



Penerapan *Support Vector Machine* (SVM) Untuk Klasifikasi Sentimen E-Tilang Pada *Twitter*

Tegar Dwi Prayuda*, Moh. Dasuki, Yusril Izzi Arlisa Amiri

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Email: tegardwi146@gmail.com*, moh.dasuki22@unmuhjember.ac.id, yusril.amiri@unmuhjember.ac.id

ABSTRAK

E-Tilang sebuah inovasi penegak hukum di bidang lalu lintas yang berbasis teknologi yang menggunakan perangkat elektronik seperti kamera pengawas *close circuit television* (CCTV) yang mampu merekam kejadian dan mendeteksi plat nomor kendaraan sebagai barang bukti. Atas kebijakan tersebut banyak masyarakat yang belum begitu bijak dalam menyikapi penerapan kemajuan teknologi yaitu E-Tilang. Hal ini ditandai adanya berbagai opini oleh para pengguna twitter tentang E-Tilang yang terdiri berbagai macam ada yang pro maupun kontra. Twitter merupakan sarana yang efektif dalam menampung opini pengguna transportasi. Metode yang diusulkan pada penelitian ini menggunakan *Support Vector Machine* (SVM). Data set yang digunakan dalam penelitian ini adalah tweet dengan bahasa Indonesia dengan *keyword* E-Tilang dan didapatkan data set sebanyak 400 tweet dan 756 tweet. Hasil dari penelitian analisis sentimen E-Tilang menggunakan klasifikasi *Support Vector Machine* (SVM) dan ekstraksi fitur TF-IDF, dengan 250 dataset didapatkan *akurasi* 64%, *presisi* 65%, *recall* 80% dan menggunakan 590 dataset didapatkan *akurasi* 70%, *presisi* 71%, *recall* 96%.

Kata Kunci: Analisis Sentimen, E-Tilang, Twitter, Support Vector Machine.

ABSTRACT

E-Tilang is a technology-based law enforcement innovation in the traffic sector that uses electronic devices such as closed-circuit television (CCTV) surveillance cameras which are capable of recording incidents and detecting vehicle number plates as evidence. Due to this policy, many people are not very wise in responding to the implementation of technological advances, namely E-Tilang. This is indicated by the existence of various opinions by Twitter users about E-Tilang which consist of various pros and cons. Twitter is an effective means of accommodating the opinions of transportation users. The method proposed in this research uses -Support Vector Machine (SVM). The data set used in this research was tweets in Indonesian with the keyword E-Tilang and a data set of 400 tweets and 756 tweets was obtained. The results of E-Tilang sentiment analysis research using Support Vector Machine (SVM) classification and TF-IDF feature extraction, with 250 datasets obtained accuracy 64%, 65% precision, 80% recall and using 590 datasets obtained accuracy 70%, precision 71%, recall 96%.

Keywords: Sentiment Analysis, E-Tilang, Twitter, Support Vector Machine.

1. PENDAHULUAN

Sistem tilang telah beralih pada sistem elektronik yang dikenal dengan istilah E-Tilang. Sistem E-Tilang dapat mengambil alih sistem tilang manual yang dicatat dengan blanko atau surat tilang. Pelanggaran aturan lalu lintas yang dilakukan pengendara akan otomatis tercatat pada aplikasi milik pihak kepolisian. Dengan adanya sistem E-Tilang, mendorong masyarakat untuk membayarkan denda lewat bank sehingga menyiasati para oknum kepolisian agar tidak melakukan pungutan liar (pungli). *Electronic Traffic Law Enforcement* (E-TLE) menjadi sebuah terobosan sistem baru para penegak hukum pada biro lalu lintas dengan basis teknologi informasi dengan perantara beragam perangkat elektronik, kamera pengawas (CCTV) dengan resolusi tinggi yang memiliki kemampuan dalam mengabadikan kejadian pelanggaran serta dapat mengetahui plat nomor kendaraan, selanjutnya isi dari rekaman kamera pengawas (CCTV) akan dijadikan barang bukti (Irawan dkk., 2022)

Analisis sentimen dapat didefinisikan terkait pemahaman dalam mengekstrak serta memproses data tekstual secara otomatis dalam memperoleh informasi sentimen yang tercantum pada suatu kalimat opini (Buntoro, 2017). Analisis sentimen termuat dari sistem yang difokuskan untuk observasi mengenai opini para pengguna yang termuat dari sosial media maupun komentar blog dan website untuk memperlihatkan tingkat kepuasan serta sikap para pengguna terhadap suatu pembahasan. Beberapa referensi terhadap beberapa penelitian terkait analisis sentimen dengan menggunakan metode *Support Vector Machine*. Seperti pada penelitian (Tineges dkk., 2020) berjudul “Analisis Sentimen

Terhadap Layanan *Indihome* Berdasarkan Twitter dengan metode klasifikasi SVM”. Penelitian tersebut bertujuan mengklasifikasi layanan *Indihome* dengan metode *support vector machine* dan didapatkan hasil *accuracy* 87%, *presisi* 86%, *recall* 95%, *error rate* 13% dan *f1-score* 90%. Selanjutnya penelitian oleh (Pangestu dkk., 2019) berjudul “Algoritma *Support Vector Machine* untuk klasifikasi Sikap Politik Terhadap Partai Politik Indonesia”. Penelitian ini bertujuan untuk klasifikasi tindakan politik pada partai politik Indonesia melalui metode *Support Vector Machine* dengan hasil *accuracy* 86% dan pengujian pada *10-Fold Cross Validation* dengan rata-rata *accuracy* 71% dan *error rate* 29%.

Pelanggaran dalam lalu lintas umumnya yaitu pelanggaran dalam marka jalan, rambu-rambu lalu lintas dan lampu pengatur lalu lintas seperti larangan berhenti, parkir di tempat-tempat tertentu, menerobos lampu merah dan tidak menggunakan helm. Meskipun demikian masih banyak di kalangan masyarakat yang belum begitu bijak dalam menyikapi penerapan kemajuan teknologi di bidang transportasi yaitu E-Tilang. Hal ini ditandai dengan adanya berbagai opini oleh para pengguna Twitter tentang E-Tilang. Tweet tentang E-Tilang terdiri atas berbagai macam ada yang pro maupun kontra dengan kebijakan penilangan secara elektronik atau E-Tilang. Sarana yang bekerja secara efektif dalam menampung opini pengguna transportasi salah satunya ialah twitter. Twitter menjadi sosial media yang cukup akrab digunakan oleh masyarakat di Indonesia, tentunya dalam hal ini akan memudahkan untuk penghimpunan opini dibandingkan dengan melakukan survey ataupun kuisioner secara langsung. Selanjutnya dilakukan analisis mengenai ulasan pengguna transportasi di twitter dalam mengetahui tanggapan dari pengguna transportasi hasil dari opini tersebut akan diklasifikasi pada komentar positif ataupun negatif, dalam evaluasi tersebut akan diolah dengan analisis sentimen. Teknik klasifikasi pada tugas akhir ini adalah pendekatan klasifikasi *Support Vector Machine*. *Support Vector Machine* ialah metode dengan mempertimbangkan tugas sederhana dalam klasifikasi. Pemilihan metode *Support Vector Machine* pada penelitian ini ialah proses ini dapat diimplikasikan dalam pengklasifikasian dokumen dengan basis teks, sehingga menghasilkan tahap akurasi yang baik serta waktu komputasinya cepat. Tujuan dari penelitian yaitu menerapkan metode *Support Vector Machine* untuk klasifikasi sentimen E-Tilang.

2. KAJIAN PUSTAKA

A. Sentimen Analisis

Analisis sentimen merupakan segmen yang luas dari proses bahasa alami, komputasi linguistik dan *text mining* dalam hal menelaah pendapat, sentimen, evaluasi, penilaian perilaku serta emosi (Bhatia dkk., 2018). Tujuan *opinion mining* untuk menetapkan paradoks gabungan teks pada tulisan yang terdapat di dokumen, kalimat, paragraf yang mengacu ke indikasi negatif, netral dan positif (Romadoni dkk., 2020).

B. E-Tilang

E-Tilang atau *Electronic Traffic Law Enforcement* (E-TLE) merupakan sebuah inovasi sistem penegakan hukum di bidang lalu lintas yang berbasis teknologi informasi menggunakan beragam perangkat elektronik, seperti kamera pengawas (*closed circuit television/CCTV*) beresolusi tinggi yang mampu merekam kejadian pelanggaran dan mendeteksi plat nomor kendaraan. Kamera pengawas beresolusi tinggi ini kemudian dapat dilengkapi dengan perangkat elektronik lainnya seperti *speed radar* yang dapat mendeteksi batas kecepatan kendaraan bermotor. Hasil rekaman kamera pengawas inilah yang akan menjadi barang bukti adanya pelanggaran (Irawan dkk., 2022).

C. *Support Vector Machine*

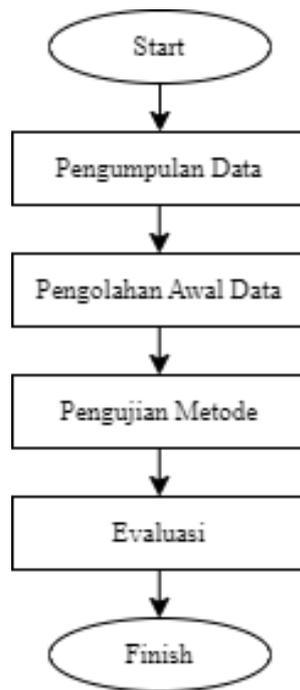
Support vector machine (SVM) adalah metode dengan mempertimbangkan tugas sederhana dalam klasifikasi. Pada umumnya digunakan dalam klasifikasi serta prediksi. Serta abstrak akan menggali *hyperlane* untuk memisahkan secara garis lurus dengan melaksanakan gambaran *nonlinier* secara tepat

menuju arah dimensi yang cukup tinggi, melalui data dua kelas bisa diurai menggunakan *hyperlane* ini dengan *support vector machine* serta *margin* kelas (Campbell, 2011).

Cara kerja *support vector machine* dengan mendeteksi *hyperlane* optimal dengan mengambil jarak atau pemisah diantaradua kelas. *Hyperlane* memiliki *margin* maksimal, jarak antara titik data yang terdekat pada *hyperlane* dinyatakan sebagai *margin*. *Vector* pendukung dinyatakan menjadi titik terdekat dengan *hyperlane* (Tineges dkk., 2020).

3. METODE PENELITIAN

Rancangan penelitian memiliki langkah penting pada penelitian ini yang berlangsung agar pelaksanaan penelitian dapat tertata dengan baik dan sistematis. Penelitian ini memiliki tahapan penelitian untuk memperoleh hasil yang maksimal. Tahapan penelitian analisis sentimen sebagai berikut:



Gambar 1. *Flowchart* Alur Proses Penelitian

a) *Pengumpulan Data*

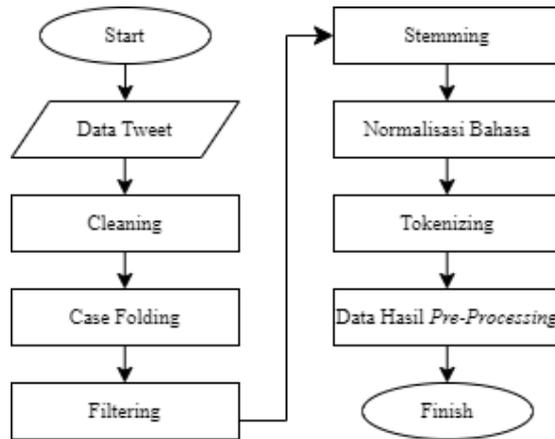
Dalam tahap ini proses pengumpulan data dilaksanakan dengan teknik *crawling* data di twitter menggunakan bahasa pemrograman *python*.

b) *Labeling Data*

Pada penelitian ini dilakukan *labeling data* secara manual yang selanjutnya dilakukan validasi oleh ahli bahasa.

c) *Pre-Processing*

Pada tahapan ini bertujuan untuk membersihkan data yang masih kotor, berikut merupakan diagram alir *pre-processing* :



Gambar 2. Diagram Alir *Text Pre-Processing*

d) Pembobotan Kata

Setelah melakukan tahapan *pre-processing* selanjutnya dilanjutkan dengan tahapan pembobotan kata. Pada tahapan pembobotan kata ini menggunakan TF-IDF, yang dimana pembobotan kata untuk memberikan nilai pada setiap kata/*term*.

$$Tf = 0.5 + 0.5 * \frac{ft,d}{\max(ft,d)} \quad (1)$$

$$IDF = \log \left(\frac{N}{df} \right) \quad (2)$$

$$W = tf * idf \quad (3)$$

Keterangan :

- d = Dokumen
- t = kata pada dokumen
- ft,d = Frekuensi kata pada d
- IDF = *invers document frequency*
- N = Total dokumen
- Df = Banyak dokumen yang mengandung *term*
- W = Bobot kata
- tf = Nilai dari *term frequency* (tf)
- idf = Nilai dari *invers document frequency* (idf)

e) *Klasifikasi Support Vector Machine*

Selanjutnya setelah melakukan proses pembobotan kata/TF-IDF. Pada klasifikasi ini sekedar menilik titik dan ruang di dokumen tersebut untuk intensi pemodelan ruang *vector* yang dipakai untuk memberikan setiap kata dalam dokumen yang diproses.

$$h(x) = \begin{cases} +1, & \text{if } w \cdot x < 0 \\ -1, & \text{if } w \cdot x \geq 0 \end{cases} \quad (4)$$

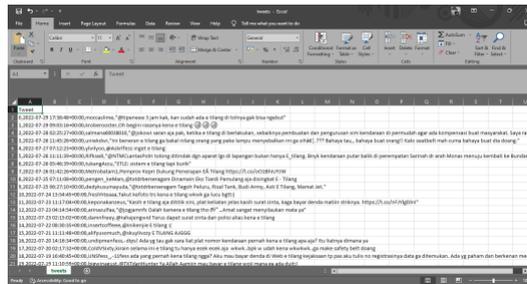
f) *Evaluasi dan Validasi*

Setelah melakukan klasifikasi selanjutnya dilakukan tahapan evaluasi dan validasi, evaluasi dan validasi dilakukan untuk mengetahui sebuah performa pada metode *support vector machine*. Pada tahapan evaluasi dan validasi dilakukan dengan menggunakan *K-Fold cross validaiton* dan dilanjutkan dengan menggunakan *confusion matrix* untuk mendapatkan nilai akurasi dari hasil validasi.

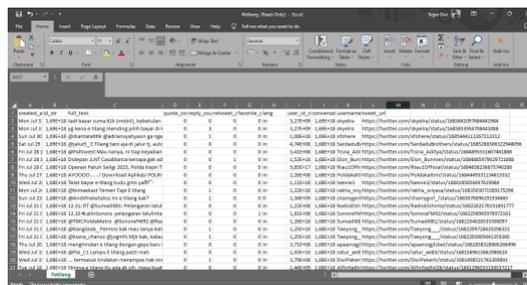
4. HASIL DAN PEMBAHASAN

A. Penambangan Data

Pada penelitian ini dilakukan diperlukan data sebagai pokok bahan analisis yang akan dilakukan. Tahap penambangan data dilakukan dengan cara *crawling*. *Crawling* data twitter dengan menggunakan bahasa pemrograman *python* dengan *keyword* E-Tilang. Pada tahapan ini dilakukan dengan *crawling* data sebanyak 2 kali yang bertujuan untuk membandingkan performa *support vector machine* dengan jumlah data yang berbeda.



Gambar 3. *Crawling* 400 Tweet



Gambar 4. *Crawling* 756 Tweet

B. Text Pre-Processing

Tahap *text pre-processing* selanjutnya diawali dengan penghapusan atribut yang tidak diperlukan seperti *emoticon*, *link* berikut hasil dari *cleansing*, selanjutnya *casefolding* berfungsi untuk merubah huruf besar menjadi kecil, *stemming* berfungsi untuk menghilangkan imbuhan pada setiap kata, *filtering/stopword removal* berfungsi untuk menghapus kata-kata yang tidak memiliki makna, normalisasi bahasa berfungsi untuk mengubah dari bahasa non baku menjadi bahasa baku dan selanjutnya *tokenizing* yang berfungsi untuk memisahkan masing-masing data dengan menggunakan tanda koma(,).

C. Pembobotan Kata TF-IDF

Pembobotan TF-IDF dilakukan setelah dataset melalui *text pre-processing*. Pembobotan TF-IDF pada python akan menggunakan *library TF-IDF Vectorize*. Berikut merupakan hasil dari pembobotan kata TF-IDF.

Tabel 1. Hasil Pembobotan Kata TF-IDF

jam	0.5060923780123743
kak	0.7371414381990228
tilang	-0.027585739746262238
tol	0.48040059804116464
gak	0.3184238342030037
ngebut	0.6216169081056986

D. Klasifikasi *Support Vector Machine*

Setelah dataset melalui proses pembobotan kata selanjutnya dataset diolah untuk klasifikasi *support vector machine* untuk mendapat hasil keputusan sentimen, dengan memanfaatkan *library* *sk-learn*.

```
from sklearn.svm import SVC
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
```

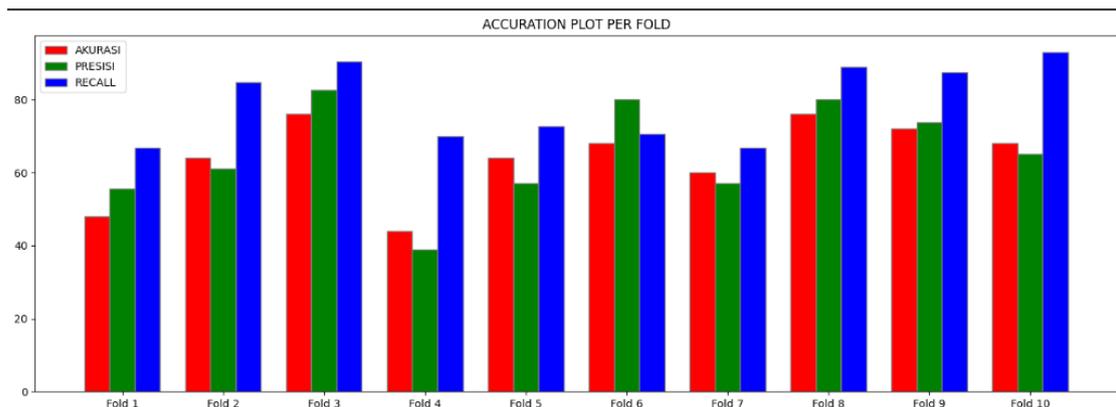
Gambar 5. *Library* *sk-learn*

E. Evaluasi dan Validasi

Pada tahapan evaluasi dan validasi dilakukan untuk mengetahui hasil akurasi dari kinerja algoritma *support vector machine*. *K-Fold* 10 dengan menggunakan data set sebanyak 250 data dihasilkan sebesar:

Tabel 2. Hasil akurasi dengan 250 data

<i>K-Fold Cross Validation</i>	Pengujian	<i>Support Vector Machine</i>
<i>K-Fold</i> 10	Pengujian 1	48%
	Pengujian 2	64%
	Pengujian 3	76%
	Pengujian 4	44%
	Pengujian 5	64%
	Pengujian 6	68%
	Pengujian 7	60%
	Pengujian 8	76%
	Pengujian 9	72%
	Pengujian 10	68%

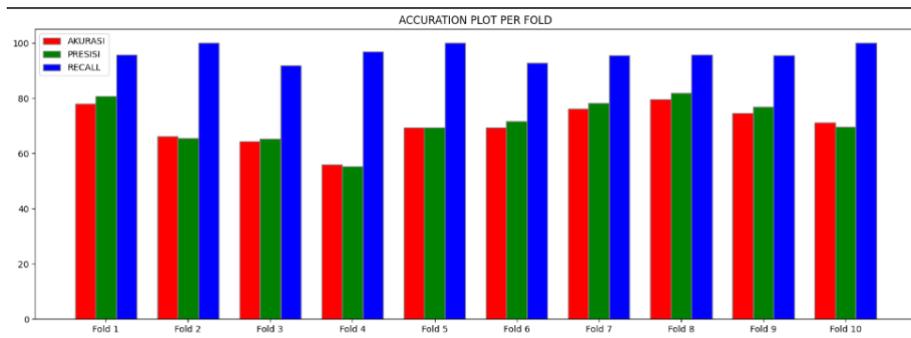


Gambar 6. Hasil akurasi per *fold*

K-Fold 10 dengan menggunakan data set sebanyak 590 data dihasilkan sebesar:

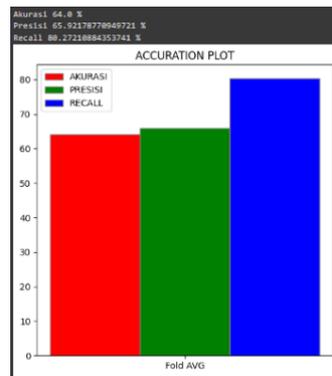
Tabel 3. Hasil akurasi dengan 590 data

<i>K-Fold Cross Validation</i>	<i>Pengujian</i>	<i>Support Vector Machine</i>
<i>K-Fold 10</i>	Pengujian 1	78%
	Pengujian 2	66%
	Pengujian 3	64%
	Pengujian 4	56%
	Pengujian 5	69%
	Pengujian 6	69%
	Pengujian 7	76%
	Pengujian 8	80%
	Pengujian 9	74%
	Pengujian 10	71%

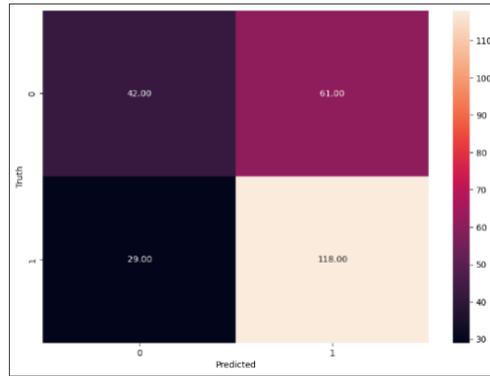


Gambar 7. Hasil akurasi per *fold*

Selanjutnya dilakukan tahapan evaluasi dengan menggunakan *confusion matrix* dengan menggunakan 250 data dengan mendapatkan hasil akurasi sebesar 64%, presisi 65% dan *recall* 80%.



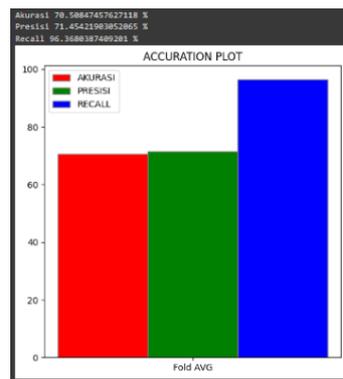
Gambar 8. Grafik hasil *confusion matrix* dengan 250 data



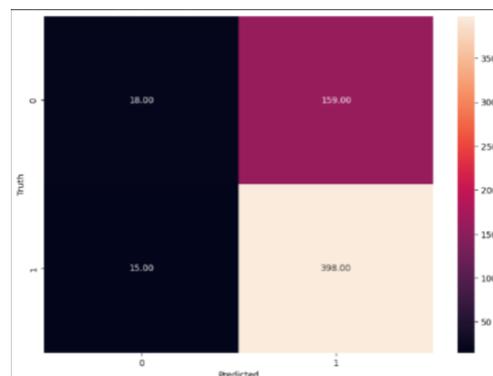
Gambar 9. Confusion matrix 250 data

Berdasarkan gambar di atas dapat dilihat bahwa nilai negatif yang diprediksi benar oleh sistem *True Negative* berjumlah 42 sedangkan untuk nilai negatif yang diprediksi salah oleh sistem *False Negative* berjumlah 29 dan dapat dilihat juga yang bernilai positif benar yang diprediksi oleh sistem bernilai *True Positive* sebanyak 118 dan untuk yang salah prediksi *False Positive* berjumlah 61.

Dan selanjutnya dilakukan dengan menggunakan data set sebanyak 590 dan didapatkan nilai akurasi sebesar 70%, presisi 71% dan recall 96%.



Gambar 10. Hasil confusion matrix dengan 590 data



Gambar 11. Confusion matrix 590 data

Berdasarkan gambar di atas dapat dilihat bahwa nilai negatif yang diprediksi benar oleh sistem *True Negative* berjumlah 18 sedangkan untuk nilai negatif yang diprediksi salah oleh sistem *False Negative*

berjumlah 15 dan dapat dilihat juga yang bernilai positif benar yang diprediksi oleh sistem bernilai *True Positive* sebanyak 398 dan untuk yang salah prediksi *False Positive* berjumlah 159.

5. KESIMPULAN

Berdasarkan hasil analisis perbandingan data set dengan jumlah 250 data dan 590 data dengan memanfaatkan algoritma *support vector machine*, serta evaluasi dan validasi dengan *cross validation* dan *confusion matrix* didapatkan kesimpulan sebagai berikut:

1. Evaluasi dan validasi dengan menggunakan 250 data didapatkan hasil akurasi sebesar 64%, presisi sebesar 65% dan *recall* sebesar 80%.
2. Evaluasi dan validasi dengan menggunakan 490 data didapatkan hasil akurasi sebesar 70%, presisi 71% dan *recall* sebesar 96%.

6. DAFTAR PUSTAKA

- Bhatia, S., Sharma, M., & Bhatia, K. K. (2018). Sentiment Analysis and Mining of Opinions. *Internet of Things and Big Data Analytics Toward Next-Generation Intelligence. Studies in Big Data*, 30(1), 503–523.
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, 2(1), 32–41.
- Colin Campbell, Y. Y. (2011). *Learning with Support Vector Machines (Synthesis Lectures on Artificial Intelligence and Machine Learning)*. Morgan & Claypool Publishers.
- Irawan, I. C., Sekarsari, L. A., Hayam, U., & Perbanas, W. (2022). Analisis Persepsi Manfaat, Sikap, dan Niat Tertib Berlalu Lintas pada Sistem E-Tilang Menggunakan TAM. *FORUM EKONOMI: Jurnal Ekonomi, Manajemen Dan Akuntansi*, 24(3), 547–555.
- Pangestu, S. Y., Astuti, Y., & Farida, L. D. (2019). Algoritma Support Vector Machine Untuk Klasifikasi Sikap Politik Terhadap Partai Politik Indonesia. *Jurnal Mantik Penusa*, 3(1), 236–241.
- Romadoni, F., Umaidah, Y., & Sari, B. N. (2020). Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 9(2), 247–253.
- Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). *Jurnal Media Informatika Budidarma*, 4(3), 650.