H. (2016). Best Fitting Hyperplanes for Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1076 - 1088.

Chandani. (2015). Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film. *Techno Inka*, 56-67.

Cresswell, J. W. (2008). *Educational Research. Third Edition.* New Jersey: Pearson Education, Inc.

Daniel, W. (1999). *Biostatistics: A Foundation for Analysis in the Health Sciences. 7th edition.* New York: John Wiley & Sons.

Dewi, S. (2016). Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan. *Jurnal Techno Nusa Mandiri*, 60-66.

Didik Garbian Nugroho, Y. H. (2016). Analisis Sentimen pada Jasa Ojek Online Menggunakan Metode Naïve Bayes. *Semnas Saintek*, 24-36.

Dorsey, J. (2010, September 3). *The Evolving Ecosystem*. Retrieved from Blog Twitter: https://blog.twitter.com/official/en_us/a/2010/the-evolving-ecosystem.html

Hadna, N. M. (2016). Perbandingan Metode untuk Proses Analisis Sentimen di Twitter. *Jitka*, 97-114.

Hall, M. (2006). A Decision Tree-Based Attribute Weighting Filter for Naive Bayes. . *KnowledgeBased Systems*, 120–126.

Han, J., & Kamber, M. (2001). *Data Mining: Concepts and Techniques Tutorial.* San Frasisco: Morgan Kaufman.

Hania, A. A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. *Jurnal Teknologi Indonesia*, 21-27.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: data mining , inference, and prediction.* New York: Springer-Verlag.

Haykin, S. (1999). *Neural Network: A Comprehensive Foundation.* New Jersey: Prentice Hall.

Huang, K. Y. (2008). *Machine Learning Modeling Data Locally And Globally.* Berlin Heidelberg: Zhejiang University Press, Hangzhou And Springer-Verlag Gmbh.

Indrayuni, E. (2018). Komparasi Algoritma Naive Bayes dan Support Vector Machine untuk Analisa Sentimen review film. *Jurnal Pilar Nusa Mandiri*, 175-182.

Isaac, S., & Michael, W. B. (1977). *Handbook in Research and Evaluations.* San Diego, California: Ediths Publisher.

Iskandar, D., & Suprapto, Y. K. (2015). Perbandingan Akurasi Klasifikasi tingkat Kemiskinan antara algoritma c 4.5 dan naïve bayes. *Jurnal Ilmiah NERO*, 37-43.

Iskandar, D., & Suprapto, Y. K. (2015). Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan antara algoritma c 4.5 dan naïve bayes. *Jurnal Ilmiah NERO* , 37-43.

Kadir, A. (2005). *Dasar Pemrograman Python.* Yogyakarta: Andi Offset.

Kemp, S. (2020). *Digital 2020: Global Digital Overview*. New York: We Are Social and Hootsuite.

Kohavi, R. (1995). *A study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection.* Paris: frostiebek.

Lestari, A. R. (2017). Analisis Tentang Opini Pilkada DKI 2017 pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes dan Pembobotan Emoji. *MetroTekno*, 70-83.

Ma'arif, A. A. (2015). Penerapan Algoritma Tf-Idf Untuk Pencarian Karya Ilmiah. *Jurusan Teknik Informatika Fakultas Ilmu Komputer. Universitas Dian Nuswantoro Semarang*, 15-22.

Maimon, O. (2010). *Data Mining And Knowledge Discovery Handbook.* London: Springer.

Melita, R., Amrizal, V., Suseno, H. B., & Dirjam, T. (2018). Penerapan Metode Term Frequency Inverse Document Frequency (tf-idf) dan Cosine Similarity Pada Sistem temu kembali informasi untuk mengetahui syarah hadits berbasis web (studi kasus: syarah umdatil ahkam). *JURNAL TEKNIK INFORMATIKA* , 149-165.

Mustafaraj, E., & Metaxas, P. (2010). From obscurity to prominence in minutes: Political speech and real-time search. *Proceedings of the WebSci10* (pp. 150-155). Berlin: Extending the Frontiers of Society On-Line.

Nasukawa, T., & Yi, J. (2003). Sentiment Analysis: Capturing Favorability Using Natural Language Processing. *Proceedings of the 2nd International Conference on Knowledge Capture*, 70–77.

Nisfiannoor, M. (2009). *Pendekatan Statistika Modern untuk Ilmu Sosial.* Jakarta: Salemba Humatika.

Nugroho, Y. S. (2011). Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro. *UDN Jurnal*, 43-61.

Nurirwan Saputra, T. B. (2014). Analisis Sentimen Data Presiden Jokowi Dengan Preprocessing Normalisasi Dan Stemming Menggunakan Metode Naive Bayes Dan Svm. *Dinamika Informatika*, 50-62.

Pitria, P. (2016). Analisis Sentimen Pengguna Twitter Pada Akun Resmi Samsung Indonesia Dengan Menggunakan Naïve Bayes. *Jurnal Ilmiah Komputer dan Informatika*, 50-57.

Prasetyo, E. (2012). *Data mining konsep dan aplikasi menggunakan matlab.* Yogyakarta: Andi.

Putra, A. A. (2016). Implementasi Text Summarization Menggunakan Metode Vector Space Model pada Artikel Berita Bahasa Indonesia. *Jurusan Teknik Informatika. Fakultas Teknik dan Ilmu Komputer. Universitas Komputer Indonesia.* , 45-52.

Qi, Z., Tian, Y., & Shi, Y. (2013). *Multi-instance classification based on regularized multiple criteria linear programming. , 23(3-4),* . Wuhan: Neural Computing and Applications.

Rauhan, A. (2019). Pengolahan Data Menggunakan Machine Learning. *Student Paper (EE)*, 12-15.

Resphati, A. (2010). *Opini Masyarakat Tentang Pemberitaan Demo 100 Hari Pemerintahan SBY-Boediono.* Jawa Timur: Surat Kabar Jawa Pos.

Rish, I. (2005). *An empirical study of the naive Bayes classifier.* New York: T.J. Watson Research Center.

Rufiqoh, U. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter. *Telkom Journal*, 120-131.

Saberi, B., & Saad, S. (2017). Sentiment Analysis or Opinion Mining: A Review. *Ijaseit* , 1660-1666.

Saputra, N., Adji, T. B., & Permanasari, A. E. (2015). Analisis Sentimen Data Presiden Jokowi dengan Preprocessing Normalisasi dan Stemming Menggunakan Metode Naive Bayes dan SVM. *Jurnal Dinamika Informatika* , 12-18.

Scholkopf, B., & Smola, A. (2002). *Learning with Kernels.* Cambridge, Massachusetts: The MIT Press.

Schwarz, J., & Morris., M. R. (2011). Augmenting Web Pages and Search Results to Support Credibility Assessment. *ACM Conference on Human Factors in Computing Systems (CHI)* (pp. 50-55). New York: ACM Press.

Sentiaji. (2015). Analisis Sentimen Terhadap Acara Televisi Berdasarkan Opini Publik. *AITIF*, 34-44.

Supranto, j. (2009). *Statistik: Teoti dan Aplikasi.* Jakarta: Penerbit Erlangga.

Taheri, S., & Mammadov, M. (2013). Learning The Naive Bayes Classifier With Optimization Models. *International Journal of Applied Mathematics and Computer Science*, 787– 795.

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 (JTIIK)*, 577-584.

Tokunaga, T., & Iwayama, M. (1994). *Text Categorization Based On Weighted Inverse Document Frequency*. Tokyo, Japan: Tokyo Institute of Technology.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.

Vieweg, S. (2012). *Microblogged contributions to the emergency arena: Discovery, interpretation and implications.* New York: Computer Supported Collaborative Work.

Wang, L. (2005). *Support Vector Machines: Theory and Applications.* Berlin: Springer.

Widystuti, W., & Darmawan, J. B. (2018). Pengaruh Jumlah Data Set terhadap Akurasi Pengenalan dalam Deep Convolutional Network . *Konferensi Nasional Sistem Informasi* , 634-639.

WIlan, W., Hanasbey, S. H., Awinero, M. R., Modouw, J. V., & Sitokdana, M. N. (2018). Analisis Sentimen Terhadap Opini Masyarakat Indonesia Mengenai Bukalapak. *Seminar Nasional Teknologi Informasi dan Multimedia*, 17-28.

Wulandini, F., & Nugroho, A. S. (2009). Text Classification Using Support Vector Machine for Webmining Based Spatio Temporal Analysis of the Spread of Tropical Diseases. *International Conference on Rural Information and Communication Technology*, 189-192.

Yusuf, A. (2016). *Metode penelitian kuantitatif, kualitatif & penelitian gabungan.* Jakarta: Prenada Media.

Zy, A. T. (2017). Comparison Algorithm Classification Naive Bayes, Decision Tree, And Neural Network For Analysis Sentiment. *Jurnal Pelita Teknologi* , 5-12.

LAMPIRAN

Lampiran 1 *Contoh Tweet Hasil Crawling*

tweet	lokasi
PSBB Tangerang Raya Dimulai Besok, Puluhan Check Point Disebar https://t.co/ncntW4us3r	jakarta
@ben_khair Haha kaga blay.... psbb diperpanjang, 2 bulan nih gua kekurung dikosan Ã°Â?Â?Â?	jakarta
psbb diperpanjang. animal crossing, kuharapkan update update ciamikmu untuk temani.	jakarta
Presiden Joko Widodo (Jokowi) resmi melarang masyarakat untuk melakukan aktivitas mudik saat hari raya Idul Fitri untuk wilayah Jabodetabek, serta wilayah yang telah menerapkan status pembatasan sosial berskala besar (PSBB). https://t.co/Ulinq9AKUQ	jakarta
Plis pengen ulang tahun terus pintu kamar di ketok, temen2 pada masuk terus bilang "SURPRISE, karantina ini cuma prank ulang tahun lo biar stay dirumah!!! Kaget kann???"	jakarta
16. Pedoman Penggunaan Kendaraan Selama PSBB di Jakarta https://t.co/WgNJ9WuNk3 Ã¢Â?Â? (17.04.20) #KlinikSepekan	jakarta
Masih Banyak Warga Tak Patuh PSBB Covid-19 di Kota Depok https://t.co/PsIxUt9FAC https://t.co/ptXDNOvHhq	jakarta
SobatQ, hari ini 20 April jadi hari tuk peringati Hak atas Perlindungan Konsumen Bagi kami,pengguna jasa karantina pertanian yg laporkan produk pertanian yg dilalulintaskan di tanah air adlh pahlawan kelestarian SDA Hayati Ã°Ã?Â?Â®Ã°Â?Â?Â© Slamat awali pekan,ttp #DirumahAja Karimin #SapaPagi https://t.co/yEymQZdz16	jakarta
Prihatin, instruksi Pak @jokowi & semua pihak ga digubris. Presiden JKW semakin dilematis utk lakukan lockdown atau karantina, sbb arus mudik sdh terlanjur terjadi. #CoronaUpdate Pandemi Corona, 1 Kecamatan di Brebes Kedatangan 923 Pemudik dari Jakarta https://t.co/2Qfcu877CA	jakarta
16.15 Check Point PSBB Kamal Jakut, menghimbau agar masyarakat tetap mengunakan Masker & mematuhi PSBB guna mencegah penularan Covid-19. https://t.co/Irw2o73xa3	jakarta

bayangin dong nangi ada atsd gw,, sumpah nangis ampe karantina selese kali	jakarta
@geanawl_ Ayok kasih tau warganya biar sadar dari per individualnya, lockdown bukan satu2nya jalan	jakarta
Check point PSBB di Kota Bandung akan tersebar di perbatasan hingga pusat kota. #kumparanNEWS https://t.co/WWeNZquq3p	jakarta
@AhmadZulfian4 @RadityaChavvah @budimandjatmiko @inovator4id @ainunnajib @amflife @mantriss @sganjara Profesor sekaligus Direktur Center Suistainable Jeffrey Sachs mengatakan salah satu penyebab AS menjadi epicentrum corona krn sikap pemerintah AS terutama presiden Trump yang intensinya adalah menyelamatkan perekonomian ? Telat Lockdown	jakarta
@asumsico Mudah2an dng PSBB bs mengurangi penularan	jawa barat
@raayisunal Walah serius? Tdnya mau pulang naek motor, tp yaa gitu sepanjang perbatasan banyak pemeriksaan, mau local lockdown jg kayanya	jawa barat
ternyata udah hari ke-35 karantina	jawa barat
@CallMejutek @Canda_Bisnis2 @MymarrowQueen @GabusKnight_13 @Dyandradie @Adr_Nnaa @ajiesara2 @MaryanaNunung @tandanVradio @AriestaRiico @cintada17 @KTabahawa @FloraSafira Jauuhhh ciiinnn....gegara kereta nya penuh PSBB Ã°Â?Â?Â? Ã°Â?Â¤Â£Ã°Â?Â¤Â£Ã°Â?Â¤Â£	jawa barat
Pertama Kali, Pemandangan Janggal di Rumah Amitabh Bachchan dan Shah Rukh Khan sejak India Lockdown https://t.co/YkWI5CHoEa	jawa barat
Sby-Gresik-Doarjo siap" mau psbb, alamat gaiso balek malang Ã°Â?Â?Â¥	jawa timur
@itssnotyours Biar di karantina Ã°Â?Â?Â?	jawa timur
@PartaiSocmed @ainunnajib Lonjakan di SG terjadi karna yg terjangkit Foreign worker dr BL yg stay d dormitry. Dormitry dgn kapasitas ribuan, terjangkit 2orng langsung sigap di karantina seluruhnya, dan di test semua. Langsung PSBB, yg bandel kena fine $300-10k bahkan penjara. Indonesia? Jauh :((	jawa timur
tips & trick menjalani psbb bagi yg rumahnya tidak dipasangi wifi	jawa timur

Local Lockdown hari ini Jumat (27/3/2020) pada pukul 19.00 WIB telah resmi berlaku di kawasan Tunjungan Surabaya. . . Sejumlah pengendara motor terlihat masih banyak yang belum mengetahui kebijakan lokalÃ¢Â?Â¦ https://t.co/JhRkndmluq	jawa timur
Rakyat lagi yg harus menderita #PSBB https://t.co/xrDdMY3Jtc	yogyakarta
Keluar2 karantina corona rambut gondrong, brewok kemana2. Fix di kira napi lepas ini	yogyakarta
Dear bapak Presiden Joko Widodo. Harapan saya, kalau nanti corona sudah selesai.. Tolong banget berlakukan peraturan atau Himbauan jangan langsung keluar rumah.. Seenggaknya kita karantina diri sampe 14 hari setelah orang terakhir sembuh.. #CoronaIndonesia #LawanCovid19	yogyakarta
Like kampungku portal udh ditutup ditulisi lockdown, dipasangi alat semprot disinfektan, tp bapak2 srawung main catur sangu kopi neng ngarep portal. Untuk apaaaa https://t.co/VaFVrEhOjE	yogyakarta
@Iseasyazxx Gini lho kalo lockdown otomatis kan persebaran virus bisa dikontrol yg sakit diobati yg ga sakit ga tertular, negara mana italy amerika emng nerapin lockdown contohnya wuhan cina lockdown berhasil.	jawa tengah
@kwaenchanayoo Dari rumah gitu Apa di kasih pas karantina ?	jawa tengah
@EarlytaRatna @achek_neggro @suara_dahlanis @Dennysiregar7 @jokowi Cuma teori saja buat yg lengin lockdown, prakteknya nol besar,, gimana bisa menghidupkan ekonomi,, matinya ekonomi ya matinya kehidupan...,,, itu teori orang yg kejam, tidak memperhatikan nasib orang yg bawah.	jawa tengah
Hotel Jimbarwana Hampir Penuh, Pemkab Jembrana Jajaki Hotel Melati https://t.co/IIDJnbNTcB #Karantina #berita #nusabali #PemkabJembrana #COVID19 #VirusCorona #HotelJimbarwana #PekerjaMigranIndonesia #TKI	bali
Karantina/Lockdown gk mau, stop aliran masuk TKA China gak brani, janji angsuran libur setahun trnyata zonk. Jadi doi ituh lg kampanye atau lg bersikap sbagai Kepala Negara sih?? >> #KesadaranMemicuPerubahan << #RezimCuciTangan >> https://t.co/6uAbPK3RfB	bali

Kadang miris lihatnya... nyalahin pemerintah tapi baru habis dari luar negeri ga karantina mandiri dan jalan2 kemana2 termasuk ngempu ponakan baru 1 tahun. Sekarang pacarnya ponakan semua yg kontak sama dia positif. Ya dia positif.	bali

Lampiran 2 *Contoh Tweet Hasil Preprocessing*

tweet
tangerang raya mulai besok puluh check point sebar
haha kagak blay panjang bulan gua kurung dikosan
panjang animal crossing harap update update ciamikmu teman
presiden joko widodo jokowi resmi larang masyarakat laku aktivitas mudik hari raya idul fitri wilayah jabodetabek wilayah telah terap status batas sosial skala besar
plis pengin ulang tahun terus pintu kamar ketok teman masuk terus bilang surprise ini cuma prank ulang tahun lo biar stay rumah kaget kan
pedoman guna kendara lama di jakarta kliniksepekan
banyak warga tak patuh covid kota depok
sobatq hari april hari ingat hak atas lindung konsumen kamipengguna jasa tani lapor produk tani dilalulintaskan tanah air pahlawan lestari sda hayati selamat awal pekanttp dirumahaja karimin
prihatin instruksi pak amp semua pihak enggak gubris presiden jkw makin lat laku atau sebab arus mudik lanjur jadi pandemi corona camat brebes datang mudik jakarta
check point kamal jakut menghimbau masyarakat tetap mengunakan masker amp patuh guna cegah tular covid
bayang nang atsd sumpah menang karantina selese kali
ayok kasih tau warga biar sadar per individual bukan satu jalan
check point kota bandung sebar di batas hingga pusat kota
profesor sekaligus direktur center suistainable jeffrey sachs kata salah satu sebab as jadi epicentrum corona sikap perintah as utama presiden trump intensi selamat ekonomi telat
jakarta kalo lockdown barang olshop masih nyampek enggak
tinggi jumlah kasus pasien konfirmasi positif coronavirus disease covid buat pemkot bogor mulai pikir lebih ekstra progresif antisipasi virus mati
minggu iii beli semua suplemen vitamin a ada zinc kalsium madu susu murni cemilan badan
makin panjang makin rawan aman banyak tak punya hasil moga perintah sadar bahaya balik ini
pilih atau darurat sipil apa sama turut kamus bapak jangan bicara rakyat pak bilang para dukung bapak sendiri tentu semua rakyat sepakat twit bapak enggak usah seret nama rakyat jangan udang balik bakwan pak

kota malang jadi wilayah pertama jatim aju dprd jatim minta wilayah hitung matang aju
wib laku batas sosial skala besar periksa pas rebo begini putri citeureup sentul selatan begitu bogor ciawi arah masukkeluar jakarta maksimal jumlah tumpang jumlah kursi guna masker pergi
india terap untuk tekan angka sebar virus corona bijak buat chaos dalam negeri hingga perdana menteri narendra modi minta maaf
andai jakarta prasetio minta pemprov dki alih anggar formula e
milik toko minta patuh atur laku

Lampiran 3 *Data Trainer ygang memberi label*

	Data Trainer	
1	Nama	Ahmad Nur Fauzi
	Pekerjaan	Mahasiswa (UIN Sunan Kalijaga)
	Umur	22
	Jenis Kelamin	Laki-laki
	Agama	Islam
	Suku	Jawa
2	Nama	Maria Josephine Vivian Chang
	Pekerjaan	Mahasiswa (Sanata Dharma)
	Umur	21
	Jenis Kelamin	Perempuan
	Agama	Katholik
	Suku	Chinese
3	Nama	Muhyidin
	Pekerjaan	Guru
	Umur	47
	Jenis Kelamin	Laki-laki
	Agama	Islam
	Suku	Jawa
4	Nama	Kristina Weningtyas
	Pekerjaan	Mahasiswa (Poltekkes)
	Umur	22
	Jenis Kelamin	Perempuan
	Agama	Katholik
	Suku	Jawa
5	Nama	Bela Kurnia Davis
	Pekerjaan	Mahasiswa (UIN Sunan Kalijaga)
	Umur	21
	Jenis Kelamin	Perempuan
	Agama	Islam
	Suku	Jawa (Sumatera)

Lampiran 4 *Contoh Tweet Hasil Labeling secara Manual*

Tweet	**Lokasi**	**Fauzi**	**Apin**	**Bapak**	**Wening**	**Bela**

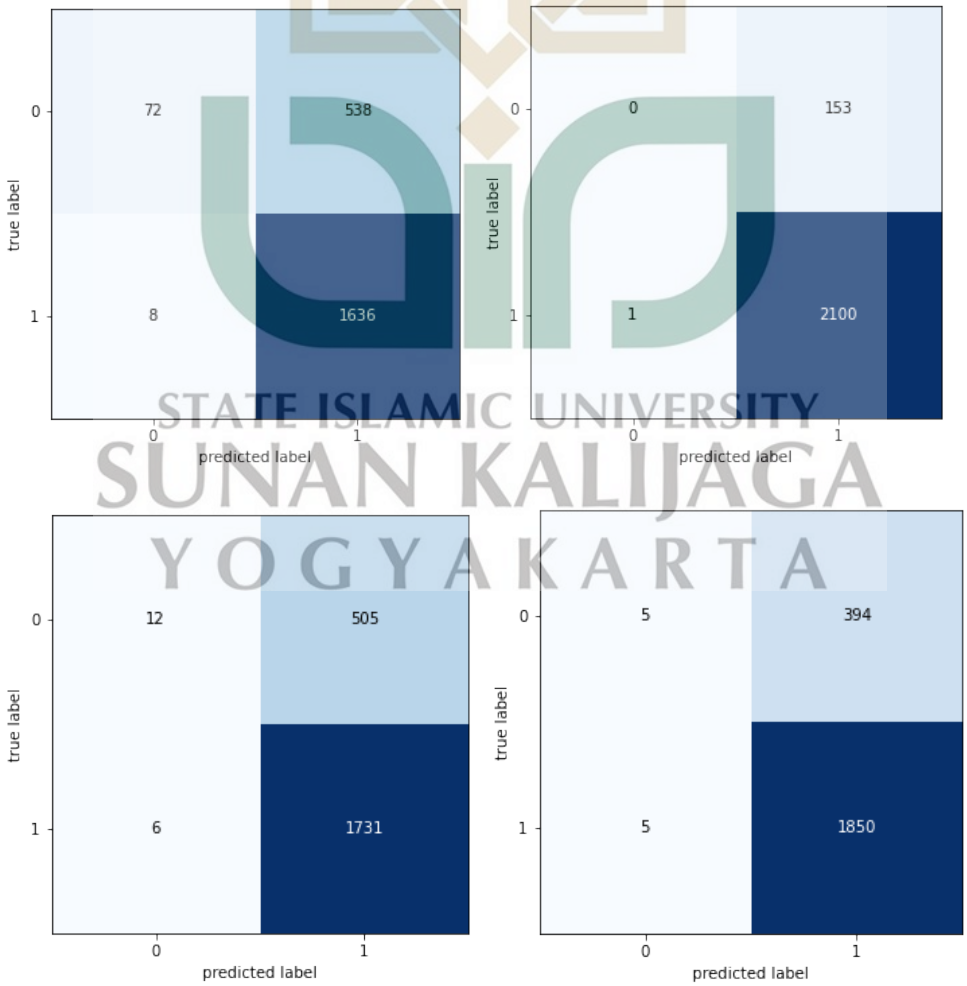
PSBB Tangerang Raya Dimulai Besok, Puluhan Check Point Disebar https://t.co/ncntW4us3r	jakarta	1	1	1	1	1
@ben_khair Haha kaga blay.... psbb diperpanjang, 2 bulan nih gua kekurung dikosan Ã°Â?Â?Â?	jakarta	0	0	1	0	1
psbb diperpanjang. animal crossing, kuharapkan update update ciamikmu untuk temani.	jakarta	0	0	0	0	0
Presiden Joko Widodo (Jokowi) resmi melarang masyarakat untuk melakukan aktivitas mudik saat hari raya Idul Fitri untuk wilayah Jabodetabek, serta wilayah yang telah menerapkan status pembatasan sosial berskala besar (PSBB). https://t.co/Ulinq9AKUQ	jakarta	0	1	1	1	1
Plis pengen ulang tahun terus pintu kamar di ketok, temen2 pada masuk terus bilang "SURPRISE, karantina ini cuma prank ulang tahun lo biar stay dirumah!!! Kaget kann???"	jakarta	1	1	0	1	1
16. Pedoman Penggunaan Kendaraan Selama PSBB di Jakarta https://t.co/WgNJ9WuNk3 Ã¢Â?Â? (17.04.20) #KlinikSepekan	jakarta	1	1	1	1	1
Masih Banyak Warga Tak Patuh PSBB Covid-19 di Kota Depok https://t.co/PsIxUt9FAC https://t.co/ptXDNOvHhq	jakarta	0	0	1	0	0

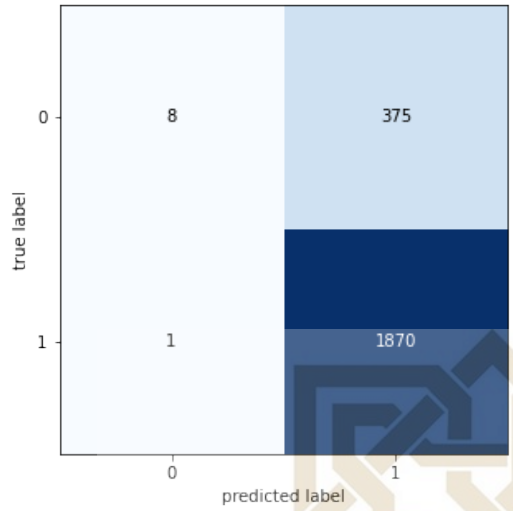
SobatQ, hari ini 20 April jadi hari tuk peringati Hak atas Perlindungan Konsumen Bagi kami,pengguna jasa karantina pertanian yg laporkan produk pertanian yg dilalulintaskan di tanah air adlh pahlawan kelestarian SDA Hayati Ã°Â?Â?Â®Ã°Â?Â?Â© Slamat awali pekan,ttp #DirumahAja Karimin #SapaPagi https://t.co/yEymQZdz16	jakarta	1	1	0	1	0
Prihatin, instruksi Pak @jokowi & semua pihak ga digubris. Presiden JKW semakin dilematis utk lakukan lockdown atau karantina, sbb arus mudik sdh terlanjur terjadi. #CoronaUpdate Pandemi Corona, 1 Kecamatan di Brebes Kedatangan 923 Pemudik dari Jakarta https://t.co/2Qfcu877CA	jakarta	0	0	0	0	1
16.15 Check Point PSBB Kamal Jakut, menghimbau agar masyarakat tetap mengunakan Masker & mematuhi PSBB guna mencegah penularan Covid-19. https://t.co/Irw2o73xa3	jakarta	1	1	0	1	1
bayangin dong nangi ada atsd gw,, sumpah nangis ampe karantina selese kali	jakarta	0	0	0	0	0
@geanawl_ Ayok kasih tau warganya biar sadar dari per individualnya, lockdown bukan satu2nya jalan	jakarta	1	0	1	0	1
Check point PSBB di Kota Bandung akan tersebar di perbatasan hingga pusat kota.	jakarta	1	1	1	1	1

#kumparanNEWS https://t.co/WWeNZquq3p						
@AhmadZulfian4 @RadityaChavvah @budimandjatmiko @inovator4id @ainunnajib @amflife @mantriss @sganjara Profesor sekaligus Direktur Center Suistainable Jeffrey Sachs mengatakan salah satu penyebab AS menjadi epicentrum corona krn sikap pemerintah AS terutama presiden Trump yang intensinya adalah menyelamatkan perekonomian ? Telat Lockdown	jakarta	0	0	1	1	0
Ini JKT kalo ke lockdown barang olshop gw masih bisa nyampek gak ya? Ã°Â?Â?Â?	jakarta	0	0	0	1	1
Tingginya jumlah kasus pasien terkonfirmasi positif Coronavirus Disease (Covid-19), membuat Pemkot Bogor mulai berpikir lebih ekstra progresif dalam mengantisipasi virus mematikan itu.Ã?Â #Bogor https://t.co/lszgzMRZap	jakarta	1	1	1	1	1
Karantina minggu III. Beli semua suplemen, vitamin A, C, D3, Zinc, Kalsium, Madu, Susu Murni, cemilan. Badan udah	jakarta	1	1	1	1	0
Semakin panjang PSBB, semakin rawan keamanan krn banyak yg tak punya penghasilan.. Semoga pemerintah sadar bahaya dibalik PSBB ini.. https://t.co/uwjUXi4Rp9	jakarta	0	0	1	0	1

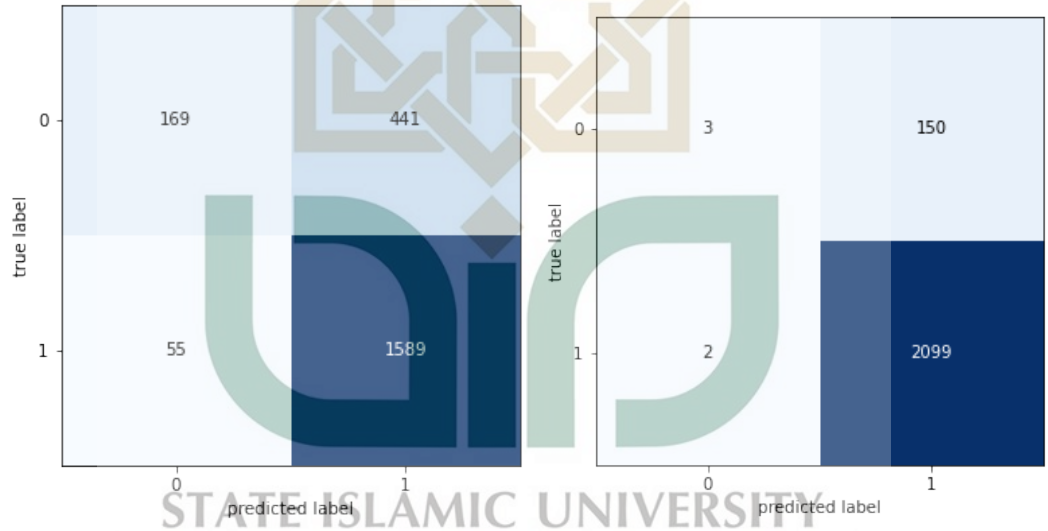
@msaid_didu Jadi @msaid_didu pilih LOCKDOWN atau DARURAT SIPIL ? Apa itu sama aja menurut kamus bapak ? Jgn bicara rakyat pak bilang aja para pendukung bapak sendiri belum tentu semua rakyat sepakat dgn twit bapak Jdi ngak usah seret2 nama rakyat Jgn ada udang dibalik bakwan pak ?	jakarta	0	0	1	0	1

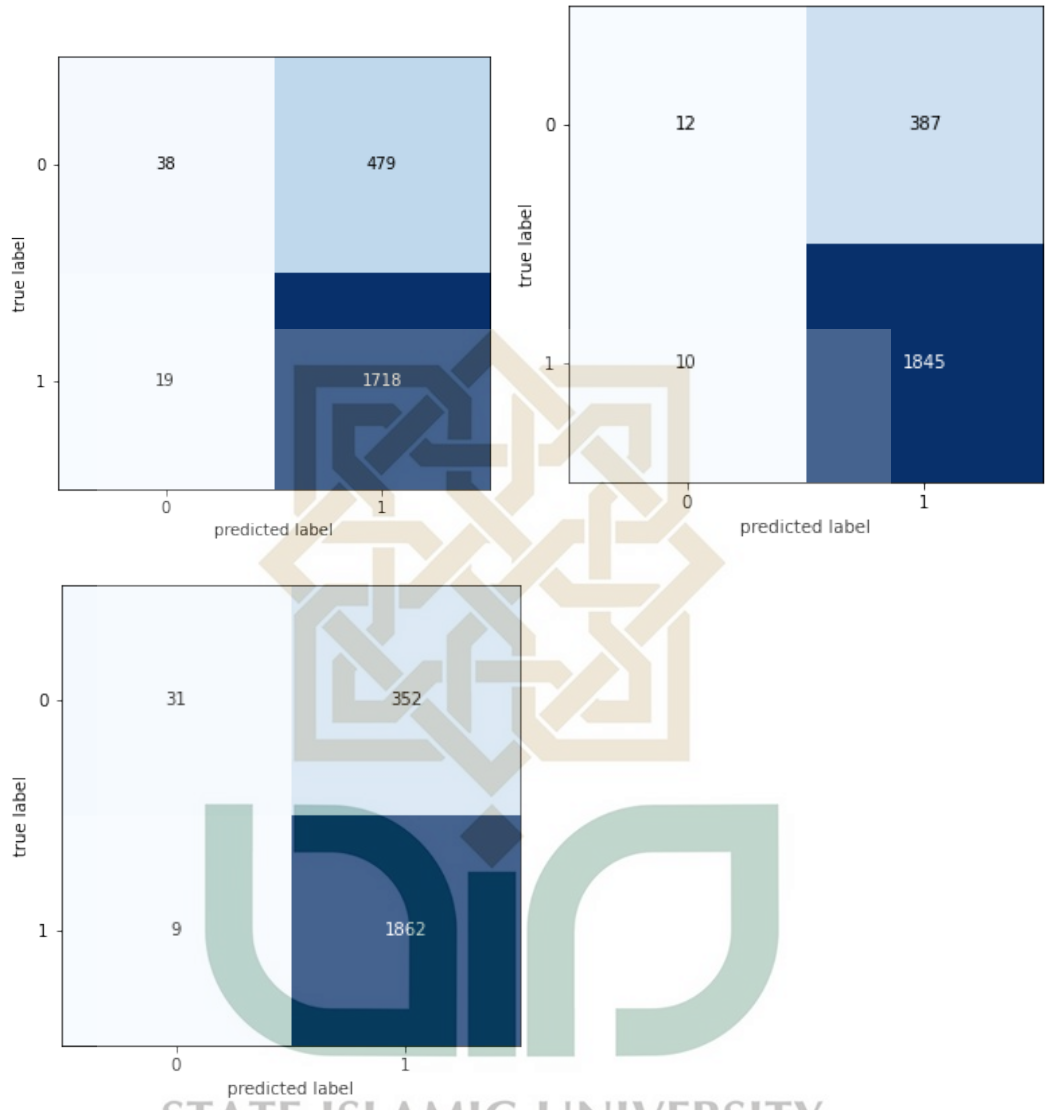
Lampiran 5 *Contoh Hasil Evaluasi dengan Confusion matrix*



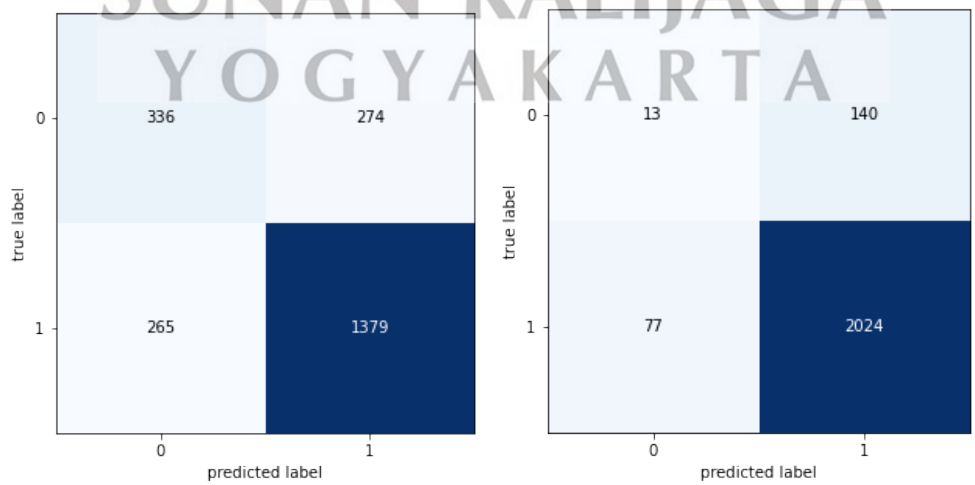


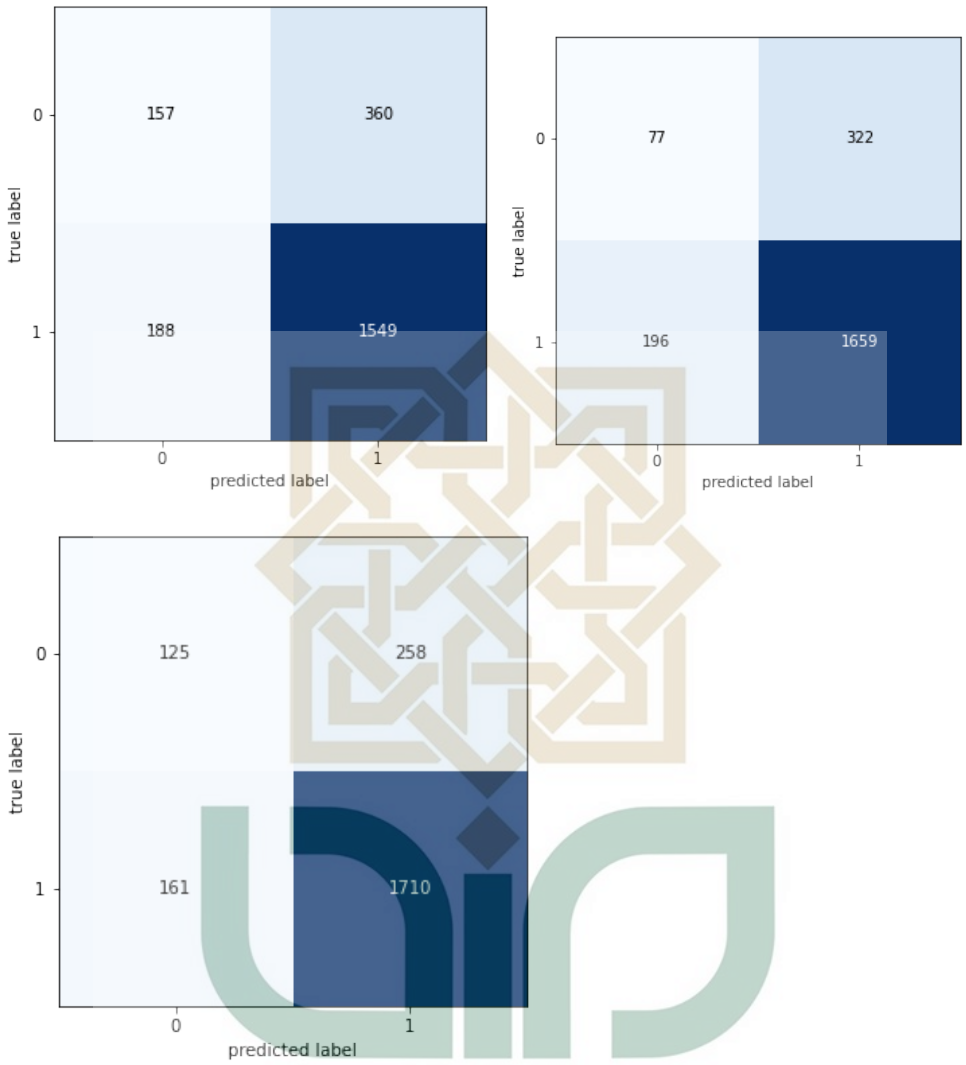
Gambar 0.1 NBC TF-IDF trainer 1 sampai 5



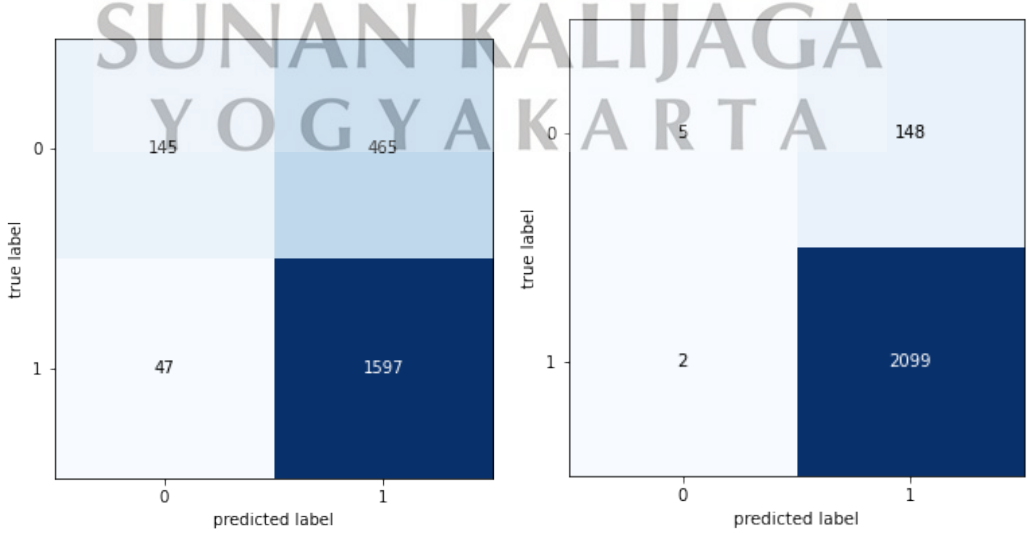


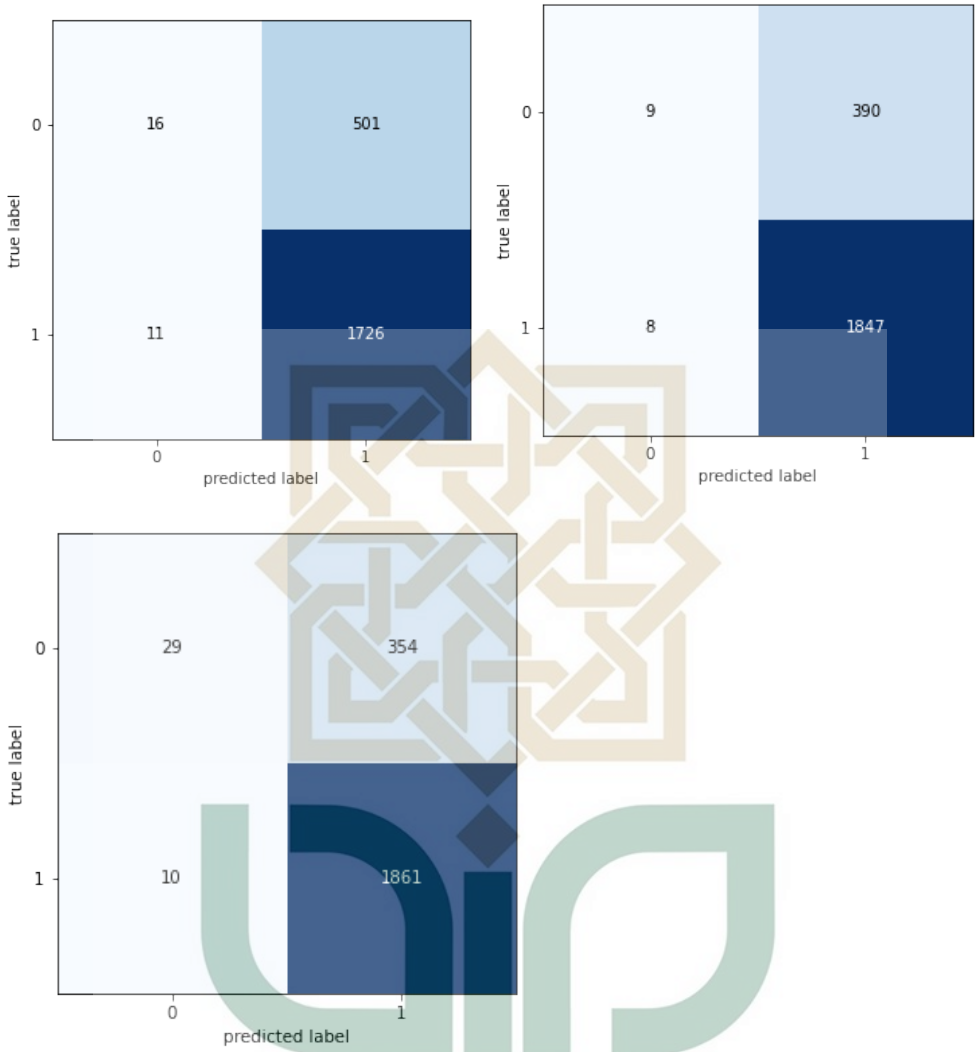
Gambar 0.2 SVM TF-IDF trainer 1 sampai 5





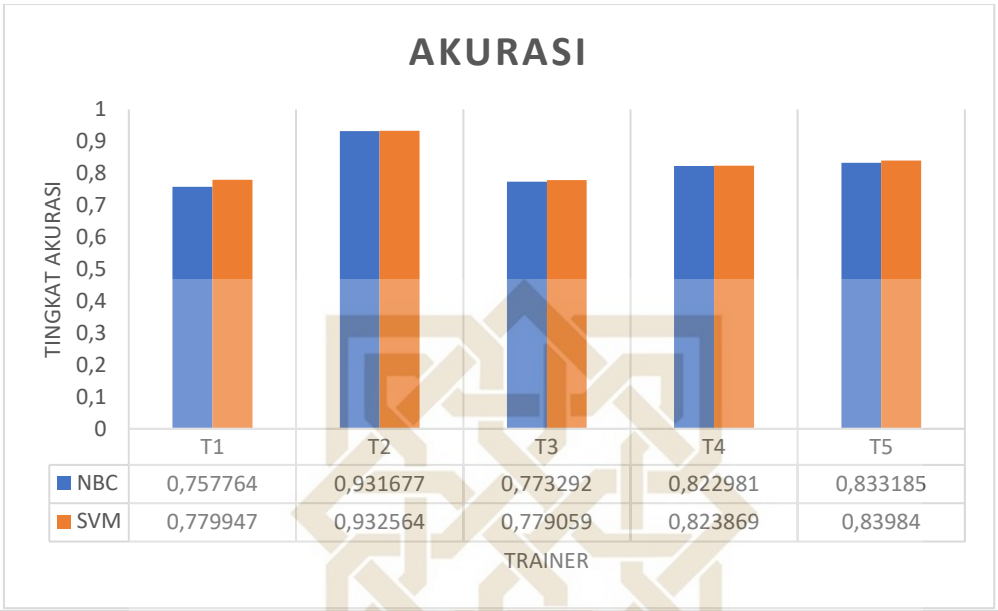
Gambar 0.3 NBC Term Frequency trainer 1 sampai 5

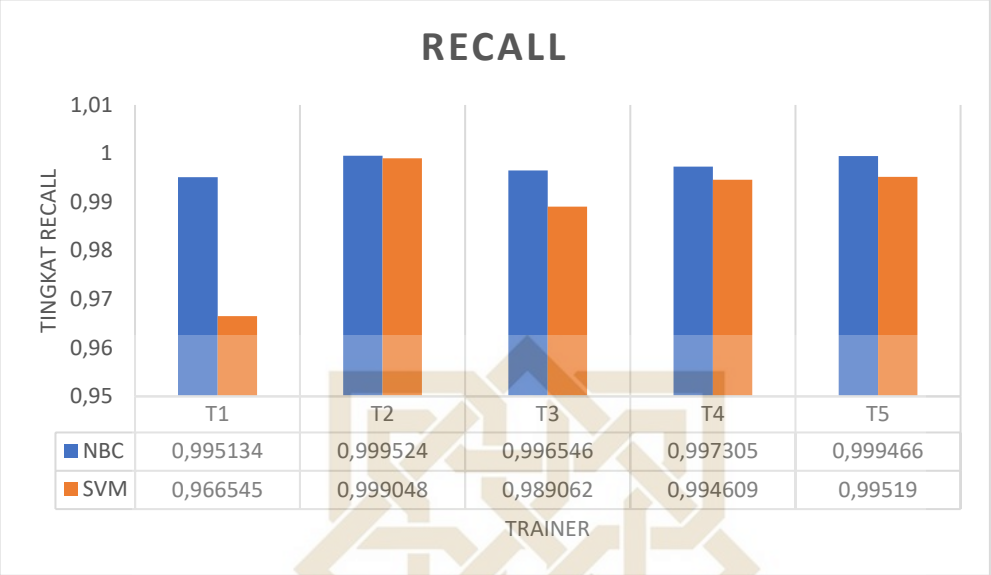


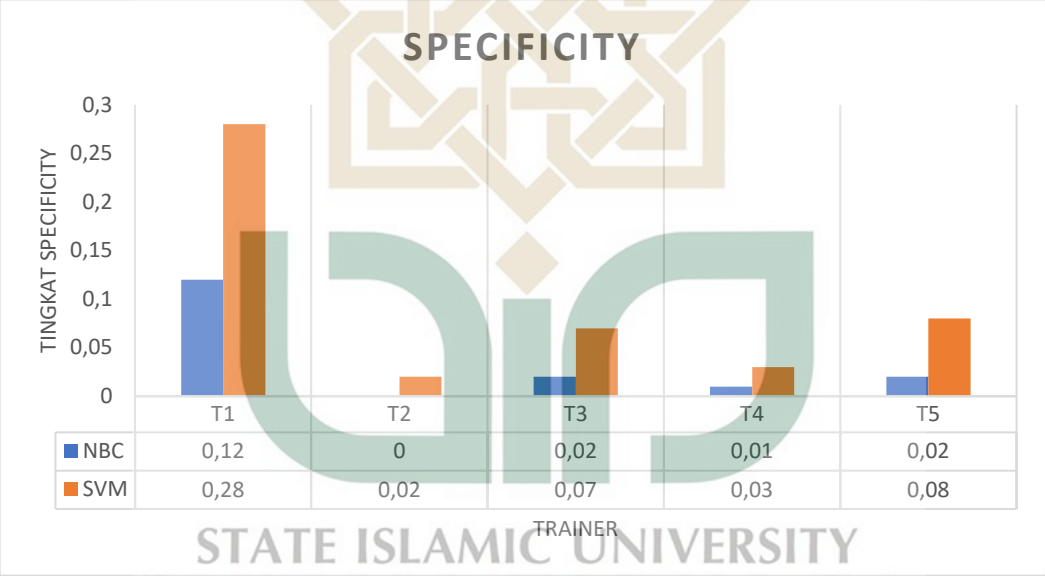


Gambar 0.4 SVM Term Frequency trainer 1 sampai 5

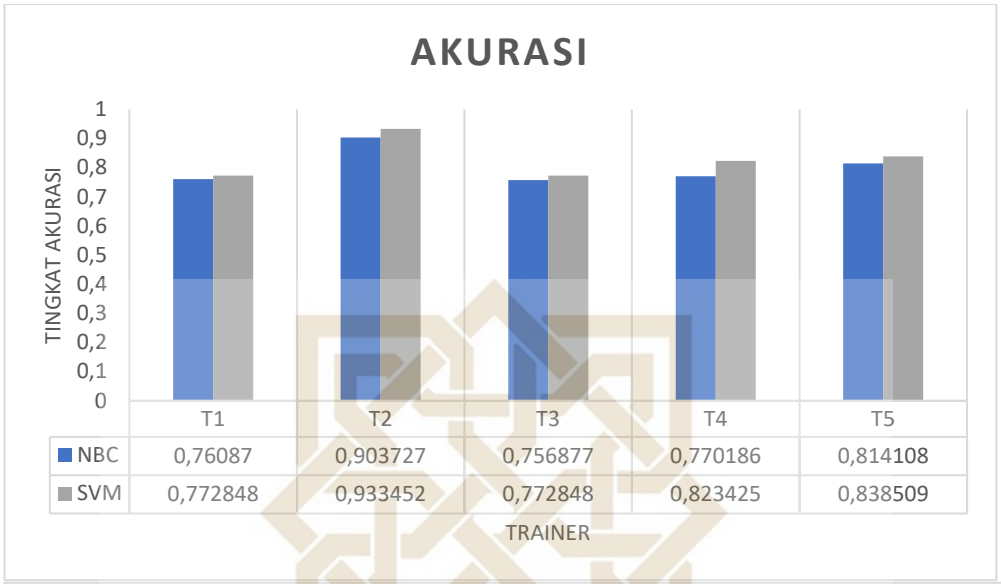
Lampiran 6 *Perbandingan hasil Confusion Matrix TF-IDF*

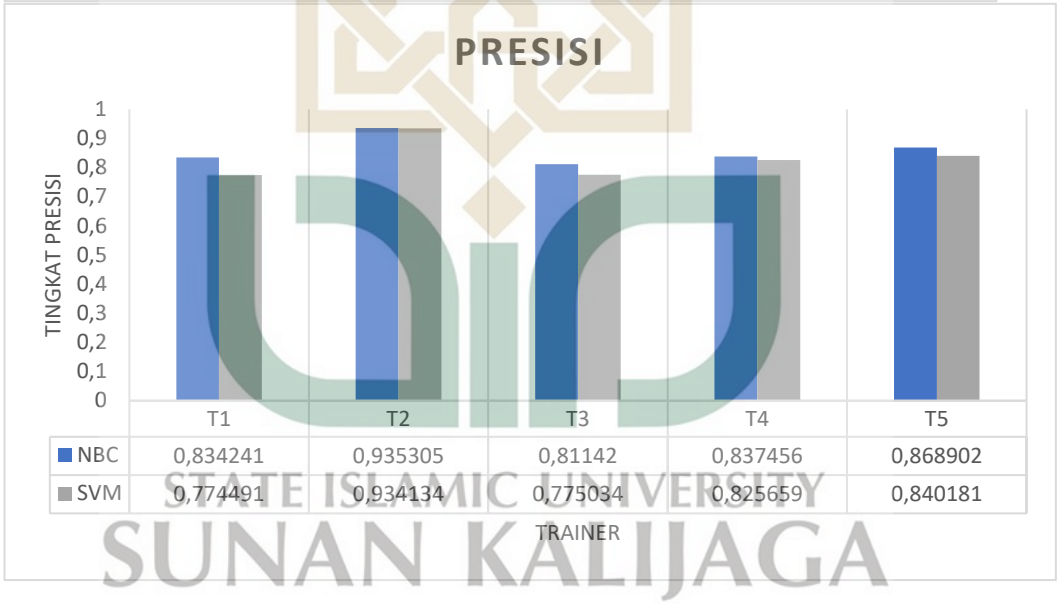


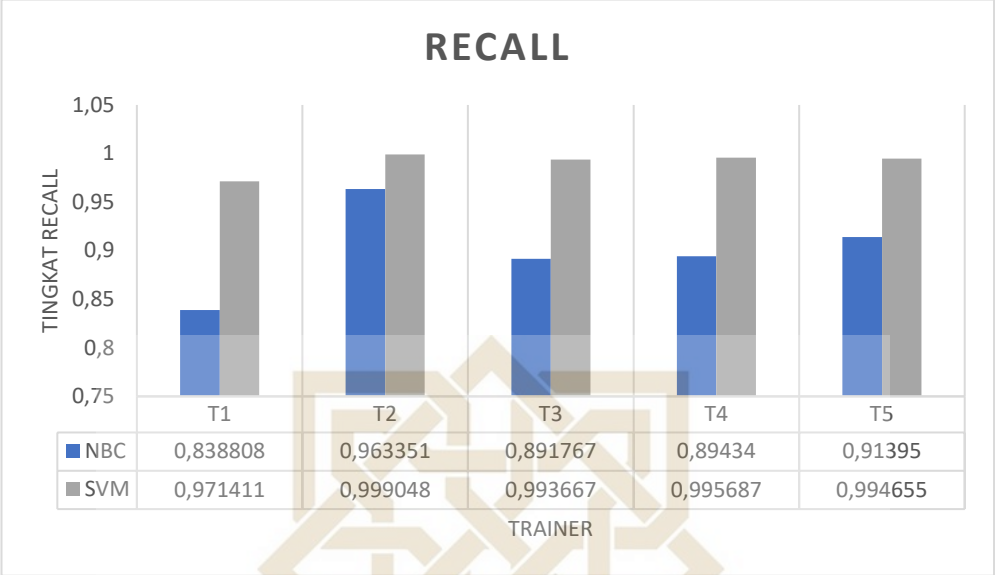


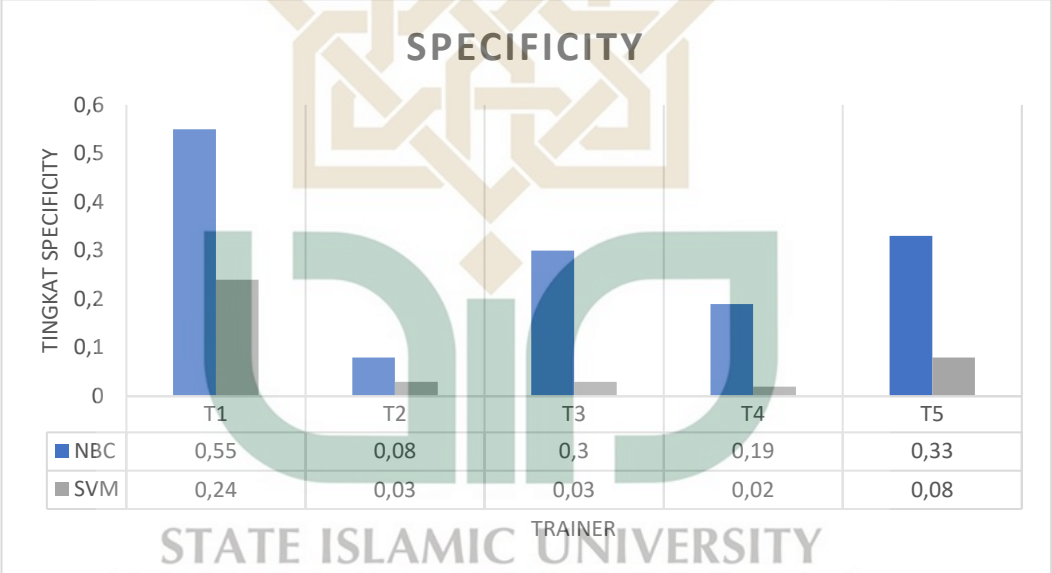


Lampiran 7 *Hasil perbandingan dengan TF menggunakan Confusion Matrix*









Lampiran 8 *Contoh Source code Program*

```
import pandas as pd
import string
from collections import Counter
import nltk
from nltk.tokenize import sent_tokenize
from nltk.tokenize import word_tokenize
import numpy as np
import matplotlib.pyplot as plt
import re
from Sastrawi.StopWordRemover.StopWordRemoverFactory import
StopWordRemoverFactory, StopWordRemover, ArrayDictionary

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
stemmer=StemmerFactory().create_stemmer()

from sklearn.feature_extraction.text import TfidfVectorizer as
tfidf
from sklearn.model_selection import train_test_split

from sklearn.feature_extraction.text import CountVectorizer as CV

from sklearn.metrics import classification_report ,
accuracy_score
%matplotlib inline
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix
from mlxtend.plotting import plot_confusion_matrix
```

```
def get_word_frequency(tweets, count_once_per_tweet=False):

    count_all = Counter()
    for tweet in tweets:
        tokens = word_tokenize(tweet)
        terms_all = [term for term in tokens]
        if count_once_per_tweet:
            terms_all = set(terms_all)
        count_all.update(terms_all) # update the counter
    return count_all

def plot_word_frequency(data, color='k', title='Twitter Word
Count'):
    words, counts = zip(*data)
    val = counts[::-1]   # sort the bars from longest to shortest
    words = words[::-1]
    pos = np.arange(len(data))+.5  # center the bars on the y axis

    plt.figure(figsize=(12,12))
    plt.barh(pos, val, align='center', color=color)
    plt.yticks(pos, words)
    plt.xlabel('Count')
    plt.title(title)
    plt.grid(True)
    plt.show()

def clean(x):
    x = str(x).lower()
    x = re.sub(r'@\w+ ', '', x) #remove mention
    x = re.sub(r'#\w+ ', '', x) #remove hastag
    x = re.sub(r'([^a-z\s])|(\w+:\/\/\S+)', ' ', x) #remove number,
punctuation, and url
    x = re.sub(r'\n+', ' ', x) #remove new line
    x = re.sub(r' +',' ',x).strip() #remove whitespaces
    return x

stop_factory = StopWordRemoverFactory().get_stop_words()
more_stopword =
['lockdown','psbb','karantina','rt','yang','shg','dlm',
'utk','krn', 'dengan', 'di', 'jg', 'scr', 'knp', 'dmn', 'nih',
'dong', 'aja', 'jd', 'jadi', 'kok', 'gue', 'aku', 'sama', 'tp',
'pd','sih']
# Merge stopword
stopword = stop_factory + more_stopword

dictionary = ArrayDictionary(stopword)
r_sw = StopWordRemover(dictionary)
```

```
def slang(x):
    x = str(x).lower()
    x = word_tokenize(x)
    ul = 0
    for a in slangword['slang']:
        w = 0
        for kt in x:
            if kt == a:
                x[w] = slangword['formal'][ul]
            w = w + 1
        ul = ul + 1
    kal = ""
    for kl in x:
        kal = kal + " " +  kl
    x = kal
    return x

data=pd.read_csv('finalfixtweet1.csv',sep=';',encoding = "ISO-
8859-1")

val = []
idx = 0
for i in data.fauzi:
    if i+data['apin'][idx]+data['bapak'][idx]+data['wening'][idx]+
data['bela'][idx] >= 3:
        val.append(1)
    else:
        val.append(0)
    idx = idx + 1
val

data['value'] = val

# <h1>Import data test
test=pd.read_csv('finalTest.csv',sep=';',encoding = "ISO-8859-1")
test

# <h1>import data Slangword
slangword = pd.read_csv('colloquial-indonesian-
lexicon.csv',encoding = "ISO-8859-1")
slangword
cl_data=data
cl_data
```

```
cl_data.tweet=cl_data.tweet.apply(lambda x : clean(x))
cl_data.tweet.to_csv('cleansing.csv',encoding='utf-
8',index=False, sep=';')
cl_data.tweet

cl_data.tweet=cl_data.tweet.apply(lambda x : slang(x))
cl_data.tweet.to_csv('slangword.csv',encoding='utf-
8',index=False, sep=';')
cl_data.tweet

cl_test.tweet=cl_test.tweet.apply(lambda x : clean(x))
cl_test.tweet

cl_test.tweet=cl_test.tweet.apply(lambda x : slang(x))
cl_test.tweet

# <h2>Stemming
cl_data.tweet=cl_data.tweet.apply(lambda x : stemmer.stem(x))

cl_data.tweet.to_csv('stemming.csv',encoding='utf-
8',index=False, sep=';')
cl_data.tweet

cl_test.tweet=cl_test.tweet.apply(lambda x : stemmer.stem(x))
cl_test.tweet

cl_test.tweet.to_csv('stemmingtest.csv',encoding='utf-
8',index=False, sep=';')

cl_data.tweet = cl_data.tweet.apply(lambda x : r_sw.remove(x))
cl_data.tweet.to_csv('stopword.csv',encoding='utf-
8',index=False, sep=';')
cl_data.tweet

cl_test.tweet = cl_test.tweet.apply(lambda x : r_sw.remove(x))
cl_test.tweet

import nltk
DataStem = pd.read_csv('stemming.csv',sep=';',encoding = "ISO-
8859-1")
DataStemTest = pd.read_csv('stemmingtest.csv',sep=';',encoding = "
ISO-8859-1")
datatest = DataStemTest
```

```
vectorizer= tfidf(stop_words= stopword)       # using TfIdf to mak
e words as features by making word vectors
x= vectorizer.fit_transform(DataStem['tweet'].values.astype('U')).
toarray()
y= data['value']

# <H2>Term Frequency</h2>
countV = CV(stop_words= stopword)
ex = countV.fit_transform(DataStem['tweet'].values.astype('U')).to
array()
Y = data['value']

xtest = vectorizer.transform(DataStemTest['tweet'].values.astype('
U')).toarray()
extest = countV.transform(DataStemTest['tweet'].values.astype('U')
).toarray()

x_train,x_test,y_train,y_test= train_test_split(x,y,random_state=
42)
ex_train,ex_test,ey_train,ey_test= train_test_split(ex,y,random_st
ate= 42)
#[Train Test Data] membagi data menjadi 90% : 10%
xtrain90, xtest90, ytrain90, ytest90 = train_test_split(x, y,rando
m_state=15,test_size=0.1)
#[Train Test Data] membagi data menjadi 80% : 20%
xtrain80, xtest80, ytrain80, ytest80 = train_test_split(x, y,rando
m_state=15,test_size=0.2)
#[Train Test Data] membagi data menjadi 70% : 30%
xtrain70, xtest70, ytrain70, ytest70 = train_test_split(x, y,rando
m_state=15,test_size=0.3)
#[Train Test Data] membagi data menjadi 60% : 40%
xtrain60, xtest60, ytrain60, ytest60 = train_test_split(x, y,rando
m_state=15,test_size=0.4)
#[Train Test Data] membagi data menjadi 90% : 10%
extrain90, extest90, eytrain90, eytest90 = train_test_split(ex, Y,
random_state=15,test_size=0.1)
#[Train Test Data] membagi data menjadi 80% : 20%
extrain80, extest80, eytrain80, eytest80 = train_test_split(ex, Y,
random_state=15,test_size=0.2)
#[Train Test Data] membagi data menjadi 70% : 30%
extrain70, extest70, eytrain70, eytest70 = train_test_split(ex, Y,
random_state=15,test_size=0.3)
#[Train Test Data] membagi data menjadi 60% : 40%
extrain60, extest60, eytrain60, eytest60 = train_test_split(ex, Y,
random_state=15,test_size=0.4)
```

```
rom sklearn.naive_bayes import MultinomialNB,BernoulliNB
import memory_profiler
from memory_profiler import profile
from memory_profiler import memory_usage
import time
tfidfnaive = []
clf=MultinomialNB()
m1 = memory_profiler.memory_usage()
startNB = time.process_time()
clf.fit(xtrain90,ytrain90)
pred= clf.predict(xtest90)
memory = memory_profiler.memory_usage()[0]-m1[0]
timeNB = time.process_time() - startNB
print('memory =',memory)
print('process time =',timeNB)
print('accuracy=',accuracy_score(ytest90,pred))
print(classification_report(ytest90, pred))
print(confusion_matrix(ytest90,pred))
tfidfnaive.append({'train':'90/10','memory' : memory,'time' : time
NB,'accuracy' : accuracy_score(ytest90,pred)})
CM = confusion_matrix(ytest90, pred)
fig, ax = plot_confusion_matrix(conf_mat=CM ,  figsize=(10, 5))
plt.show()

from sklearn.svm import SVC
tfidfsvm = []
m1 = memory_profiler.memory_usage()
startsvm = time.process_time()
svclassifier = SVC(C=1,kernel='rbf')
svclassifier.fit(xtrain90, ytrain90)
pred = svclassifier.predict(xtest90)
memory = memory_profiler.memory_usage()[0]-m1[0]
timeSVM = time.process_time() - startsvm
print("time proses : " + str(timeSVM))
print('memory =',memory)
print('process time =',timeSVM)
print('accuracy=',accuracy_score(ytest90,pred))
print(classification_report(ytest90, pred))
print(confusion_matrix(ytest90,pred))
tfidfsvm.append({'train':'90/10','memory' : memory,'time' : timeSV
M,'accuracy' : accuracy_score(ytest90,pred)})
CM = confusion_matrix(ytest90, pred)
fig, ax = plot_confusion_matrix(conf_mat=CM ,  figsize=(10, 5))
plt.show()
```

```
naiveB = pd.DataFrame(tfidfnaive, columns=['train','memory','time'
,'accuracy'])
naiveB
SVm = pd.DataFrame(tfidfsvm, columns=['train','memory','time','acc
uracy'])
SVm
plt.plot( 'train', 'accuracy', data=naiveB, marker='o', markerface
color='blue', markersize=12, color='skyblue', linewidth=4,label="N
aive")
plt.plot( 'train', 'accuracy', data=SVm, marker='o', markerfacecol
or='orange', markersize=12, color='yellow', linewidth=4,label="SVM
")
plt.title('Accuracy of method')
plt.legend()
plt.show

from sklearn.model_selection import KFold
from sklearn.model_selection import cross_validate
from sklearn.model_selection import cross_val_score
from sklearn.metrics import make_scorer, accuracy_score, precision
_score, recall_score, f1_score
kf = KFold(n_splits=10, shuffle=True, random_state=42)
clf = MultinomialNB()
scoring = {'accuracy' : make_scorer(accuracy_score),
           'precision' : make_scorer(precision_score),
           'recall' : make_scorer(recall_score),
           'f1_score' : make_scorer(f1_score)}
tfidfnaivescore = cross_validate(clf, x, Y,cv=kf,scoring=scoring)
naiveBayKfold = pd.DataFrame(tfidfnaivescore, columns=['fit_time',
'score_time','test_accuracy','test_precision','test_recall','test_
f1_score'])
naiveBayKfold.to_csv('naiveBayKfold.csv',encoding='utf-
8',index=False, sep=';')
naiveBayKfold['K'] = ['1','2','3','4','5','6','7','8','9','10']
tfidfsvmscore = cross_validate(svclassifier, x, y,cv=kf,scoring=sc
oring)
svmKfold = pd.DataFrame(tfidfsvmscore, columns=['fit_time','score_
time','test_accuracy','test_precision','test_recall','test_f1_scor
e'])
svmKfold.to_csv('svmKfold.csv',encoding='utf-
8',index=False, sep=';')
svmKfold['K'] = ['1','2','3','4','5','6','7','8','9','10']
```

CURRICULUM VITAE

A. Biodata Diri

Nama Lengkap : Ahmad Nur Fauzi

Jenis Kelamin : Laki-laki

Tempat, Tanggal Lahir : Sleman, 4 Maret 1998

Alamat Asal : Bangkong, Wukirsari, Cangkringan, Sleman, DIY.

Alamat Tinggal : Bangkong, Wukirsari, Cangkringan, Sleman, DIY.

E-mail : ozianf001@gmail.com

B. Latar Belakang Pendidikan Formal

Jenjang	Nama Sekolah	Tahun
TK	TK NEGERI 2 SLEMAN	2003 – 2004
SD	SD NEGERI KIYARAN 1	2004 – 2010
SMP	SMP NEGERI 1 PAKEM	2010 – 2013
SMA	SMA NEGERI 2 YOGYAKARTA	2013 – 2016