

ANALISIS SENTIMEN CUSTOMER REVIEW DI E-COMMERCE TOKOPEDIA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) DAN ADAPTIVE SYNTHETIC SAMPLING (ADASYN)

*Diajukan Untuk Menyusun Skripsi
Di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI*



Oleh :

MUHAMMAD GANDA RIZKY

NIM : 09021382025112

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024**

LEMBAR PENGESAHAN SKRIPSI

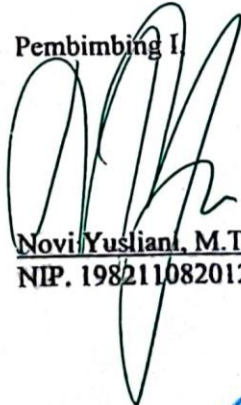
ANALISIS SENTIMEN CUSTOMER REVIEW DI E-COMMERCE TOKOPEDIA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) DAN ADAPTIVE SYNTHETIC SAMPLING (ADASYN)

Oleh :

Muhammad Ganda Rizky
NIM: 09021382025112

Palembang, 21 Juni 2024

Pembimbing I,



Novi Yusliani, M.T.
NIP. 198211082012122001

Pembimbing II,



Anggina Primanita, M.IT, Ph.D
NIP. 198908062015042002

Mengetahui,
Ketua Jurusan Teknik Informatika



Dr. M. Fachrurrozi, S.Si., M.T.

NIP. 198005222008121002

TANDA LULUS UJIAN KOMPREHENSIF

Pada hari Jumat tanggal 14 Juni 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Muhammad Ganda Rizky

NIM : 09021382025112

Judul : Analisis Sentimen Customer Review Di E-Commerce Tokopedia Menggunakan Metode Support Vector Machine (SVM) dan Adaptive Synthetic Sampling (ADASYN)

dan dinyatakan LULUS.

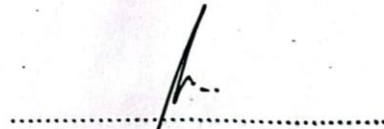
1. Ketua Penguji

Al Farissi, S.Kom., M.Cs
NIP. 198512152014041001



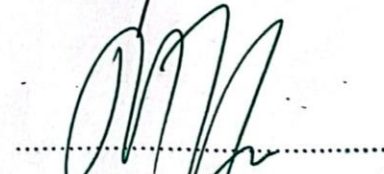
2. Penguji I

Rizki Kurniati, S.Kom., M.T
NIP. 199107122019032016




3. Pembimbing I

Novi Yusliani, M.T.
NIP. 198211082012122001




4. Pembimbing II

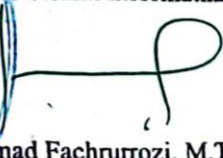
Anggina Primanita, M.IT, Ph.D
NIP. 198908062015042002



Mengetahui,
Ketua Jurusan Teknik Informatika



Dr. Muhammad Fachrurrozi, M.T.
NIP. 198005222008121002



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Muhammad Ganda Rizky

NIM : 09021382025117

Program Studi : Teknik Informatika Bilingual

Judul : Analisis Sentimen Customer Review Di E-Commerce Tokopedia Menggunakan Metode Support Vector Machine (Svm) Dan Adaptive Synthetic Sampling (Adasyn)

Hasil Pengecekan Software *iThenticate/Turnitin*: 17%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari siapapun.



Palembang, 12 Juni 2024



Muhammad Ganda Rizky

NIM. 09021382025112

MOTTO DAN PERSEMBAHAN

“Amun nak kuliah, kuliahlah kudai sampai lulus...amun kaba nak nikah ka kubelika emas 20 suku, jangan pas kaba kuliah mangke nikah tengah jalan” – Nek ibuk

“Pacaklah kaba asak selesai” – Nek Bak

*“And When You Die the only Kingdom you’ll see is 2 ft wide and 6 ft deep”
– Bring Me The Horizon*

ABSTRACT

Sentiment analysis is the process of automatically understanding, extracting, and processing textual data to uncover information about opinions contained within the data, determining whether they are negative or positive. This study aims to perform sentiment analysis and evaluate the performance of algorithms in classifying customer reviews on the Indonesian e-commerce platform, Tokopedia, using the Support Vector Machine (SVM) algorithm and Adaptive Synthetic Sampling (ADASYN). The SVM algorithm was chosen for its ability to handle complex classification problems, while the ADASYN technique was employed to address the imbalance between the majority and minority classes in the dataset, which consists of 2,819 negative reviews and 450 positive reviews. The results show that the combination of the SVM and ADASYN algorithms have the highest accuracy with an accuracy rate of 93.7%, a recall of 76.6%, and an f1-score of 79%.

Keywords : *Adaptive Synthetic Sampling, algorithm, dataset, Sentiment analysis, Support Vector Machine.*

ABSTRAK

Analisis sentimen adalah proses memahami, mengekstraksi, dan mengolah data teks secara otomatis untuk menemukan informasi mengenai opini dalam data tersebut, apakah mengandung sentimen negatif atau positif. Penelitian ini bertujuan untuk melakukan analisis sentimen serta mengevaluasi kinerja algoritma dalam mengklasifikasikan ulasan pelanggan di platform *e-commerce* Indonesia, Tokopedia, dengan menggunakan algoritma *Support Vector Machine* (SVM) dan *Adaptive Synthetic Sampling* (ADASYN). Algoritma SVM dipilih karena kemampuannya dalam menangani masalah klasifikasi pada dataset yang kompleks, sedangkan teknik ADASYN digunakan untuk menangani ketidakseimbangan antara kelas mayoritas dan minoritas dalam dataset yang terdiri dari 2.819 ulasan negatif dan 450 ulasan positif. Hasil penelitian menunjukkan bahwa kombinasi algoritma SVM dan ADASYN memberikan akurasi tertinggi dengan nilai *accuracy* sebesar 93,7%, *recall* sebesar 76,6%, dan *f1-score* sebesar 79%.

Kata Kunci : *Adaptive Synthethic Sampling*, algoritma, Analisis sentimen, *dataset*, *Support Vector Machine*.

KATA PENGANTAR

Segala puji dan syukur atas kehadiran Tuhan Yang Maha Esa atas segala berkat dan Rahmat-Nya sehingga dapat menyelesaikan Skripsi dengan judul “Analisis Sentimen *Customer Review* Di *E-Commerce* Tokopedia Menggunakan Metode *Support Vector Machine* (SVM) Dan *Adaptive Synthetic Sampling* (Adasyn)”. Skripsi ini disusun sebagai salah satu syarat dalam menyelesaikan Pendidikan dan mendapatkan gelar Sarjana Komputer di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Dalam penulisan skripsi ini, penulis mengucapkan rasa syukur kepada Tuhan Yang Maha Esa dan penulis mengucapkan banyak terima kasih kepada :

1. Kepada Papa yiyi, Heri Widijantoro yang selalu memberikan dukungan, dan doa tiada henti. Beliau pada masa mudanya berkuliah di salah satu kampus swasta sambil bekerja untuk memenuhi kebutuhan serta membayar biaya kuliah sendiri, namun beliau mampu memenuhi tanggung jawabnya sebagai kepala keluarga dan memberikan Pendidikan kepada saya hingga jenjang perkuliahan.
2. Pintu surgaku Mama yiyi, Novaridah. Beliau sangat berperan penting dalam menyelesaikan program studi penulis, beliau juga tidak bersal dari universitas negeri namun mampu memberikan pendidikan serta moral yang tegas kepada untuk tetap semangat hingga menyelesaikan proses perkuliahan saya, rasa kasih sayangnya serta rasa cintanya yang selalu menjadi doa untuk kesuksesan anak-anaknya.

3. Kepada adikku yang kusayang dan kubanggakan, Muhammad Brilian Rizky yang telah memberikan doa dan motivasi sampai akhirnya penulis dapat menyelesaikan skripsi ini.
4. Kepada Nenek bak yang tidak sempat melihat cucu-nya wisuda dan Nek ibuk yang telah menyemangati dan sabar menunggu cucu-nya menyelesaikan skripsi ini
5. Prof. Dr. Erwin, S.Si., M.Si., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Dr. M. Fachrurrozi, S.Si., M. T. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Ibu Novi Yusliani, M.T. M. T. dan Bapa Anggina Primanita, M.IT, Ph.D Selaku Dosen Pembimbing yang telah membimbing dan memberi arahan, masukan, saran dan motivasi kepada penulis dalam proses penyelesaian skripsi ini.
8. Mba Rika selaku admin Jurusan Teknik Informatika yang telah membantu mengurus berkas administrasi penulis.
9. Rekan-rekan Infixion Teknik Informatika Angkatan 2020, terutama rekan-rekan Mbah Yono.
10. Rekan-rekan yang membantu proses pembuatan ini mulai dari Jerry, Faiq, Bil Hakki, Fauzan, Don Rio (BUMN), Jidan, Ustad Rama, An Ichasan, Nazifah, Anggi, Daffa.

11. Terimakasih kepada Muhammad Qinthar Fulky Thiasif telah menemani penulis untuk main DOTA 2 selama pengerjaan skripsi ini.

12. Terimakasih Kepada Bring Me The Horizon, Baby Metal, Rainych Ran, Nirvana, Slipknot, The First Take, Rammstein, Ado, The Suicide Silence. Karya mereka telah menemani penulis dalam penulisan skripsi ini.

Penulis menyadari dalam penulisan skripsi ini masih banyak terdapat kekurangan, baik dalam hal isi maupun dalam penulisannya. Oleh karena itu, penulis membutuhkan kritik dan saran yang dapat membangun skripsi yang lebih baik. Penulis juga mengharapkan skripsi ini dapat bermanfaat bagi yang membutuhkan.

Palembang, 03 Juli 2024

Muhammad Ganda Rizky

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN SKRIPSI.....	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vii
ABSTRAK	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	xii
DAFTAR GAMBAR	xvi
DAFTAR TABEL.....	xvii
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah.....	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-5
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-5
1.8 Kesimpulan.....	I-6
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori.....	II-1

2.2.1	Analisis Sentimen.....	II-1
2.2.2	<i>Text Preprocessing</i>	II-2
2.2.3	<i>Review</i>	II-4
2.2.4	<i>Support Vector Machine (SVM)</i>	II-5
2.2.5	<i>Oversampling</i>	II-10
2.2.6	<i>Adaptive Synthetic Sampling (ADASYN)</i>	II-11
2.2.7	TF-IDF.....	II-14
2.2.8	<i>Confusion Matrix</i>	II-15
2.3	Penelitian Lain yang Relevan.....	II-20
2.4	Kesimpulan.....	II-21
BAB III METODOLOGI PENELITIAN.....		III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan data	III-1
3.2.1	Jenis dan Sumber Data	III-1
3.3	Tahapan Penelitian	III-2
3.3.1	Studi Literatur.....	III-3
3.3.2	Pengumpulan Data.....	III-3
3.3.3	Kerangka Kerja.....	III-3
3.3.4	Format Data Pengujian.....	III-5
3.3.7	Hasil dan Analisis.....	III-6
3.4	Metode Pengembangan Perangkat Lunak	III-7
3.4.1	Fase Insepsi	III-7
3.4.2	Fase Elaborasi.....	III-7
3.4.4	Fase Transisi.....	III-8
3.5	Jadwal Penelitian.....	III-8

3.6	Kesimpulan.....	III-10
BAB IV PENGEMBANGAN PERANGKAT LUNAK.....		IV-1
4.1	Pendahuluan	IV-1
4.2	Fase <i>Inception</i>	IV-1
4.2.1	Pemodelan Bisnis	IV-1
4.2.2	Kebutuhan Fungsional.....	IV-2
4.2.3	Analisis dan Desain	IV-3
4.3	Fase Elaborasi	IV-33
4.4	Fase Konstruksi	IV-41
4.4.1	<i>Class Diagram</i>	IV-42
4.4.2	Implementasi	IV-42
4.5	Fase Transisi	IV-44
4.5.1	Pemodelan Bisnis	IV-44
4.5.2	Rencana Pengujian	IV-44
4.5.3	Implementasi	IV-46
4.6	Kesimpulan.....	IV-48
BAB V HASIL PENELITIAN DAN PEMBAHASAN		V-1
5.1	Pendahuluan	V-1
5.2	Data Hasil Penelitian.....	V-1
5.2.1	Konfigurasi Percobaan	V-1
5.2	Analisis Hasil Penelitian	V-5
5.4	Kesimpulan.....	V-11
BAB VI KESIMPULAN DAN SARAN		VI-1
6.1	Pendahuluan	VI-1
6.2	Kesimpulan.....	VI-1

6.3	Saran.....	VI-2
	DAFTAR PUSTAKA	xvii
	LAMPIRAN.....	xxi

DAFTAR GAMBAR

Gambar II-1. Ilustrasi SVM	II-6
Gambar II-2. Ilustrasi <i>Undersampling</i> dan <i>Oversampling</i>	II-11
Gambar II-3. Arsitektur metode	II-19
Gambar III-1. Diagram Tahapan Penelitian.....	III-2
Gambar III-2. Rancangan Kerangka Kerja Penelitian	III-4
Gambar IV- 1. Use Case Diagram	IV-26
Gambar IV- 2. Rancangan Interface Halaman Utama	IV-34
Gambar IV- 3. Activity Diagram Menginput Dataset Customer Review	IV-36
Gambar IV- 4. Activity Diagram Hasil Analisis Sentimen SVM.....	IV-37
Gambar IV- 5. Activity Diagram Hasil Analisis Sentimen SVM & ADASYN	IV-38
.....	IV-38
Gambar IV-6. Sequence Diagram Input Data	IV-39
Gambar IV-7. Sequence Diagram Analisis Sentimen Customer Review Menggunakan SVM	IV-40
Gambar IV-8. Sequence Diagram Analisis Sentimen Customer Review Menggunakan SVM dan ADASYN.....	IV-41
Gambar IV-9. Class Diagram.....	IV-42
Gambar IV-10. User Interface Perangkat Lunak	IV-44
Gambar V-1. Grafik Perbandingan Nilai Accuracy.....	V-6
Gambar V-2. Grafik Perbandingan Nilai Precision	V-7
Gambar V-3. Grafik Perbandingan Nilai Recall	V-8
Gambar V-4. Grafik Perbandingan Nilai F1-Score.....	V-9

DAFTAR TABEL

Tabel II-1. Proses Cleaning.....	II-2
Tabel II-2. Proses Case Folding.....	II-3
Tabel II- 3. Proses Tokenizing	II-3
Tabel II-4. Proses Stemming.....	II-4
Tabel II- 5. Proses Stopwords Removal.....	II-4
Tabel II- 6. Confusion Matrix Dua Kelas	II-16
Tabel III-1. Contoh Dataset Customer Review Tokopedia.....	III-1
Tabel III-2. Format Tabel Pengujian Confusion Matrix	III-5
Tabel III-3. Format Tabel Pengujian.....	III-6
Tabel III-4. Jadwal Rancangan Penelitian	III-8
Tabel IV- 1. Kebutuhan Fungsional.....	IV-2
Tabel IV- 2 Kebutuhan Non-Fungsional.....	IV-3
Tabel IV- 3. Contoh data <i>Customer Review</i>	IV-5
Tabel IV- 4. Hasil proses <i>Cleaning</i>	IV-6
Tabel IV- 5 Hasil proses <i>Case Folding</i>	IV-7
Tabel IV- 6 Hasil Proses <i>Tokenizing</i>	IV-8
Tabel IV- 7 Hasil proses <i>Stemming</i>	IV-9
Tabel IV- 8 Hasil proses <i>Stopword Removal</i>	IV-10
Tabel IV- 9. Hasil perhitungan TF dan IDF.....	IV-11
Tabel IV- 10. Hasil Akhir TF-IDF.....	IV-13
Tabel IV- 11. Hasil Proses Pembentukan KNN Data Minoritas.....	IV-16
Tabel IV- 12. Hasil Proses Pembentukan Data Synthetic.....	IV-20
Tabel IV- 13. Hasil Proses Nilai <i>Dot Product</i>	IV-22
Tabel IV- 14. Hasil Klasifikasi D4 Menggunakan SVM & ADASYN	IV-25
Tabel IV- 15. Tabel Definisi Aktor.....	IV-26
Tabel IV- 16. Tabel Definisi Use Case	IV-27
Tabel IV- 17. Skenario Input Data.....	IV-27
Tabel IV- 18. Analisis Sentimen Customer Review Menggunakan SVM	IV-28

Tabel IV- 19. Analisis Sentimen Customer Review Menggunakan SVM dan ADASYN	IV-31
Tabel IV- 20. Tabel Implementasi Kelas	IV-42
Tabel IV- 21. Rencana Pengujian Use Case Input Data	IV-45
Tabel IV- 22. Rencana Pengujian Use Case Analisis Sentimen Customer Review Menggunakan SVM	IV-45
Tabel IV- 23. Rencana Pengujian Use Case Melakukan Analisis Sentimen Customer Review Menggunakan SVM dan ADASYN	IV-45
Tabel IV- 24. Pengujian Use Case Input Data	IV-46
Tabel IV- 25. Pengujian Use Case Analisis Sentimen Customer Review Menggunakan SVM	IV-46
Tabel IV- 26. Pengujian Use Case Analisis Sentimen Customer Review Menggunakan SVM dan ADASYN	IV-47
Tabel V- 1 Tabel konfigurasi confusion matrix	V-2
Tabel V- 2 Confusion Matrix Klasifikasi SVM dengan Nilai C = 0,1	V-2
Tabel V- 3. Data Hasil Evaluasi Klasifikasi SVM dengan Nilai C = 0,1	V-2
Tabel V- 4. Confusion Matrix Klasifikasi SVM dengan Nilai C = 1	V-3
Tabel V- 5. Data Hasil Evaluasi Klasifikasi SVM dengan Nilai C = 1	V-3
Tabel V- 6. Confusion Matrix Klasifikasi SVM dengan Nilai C = 10	V-3
Tabel V- 7. Data Hasil Evaluasi Klasifikasi SVM dengan Nilai C = 10	V-3
Tabel V- 8. Confusion Matrix Klasifikasi SVM dan ADASYN dengan Nilai C = 0,1	V-4
Tabel V- 9. Data Hasil Evaluasi Klasifikasi SVM dan ADASYN dengan Nilai C = 0,1	V-4
Tabel V- 10. Confusion Matrix Klasifikasi SVM dan ADASYN dengan Nilai C = 1	V-4
Tabel V- 11. Data Hasil Evaluasi Klasifikasi SVM dan ADASYN dengan Nilai C = 1	V-5
Tabel V- 12. Confusion Matrix Klasifikasi SVM dan ADASYN dengan Nilai C = 10	V-5

Tabel V- 13. Data Hasil Evaluasi Klasifikasi SVM dan ADASYN dengan Nilai C = 10	V-5
Tabel V-14. Perbandingan Hasil Nilai Accuracy.....	V-6
Tabel V- 15. Perbandingan Hasil Nilai Precision	V-7
Tabel V- 16. Perbandingan Hasil Nilai Recall.....	V-8
Tabel V- 17. Perbandingan Hasil Nilai F1-Score	V-9
Tabel V- 18. Hasil Evaluasi SVM.....	V-10
Tabel V- 19. Hasil Evaluasi SVM + ADASYN.....	V-11

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini akan membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan. Selain itu, bagian ini juga menyajikan gambaran umum dari keseluruhan kegiatan penelitian yang berjudul “Analisis Sentimen *Customer Review* di *E-commerce* Tokopedia Menggunakan *Support Vector Machine* (SVM) dan *Adaptive Synthetic Sampling* (ADASYN).

1.2 Latar Belakang Masalah

Penggunaan internet dan canggihnya teknologi saat ini telah banyak memberikan perubahan dalam kehidupan bermasyarakat. Segala aktivitas dapat dilakukan secara *mobile/online* dalam keadaan apapun dan dimanapun. Pada Januari 2023 laporan *We Are Social* menunjukkan jumlah pengguna internet di Indonesia telah mencapai 212 juta jiwa yang berarti sekitar 77% dari populasi Indonesia telah menggunakan internet. Jumlah tersebut lebih tinggi 3,85% dibanding tahun lalu yang tercatat sebanyak 205 juta jiwa (Rizaty, 2023). Dengan banyaknya pengguna internet di Indonesia, kini para pebisnis memanfaatkan *e-commerce* sebagai wadah untuk menawarkan produk atau jasanya secara *mobile/online* dan terbukti dapat menjangkau lebih banyak pelanggan. Salah satu yang menjadi bagian dari *e-commerce* adalah *marketplace* (Servanda et al., 2019).

Tokopedia menjadi salah satu *marketplace* yang paling sering dikunjungi di Indonesia (Dihni, 2022). Hal tersebut membuat banyaknya masyarakat

melakukan transaksi jual beli di Tokopedia. Sebelum melakukan proses pembelian keranjang belanja, calon pelanggan biasanya terlebih dahulu melihat *rating* dari produk tersebut. Namun ternyata *rating* yang bagus tidak dapat menjamin toko tersebut mendapatkan *review* yang baik juga., *review* produk dapat membantu calon pelanggan mendapatkan informasi tentang produk dari pelanggan lain yang telah merasakan manfaat dari produk yang dijual, yang pada gilirannya dapat memengaruhi calon pelanggan untuk memutuskan untuk membeli produk atau tidak. (Atimi & Enda, 2022).

Oleh itu karena itu perlu adanya cara untuk melihat apakah *rating* bagus juga banyak yang memberikan *review* yang bagus juga, salah satu algoritma untuk menyelesaikan analisis sentimen pada penelitian ini adalah *support vector machine* (SVM) (Efrizoni et al., 2022). Konsep algoritma SVM dapat disederhanakan sebagai upaya dalam menemukan garis pemisah terbaik antara dua kelas dalam ruang *input*. SVM bertujuan untuk menemukan *hyperplane* pemisah dengan memperbesar jarak antara kedua kelas. (Suyanto, 2017).

Penelitian sebelumnya yang dilakukan oleh Aldisa & Maulana (2022) membandingkan metode *Support Vector Machine*, *Decision Tree*, *Naïve Bayes* untuk analisis sentimen opini masyarakat yang berkaitan dengan vaksinasi *Booster* COVID-19. Hasil penelitian menunjukkan bahwa SVM memiliki akurasi lebih tinggi sebesar 83,33% daripada *Decision Tree* sebesar 79,00% dan *Naive Bayes* sebesar 70,00%.

Ketika melakukan penelitian dengan data, seringkali terjadi ketidakseimbangan atau *imbalance class* pada data kelas yang ingin

diklasifikasikan (Cao et al., 2019). *Imbalance class* terjadi ketika suatu kelas memiliki jumlah data sangat kecil atau besar terhadap kelas lainnya yang menyebabkan penurunan kinerja klasifikasi (Kumar et al., 2021). Klasifikasi pada imbalance class akan cenderung mengabaikan kelas yang memiliki jumlah sampel yang sedikit sehingga dapat berpengaruh buruk terhadap performa dari algoritma klasifikasi (Triyanto & Kusumaningrum, 2017).

Adaptive Synthetic Sampling (ADASYN) adalah teknik *oversampling* yang bertujuan untuk menangani ketidakseimbangan data dalam masalah klasifikasi. Prinsipnya adalah memberikan bobot yang berbeda pada contoh-contoh dari setiap kelas berdasarkan tingkat kesulitan. Kelas yang kurang representatif akan diperkuat dengan sampel sintetis, sehingga menciptakan distribusi sampel yang lebih seimbang antara kelas-kelas yang terlibat. (Li et al., 2021), ADASYN meningkatkan proses pembelajaran dengan dua pendekatan. Pertama, mengurangi bias yang timbul akibat ketidakseimbangan antar kelas. Kedua, dengan menyesuaikan secara adaptif batas keputusan klasifikasi terhadap tingkat kesulitan data.

Beberapa penelitian sebelumnya yang menerapkan metode ADASYN menunjukkan hasil akurasi cenderung meningkat dibandingkan tanpa menggunakan metode ADASYN. Berdasarkan penelitian Chamidah et al., (2020) tentang dampak penerapan ADASYN. Hasil data klasifikasi yang diperoleh tanpa menggunakan metode ADASYN sebesar 84%, sedangkan dengan ADASYN sebesar 91%. Kemudian juga mengacu pada penelitian Magnolia et al., (2022) tentang pengelolaan dataset yang tidak seimbang untuk mengklasifikasikan

komentar program kampus merdeka di aplikasi Twitter. Hasil akurasi yang diperoleh dari data tanpa menggunakan metode ADASYN sebesar 70,2%, sedangkan dengan ADASYN sebesar 89,5%.

Berdasarkan uraian tersebut, ADASYN mampu mengatasi ketidakseimbangan data dan meningkatkan kinerja klasifikasi data maka penelitian ini akan menggunakan teknik *oversampling* ADASYN dan melakukan klasifikasi menggunakan SVM.

1.3 Rumusan Masalah

Berdasarkan uraian dari latar belakang maka rumusan masalah dari penelitian ini adalah sebagai berikut :

1. Bagaimana menerapkan metode *Support Vector Machine* dan *Adaptive Synthetic Sampling* dalam mengklasifikasikan analisis sentimen pada *customer review* di *e-commerce* Tokopedia?
2. Bagaimana kinerja analisis sentimen *customer review* di Tokopedia menggunakan *Support Vector Machine* dan *Adaptive Synthetic Oversampling*

1.4 Tujuan Penelitian

Berikut tujuan penelitian ini:

1. Menghasilkan model analisis sentimen *customer review* pada *e-commerce* Tokopedia menggunakan *Support Vector Machine* (SVM) dan *Adaptive Synthetic Oversampling* (ADASYN).

2. Mengetahui kinerja *accuracy*, *precision*, *recall*, dan *f1-score* pada analisis sentimen *customer review* di Tokopedia menggunakan *Support Vector Machine* (SVM) dan *Adaptive Synthetic Oversampling* (ADASYN).

1.5 Manfaat Penelitian

Berikut manfaat penelitian ini :

1. Memahami penggunaan metode SVM yang dipadukan dengan teknik ADASYN untuk menangani ketidakseimbangan data.
2. Hasil penelitian ini dapat digunakan sebagai acuan bagi peneliti lain dalam mengatasi ketidakseimbangan data dengan memanfaatkan metode ADASYN.

1.6 Batasan Masalah

Berikut batasan masalah dari penelitian ini:

1. Penggunaan dataset yang berbahasa Indonesia.
2. Analisis sentimen dibagi menjadi dua kategori: *positive* (0) dan *negative* (1).

1.7 Sistematika Penulisan

Agar penyusunan dan pemahaman mengenai penelitian ini lebih mudah, setiap bab dalam laporan akan diatur sesuai dengan struktur penulisan berikut:

BAB I PENDAHULUAN

Bab ini peneliti menjelaskan latar belakang dari penelitian ini, rumusan masalah, tujuan, manfaat penelitian, batasan masalah, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini mencakup prinsip dasar atau teori yang diterapkan pada penelitian ini serta juga literatur yang diperoleh dari penelitian yang akan dilakukan.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan setiap langkah yang diambil dalam penelitian ini dijelaskan secara rinci dengan menggunakan kerangka kerja.

BAB IV PENGEMBANGAN PERANGKAT LUNAK

Bab ini membahas pengembangan perangkat lunak dan pembangunan perangkat lunak yang didasarkan pada *Rational Unified Process* (RUP).

BAB V HASIL DAN ANALISIS PENELITIAN

Bab ini menampilkan hasil dari pengembangan perangkat lunak yang telah diselesaikan, serta menjelaskan hasil dari pengujian model yang telah dievaluasi.

BAB VI KESIMPULAN DAN SARAN

Bab ini menjabarkan ringkasan dari penelitian dan saran untuk pengembangan penelitian di masa mendatang.

1.8 Kesimpulan

Berdasarkan latar belakang penelitian yang telah diuraikan maka penelitian ini akan menguji pengaruh *Adaptive Synthetic Sampling* (ADASYN) pada analisis sentimen *customer review* di *e-commerce* tokopedia menggunakan algoritma SVM.

BAB II

KAJIAN LITERATUR

2.1 Pendahuluan

Bab ini menjelaskan teori-teori yang mendukung topik penelitian ini dan memberikan contoh gambaran menyeluruh tentang permasalahan, metode, dan teknik yang terkait dengan penelitian ini.

2.2 Landasan Teori

2.2.1 Analisis Sentimen

Analisis sentimen adalah gabungan antara data mining dan text mining, yang merupakan suatu metode untuk mengelola pendapat-pendapat yang diberikan oleh konsumen melalui berbagai saluran, terkait dengan suatu produk, layanan, atau instansi. (Sipayung et al., 2016). Analisis sentimen adalah tahapan dimana pendapat, sentimen, evaluasi, dan emosi individu terhadap topik tertentu yang diekspresikan secara tertulis diekstraksi menggunakan metode pemrosesan bahasa alami (Idris et al., 2023).

Proses analisis sentimen terbagi menjadi beberapa tahap. Menurut penelitian dari Magnolia et al., (2022) Analisis sentimen melibatkan tiga tahapan komponen inti. Tahap awal adalah *preprocessing* data, yang bertujuan untuk merapikan data agar lebih terstruktur, memudahkan proses analisis. Tahap berikutnya adalah pemberian bobot pada kata-kata menggunakan metode TF-IDF. Tahap terakhir adalah klasifikasi data menggunakan algoritma *machine learning*.

DAFTAR PUSTAKA

- Aldisa, R. T., & Maulana, P. (2022). Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM. *Building of Informatics, Technology and Science (BITS)*, 4(1), 106–109.
- Atimi, R. L., & Enda, E. P. (2022). Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia. *Jurnal Sains Dan Informatika*, 8(1), 88–96. <https://doi.org/10.34128/jsi.v8i1.419>
- Badr, W. (n. d.). (2019, February 22). *Badr, W. (n.d.)*. <https://Towardsdatascience.Com/Havingan-Imbalanced-Dataset-Here-Is-Howyou-Can-Solve-It-1640568947eb> .
- Cao, K., Wei, C., Gaidon, A., Arechiga, N., & Ma, T. (2019). Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in Neural Information Processing Systems*, 32.
- Chamidah, N., Santoni, M. M., & Matondang, N. (2020). The effect of oversampling on the classification of hypertension with the naive bayes algorithm, decision tree, and artificial neural network (ann). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(4), 635–641.
- Dihni, V. A. (2022, April 12). *Situs E-Commerce dengan Jumlah Pengunjung Terbanyak di Indonesia (2021)*. Databoks.Katadata.Co.Id.
- Efrizoni, L., Defit, S., Tajuddin, M., & Anggrawan, A. (2022). Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(3), 653–666. <https://doi.org/10.30812/matrik.v21i3.1851>
- Farki, A., & Baihaqi, I. (2016). Pengaruh online customer review dan rating terhadap kepercayaan dan minat pembelian pada online marketplace di Indonesia. *Jurnal Teknik ITS*, 5(2), A614–A619.

- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan metode naïve bayes dan support vector machine pada analisis sentimen twitter. *SMATIKA Jurnal: STIKI Informatika Jurnal*, 10(02), 71–76.
- Hananti, H., & Sari, K. (2021). Perbandingan Metode Support Vector Machine (SVM) dan Artificial Neural Network (ANN) pada Klasifikasi Gizi Balita. *Seminar Nasional Official Statistics*, 2021(1), 1036–1043.
- Hariyanto, H. T., & Trisunarno, L. (2021). Analisis pengaruh online customer review, online customer rating, dan star seller terhadap kepercayaan pelanggan hingga keputusan pembelian pada toko online di Shopee. *Jurnal Teknik ITS*, 9(2), A234–A239.
- He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 1322–1328.
- Idris, I. S. K., Mustofa, Y. A., & Salihi, I. A. (2023). Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM). *Jambura Journal of Electrical and Electronics Engineering*, 5(1), 32–35.
- Kumar, P., Bhatnagar, R., Gaur, K., & Bhatnagar, A. (2021). Classification of imbalanced data: review of methods and applications. *IOP Conference Series: Materials Science and Engineering*, 1099(1), 012077.
- Li, Y., Xu, W., Li, W., Li, A., & Liu, Z. (2021). Research on hybrid intrusion detection method based on the ADASYN and ID3 algorithms. *Math. Biosci. Eng*, 19(2), 2030–2042.
- Magnolia, C., Nurhopipah, A., & Kusuma, B. A. (2022). Penanganan Imbalanced Dataset untuk Klasifikasi Komentar Program Kampus Merdeka Pada Aplikasi Twitter. *Edu Komputika Journal*, 9(2), 105–113.
- Melita, R., Amrizal, V., Suseno, H. B., & Dirjam, T. (n.d.). *PENERAPAN METODE TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DAN COSINE SIMILARITY PADA SISTEM TEMU KEMBALI*

*INFORMASI UNTUK MENGETAHUI SYARAH HADITS BERBASIS WEB
(STUDI KASUS: SYARAH UMDATIL AHKAM).*

- Mualim, E. P. (2022). *PENDEKATAN OVERSAMPLING SMOTE UNTUK IMBALANCED DATASET AKSARA LAMPUNG DAN KLASIFIKASI MENGGUNAKAN SVM.*
- Mutawalli, L., Zaen, M. T. A., & Bagye, W. (2019). Klasifikasi Teks Sosial Media Twitter Menggunakan Support Vector Machine (Studi Kasus Penusukan Wiranto). *Jurnal Informatika Dan Rekayasa Elektronik*, 2(2), 43–51.
- Nuraliza, H., Pratiwi, O. N., & Hamami, F. (2022). Analisis Sentimen IMDb Film Review Dataset Menggunakan Support Vector Machine (SVM) dan Seleksi Feature Importance. *Jurnal Mirai Management*, 7(1), 1–17.
- Perwitasari, R., Afawani, R., & Anjarwani, S. E. (2020). Penerapan Metode Rational Unified Process (RUP) Dalam Pengembangan Sistem Informasi Medical Check Up Pada Citra Medical Centre. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTIKA)*, 2(1), 76–88.
- Ramadhan, N. G. (2021). Comparative Analysis of ADASYN-SVM and SMOTE-SVM Methods on the Detection of Type 2 Diabetes Mellitus. *Sci. J. Informatics*, 8(2), 276–282.
- Rizaty, M. A. (2023, February 3). *Pengguna Internet di Indonesia sentuh 212 juta pada 2023*. DataIndonesia.Id.
- Servanda, I. R. S., Sari, P. R. K., & Ananda, N. A. (2019). Peran Ulasan Produk Dan Fot Produk Yang Ditampilkan Penjual Pada Marketplace Shopee Terhadap Minat Beli Pria Dan Wanita. *Jurnal Manajemen Dan Bisnis*, 2(2), 69–79.
- Shevira, S., Suarjaya, I., & Buana, P. W. (2022). Pengaruh Kombinasi dan Urutan Pre-Processing pada Tweets Bahasa Indonesia. *JITTER-Jurnal Ilmiah Teknologi Dan Komputer*, 3(2).
- Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. *JSI: Jurnal Sistem Informasi (E-Journal)*, 8(1).

- Sitohang, E. H., Setiabudi, D. H., & Ananda, S. A. (2021). Penerapan Modified ADASYN untuk Meningkatkan Akurasi Pendeteksian Pola Fraud pada Transaksi Kartu Kredit. *Jurnal Infra*, 9(2), 84–90.
- Suyanto, D. (2017). Data Mining untuk klasifikasi dan klusterisasi data. *Bandung: Informatika Bandung*.
- Triyanto, A. Y., & Kusumaningrum, R. (2017). Implementasi Teknik Sampling untuk Mengatasi Imbalanced Data pada Penentuan Status Gizi Balita dengan Menggunakan Learning Vector Quantization. *JURNAL IPTEKKOM (Jurnal Ilmu Pengetahuan & Teknologi Informasi)*, 19(1), 39–50.
- Wu, S., & Gao, Y. (2022). Happy or grumpy? A Machine Learning Approach to Analyze the Sentiment of Airline Passengers' Tweets. *ArXiv Preprint ArXiv:2209.14363*.
- Yulietha, I., Faraby, S., & Adiwijaya, A. (2017). Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine. *EProceedings of Engineering*, 4(3).