

Twitter User Sentiment Analysis Of TIX ID Applications Using Support Vector Machine Algorithm

Asyfh Nabillah^{1*}, Syariful Alam², Mochzen Gito Resmi³

Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana

Jl. Cikopak No.53, Sa dang Purwakarta, Ja wa Barat – Indonesia

e-mail : asyahnabillah33@wastukencana.ac.id¹, syarifulalam@wastukencana.ac.id²,
mochzen@wastukencana.ac.id³

Abstract: *The cinema is a place to watch movies using the big screen. Various comments on the TIX ID application service can be used to reference the company's evaluation material in assessing the level of service quality that has been provided, so that later the TIX ID application can be used optimally by application users and the company, as for the purpose of this study to find out responses and find out the sentiment analysis stage on twitter social media using the vector machine support algorithm for the TIX ID application. This algorithm is commonly used for text mining by going through the data collection stage, cleaning and labelling data stage, training and testing data sharing stage with 3 comparison scenarios, namely 70:30, 80:20, and 90:10 using 3 kernels, namely dot, radial, and polynomial, then through the text preprocessing stage, the TF-IDF word weighting stage, the data modeling stage, and the evaluation stage. The preprocessing stage consists of transform case, tokenize, and stopwords filters. The result of this study is that the support vector machine algorithm has an accuracy value of 74.17%. The research concludes that the support vector machine algorithm with a ratio of 80:20 training and testing data ratio scenario produces the highest accuracy.*

Keywords : *Support Vector Machine, cinema, TIX ID, text mining, Algorithm*

Abstrak: *Bioskop merupakan tayangan layar lebar untuk menonton film. Berbagai komentar terhadap layanan aplikasi TIX ID dapat digunakan untuk acuan bahan evaluasi perusahaan dalam menilai tingkat kualitas layanan yang telah diberikan, agar nantinya aplikasi TIX ID dapat digunakan secara maksimal oleh pengguna aplikasi dan perusahaan, Tujuan penelitian dilakukan untuk mengetahui tanggapan dan mengetahui tahap analisis sentimen di media sosial twitter dengan menggunakan algoritma *support vector machine* terhadap aplikasi TIX ID. Algoritma ini biasa digunakan untuk *text mining* dengan melalui tahap pengumpulan data, tahap *cleaning* dan data *labelling*, tahap pembagian data *training* dan *testing* dengan 3 skenario perbandingan yaitu 70:30, 80:20, dan 90:10 dengan menggunakan 3 *kernel* yaitu *dot*, *radial*, dan *polynomial*, kemudian melalui tahap *text preprocessing*, tahap pembobotan kata TF-IDF, tahap pemodelan data, dan tahap evaluasi. Pada tahap *preprocessing* terdiri dari proses *transform case*, *tokenize*, dan *filter stopwords*. Hasil penelitian ini yaitu algoritma *support vector machine* memiliki nilai akurasi sebesar 74,17%. Penelitian memiliki kesimpulan bahwa algoritma *support vector machine* dengan skenario rasio perbandingan data *training* dan *testing* 80:20 menghasilkan akurasi tertinggi.*

Kata kunci: *Support Vector Machine, bioskop, TIX ID, text mining, Algoritma*

PENDAHULUAN

Menurut survei Asosiasi Penyelenggara Jasa Internet di Indonesia, Indonesia adalah penduduk dengan pengguna internet terbanyak untuk saat ini di dunia. Di awal tahun 2021 jumlah penduduk yang menggunakan internet sebanyak 202,6 juta dengan peningkatan 27 juta pada beberapa tahun kebelakang (Ismail et al., 2022). Dengan bertambahnya pengguna internet tersebut

membuat para pelaku bisnis atau perusahaan-perusahaan memanfaatkan teknologi informasi untuk menciptakan bisnis baru dan mempermudah proses transaksi (Cahyani & Raya Sulistyowati, 2021). Bioskop merupakan tempat untuk menonton film dengan tampilan layar berukuran lebar. Aplikasi TIX ID berada di posisi jumlah pengunduh terbanyak dengan peringkat tertinggi, aplikasi TIX ID merupakan aplikasi *m-ticketing* yang bekerja sama dengan perusahaan layanan pertunjukan bioskop di Indonesia seperti XXI, CGV, dan Cinepolis. Aplikasi TIX ID memudahkan pelayanan dengan menyediakan tiket menonton secara satu pintu, sehingga sangat memudahkan dan juga bermanfaat dalam memperoleh informasi mengenai jam tayang, menyediakan fitur untuk beli atau sewa film favorit untuk pengguna secara online, menyediakan penawaran menarik melalui voucher, menyediakan fitur detail lebih lanjut tentang film, dan menyediakan fitur LOYALTIX Points untuk mendapatkan berbagai promo menarik yang diberikan secara khusus. Media sosial Twitter dapat digunakan sebagai alat untuk mengumpulkan data analisis sentimen. Analisis Sentimen digunakan untuk melihat komentar yang masuk kedalam isu positif atau negatif untuk acuan dalam meningkatkan kualitas layanan (Sari, 2020). Aplikasi merupakan suatu perangkat penting yang ada di *smartphone* dan dikembangkan dengan menggunakan bahasa pemrograman. Aplikasi dirancang agar perintah pengguna dijalankan untuk mendapatkan masukan dan mencapai hasil yang diinginkan (Sarmidi, 2018). Aplikasi yang dapat digunakan untuk melakukan transaksi mengenai tiket bioskop disebut aplikasi *mobile ticketing*. Aplikasi *mobile ticketing* merupakan terobosan dari aplikasi *e-commerce* yang menerapkan layanan berbasis aplikasi *mobile* dalam bidang per tiket-an. Aplikasi *mobile ticketing* dapat melakukan transaksi tanpa harus datang ke tempat pembelian tiket salah satu manfaat adanya aplikasi *mobile ticketing* ini yaitu mempermudah pengguna (Murdiono et al., 2018). Dengan adanya aplikasi TIX ID menghasilkan berbagai ulasan pengguna mengenai aplikasi baik ulasan positif ataupun negatif. Semakin banyak ulasan maka semakin sulit pula bagi pengguna lain untuk menyimpulkan hasil ulasan pengguna aplikasi. Dalam penerapan analisis sentimen digunakannya metode algoritma, maka digunakannya algoritma *Support Machine Vector* dalam penelitian untuk mengetahui hasil ulasan dari pengguna. Algoritma *Support Machine Vector* ini memungkinkan perhitungan untuk masalah linear. Analisis sentimen ini digunakan untuk mengetahui sebuah tweet dengan metode *Support Vector Machine* (SVM) dan Bagaimana hasil dari penerapan metode *Support Vector Machine* (SVM) terhadap tweet dengan kata kunci TIX ID di Media Sosial Twitter.

Penelitian sebelumnya yang telah dilakukan memiliki perbedaan. Pada penelitian (Pravina et al., 2019) Maksimalkan analisis sentimen maskapai di Twitter menggunakan algoritma mesin vektor dukungan dengan fitur berbasis kamus. Parameter *learning rate* mengambil nilai 0,03, C mengambil nilai 10, dan nilai parameter optimal. Penelitian selanjutnya (Rahutomo et al., 2018) menggunakan algoritma *support vector machine* yang terbukti bermanfaat untuk

mengklasifikasikan tweet tentang review film, menggunakan jumlah data latih yang berbeda dalam pengujian ini mengakibatkan penggunaan data latih lebih banyak, dan nilai akurasi dan presisi cenderung meningkat.. Pada penelitian (Petiwi et al., 2022) metode *Support Vector Machine* mendapatkan nilai akurasi yang besar yaitu 83% dari 5000 tweet dengan bahasa Python dan 98,5% dari 1000 tweet yang dilakukan dengan uji coba antarmuka. Sedangkan hasil dari *Naïve Bayes* menghasilkan nilai 74,6% memakai Python dan 91,5% melalui uji coba antarmuka. Dapat diketahui bahwa metode *Support Vector Machine* memiliki hasil lebih akurat.. Dalam penelitian (Haditira et al., 2022) telah dibangun sistem klasifikasi untuk mengenal suatu data *record* menjadi bentuk data fitur berdasarkan ulasan atau tweet pada game yang ada di *Steam*, penggunaan metode *Multinomial Naïve Bayes* dengan seleksi Fitur Gini Index Text dapat menghasilkan akurasi, *presisi*, *recall*, dan skor-F1. Hasil klasifikasi pada 5 kali percobaan *training* dan *testing*, untuk porsi Rasio data 90:10, 80:20, 70:30, pada 4000 data dengan mengambil 100 fitur terbaik (skor tertinggi) dan dilakukan penambahan fitur setiap iterasinya sebagai akurasi terbaik, nilai Rata-rata akurasi MNNB Tanpa GIT didapat 81.38%, 83.30%, 80.08%, 81.04%.

Dari uraian latar belakang diatas, peneliti akan melakukan analisis sentimen terhadap aplikasi TIX ID menggunakan metode *Support Vector Machine* (SVM) menggunakan *kernel dot*, *radial* dan *polynomial*. Dari proses pengujian data pada Rapidminer dimulai dengan pembentukan model data pada bagian pertama pembagian data *training* dan data *testing*, melalui tahap pemodelan data dengan algoritma *support vector machine* menggunakan *kernel dot*, *radial*, dan *polynomial*. Dapat diketahui akurasi manakah yang memiliki tingkat akurasi yang paling tinggi. Penelitian ini bertujuan untuk mengetahui tanggapan dari pengguna twitter terhadap aplikasi TIX ID dengan melakukan analisis sentimen.

KAJIAN PUSTAKA

Analisis Sentimen

Menurut (Gifari et al., 2022) Analisis sentimen adalah teknik untuk secara otomatis memahami, mengekstraksi, dan memanipulasi informasi tekstual untuk menemukan informasi emosional ketika mengekspresikan pikiran. Analisis sentimen digunakan untuk melihat kesamaan pendapat atau pendapat orang tentang suatu masalah atau objek yang mengarah pada pendapat positif atau negatif.

Analisis sentimen adalah pendeteksian sikap terhadap objek atau orang. Dari miliaran tweet di Twitter, Analisis sentimen dilakukan agar mengetahui seberapa banyak komentar positif dan negatif terhadap situasi individu atau tertentu (Alhaq et al., 2021).

Aplikasi

Aplikasi merupakan kebutuhan pengguna yang dapat dikembangkan untuk memenuhi kebutuhan. Aplikasi adalah perangkat lunak yang dibuat oleh komputer agar memudahkan pekerjaan yang dilakukan oleh pengguna (Agustina, 2021). Aplikasi adalah perangkat lunak yang dipakai untuk kebutuhan manusia yang sering digunakan untuk pelayanan masyarakat dari setiap aktivitas yang dijalani (Rudiansyah et al., 2022).

Support Vector Machine

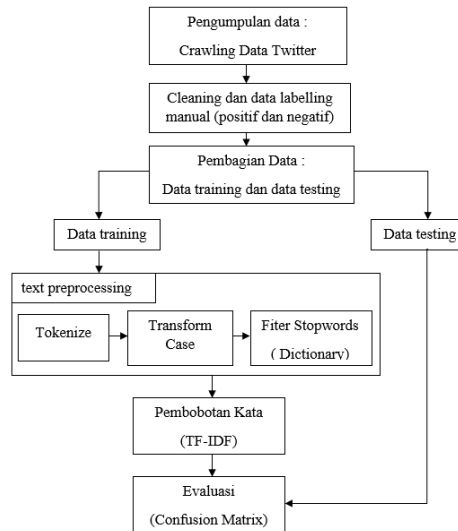
Menurut (Tineges et al., 2020) Pendekatan SVM memungkinkan untuk menghitung masalah linier dengan menerapkan transformasi matematis ke ruang belajar menggunakan fungsi perkalian. Metode SVM memiliki konsep dasar dalam klasifikasi data, yaitu menentukan hyperplane yang optimal untuk memberikan jarak atau jarak yang telah ditentukan sebelumnya antara dua kelas.

Mobile Ticketing

Menurut (Rahayu et al., 2019) *Mobile ticketing* adalah cara untuk mendokumentasikan proses penjualan acara pelanggan tanpa perlu dokumentasi secara langsung.

METODE PENELITIAN

Metode penelitian ini dilakukan untuk memperoleh data yang sesuai untuk menemukan pengetahuan (Darna et al., 2018). Penelitian ini menggunakan algoritma *Support Vector Machine* untuk algoritma klasifikasi. Pada tahap pertama pengumpulan data didapatkan dari hasil *crawling* melalui media sosial yaitu Twitter. Setelah data didapatkan maka tahap selanjutnya membersihkan data atau *Cleaning* dan memberikan label pada setiap ulasan secara manual untuk mengetahui ulasan dari pengguna masuk kedalam label positif atau negatif. Kemudian data dibagi mejadi dua untuk diolah diantaranya ada data *Testing* dan *Data Training*. Setelah tahapan membagi data selesai, maka akan dilakukan tahapan *Text Preprocessing* yang memiliki proses diantaranya *Tokenize*, *Transform Case* dan *Filter Stopwords*. Tahap berikutnya yang dilakukan yaitu Pembobotan TF-IDF, Pemodelan Data, dan evaluasi. Rancangan penelitian dapat dilihat pada Gambar 1 dibawah ini :



Gambar 1. Diagram Alur Penelitian

Objek Penelitian

Objek yang digunakan mengenai data tweet yang didapat dari proses pencarian dengan kata kunci TIX ID yang berasal dari media sosial twitter. Data yang didapatkan berjumlah 3.365 data. Data tersebut akan digunakan sebagai objek penelitian untuk mengetahui sentimen masyarakat dan akurasi algoritma.

Pengumpulan Data

Data yang digunakan yaitu data dari tweet pengguna yang ada di media sosial twitter yang meliputi sebuah aplikasi *m-ticketing* yaitu aplikasi TIX ID. Data diambil melalui proses *Crawling* data dengan menggunakan python dan kata kunci pencarian tweet yaitu TIX ID dari mulai tanggal 18 Mei 2022 sampai dengan 8 Juni 2022.

Crawling Data

Crawling data yaitu tahap pertama untuk proses pengumpulan data yang di dapatkan melalui situs web ataupun media sosial menggunakan API yang telah disediakan (Homepage et al., 2021). *Crawling* data dilakukan untuk mengambil kumpulan data dalam proses analisis dengan Twitter API. Pada proses ini data yang di dapatkan sebanyak 3.365 tweet.

Cleaning dan Data Labeling

Setelah tahap *crawling* data selesai, Pada tahap ini dilakukan untuk membersihkan atau menghilangkan komponen-komponen atau fitur yang terdapat pada data hasil *crawling* seperti url, @username, RT, dan lain-lain, kemudian informasi yang terkandung dalam tweet berupa sentimen positif atau negatif diambil guna memudahkan proses olah data, dalam tahap labelling data analisa sentimen dilakukan manual dengan cara memahami makna apa yang terkandung

dalam suatu kalimat, sehingga setelah mendapatkan maksud dari kalimat tersebut dapat diberi nilai bahwa kalimat tersebut merupakan sentimen positif atau negatif.

Data Training dan Data Testing

Proses ini dilakukan secara manual, Data *Training* digunakan untuk memberi pelatihan dataset supaya algoritma dapat mengetahui data yang dipilih termasuk kedalam kelas positif atau negatif. Penelitian ini dibuat untuk mencari akurasi yang terbaik, Proses Data *Training* dan Data *Testing* dibagi menjadi 3 (tiga) nilai perbandingan untuk memperkuat nilai akurasi yang diperoleh. Nilai perbandingan pertama adalah rasio 70% dan 30%, rasio kedua adalah 80% dan 20%, dan rasio ketiga adalah 90% dan 10%. Setelah data *sharing*, tahap selanjutnya adalah melatih data training dari hasil data *sharing* dengan algoritma SVM, hasil dari training dapat dipakai pada klasifikasi sentimen dan data testing agar diketahui nilai terbaik.

Preprocessing

Proses pengolahan data merupakan langkah awal dalam *Preprocessing* supaya memperoleh kata-kata yang bisa diklasifikasikan (Gunawan et al., 2020). Setelah didapatkan data hasil crawling maka selanjutnya dokumen tweet pengguna ini akan dilakukan tahapan *preprocessing*. Proses pada tahapan *preprocessing* penelitian ini yaitu *Tokenize*, *Transform case* dan *Filter Stopword*. Berikut tabel penjelasan tahapan *Preprocessing* pada tabel 1 dibawah ini :

Tabel 1. Penjelasan Tahapan *Preprocessing*

No	Tahapan	Penjelasan
1	<i>Tokenize</i>	Untuk pemecahan kalimat dan menghilangkan tanda baca
2	<i>Transform Case</i>	Untuk mengubah semua tulisan menjadi huruf kecil
3	<i>Filter Stopword</i>	Menghilangkan teks yang tidak saling berhubungan

Pembobotan TF-IDF

Setelah tahap *Preprocessing* selesai, maka dilakukan pemberian bobot pada setiap kata. Nilai bobot yang diberikan akan memberi pengaruh kemiripan saat mengukur kata dari suatu dokumen. Pembobotan TF-IDF ini melihat dari jumlah label positif dan negatif yang sudah ditentukan. Menurut (Aziz Syahputro et al., 2019) TF-IDF yaitu cara untuk pemberian bobot dalam suatu kata yang berhubungan dengan dokumen. TF-IDF memiliki dua konsep yang digabungkan untuk menghitung bobot yaitu, untuk mengetahui seberapa banyak kata yang selalu muncul didalam dokumen dan seberapa banyak dokumen yang termasuk kata tersebut. Tujuan dari pembobotan adalah memberikan nilai frekuensi supaya bisa diproses oleh algoritma *Support Vector Machine* terhadap kata yang dianggap bobot. Menghitung nilai *term frequency* setiap kata adalah tahap pertama, kemudian tahap kedua adalah nilai *document frequency* setiap kata yang akan dihitung, tahap ketiga yang akan dihitung yaitu nilai *inverse document frequency*, dan tahap terakhir akan

dihitung bobot yang dihasilkan dari perkalian *term frequency* dengan *inverse document frequency* (Ananda & Pristyanto, 2021). Pembobotan ini memiliki rumus untuk menghitung setiap token pada dokumen.

TF = jumlah munculnya term pada dokumen / jumlah term pada dokumen

IDF = $\log(\text{jumlah seluruh dokumen} / \text{jumlah dokumen term yang muncul})$

Pemodelan Data

Algoritma yang dipakai akan ditentukan pada tahapan pemodelan data, pada tahap ini juga dilakukan *10 fold cross validation*. Algoritma *support vector machine* akan digunakan saat ini dibagi menjadi data training dan testing terdapat proses didalamnya berupa *apply model* dan *performance*.

Evaluasi

Setelah melakukan tahap pemodelan data menggunakan *Support Vector Machine*, maka tahap selanjutnya yaitu Evaluasi. Menurut (Fatmawati, 2016) Pada tahap ini akan dilakukan evaluasi terhadap performa *support vector machine* dalam menentukan klasifikasi berdasarkan sampel yang telah diberikan. Untuk mengevaluasi model digunakan *Confusion Matrix* menggunakan *Precision, Recall, Accuracy Classification*. Dibawah ini merupakan tabel confusion matrix pada tabel 2 :

Tabel 2. Confusion Matrix

Classification		Actual Class	
		True Positif	False Negative
Prediksi	True Positif	TP	FP
	False Negative	FN	TN

HASIL DAN PEMBAHASAN

Pengumpulan Data

Data yang diambil diperoleh dengan cara *crawling* dan bantuan dari kunci *Application Programming Interface* yang ada dalam twitter. Cara ini memperoleh ulasan pengguna twitter dalam sekumpulan teks. Pengumpulan data dilakukan dengan menggunakan tools Google Colab dan bahasa pemrograman python dengan memasukan *access token* yang di dapat melalui pendaftaran di Twitter Developer.

Cleaning dan Data Labeling

Pada tahapan ini dilakukan pembersihan dan memberi *Labeling* pada data untuk menentukan data tersebut termasuk kedalam ulasan positif atau negatif. Hasil dari keseluruhan data sudah diberi

label positif berjumlah 146 dan label negatif berjumlah 231. Berikut merupakan hasil *cleaning* dan data *labeling* pada tabel 3 dibawah ini :

Tabel 3. Hasil *Cleaning* dan Data *Labeling*

Tweet	sentimen
TIX ID gue gedig juga lu, lama banget updetnya	Negatif
Aku lebih nyaman di TIX ID sih, karna ya bisa sekalian nyesuain mau pesen xxi, cine apa cgv di satu ap	Positif
Dari tadi mantengin TIX ID reset lama banget	Negatif
TIX ID mau beli tiket penuh terus astagaa, masa tiap jamnya ga ada yg kosong	Negatif
Pake TIX ID bisa milih mau nonton dimana	Positif

Data *Training* dan Data *Testing*

Proses Data *Training* dan Data *Testing* dibagi menjadi 3 (tiga) nilai perbandingan untuk memperkuat nilai akurasi yang diperoleh. Berikut data *training* dan data *Testing* pada tabel 4 dibawah ini :

Tabel 4. Pembagian Data *Training* dan Data *Testing*

Skenario	Training	Testing
70:30	265	112
80:20	302	75
90:10	340	37

Preprocessing

Pada tahap *Preprocessing* menggunakan *tools* Rapidminer untuk membantu dalam proses pengerjaannya meliputi *Tokenize*, *Transform Case*, dan *Filter Stopwords*. *Tokenize* merupakan proses pemecahan setiap kata atau teks ulasan sekaligus proses untuk menghilangkan tanda baca (Fitriyani & Arifin, 2020). Pada proses *tokenize* beberapa kata data tweet akan dipecah, tanda baca dan angka akan dihilangkan. Bisa dilihat pada tabel 5 dibawah ini :

Tabel 5. Hasil *Tokenize*

Tweet	<i>Tokenize</i>
TIX ID gue gedig juga lu, lama banget updetnya	TIX ID gue gedig juga lu lama banget updetnya
Aku lebih nyaman di TIX ID sih, karna ya bisa sekalian nyesuain mau pesen xxi, cine apa cgv di satu ap	Aku lebih nyaman di TIX ID sih karna ya bisa sekalian nyesuain mau pesen xxi cine apa cgv di satu ap
Dari tadi mantengin TIX ID reset lama banget	Dari tadi mantengin TIX ID reset lama banget
TIX ID mau beli tiket penuh terus astagaa, masa tiap jamnya ga ada yg kosong	TIX ID mau beli tiket penuh terus astagaa masa tiap jamnya ga ada yg kosong
Pake TIX ID bisa milih mau nonton dimana	Pake TIX ID bisa milih mau nonton dimana

Setelah tahap *Tokenize* selesai, maka tahap selanjutnya yaitu *transform case* untuk merubah tulisan dari huruf besar menjadi kecil agar mempermudah dalam pengolahan data (Subagja et al., 2021). Berikut merupakan hasil *Transform Case* bisa dilihat pada tabel 6 sebagai berikut :

Tabel 6. Hasil Proses Transform Case

<i>Tokenize</i>	<i>Transform case</i>
TIX ID gue gedig juga lu lama banget updetny	TIX ID gue gedig juga lu lama banget updetny
Aku lebih nyaman di TIX ID sih karna ya bisa sekalian nyesuain mau pesen xxi cine apa cgv di satu ap	aku lebih nyaman di TIX ID sih karna ya bisa sekalian nyesuain mau pesen xxi cine apa cgv di satu ap
Dari tadi mantengin TIX ID reset lama banget	dari tadi mantengin TIX ID reset lama banget
TIX ID mau beli tiket penuh terus astagaa masa tiap jamnya ga ada yg kosong	TIX ID mau beli tiket penuh terus astagaa masa tiap jamnya ga ada yg kosong
Pake TIX ID bisa milih mau nonton dimana	pake TIX ID bisa milih mau nonton dimana

Setelah tahap *Transform Case* selesai, maka proses terakhir pada tahap *preprocessing* dalam penelitian ini yaitu *filter stopwords* Teknik untuk menghapus konjungsi atau kata berimbuhan seperti ‘dan’, ‘yang’, atau ‘ke’ hingga hanya menempatkan kata-kata bermakna yang memiliki arti penting untuk menghilangkan *stopwords* (Doloksaribu & Samuel, 2022). perubahan untuk *stopwords* yang dihilangkan bisa dilihat pada tabel 7 dibawah ini :

Tabel 7. Hasil Proses Filter Stopwprds

<i>Transform case</i>	<i>Filter Stopwords</i>
TIX ID gue gedig juga lu lama banget updetny	TIX ID gue gedig juga lu updetny
aku lebih nyaman di TIX ID sih karna ya bisa sekalian nyesuain mau pesen xxi cine apa cgv di satu ap	aku di TIX ID sih karna ya nyesuain mau pesen xxi cine apa cgv di satu ap
dari tadi mantengin TIX ID reset lama banget	dari tadi mantengin TIX ID reset
TIX ID mau beli tiket penuh terus astagaa masa tiap jamnya ga ada yg kosong	TIX ID mau beli tiket terus masa tiap jamnya ga ada yg kosong
pake TIX ID bisa milih mau nonton dimana	pake TIX ID nonton dimana

Pembobotan TF-IDF

Pada tahap ini peneliti menggunakan data hasil hasil *Preprocessing*. Dibawah ini merupakan contoh dataset yang digunakan dalam proses pembobotan TF-IDF pada tabel 8 Dibawah ini :

Tabel 8. Dataset TF-IDF

Tweet	Sentimen
selalu make TIX ID lebih gampang	Positif
paling enak emang pake TIX ID	Positif
TIX ID bisa milih bioskop	Positif
TIX ID suka diskon	Positif
TIX ID banyak promo	Positif
TIX ID gangguan mulu	Negatif
TIX ID ga guna	Negatif
TIX ID error mulu	Negatif
capek sama TIX ID	Negatif
TIX ID gue gedig juga lu lama banget updatenya	Negatif

dilakukan perhitungan kemunculan kata dari berbagai kalimat menggunakan pembobotan TF-IDF. Pada tahap *Term Frequency* jumlah munculnya kata semakin besar (*TF*) maka bobotnya akan besar pada dokumen. Kemudian pada tahap selanjutnya melakukan perhitungan berapa jumlah kata yang muncul dalam dokumen tersebut atau dengan istilah lain menghitung *document frequency (DF)* lalu total dari semua dokumen dibagi oleh nilai *Document Frequency* untuk menghasilkan nilai *Inverse Document Frequency*.

TF-IDF menggunakan Rapidminer

Pembobotan kata TF-IDF pada tools rapidminer, tahap ini dilakukan dengan operator *process document from data* dengan memilih TF-IDF pada parameter *vector creation*.

Pemodelan data

Pada tahap pemodelan data, ada beberapa proses yang dilakukan yaitu *cross Validation*, Pemodelan menggunakan *Support Vector Machine* dan pengujian menggunakan tools Rapidminer.

Cross Validation

Tahap selanjutnya yang dilakukan adalah Proses cross validation melalui algoritma SVM dibagi menjadi data training dan testing berbentuk kinerja dan model yang diterapkan. Untuk mengisi parameter number of folds diisi dengan angka 10. Proses *Cross Validation* terdiri dua sub proses diantaranya *training* dan *testing*, data dilatih dengan menggunakan algoritma SVM.

Pemodelan menggunakan Support Machine Vector

Pada tahap proses pemodelan dengan menggunakan algoritma support vector dengan menggunakan tools RapidMiner, pada penelitian ini menggunakan kernel dot, radial, dan polynomial pada setiap skenario pembagian data set. Di bawah ini merupakan tabel hasil pemodelan *support vector machine* menggunakan *kernel dot*, *radial*, dan *polynomial* bisa dilihat pada tabel 9 dibawah ini :

Tabel 9. Hasil Akurasi Rasio Data dengan Kernel

Rasio Pembagian Data	Akurasi		
	Kernel		
	Dot	Radial	polynomial
70:30	53.58%	66.42%	67.92%
80:20	74.17%	65.89%	65.23%
90:10	65.88%	67.35%	65.00%

Evaluasi

Pada tahap terakhir yang dilakukan adalah perhitungan manual dengan *confusion matrix* untuk menguji hasil. Selain dari hasil *confusion matrix* akan ditampilkan juga hasil *wordcloud* berdasarkan dataset yang sudah diolah pada tahap sebelumnya. Dari proses cross validation menggunakan algoritma *support vector machine* dengan *kernel dot* dengan skenario rasio perbandingan data *training* dan testing 80:20 untuk analisis sentimen pengguna tweet terhadap aplikasi TIX ID menghasilkan nilai *accuracy* tertinggi sebesar 74,17%. Dibawah ini merupakan gambar nilai *accuracy* algoritma SVM untuk analisis sentimen pengguna twitter terhadap aplikasi TIX ID menggunakan tools RapidMiner :

accuracy: 74.14% +/- 6.55% (micro average: 74.17%)

	true positif	true negatif	class precision
pred. positif	65	23	73.86%
pred. negatif	55	159	74.30%
class recall	54.17%	87.36%	

Gambar 2. Nilai Akurasi Algoritma SVM

Confusion Matrix

Confusion matrix yaitu tabel yang memberi informasi klasifikasi jumlah benar data yang diuji dan jumlah salah data yang diuji (Normawati & Prayogi, 2021). *confusion matrix* memiliki informasi yang dapat dibandingkan hasil algoritma yang diperoleh sistem yang seharusnya. Berikut merupakan rumus yang digunakan dalam *Confusion Matrix* bisa dilihat pada tabel 10 dibawah ini :

Tabel 10. Rumus Confusion Matrix

	True Class	
Predicted Class	True Positif (TP)	False Positif (FP)
	False Negatif (FN)	True Negatif (TN)

Berikut merupakan hasil *confusion matrix* dari algoritma *Support Vector Machine* bisa dilihat pada tabel 11. dibawah ini :

Tabel 11. Hasil Confusion Matrix

	True Class	
Predicted Class	65	23
	55	159

Berikut ini merupakan hasil perhitungan *confusion matrix* yang dilakukan secara manual :

$$\text{Rumus mencari Akurasi} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (1)$$

Wordcloud

Word Cloud merupakan visualisasi kata-kata yang sering muncul atau digunakan dari masing-masing label sentimen positif ataupun negatif yang dianalisis (Adiyana et al., 2015). Kata-kata sentimen positif disajikan dalam bentuk *word cloud* berwarna biru. Beberapa kata yang memiliki font lebih besar artinya lebih sering dituliskan seperti “TIX ID”, “di tix”, “id sering”, “enakan tix”,

“lebih nyaman” dan lainnya Dibawah ini merupakan gambar visualisai *word cloud* dari sentimen positif pengguna aplikasi TIX ID bisa dilihat pada gambar 3 dibawah ini :



Gambar 3. Wordcloud Sentimen Positif

Sedangkan untuk sentimen negatif akan diberikan visualisasi *word cloud* berwarna merah. Beberapa kata yang lebih sering dituliskan seperti “di tix”, “TIX ID”, “tix”, “pake tix”, “mantengin tix”, “penuh terus”, “ke refund”, “ga update”, “suka error”, “tix selalu penuh” dan lainnya Dibawah ini merupakan gambar visualisai *word cloud* dari sentimen negatif pengguna aplikasi TIX ID bisa dilihat pada gambar 4. Dibawah ini :



Gambar 4. Wordcloud sentimen Negatif

Berdasarkan visualisasi sentimen menggunakan *word cloud* digunakan untuk mengetahui berbagai komentar yang terdapat dari media sosial dapat digunakan sebagai bahan acuan untuk evaluasi bagi pihak perusahaan dalam upaya meningkatkan kualitas layanan aplikasi tix id terhadap pengguna dan bagi calon pengguna dapat dijadikan sebagai daya tarik serta bahan pertimbangan sebelum menggunakan suatu produk/jasa karena tentu saja calon pengguna pasti akan mencari mencari informasi terlebih dahulu terkait produk/jasa yang akan mereka gunakan.

KESIMPULAN DAN SARAN

Hasil penelitian dengan analisis sentimen terhadap tweet TIX ID dapat disimpulkan bahwa akurasi tertinggi diperoleh algoritma *support vector machine* dengan *kernel dot* dan skenario pembagian data 80:20 akurasi yang dihasilkan sebesar 74,17%. Sehingga dapat disimpulkan bahwa algoritma *support vector machine* memperoleh nilai cukup baik. Analisis sentimen pengguna twitter terhadap aplikasi TIX ID menghasilkan 146 sentimen positif dan 231 sentimen negatif. Berdasarkan nilai tersebut membuktikan sentimen masyarakat mengenai aplikasi TIX ID di media sosial twitter tergolong negatif dan pengguna masih kurang puas dengan aplikasi karena update jadwal yang tidak pasti, aplikasi sering error, dan terjadi nya *refund*. Pada penelitian berikutnya dapat melakukan penambahan waktu dalam proses pengumpulan data. Diharapkan dapat menganalisis dengan metode lain atau membandingkan dengan algoritma lainnya untuk bahan evaluasi agar bisa ditingkatkan.

DAFTAR PUSTAKA

- Adiyana, I., Hakim, R. B. F., Si, S., & Si, M. (2015). *IMPLEMENTASI TEXT MINING PADA MESIN PENCARIAN TWITTER UNTUK MENGANALISIS TOPIK – TOPIK TERKAIT “ KPK dan JOKOWI .”* 570–581.
- Agustina, S. P. M. (2021). *Aplikasi pengelolaan atk pada dinas kebudayaan dan pariwisata provinsi sumatera selatan 1.* 156–160.
- Alhaq, Z., Mustopa, A., Mulyatun, S., & Santoso, J. D. (2021). Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter. *Journal of Information System Management (JOISM)*, 3(2), 44–49. <https://doi.org/10.24076/joism.2021v3i2.558>
- Ananda, F. D., & Pristyanto, Y. (2021). Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(2), 407–416. <https://doi.org/10.30812/matrik.v20i2.1130>
- Aziz Syahputro, R., Widodo, & Ajie, H. (2019). Bayes Classifier dan Support Vector Machine dalam Klasifikasi Judul Karya Akhir Mahasiswa Program Studi PTIK UNJ. *PINTER : Jurnal Pendidikan Teknik Informatika Dan Komputer*, 3(1), 54–62. <https://doi.org/10.21009/pinter.3.1.9>
- Cahyani, W., & Raya Sulistyowati. (2021). PENGARUH SALES PROMOTION DAN ELECTRONIC WORD OF MOUTH TERHADAP KEPUTUSAN PEMBELIAN TIKET BIOSKOP MELALUI APLIKASI TIX ID (Studi pada Pengguna Aplikasi TIX ID di Surabaya) Wiwik Cahyani Raya Sulistyowati. *Jurnal Pendidikan Tata Niaga (JPTN)*, 9(1), 1055–1061.
- Darna, N., Herlina, E., Tetap, D., Studi, P., Fakultas, M., Universitas, E., & Ciamis, G. (2018). *Memilih metode penelitian yang tepat: bagi penelitian bidang ilmu manajemen.* 5(April), 287–292.
- Doloksaribu, H. P., & Samuel, Y. T. (2022). Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi. *Jurnal Teknologi Informasi: Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika*, 16(1), 1–11.
- Fatmawati. (2016). Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes. *Jurnal Techno Nusa Mandiri*, XIII(1), 50.
- Fitriyani, F., & Arifin, T. (2020). Penerapan Word N-Gram Untuk Sentiment Analysis Review Menggunakan Metode Support Vector Machine (Studi Kasus: Aplikasi Sambara). *Sistemasi*, 9(3), 610. <https://doi.org/10.32520/stmsi.v9i3.954>
- Gifari, O. I., Adha, M., Hendrawan, I. R., Freddy, F., & Durrand, S. (2022). *Analisis Sentimen*

- Review Film Menggunakan TF-IDF dan Support Vector Machine.* 2(1), 36–40.
- Gunawan, D., Riana, D., Ardiansyah, D., Akbar, F., & Alfarizi, S. (2020). *Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023.* VI(1), 2442–2436. <https://doi.org/10.31294/jtk.v4i2>
- Haditira, R., Informatika, F., Telkom, U., Review, S., Murdiansyah, D. T., Informatika, F., Telkom, U., Astuti, W., Informatika, F., & Telkom, U. (2022). *Analisis Sentimen Pada Steam Review Menggunakan Metode Multinomial Naïve Bayes dengan Seleksi Fitur Gini Index Text.* 9(3), 1793–1799.
- Homepage, J., Hardian, R. W., Prasetyo, E., Khaira, U., Suratno, T., Sains, F., & Teknologi, D. (2021). *MALCOM: Indonesian Journal of Machine Learning and Computer Science Online Lecture Sentiment Analisis On Twitter Social Media During The Covid-19 Pandemic Using Sentistrength Algorithm Analisis Sentiment Kuliah Daring di Media Sosial Twitter Selama Pandem.* 1(October), 138–143.
- Ismail, I., Prasetyo, I., Kusuma, A. O., Putri, A. F., Yunitasari, I., Manthovani, S., Nurhidayat, Y., AS, M. D. D., Rahmawati, Y., & Kumiawan, A. (2022). *Sosialisasi Penggunaan Media Sosial Yang Bijak Dan Aman Di SMK Sasmita Jaya 1.* *JATIMIKA: Jurnal Kreativitas Mahasiswa Informatika,* 2(1), 81–84. <http://openjournal.unpam.ac.id/index.php/JATIMIKA/article/view/14465>
- Murdiono, R. A., Tolle, H., & Kharisma, A. P. (2018). *Evaluasi User Experience Pada Aplikasi Mobile Penjualan Tiket Online.* *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer,* 2(5), 2078–2085. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1544>
- Normawati, D., & Prayogi, S. A. (2021). *Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter.* *Jurnal Sains Komputer & Informatika (J-SAKTI),* 5(2), 697–711.
- Petiwi, M. I., Triayudi, A., & Sholihati, I. D. (2022). *Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine.* *Jurnal Media Informatika Budidarma,* 6(1), 542. <https://doi.org/10.30865/mib.v6i1.3530>
- Pravina, A. M., Cholissodin, I., & Adikara, P. P. (2019). *Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM).* *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer,* 3(3), 2789–2797. <http://j-ptiik.ub.ac.id>
- Rahayu, S., Alfeno, & Wahyono, N. (2019). *Rancang Bangun Aplikasi Sistem.* 5(1), 1–11.
- Rahutomo, F., Saputra, P. Y., & Fidyawan, M. A. (2018). *Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine.* *Jurnal Informatika Polinema,* 4(2), 93. <https://doi.org/10.33795/jip.v4i2.152>
- Rudiansyah, S., Selvia, N., & Alfarizi, S. (2022). *Perancangan Aplikasi Penjualan Toko Citra Bahari Gypsum Berbasis Java.* *Jurnal Riset Dan Aplikasi Mahasiswa Informatika (JRAMI),* 3(01), 141–148. <https://doi.org/10.30998/jrami.v3i01.3045>
- Sari, R. (2020). *Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn).* *EVOLUSI : Jurnal Sains Dan Manajemen,* 8(1), 10–17. <https://doi.org/10.31294/evolusi.v8i1.7371>
- Sarmidi, S. I. R. (2018). *Aplikasi Scanner Berbasis Android Untuk Menampilkan Data Id Card Menggunakan Barcode.* *Jurnal Manajemen Dan Teknik Informatika,* 02(01), 181–190. <http://jurnal.stmik-dci.ac.id/index.php/jumantaka/article/view/418/472>
- Subagja, R. A., Widiastiwi, Y., & Chamidah, N. (2021). *Klasifikasi Ulasan Aplikasi Jenius pada Google Play Store Menggunakan Algoritma Naive Bayes.* 4221(3), 197–208.
- 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. <https://doi.org/10.30865/mib.v4i3.2181>