

Perbandingan Analisis Sentimen Aplikasi Traveloka dan Tiket.com pada Twitter dengan Metode *Support Vector Machine*

Putri Utami Rukmana

Fakultas Rekayasa Industri

Universitas Telkom

Bandung, Indonesia

putriutamirukmana@gmail.com

Oktariani Nurul Pratiwi

Fakultas Rekayasa Industri

Universitas Telkom

Bandung, Indonesia

onurulp@telkomuniversity.ac.id

Hanif Fakhruroja

Fakultas Rekayasa Industri

Universitas Telkom

Bandung, Indonesia

haniff@telkomuniversity.ac.id

Abstract— The emergence of the COVID-19 pandemic in Indonesia resulted in an economic crisis, including in the world of tourism, which caused a decline in the national economy. With the existence of Online Travel Agencies (OTA) such as Traveloka and Tiket.com, it is hoped that they can help improve the tourism sector for the Indonesian economy. As a popular OTA and to see the opinion of the Indonesian people, it can be seen from public opinion in the form of tweets on the Twitter application. The tweets data will be taken and sentiment analysis will be carried out on the OTA Traveloka and Tiket.com applications which will be classified into certain classes based on opinions and modeling will be carried out using the Support Vector Machine (SVM) algorithm method. This research aims to determine the level of accuracy of the SVM algorithm and find out how sentiment analysis compares between Traveloka and Tiket.com. In the sentiment analysis comparison, in terms of price, Traveloka is superior and in terms of service, Tiket.com is superior. After modeling by comparing splitting data and handling imbalanced data using Synthetic Minority Oversampling Technique (SMOTE), the best SVM accuracy results for the Tiket.com price dataset were 68%, for Traveloka prices it was 97%, for Tiket.com services it was 92%, and for Traveloka services it is 89%.

Keywords— Sentiment Analysis, Support Vector Machine (SVM), Traveloka, Tiket.com, Online Travel Agency (OTA)

I. PENDAHULUAN

Adanya pandemi COVID-19 di Indonesia menimbulkan krisis ekonomi termasuk dunia pariwisata. Salah satu dampak pandemi terhadap industri pariwisata adalah penurunan pariwisata dalam perekonomian nasional di berbagai negara, termasuk Indonesia [1]. Untuk membantu pemulihan sektor pariwisata di Indonesia dapat dibantu dengan kehadiran *Online Travel Agent* (OTA). Dari data survei yang dilakukan oleh DailySocial.id pada kuartal 1 tahun 2022 bahwa aplikasi Traveloka dan Tiket.com sebagai startup hotel dan perjalanan terpopuler di Indonesia [2]. Untuk melihat opini masyarakat mengenai aplikasi tersebut dapat dilihat melalui Twitter. Berdasarkan survei yang dilakukan oleh Kepios didapatkan bawa Twitter di Indonesia masuk ke dalam lima teratas platform media sosial favorit yang digunakan masyarakat [3].

Maka dari itu, Traveloka dan Tiket.com sebagai aplikasi OTA yang sangat populer di Indonesia menawarkan berbagai layanan yang menarik kepada customer. Sehingga, Traveloka dan Tiket.com bersaing untuk mendapatkan customer, keduanya sering kali menawarkan berbagai layanan dan promo yang ditawarkan untuk menarik masyarakat dan untuk meningkatkan kepuasan pelanggan. Karena pemberian layanan yang baik berpengaruh secara signifikan terhadap kepuasan pelanggan serta adanya promo penjualan dapat mempengaruhi tingkat kepuasan pelanggan [4]. Untuk melihat bagaimana perbandingan antara layanan dan harga yang ditawarkan dengan kenyataan yang dirasakan oleh customer dilakukan analisis sentimen untuk menggali informasi mengenai suatu entitas dan mengidentifikasi subjektivitas entitas tersebut untuk menentukan teks atau pendapat yang dihasilkan oleh user itu positif, negatif, atau netral [5].

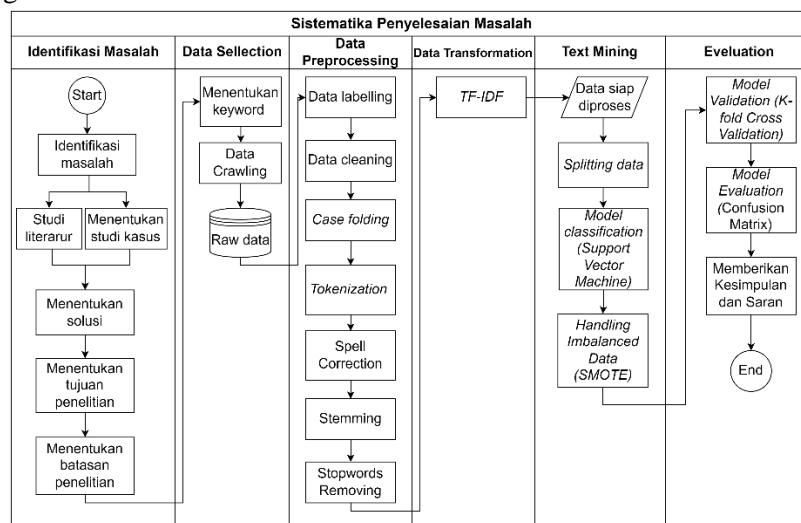
Dalam penelitian ini, penulis menggunakan metode *Support Vector Machine* (SVM) yang merupakan metode *supervised learning* yang populer dalam melakukan klasifikasi

teks termasuk tweets menjadi kategori positif atau negatif. Karena, SVM dapat membentuk hyperlane yang berfungsi sebagai pemisah antar kategori [6].

Berdasarkan hal di atas, penulis melakukan penelitian terkait sentimen analisis pada aplikasi OTA Traveloka dan Tiket.com dengan data yang diambil dari Twitter yang akan diproses menggunakan metode algoritma SVM. Penelitian ini bertujuan untuk mengetahui akurasi algoritma SVM dan mengetahui bagaimana perbandingan sentimen antara Traveloka dan Tiket.com.

II. METODE PENELITIAN

Penelitian ini menggunakan metode yang diadaptasi dari *Knowledge Discovery in Database* (KDD) sebagai salah satu pengetahuan dalam bidang *database* yang dapat membantu dalam mengenali pola dalam kumpulan data yang meningkat [7]. Maka dari itu, KDD akan diadopsi dalam penelitian ini yang akan disesuaikan tahapannya secara rinci, sebagai berikut:



Gambar 1. Sistematika Penyelesaian Masalah

A. Data Selection

Pengumpulan data menggunakan metode *data crawling* dengan library *sns scrape* pada Python. Data yang diambil dari Twitter yang berupa *tweets* Bahasa Indonesia terkait harga dan layanan pada Traveloka dan Tiket.com. Data yang dikumpulkan *tweets* pada rentang waktu dari bulan Juni 2022 sampai bulan Februari 2023.

B. Data Preprocessing

Data preprocessing bertujuan membersihkan *tweet* dari *noise* untuk mempermudah proses selanjutnya sehingga proses klasifikasi mendapatkan akurasi yang lebih akurat [8], atau untuk mengubah data tidak terstruktur menjadi data terstruktur [9]. Adapun langkah-langkah yang digunakan dalam penelitian ini, sebagai berikut:

1) Data Labelling

Proses ini dilakukan secara manual dengan mengidentifikasi label positif atau negatif atau netral. Namun, label netral tidak digunakan karena tidak relevan serta tidak memberikan informasi signifikan.

2) Data Cleaning

Membersihkan data teks dari sesuatu yang tidak berarti yang bukan *natural language* [10]. Terdapat beberapa langkah seperti: menghapus data duplikasi, menghapus username Twitter, menghapus *hashtag*, menghapus *link https/ http/ www/ html*, menghapus non-ASCII dan simbol, menghapus baris baru, menghapus spasi di awal dan di akhir *tweets*, menghapus angka, menghapus *punctuation* atau tanda baca, dan menghapus data yang memiliki label netral.

3) Case Folding

Melakukan pengubahan karakter menjadi huruf kecil semua pada teks. Perbedaan huruf kapital dan huruf kecil pada suatu kata yang sama akan dianggap berbeda oleh sistem [10].

4) Tokenization

Melakukan pemisahan data teks menjadi token-token yang berbentuk kata, karakter spesial, atau tanda baca [10]. Dalam penelitian ini memanfaatkan fungsi `word_tokenize()` dari *library* nltk.

5) Spell Correction

Digunakan untuk memperbaiki kesalahan kata pada dataset seperti memperbaiki penulisan yang salah atau *typo*. Tahap ini dilakukan secara manual.

6) Stemming

Mengubah suatu kata menjadi bentuk kata dasarnya dengan menghilangkan imbuhan [10]. Tahap ini memanfaatkan *library* StemmerFactory dari sastrawi untuk *stemming* bahasa Indonesia.

7) Stopwords Removing

Melakukan penghapusan pada kata umum yang sering muncul dengan peran yang tidak penting [10]. Pada tahap ini menggunakan *library* nltk dengan fungsi `stopwords.words('indonesian')` dan *dictionary stopwords* tambahan.

Adapun contoh dari proses *data preprocessing*, sebagai berikut:

Tabel 1
Contoh Data Preprocessing

Tweets Sebelum	Tweets Sesudah
@JennyJusuf Traveloka, karena ga pake tiketcom. pernah salah tanggal booking langsung telpon cs di bantu proses refund sampe kembali 100%, pernah juga salah tf ke akunnya traveloka, cepet2 email cs dan langsung di proses 🎉	['ga', 'pakai', 'tiketcom', 'pernah', 'salah', 'tanggal', 'pesan', 'langsung', 'telepon', 'customer service', 'bantu', 'proses', 'refund', 'pernah', 'salah', 'transfer', 'akun', 'cepat', 'email', 'customer service', 'langsung', 'proses']

C. Data Transformation

Data akan diubah menjadi angka dengan pembobotan kata menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) dimana setiap kata diubah menjadi representasi vektor. Pada penelitian ini, untuk menghitung TF-IDF memanfaatkan fungsi `TfidfVectorizer` dari *library* scikit-learn. Adapun perhitungan TF-IDF, sebagai berikut [11]:

$$idf(t) = \ln \frac{1+n}{1+df(t)} + 1 \quad (1)$$

$$tf - idf(t, d) = tf(t, d) \times idf(t) \quad (2)$$

$$v_{norm} = \frac{v}{\|v\|_2} = \frac{v}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}} \quad (3)$$

$tf(t, d)$ adalah *term frequency*, $idf(t)$ adalah *inverse document frequency*, $tf - idf(t, d)$ adalah bobot term (t) dalam dokumen (d), v_{norm} adalah vektor normalisasi, v adalah vektor TF-IDF awal, dan n adalah jumlah total dokumen.

D. Text Mining

Pada tahap ini dilakukan dahulu *splitting data*. Pada penelitian ini, untuk hasil terbaik membandingkan rasio *data train : data test* 80:20, 70:30, dan 60:40. Lalu untuk menangani *imbalanced data* menggunakan metode *oversampling* dengan algoritma *Synthetic Minority Oversampling Technique* (SMOTE). Didapatkan hasil akurasi dari hasil klasifikasi oleh *Support Vector Machine* (SVM) dengan kernel linear.

Tabel 2
Distribusi Data

Dataset	Rasio	Jumlah Data Train				Jumlah Data Test	
		Sebelum SMOTE		Sesudah SMOTE		Positif	Negatif
		Positif	Negatif	Positif	Negatif	Positif	Negatif
Harga Tiket.com	80:20	164	58	164	164	40	16
	70:30	145	49	145	145	59	25
	60:40	126	40	126	126	78	34
Harga Traveloka	80:20	3854	114	3854	3854	967	26
	70:30	3370	102	3370	3370	1451	38
	60:40	2890	86	2890	2890	1931	54
Layanan Tiket.com	80:20	119	138	138	138	32	33
	70:30	103	122	122	122	48	49
	60:40	84	109	109	109	67	62
Layanan Traveloka	80:20	114	294	294	294	26	76
	70:30	99	258	258	258	41	112
	60:40	83	223	223	223	57	147

E. Evaluation

Dalam penelitian ini metode evaluasi menggunakan *K-fold Cross Validation* dan *Confusion Matrix*.

1) K-fold Cross Validation

Metode validasi yang memeriksa *overfitting* dalam pemodelan klasifikasi dimana dataset yang ada secara acak dibagi kedalam data uji dan data latih sebanyak K bagian yang sama besar [8].

2) Confusion Matrix

Menampilkan informasi perbandingan hasil klasifikasi prediksi dan hasil klasifikasi sebenarnya dan jumlah data uji yang telah diklasifikasikan dengan benar dan salah [6]. Untuk Harga Tiket.com akurasi terbaik pada rasio *data splitting* 60:40, untuk Harga Traveloka pada rasio 60:40, untuk Layanan Tiket.com pada rasio 80:20, dan untuk Layanan Traveloka pada rasio 70:30.

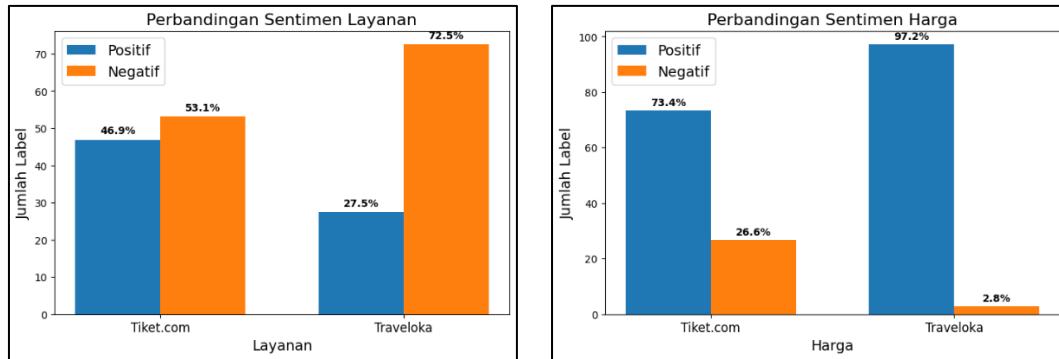
Tabel 3
Confusion Matrix

Dataset	TP	TN	FN	FP
Harga Tiket.com	68	8	10	26
Harga Traveloka	1916	19	15	35
Layanan Tiket.com	27	33	5	0
Layanan Traveloka	34	102	7	10

III. HASIL DAN PEMBAHASAN

A. Analisis Data

Untuk melihat gambaran perbandingan Traveloka dan Tiket.com dalam distribusi label positif dan negatif pada bagian harga dan layanan, sebagai berikut:



Gambar 2. Perbandingan Sentimen Harga dan Layanan

Jumlah data dalam dataset layanan Traveloka cenderung memiliki data yang lebih banyak dibandingkan data layanan Tiket.com. Untuk persentase data positif Tiket.com sebanyak 46,89% dengan jumlah data 151 sedangkan Traveloka memiliki data positif 27,45% dengan jumlah data 140. Untuk data negatif Traveloka yaitu 72,55% dengan jumlah data 370 sedangkan Tiket.com sebesar 53,11% dengan jumlah data 171.

Jumlah data dalam dataset haga Traveloka cenderung memiliki data yang lebih banyak dibandingkan data layanan Tiket.com. Untuk data positif Traveloka memiliki persentase data positif sebanyak 97,18% dengan jumlah data 4821 sedangkan Tiket.com memiliki data positif sebanyak 73,38% dengan jumlah data 204. Untuk data negatif Tiket.com memiliki persentase data yang lebih besar sebanyak 26,62% dengan jumlah data 741 sedangkan Traveloka hanya sebanyak 2,82% dengan jumlah data 140.

B. Word Clouds

Word clouds merupakan teknik untuk visualisasi data untuk memudahkan dalam mengenal atau mengidentifikasi konten dari kumpulan dokumen tekstual yang besar menjadi lebih cepat [12]. Setiap dataset divisualisasikan *word clouds* berdasarkan data positif dan data negatif.



Gambar 3. Word Clouds Positif Aspek Harga



Gambar 4. Word Clouds Negatif Aspek Harga



Gambar 5. Word Clouds Positif Aspek Layanan



Gambar 6. Word Clouds Negatif Aspek Layanan

C. N-gram

N-gram digunakan untuk menampilkan kata yang merepresentasikan setiap sentimen positif dan negatif dari setiap dataset.

```
--- N-grams yang Merepresentasikan Sentimen Positif Harga Tiket.com --
tiketcom harga gledek : 5
tiket konser raisa : 4
konser raisa gratis : 4
tiket gratis konser : 4
gratis konser raisa : 4
raya ulang hadir : 3
ulang hadir ragam : 3
hadir ragam ulang tahun : 3
harga gledek jam : 3
harga gledek tiketcom : 3

--- N-grams yang Merepresentasikan Sentimen Negatif Harga Tiket.com --
harga gledek tiket : 3
gledek harga tiket : 2
admin harga gledek : 2
harga gledek diskon : 2
sale kalah admin : 1
kalah admin aplikasi : 1
admin aplikasi sebelah : 1
aplikasi sebelah harga : 1
sebelah harga super : 1
harga super gledek : 1
```

Gambar 7. N-gram Aspek Harga Tiket.com

Berdasarkan gambar di atas mengenai aspek harga di Tiket.com dari sentimen positif menunjukkan pengguna menyukai adanya harga gledek yang ditawarkan, adanya penjualan tiket konser raisa dan Tiket.com yang membagikan tiket konser raisa gratis, dan adanya promo saat ulang tahun Tiket.com. Sedangkan, berdasarkan sentimen negatif menunjukkan kekecewaan beberapa pengguna karena tidak mendapatkan harga tiket saat harga gledek dan pengguna yang merasa sale atau promo di Tiket.com kalah dengan promo yang ditawarkan pada aplikasi lain.

```
--- N-grams yang Merepresentasikan Sentimen Positif Harga Traveloka ---
suka staycation hotel : 1858
staycation hotel suka : 1855
hotel suka staycation : 1855
hotel holiday stays : 584
inap hotel holiday : 350
ya inap hotel : 348
holiday stays pakai : 348
stays pakai diskon : 346
libur ya inap : 320
pakai diskon libur : 270

--- N-grams yang Merepresentasikan Sentimen Negatif Harga Traveloka ---
maling duit pelanggan : 10
rela pikir maling : 6
pikir maling duit : 6
duit hasil maling : 5
duit pelanggan program : 5
program binatang tanggung : 5
binatang tanggung rela : 5
tanggung rela pikir : 5
duit hasil tipu : 4
pelanggan program binatang : 4
```

Gambar 8. N-gram Harga Traveloka

Berdasarkan gambar di atas mengenai aspek harga di Traveloka pada sentimen positif menunjukkan pengguna menyukai staycation di hotel karena Traveloka menawarkan harga yang terjangkau dan pengguna memanfaatkan harga hotel yang terjangkau untuk liburan. Sedangkan, berdasarkan sentimen negatif menunjukkan kekecewaan beberapa pengguna yang berpendapat bahwa Traveloka mengambil uang pelanggan dan merasa ditipu karena uang mereka tidak dikembalikan atau proses refund yang tidak baik.

```
--- N-grams yang Merepresentasikan Sentimen Positif Layanan Tiket.com ---
pesan tiket libur : 23
konser raisa gratis : 21
ayo pesan tiket : 16
tiket konser raisa : 13
tiket gratis konser : 8
gratis konser raisa : 8
kasih tiket konser : 6
konser gratis raisa : 6
kalo pesan tiket : 6
libur pesan tiket : 6

--- N-grams yang Merepresentasikan Sentimen Negatif Layanan Tiket.com ---
telepon call center : 4
server jelek banget : 3
ya beli tiket : 2
server tiketcom kesal : 2
customer service tiketcom beda : 2
server tiketcom down : 2
tiketcom down gimana : 2
ga pesan tiket : 2
tiket telepon call : 2
suruh tunggu email : 2
```

Gambar 9. N-gram Layanan Tiket.com

Berdasarkan gambar di atas mengenai aspek layanan di Tiket.com pada sentimen positif menunjukkan pengguna senang dengan layanan pemesanan tiket untuk liburan dan adanya tiket konser raisa gratis. Sedangkan, berdasarkan sentimen negatif menunjukkan kekecewaan beberapa pengguna terhadap layanan *call center* dan *customer service* dan merasa kesal dengan server Tiket.com yang jelek atau *down*.

```

--- N-grams yang Merepresentasikan Sentimen Positif Layanan Traveloka ---
senang ayo lihat : 21
ayo lihat akomodasi : 21
lihat akomodasi pengalaman : 16
akomodasi pengalaman menginap : 13
pesan tiket pesawat : 12
pesan tiket bus : 5
pengalaman menginap hotel : 4
pesan tiket hotel : 3
hotel senang ayo : 3
ambah pengalaman mumpung : 3

--- N-grams yang Merepresentasikan Sentimen Negatif Layanan Traveloka ---
pesan tiket pesawat : 13
telepon call center : 10
hubung call center : 7
suruh tunggu jam : 6
admin pesan tiket : 5
ya pesan tiket : 5
pesan tiket via : 4
pesan tiket kereta : 4
admin tolong bantu : 4
customer service suruh tunggu : 4

```

Gambar 10. N-gram Layanan Tiket.com

Berdasarkan gambar di atas mengenai aspek layanan di Traveloka pada sentimen positif menunjukkan pengguna senang untuk mempersiapkan akomodasi untuk menginap dari Traveloka, pengguna senang untuk melakukan pesan tiket pesawat, bus, dan hotel, serta pengguna senang dengan pengalaman menginap di hotel dengan memesan tiket melalui Traveloka. Sedangkan, berdasarkan sentimen negatif menunjukkan kekecewaan beberapa pengguna yang tidak puas dengan pemesanan tiket pesawat dan kurang puas dengan layanan *call center* ketika mereka membutuhkan bantuan untuk mengatasi keluhan pemesanan tiket namun harus menunggu.

D. Text Mining

Berdasarkan hasil implementasi oversampling menggunakan SMOTE didapatkan hasil perbandingan akurasi pemodelan SVM sebelum menggunakan SMOTE dan sesudah menggunakan SMOTE, sebagai berikut:

Tabel 4
Hasil Akurasi SVM

Dataset	Akurasi					
	80:20		70:30		60:40	
	Sebelum SMOTE	Sesudah SMOTE	Sebelum SMOTE	Sesudah SMOTE	Sebelum SMOTE	Sesudah SMOTE
Harga Tiket.com	0,679	0,589	0,690	0,667	0,696	0,679
Harga Traveloka	0,977	0,941	0,978	0,942	0,976	0,975
Layanan Tiket.com	0,923	0,923	0,887	0,876	0,884	0,884
Layanan Traveloka	0,902	0,853	0,902	0,889	0,872	0,848

E. Evaluation

1) K-fold Cross Validation

Adapun hasil validasi menggunakan *K-fold Cross Validation*, sebagai berikut:

Tabel 5

Hasil K-fold Cross Validation

Dataset	Akurasi					
	K1	K2	K3	K4	K5	Rata-rata
Harga Tiket.com	0,679	0,679	0,786	0,782	0,690	0,723
Harga Traveloka	0,977	0,976	0,980	0,972	0,983	0,977
Layanan Tiket.com	0,923	0,877	0,875	0,875	0,828	0,876
Layanan Traveloka	0,902	0,794	0,882	0,872	0,823	0,855

2) Confusion Matrix

Dari hasil *confusion matrix* didapatkan metrik evaluasi untuk *precision*, *recall*, *f1-score*, dan *accuracy*, sebagai berikut:

Tabel 6
Perhitungan Metrik Evaluasi

Dataset	Precision	Recall	F1-Score	Accuracy
Harga Tiket.com	0,72	0,87	0,79	0,68
Harga Traveloka	0,98	0,99	0,99	0,97
Layanan Tiket.com	1,00	0,84	0,92	0,92
Layanan Traveloka	0,77	0,83	0,80	0,89

IV. KESIMPULAN

Berdasarkan hasil analisis dari penelitian yang telah dilakukan. Maka, diperoleh kesimpulan mengenai harga Traveloka memiliki sentimen positif yang lebih baik sebesar 97,18% sedangkan Tiket.com 73,38%. Sedangkan, perbandingan dalam sentimen negatif mengenai harga Tiket.com memiliki persentase data negatif lebih besar yaitu 26,62% sedangkan Traveloka 2,82%. Jika dilihat dari hasil analisis *word clouds* dan N-gram yang merepresentasikan sentimen positif terkait harga Traveloka dapat disimpulkan pengguna senang dengan harga terjangkau yang ditawarkan untuk penawaran *staycation* di hotel dengan diskon saat liburan serta untuk harga Tiket.com dapat disimpulkan pengguna menyukai adanya promo harga gledek, promo saat ulang tahun Tiket.com, harga yang relatif murah, dan customer yang antusias untuk mendapatkan tiket konser raisa dan untuk mendapatkan tiket konser raisa gratis. Jika dilihat dari hasil analisis dengan *word clouds* dan N-gram yang merepresentasikan sentimen negatif terkait harga untuk Tiket.com dapat disimpulkan ketidakpuasan pengguna pada harga tiket yang disediakan tidak mendapatkan harga gledek atau tidak mendapatkan potongan harga dan pengguna merasa *sale* atau promo di Tiket.com kalah dengan promo yang ditawarkan pada aplikasi lain serta untuk harga Traveloka dapat disimpulkan pelanggan kecewa terhadap harga tiket dan merasa ada yang ditipu bahwa Traveloka mengambil uang pelanggan dengan tidak mengembalikan uang mereka atau proses *refund* yang tidak baik. Maka, dapat disimpulkan bahwa jika dari segi harga Traveloka lebih unggul dibandingkan Tiket.com. Untuk perbandingan analisis sentimen mengenai layanan, Tiket.com memiliki persentase data positif lebih banyak sebesar 46,89% sedangkan Traveloka 27,45%. Sedangkan perbandingan dalam sentimen negatif Traveloka memiliki persentase data negatif lebih tinggi sebesar 72,55% dibanding Tiket.com sebesar 53,11%. Hal tersebut, jika dilihat dari analisis *word clouds* dan N-gram yang merepresentasikan sentimen positif terkait layanan Tiket.com dapat disimpulkan pengguna yang puas terhadap layanan pemesanan tiket, menunjukkan antusias pengguna untuk mendapatkan tiket konser raisa, dan Tiket.com memberikan layanan berbagi tiket konser raisa gratis serta untuk Traveloka menunjukkan pengalaman pemesanan tiket di Traveloka disenangi oleh pengguna dan untuk mempersiapkan akomodasi untuk menginap dari Traveloka, pelanggan senang untuk melakukan pesan tiket pesawat, bus, dan hotel. Jika dilihat dari hasil analisis dengan *word clouds* dan N-gram yang merepresentasikan sentimen negatif terkait layanan untuk Traveloka bahwa pelanggan tidak puas dengan pemesanan tiket pesawat dan kurang puas dengan layanan *call center* ketika mereka membutuhkan bantuan untuk mengatasi keluhan pemesanan tiket namun harus menunggu serta proses *refund* yang kurang baik serta untuk Tiket.com bahwa pengguna merasa *server* sering terjadi eror atau *down* dan *customer service* masih kurang baik serta *refund* yang kurang baik. Berdasarkan hal di atas, dapat disimpulkan bahwa jika dari segi layanan Tiket.com lebih unggul dibandingkan Traveloka.

Berdasarkan hasil implementasi penerapan algoritma klasifikasi *machine learning* menggunakan *Support Vector Machine* (SVM) untuk mendapatkan akurasi yang terbaik penulis membandingkan rasio untuk *splitting data* yaitu 80:20, 70:30, dan 60:40 serta menggunakan teknik *oversampling* untuk mengatasi *imbalanced data* menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) yang diterapkan pada seluruh dataset. Untuk dataset harga Tiket.com akurasi terbaik terdapat pada rasio 60:40 dengan hasil sebelum melakukan SMOTE memiliki akurasi 0,696 atau 70% sedangkan setelah melakukan SMOTE memiliki akurasi 0,679 atau 68%. Untuk dataset harga Traveloka akurasi terbaik terdapat pada rasio 60:40 dengan hasil sebelum melakukan SMOTE memiliki akurasi 0,976 atau 98% sedangkan setelah melakukan SMOTE memiliki akurasi 0,975 atau 97%. Untuk dataset layanan Tiket.com akurasi terbaik terdapat pada rasio 80:20 sebelum melakukan SMOTE memiliki akurasi 0,923 atau 92% sedangkan setelah melakukan SMOTE memiliki akurasi 0,923 atau 92%. Untuk dataset layanan Traveloka akurasi terbaik terdapat pada rasio 70:30 sebelum melakukan SMOTE memiliki akurasi 0,902 atau 90% sedangkan setelah melakukan SMOTE memiliki akurasi 0,889 atau 89%.

REFERENSI

- [1] Badan Pusat Statistik Indonesia, *Tourism Satellite Account Indonesia 2016-2019*. Jakarta: BPS-Statistics Indonesia, 2022.
- [2] D. Angelia, “6 Startup Hotel Perjalanan Paling Populer di Indonesia,” *DailySocial.id*, 2022. <https://goodstats.id/article/sering-liburan-intip-6-startup-hotel-dan-perjalanan-terpopuler-di-indonesia-q1-2022-HsAss> (accessed Oct. 14, 2022).
- [3] S. Kemp, “Digital 2022: Indonesia,” 2022. <https://datareportal.com/reports/digital-2022-indonesia>
- [4] B. Bambang and T. A. Wahyudi, “Analisis pengaruh citra perusahaan dan kualitas layanan terhadap loyalitas pelanggan melalui kepuasan pelanggan,” *J. Manaj. Strateg. dan Apl. Bisnis*, vol. 2, no. 1, pp. 61–70, 2019, doi: 10.36407/jmsab.v2i1.49.
- [5] P. Lin, X. Luo, and Y. Fan, “Sentiment Analysis Based on Deep Learning: A Comparative Study,” *Int. J. Comput. Inf. Eng.*, vol. 14, no. 12, pp. 473–485, 2020.
- [6] M. I. Fikri, T. S. Sabrina, and Y. Azhar, “Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter,” *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.
- [7] Nwagu, C. Kenneth, Omankwu, O. Chinecherem, and Inyiama, “Knowledge discovery in databases: An overview,” *Int. J. Comput. Sci. Inf. Secur.*, vol. 15, no. 12, pp. 12–16, 2017.
- [8] M. R. A. Nasution and M. Hayaty, “Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter,” *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.
- [9] R. A. Raharjo, I. M. G. Sunarya, and D. G. H. Divayana, “Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Kasus Analisis Sentimen Terhadap Data Vaksin,” *J. Ilm. Elektron. dan Komput.*, vol. 15, no. 2, pp. 456–464, 2022.
- [10] M. R. Faisal, D. Kartini, A. R. Arrahimi, and T. Hamonangan, *Text Mining Untuk Pemula*, 1st ed. Banjarbaru,: Scripta Cendekia, 2021.
- [11] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [12] Y. Kalmukov, “Using Word Clouds for Fast Identification of Papers’ Subject Domain and Reviewers’ Competences 15,” *Proc. Univ. Ruse*, vol. 60, pp. 114–119, 2021, [Online]. Available: www.compsystech.org