

Analisis Sentimen Terhadap Twitter Direktorat Jenderal Bea dan Cukai Menggunakan komparasi Algoritma Naïve Bayes dan Support Vector Machine

Dedi Dwi Saputra^{1*}, Riza Fahlapi², Antonius Yadi Kuntoro³, Hermanto⁴, Taufik Asra⁵

¹ Teknologi Informasi , Universitas Siber Indonesia, Kota Jakarta Selatan, Prov. D.K.I. Jakarta, Indonesia

² Teknologi Informasi, Universitas Bina Sarana Informatika, Jl. Ringroad Barat, Gamping Kidul, Ambarketawang, Kec. Gamping, Kabupaten Sleman, Daerah Istimewa Yogyakarta, 55184, Indonesia

³ Sistem Informasi, Universitas Nusa Mandiri, Jl. Jatiwaringin No. 2, Cipinang Melayu, Makasar Jakarta Timur, 13620, Indonesia

⁴ Sistem Informasi, Universitas Bina Sarana Informatika, Jl. Ringroad Barat, Gamping Kidul, Ambarketawang, Kec. Gamping, Kabupaten Sleman, Daerah Istimewa Yogyakarta, 55184, Indonesia

⁵ Rekayasa Perangkat Lunak, Universitas Bina Sarana Informatika, Jl. Ringroad Barat, Gamping Kidul, Ambarketawang, Kec. Gamping, Kabupaten Sleman, Daerah Istimewa Yogyakarta, 55184, Indonesia

***Email Korespondensi:**

dedi.dwi.s@cyber-univ.ac.id

Abstrak

Direktorat Jenderal Bea & Cukai (DJBC) merupakan lembaga instansi pemerintah yang bertugas mengawal dan melayani kegiatan ekspor dan import di Indonesia sejak berdiri pada tahun 1946 yang diharapkan oleh masyarakat sebagai garda terdepan dalam mengayomi masyarakat dalam bidang ini, namun beberapa belakangan ini banyak sekali kasus yang melibatkan institusi Direktorat Jenderal Bea & Cukai yang membuat institusi ini dapat mempengaruhi pandangan kinerja institusi ini. Dengan deskripsi permasalahan diatas sangat menarik untuk dilakukan penelitian pandangan masyarakat menggunakan tweet dari twitter @bravobeacukai dan @beacukaiRI yang dimiliki dan di olah Direktorat Jenderal Bea & Cukai sebagai tempat penyaluran opini dan pandangan masyarakat terhadap institusi ini. Penelitian ini menggunakan metode Smote dengan Naïve Bayes dan dikomparasi dengan metode Support vector machine untuk hasil tersebut dapat dibandingkan tingkat akurasi. Hasil dari penelitian ini didapat bahwa menggunakan metode Smote dengan Naïve Bayes didapati nilai Accuracy 78.95%, Precision 74.01%, Recall 89.41%, dan AUC 0.650 sedangkan untuk metode Smote dengan Support vector machine bernilai Accuracy 73.35%, Precision 67.88%, Recall 88.95%, dan AUC 0.853. Berdasarkan hasil penelitian, metode Smote dengan Naïve Bayes mempunyai hasil terbesar dan bernilai efektif dengan dataset yang diteliti.

Kata Kunci : Naïve Bayes; Twitter; Smote; Support Vector Machine

Abstract

Direktorat Jenderal Bea & Cukai (DJBC) is a government agency in charge of guarding and serving export and import activities in Indonesia since its establishment in 1946 which is expected by the community as the front guard in protecting the community in this field, but in recent times there have been many cases involving the institution of the Directorate General of Customs & Excise which make this institution can affect the view of the performance of this institution. With the description of the problem above, it is very interesting to conduct research on public views using tweets from twitter @bravobeacukai and @beacukaiRI which are owned and processed by Direktorat Jenderal Bea & Cukai as a place to channel public opinions and views on this institution. This research uses the Smote method with Naïve Bayes and compared with Support vector machine methods for these results to compare the level of accuracy. The results of this study found that using the Smote method with Naïve Bayes

©2024 J-Intech. Published by LPPM STIKI Malang

This is an open access article under the CC BY SA license. (<https://creativecommons.org/licenses/by-sa/4.0/>)

obtained an Accuracy value of 78.95%, Precision 74.01%, Recall 89.41%, and AUC 0.650 while for the Smote method with Support vector machine is worth 73.35% Accuracy, Precision 67.88%, Recall 88.95%, and AUC 0.853. Based on the research results, the smote method with Naïve Bayes has the greatest results and is effective with the dataset studied.

Keyword : Naïve Bayes; Twitter; Smote; Support Vector Machine

1. Pendahuluan

Seiring dengan kemajuan teknologi dan informasi, social media tidak lagi hanya menjadi sampingan melainkan sudah menjadi bagian penting dalam kehidupan masyarakat saat ini, social media tidak hanya digunakan untuk bersosialisasi dan berinteraksi bisa juga sebagai media mempromosikan jasa dan produk ataupun sebagai sumber media data masukan dan opini terhadap suatu produk, jasa bahkan instansi pemerintahan dimana kasus ini kami mengangkat salah satu instansi pemerintahan yaitu Direktorat Jenderal Bea Cukai (DJBC) sebagai instansi yang bertanggung jawab dan berwenang atas pengawasan dan pemeriksaan terhadap barang-barang yang masuk atau keluar dari negara baik sebagai impor dan ekspor Bea cukai adalah sumber pendapatan negara yang signifikan di Indonesia. Oleh karena itu, penting untuk memastikan bahwa proses pemungutan bea cukai dilakukan secara efektif dan efisien untuk memaksimalkan pendapatan negara.

Dalam beberapa tahun terakhir, Bea Cuka telah memanfaatkan social media sebagai wadah pemberian informasi tentang peraturan dan prosedur perihal bea dan cukai yang memberikan hal positif kepada masyarakat, namun Dirjen Bea Cukai (DJBC) tidak luput dari beberapa hal negatif salah satu keNEGATIFAN nya yang kami akan bahas adalah dimana informasi perihal kebijakan yang abu abu dan performa kinerja pimpinan dan anggota bea cukai yang dipertanyakan yang membuat opini dan kepercayaan masyarakat dalam dunia social media menjadi negatif dan menurun, ditambah dengan kasus menghebohkan beberapa waktu yang lalu perihal biaya bea cukai terhadap barang gratis atau peimbeiriaan tanpa ada harga tetap yang tidak wajar dan banyaknya kasus dugaan korupsi dari para petinggi dan anggota Dirjen Bea Cukai(DJBC) (Hardiantoro & Firdaus, n.d.).

Dalam penelitian ini bertujuan untuk mengukur beberapa opini dan masukan masyarakat dalam social media terhadap kinerja atau berita tentang bea cukai apakah positif maupun negatif dengan menggunakan twitter dengan @bravobeacukai dan @beacukaiRI sebagai kunci pencarian tentang data yang kami gunakan dalam melakukan penelitian pada jurnal ini, peneliti menggunakan pendekatan komparatif antara algoritma Naïve Bayes yang ditambahkan metode smote dan dibandingkan dengan Algoritma Support Vector Machine (SVM). Menurut jurnal (Yudistira & Wicaksana, 2022), Algoritma Support Vector Machine dapat menjadi solusi yang efektif untuk melakukan klasifikasi suatu data yang akan masuk kedalam kategori positif atau negative (2 class). Tujuan uji kompartif algoritma Naïve Bayes yang ditambahkan metode smote dan dibandingkan dengan Algoritma Support Vector Machine (SVM) ini berguna untuk mencari metode yang paling efektif agar dapat digunakan untuk membangun model klasifikasi yang berdasarkan pada analisis statistik dari data historis hasil mining data twitter sosial media Dirjen Bea Cukai (DJBC). Harapannya dengan melakukan uji komparatif dengan menggunakan kedua model tersebut dapat mengetahui metode yang efektif dan mempunyai nilai akurasi yang tinggi, agar dapat di implmentasikan ke Bea Cukai agar dapat membuat keputusan yang lebih akurat dan efisien dalam menentukan kebijakan dan aturan didalam korporasi instansi yang ada didalam Dirjen Bea Cukai (DJBC).

Berdasarkan masalah tersebut diatas, maka penulis tertarik untuk melakukan penelitian uji komparasi dengan judul : " Analisis Sentimen Terhadap Twitter Direktorat Jenderal Bea dan Cukai Menggunakan Perbandingan Algoritma Naïve Baye dan Support Vector Machine". Mengacu terhadap penelitian sebelum jurnal ini dibuat, telah banyak dilakukan penelitian mengenai pengukuran analisa sentimen antara beberapa algortima seperti Naive Bayes, K-Nearest Neighbor, Decision Tree dan SVM (Support Vector Machine) (Firmansyah, Gaol, & Susilo, 2019)(Supriyatna & Mustika, 2018)(Devita, Herwanto, & Wibawa, 2018), yang dimana analisis tersebut memiliki hasil yang mengarahkan antara Naïve Bayes dan SVM (Support Vector Machine) memiliki tingkat akurasi tinggi.

Tweet adalah sebuah teks yang diketik oleh sebuah user dalam memposting suatu hal didalam platform twitter, dimana mayoritas kegunaan twitter digunakan untuk saling berbagi informasi via internet namun terkadang ada tweet yang berbentuk opini atau kritik terhadap suatu hal oleh karena deskripsi ini menjadikan tweet sebagai target data dalam penulisan ini, dimana untuk mengelola dan mengambil data dari tweet dari twitter penulis menggunakan Rapidminer yang merupakan platform analisis data bebas pemrograman, yang memungkinkan pengguna untuk merancang proses analisis data dengan cara plug-and-play dengan memasang operator. Disisi lain, fungsionalitas dapat ditambahkan ke RapidMiner dengan mengembangkan ekstensi, yang tersedia di RapidMiner Marketplace. Ekstensi RapidMiner Linked Open Data menambahkan operator untuk memuat data dari set data dalam Linked Open Data, serta secara otonom mengikuti tautan RDF ke set data lain dan mengumpulkan data tambahan dari sana. Selain itu, ekstensi ini mendukung pencocokan skema untuk data yang dikumpulkan dari dataset yang berbeda (Ristoski, Bizer, & Paulheim, 2015). Penelitian memfokuskan diri uji komparatif pada teknik Natural Language Processing yang merupakan serangkaian teknik komputasi yang dimotivasi oleh teori untuk analisis dan representasi otomatis bahasa manusia. Penelitian NLP telah berevolusi dari era kartu punch dan pemrosesan batch (di mana analisis sebuah kalimat dapat memakan waktu hingga 7 menit) hingga era Google dan sejenisnya (di mana jutaan halaman web dapat diproses dalam waktu kurang dari satu detik). Penelitian ini mengacu pada perkembangan terbaru dalam penelitian NLP untuk melihat masa lalu, saat ini, dan masa depan teknologi NLP dengan sudut pandang yang baru. Meminjam paradigma 'kurva lompatan' dari bidang manajemen bisnis dan prediksi pemasaran, artikel survei ini menafsirkan evolusi penelitian NLP sebagai perpotongan tiga kurva yang tumpang tindih yaitu Kurva Sintaksis, Semantik, dan Pragmatik-yang pada akhirnya akan mengarahkan penelitian NLP untuk berevolusi menjadi pemahaman bahasa alami (Cambria & White, 2014). Dalam analisis sentimen, memilih algoritma yang tepat sangatlah penting karena setiap algoritma memiliki kekuatan dan kelemahannya masing-masing. Dimana pada penelitian ini ketiga algoritma Naive bayes dan Support Vector Machine (SVM) dengan menggunakan metode SMOTE untuk mengatasi masalah ketidakseimbangan data akan memberikan gambaran yang lebih komprehensif mengenai metode mana yang paling efektif untuk analisis sentimen terhadap data twitter Dirjen Bea Cukai (DJBC). Hasil perbandingan ini diharapkan dapat meningkatkan akurasi dan efisiensi dalam pengambilan keputusan terkait kebijakan dan aturan di dalam Dirjen Bea Cukai (DJBC).

2. Metode Penelitian

Metode penelitian yang digunakan adalah metode deskriptif dengan pendekatan kuantitatif. Metode deskriptif ini melibatkan pengumpulan data untuk menguji hipotesis atau menjawab pertanyaan tentang pendapat orang atas sebuah isu atau topik. Penelitian kuantitatif merupakan penelitian yang didasarkan pada pengumpulan dan analisis data berbentuk angka (numerik) untuk menjelaskan, memprediksi, dan mengontrol fenomena yang diminati. Dengan metode kuantitatif akan diperoleh signifikansi hubungan antar variabel dengan menggunakan sumber data primer berupa jawaban kuesioner dari Twitter @bravobeacukai dan @beacukaiRI yang diperoleh secara online dan data sekunder yang berasal dari instansi terkait guna mendukung analisis. Responden merupakan Tweet dari User Twitter se-Indonesia sehingga jumlah respon adalah 2000 Tweet yang dianggap mewakili masyarakat (Dharmawan, Raharjo, & Kusumawardhani, 2019).

Saat ini Pesan teks bahasa Indonesia dari media sosial seperti Twitter atau Facebook, orang cenderung menggunakan kata-kata tidak formal daripada yang formal seperti menggunakan angka untuk mengganti alfabet, karakter berulang vokal, dan menggunakan kata-kata informal yang umum untuk menggantikan kata-kata resmi. Untuk memproses kata-kata seperti itu, maka harus dilakukan tahapan-tahapan *preprocessing*.

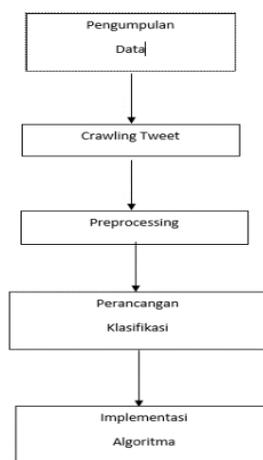
Studi tentang klasifikasi teks dengan menggunakan algoritma Naive Bayes bukanlah sebuah pendekatan baru. Pada penelitian-penelitian sebelumnya, Naive Bayes telah menunjukkan performa yang unggul dalam tugas klasifikasi teks, terutama karena kesederhanaan dan efisiensinya dalam menangani data dengan jumlah fitur yang besar. Misalnya, penelitian yang dilakukan oleh (Jiang et al, 2009) dalam 'Text Classification Using Naive Bayes' menggarisbawahi keunggulan algoritma ini dalam mengelompokkan dokumen berdasarkan topik secara efektif.

Penelitian terbaru menunjukkan bahwa baik Naive Bayes maupun Support Vector Machine (SVM) tetap menjadi metode yang kompetitif dalam tugas klasifikasi teks. Dalam konteks klasifikasi spam email, Naive Bayes mencapai akurasi sebesar **89.6%** sebagaimana dilaporkan oleh (Abiodun & Omotosho, 2021), yang menekankan efektivitas algoritma ini dalam menangani teks dengan fitur yang bervariasi (Khan & Kamaruddin, 2020) juga mencatat bahwa Naive Bayes, ketika dikombinasikan dengan teknik pengurangan fitur, dapat mencapai akurasi **87.3%** dalam klasifikasi teks pendek.

Di sisi lain, SVM menunjukkan performa yang sangat baik, dengan melaporkan akurasi **93.7%** dalam klasifikasi teks legal, dan (Singh, 2016) menemukan bahwa SVM mampu mencapai akurasi **91.2%** dalam analisis sentimen Twitter, menunjukkan keunggulannya dalam dataset yang tidak seimbang. Dalam deteksi spam di media sosial, (Al-Qurishi, Rana, & Rajarajan, 2019) menemukan bahwa Naive Bayes mencapai akurasi **90.4%**, sementara algoritma SVM mencapai akurasi **92.5%** dalam klasifikasi email spam, menggarisbawahi kemampuan SVM dalam memisahkan kelas yang tumpang tindih.

Selain itu, (Jain & Gupta, 2022) melaporkan pada jurnalnya akurasi **85.9%** dari Naive Bayes dalam analisis sentimen Twitter, yang menggarisbawahi relevansinya dalam teks pendek dan dinamis. Sebaliknya, bahwa algoritma SVM berhasil mencapai akurasi **88.9%** dalam klasifikasi teks multi-label, menegaskan efektivitasnya dalam tugas yang lebih kompleks. Terakhir, baik Naive Bayes maupun SVM menunjukkan performa yang baik dalam konteks teks khusus; Naive Bayes dengan akurasi **82.4%** dalam analisis sentimen Twitter (Sangeetha & Vijayakumar, 2018) dan SVM dengan akurasi **90.1%** dalam klasifikasi teks. Didalam penelitian (Bagus Panuluh Yudha Perkasa & Eka Purwiantono, n.d.) metode Naive Bayes Gaussian berhasil mengklasifikasikan data dengan akurasi mencapai 87,85%. Kombinasi hasil ini menunjukkan bahwa baik Naive Bayes maupun SVM memiliki keunggulan masing-masing dalam berbagai aplikasi klasifikasi teks, tergantung pada konteks dan kompleksitas tugas yang dihadapi.

Pada rancangan desain penelitian ini dimulai dengan secara teknis membicarakan masing-masing bagian konstruksi desain penelitian seperti judul penelitian, latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, tinjauan pustaka, hipotesis, konsep-konsep penelitian, penentuan variabel dan indikator variabel, pengukuran, sumber data, metode pengumpulan data, rancangan analisis dan metode analisis data (Haqiqi, Santika, & Yusmalina, 2021).

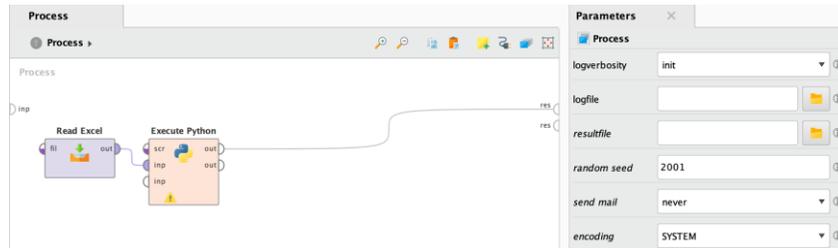


Gambar 1. Tahapan Penelitian

Pada penelitian ini, Digunakan data Sekunder yang diambil dari akun Twitter resmi @bravobeacukai dan @beacukaiRI. Data yang diambil akan dijadikan sebagai data training untuk membuat kamus opini. Pembuatan kamus data opini tersebut dengan mengumpulkan data opini melalui media data online. Kamus data opini yang dikumpulkan diklasifikasi menjadi kamus opini negatif dan positif. Untuk melakukan klarifikasi peneliti menggunakan bantuan secara online maupun offline untuk melakukan klarifikasi agar data kamus sesuai dan

meminimalkan subjektifitas.

satu ini dilakukan melalui bantuan website *gataframework*. Diawali dengan membuat format excel 2003 dengan aturan kolom yang terdiri dari No, Text dan Status dimana akan diisi dari data hasil metode pengumpulan data sebelumnya lalu membuat koding python untuk dapat mengupload data ke website untuk diproses didalam rapidminer.



Sumber: *gataframework.com* (2017)
Gambar 2. Struktur Operator di Rapidminer

```
Edit Parameter Text: script
The python script to execute.

1 import pandas
2 import urllib.parse
3 import urllib.request
4 import time
5
6 urlgataframework = "http://www.gataframework.com/textmining/index.php?model=transaction_text&action=processTextPublic&techniques="
7
8 def sendToGataFrameworkTextMining(text, teknik, counttime):
9     text = urllib.parse.quote_plus(text);
10    url = urlgataframework + teknik + "&textbefore=" + text
11    try:
12        textresult = urllib.request.urlopen(url)
13        text = textresult.read().decode('utf-8')
14    except:
15        if (counttime <= 3): #diakses sampai 4 kali percobaan
16            counttime += counttime
17            text = sendToGataFrameworkTextMining(text, teknik, counttime)
18        else:
19            text = "time out"
20
21    time.sleep(.03)
22    return text
23
24 def rm_main(data):
25
26    for id, row in data.iterrows():
27        data.at[id,'remove_annotation'] = sendToGataFrameworkTextMining(data.at[id,'text'],'9', 0) #Remove @annotation
28        data.at[id,'remove_hashtag'] = sendToGataFrameworkTextMining(data.at[id,'remove_annotation'],'15', 0) #Remove Hashtag
29        data.at[id,'remove_url'] = sendToGataFrameworkTextMining(data.at[id,'remove_hashtag'],'7', 0) #Remove URL
30        data.at[id,'regexp'] = sendToGataFrameworkTextMining(data.at[id,'remove_url'],'8', 0) #Remove Regexp
31        data.at[id,'indonesian_stemming'] = sendToGataFrameworkTextMining(data.at[id,'regexp'],'3', 0) #Stemming
32        data.at[id,'indonesian_stopword_removal'] = sendToGataFrameworkTextMining(data.at[id,'indonesian_stemming'],'2', 0) #Stopword
33
34    return data
```

Sumber: *gataframework.com* (2017)
Gambar 3. Code Python untuk Pre-Processing tahap 1 via *gataframework*

Dalam tahap ini *gataframework* melakukan 5 teknik pembersihan yaitu (Syah et al., 2023):

1. @ Anotation Removal
2. Transformation Remove URL
3. Regexp
4. Stemming (Indonesian Dictionary)
5. Stop word Removal ((Indonesian Dictionary)

Yang membantu dalam pembersihan data dari simbol-simbol, tautan web, *emoticon*, bermakna ganda dan kata yang tak berarti yang memabntu meningkatkan persentasi akurasi dan presisi data

Tabel 1. Contoh Kosakata dalam Proses Stemming

Sebelum	Sesudah
Berlebihan	Lebih
Bicarakan	Bicara
Dianggap	Anggap

Tabel 2. Contoh Kosakata Stopword

Yang	Tentu
Di	Tapi
Apkh	Dan
Ini	Itu

Kemudian setelah data selesai melakukan Pre-Processing satu selanjutnya akan melakukan Pre-processing dua menggunakan aplikasi Rapidminer dalam melakukannya dimana akan memfokuskan kepada Stemming dan Stop word, diawali dengan proses stemming dimana kosakat imbuhan bahasa indonesia dihilangkan pada bagian awal dan akhir kosakata demi meningkatkan akurasi dan presisi kosakata dalam dataset, dilanjutkan dengan melakukan stop word dimana kosakata yang tidak berpengaruh kedalam data set yang bertujuan sama seperti stemming.

Dalam kegiatan analisis sentimen ini menggunakan dataset tweet dari twitter @bravobeacukai dan @beacukaiRI dilakukan dengan metode Smote yang digabungkan dengan 2 algoritma yang ingin di bandingkan yaitu Naive Bayes dan Supprt Vector Machine.

Dalam jurnal (Yulita, 2021a) dijelaskan, bahwa algoritma Naive Bayes adalah salah satu metode yang digunakan dalam machine learning dan statistika untuk klasifikasi. Algoritma ini didasarkan pada Teorema Bayes, yang menghubungkan probabilitas kondisi suatu peristiwa dengan probabilitas kondisi lainnya.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (1)$$

dimana:

$P(A|B)$ adalah probabilitas hipotesis A benar diberikan bahwa bukti B telah terjadi.

$P(B|A)P(B|A)P(B|A)$ adalah probabilitas bukti B terjadi jika hipotesis A benar.

$P(A)P(A)P(A)$ adalah probabilitas a priori dari hipotesis A.

$P(B)P(B)P(B)$ adalah probabilitas dari bukti B.

Naïve Bayes merupakan Algoritma yang menggunakan perhitungan probabilitas yang menggunakan konsep Bayes. Penggunaan teorema Bayes pada algoritma Naïve Bayes adalah dengan menyatukan probabilitas awal dan probabilitas bersyarat dalam suatu rumus yang dapat digunakan untuk menghitung kemungkinan dari setiap kemungkinan klasifikasi (Yulita, 2021a).

Dalam jurnal (Pravina, Cholissodin, & Adikara, 2019) Support Vector Machine (SVM) adalah salah satu algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. SVM paling dikenal karena kemampuannya dalam memecahkan masalah klasifikasi dengan mencari hyperplane terbaik yang memisahkan kelas-kelas yang berbeda dalam data.

$$\text{Minimize} = \frac{1}{2} \|w\|^2 \text{ subject to } y_i(w \cdot x_i + b) \geq 1 \text{ for all } i \quad (2)$$

dimana:

ω adalah vektor bobot yang menentukan orientasi hyperplane.

β adalah bias atau offset dari hyperplane.

γ_i adalah label kelas untuk data x_i

©2024 J-Intech. Published by LPPM STIKI Malang

This is an open access article under the CC BY SA license. (<https://creativecommons.org/licenses/by-sa/4.0/>)

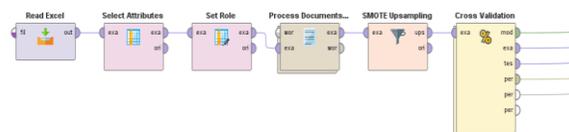
Sedangkan Support Vector Machine (SVM) Machine Learning (supervised learning) yang dapat memprediksi kelas berdasarkan hasil dari proses pelatihan. Dengan melakukan pelatihan menggunakan data input berupa angka dan hasil ekstraksi fitur, maka diperoleh suatu pola yang akan digunakan dalam proses pelabelan (Pravina et al., 2019).

Evaluasi yang ada dalam penelitian ini digunakan dalam mendapatkan hasil akurasi, presisi, recall dan AUC sebuah metode yang akan diteliti yaitu Smote menggunakan Naive Bayes atau Support Vector Machine dalam perbandingan hasil yang paling maksimal dalam menggunakan dataset yang sudah diolah sebelumnya.

3. Hasil

Setelah melakukan persiapan maka hasil yang didapatkan dari penelitian ini. Dataset yang digunakan dalam penelitian ini adalah hasil tweet masyarakat dalam twitter di @bravobeacukai dan @beacukaiRI pada tahun 2023 di Twitter, dimana telah mengumpulkan sekitar 2000 dataset, lalu dalam pemrosesan labelisasi didapatkan hasil bahwa tweet yang termasuk kedalam class positif adalah 1285 tweet dan untuk class negatif yaitu 715 tweet.

Dalam bagian ini penulis akan menjelaskan hasil pengujian dari semua proses yang sudah disampaikan dalam penelitian ini, setelah melakukan Pre-Processing kedua, akan langsung melakukan analisis perbandingan antara Algoritma Naive Bayes dan Support Vector Machine (SVM) dengan melakukannya melalui Rapidminer dengan alur operator seperti dibawah ini



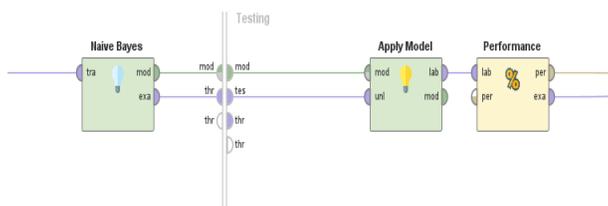
Gambar 4. Struktur Operator Pengujian di Rapidminer

Gambar 4 ini menunjukkan alur pengujian menggunakan RapidMiner untuk analisis sentimen. Alur ini mengintegrasikan operator pengambilan data, pemrosesan, dan pengklasifikasian data menggunakan algoritma yang diuji (Naive Bayes dan SVM). Struktur ini menjadi dasar evaluasi akurasi, presisi, recall, dan AUC yang dibandingkan pada penelitian ini.



Gambar 5. Struktur Operator Process Documents from Data di Rapidminer

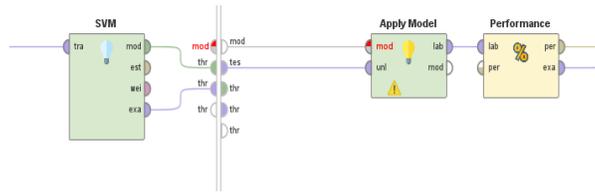
Gambar 5 menjelaskan langkah-langkah pemrosesan dokumen dari data mentah hingga menjadi input yang siap digunakan untuk algoritma klasifikasi. Tahapan ini melibatkan pembersihan data (preprocessing), seperti stemming dan stopword removal.



Gambar 6. Penggunaan metode Naive Bayes

Gamabr 6 ini mengilustrasikan penerapan algoritma Naive Bayes dalam RapidMiner untuk mengklasifikasikan

sentimen dari data Twitter. Naïve Bayes memiliki keunggulan dalam menangani data dengan fitur besar secara efisien.



Gambar 7. Penggunaan metode Support Vector Machine

Pada Gambar 7 memperlihatkan bagaimana SVM diterapkan sebagai pembanding dengan Naïve Bayes. SVM dikenal unggul dalam mengatasi data yang tidak seimbang, seperti pada dataset penelitian ini. Pada Penggunaan metode algoritma Naive bayes mendapatkan akurasi sebesar 78.95%, presisi 74.01%, Recall 89.41% dan AUC 0.650.

4. Pembahasan

Pada gambar 8 hingga 11, hasil pengujian Naïve Bayes menunjukkan akurasi sebesar 78.95%, presisi 74.01%, recall 89.41%, dan AUC 0.650. Hal ini menunjukkan bahwa algoritma ini efektif untuk analisis sentimen pada dataset yang digunakan.

Criterion	Value
accuracy	78.95%
precision	74.01%
recall	89.41%
AUC (optimistic)	0.650
AUC (pessimistic)	0.650

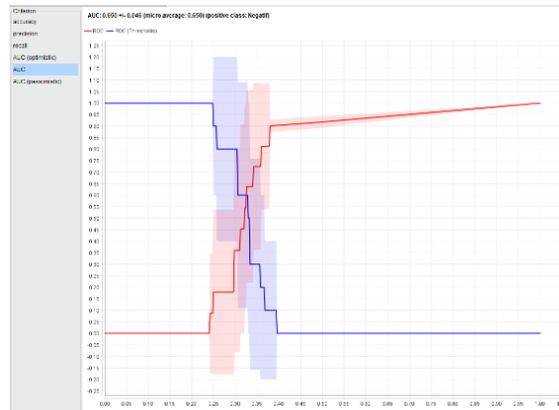
Gambar 8. Hasil Accuracy pada Metode Algoritma Naive Bayes

Criterion	Value
precision	74.01%
recall	89.41%
AUC (optimistic)	0.650
AUC (pessimistic)	0.650

Gambar 9. Hasil Precision pada Metode Algoritma Naive Bayes

Criterion	Value
recall	89.41%
precision	74.01%
AUC (optimistic)	0.650
AUC (pessimistic)	0.650

Gambar 10. Hasil Recall pada Metode Algoritma Naive Bayes



Gambar 11. Hasil AUC pada metode Algoritma Naive Bayes

Pada Gambar 12 hingga 15 menunjukkan hasil pengujian SVM mencatat akurasi sebesar 73.35%, presisi 67.88%, recall 88.95%, dan AUC 0.853. Meskipun SVM memiliki keunggulan pada AUC, hasil lainnya sedikit di bawah Naive Bayes.

Criterion: accuracy

accuracy: 73.35% +/- 3.37% (micro average: 73.35%)

	true Positif	true Negatif	class precision
pred. Positif	742	142	83.94%
pred. Negatif	543	1143	67.79%
class recall	57.74%	88.95%	

Gambar 12. Hasil Accuracy pada metode Algoritma SVM

Criterion: precision

precision: 67.88% +/- 3.33% (micro average: 67.79% (positive class: Negatif))

	true Positif	true Negatif	class precision
pred. Positif	742	142	83.94%
pred. Negatif	543	1143	67.79%
class recall	57.74%	88.95%	

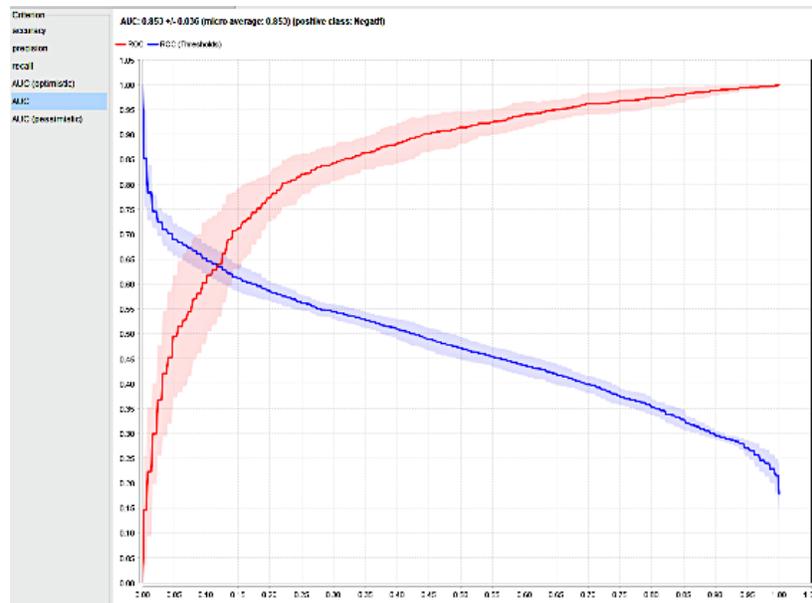
Gambar 13. Hasil Precision pada Metode Algoritma SVM

Criterion: recall

recall: 88.95% +/- 2.56% (micro average: 88.95% (positive class: Negatif))

	true Positif	true Negatif	class precision
pred. Positif	742	142	83.94%
pred. Negatif	543	1143	67.79%
class recall	57.74%	88.95%	

Gambar 14. Hasil Recall pada metode Algoritma SVM



Gambar 15. Hasil AUC pada metode Algoritma SVM

Dengan hasil penelitian diatas antara 2 algoritma ini dalam pengukuran Akurasi, Presisi dan Recall, dimana dapat dihasilkan algoritma Naive Bayes yang ditambahkan dengan smote lebih unggul daripada algoritma SVM, namun AUC Navie Bayes lebih rendah daripada SVM, dengan keunggulan dalam 3 parameter menjadikan Algoritma Naive Bayes yang ditambahkan smote lebih unggul daripada Algoritma Support Vector Machine (SVM).

Penelitian ini sejalan dengan studi lain yang membandingkan Naïve Bayes dan SVM dalam analisis sentimen. Naïve Bayes efektif untuk analisis sentimen Twitter dengan akurasi tinggi pada teks pendek (Toy, Sari, & Cholissodin, 2021; Yulita, 2021b). SVM unggul dalam memisahkan kelas data yang tidak seimbang, tetapi performa tergantung pada konfigurasi parameter dan dataset. (Lumbanraja, Gaol, Shofiana, & Junaidi, 2024; Qorry Meidianingsih, Wardani, Salsabila, Nafisah, & Mutia, 2023)

Dengan hasil tersebut maka, saran dari penelitian ini terhadap Direktorat Jenderal Bea dan Cukai, hasilnya dapat di implementasikan dengan melanjutkan ke tahap deployment modelling dengan mengimplementasikan model Algoritma Naïve Bayes yang ditambahkan metode SMOTE, Direktorat Jenderal Bea & Cukai (DJBC) dapat memperoleh pemahaman yang lebih akurat tentang bagaimana masyarakat melihat dan merespons kebijakan serta kinerja mereka di media sosial. Hal ini penting untuk mengidentifikasi isu-isu yang sering muncul dan meresponsnya dengan cepat dan dapat memperbaiki area-area yang menjadi perhatian publik dan meningkatkan transparansi, yang pada akhirnya dapat meningkatkan kepercayaan masyarakat terhadap instansi ini serta informasi berdasarkan hasil modelling ini klasifikasi ini dapat membantu dalam merancang kampanye informasi yang lebih tepat sasaran, menjelaskan kebijakan secara lebih efektif, dan menangani kritik serta isu negatif dengan lebih baik.

5. Kesimpulan

Hasil penelitian ini menunjukkan bahwa penggunaan metode SMOTE dengan Naive Bayes menghasilkan nilai Accuracy sebesar 78.95%, Precision sebesar 74.01%, Recall sebesar 89.41%, dan AUC sebesar 0.853. Sebaliknya, metode SMOTE dengan Support Vector Machine (SVM) menghasilkan nilai Accuracy sebesar 73.35%, Precision sebesar 67.88%, Recall sebesar 88.95%, dan AUC sebesar 0.650. Berdasarkan hasil penelitian ini, metode SMOTE dengan Naive Bayes menunjukkan performa terbaik dan dinilai paling efektif untuk dataset yang diteliti. Metode algoritma Naive Bayes memiliki keunggulan pada parameter yang dibandingkan, sehingga dinilai efektif terhadap dataset yang digunakan dalam penelitian ini.

Penelitian ini bertujuan untuk mempermudah analisis sentimen pada media sosial twitter Direktorat Jendral Bea & Cukai (DJBC). Oleh karena itu, dapat disimpulkan bahwa metode SMOTE dengan Naive Bayes adalah pilihan yang lebih unggul dibandingkan dengan metode SMOTE dengan SVM dalam konteks klasifikasi sentimen pada dataset yang digunakan. Diharapkan penelitian selanjutnya dapat memperluas analisis dengan mempertimbangkan penggunaan metode atau algoritma tambahan sebagai pembanding dan memasukkan label netral dalam analisis sentimen. Hal ini akan memberikan wawasan yang lebih mendalam dan berguna sebagai saran dan rujukan dalam penelitian berikutnya dengan mempertimbangkan variasi sumber dan referensi data yang berbeda.

Referensi

- Abiodun, O. I., & Omotosho, O. M. (2021). A Comparative Analysis of Naive Bayes and Support Vector Machines for Spam Classification. *Journal of Applied Sciences and Environmental Management*, 25(4), 633–639.
- Al-Qurishi, M., Rana, O., & Rajarajan, M. (2019). Real-time Spam Detection in Online Social Media Using Naive Bayes and Decision Tree. *International Journal of Machine Learning and Cybernetics*, 10, 2211–2221.
- Bagus Panuluh Yudha Perkasa, K., & Eka Purwiantono, F. (n.d.). Sistem Rekomendasi Jurusan Menggunakan Algoritma Naive Bayes Gaussian Berbasis Web. *Jl. Raya Tidar*, 65146(100). Retrieved from <https://jurnal.stiki.ac.id/J-INTECH/article/view/1090>
- Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9(2), 48–57. <https://doi.org/10.1109/MCI.2014.2307227>
- Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 427–434. <https://doi.org/10.25126/jtiik.201854773>
- Dharmawan, A., Raharjo, S. T., & Kusumawardhani, A. (2019). Analisis Kinerja Bea Cukai Indonesia Yang Dipengaruhi Oleh Budaya Organisasi, Partisipasi Stakeholder, Teknologi Informasi, Dan Transfer of Knowledge. *Jurnal Bisnis Strategi*, 27(2), 110. <https://doi.org/10.14710/jbs.27.2.110-122>
- Firmansyah, S., Gaol, J., & Susilo, S. B. (2019). Comparison of SVM and Decision Tree Classifier with Object Based Approach for Mangrove Mapping to Sentinel-2B Data on Gili Sulat, Lombok Timur. *Jurnal Pengelolaan Sumberdaya Alam Dan Lingkungan*, 9(3), 746–757. <https://doi.org/10.29244/jpsl.9.3.746-757>
- Haqiqi, F., Santika, M., & Yusmalina. (2021). Analisis Pengaruh Realisasi Penerimaan Bea Masuk Terhadap Target Penerimaan Bea Masuk (Pada Kantor Pengawasan Dan Pelayanan Bea Cukai Tipe Madya B Tanjung Balai Karimun Periode 2017-2019). *Jurnal Cafeteria*, 2(1), 48–61. <https://doi.org/10.51742/akuntansi.v2i1.282>
- Hardiantoro, A., & Firdaus, F. (n.d.). Ramai soal Barang Impor Gratis Masih Dikenai Pajak Bea Masuk, Ini Kata Ditjen Bea Cukai - Kompas.com. Retrieved June 6, 2024, from [kompas.com website: https://www.kompas.com/tren/read/2023/03/18/103000465/ramai-soal-barang-impor-gratis-masih-dikenai-pajak-bea-masuk-ini-kata?page=all#google_vignette](https://www.kompas.com/tren/read/2023/03/18/103000465/ramai-soal-barang-impor-gratis-masih-dikenai-pajak-bea-masuk-ini-kata?page=all#google_vignette)
- Jain, A., & Gupta, B. (2022). An Enhanced Naive Bayes Model for Sentiment Analysis of Tweets. *Applied Computing and Informatics*.
- Jiang et al. (2009). *text classification with Naive Bayes*.
- Khan, M., B. B., & Kamaruddin, H. (2020). Efficient Classification of Short Text with Deep Learning and Feature Reduction Techniques. *IEEE*, 8, 134479–134491.
- Lumbanraja, F. R., Gaol, E. C. L., Shofiana, D. A., & Junaidi, A. (2024). Implementasi SMOTE dan Support Vector Machine Pada Klasifikasi Data Tidak Seimbang Metilasi Arginin. *Jurnal Pepadun*, 5(1), 27–37. <https://doi.org/10.23960/pepadun.v5i1.209>
- Pravina, A. M., Cholissodin, I., & Adikara, P. P. (2019). Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2789–2797.
- Qorry Meidianingsih, Wardani, D. E., Salsabila, E., Nafisah, L., & Mutia, A. N. (2023). Perbandingan Performa Metode Berbasis Support Vector Machine untuk Penanganan Klasifikasi Multi Kelas Tidak Seimbang. *STATISTIKA Journal of Theoretical Statistics and Its Applications*, 23(1), 8–18. <https://doi.org/10.29313/statistika.v23i1.1660>

- Ristoski, P., Bizer, C., & Paulheim, H. (2015). Mining the Web of Linked Data with RapidMiner. *Journal of Web Semantics*, 35, 142–151. <https://doi.org/10.1016/j.WEBSEM.2015.06.004>
- Sangeetha, S., & Vijayakumar, V. . (2018). Performance Analysis of Naive Bayes Algorithm in Classifying Twitter Sentiment with Different Feature Extraction Methods. *International Journal of Engineering and Technology (IJET)*, 7, 580–584.
- Singh, M. (2016). *Sentiment Analysis and Similarity Evaluation for Heterogeneous-Domain Product Reviews*. 144(2), 16–19.
- Supriyatna, A., & Mustika, W. P. (2018). Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 2(2), 152. <https://doi.org/10.30645/j-sakti.v2i2.78>
- Syah, F., Fajrin, H., Afif, A. N., Saeputra, M. R., Mirranty, D., & Saputra, D. D. (2023). Analisa Sentimen Terhadap Twitter IndihomeCare Menggunakan Perbandingan Algoritma Smote, Support Vector Machine, AdaBoost dan Particle Swarm Optimization. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 7(1), 53–58. <https://doi.org/10.35870/jtik.v7i1.686>
- Toy, K. V. S., Sari, Y. A., & Cholissodin, I. (2021). Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(11), 5068–5074. Retrieved from <http://j-ptiik.ub.ac.id>
- Yudistira, A., & Wicaksana, S. T. (2022). Studi Kasus Implementasi SNI ISO 37001:2016 Dalam Pencegahan Korupsi Pada KPU BC Tipe A Tanjung Priok. *Jurnal Perspektif Bea Dan Cukai*, 6(2), 273–283.
- Yulita, W. (2021a). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naive Bayes Classifier. *Jurnal Data Mining Dan Sistem Informasi*, 2(2), 1. <https://doi.org/10.33365/jdmsi.v2i2.1344>
- Yulita, W. (2021b). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naive Bayes Classifier. *Jurnal Data Mining Dan Sistem Informasi*, 2(2), 1. <https://doi.org/10.33365/jdmsi.v2i2.1344>