



Analisis Sentimen Tanggapan Publik Mengenai *E-Tilang* Melalui *Twitter* Menggunakan Algoritma *K-Nearest Neighbor*

Ade Septiansyah*¹, Mesra Betty Yel², Yuma Akbar³

*^{1,2,3}Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Jakarta, Indonesia

*Email Penulis Korespondensi: adetians@stikomcki.ac.id

Abstrak

Suatu kota harus memiliki sistem transportasi yang baik. Sistem tilang elektronik pemerintah untuk mengawasi lalu lintas disebut *e-tilang*. Banyak pelanggaran lalu lintas yang terjadi di jalan raya yang cenderung menyebabkan kecelakaan dan menambah kemacetan. Beberapa faktor penyebab pelanggaran lalu lintas termasuk ketidakpatuhan individu terhadap peraturan lalu lintas, misalnya larangan berhenti dan parkir di mana pun. Tujuan dari penelitian adalah untuk mengidentifikasi bagaimana pandangan masyarakat Indonesia tentang tilang elektronik dalam penelitian ini digambarkan melalui analisis sentimen tentang tilang elektronik dan mengelompokkan berbagai jenis komentar. *Twitter* adalah platform media sosial yang ideal untuk menyampaikan opini karena mudah digunakan, topik terkini, dan memungkinkan komentar. Setelah data dikumpulkan dari *Twitter*, *preprocessing* dilakukan untuk membersihkan data, *tokenizing*, *case normalization*, *stopword*, dan *stemming*. Selain itu, seleksi mengurangi noise dari label yang tidak relevan. Dataset terdiri dari 1.762 baris record tweets diklasifikasi menghasilkan dua kelas data kelas data positif dan kelas data negatif dengan total 1.342 data. Dalam studi ini, algoritma mesin pengklasifikasi *K-Nearest Neighbor* digunakan data yang sudah ada dari kelas data ini digunakan sebagai data pelatihan untuk mesin pengklasifikasi. Hasil evaluasi menunjukkan akurasi rata-rata 75,41%.

Kata Kunci — *E-Tilang*, Algoritma *K-Nearest Neighbor*, *Text Mining*

Abstract

A city must have a good transportation system. The government's electronic ticketing system for monitoring traffic is called *e-tilang*. Many traffic violations occur on highways which tend to cause accidents and increase traffic jams. Some factors that cause traffic violations include individual non-compliance with traffic rules, for example the prohibition on stopping and parking anywhere. The aim of the research is to identify how Indonesian people's views about electronic tickets in this research are described through sentiment analysis about electronic tickets and grouping various types of comments. *Twitter* is an ideal social media platform for expressing opinions because it is easy to use, current topics and allows comments. After data is collected from *Twitter*, *preprocessing* is carried out to clean the data, *tokenizing*, *case normalization*, *stopwords*, and *stemming*. Additionally, selection reduces noise from irrelevant

labels. The dataset consists of 1,762 rows of tweets records which are classified to produce two data classes, positive data class and negative data class, with a total of 1,342 data. In this study, the K-Nearest Neighbor classifier machine algorithm uses existing data from this data class as training data for the classifier machine. The evaluation results show an average accuracy of 75.41%.

Keywords — *E-Tilang, Algorithm K-Nearest Neighbor, Text Mining*

1. PENDAHULUAN

Sistem tilang elektronik yang digunakan pemerintah untuk mengawasi lalu lintas disebut *e-tilang*. Metode algoritma KNN digunakan untuk membagi data berdasarkan jarak terpendek dari objek data. Data saat ini digunakan untuk menentukan nilai K terbaik untuk algoritma ini. Karena kecepatan dan ketepatan dalam mengklasifikasikan teks, metode ini telah banyak digunakan dalam analisis sentimen. Pemerintah telah menerapkan *E-Tilang* sebagai langkah baru dalam mewujudkan *Good Governance*. Ini akan membuat layanan publik lebih efisien, transparan, dan akuntabel. *E-tilang* adalah digitalisasi proses tilang. Teknologi ini diharapkan menjadikan seluruh proses tilang lebih efisien. Saat ini, tilang menggunakan sistem elektronik, yang sekarang dikenal sebagai sistem tilang elektronik atau *Electronic Traffic Law Enforcement* (ETLE), sebagai hasil dari kemajuan teknologi dan informasi. Oleh karena itu, diharapkan bahwa seluruh proses tilang akan berjalan lebih efisien dan efektif daripada tilang umum yang biasa dilakukan pada masa lalu.

Dalam sistem ini, Aplikasi *E-tilang*, yang dapat diunduh di *Google Playstore*, memungkinkan pelanggar lalu lintas untuk mengunduh dan membayar denda tilang melalui *M-banking* atau *E-banking*, atau mereka juga dapat membayar melalui Bank Rakyat Indonesia (BRI). Polantas memasukkan data pelanggaran ke dalam aplikasi, yang kemudian memberikan nomor registrasi tilang kepada pengendara. Aplikasi *Etilang*, juga dikenal sebagai Tilang Elektronik, telah diberikan kepada setiap petugas lalu lintas. Sesuai dengan undang-undang lalu lintas, aplikasi ini terhubung ke sistem yang dikembangkan oleh Bank Rakyat Indonesia (BRI). Ketika petugas melakukan penilaian, mereka harus memasukkan data pelanggar dan jenis pelanggaran yang dilakukan masyarakat ke dalam sistem, dan data ini akan langsung terhubung ke bank BRI untuk menentukan jumlah tilang yang harus dibayar oleh pelaku [1].

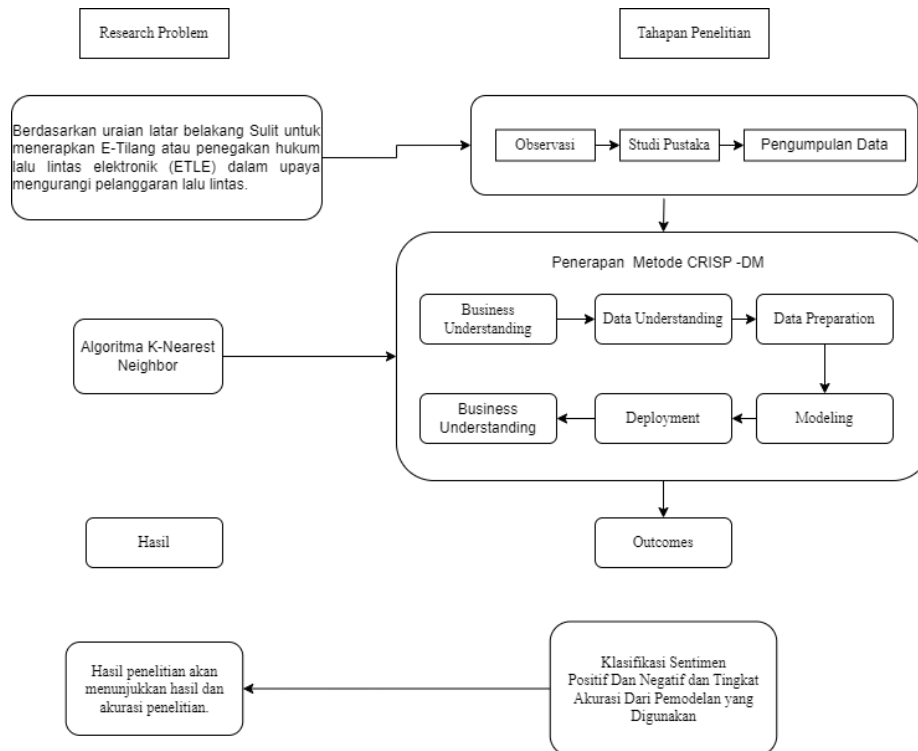
Dalam Penelitian Sebelumnya, Hasil dari penelitian yang melakukan analisis sentimen terhadap sistem *e-tilang* menunjukkan bahwa reaksi masyarakat cenderung netral. Setelah menggunakan proses *Confusion Matrix*, presisi netral 85% dan *recall* netral 95% diperoleh. Namun, presisi ini harus ditingkatkan untuk kelas negatif dan positif. Keseluruhan akurasi adalah 80%. Penelitian ini sangat membantu pemerintah dalam menilai program yang dijalankan, khususnya yang berkaitan dengan sistem *e-tilang*, karena analisis ini memberikan informasi penting bagi pemerintah dalam memahami respons masyarakat terhadap sistem *e-tilang*. Dengan melakukan analisis sentimen masyarakat terhadap sistem *e-tilang*, pemerintah dapat mendapatkan informasi penting tentang pandangan dan reaksi masyarakat terhadap penerapan sistem ini [2].

Indonesia saat ini melakukan persiapan penggunaan teknologi dalam bidang penegakan hukum. Tujuan akhir kami adalah untuk memastikan bahwa lalu lintas aman, tertib, dan lancar (KAMSELTIBCAR). Penelitian ini menggunakan algoritma KNN untuk mengoptimalkan posting sistem *e-Tilang* di media sosial *Twitter* dengan *text mining* dan fitur seleksi *Rapidminer*. *Text mining* adalah teknik untuk menggali kumpulan dokumen teks untuk mendapatkan intisarinnya [3].

2. METODE PENELITIAN

Pada tahap pengumpulan data, data yang dikumpulkan adalah *E-Tilang* pada tahun 2022 - 2024 data didapatkan melalui *Crawling Data Google Colab*. Dataset terdiri dari 1.762 baris *record tweets* diklasifikasi menghasilkan dengan 1.342 data kualitatif yang digunakan dalam

penelitian, dua kelas data positif dan kelas data negatif digunakan. Selanjutnya, menerapkan model proses *data mining standard cross-industry* (CRISP-DM). Untuk memproses data dan membuat pengujian model, CRISP-DM akan digunakan. *Understanding Business, Data Pemahaman, Persiapan, Modeling, Evaluasi, dan Penggunaan* adalah komponen CRISP-DM. Hasil penelitian akan menunjukkan hasil dan akurasi penelitian. Gambar 1 merupakan diagram tahapan penerapan metodologi.



Gambar 1 Diagram Penerapan Metodologi

2.1 Rancangan Pengujian

Pada tahap ini, untuk memecahkan masalah, metode *Cross Industry Standard for Data Mining* (CRISP-DM) digunakan. Metode ini menggunakan model proses pengembangan data yang sering digunakan oleh para ahli. Dalam penelitian ini, CRISP-DM terdiri dari enam tahap, yang diuraikan sebagai berikut.

2.2 Pemahaman Bisnis

Tujuan dari Dalam tahap pemahaman bisnis, yang juga dikenal sebagai "pemahaman bisnis", fokus ditempatkan pada tujuan dan kebutuhan, apakah itu berkaitan dengan bisnis, penelitian, atau kebutuhan lain [4]. Tujuan penelitian ini adalah untuk menggunakan algoritma *K-Nearest Neighbor* untuk mengkategorikan tanggapan masyarakat pada *E-Tilang* di *Twitter*. Selanjutnya, tingkat akurasi algoritma ini diukur untuk mengklasifikasikan sentimen positif dan negatif.

2.3 Pemahaman Data

Dalam langkah ini, pemahaman data (*data understanding*) memberikan fondasi analitik untuk penelitian dengan menguraikan dan menemukan masalah potensial yang terdapat di dalam data. Data *E-Tilang*, yang dikumpulkan dari 1.762 *record* tweets, dikumpulkan melalui *Crawling Data Google Colab* dari tahun 2022–2024, memiliki enam atribut: *Username, Hastags, Time Stamp, Replies, Tweet Text, dan Retweet*. Data ini akan diproses sebelum diproses untuk dibersihkan dan diubah [5].

2. 4 Data Preparation

Selanjutnya adalah persiapan data, yang mencakup semua tindakan yang diperlukan untuk menghasilkan data akhir atau data yang akan digunakan dalam alat pemodelan dari data sebelumnya. Pada tahap ini, pembersihan data dan transformasi data juga dilakukan. Tahapan persiapan data adalah sebagai berikut [6]:

1. Pengumpulan data Dimulai dengan mengumpulkan data *record* tanggapan *E-Tilang* dari masyarakat melalui *Twitter*.
2. Pembersihan data adalah proses memilih kembali data yang telah diperoleh untuk mengidentifikasi data yang salah, tidak lengkap, atau tidak benar dan untuk meningkatkan kualitas data. Dalam proses klasifikasi, nama, jenis kelamin, dan tanggal pengukuran akan dihapus dari data penelitian. Pada aplikasi Rapidminer, *Operator Replace* digunakan untuk menyelesaikan proses pembersihan data.
3. Analisis Data: Setelah data dikumpulkan, perlu dilakukan analisis untuk menemukan jawaban atas masalah penelitian. Bergantung pada jenis data yang dikumpulkan, peneliti dapat menggunakan teknik analisis statistik.

2. 5 Permodelan (Modelling)

Proses pemodelan pertama akan melibatkan metode klasifikasi menggunakan algoritma *K-Nearest Neighbor* seperti diperlihatkan pada formula (1). Setelah itu, model klasifikasi data yang terdiri dari pelatihan dan pengujian data akan dibuatkan. Selanjutnya, tingkat akurasi akan diukur dengan algoritma *K-Nearest Neighbor* dengan menggunakan *Confusion Matrix*. Sebelum melakukan data uji ada beberapa tahap yang perlu dilakukan proses pengambilan data melalui *google colab*, *preprocessing data rapidminer*, klasifikasi algoritma *K-Nearest Neighbor*, yang merupakan produk dari model *K-Nearest Neighbor* [7].

$$d_i = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan: d_i adalah jarak, x adalah data latihan, y adalah data uji, n adalah dimensi data, dan i adalah variabel data.

2. 6 Evaluasi

Peneliti menggunakan *RapidMiner* untuk melakukan proses eksperimen pada tahap pengujian metode. Data pelatihan diambil dari komentar video *E-Tilang* di *Twitter*. Data tersebut kemudian dibagi menjadi pendapat yang positif dan negatif [8]. Peneliti menggunakan algoritma KNN, yang menghasilkan *vektor performance* dengan *matrix confusion*. Selanjutnya, nilai AUC dihitung dengan menggunakan kurva ROC.

1. Confusion Matrix

Confusion Matrix memberikan hasil prediksi secara keseluruhan dari nilai akurasi dan melihat kinerja pengklasifikasi, yaitu seberapa sering kasus *class X* yang benar diklasifikasikan sebagai *class X* atau kesalahan klasifikasi *class* lainnya. Tabel 1 memperlihatkan *dataset* dengan hanya dua kelas: positif dan negatif. Jumlah *record* positif dalam *dataset* yang diklasifikasikan positif (TP), jumlah *record* negatif dalam *dataset* yang diklasifikasikan negatif (TN), dan jumlah *record* positif dalam *dataset* yang diklasifikasikan negatif (FN).

Tabel 1 *Confusion Matrix*

Correct Classification	Classified As	
	+	-
+	True Positif	False Negatif
-	False Positif	True Negatif

2. Kurva ROC

Kurva ROC untuk *Kurva ROC*, yang menunjukkan akurasi dan perbandingan klasifikasi secara visual, digunakan untuk mengklasifikasikan keakuratan pengujian menggunakan AUC. *Kurva ROC* adalah grafik dua dimensi, dengan *false positif* ditunjukkan sebagai garis horizontal

dan *true positif* ditunjukkan sebagai garis vertical [9]. Rentang nilai dapat dikelompokkan dalam batasan sebagai berikut:

0.90 -1.00 = Peringkat Klasifikasi yang sangat baik.

0.80 -0.90 = Klasifikasi yang baik.

0.70 -0.80 = Klasifikasi Wajar.

0.60 -0.70 = Klasifikasi Buruk.

0.50 -0.60 = Kegagalan.

2.7 Penyebaran (Deployment)

Proses CRISP-DM yang paling penting adalah tahap *deployment*, atau rencana penggunaan model. Perencanaan *deployment* dimulai selama pemahaman bisnis dan harus mencakup teknik untuk membuat nilai model, mengubah skor keputusan, dan memasukkan keputusan ke dalam sistem operasional [10]. Pada akhirnya, rencana pengembangan sistem mengakui bahwa tidak ada model yang statis; model terdiri dari data yang diwakili sebagai data yang telah terjadi pada waktu tertentu, sehingga karakteristik data dapat berubah dengan waktu. Selain itu, model harus dipantau dan mungkin diganti dengan model yang lebih baik. Pada tahap ini, hasil penelitian dan data yang dikumpulkan didistribusikan untuk digunakan dalam praktik laporan dan komunikasi dari pengetahuan yang telah diperoleh melalui pemodelan dan evaluasi. Proses penyebaran mencakup penciptaan laporan dan jurnal tentang proses *data mining*.

3. HASIL DAN PEMBAHASAN

3.1 Alat dan Bahan Penelitian

Membutuhkan alat untuk mendukung penelitian. Perangkat lunak dan perangkat keras digunakan untuk melakukan penelitian.

3.1.1 Perangkat Lunak (Software)

Tabel 2 menunjukkan alat perangkat lunak (*software*), versi, dan fungsinya yang digunakan dalam penelitian ini.

Tabel 2 Spesifikasi Software

No	Software	Versi	Fungsi
1	Microsoft Excel	2019	Melabeli sentimen yang diperoleh dari hasil <i>crawling data</i> di <i>Twitter</i> untuk digunakan dalam data latih dan menyimpan data.
2	Rapidminer Studio	10.0	Untuk menerapkan teknik <i>text mining</i> yang telah digunakan dalam penelitian ini.

3.1.2 Perangkat Keras (Hardware)

Tabel 3 menunjukkan alat perangkat keras (*Hardware*), jenisnya, dan spesifikasinya yang digunakan dalam penelitian ini.

Tabel 3 Spesifikasi Hardware

Jenis Hardware	Spesifikasi
Model	Toshiba Dynabook R732/H
Processor	Intel Core i3-3120M
Memory	DDR3 4Gb
GPU	HD Graphics 4000
Storage	SSD 128Gb, HDD 128Gb
Display	1.366 x 768 piksel pada 13.3 inci <i>High Definition</i> (HD).

3.2 Implementasi Dan Pengujian

Data dikumpulkan dari *platform* media sosial *Twitter* menggunakan alat pengumpul data *Google Colab* yang disebut *Auth-Token*. Data yang dikumpulkan mencakup 1.762 komentar pengguna pada *Tweet* tanggapan masyarakat publik mengenai berbagai *E-Tilang*. *Dataset* publik digunakan dalam proses pengumpulan data penelitian ini [11].

3.2.1 CRISP-DM

Metode *Data Mining Standar Cross-Industri* (CRISP-DM), yang terdiri dari 6 tahapan, dapat digunakan untuk melakukan pengujian ini.

3.2.2 Pemahaman Bisnis

1. Tujuan

Penelitian ini bertujuan untuk menentukan tanggapan masyarakat publik terhadap *e-tilang* antara item. Hal ini dilakukan untuk membantu meningkatkan pengambil keputusan dalam menentukan sentimen mana yang mendekati jarak [12].

2. Menilai Situasi

Evaluasi situasi *Google Colab Python* saat ini menunjukkan bahwa penelitian ini belum dapat menentukan tanggapan masyarakat publik terhadap *E-Tilang* antar item. Oleh karena itu, harus memahami tujuan pelabelan data untuk *data mining*.

3.2.3 Pemahaman Data (Data Understanding)

Untuk mencapai tujuan menentukan efektivitas metode CRISP-DM, tahap selanjutnya adalah pemahaman data dan efisien tanggapan masyarakat terhadap *E-Tilang*. Tahap ini dimulai dengan data yang telah dikumpulkan. Data tanggapan tersebut merupakan *dataset* yang dikumpulkan dari *Crawling Data Twitter Google Colab Python*. Dokumen tersebut berbentuk csv dengan 1.762 *record*, dan 1.342 *record* difilter. Tabel 4 menunjukkan langkah-langkah dalam fase pemahaman data (fase pemahaman data).

Tabel 4 Sampel Dataset

Attribut	Keterangan
<i>Username</i>	Pengguna
<i>Hashtags</i>	Tagar untuk menghubungkan
<i>Time Stamp</i>	Ketika <i>tweet</i> dikirim
<i>Replies</i>	Respon terhadap pesan
<i>Tweet Text</i>	Tujuan dari <i>tweet</i>
<i>Retweet</i>	Berbagi <i>tweet</i>

3.2.4 Data Preparation

Untuk tahap mendeskripsikan data, *dataset* tanggapan masyarakat terhadap *E-Tilang*, yang dikumpulkan selama tiga tahun, mencakup 1.762 *record* dan memiliki enam atribut: *Username*, *Hastags*, *Time Stamp*, *Replies*, *Tweet Text*, dan *Retweet*. Hanya dua atribut dari enam atribut tersebut digunakan seperti diperlihatkan pada Gambar 2.

	A	B	C	D	E	F	G
1	Text	Status					
2	Cctv etilang dimananya teh	negatif					
3	siap nerobos etilang	positif					
4	Ok infonya tapi kurang jelas Jam berapa mulai ganjil jam berapa mulai genap Kok tiba dapet surat Etilang kasih tau dhu lah biar kita macetkan pantura saja kalo ada ganjil genap	positif					
5	Mgkin kenaikan pajak atv ad denda etilang	positif					
6	Eh trs pake etilang buset org mudik dirampok juga maaf bgt gue emosi	positif					
7	Ribet bgt ngurus etilang	positif					
8	Sekarang saja di perempatan Aslamah sama banjo ahn lama ada cctv etilang Mending pake helm kata aku	positif					
9	Apalagi semenjak ada etilang kan	positif					
10	ada yang pernah kena etilang tapi salah gak	negatif					
11	Ga komplek bang kena etilang tapi gak ada suratnya keramah	positif					
12	Khusus jakarta ya app nya saya bukan di jakarta Itu di app nya ada keterangannya mas kalau terblokir karena etilang	negatif					
13	Usahkan jalan di sudirman sampai kefoto etilang	positif					
14	Dirimi dari arah mana juju Perasaan etilang tidak diberlakukan	positif					
15	Ngga bisa pake seatbelt itu woyy Kalau kejeperet eTilang gimana	negatif					
16	hidup lagi seru serunya diingetin harus bayar etilang	positif					
17	Udah keliatan kok pak dari dhu penegakan hukum emang kurang contohnya buang sampah sembarangan penerapan etilang bukannya pengendaranya jadi terib malah plat nomornya pada diganti antrian itu akibat dari hukum yang berlaku antrian tak hanya berkendara tapi dalam banyak hal selama bisa menyus maka antrian tak berlaku	positif					

Gambar 2 Data yang digunakan (Data uji)

3.2.5 Pemodelan Data Mining

Tahap ini secara langsung melibatkan metode pengolahan data dengan memilih metode dan memilih algoritma yang akan digunakan. Dalam penelitian ini, pemodelan *data mining*

menggunakan *csv* untuk menghitung nilai Algoritma *K-Nearest Neighbor* kemudian akan digunakan untuk mengukur tingkat akurasi dengan *RapidMiner versi 10.0* dan *Confusion Matrix*. Pada tahap ini, nilai dukungan item digunakan untuk mencari kombinasi item yang memenuhi minimal nilai dukungan yang ada dalam database [13].

3.2.6 Evaluasi

Pada tahap ini, analisis atau pengukuran ketepatan pemodelan akan dilakukan. Perhitungan akurasi konsep data mining dilakukan dengan metode *Confusion Matrix*. Metode ini menghasilkan nilai *accuracy*, *precision*, *recall*, dan *f-measure* untuk penelitian. Untuk memastikan bahwa pemodelan yang digunakan sesuai dan tepat dengan rencana awal penelitian, evaluasi dilakukan. Hasil evaluasi juga akan membantu menentukan apakah penelitian harus dilakukan lagi atau diulang dari awal karena tidak sesuai dengan rencana awal.

3.2.7 Deployment

Proses *CRISP-DM* yang paling penting adalah tahap *deployment*, atau rencana penggunaan model. *Deployment* adalah tahap akhir dalam proses pembuatan laporan hasil *data mining*. Penelitian ini berisi pola atau informasi yang ditemukan selama proses *data mining*. Pada titik ini, data dan hasil penelitian didistribusikan untuk digunakan dalam laporan dan presentasi dari pengetahuan yang diperoleh dari pemodelan dan evaluasi. Pembuatan laporan dan jurnal tentang proses *data mining* adalah bagian dari proses penyebaran [14].

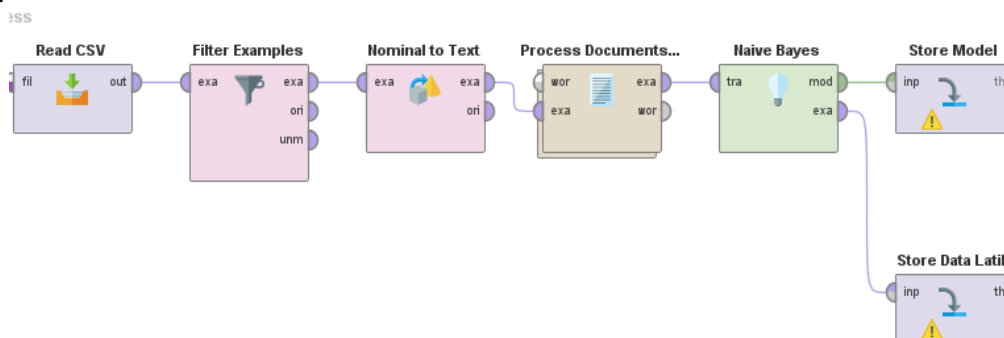
3.3 Hasil Akhir Pengujian

3.3.1 Pengumpulan Data

Tahap pertama data dikumpulkan dari *platform* media sosial *Twitter* dengan menggunakan alat pengumpul data *auth_token*. Data yang dikumpulkan pada Data publik yang dikumpulkan akan digunakan dalam penelitian ini. *tweet* pengguna *Twitter* menggunakan *Google Colab Python* dengan kata kunci "*E-tilang*" sebanyak 1.762 data yang kemudian *filtered* *untagged*. Data yang dikumpulkan meliputi komentar pengguna pada *Tweet* tanggapan masyarakat terhadap berbagai *E-tilang*.

3.3.2 Pelabelan Dataset

125 data akan dilabeli secara manual untuk digunakan sebagai data latih, dan 1.217 akan digunakan sebagai data model yang belum terlabeli. Untuk melakukannya, Penulis menggunakan *RapidMiner* untuk melebeli data model, atau data kosong, secara otomatis. Data diberi label "positif" dan "negatif". Gambar 3 memperlihatkan tahapan dalam penentuan data model dan data latih.

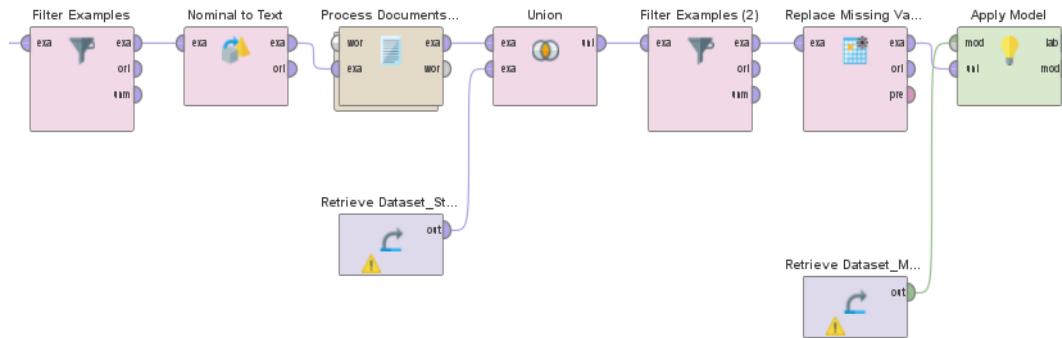


Gambar 3 Menghasilkan Data Latih dan Model

Tujuan tahapan ini adalah untuk membuat sistem belajar yang menggunakan data (dalam hal ini label atau sentimen). Gambar 4 dan Gambar 5 merupakan proses pelabelan yang dilakukan dalam penelitian ini. Jadi, *operator union* secara otomatis melabeli *dataset* yang digunakan di *RapidMiner Studio* oleh peneliti terlebih dahulu dibatasi oleh beberapa kriteria [15]:

1. *Tweets* yang mengandung kata-kata positif seperti "baik", "kompeten", "menakutkan", dan kata-kata positif lainnya akan dikategorikan sebagai positif.

2. *Tweets* yang mengandung maksud bahwa pengguna setuju mengikuti Proses *E-tilang* akan dikategorikan sebagai positif.



Gambar 4 Prediksi Label Data Kosong

Result History: ExampleSet (Apply Model)

Open in: Turbo Prep, Auto Model

Filter (1,217 / 1,217 examples): all

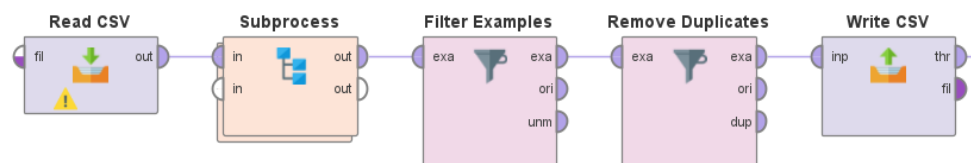
Row No.	Status	prediction(S...	confidence(...	confidence(...	text	Minggu	Neger	abalkan	abang
1	?	positif	1	0	dips area te...	0	0	0	0
2	?	positif	1	0	nebeng chan...	0	0	0	0
3	?	positif	1	0	nolif kena eti...	0	0	0	0
4	?	negatif	0	1	urus kerjaan f...	0	0	0	0
5	?	negatif	0	1	mobil temen ...	0	0	0	0
6	?	positif	1	0	oknuh lindak t...	0	0	0	0
7	?	positif	1	0	banyaknya la...	0	0	0	0
8	?	negatif	0	1	anjir keknya e...	0	0	0	0
9	?	positif	1	0	alamat mala...	0	0	0	0
10	?	positif	1	0	selamat ma...	0	0	0	0
11	?	negatif	0	1	konfirmasi et...	0	0	0	0
12	?	negatif	0	1	kalo fokus tla...	0	0	0	0
13	?	positif	1	0	kana eliang ...	0	0	0	0

ExampleSet (1,217 examples, 5 special attributes, 3,524 regular attributes)

Gambar 5 Hasil Prediksi Label Data

3.3.3 Tahap Pembersihan Data

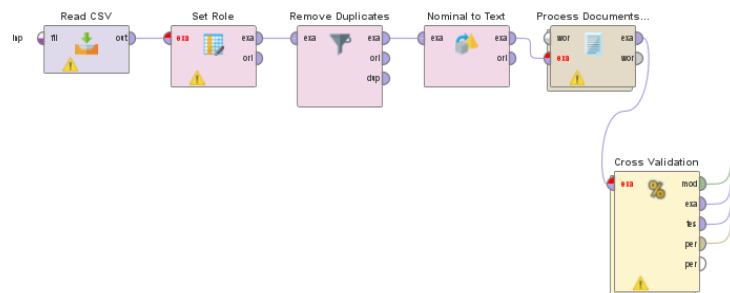
Tahap Pembersihan Data (*Cleansing Data*) mencakup penghapusan *tweet* dari kata-kata yang tidak diperlukan, seperti karakter *hashtag* "#", pernyataan "@", dan URL atau simbol yang tidak dibutuhkan selama proses sentimen analisis. Didalam Subproses terdapat kumpulan Data *Replace* seperti *Replace URL*, *Replace URL 2*, *Replace Mention*, Dan *Replace Simbol*. *Filter Examples* memfilter *text data* yang sudah ada atau terisi, *Remove duplikat* menghapus data yang sama. Gambar 6 memperlihatkan tahapan yang dilakukan dalam *Cleaning Data*.



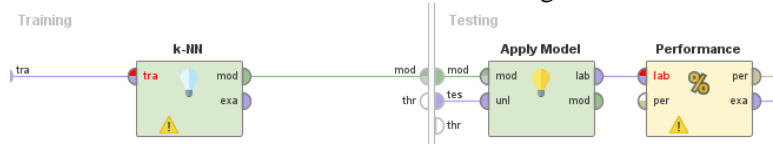
Gambar 6 *Cleaning Data*

3.3.4 Tahap Nilai Akurasi K-Nearest Neighbor

Hasil perhitungan Dari 1.342 data uji, 447 menunjukkan sentimen negatif dan 895 menunjukkan sentimen positif, menunjukkan nilai akurasi 75,41%. Gambar 7 dan Gambar 8 memperlihatkan visualisasi implementasi K-NN dalam *RapidMiner*.



Gambar 7 Model Algoritma KNN



Gambar 8 Cross Validation

accuracy: 75.41% +/- 3.23% (micro average: 75.41%)

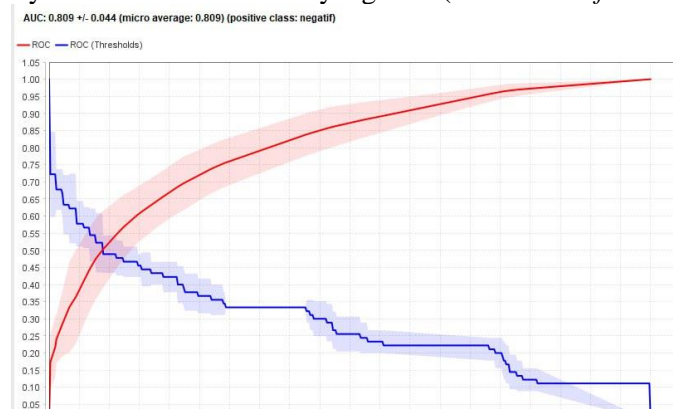
	true positif	true negatif	class precision
pred. positif	858	293	74.54%
pred. negatif	37	154	80.63%
class recall	95.87%	34.45%	

Gambar 9 Confusion Matrix Algoritma KNN

Mengacu pada Gambar 9 menunjukkan bahwa akurasi *K-Nearest Neighbor* dengan hasil uji coba sebesar 75,41%. *Perofmance Vector* juga menunjukkan jumlah dan jenis data mana yang sesuai atau tidak sesuai antara hasil uji coba dan perhitungan *K-Nearest Neighbor*. Hasil perhitungan dapat dikategorikan menjadi *true positif*, *true negatif*, *false positif*, dan *false negatif*. Setiap ukuran evaluasi memiliki kekuatan dan kelemahan yang berbeda, menggunakan kombinasi dari beberapa ukuran evaluasi dapat memberikan gambaran yang lebih komprehensif tentang performa model. Pada *Perfomance vector* diketahui bahwa *True Positif* : 858, *False Positif*: 37, *False Negatif* : 293, *True Negatif*: 154

3.3.5 Tahap Nilai Kurva ROC

Melihat Gambar 10 terlihat algoritma *K-Nearest Neighbor* memiliki kurva ROC sebesar 0,809 berdasarkan hasilnya dikatakan Klasifikasi yang baik (*Good Classification*).



Gambar 10 Hasil Kurva ROC Algoritma KNN

3.3.6 Hasil Akhir

Dari 1.342 data uji, 895 menunjukkan sentimen negatif dan 447 menunjukkan sentimen positif, hasil akhir *K-Nearest Neighbor* (KNN) menunjukkan nilai akurasi sebesar 75,41%. Hasil

-
- [3] Z. Nanda Aulia, G. Kuncoro Jati, dan I. Santoso, “ANALISIS SENTIMEN TANGGAPAN PUBLIC MENGENAI E-TILANG MELALUI MEDIA SOSIAL YOUTUBE MENGGUNAKAN ALGORITMA NAIVE BAYES,” Jakarta, 2023.
- [4] W. Pramusinto, “3 rd Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) 30 Agustus 2023-Jakarta,” 2023.
- [5] B. Adi Prayogo, “ANALISIS TERHADAP INDIKATOR PELANGGARAN DALAM SISTEM TILANG ELEKTRONIK WILAYAH HUKUM POLRESTA SIDOARJO ANALYSIS OF VIOLATION INDICATORS IN THE SIDOARJO POLRESTA ELECTRONIC SYSTEM,” 2023.
- [6] M. S. Amrullah dan S. Fachri Pane, “Systematic Literature Review: Analisa Sentimen Masyarakat terhadap Penerapan Peraturan E-TLE,” JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST), vol. 4, no. 1, hlm. 2723–1453, 2023.
- [7] D. Maulina Putri, H. E. Rosyadi, dan S. Cahyani, “ANALYSIS OF E-TILANG WEBSITE AT STATE ATTORNEY AT BATU CITY,” Antivirus : Jurnal Ilmiah Teknik Informatika, vol. 17, no. 1, hlm. 78–86, Agu 2023.
- [8] D. Oktavia dan Y. R. Ramadahan, “Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM),” Media Online), vol. 4, no. 1, hlm. 407–417, 2023.
- [9] F. Haradongan, “KAJIAN PELAKSANAAN E-TILANG UNTUK ANGKUTAN BARANG,” Jurnal Transportasi, Logistik, dan Aviassi, vol. 1, no. 1, hlm. 110–116, Jul 2021.
- [10] S. K. Dirjen, P. Riset, D. Pengembangan, R. Dikti, dan I. Firman Maulana, “Terakreditasi SINTA Peringkat 2 Penerapan Firebase Realtime Database pada Aplikasi E-Tilang Smartphone berbasis Mobile Android,” masa berlaku mulai, vol. 1, no. 3, hlm. 854–863, 2021.
- [11] A. Putra Wardana, A. Nugraha, K. Kunci, : Efektifitas, T. Elektronik, dan K. Hukum, “EFEKTIFITAS PENERAPAN PROGAM E-TILANG DALAM MENINGKATKAN KESADARAN HUKUM MASYARAKAT DALAM BERLALU LINTAS DI KOTA KLATEN,” JURNAL MADANI HUKUM - Jurnal Ilmu Sosial Dan Hukum, hlm. 72–80, 2023.
- [12] Lutfina Zunia Apriliana, “EFEKTIVITAS PENGGUNAAN E-TILANG TERHADAP PELANGGARAN LALU LINTAS DI POLRES MAGELANG,” Jurnal Komunikasi Hukum (JKH), vol. 5(2), hlm. 1–11, 2019.
- [13] E. R. Efendi, “PENERAPAN HUKUM TERHADAP E - TILANG DALAM UPAYA PENERTIBAN LALU LINTAS PADA UNDANG - UNDANG NO 22 TAHUN 2009 TENTANG LALU LINTAS DAN ANGKUTAN JALAN,” Bureaucracy Journal : Indonesia Journal of Law and Social-Political Governance, vol. 2, no. 2, hlm. 521–538, Agu 2022.
- [14] Sherly Ayu Puspita Sari, Marisa Elsera, dan Emmy Solina, “Tindakan Pengendara Motor Terhadap Pelanggaran Aturan Lalu Lintas Dalam Pelaksanaan E-Tilang Di Kota Batam ,” JISHUM (Jurnal Ilmu Sosial dan Humaniora), vol. Vol. 1, No. 3, hlm. 565–578, 2023.
- [15] S. Tinggi dan I. Kepolisian, “Penegakan Hukum Lalu Lintas Melalui Sistem e-Tilang Agung Asmara A Wahyurudhanto Sutrisno,” 2019.