

**Hak Cipta Dilindungi Undang-1**

1. Dilarang mengutip sebagian

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



mencantumkan dan menyebutkan sumber:

**PENERAPAN ALGORITMA *SUPPORT VECTOR MACHINE*
(SVM) UNTUK KLASIFIKASI SENTIMEN ANALISIS
TERHADAP OPINI POSTINGAN TWITTER (MAXIM)**

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

ANGGI PRANATA

11551102853



UIN SUSKA RIAU

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU**

PEKANBARU

2022

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN**PENERAPAN ALGORITMA *SUPPORT VECTOR MACHINE*
(SVM) UNTUK KLASIFIKASI SENTIMEN ANALISIS
TERHADAP OPINI POSTINGAN TWITTER (MAXIM)****TUGAS AKHIR**

Oleh

ANGGI PRANATA**11551102853**

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 23 Juni 2022

Pembimbing,

ELVIA BUDIANITA, ST, M.Cs**NIP. 19801018 200710 1 002**



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

**PENERAPAN ALGORITMA *SUPPORT VECTOR MACHINE*
(SVM) UNTUK KLASIFIKASI SENTIMEN ANALISIS
TERHADAP OPINI POSTINGAN TWITTER (MAXIM)**

Oleh

ANGGI PRANATA

11551102853

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 23 Juni 2022

Mengesahkan,
Ketua Jurusan,



Dr. Hartono, M.Pd

NIP. 19640301 199203 1 003

Iwan Iskandar, M.T

NIP. 19821216 201503 1 003

DEWAN PENGUJI

- | | |
|--------------|---------------------------------|
| Ketua | : Iwan Iskandar, M.T |
| Pembimbing I | : Elvia Budianita, S.T, M.Cs |
| Penguji I | : Yusra, S.T, M.T |
| Penguji II | : Eka Pandu Cynthia, S.T, M.Kom |



LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan didalam daftar pustaka.

Pekanbaru, 23 Juni 2022

Yang membuat pernyataan,

ANGGI PRANATA

NIM. 11551102853

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Allah akan meninggikan derajat orang-orang yang beriman di antaramu dan orang-orang yang diberi ilmu pengetahuan.

Alhamdulillah...

Rasa syukur kuhaturkan kepada-Mu, Yaa Allah yang Maha Ber-Ilmu, hanya karena karuniaMu sajalah hamba-Mu akhirnya dapat menyelesaikan Tugas Akhir ini

Tugas Akhir ini kupersembahkan untuk Papa dan Mama. Tentulah tidak akan tergantung semua jasa, pengorbanan, tetes keringat dan rasa letih itu, hanya dengan karya kecil dariku ini... Namun semoga dengan ini, aku dapat mengukir sebaris senyum bahagia di hati Papa dan Mama... Jika boleh kujabarkan cinta,, tentulah tidak pernah dapat seindah rasa syukurku menjadi anakmu... Terimakasih untuk semua rangkaian do'a, kasih sayang serta ilmu yang berharga...

Dan tidak lupa kupersembahkan untuk kakak dan adik ku tersayang, terimakasih untuk semua dukungan dan doa yang telah diberikan selama ini...

Juga, kupersembahkan untuk semua keluargaku dan kerabat... Semua kesulitan seolah lenyap saat mengingat bahwa aku memiliki dukungan darimu semua.

Dan yang terakhir Tugas Akhir kupersembahkan kepada Pujaan Hati yang jauh disana

Alhamdulillah, Allah menganugerahiku keluarga yang indah...

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Klasifikasi Sentimen Terhadap Maxim Menggunakan Algoritma SVM Pada Media Sosial Twitter

Anggi Pranata^{1*}, Elvia Budianita², Yusra³ dan Eka Pandu Cynthia⁴

Teknik Informatika UIN Sultan Syarif Kasim Riau

Jl. Subrantas Km.15, Pekanbaru, 28293, 0761-562223

*e-mail: anggi.pranata1@students.uin-suska.ac.id

Abstrak - Twitter merupakan media sosial yang penggunanya paling pesat. Seiring perkembangan waktu, setiap ojek online memiliki popularitas di masing-masing user. Contoh saja Maxim, pendatang baru yang menyediakan layanan yang berbeda dengan ojek lain. Salah satu aktivitas yang biasa dilakukan para fans atau haters kepada akun twitter ojek online yaitu memberikan komentar pada tweet, untuk mengetahui komentar para fans dan haters diperlukan klasifikasi dengan menggunakan *machine learning*. Pada penelitian ini, langkah pertama adalah dilakukan proses sortir dan pemberian label pada data tersebut. Hasilnya akan tercipta 3 label yaitu label data positif, netral dan label data negatif dengan jumlah 1200 data. Selanjutnya melakukan analisa *preprocessing* data yang meliputi *case folding*, *cleansing data*, *tokenizing*, *filtering* dan *stemming*. Lalu dilakukan pembobotan dengan metode TF-IDF dan diklasifikasikan dengan metode *Support Vector Machine*. Hasil pengujian dilakukan dengan metode *Confusion Matrix*, berdasarkan hasil pengujian diperoleh akurasi terbaik pada perbandingan data 90:10 sebesar 85% dengan menggunakan Kernel *RBF* dan *Polynomial*, dilanjutkan dengan kernel *Sigmoid* sebesar 82,5% dimana hasil klasifikasi didominasi kalimat positif.

Kata kunci: analisis sentiment, klasifikasi, ojek online, support vector machine, twitter

Abstract - Twitter is one of the Social Medias which has a rapid user. Over the time, every Ojek Online, has its own popularity among their users. Maxim, for instance, a newcomer which provides a different service from other online motorcycle taxis. One of the activities which is always do by the fans or haters toward twitter account of online motorcycle taxis is giving comments on tweet. To identify the comments from fans or haters is required classification by using Machine Learning. In this research, the first step was sorting process and labelling the data. As the result 3 labels would have created, which were positive data label, neutral data label, and negative data label with total of 1200 data. The next step was conducting the analysis of preprocessing data which included case folding, data cleansing, tokenizing, filtering and stemming. Then, the weighting was carried out using the TF-IDF method and classified by the Support Vector Machine method. The test results were carried out using the Confusion Matrix method, based on the test results, the best accuracy was obtained at a data comparison of 90:10 by 85% using the RBF Kernel and Polynomial, followed by the Sigmoid kernel of 82.5% where the classification results were dominated by positive sentences.

Key word: classification, Ojek Online, sentiment analysis, support vector machine, twitter

1. Pendahuluan

Perkembangan internet di Indonesia cukup pesat, hal ini dibuktikan dengan semakin mudahnya akses internet dengan semakin banyaknya akses wireless yang beredar, perkembangan teknologi smartpone dan teknologi jaringan 4G membuat masyarakat lebih mudah dan cepat mengakses informasi melalui internet. Internet merupakan media yang digunakan sebagai sarana komunikasi yang lambat laun telah menyingkirkan media komunikasi seperti telepon atau sms (*short message send*) [1]. Kemudahan akses internet ini mengubah kebiasaan masyarakat yang biasanya mengakses informasi melalui media cetak menjadi melalui internet.

Perkembangan teknologi informasi dan komunikasi juga memudahkan orang untuk saling terhubung dan berbagi informasi, salah satunya melalui media jejaring sosial. Menurut [2], internet tidak terlepas dari yang namanya media jejaring sosial. Twitter, Facebook, dan Youtube merupakan salah satu media jejaring sosial sebagai media perangkat komunikasi terpopuler yang ada dikalangan pengguna internet saat ini. *Trend* penggunaan yang tinggi baik frekuensi dan intensitasnya memang sejalan dengan fenomena saat ini. *Trend* pengguna Twitter di Indonesia yang menunjukkan bahwa pengguna Indonesia merupakan ranking empat penyumbang traffic terbesar ke situs media sosial terpopuler selama tahun 2010 [3]. Salah satu jejaring sosial yang paling banyak digunakan di Indonesia adalah Twitter. Hal tersebut dikarenakan Twitter merupakan salah satu media jejaring sosial dengan pengguna terbanyak diantara beberapa situs jejaring sosial yang ada. Twitter terus mengalami peningkatan pengguna sejak kemunculannya pada tahun 2006 [4]. Data statistik menunjukkan bahwa Twitter memiliki lebih dari 200 juta pengguna aktif secara global [5].

Twitter memungkinkan penggunanya untuk menulis berbagai topik dan mendiskusikan masalah yang terjadi. Tidak seperti beberapa media sosial lain yang membutuhkan persetujuan kedua belah pihak untuk

50%. Kernel *sigmoid* memiliki tingkat *recall* paling baik sekitar 85% dilanjutkan dengan Kernel *Linear* sekitar 76%. Untuk Kernel *Gaussian* dan *Polinom* nilai akurasi sekitar 50%. Kernel *Sigmoid* memiliki tingkat akurasi paling baik sekitar 81% dilanjutkan dengan Kernel *Linear* sekitar 77%. Untuk Kernel *Gaussian* dan *Polinom* nilai akurasi sekitar 50% [10].

Berdasarkan permasalahan tersebut, maka penelitian ini akan menerapkan algoritma SVM dalam mengklasifikasikan sentiment positif, netral dan negatif masyarakat terhadap Maxim berdasarkan komentar di media sosial Twitter serta mengetahui akurasi klasifikasi yang dihasilkan metode SVM.

2. Tinjauan Pustaka

2.1 Text Mining

Text mining adalah proses semi otomatis penggalian pola (pengetahuan dan informasi yang berguna) dari sejumlah sumber data berukuran besar yang tidak terstruktur berupa teks [11][12]. *Text mining* menggunakan *natural language processing* untuk memasukkan struktur ke dalam kumpulan teks. *Natural language processing* berupaya memecahkan masalah untuk memahami bahasa alami manusia dan mengubah bahasa tersebut menjadi representasi formal yang dapat diproses oleh komputer.

Proses awal dari tahapan *text mining* yaitu *text preprocessing*. Tahap ini merupakan proses untuk mempersiapkan data teks dokumen atau dataset. Proses ini berfungsi untuk mengubah data teks yang tidak terstruktur menjadi data yang terstruktur. Adapun proses yang akan dilakukan pada tahap ini adalah:

- a. *Cleaning*
Cleaning merupakan proses pembersihan atribut-atribut baik itu kata maupun karakter pada suatu dokumen yang tidak berhubungan dengan informasi atau yang tidak diperlukan untuk mengurangi noise pada proses klasifikasi.
- b. *Case Folding*
Case folding adalah proses untuk merubah semua huruf dari “a” sampai dengan “z” dalam dokumen menjadi huruf kecil atau *lowercase*.
- c. *Tokenizing*
Tokenizing merupakan tahapan pemotongan string input berdasarkan tiap kata yang menyusunnya.
- d. *Normalisasi*
 Merupakan proses pengembalian kata-kata yang tidak baku kedalam bahasa baku dalam kamus bahasa Indonesia.
- e. *Remove Stopword*
Remove stopwords adalah kata umum atau berupa kata ganti dan kata sambung yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna, sehingga tidak akan mempengaruhi. Misalnya kata penghubung seperti “dan”, “atau”, “tapi”, “di”, “ke”, “dari”, “yang”, “ada”, “dengan”, “akan”, “itu”, “tidak”, “mau”, dan lain-lain. Tujuan utama dari *stopword remove* ini untuk mengurangi jumlah kata dalam sebuah dokumen yang nantinya akan berpengaruh dalam kecepatan dan performa dalam kegiatan *Natural Language Processing (NLP)*.
- f. *Stemming*
Stemming adalah inti pemrosesan bahasa alami teknik pengambilan informasi yang efisien dan efektif, dan salah satu yang diterima secara luas oleh pengguna. *Stemming* mengubah varian kata menjadi akar kata yang sama dengan menerapkannya dalam banyak kasus aturan morfologi [13]. Algoritma Nazief & Adriani dikembangkan pertama kali oleh Bobby Nazief dan Mirna Adriani. Algoritma ini berdasarkan pada aturan morfologi bahasa Indonesia yang luas, yang dikumpulkan menjadi satu grup dan di-enkapsulasi pada imbuhan/affixes yang diperbolehkan (*allowed affixes*) dan imbuhan/affixes yang tidak diperbolehkan (*disallowed affixes*). Algoritma ini menggunakan kamus kata dasar dan mendukung *recoding*, yakni penyusunan kembali kata-kata yang mengalami proses *stemming* berlebih.

2.2 Klasifikasi

Klasifikasi merupakan salah satu tahap penting dalam data mining. Klasifikasi adalah pengelompokan data atau objek baru ke dalam kelas atau label berdasarkan atribut-atribut tertentu [14]. Dalam klasifikasi terdapat dua proses yang dilakukan yaitu dengan membangun model untuk disimpan sebagai memori dan menggunakan model tersebut untuk melakukan pengenalan atau klasifikasi atau prediksi pada suatu data lain supaya diketahui di kelas mana objek data tersebut dimasukkan berdasarkan model yang telah disimpan dalam memori.

Klasifikasi terdiri dari tiga tahap, yaitu pembangunan model, penerapan model, dan evaluasi. Pembangunan model adalah membangun model menggunakan data latih yang telah memiliki atribut dan kelas. Kemudian, data-data tersebut diterapkan untuk menentukan kelas dari data atau objek yang baru. Setelah itu, data dievaluasi untuk melihat tingkat akurasi dari pembangunan dan penerapan model terhadap data baru [15]. Klasifikasi adalah Metode supervised learning yang mencoba menemukan relasi antara atribut input dan atribut

target [16]. Proses dalam memprediksi label kelas dan mengklasifikasi data dimulai dari *Data Cleaning*, *Relevance Analysis*, *Data transformation* dan *reduction*.

3.3 Pembobotan Kata

Pembobotan atau *term weighting* adalah proses untuk mendapatkan nilai suatu kata yang telah diproses sebelumnya [17]. Ada beberapa metode pembobotan data yang dapat digunakan diantaranya adalah *document frequency* (DF), *term frequency* (TF), dan *inverse document frequency* (IDF). Pada penelitian ini pembobotan kata yang digunakan yaitu gabungan dari *term frequency* dan *inverse document frequency*. Menurut penelitian yang dilakukan oleh [18] fungsi metode TF-IDF adalah untuk mencari representasi nilai dari tiap-tiap dokumen dari suatu kumpulan data *training (training set)* dimana nantinya dibentuk suatu vektor Antara dokumen dengan kata (*documents with terms*) yang kemudian untuk kesamaan antar dokumen dengan *cluster* akan ditentukan oleh sebuah *prototype* vektor yang disebut juga dengan *cluster centroid*.

3.4 Support Vector Machine (SVM)

SVM adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan *learning bias* yang berasal dari teori pembelajaran statistik. SVM dikembangkan oleh Broser, Guyon, Vapnik, dan pertama kali dipresentasikan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. Support Vector Machine (SVM) termasuk machine learning (*supervised learning*) yang dapat memprediksi kelas berdasarkan dari hasil proses pelatihan. Dengan melakukan pelatihan menggunakan data masukan dalam bentuk numerik dan hasil dari ekstraksi fitur didapatkan sebuah pola yang nantinya akan digunakan dalam proses pelabelan. Nilai atau pola yang dihasilkan dari Metode Support Vector Machine sebenarnya adalah sebuah garis pemisah yang disebut dengan *hyperplane*, yang mana garis tersebut berperan dalam memisahkan tweet dengan sentimen positif dengan tweet yang memiliki sentimen negative dan netral. Dalam mengambil keputusan dengan metode SVM, digunakan fungsi kernel. Pada penelitian ini akan menggunakan kernel *polynomial*, *RBF* dan *Sigmoid* yang ditunjukkan pada pengujian [19]. Tujuan utama dari metode ini adalah untuk membangun OSH (*Optimal Separating Hyperplane*), yang membuat fungsi pemisahan optimum yang dapat digunakan untuk klasifikasi.

3.5 Confusion Matrix

Confusion Matrix adalah metode yang bias digunakan untuk perhitungan *accuracy*, *recall*, *precision* dan *error rate*. Dimana, *precision* mengevaluasi kemampuan sistem untuk menemukan peringkat yang paling relevan, dan didefinisikan sebagai presentase dokumen yang di *retrieve* dan benar-benar relevan terhadap query. *Recall* mengevaluasi kemampuan sistem untuk menemukan semua item yang relevan dari koleksi dokumen dan didefinisikan sebagai presentase dokumen yang relevan terhadap query. *Accuracy* merupakan perbandingan kasus yang diidentifikasi benar dengan jumlah seluruh kasus dan *error rate* merupakan kasus yang diidentifikasi salah dengan jumlah seluruh kasus [20].

Klasifikasi yang benar	Diklasifikasikan sebagai		
	Positif	Negatif	Netral
Positif	<i>True Positive</i>	<i>False Positive Negatif</i>	<i>False Positif Netral</i>
Negatif	<i>False Negative Positif</i>	<i>True Negative</i>	<i>False Negatif Netral</i>
Netral	<i>False Netral Positif</i>	<i>False Negatif Netral</i>	<i>True Netral</i>

Tabel 1. Confusion Matrix

3. Metode Penelitian

Metode penelitian terdiri atas beberapa tahapan diantaranya :

3.1 Pengumpulan Data

Tahapan ini merupakan tahapan pengumpulan data yang diperlukan untuk menganalisa dan memperoleh data-data serta informasi yang berhubungan dengan penelitian yang akan dilakukan. Data yang dikumpulkan berasal dari tweet pada media sosial Twitter terhadap akun @Maxim_Indonesia serta keyword yang berkaitan dengan akun tersebut. Pengambilan data dilakukan dari tanggal 1 Juni 2019 sampai dengan 27 Maret 2022, data dikumpulkan dari server Twitter menggunakan Twitter API (*Application Programming Interface*). Data yang diperoleh sebanyak 1200 data. Data yang telah dikumpulkan disimpan dan di *export* kedalam format *Comma Separated Values* (CSV) untuk dilabeli oleh pakar.

3.2 Pelabelan Data

Setelah mengumpulkan data komentar serta menyeleksinya, tahap selanjutnya adalah memberi label pada seluruh data komentar sesuai dengan kategori yang telah ditentukan secara manual dengan cara memberikan label untuk setiap komentar. Pelabelan dilakukan oleh Rosmiati, S.Pd sebagai Guru Bahasa Indonesia SMK Teknologi Riau. Ada tiga kategori yang ditentukan untuk klasifikasi berdasarkan informasi yang

Hak Cipta Dilindungi Undang-Undang
 © Hak Cipta Dilindungi Undang-Undang
 State Islamic University of Sultan Firdausy of Pekanbaru
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dikandungnya, antara lain :

- a. Kategori Opini Positif
- b. Kategori Opini Netral
- c. Kategori Opini Negatif

No	Text	Kelas
1	Besok pergi kerja naik maxim lah. Lebih murah	Positif
2	Risaunya kalau dpt driver maxim bapak tua. Tak percaya diri cara ngendarai motornya nya hm.	Negatif
3	@adelliaaaaa maxim food go food	Netral

Tabel 2. Pelabelan Manual

3.3 Text Preprocessing

Setelah proses pelabelan sudah dilakukan, tahapan selanjutnya yaitu *Preprocessing Text*. *Preprocessing* ialah proses mempersiapkan data agar data siap untuk diolah, tahapan ini adalah tahapan yang paling penting dalam melakukan *mining* teks. Ada beberapa tahapan *preprocessing* dalam melakukan penelitian ini, yaitu :

a. Case Folding

Case folding merupakan perubahan huruf yang ada di kata maupun kalimat menjadi huruf kecil secara keseluruhan.

b. Cleaning

Cleaning merupakan proses penghapusan kata dan karakter yang tidak dibutuhkan. Seperti mention username (@), link, hashtag (#), emoticon maupun simbol (@#\$\$%^&*()+":{ }<>?!~/[] dan angka.

c. Tokenizing

Tokenizing merupakan proses pemisahan kalimat menjadi per kata. Pemisahan kata berdasarkan karakter angka, huruf dan spasi.

d. Filtering

Filtering merupakan proses penghapusan kata yang tidak penting.

e. Stemming

Stemming merupakan perubahan kata dasar yang ber-imbuh. algoritma *stemming* yang peneliti gunakan yaitu Nazief dan Adriani.

3.4 Pembobotan TF-IDF

Tahapan pembobotan dan pemilihan fitur ini dilakukan dalam proses ekstraksi fitur untuk menghasilkan fitur list dan data train. Dalam proses ini, data token akan dihitung jumlah kemunculannya pada setiap dokumen dan nilainya akan menjadi bobot token. Pembobotan ini dilakukan untuk merangking kata selama digunakan dalam proses SVM.

3.5 Klasifikasi Support Vector Machine

Pembelajaran dan uji data akan di lakukan pada tahap ini. Pengujian pada model dilakukan dengan menggunakan data testing untuk mengetahui nilai akurasi dan klasifikasinya. Hasil dari akurasi metode SVM tidak bisa langsung digunakan, karena dalam SVM ada banyak faktor yang bisa meningkatkan hasil akurasi dengan cara melakukan cek performa dari berbagai faktor. Pada penelitian ini peneliti menggunakan kernel *linear*. Proses Klasifikasi dibagi dalam 2 proses, yaitu proses latih data (*train*) dan proses klasifikasi data (*classifier*). Proses *training* data dilakukan untuk menghasilkan model fitur yang nantinya digunakan sebagai acuan perhitungan untuk mengklasifikasikan kalimat ke dalam tiga kelas yaitu positif, negatif dan netral.

3.6 Pengujian Confusion Matrix

Pengujian *Confusion Matrix* dilakukan untuk mendapatkan hasil akurasi dari proses SVM yang telah dilakukan.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 1 Tahapan klasifikasi Sentimen Terhadap Maxim Menggunakan Algoritma SVM Pada Media Sosial Twitter

4. Hasil dan Pembahasan

Berikut merupakan hasil dari setiap proses dalam penelitian ini, yaitu klasifikasi Sentimen Terhadap Maxim Menggunakan Algoritma SVM Pada Media Sosial Twitter pada 3 kalimat sentiment.

4.1 Text Preprocessing

a. Case Folding

Berikut merupakan hasil tahapan *Case folding* pada tabel 3 berikut :

Data	Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
1	Besok pergi kerja naik maxim lah. Lebih murah	besok pergi kerja naik maxim lah. lebih murah
2	Risaunya kalau dpt driver maxim bapak tua. Tak percaya diri cara ngendarai motornya nya hm.	risaunya kalau dpt driver maxim bapak tua. tak percaya diri cara ngendarai motornya nya hm.
3	@adelliaaaaa maxim food go food	@adelliaaaaa maxim food go food

Tabel 3 Hasil Tahapan *Case Folding*

Hak Cipta Dilindungi Undang-Undang

c. Tokenizing

Data	Sebelum Cleaning	Setelah Cleaning
1	besok pergi kerja naik maxim lah. lebih murah	besok pergi kerja naik maxim lah lebih murah
2	risaunya kalau dpt driver maxim bapak tua. tak percaya diri cara ngendarai motornya nya hm.	risaunya kalau dpt driver maxim bapak tua tak percaya diri cara ngendarai motornya nya hm
3	@adelliaaaaa maxim food go food	maxim food go food

Tabel 4 Hasil Tahapan Cleaning

Data	Sebelum Tokenizing	Hasil Tokenizing
1	besok pergi kerja naik maxim lah lebih murah	besok pergi kerja naik maxim lah lebih murah
2	risaunya kalau dpt driver maxim bapak tua tak percaya diri cara ngendarai motornya nya hm	risaunya kalau dpt driver maxim bapak tua tak percaya diri cara ngendarai motornya nya hm
3	maxim food go food	maxim food go food

Tabel 5 Hasil Tahapan Tokenizing

d. Filtering

Data	Sebelum Filtering	Hasil Filtering
1	besok pergi kerja naik maxim lah lebih murah	besok pergi kerja maxim murah
2	risaunya kalau dpt driver maxim bapak tua tak percaya diri cara ngendarai motornya nya hm	risaunya driver maxim tua percaya ngendarai motornya
3	maxim food go food	maxim food go food

Tabel 6 Hasil Tahapan Filtering

e. Stemming

Data	Sebelum Stemming	Hasil Stemming
1	besok pergi kerja maxim murah	besok pergi kerja maxim murah
2	risaunya driver maxim tua percaya ngendarai motornya	risau driver maxim tua percaya ngendarai motor
3	maxim food go food	maxim food go food

Tabel 7 Hasil Tahapan Stemming

4.2 Pelebotan Kata TF-IDF

Berikut hasil perhitungan TF-IDF pada tabel 8 berikut :

Term (t)	TF			DF	IDF	TF.IDF		
	D1	D2	D3			D1	D2	D3
besok	1	0	0	1	$\log(3/1) = 0,477$	0,477	0	0
pergi	1	0	0	1	$\log(3/1) = 0,477$	0,477	0	0
kerja	1	0	0	1	$\log(3/1) = 0,477$	0,477	0	0
maxim	1	1	1	3	$\log(3/3) = 0$	0,477	0,477	0,477
murah	1	0	0	1	$\log(3/1) = 0,477$	0,477	0	0
risau	0	1	0	1	$\log(3/1) = 0,477$	0	0,477	0
driver	0	1	0	1	$\log(3/1) = 0,477$	0	0,477	0
tua	0	1	0	1	$\log(3/1) = 0,477$	0	0,477	0
percaya	0	1	0	1	$\log(3/1) = 0,477$	0	0,477	0

ngendarai	0	1	0	1	$\log(3/1) = 0,477$	0	0,477	0
motor	0	1	0	1	$\log(3/1) = 0,477$	0	0,477	0
food	0	0	2	2	$\log(3/2) = 0,176$	0	0	0,352
go	0	0	1	1	$\log(3/1) = 0,477$	0	0	0,477
					Total	2,385	3,339	1,306

Tabel 8 Hasil TF-IDF

4.3 Klasifikasi SVM

Hasil yang didapat dari Klasifikasi ialah data baru yang telah diproses pada tahapan *preprocessing*. Data tersebut tidak perlu dilakukan lagi proses pelabelan manual, karena data tersebut akan diproses oleh SVM itu sendiri. Klasifikasi SVM dapat menggunakan beberapa Kernel. Kernel yang digunakan pada penelitian ini menggunakan *Gaussian Kernel Radial Basis Function (RBF)*, Kernel *Sigmoid*, Kernel *Polinomial* dan Kernel *Linear*.

4.4 Pengujian menggunakan tabel Confusion Matrix

Tabel 9 adalah hasil akurasi dari klasifikasi dengan pembagian data kedalam 3 proses perbandingan data latih dan data uji dengan jumlah data 1200, yaitu 90% : 10%, 80% : 20% serta 70% : 30% dengan menggunakan kernel *RBF*, *Sigmoid* dan *Polinomial* yang tersaji dalam tabel 9 berikut.

No	Kernel	Perbandingan		
		90 : 10	80 : 20	70 : 30
1	RBF	85%	81,25%	81,67%
2	Sigmoid	82,5%	79,58%	79,72%
3	Polinomial	85%	81,67%	81,94%

Tabel 9 Perbandingan akurasi pengujian dataset

Pada tabel 9 Kernel *RBF* dan *Polinomial* memiliki akurasi terbaik pada perbandingan data latih dan data uji 90% : 10%. Berikut Perhitungan pengujian tabel *Confusion Matrix* menggunakan kernel *RBF* :

- Hasil percobaan pertama menggunakan Algoritma *Support Vektor Machine* pada perbandingan data 90:10 dengan jumlah data 120 komentar, akurasi yang didapat ialah 85%. Hasil prediksi menggunakan *confusion matrix* dapat dilihat pada Tabel 10.

Aktual	Prediksi		
	False Negatif	False Netral	False Positif
True Negatif	0	0	10
True Netral	0	2	7
True Positif	1	0	100

Tabel 10 Hasil Akurasi Confusion Matrix SVM Data Uji 90 : 10

$$\begin{aligned} \text{Perhitungan Akurasi} &= \frac{0+2+100}{0+0+10+0+2+7+1+0+100} \times 100 \\ &= \frac{102}{120} \times 100 \\ &= 85 \end{aligned}$$

- Hasil percobaan kedua menggunakan Algoritma *Support Vektor Machine* pada perbandingan data 80:20 dengan jumlah data 240 komentar, akurasi yang didapat ialah 81,25%. Hasil prediksi menggunakan *confusion matrix* dapat dilihat pada Tabel 11.

Aktual	Prediksi		
	False Negatif	False Netral	False Positif
True Negatif	1	0	25
True Netral	0	3	18
True Positif	2	0	191

Tabel 11 Hasil Akurasi Confusion Matrix SVM Data Uji 80 : 20

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak Cipta milik UIN Suska Riau
State Islamic University of Sultan Syarif Kasim Riau

$$\begin{aligned} \text{Perhitungan Akurasi} &= \frac{1+3+191}{1+0+25+0+3+18+2+0+191} \times 100 \\ &= \frac{195}{240} \times 100 \\ &= 81,25 \end{aligned}$$

c. Hasil percobaan ketiga menggunakan Algoritma *Support Vektor Machine* pada perbandingan data Uji 70:30 dengan jumlah data 360 komentar, akurasi yang didapat ialah 81,67%. Hasil prediksi menggunakan *confusion matrix* dapat dilihat pada Tabel 12.

Aktual	Prediksi		
	False Negatif	False Netral	False Positif
True Negatif	3	0	36
True Netral	0	6	28
True Positif	2	0	285

Tabel 12 Hasil Akurasi Confusion Matrix SVM Data Uji 70 : 30

$$\begin{aligned} \text{Perhitungan Akurasi} &= \frac{3+6+285}{3+0+36+0+6+28+2+0+285} \times 100 \\ &= \frac{294}{360} \times 100 \\ &= 81,667 \end{aligned}$$

Penulis juga mencoba pengujian klasifikasi tanpa menggunakan tahapan *filtering*, hasil akurasi dapat dilihat pada tabel 13 berikut.

No	Kernel	Perbandingan		
		90 : 10	80 : 20	70 : 30
1	RBF	85%	81,25%	81,67%
2	Sigmoid	82,5%	79,58%	79,72%
3	Polinomial	85%	81,67%	81,94%

Tabel 13 Perbandingan akurasi pengujian dataset tanpa tahapan *filtering*

5. Kesimpulan

1. Metode *Support Vector Machine* dapat digunakan untuk proses klasifikasi data teks terhadap komentar tweet pada media sosial twitter.
2. Metode *Support Vector Machine* untuk klasifikasi komentar tweet yang dibagi menjadi 3 pembagian data yakni positif, negatif dan netral mendapat akurasi klasifikasi tertinggi pada kelompok pembagian 90:10 yaitu 85% dengan menggunakan Kernel *Gaussian Kernel Radial Basis Function (RBF)* dan *Polynomial Kernel*.
3. Dari Analisa yang dilakukan terhadap kasus maxim, terdapat banyak kalimat berkategori positif.
4. Jika klasifikasi dilakukan tanpa melakukan tahapan *filtering*, hasilnya tidak berbeda.

Daftar Pustaka

- [1] R. E. Sari, A. Saleh, and M. R. Afandi, "Pemanfaatan Website Dalam Mengembangkan Kreativitas Siswa Di Dunia Usaha," *J. TUNAS*, pp. 43–47, 2021, [Online]. Available: <http://tunasbangsa.ac.id/abdimas/index.php/tunasabdimas/article/view/29>.
- [2] A. Nugroho, "Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstrasi Fitur N-Gram," *J-SAKTI (Jurnal Sains Komput. dan Inform., vol. 2, no. 2, pp. 200–209, 2018.*
- [3] E. A. Sosiawan, "Penggunaan situs jejaring sosial sebagai media interaksi dan komunikasi di kalangan mahasiswa," *J. Ilmu Komun., vol. 9, no. 1, pp. 60–75, 2020.*
- [4] Y. I. Claudy, R. S. Perdana, and M. A. Fauzi, "Klasifikasi Dokumen Twitter untuk Mengetahui Karakter Calon Karyawan Menggunakan Algoritme K-Nearest Neighbor (KNN)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN, vol. 2548, p. 964X, 2018.*
- [5] H. Hartanto, "TEXT MINING DAN SENTIMEN ANALISIS TWITTER PADA GERAKAN LGBT,"

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

[6] Intuisi *J. Psikol. Ilm.*, vol. 9, no. 1, pp. 18–25, 2017.

[7] Oryza Habibie Rahman, Gunawan Abdillah, and Agus Komarudin, “Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 17–23, 2021, doi: 10.29207/resti.v5i1.2700.

[8] S. Hanganararas, “Analisis Kelayakan Layanan GOJEK Sebagai Moda Transportasi di Yogyakarta,” *J. Tek. Sipil*, vol. 7, 2017.

[9] B. M. Pintoko and K. M. Lhaksmana, “Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier,” *eProceedings Eng.*, vol. 5, no. 3, 2018.

[10] I. P. Windasari, F. N. Uzzi, and K. I. Satoto, “Sentiment analysis on Twitter posts: An analysis of positive or negative opinion on GoJek,” *Proc. - 2017 4th Int. Conf. Inf. Technol. Comput. Electr. Eng. ICITACEE 2017*, vol. 2018-Janua, pp. 266–269, 2017, doi: 10.1109/ICITACEE.2017.8257715.

[11] I. Taufik, “Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma Support Vector Machine (SVM).” Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta, 2018.

[12] S. Hilda Kusumahadi, H. Junaedi, and J. Santoso, “Klasifikasi Helpdesk Menggunakan Metode Support Vector Machine,” *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 54–60, 2019, doi: 10.30591/jpit.v4i1.1125.

[13] N. R. S. Purba and F. Riandari, “Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Analisis Keranjang Belanja Transaksi Penjualan Pada PT Madu Kembang Joyo,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 4, no. 1, 2021.

[14] D. Jatikusumo and H. D. Wijaya, “Social Media dengan Kombinasi Algoritma Stemming,” *Indones. J. Comput. Inf. Technol.*, vol. 6, no. April, pp. 77–81, 2021.

[15] D. A. Nasution, H. H. Khotimah, and N. Chamidah, “Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN,” *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.

[16] F. F. Maulana and N. Rochmawati, “Klasifikasi Citra Buah Menggunakan Convolutional Neural Network,” *J. Informatics Comput. Sci.*, vol. 1, no. 02, pp. 104–108, 2020, doi: 10.26740/jinacs.v1n02.p104-108.

[17] S. Hendrian, “Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan,” *Fakt. Exacta*, vol. 11, no. 3, 2018.

[18] A. A. Puspitasari, “Klasifikasi Dokumen Tumbuhan Obat Menggunakan Metode Improved k-Nearest Neighbor.” Universitas Brawijaya, 2017.

[19] N. K. Widyasanti, I. K. G. Darma Putra, and N. K. Dwi Rusjyanthi, “Seleksi Fitur Bobot Kata dengan Metode TFIDF untuk Ringkasan Bahasa Indonesia,” *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 6, no. 2, p. 119, 2018, doi: 10.24843/jim.2018.v06.i02.p06.

[20] A. M. Pravina, I. Cholissodin, and P. P. Adikara, “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>.

[20] A.- Arini, L. K. Wardhani, and D.- Octaviano, “Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Pada Tweet Hashtag #2019gantipresiden,” *Kilat*, vol. 9, no. 1, pp. 103–114, 2020, doi: 10.33322/kilat.v9i1.878.

UIN SUSKA RIAU

