

**ANALISIS SENTIMEN PADA REVIEW OBJEK WISATA  
TKL ECOPARK MENGGUNAKAN  
ALGORITME SUPPORT VECTOR MACHINE DAN  
TEKNIK BALANCING DATASET**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Ilmu Komputer



disusun oleh

**PANDU ANGGIT WIRABUANA**

**19.11.3224**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**ANALISIS SENTIMEN PADA REVIEW OBJEK WISATA  
TKL ECOPARK MENGGUNAKAN  
ALGORITME SUPPORT VECTOR MACHINE DAN  
TEKNIK BALANCING DATASET**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Ilmu Komputer



disusun oleh

**PANDU ANGGIT WIRABUANA**

**19.11.3224**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**HALAMAN PERSETUJUAN  
SKRIPSI**

**ANALISIS SENTIMEN PADA REVIEW OBJEK WISATA  
TKL ECOPARK MENGGUNAKAN  
ALGORITME SUPPORT VECTOR MACHINE DAN  
TEKNIK BALANCING DATASET**

yang disusun dan diajukan oleh

**Pandu Anggit Wirabuana**

**19.11.3224**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 16 Agustus 2023

**Dosen Pembimbing,**



**Theopilus Bayu Sasongko, S.Kom, M.Eng**

**NIK. 190302375**

**HALAMAN PENGESAHAN**

**SKRIPSI**

**ANALISIS SENTIMEN PADA REVIEW OBJEK WISATA  
TKL ECOPARK MENGGUNAKAN  
ALGORITME SUPPORT VECTOR MACHINE DAN  
TEKNIK BALANCING DATASET**

yang disusun dan diajukan oleh

**Pandu Anggit Wirabuana**

**19.11.3224**

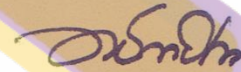
Telah dipertahankan di depan Dewan Penguji  
pada tanggal 16 Agustus 2023

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

Windha Mega Pradnya D, M.Kom  
NIK. 190302185



Muhammad Rudyanto Arief, M.T  
NIK. 190302098



Theopilus Bayu Sasongko, S.Kom, M.Eng  
NIK. 190302375



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 16 Agustus 2023

**DEKAN FAKULTAS ILMU KOMPUTER**



Hanif Al Fatta, S.Kom., M.Kom.

NIK. 190302096

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

**Nama mahasiswa : Pandu Anggit Wirabuana**  
**NIM : 19.11.3224**

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Sentimen Pada Review Objek Wisata TKL Ecopark Menggunakan Algoritme Support Vector Machine dan Teknik Balancing Dataset**

Dosen Pembimbing : Theopilus Bayu Sasongko, S.Kom, M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 16 Agustus 2023

Yang Menyatakan,



Pandu Anggit Wirabuana

## HALAMAN PERSEMBAHAN

Puji syukur kehadiran Allah SWT, atas limpahan rahmat dan karunia-Nya sehingga saya dapat menyelesaikan laporan skripsi ini. Hasil karya ini saya persembahkan kepada diri saya sendiri yang telah berjuang semaksimal mungkin di bangku perkuliahan dan kepada rang-orang yang telah berjasa memberi dukungan tiada henti :

1. Kedua orang tua saya, Bapak Darmadi dan Ibu Dewi yang senantiasa mendoakan setiap langkah serta berbagai macam dukungan untuk mengantarkan saya menuju kesuksesan.
2. Seluruh keluarga besar saya yang senantiasa mendoakan, mendukung, dan menantikan kelulusan saya.
3. Dosen pembimbing saya, Theopilus Bayu Sasongko, S.Kom, M.Eng, yang senantiasa memberikan arahan, perbaikan selama penyusunan dan pelaksanaan penelitian.
4. Orang-orang terkasih dan para sahabat terdekat saya yang tidak dapat saya sebutkan satu persatu yang telah selalu memberikan dukungan, menghibur dan menemani. Terutama kepada Zahra Diniatul Arimbi yang telah berperan penting dalam mendukung dan membantu kegiatan penelitian ini.
5. Teman-teman mahasiswa Informatika yang tidak dapat saya sebutkan satu persatu.

## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi dengan judul "Analisis Sentimen Pada Review Objek Wisata TKL Ecopark Menggunakan Algoritme Support Vector Machine dan Teknik Balancing Dataset". Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan studi di Program Sarjana Universitas Amikom Yogyakarta.

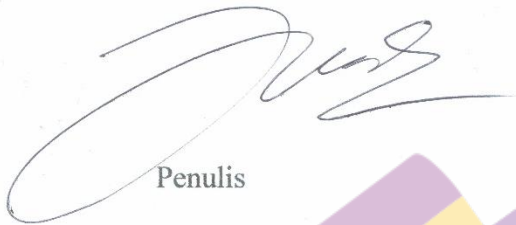
Dengan selesainya skripsi ini, maka penulis tidak lupa mengucapkan terimakasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta.
2. Hanif Al Fatta, S.Kom., M.Kom, selaku Dekan Fakultas Ilmu Komputer.
3. Theopilus Bayu Sasongko, S.Kom, M.Eng, selaku dosen pembimbing saya yang telah banyak memberikan pengarahan bagi penulis dalam pembuatan skripsi.
4. Bapak-Ibu Dosen Universitas Amikom Yogyakarta, atas ilmu dan pengetahuan yang telah diberikan kepada penulis selama duduk di bangku kuliah.
5. Kedua orang tua saya, Bapak Darmadi dan Ibu Dewi yang tak pernah berhenti untuk selalu memberikan selamat serta doa.
6. Seluruh teman-teman angkatan 2019, terutama untuk kelas IF 10, terimakasih atas kerjasamanya selama ini.
7. Teman-teman seperjuangan (ilham, fahmudin, filah, bagas, aji, damar, anton, adli, farhan dan yohan) yang telah memberikan dukungan dan terimakasih atas momen momen yang tak pernah terlupakan.

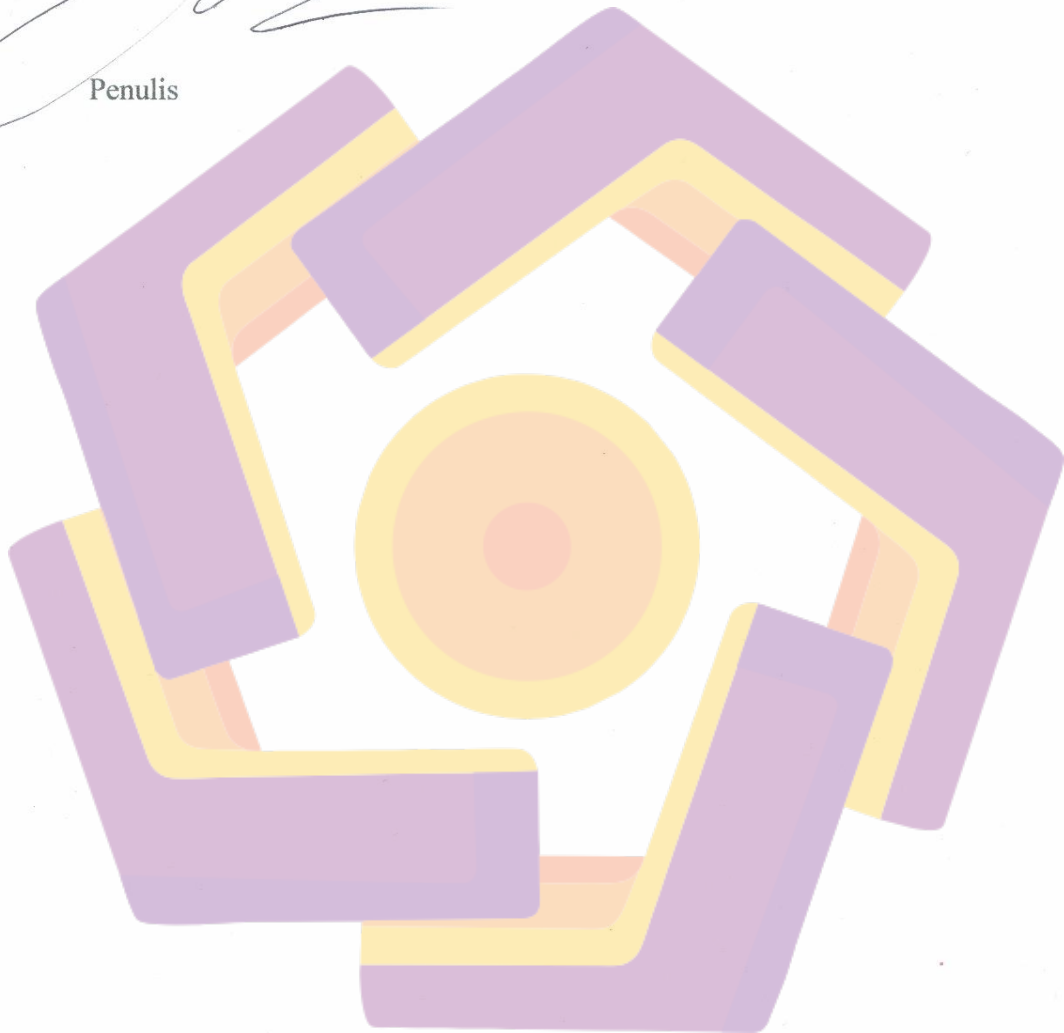
Penulis tentunya menyadari bahwa pembuatan skripsi ini masih banyak sekali kekurangan-kekurangan dan kelemahan-kelemahannya. Oleh karena itu penulis berharap kepada semua pihak agar dapat menyampaikan kritik dan saran

yang membangun untuk menambah kesempurnaan skripsi ini. Namun penulis tetap berharap skripsi ini akan bermanfaat bagi semua pihak yang membacanya.

Yogyakarta, 16 Agustus 2023



Penulis

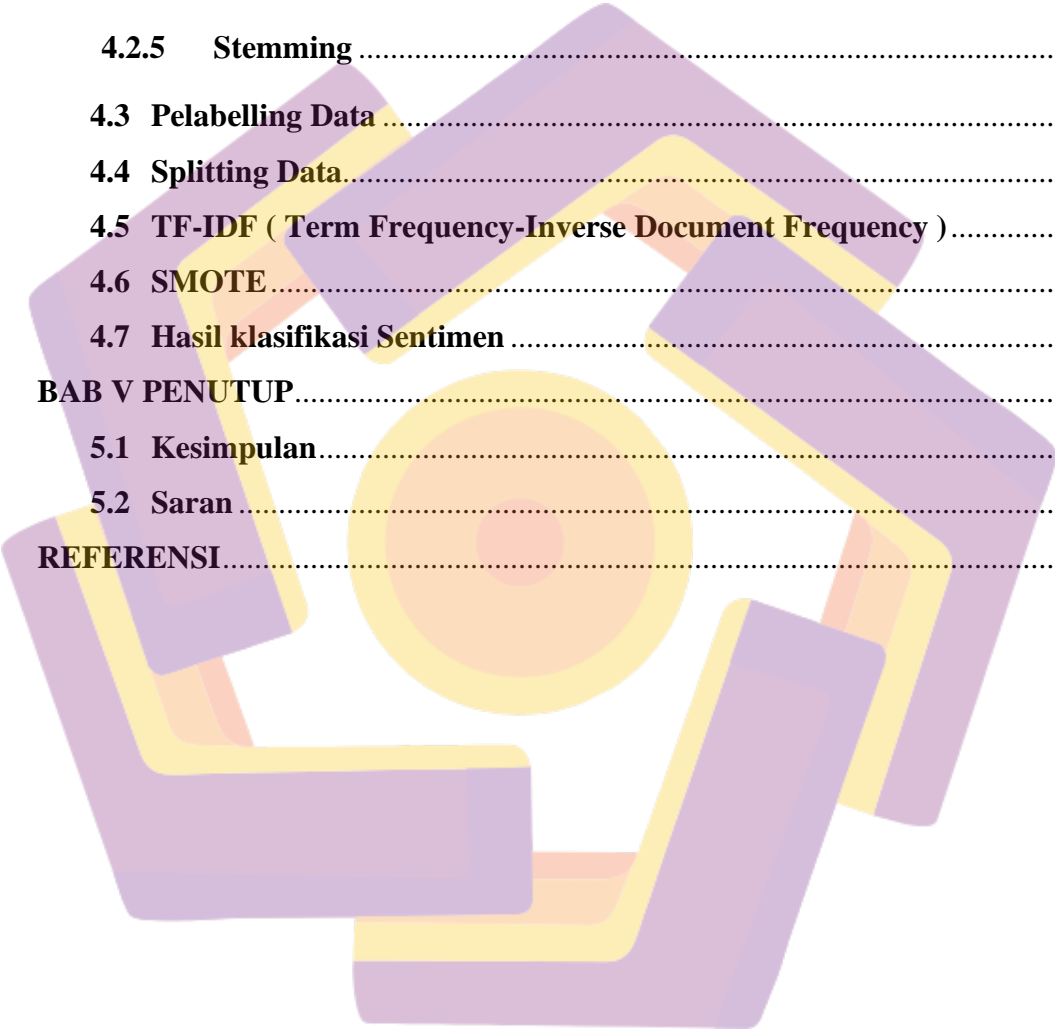




## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>ii</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>iii</b>
<b>HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....</b>	<b>iv</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>v</b>
<b>KATA PENGANTAR.....</b>	<b>vi</b>
<b>DAFTAR ISI.....</b>	<b>viii</b>
<b>DAFTAR TABEL .....</b>	<b>xi</b>
<b>DAFTAR GAMBAR.....</b>	<b>xii</b>
<b>INTISARI .....</b>	<b>xiv</b>
<b>ABSTRACT .....</b>	<b>xv</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
<b>1.1 Latar belakang .....</b>	<b>1</b>
<b>1.2 Rumusan masalah.....</b>	<b>2</b>
<b>1.3 Batasan masalah .....</b>	<b>2</b>
<b>1.4 Tujuan Penelitian.....</b>	<b>3</b>
<b>1.5 Manfaat Penelitian.....</b>	<b>3</b>
<b>1.6 Sistematika Penulisan.....</b>	<b>4</b>
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>5</b>
<b>2.1 Studi Literatur .....</b>	<b>5</b>
<b>2.2 Dasar Teori.....</b>	<b>15</b>
<b>2.2.1 Machine Learning.....</b>	<b>15</b>
<b>2.2.2 Word Cloud .....</b>	<b>15</b>
<b>2.2.3 Text Mining .....</b>	<b>15</b>
<b>2.2.4 Analisis Sentimen .....</b>	<b>16</b>
<b>2.2.5 Teknik Balancing Dataset .....</b>	<b>16</b>
<b>2.2.5.1 SMOTE .....</b>	<b>16</b>

2.2.6	Preprocessing Data .....	17
2.2.7	Lexicon .....	18
2.2.8	Support Vector Machine .....	18
2.2.9	TF-IDF (Term Frequency Inverse Document Frequency) .....	20
2.2.10	Confusion Matrix .....	21
<b>BAB III METODE PENELITIAN .....</b>		<b>24</b>
3.1	Objek Penelitian .....	24
3.2	Alur Penelitian .....	24
3.2.1	Pengumpulan Data .....	26
3.2.2	Preprocessing .....	26
3.2.3	Case Folding .....	26
3.2.4	Cleansing .....	27
3.2.5	Normalization .....	27
3.2.6	Stopword .....	27
3.2.7	Stemming .....	27
3.2.8	Pelabelan Data .....	28
3.2.9	Splitting Data .....	28
3.2.10	TF-IDF .....	28
3.2.11	SMOTE .....	28
3.2.12	Support Vector Machine .....	29
3.2.13	Confusion Matrix .....	30
3.3	Alat dan Bahan .....	30
3.3.1	Alat Penelitian .....	30
3.3.2	Bahan Penelitian .....	30
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>31</b>
4.1	Pengumpulan Data .....	31



<b>4.2 Preprocessing Data</b> .....	33
<b>4.2.1 Case Folding</b> .....	33
<b>4.2.2 Cleansing Data</b> .....	35
<b>4.2.3 Normalization</b> .....	37
<b>4.2.4 Stopword</b> .....	38
<b>4.2.5 Stemming</b> .....	40
<b>4.3 Pelabelling Data</b> .....	41
<b>4.4 Splitting Data</b> .....	43
<b>4.5 TF-IDF ( Term Frequency-Inverse Document Frequency )</b> .....	44
<b>4.6 SMOTE</b> .....	46
<b>4.7 Hasil klasifikasi Sentimen</b> .....	47
<b>BAB V PENUTUP</b> .....	52
<b>5.1 Kesimpulan</b> .....	52
<b>5.2 Saran</b> .....	52
<b>REFERENSI</b> .....	54

## DAFTAR TABEL

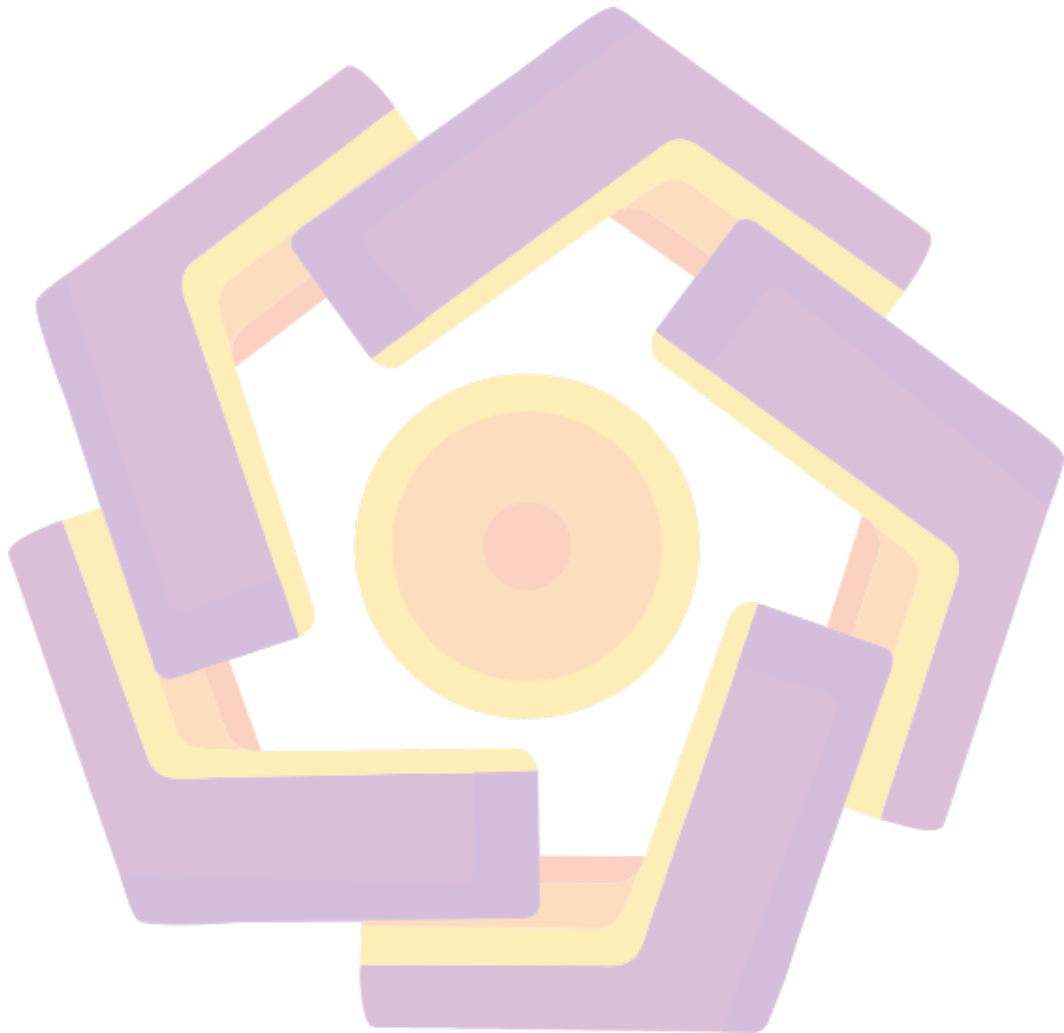
Tabel 2.1 Keaslian Penelitian .....	8
Tabel 2.2 Confusion Matrix .....	21
Tabel 3.1 Tabel Alat Penelitian .....	30
Tabel 4.1 Contoh Dataset Mentah.....	32
Tabel 4.2 Data sesudah dan sebelum Case Folding .....	35
Tabel 4.3 Data sesudah dan sebelum Cleansing Data.....	37
Tabel 4.4 Data sesudah dan sebelum Normalization .....	38
Tabel 4.5 Data sebelum dan sesudah Stopword.....	40
Tabel 4.6 Data sebelum dan sesudah Stemming.....	41
Tabel 4.7 Data sesudah dan sebelum labelling .....	43
Tabel 4.8 Perhitungan Manual TF-IDF .....	45

## DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian .....	25
Gambar 3.2 Kernel Linier .....	29
Gambar 4.1 Scrapping Data Ulasan pada website Apify.....	31
Gambar 4.2 Dataset Mentah.....	32
Gambar 4.3 Alur Preprocessing .....	33
Gambar 4.4 Coding Case Folding.....	34
Gambar 4.5 Data hasil dari Case Folding .....	34
Gambar 4.6 Coding Cleansing Data (1).....	35
Gambar 4.7 Coding Cleansing Data (2).....	36
Gambar 4.8 Data hasil dari Cleansing Data.....	36
Gambar 4.9 Coding Normalization.....	37
Gambar 4.10 Data hasil dari Normalization .....	38
Gambar 4.11 Coding Stopword .....	39
Gambar 4.12 Data hasil dari Stopword.....	39
Gambar 4.13 Coding Stemming.....	40
Gambar 4.14 Data hasil dari Stemming .....	41
Gambar 4.15 Coding Lexicon.....	42
Gambar 4.16 Diagram Labelling Lexicon .....	43
Gambar 4.17 Coding Splitting Data.....	44
Gambar 4.18 Coding TF-IDF .....	45
Gambar 4.19 Visualisasi data Word Cloud.....	46
Gambar 4.20 Jumlah Data sebelum proses SMOTE.....	47
Gambar 4.21 Jumlah Data sesudah proses SMOTE .....	47
Gambar 4. 22 Diagram data SMOTE .....	47
Gambar 4. 23 Hasil Percobaan Pertama.....	48
Gambar 4.24 Hasil Percobaan Kedua .....	49
Gambar 4.25 Hasil Percobaan Ketiga.....	49
Gambar 4.26 Hasil Percobaan Keempat .....	50

Gambar 4.27 Hasil Percobaan Kelima.....50

Gambar 4.28 Hasil Percobaan Keenam .....51



## INTISARI

Dalam era digital seperti sekarang, ulasan dan komentar tentang pengalaman berkunjung ke objek wisata salah satunya TKL Ecopark semakin banyak diungkapkan oleh pengunjung. Ulasan ini dapat menjadi sumber informasi berharga bagi manajemen objek wisata dalam meningkatkan kualitas pelayanan dan mengembangkan objek wisata mereka. Namun, jumlah review yang banyak dan beragam serta sulitnya menentukan keakuratan dari setiap review membuat pengolahan informasi menjadi sulit.

Penelitian ini mengambil data dari review google maps objek wisata TKL Ecopark yang akan diklasifikasikan menggunakan SVM. Sebelum proses klasifikasi dimulai ada alur proses untuk mendapatkan data yang bersih sehingga dapat diklasifikasikan dengan baik, alur proses tersebut yaitu proses pengumpulan data pada review objek wisata TKL Ecopark, lalu data yang sudah di kumpulkan akan dibersihkan melalui preprocessing supaya data mudah untuk diolah dan diklasifikasikan, setelah data melalui tahap preprocessing data akan dilabelling menjadi dua kelas menggunakan lexicon yaitu kelas positif dan negatif, data yang sudah dilabelling tadi lalu akan dibagi menjadi data latih dan data uji secara random dan dengan rasio tertentu, setelah itu data akan diberikan pembobotan kata menggunakan TF-IDF, setelah data melalui tahapan-tahapan tersebut baru akan dimasukan kedalam proses SMOTE yaitu menyeimbangkan data positif dan negatif lalu data akan diklasifikasikan menggunakan Support Vector Machine (SVM), data juga akan di klasifikasikan tanpa menggunakan SMOTE supaya peneliti dapat membandingkan apakah SMOTE dapat menaikkan nilai akurasi SVM setelah itu akurasi akan dihitung kembali akurasinya menggunakan confusion matrix.

Penelitian ini dapat memberikan manfaat kepada calon pengunjung dan pengurus objek wisata TKL Ecopark dalam memantau analisis sentimen terkait data ulasan yang sudah diolah. Selain itu Hasil dari penelitian ini dapat menjadi acuan untuk peneliti selanjutnya dalam pengembangan metode analisis sentimen menggunakan SVM. Hasil pengujian pada proses labeling didapatkan sentimen positif sebesar 70,2% dan sentimen negatif 29,8% berdasarkan review google maps pada TKL Ecopark memiliki tanggapan positif pada rentan 01 Januari 2022 hingga 01 Februari 2023. Pengujian SVM pada penelitian ini dilakukan sebanyak 6 kali percobaan menggunakan SMOTE dan tanpa SMOTE dan menghasilkan tingkat rata-rata akurasi SVM lebih dari 79%. Dari data yang sudah diolah akurasi SVM mengalami kenaikan akurasi pada saat menggunakan SMOTE, hal ini juga dipengaruhi pembagian data sesuai rasio yang sudah ditentukan.

**Kata kunci:** SVM, SMOTE, Analisis Sentimen, Objek Wisata

## ABSTRACT

In today's digital era, reviews and comments about the experience of visiting tourist attractions, including TKL Ecopark, are increasingly being expressed by visitors. These reviews can be a valuable source of information for attraction management in improving service quality and developing their attractions. However, the large number and variety of reviews and the difficulty of determining the accuracy of each review make information processing difficult.

This research takes data from the review of google maps of TKL Ecopark tourist attractions which will be classified using SVM. Before the classification process begins there is a process flow to get clean data so that it can be classified properly, the process flow is the process of collecting data on TKL Ecopark tourist attraction reviews, then the data that has been collected will be cleaned through preprocessing so that the data is easy to process and classify, after the data has gone through the preprocessing stage the data will be labeled into two classes using a lexicon, namely positive and negative classes, the data that has been labeled earlier will then be divided into training data and test data randomly and with a certain ratio, after that the data will be given word weighting using TF-IDF, after the data has gone through these stages, it will be entered into the SMOTE process, namely balancing positive and negative data and then the data will be classified using Support Vector Machine (SVM), the data will also be classified without using SMOTE so that researchers can compare whether SMOTE can increase the accuracy value of SVM after that the accuracy will be recalculated using the confusion matrix.

This research can provide benefits to prospective visitors and administrators of TKL Ecopark tourist attractions in monitoring sentiment analysis related to the processed review data. In addition, the results of this study can be a reference for further researchers in developing sentiment analysis methods using SVM. The test results in the labeling process obtained a positive sentiment of 70.2% and a negative sentiment of 29.8% based on google maps reviews on TKL Ecopark have positive responses in the vulnerable 01 January 2022 to 01 February 2023. SVM testing in this study was carried out 6 times using SMOTE and without SMOTE and resulted in an average SVM accuracy rate of more than 79%. From the data that has been processed, the accuracy of SVM has increased when using SMOTE, this is also influenced by the distribution of data according to a predetermined ratio.

**Keyword:** SVM, SMOTE, Sentiment Analyst, Tourism Objects