

## ANALISIS SENTIMEN TERHADAP BEA CUKAI MENGGUNAKAN SUPPORT VECTOR MACHINE DAN K-FOLD CROSS VALIDATION

Nataniel Blesyova, Fuad Nur Hasan

Universitas Bina Sarana Informatika

Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Indonesia

*nataniel515@gmail.com*

### ABSTRAK

Penelitian ini dilatarbelakangi masalah yang disebabkan oleh Bea Cukai terkait penahanan donasi alat taptilo dan korban dimintai sejumlah uang untuk penebusan barang, sehingga menimbulkan berbagai opini/emosi di masyarakat terhadap Bea Cukai. Tujuan utama penelitian adalah mengevaluasi kinerja Support Vector Machine (SVM) dalam klasifikasi sentimen komentar media sosial terkait Bea Cukai, sebagai algoritma machine learning yang efektif. Metode yang digunakan melibatkan beberapa tahap, yaitu pengumpulan data dengan metode crawling data, pre-processing data yang terdiri dari case folding, cleaning, stemming, stopword, dan tokenize, penerapan algoritma SVM, evaluasi menggunakan confusion matrix, dan pengujian menggunakan teknik K-Fold Cross Validation, semua tahapan tersebut akan dilakukan menggunakan aplikasi Google Colab menggunakan bahasa Python. Hasil penelitian menunjukkan bahwa algoritma SVM menghasilkan akurasi sebesar 94% dari 400 data test dan 1600 data train, dengan precision nilai 88%, class recall dengan nilai 95%, class f1-score dengan nilai 91.36%. Pengujian dengan K-Fold Cross Validation memberikan akurasi tertinggi sebesar 98.85%, menunjukkan peningkatan signifikan. Cloudword dapat divisualisasikan per kategori data untuk melihat kata apa yang frekuensinya sering muncul. Model ini juga menunjukkan kemampuan baik dalam mengklasifikasikan sentimen positif dan negatif secara konsisten. Kesimpulannya, algoritma SVM terbukti efektif dalam klasifikasi sentimen komentar media sosial terkait Bea Cukai.

**Kata kunci :** Analisis Sentimen, Support Vector Machine, Machine Learning

### 1. PENDAHULUAN

Direktorat Jendral Bea dan Cukai merupakan instansi pemerintah yang berada dibawah naungan dan bertanggung jawab kepada Menteri Keuangan. Direktorat Jendral Bea dan Cukai memiliki tugas dalam perumusan dan pelaksanaan kebijakan dibidang pengawasan, penegakan hukum, pelayanan dan optimasi penerimaan negara negara dalam bidang kepabeanan dan cukai sesuai dengan ketentuan peraturan perundang-undangan.

Baru-baru ini terjadi kasus yang viral dari media sosial X dimana alat taptilo bantuan dari perusahaan Korea Selatan ditahan Kantor Bea dan Cukai Bandara Soekarno-Hatta (Soetta). Saat pemilik akun ingin mengambil barang tersebut, yang bersangkutan malah ditagih senilai ratusan juta rupiah. Kasus alat taptilo yang melibatkan pihak Bea Cukai tentunya menghadirkan berbagai sentimen di masyarakat luas, ada yang memberikan sentimen positif ataupun negatif. Dalam hal ini penulis akan memanfaatkan platform X untuk melakukan analisis sentimen berdasarkan data dari komentar dan balasan pengguna X terkait dengan tagar Bea Cukai.

“Analisis sentimen adalah teknik untuk mengolah data tekstual untuk memperoleh informasi pada teks”[1]. Ide dasar pada analisis sentimen yaitu menemukan polaritas dari dokumen dan mengklasifikasikannya menjadi positif atau negatif. Analisis sentimen memiliki beberapa metode yang telah banyak digunakan dalam kasus mengklasifikasikan review produk, antara lainnya seperti *Naive Bayes*, *Support Vector Machine*, dan *K-*

*Nearest Neighbour* dan ada metode pengujian yang digunakan seperti *K-Fold Cross Validation*. *K-Fold Cross Validation* merupakan teknik validasi dalam Machine Learning yang bertujuan memperkirakan performa model secara akurat dan mengurangi bias. *K-Fold* bekerja dengan cara membagi data menjadi *K sub-fold*, melatih model pada *K-1 sub-fold*, mengevaluasinya pada *sub-fold* yang tersisa, dan mengulangi proses ini *K* kali. SVM telah banyak digunakan dalam berbagai kasus studi klasifikasi teks, salah satu kelebihanannya adalah mengatasi masalah klasifikasi teks yang bebas dan beragam.

Penelitian analisis sentimen yang menggunakan data ulasan juga sudah diteliti sebelumnya mulai dari data ulasan Twitter [2], ulasan aplikasi Shopee [3] dan ulasan film [4]. Maka penelitian ini akan menggunakan metode serupa dengan penelitian yang sudah ada yaitu melakukan analisis sentimen menggunakan model *Support Vector Machine* menggunakan data berbeda yaitu tentang ulasan dari komentar media sosial X tagar Bea Cukai.

Banyak penelitian sebelumnya telah dilakukan terkait analisis sentimen. Berikut merupakan beberapa penelitian yang sudah dilakukan :

Penelitian dari [5], menggunakan dataset sebesar 3350 tweet masyarakat terhadap pengungsi Rohingya di Indonesia. Menggunakan metode *Naive Bayes* dan SVM didapatkan hasil model *Naive Bayes* memiliki akurasi 70% sedangkan model SVM memiliki akurasi 76%.

Penelitian dari [6], menggunakan data 2462 ulasan aplikasi Gojek di Google Playstore dan

menggunakan metode KNN dan SVM dengan hasil penelitian metode KNN memperoleh nilai akurasi, presisi, dan recall berturut-turut sebesar 82%, 82%, 95%, sedangkan metode SVM memperoleh nilai 87%, 88%, 95%.

Penelitian dari [7], menggunakan data 2750 komentar dari Instagram terhadap PSSI, metode yang digunakan adalah Decision Tree dan SVM, model Decision Tree menghasilkan nilai akurasi 87%, precision 87%, recall 91%, dan F1-Score 89%, sedangkan model SVM mendapatkan nilai akurasi 94%, precision 96%, recall 94%, dan F1-Score 95%.

Penelitian ini bertujuan untuk memahami pandangan masyarakat terhadap Bea Cukai yang terjerat kasus penahanan sumbangan alat taptilo dari Korea. Dengan menggunakan metode Support Vector Machine untuk klasifikasi data, dan juga K-Fold Cross Validation untuk metode pengujian model.

## 2. TINJAUAN PUSTAKA

Dalam penulisan skripsi ini peneliti melakukan riset dari beberapa penelitian sebelumnya sebagai bahan perbandingan. Selain itu, peneliti juga melakukan riset dari buku-buku, jurnal ataupun skripsi dalam rangka mendapatkan informasi terkait tentang landasan teori yang berhubungan dengan tema penelitian penulis.

### 2.1. Teks Mining

Klasifikasi termasuk pada model supervised. Pada persoalan klasifikasi kita memiliki sampel data dan memprediksi beberapa class yang ada berdasarkan sampel yang ada. Hanya satu atribut di antara banyak atribut yang disebut dengan atribut target [8].

### 2.2. Python

Python merupakan high level and multi-purpose programming language yang dikembangkan oleh Guido Van Rossum pada tahun 1990 di Stichting Mathematisch Centrum (CWI), Amsterdam sebagai kelanjutan dan bahasa pemrograman ABC. Sintaks bahasa Python relatif paling sederhana dan mudah dipelajari oleh mereka yang bukan programmer tulen. Python mendukung model pengembangan berbasis objek (object-oriented atau OOP) dan juga prosedur (procedural) [9].

### 2.3. Pre-Processing

Tujuan utama dari data preprocessing adalah untuk meningkatkan kualitas data, memastikan keakuratan hasil analisis, dan mengatasi masalah atau kekurangan yang mungkin muncul dalam data mentah [10].

Pada tahapan ini, data akan dilanjutkan ke tahapan *pre-processing* untuk dilakukan *text cleansing* melalui *Google Colab* dengan menggunakan *library* yang tersedia, terdiri dari sebagai berikut :

#### a. Case Folding

*Case folding* adalah proses mengubah semua huruf dalam teks menjadi huruf kecil. Hal ini dilakukan untuk menghilangkan perbedaan makna yang hanya disebabkan oleh kapitalisasi.

#### b. Cleaning

*Cleaning* adalah proses membersihkan teks dari noise atau elemen yang tidak relevan. Noise ini dapat berupa tag HTML, URL, karakter khusus, tanda baca yang berlebihan, dan whitespace yang tidak perlu.

#### c. Stemming

*Stemming* adalah proses mereduksi kata menjadi bentuk dasarnya. Hal ini dilakukan untuk mengelompokkan kata-kata yang memiliki makna yang sama meskipun memiliki bentuk yang berbeda.

#### d. Stopword

*Stopword* adalah kata-kata yang umum muncul dalam bahasa Indonesia dan tidak memiliki makna yang signifikan dalam konteks tertentu. Kata-kata ini dihapus dari teks untuk meningkatkan fokus pada kata-kata yang lebih penting.

#### e. Tokenize

*Tokenize* adalah proses memecah teks menjadi unit-unit yang lebih kecil, seperti kata, kata sifat, atau kata kerja. Token-token ini dapat digunakan untuk analisis teks lebih lanjut, seperti sentiment analysis atau topic modeling.

## 2.4. Support Vector Machine

Metode SVM adalah salah satu cara untuk mengklasifikasikan opini menjadi beberapa bagian. Support Vector Machine adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (feature space) yang berdimensi tinggi dan mengimplementasikan learning bias[11].

## 2.5. K-Fold Cross Validation

K-Fold Cross Validation adalah salah satu dari jenis pengujian Cross Validation yang berfungsi untuk menilai kinerja proses sebuah metode algoritma dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai K [12].

## 2.6. Confusion Matrix

*Confusion matrix* adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. *Confusion matrix* digambarkan dengan tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan [13]

Tabel 1. Tabel Confusion Matrix

Correct Classification	Classified as	
	Predicted “+”	Predicted “-”
Actual “+”	True Positives	False Negatives
Actual “-”	False Positives	True Negatives

Berdasarkan tabel Confusion Matrix diatas:

- True Positives (TP)* adalah jumlah data positif yang diklasifikasikan sebagai nilai positif.
- False Positives (FP)* adalah jumlah data positif yang diklasifikasikan sebagai nilai negatif.
- False Negatives (FN)* adalah jumlah data negatif yang diklasifikasikan sebagai nilai positif.
- True Negatives (TN)* adalah jumlah data negatif yang diklasifikasikan sebagai nilai negatif.

Dalam penghitungan akurasi menggunakan *Confusion Matrix* terdapat *accuracy*, *precision*, *recall*, dan *f1-score* yang dapat dijelaskan sebagai berikut:

- Accuracy* merupakan perbandingan antara data yang diklasifikasikan benar dengan keseluruhan data hasil klasifikasi.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (1)$$

- Precision* merupakan seluruh data dengan nilai positif yang diklasifikasikan secara benar dibagi dengan data nilai positif dan data diklasifikasikan positif.

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (2)$$

- Recall* merupakan merupakan data dengan nilai positif yang diklasifikasikan dengan benar dibagi seluruh data yang sebenarnya positif.

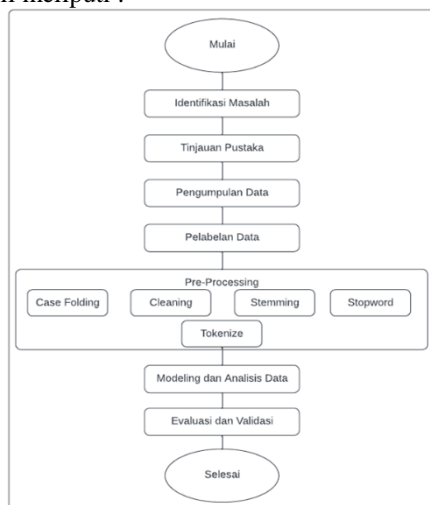
$$Recall = \frac{TP}{TP+FN} \times 100 \quad (3)$$

- F1-Score* mengindikasikan bahwa model klasifikasi memiliki *precision* dan *recall* yang baik, hasil dapat dilakukan dengan persamaan.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100 \quad (4)$$

### 3. METODE PENELITIAN

Dalam penelitian ini, langkah penelitian yang diambil meliputi :



Gambar 1. Flowchart penelitian

#### 3.1. Identifikasi Masalah

Agar hasil penelitian sesuai dengan alur penelitian opini masyarakat terhadap Bea Cukai maka dilakukan tahapan sebagai berikut

- Latar Belakang : Acuan latar belakang pada penelitian ini adalah berbagai macam kejadian yang terjadi belakangan ini menyangkut pihak Bea Cukai meninggalkan berbagai macam kesan di masyarakat luas, ada masyarakat yang memberikan komentar negatif dan ada juga yang memberikan komentar positif terhadap Bea Cukai.
- Tujuan Penelitian : Tujuan dari penelitian ini adalah mengklasifikasikan sentimen masyarakat kedalam kategori negatif atau positif dengan tingkat akurasi yang tinggi
- Rumusan Masalah : Rumusan masalah dalam penelitian ini meliputi metode yang akan di aplikasikan dalam penelitian, bagaimana cara pengaplikasiannya, bagaimana memperoleh data komentar masyarakat

#### 3.2. Tinjauan Pustaka

Proses mendapatkan materi menurut para ahli. Pada tahapan dimaksud sangat berguna dalam mendukung penelitian ini berdasarkan para ahli sehingga penelitian bersifat logis dan terarah.

#### 3.3. Pengumpulan Data

Dalam proses ini, data yang digunakan adalah data sekunder, yaitu data diambil dari opini terkait Bea Cukai pada media sosial X dilakukan dengan teknik pengumpulan data metode data crawling untuk memastikan sumber data merupakan sumber yang asli dan terpercaya, opini yang diambil adalah tweet yang dibuat pada tanggal 18 Mei 2024 sampai 22 Mei 2024, dengan total data sebanyak 2000 data.

#### 3.4. Pelabelan Data

Data yang sudah melalui tahapan pre-processing kemudian akan diberi label menjadi 2 kategori, yaitu positif dan negatif.

#### 3.5. Pre-Processing

Pada tahapan ini, data akan dilanjutkan ke tahapan *pre-processing* untuk dilakukan *text cleansing* melalui *Google Colab* dengan menggunakan *library* yang tersedia, terdiri dari sebagai berikut :

#### 3.6. Modeling dan Analisis Data

Setelah mempersiapkan data melalui proses pre-processing dan pelabelan, tahapan berikutnya adalah membuat model analisis yang akan digunakan untuk menganalisis sentimen masyarakat terhadap Bea Cukai, model analisis yang dibuat akan menggunakan metode Support Vector Machine, aplikasi berbasis web Google Colab, dan Bahasa pemrograman yang digunakan adalah Python. Data kemudian akan di bagi menjadi 80% data training dan 20% data test dan akan di aplikasikan ke model analisis sentiment machine learning dengan metode Support Vector Machine



Gambar 2. Flowchart SVM

### 3.7. Evaluasi dan Validasi



Gambar 3. Flowchart K-Fold

Metode yang dipakai untuk melakukan evaluasi pada hasil dari analisis sentiment menggunakan metode Support Vector Machine adalah Confusion Matrix dan proses pengujian validasi akan menggunakan K-Fold Cross Validation, dengan K=10 untuk pengujian dan Kembali menggunakan Confusion Matrix untuk mengevaluasi kinerja dan akurasi.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Pengumpulan Data

Dataset yang digunakan berasal dari sosial media X menggunakan aplikasi *Google Colab*. Pengambilan data dilakukan dengan kurun waktu 18 Mei 2024 sampai dengan 22 Mei 2024, dengan search keyword beacukai. Hasil crawling data sebanyak 2.000 data

akan disimpan dalam format *Comma Separated Values (CSV)* dan di download.

```

cristin, at bel_jat
Wed May 22 12:34:00 +0000 2024 Bea cukai - gembolan pernyaman... Wlanta ini bel batin dari Thailand Rp10 ribu tapi duhah kunya oleh petugas Bea Cukai jadi Rp2300 ribu loh bisa? https://t.me/maulanaq
Wed May 22 12:33:42 +0000 2024 @bel_maulanaq @comomandom @dokterparu Mikin solusi permasalahan bea cukai
Wed May 22 12:20:06 +0000 2024 Bea Cukai terus lakukan perbaikan pada pelayanan agar menjadi lebih baik. #BeaCukaiIndonesia #BeaCukai #BeaCukaiMakinBerk #BeaCukaiMelayan #BeaC
Wed May 22 12:17:09 +0000 2024 @ine_surbadi Sampai kenebak2 isi sampai jage dibongkar. Bismillah jernih dlmpru zoro tu bea cukai @beacukaiRF assu
Wed May 22 12:15:13 +0000 2024 @Wlanta jangn angan-angan ko uah ga larut karena kenebak di beacukai karna ga bisa bayar
Wed May 22 12:15:58 +0000 2024 @jllagomercu Kalopun bisa kenebak bea cukai sih wkwk
Wed May 22 12:08:19 +0000 2024 Ketakutannya mrip bea cukai
Wed May 22 12:05:08 +0000 2024 Stress banget kalo kudu bea cukai yang beres di Indonesia. Dari mulai UKT naik paji jadi tenaga kerja kesehatan bea cukai layanan kesehatan. Negara ini b
Wed May 22 11:55:07 +0000 2024 Dengan adanya petak disiplin ini semua pegawai diharapkan dapat menyadari akan pentingnya menjaga kedisiplinan dan berprestasi untuk terus meningkatkan ke
Wed May 22 11:55:56 +0000 2024 @Puthusa Karna ga gaj pejabat aja yg gedhe2 itu dipotong pejabat bea cukai misal nya atau pejabat kementerian misal nya atau gaj pensiun dor apa dihapus da
Wed May 22 11:51:00 +0000 2024 Didik Waydyd https://t.co/G5dRGAjXg -Harten. M.M. Kepala Seksi Intelijen I Bea Cukai Tipe B Batam Didik Mulyono Kabid Propom Podda Keri Komdes
Wed May 22 11:44:08 +0000 2024 @memmoo Karna bapa kudu mematuin hukum ini ya pejabat negara ya? Na 100 anggota DPR yang aji? Na anggota DPR yang TIDAK TERPILU? Na
Wed May 22 11:42:32 +0000 2024 @ine_surbadi Ngasin-ngasin bar judi online terbak se duna korupsi Ter the best se asia tenggara sakti bea cukai. Terassusss ae nng gae lah glan kyokw ke
Wed May 22 11:37:36 +0000 2024 @coohiyena Wlanta pamo Karna bea cukai kebanyakan tinglah
Wed May 22 10:39:09 +0000 2024 Cara orang biasa ditangan ya viral dula Salah sendiri cuma pejabat sama orang berduit yang diurusin sat set bat bet Kasus terbaru : tas enzy bea cukai Coba yang nyap
Wed May 22 10:38:33 +0000 2024 Gaj dan Tunjangan PNS Bea Cukai Segen Kluarnya https://t.co/Dvst7BRLW
Wed May 22 10:31:11 +0000 2024 Jika nulis harga dibawah harga sebenarnya dikenakan pajak+denda oleh beacukai. Jika nulis harga sebenarnya malah dttar lebih mahal bar pajaknya besar. H
Wed May 22 10:28:29 +0000 2024 @dondopamin Lah langsung pesen biar sedekah ke bea cukai
Wed May 22 10:28:06 +0000 2024 Kalau terlampaui banyak OKNUM di Bea Cukai atau paji membayar paji akan terasa beban besar di pundak rakyat. Untuk apa? habis dikorupsi saja.
Wed May 22 10:27:24 +0000 2024 @ine_surbadi be bea cukai lebih ksh gmn amat dan dlatr ngat gmn? jgn sampe ky kasus peti mati yg di dimintan biaya beacukai.
Wed May 22 10:26:29 +0000 2024 @tanjanyes Abis blngnya lagi yg kmren blm fix salah tangkep kan? jgn sampe ky kasus peti mati yg di dimintan biaya beacukai.
Wed May 22 10:23:52 +0000 2024 @tanjanyes76 Bea cukai slap2 mlok ada jelpo setelah pulang sekolah? ngapao pejabat jangn rayat meluk yg di cekek
Wed May 22 10:13:29 +0000 2024 @ine_surbadi Bea cukai udah lebih ksh dr anjing ksh endusan nya lebih tajam sampe tahu barang2 bermerek
Wed May 22 10:10:05 +0000 2024 @mst781032400 @GPK_Ri Klu Kepala Bea Cukai emang bosen KTP swasta buat smua pengurusan bea? Klu idek 2 Brati salatnya dimana?
Wed May : Cukai dan Akademisi. Maksud dan tujuan dilaksanakan
Wed May 22 10:08:58 +0000 2024 @BNiduna10 Keren bea Cukai
  
```

Gambar 4. Hasil Pengumpulan Data

### 4.2. Pelabelan Data

Setelah proses pengumpulan data dari X, langkah selanjutnya adalah *labeling* untuk klasifikasi data dengan 2 kelas kategori yaitu positif dan negatif seperti tabel dibawah ini.

Tabel 2. Hasil Pelabelan Data

Text	Sentimen
Cara orang biasa ditangani ya viral dulu Salah sendiri cuma pejabat sama orang berduit yang diurusin sat set bat bet Kasus terbaru : tas enzy bea cukai Coba yang nyap nyap orang biasa Gak bakal digubris	Negatif
Kalau terlampaui banyak OKNUM di Bea Cukai atau pajak membayar pajak akan terasa beban besar di pundak rakyat. Untuk apa? habis dikorupsi saja.	Negatif

### 4.3. Pre-Processing Data

Setelah proses pemberian label terhadap dataset, tahap selanjutnya adalah dilakukan *pre-processing* dengan mengimport dan mendownload library yang dibutuhkan didalam *google colab* seperti menimport *library pandas, numpy, nltk, string, StemmerFactory*, dan lain-lain, adapun juga *library* yang harus didownload seperti *stopwords, wordnet, punkt, Sastrawi pandas*. Semua *library* tersebut diperlukan untuk dapat dilakukan *pre-processing* pada dataset dengan tahapan-tahapan *pre-processing* adalah *Case Folding, Cleaning, Stemming, Stopword*, dan *Tokenize*.

Tabel 3. Hasil Case Folding

Sebelum	Sesudah
@50Kuasa Bukan di jakarta aja tp bea cukai seluruh Indonesia sarang KKN dari atasan sampai bawahan.	@50kuasa bukan di jakarta aja tp bea cukai seluruh indonesia sarang knn dari atasan sampai bawahan,

Tabel 4. Hasil Cleaning

Sebelum	Sesudah
abis blngnya lagi yg kmren blm fix salah tangkep kan? jgn sampe ky kasus peti mati yg dimintan biaya beacukai.	abis blngnya lagi yg kmren blm fix salah tangkep kan jgn sampe ky kasus peti mati yg dimintan biaya beacukai.

Tabel 5. Hasil Stemming

Sebelum	Sesudah
kalau diperiksa gak bakal ada sisa ni pejabat bea cukai kang palak rampok pula.	kalau periksa gak bakal ada sisa ni jabat bea cukai kang palak rampok pula.

#### 4.4. Modeling

Pada pemodelan penelitian ini akan digunakan metode *Support Vector Machine* untuk klasifikasi sentiment dengan pembagian dataset 80% data *training* dan 20% data *testing*, *Confusion Matrix* untuk *Evaluasi*, dan *K-Fold Cross Validation* yang akan digunakan pada tahapan validasi untuk mengukur kinerja model dengan lebih akurat. *Support Vector Machine* adalah salah satu algoritma teknik *machine learning* yang populer untuk permasalahan klasifikasi teks yang memiliki performa yang baik.

```

X = data['Teks']
y = data['Sentimen']

# Text Vectorization menggunakan TF-IDF
vectorizer = TfidfVectorizer(max_features=5000)
X_tfidf = vectorizer.fit_transform(X)

# Pembagian data tes dan data train
X_train, X_test, y_train, y_test = train_test_split(X_tfidf, y, test_size=0.2, random_state=42)

# Melatih model SVM
svm_model = SVC(kernel='linear')
svm_model.fit(X_train, y_train)

# Evaluasi
y_pred = svm_model.predict(X_test)

```

Gambar 5. Pemodelan SVM

Dengan algoritma *Support Vector Machine* klasifikasi akan dilakukan dengan melakukan perhitungan untuk menemukan *hyperlane* terbaik yang memisahkan dua kelas secara langsung, *kernel linier* menghitung antar *vector* fitur data dengan kemiripan antara dua titik, berikut ini adalah persamaan yang digunakan dalam perhitungan algoritma *Support Vector Machine* :

$$f(x) = w \cdot x + b \quad (5)$$

$$f(x) = \sum a_i \cdot y_i \cdot K(x, x_i) + b \cdot m, i = 1 \quad (6)$$

$$K(x, x_i) = x \cdot y \quad (7)$$

Keterangan :

w = Parameter hyperlane  
 x = Titik data  
 b = Nilai bias  
 ai = Nilai bobot titik data  
 K(x,xi) = Fungsi kernel linier

#### 4.5. Evaluasi

Evaluasi merupakan tahapan untuk mengevaluasi suatu model yang bertujuan untuk mengukur akurasi dari sebuah metode. Hasil dari penelitian ini adalah penerapan algoritma *Support Vector Machine* dalam analisa sentiment berdasarkan evaluasi nilai *Accuracy* menggunakan *Confusion Matrics*.

Tabel 6. Confusion Matrix SVM

Prediksi	Hasil Prediksi
True Positif	23
False Positif	3
True Negatif	373
False Negatif	1
Total Seluruh Data	400

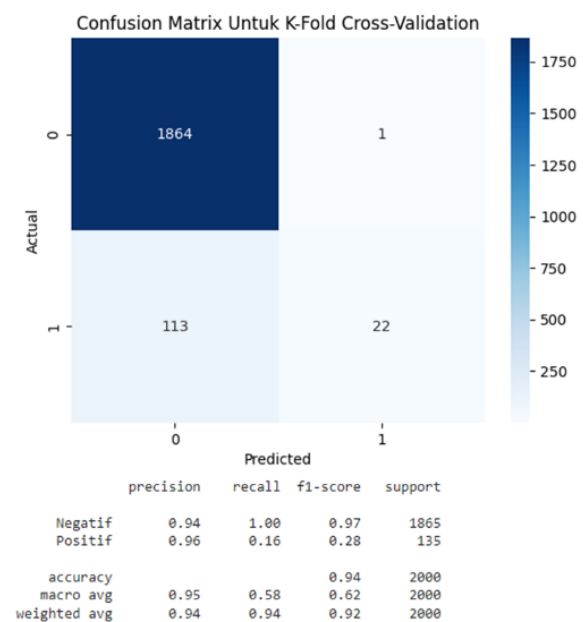
Berdasarkan hasil yang didapatkan dari evaluasi *confusion matrix* menunjukkan dari 400 data test analisis menggunakan SVM menghasilkan 23 data *true positif*, 3 data *false positif*, 373 data *true negatif*, dan 1 *false negatif* dengan nilai akurasi 94%, *precision* 93%, *recall* 94%, *f1-score* 92% yang bisa dilihat dari kinerja sesuai gambar diatas.

#### 4.6. Validasi

Tabel 7. Pengujian K-Fold Cross Validation

Fold	Hasil Akurasi
Fold 1	0.9550
Fold 2	0.9250
Fold 3	0.9200
Fold 4	0.9550
Fold 5	0.9350
Fold 6	0.9450
Fold 7	0.9600
Fold 8	0.9300
Fold 9	0.9500
Fold 10	0.9550

Dari gambar 4 diperoleh nilai akurasi paling tinggi berada pada *fold* ke 7 dengan nilai akurasi yang sama yaitu 0.9600, akurasi terendah berada pada *fold* ke 3 dengan nilai akurasi yang sama 0.9200, dan presentase negatif 93,50% sedangkan positif 6,50%.



Gambar 6. Evaluasi K-Fold

Berdasarkan hasil yang didapatkan dari evaluasi confusion matrix pengujian *K-Fold Cross Validation* pada dataset menghasilkan 113 data *true* positif, 22 data *false* positif, 1864 data *true* negatif, dan 1 *false* negatif dengan nilai akurasi 94% *precision* 94%, *recall* 94%, *f1-score* 92% yang bisa dilihat sesuai gambar diatas.

## 5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa algoritma *Support Vector Machine* (SVM) yang dikombinasikan dengan teknik *K-Fold Cross Validation* memberikan hasil yang sangat akurat dalam klasifikasi sentimen terhadap Bea Cukai. Nilai akurasi yang dicapai pada evaluasi pemodelan SVM dengan nilai akurasi 94%, *precision* 93%, *recall* 94%, *f1-score* 92% dan juga evaluasi pengujian model dengan *K-Fold Cross Validation* menunjukan hasil nilai akurasi 94% *precision* 94%, *recall* 94%, *f1-score* 92%. Hasil klasifikasi menunjukan 113 data *true* positif, 22 data *false* positif, 1864 data *true* negatif, dan 1 *false* negatif. Untuk meningkatkan hasil analisis sentimen di masa mendatang, disarankan untuk memperluas dataset dengan menambahkan lebih banyak data dari berbagai platform media sosial. Selain itu, menggabungkan SVM dengan teknik machine learning lainnya, bisa dieksplorasi untuk meningkatkan akurasi dan efisiensi model.

## DAFTAR PUSTAKA

- [1] M. Diki Hendriyanto, A. A. Ridha, and U. Enri, "ANALISIS SENTIMEN ULASAN APLIKASI MOLA PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE SENTIMENT ANALYSIS OF MOLA APPLICATION REVIEWS ON GOOGLE PLAY STORE USING SUPPORT VECTOR MACHINE ALGORITHM," *J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, 2022.
- [2] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 3, p. 650, Jul. 2020, doi: 10.30865/mib.v4i3.2181.
- [3] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Pengguna Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *J. Inf. Sci.*, vol. 36, no. 6, pp. 823–848, Dec. 2023, doi: 10.1177/0165551510388123.
- [4] O. I. Gifari, M. Adha, I. Rifky Hendrawan, F. Freddy, and S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *JIFOTECH (JOURNAL Inf. Technol.*, vol. 2, no. 1, 2022.
- [5] D. Ananda and R. R. Suryono, "Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes," vol. 8, no. April, pp. 748–757, 2024, doi: 10.30865/mib.v8i2.7517.
- [6] M. N. Muttaqin and I. Kharisudin, "Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor," *UNNES J. Math.*, vol. 10, no. 2, pp. 22–27, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- [7] M. F. Asshiddiqi and K. M. Lhaksmana, "Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI," *J. Inform.*, vol. 7, no. D, p. 9936, 2020.
- [8] N. L. W. S. R. Ginantra et al., *Data Mining dan Penerapan Algoritma*. 2021.
- [9] Bedy Purnama dan Iwan Sofana, *Implementasi Artificial Intelligence dan Machine Learning*. Bandung: Bi-obses, 2021. [Online]. Available: <https://www.gramedia.com/products/implementasi-artificial-intelligence-dan-machine-learning-d>
- [10] I. G. I. Sudipa, I. G. M. Darmawiguna, I. M. Dendi, and M. Sanjaya, *Buku ajar data mining (Issue January)*, no. January. 2024.
- [11] S. F. Pane, A. Owen, and C. Prianto, "Analisis Sentimen UU Omnibus Law pada Twitter Menggunakan Metode Support Vector Machine," *J. Telekomun. dan Komput.*, vol. 11, no. 2, p. 130, 2021, doi: 10.22441/incomtech.v11i2.10874.
- [12] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, "Aplikasi Naïve Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *J. Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, doi: 10.14710/j.gauss.v11i1.33991.
- [13] M. F. Rahman, D. Alamsah, M. I. Darmawidjadja, and I. Nurma, "Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN)," *J. Inform.*, vol. 11, no. 1, p. 36, 2020, doi: 10.26555/jifo.v11i1.a5452.