

**KLASIFIKASI IKLAN LOWONGAN KERJA
MENGUNAKAN *SUPPORT VECTOR MACHINE (SVM)* DAN
*BORDERLINE SYNTHETIC MINORITY OVER-SAMPLING
TECHNIQUE (BORDERLINE-SMOTE)***

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh :

JERRY BERLIN
NIM: 09021282025036

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

LEMBAR PENGESAHAN SKRIPSI

**KLASIFIKASI IKLAN LOWONGAN KERJA MENGGUNAKAN
SUPPORT VECTOR MACHINE (SVM) DAN BORDERLINE SYNTHETIC
MINORITY OVER-SAMPLING TECHNIQUE (BORDERLINE-SMOTE)**

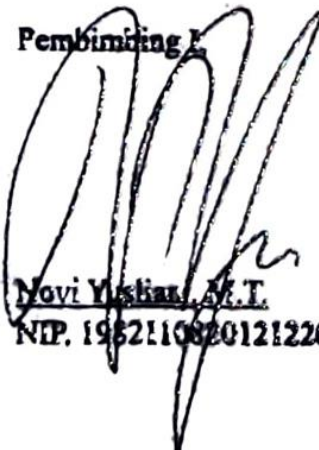
Oleh:

Jerry Berlin

NIM: 09021282025036


Palembang, 22 April 2024

Pembimbing I



Novi Yustian, M.T.
NIP. 196211082012122001

Pembimbing II



Desty Rodiah, M.T.
NIP. 198912212020122011

Mengetahui,

Ketua Jurusan Teknik Informatika



Dr. Muhammad Fachrurrozi, M.T.

NIP. 198005222008121002

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Jumat tanggal 22 Maret 2024 telah dilaksanakan ujian siding skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Jerry Berlin

NIM : 09021282025036

Judul : *Klasifikasi Iklan Lowongan Kerja Menggunakan Support Vector Machine (SVM) dan Borderline Synthetic Minority Over-Sampling Technique (Borderline-SMOTE)*

1. Ketua Penguji

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



2. Penguji I

Rizki Kurniati, M.T.
NIP. 199107122019032016



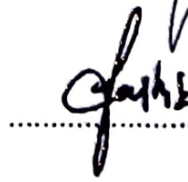
3. Pembimbing I

Novi Yusliani, M.T.
NIP. 198211082012122001



4. Pembimbing II

Desty Rodiah, M.T.
NIP. 198912212020122011



Mengetahui,

Ketua Jurusan Teknik Informatika



Dr. Muhammad Fachrurrozi, M.T.

NIP. 198005222008121002



HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Jerry Berlin

NIM : 09021282025036

Program Studi : Teknik Informatika Bilingual

Judul : Klasifikasi Iklan Lowongan Kerja Menggunakan *Support Vector Machine (SVM)* dan *Borderline Synthetic Minority Over-Sampling Technique (Borderline-SMOTE)*

Hasil Pengecekan Software iThenticate/Turnitin : 17%

Menyatakan bahwa laporan skripsi saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan skripsi ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

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



Palembang, 3 Mei 2024



Jerry Berlin

NIM. 09021282025036

MOTTO DAN PERSEMBAHAN

“Don't cry because it's over. Smile because it's happened.” – Dr. Seuss

Kupersembahkan karya tulis ini kepada:

- Tuhan Yang Maha Esa
- Keluarga
- Sahabat
- Universitas Sriwijaya

ABSTRACT

The growth of the internet has made it easier to recruit workers through the publication of online job advertisements. However, this convenience also brings the risk of fraud in job advertisements that can harm both job seekers and companies. To overcome this problem, classification of job advertisements is required. One of the main challenges in text classification, especially in job advertisement data, is the significant imbalance of data between the majority and minority classes, with the majority class reaching 3500 data and the minority class only 500 data. This research aims to classify job advertisements using Support Vector Machine (SVM) and Borderline Synthetic Minority Over-Sampling Technique (Borderline-SMOTE) to overcome data imbalance. Tests were conducted to see the effect of job advertisements classification performance using SVM without Borderline-SMOTE and SVM with Borderline-SMOTE, where the parameter C used was 0.1, 0.25, 0.50, 0.75, and 1. The results showed that SVM with Borderline-SMOTE had better performance especially in recall and f -measure. In particular, at setting parameter $C = 1$ in SVM with Borderline-SMOTE, the most optimal results were obtained, with a precision value of 0.92, recall of 0.80, and f -measure of 0.86.

Keywords: *Borderline Synthetic Minority Over-Sampling Technique, Data Imbalance, Support Vector Machine, Text Classification*

ABSTRAK

Pertumbuhan internet mempermudah proses perekrutan tenaga kerja melalui publikasi iklan lowongan kerja secara online. Namun, kemudahan ini juga membawa risiko penipuan dalam iklan lowongan kerja yang dapat merugikan baik bagi pencari kerja maupun perusahaan. Untuk mengatasi permasalahan ini, diperlukan klasifikasi terhadap iklan lowongan kerja. Salah satu tantangan utama dalam klasifikasi teks, khususnya pada data iklan lowongan kerja, adalah ketidakseimbangan data yang signifikan antara kelas mayoritas dan minoritas, dengan kelas mayoritas mencapai 3500 data dan kelas minoritas hanya 500 data. Penelitian ini bertujuan untuk melakukan klasifikasi terhadap iklan lowongan kerja menggunakan *Support Vector Machine* (SVM) dan *Borderline Synthetic Minority Over-sampling Technique* (*Borderline-SMOTE*) untuk mengatasi ketidakseimbangan data. Pengujian dilakukan untuk melihat pengaruh kinerja klasifikasi iklan lowongan kerja dengan menggunakan SVM tanpa *Borderline-SMOTE* dan SVM dengan *Borderline-SMOTE*, di mana parameter C yang digunakan yaitu 0.1, 0.25, 0.50, 0.75, dan 1. Hasil penelitian menunjukkan bahwa SVM dengan *Borderline-SMOTE* memiliki kinerja yang lebih baik terutama dalam *recall* dan *F-measure*. Secara khusus, pada pengaturan parameter $C=1$ pada SVM dengan *Borderline-SMOTE*, diperoleh hasil yang paling optimal, dengan nilai *precision* sebesar 0.92, *recall* sebesar 0.80, dan *F-measure* sebesar 0.86.

Kata Kunci: *Borderline Synthetic Minority Over-sampling Technique*, Ketidakseimbangan Data, Klasifikasi Teks, *Support Vector Machine*.

KATA PENGANTAR

Puji dan syukur penulis ucapkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya penulis dapat menyelesaikan Tugas Akhir dengan judul “**Klasifikasi Iklan Lowongan Kerja Menggunakan *Support Vector Machine (SVM)* dan *Borderline Synthetic Minority Over-Sampling Technique (Borderline-SMOTE)***”. Penyusunan Tugas Akhir ini untuk menyelesaikan program Strata-1 Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun tidak langsung. Untuk itu penulis mengucapkan terima kasih setulus-tulusnya kepada:

1. Orang Tua saya atas doa dan segala dukungannya.
2. Ketiga Kakak saya atas doa dan segala dukungannya.
3. Bapak Dr. Muhammad Fachrurrozi, M.T. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Ibu Novi Yusliani, M.T. selaku Dosen Pembimbing I yang telah memberikan bimbingan, kritik, serta saran sehingga penulis dapat menyelesaikan Tugas Akhir.
5. Ibu Desty Rodiah, M.T. selaku Dosen Pembimbing II yang telah memberikan bimbingan, kritik, serta saran sehingga penulis dapat menyelesaikan Tugas Akhir.

6. Seluruh Dosen Jurusan Teknik Informatika dan Dosen Fakultas Ilmu Komputer yang telah membagikan ilmu dan pengetahuannya selama masa perkuliahan.
7. Seluruh Staf Fakultas Ilmu Komputer yang telah membantu dalam urusan administrasi dan akademik selama perkuliahan.
8. Teman-teman kelas Bilingual A, yang sudah kebersamai penulis selama masa perkuliahan dan kebersamai dalam belajar serta berbagi informasi terkait akademisi di Universitas Sriwijaya.
9. Teman-teman Angkatan ke-16 Teknik Informatika, yang telah kebersamai selama perkuliahan di Universitas Sriwijaya.
10. Semua pihak lainnya yang selalu mendukung dan membantu selama proses penyusunan Tugas Akhir ini yang tidak bisa disebutkan satu-persatu.

Penulis menyadari banyak kekurangan dalam penelitian Tugas Akhir ini. Oleh karena itu, semua kiritik dan saran yang membangun sangat dibutuhkan agar Tugas Akhir ini dapat menjadi karya tulis yang sempurna dan dapat memberikan manfaat bagi banyak orang, serta dapat digunakan dalam pengembangan berkelanjutan di masa depan.

Palembang, 29 April 2024



Jerry Berlin

NIM. 09021282025036

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN SIDANG SKRIPSI.....	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	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-5
1.5 Manfaat Penelitian	I-5
1.6 Batasan Masalah.....	I-6
1.7 Sistematika Penulisan	I-6
1.8 Kesimpulan	I-7
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 <i>Text Classification</i>	II-1

2.2.2	<i>Text Preprocessing</i>	II-3
2.2.3	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	II-5
2.2.4	Ketidakseimbangan Data	II-6
2.2.5	<i>Borderline Synthetic Minority Over-sampling Technique (Borderline-SMOTE)</i>	II-7
2.2.6	<i>Support Vector Machine (SVM)</i>	II-10
2.2.7	<i>Confusion Matrix</i>	II-13
2.2.8	<i>Rational Unified Process</i>	II-15
2.3	Penelitian Lain yang Relevan.....	II-17
2.4	Kesimpulan	II-18
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-2
3.3.2	Pengumpulan Data	III-3
3.3.3	Kerangka Kerja	III-3
3.3.4	Format Data Pengujian.....	III-4
3.3.5	Alat Bantu Penelitian	III-4
3.3.6	Pengujian Penelitian.....	III-5
3.3.7	Hasil dan Analisis.....	III-5
3.4	Metode Pengembangan Perangkat Lunak.....	III-6
3.4.1	Insepsi	III-6
3.4.2	Elaborasi.....	III-6

3.4.3	Konstruksi	III-7
3.4.4	Transisi	III-7
3.5	Manajemen Proyek Penelitian	III-7
3.6	Kesimpulan	III-9
BAB IV	PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1
4.2.1	Pemodelan Bisnis	IV-1
4.2.2	Kebutuhan Sistem	IV-2
4.2.3	Analisis dan Desain.....	IV-3
4.3	Fase Elaborasi	IV-32
4.3.1	Pemodelan Bisnis	IV-32
4.3.2	Perancangan Data.....	IV-32
4.3.3	Perancangan <i>User Interface</i>	IV-33
4.3.4	Kebutuhan Sistem	IV-33
4.3.5	<i>Activity Diagram</i>	IV-34
4.3.6	<i>Sequence Diagram</i>	IV-36
4.4	Fase Konstruksi.....	IV-38
4.4.1	<i>Class Diagram</i>	IV-39
4.4.2	Implementasi	IV-39
4.5	Fase Transisi.....	IV-41
4.5.1	Pemodelan Bisnis	IV-41
4.5.2	Rencana Pengujian	IV-41
4.5.3	Implementasi	IV-43
4.6	Kesimpulan	IV-46

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
5.2.2	Analisis Hasil Penelitian	V-8
5.3	Kesimpulan	V-10
BAB VI	KESIMPULAN DAN SARAN.....	VI-1
6.1	Pendahuluan	VI-1
6.2	Kesimpulan	VI-1
6.3	Saran.....	VI-2
DAFTAR PUSTAKA	xix
LAMPIRAN	xxiv

DAFTAR TABEL

	Halaman
Tabel II-1. Proses <i>Cleaning</i>	II-3
Tabel II-2. Proses <i>Case Folding</i>	II-4
Tabel II-3. Proses <i>Tokenization</i>	II-4
Tabel II-4. Proses <i>Stopwords Removal</i>	II-4
Tabel II-5. Proses <i>Lemmatization</i>	II-5
Tabel II-6. <i>Confusion Matrix</i> Dua Kelas.....	II-14
Tabel III-1. Contoh <i>Dataset</i> Iklan Lowongan Kerja	III-1
Tabel III-2. Format Tabel Pengujian <i>Confusion Matrix</i>	III-4
Tabel III-3. Format Tabel Hasil 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-2
Tabel IV-3. Atribut <i>Dataset</i>	IV-4
Tabel IV-4. Contoh Data Iklan Lowongan Kerja	IV-9
Tabel IV-5. Hasil Proses <i>Cleaning</i>	IV-9
Tabel IV-6. Hasil Proses <i>Case Folding</i>	IV-10
Tabel IV-7. Hasil Proses <i>Tokenization</i>	IV-11
Tabel IV-8. Hasil Proses <i>Stopwords Removal</i>	IV-12
Tabel IV-9. Hasil Proses <i>Lemmatization</i>	IV-13
Tabel IV-10. Hasil Perhitungan TF dan IDF	IV-14
Tabel IV-11. Tabel Hasil Akhir TF-IDF	IV-16
Tabel IV-12. Data Ilustrasi Perhitungan Jarak <i>Euclidean</i>	IV-19
Tabel IV-13. Data Ilustrasi m Tetangga Terdekat	IV-20
Tabel IV-14. Data Ilustrasi k Tetangga Terdekat	IV-21
Tabel IV-15. Ilustrasi Sampel/Data Sintetis	IV-22
Tabel IV-16. Data Latih SVM	IV-23
Tabel IV-17. Definisi Aktor.....	IV-26
Tabel IV-18. Definisi <i>Use Case</i>	IV-27

Tabel IV-19. Skenario Memasukkan <i>Dataset</i>	IV-27
Tabel IV-20. Skenario Klasifikasi Menggunakan Metode SVM	IV-28
Tabel IV-21. Skenario Klasifikasi Menggunakan Metode SVM dengan <i>Borderline-SMOTE</i>	IV-30
Tabel IV-22. Tabel Implementasi Kelas	IV-39
Tabel IV-23. Rencana Pengujian <i>Use Case</i> Memasukkan <i>Dataset</i>	IV-41
Tabel IV-24. Rencana Pengujian <i>Use Case</i> Klasifikasi Menggunakan Metode SVM	IV-42
Tabel IV-25. Rencana Pengujian <i>Use Case</i> Klasifikasi Menggunakan Metode SVM dan <i>Borderline-SMOTE</i>	IV-42
Tabel IV-26. Pengujian <i>Use Case</i> Klasifikasi Menggunakan Metode SVM ...	IV-43
Tabel IV-27. Pengujian <i>Use Case</i> Klasifikasi Menggunakan Metode SVM ...	IV-44
Tabel IV-28. Pengujian <i>Use Case</i> Klasifikasi Menggunakan Metode SVM dan <i>Borderline-SMOTE</i>	IV-45
Tabel V-1. <i>Confusion Matrix</i> Klasifikasi SVM dengan Nilai $C = 0,1$	V-2
Tabel V-2. Data Hasil Evaluasi Klasifikasi SVM dengan Nilai $C = 0,1$	V-2
Tabel V-3. <i>Confusion Matrix</i> Klasifikasi SVM dengan Nilai $C = 0,25$	V-2
Tabel V-4. Data Hasil Evaluasi Klasifikasi SVM dengan Nilai $C = 0,25$	V-3
Tabel V-5. <i>Confusion Matrix</i> Klasifikasi SVM dengan Nilai $C = 0,5$	V-3
Tabel V-6. Data Hasil Evaluasi Klasifikasi SVM dengan Nilai $C = 0,5$	V-3
Tabel V-7. <i>Confusion Matrix</i> Klasifikasi SVM dengan Nilai $C = 0,75$	V-4
Tabel V-8. Data Hasil Evaluasi Klasifikasi SVM dengan Nilai $C = 0,75$	V-4
Tabel V-9. <i>Confusion Matrix</i> Klasifikasi SVM dengan Nilai $C = 1$	V-4
Tabel V-10. Data Hasil Evaluasi Klasifikasi SVM dengan Nilai $C = 1$	V-4
Tabel V-11. <i>Confusion Matrix</i> Klasifikasi SVM dengan <i>Borderline-SMOTE</i> dengan Nilai $C = 0,1$	V-5
Tabel V-12. Data Hasil Evaluasi Klasifikasi SVM dengan <i>Borderline-SMOTE</i> dengan Nilai $C = 0,1$	V-5
Tabel V-13. <i>Confusion Matrix</i> Klasifikasi SVM dengan <i>Borderline-SMOTE</i> dengan Nilai $C = 0,25$	V-6

Tabel V-14. Data Hasil Evaluasi Klasifikasi SVM dengan <i>Borderline</i> -SMOTE dengan Nilai C = 0,25	V-6
Tabel V-15. <i>Confusion Matrix</i> Klasifikasi SVM dengan <i>Borderline</i> -SMOTE dengan Nilai C = 0,5	V-6
Tabel V-16. Data Hasil Evaluasi Klasifikasi SVM dengan <i>Borderline</i> -SMOTE dengan Nilai C = 0,5	V-7
Tabel V-17. <i>Confusion Matrix</i> Klasifikasi SVM dengan <i>Borderline</i> -SMOTE dengan Nilai C = 0,75	V-7
Tabel V-18. Data Hasil Evaluasi Klasifikasi SVM dengan <i>Borderline</i> -SMOTE dengan Nilai C = 0,75	V-7
Tabel V-19. <i>Confusion Matrix</i> Klasifikasi SVM dengan <i>Borderline</i> -SMOTE dengan Nilai C = 1	V-8
Tabel V-20. Data Hasil Evaluasi Klasifikasi SVM dengan <i>Borderline</i> -SMOTE dengan Nilai C = 1	V-8

DAFTAR GAMBAR

	Halaman
Gambar II-1. Strategi Umum Untuk Klasifikasi Teks (Dalal & Zaveri, 2011)...	II-2
Gambar II-2. Contoh Distribusi Sampel (Kummer dkk., 2022)	II-8
Gambar II-3. Ilustrasi <i>Borderline</i> Data (Han dkk., 2005).....	II-9
Gambar II-4. Ilustrasi <i>Borderline</i> -SMOTE (Han dkk., 2005).....	II-10
Gambar II-5. Ilustrasi SVM Linier (Jwo dkk., 2021)	II-10
Gambar II-6. Arsitektur RUP (Perwitasari dkk., 2020).....	II-15
Gambar III-1. Diagram Tahapan Penelitian	III-2
Gambar III-2. Rancangan Kerangka Kerja Penelitian	III-3
Gambar IV-1. Perbandingan Iklan Lowongan Pekerjaan Asli vs Palsu	IV-5
Gambar IV-2. Grafik <i>Telecommuting, Has Company Logo, Has Questions</i>	IV-6
Gambar IV-3. Grafik <i>Employment Type</i>	IV-6
Gambar IV-4. Grafik <i>Required Experience</i>	IV-7
Gambar IV-5. Grafik <i>Required Education</i>	IV-7
Gambar IV-6. Grafik Jumlah Kata dalam Teks	IV-8
Gambar IV-7. <i>Use Case Diagram</i>	IV-26
Gambar IV-8. Rancangan <i>User Interface</i> Perangkat Lunak	IV-33
Gambar IV-9. <i>Activity Diagram</i> Memasukkan <i>Dataset</i>	IV-34
Gambar IV-10. <i>Activity Diagram</i> Klasifikasi Menggunakan Metode SVM....	IV-35
Gambar IV-11. <i>Activity Diagram</i> Klasifikasi Menggunakan Metode SVM dan <i>Borderline</i> -SMOTE.....	IV-35
Gambar IV-12. <i>Sequence Diagram</i> Memasukkan <i>Dataset</i>	IV-36
Gambar IV-13. <i>Sequence Diagram</i> Klasifikasi Menggunakan Metode SVM. IV-37	IV-37
Gambar IV-14. <i>Sequeunce Diagram</i> Klasifikasi Menggunakan Metode SVM dan <i>Borderline</i> -SMOTE.....	IV-38
Gambar IV-15. <i>Class Diagram</i>	IV-39
Gambar IV-16. <i>User Interface</i> Perangkat Lunak	IV-40
Gambar V-1. Hasil Kinerja Model	V-9

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini akan membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan. Bagian ini juga memuat penjelasan mengenai gambaran umum dari keseluruhan kegiatan penelitian yang berjudul “Klasifikasi Iklan Lowongan Kerja Menggunakan *Support Vector Machine* (SVM) dan *Borderline Synthetic Minority Over-sampling Technique* (Borderline-SMOTE)”.

1.2 Latar Belakang Masalah

Pertumbuhan internet membuat proses perekrutan tenaga kerja baru menjadi lebih mudah (Anita dkk., 2021). Banyak perusahaan yang memilih untuk mempublikasikan lowongan pekerjaan secara online sehingga lebih mudah untuk diakses bagi pencari kerja (Dutta & Bandyopadhyay, 2020). Perekrutan secara online juga membantu perusahaan untuk menemukan kandidat - kandidat yang memiliki potensial yang terhalang oleh jarak (Anita dkk., 2021). Namun, bersamaan dengan kemudahan ini, tidak semua iklan lowongan pekerjaan dapat dipercaya (Fajar dkk., 2022). Pada tahun 2020, berdasarkan data dari Patroli Siber terdapat 649 laporan dari masyarakat seputar penipuan online yang di dalamnya termasuk laporan penipuan lowongan kerja (Chrisendo dkk., 2023).

Penipuan lowongan kerja merugikan bagi pencari kerja dan menimbulkan masalah bagi perusahaan (Reddy & Reddy, 2021). Bagi pencari kerja, kerugian yang terjadi berupa pencurian informasi atau data, seperti alamat, rekening bank, nomor jaminan sosial, dan lain-lain (Sabita dkk., 2021). Informasi atau data pencari kerja dapat digunakan penipu untuk dijual ke *dark web* atau disalahgunakan untuk memperoleh keuntungan (Anita dkk., 2021). Bagi perusahaan, masalah yang terjadi berupa rusaknya reputasi perusahaan dan meninggalkan dampak negatif di benak pencari kerja tentang perusahaan tersebut (Reddy & Reddy, 2021).

Permasalahan tersebut dapat diatasi dengan melakukan klasifikasi terhadap iklan lowongan kerja dengan *dataset* telah dianalisis dan dibersihkan serta menerapkan algoritma *machine learning* yang sesuai (Anita dkk., 2021). Banyak algoritma *machine learning* tradisional yang telah berhasil dalam melakukan klasifikasi teks seperti *Naïve Bayes*, *Support Vector Machine*, *Decision Tree*, *K-Nearest Neighbors*, *Random Forest*, dan *Logistic Regression* (Efrizoni dkk., 2022). *Support Vector Machine* merupakan salah satu metode klasifikasi teks yang terkenal kuat dan akurat (Qadrini dkk., 2022). Algoritma *Support Vector Machine* memisahkan antara suatu kelas dengan kelas lainnya dengan mencari *hyperplane* (Fitri dkk., 2020).

Penelitian sebelumnya dari Ilmawan & Mude (2020) membandingkan metode *Support Vector Machine* dan *Naïve Bayes* untuk analisis sentimen pada ulasan tekstual di Google Play Store. Hasil penelitian menunjukkan bahwa *Support Vector Machine* mendapatkan akurasi lebih baik sebesar 81,46% dibandingkan dengan akurasi *Naïve Bayes* sebesar 75,41%.

Dalam melakukan klasifikasi teks muncul masalah yang krusial yaitu ketidakseimbangan data atau *imbalanced data* (Peranginangin dkk., 2020). *Imbalanced data* terjadi ketika salah satu kelas memiliki jumlah yang jauh lebih besar dibanding kelas lainnya sehingga menyebabkan menurunnya kinerja klasifikasi pada kelas minoritas (Fitriani dkk., 2021). Masalah utama yang timbul akibat data yang tidak seimbang adalah proses klasifikasi cenderung mengarah pada klasifikasi ke dalam kelas mayoritas. (Johnson & Khoshgoftaar, 2019).

Borderline Synthetic Minority Over-sampling Technique (Borderline-SMOTE) merupakan metode *oversampling* yang dapat diterapkan untuk menangani permasalahan *imbalanced data* (Indrawati, 2021). Metode ini merupakan varian lain dari metode *Synthetic Minority Over-sampling Technique (SMOTE)*. Ide dasar dari teknik SMOTE adalah membuat sampel sintetis baru untuk kelas minoritas untuk mencapai distribusi *dataset* yang seimbang. Tidak seperti SMOTE, *Borderline-SMOTE* hanya menghasilkan sampel sintetis di area *borderline* (Smiti & Soui, 2020). Sampel yang berada di area *borderline* dan sekitarnya cenderung salah diklasifikasikan daripada yang jauh dari *borderline* (Han dkk., 2005). Terdapat dua tipe dari *Borderline-SMOTE* yaitu tipe 1 dan tipe 2. *Borderline-SMOTE* tipe 1 menggunakan tetangga terdekat dari kelas minoritas untuk menciptakan sampel sintetis. *Borderline-SMOTE* tipe 2 tidak hanya menggunakan tetangga terdekat dari kelas minoritas, tetapi juga tetangga terdekat dari kelas mayoritas (Maciejewski & Stefanowski, 2011).

Penelitian sebelumnya dari dari Feng (2022) menerapkan *Borderline-SMOTE* dengan algoritma *Random Forest*, *Naïve Bayes*, *Decision Tree* dan *K-*

Nearest Neighbors. Penelitian tersebut menunjukkan bahwa *Borderline-SMOTE* dapat mengatasi ketidakseimbangan data dan meningkatkan kinerja algoritma *machine learning*. Han dkk. (2005) melakukan perbandingan teknik *oversampling* untuk mengatasi ketidakseimbangan data, dengan memfokuskan perhatian pada *recall* dan *F-value*. Teknik *oversampling* yang dibandingkan adalah *SMOTE*, *random oversampling*, dan *Borderline-SMOTE*. Hasil penelitian menunjukkan *Borderline-SMOTE* lebih unggul daripada *SMOTE* dan *random oversampling*.

Berdasarkan uraian penelitian tersebut, *Borderline-SMOTE* mampu mengatasi ketidakseimbangan data dan meningkatkan kinerja klasifikasi. Dengan demikian, penelitian ini akan mengembangkan teknik *oversampling* yaitu *Borderline-SMOTE* dan *Support Vector Machine* untuk melakukan klasifikasi iklan lowongan kerja.

1.3 Rumusan Masalah

Berdasarkan informasi yang telah diuraikan dalam latar belakang, perumusan masalah dapat dirumuskan sebagai berikut:

1. Bagaimana mengklasifikasikan iklan lowongan kerja menggunakan *Support Vector Machine* dan *Borderline Synthetic Minority Oversampling Technique*?
2. Bagaimana kinerja *Support Vector Machine* dalam melakukan klasifikasi iklan lowongan kerja berdasarkan *precision*, *recall*, dan *f-measure*?

3. Bagaimana kinerja *Support Vector Machine* dengan *Borderline Synthetic Minority Over-sampling Technique* dalam melakukan klasifikasi iklan lowongan kerja berdasarkan *precision*, *recall*, dan *f-measure*?

1.4 Tujuan Penelitian

Berikut tujuan penelitian ini:

1. Menghasilkan klasifikasi iklan lowongan kerja menggunakan *Support Vector Machine* dan *Borderline Synthetic Minority Over-sampling Technique*.
2. Mengetahui kinerja *Support Vector Machine* dalam klasifikasi iklan lowongan kerja berdasarkan *precision*, *recall*, dan *f-measure*.
3. Mengetahui kinerja *Support Vector Machine* dengan *Borderline Synthetic Minority Over-sampling Technique* dalam klasifikasi iklan lowongan kerja berdasarkan *precision*, *recall*, dan *f-measure*.

1.5 Manfaat Penelitian

Berikut manfaat dari penelitian ini:

1. Menjadi alat bantu bagi pencari kerja untuk mengurangi risiko terkena *scam* melalui iklan lowongan kerja yang tersedia di internet.
2. Hasil penelitian dapat digunakan sebagai sumber referensi bagi peneliti lain yang menggunakan *Support Vector Machine* dan *Borderline Synthetic Minority Over-sampling Technique*.

1.6 Batasan Masalah

Berikut adalah batasan masalah dari penelitian ini:

1. *Dataset* yang digunakan dalam penelitian bersumber dari <https://www.kaggle.com/>.
2. *Dataset* menggunakan bahasa Inggris.
3. Klasifikasi dibagi menjadi dua kategori: asli dan palsu.
4. *Kernel* yang digunakan pada *Support Vector Machine* yaitu *linear*.
5. Nilai *C* yang digunakan adalah 0,1, 0,25, 0,50, 0,75, dan 1.
6. Metode pemberian bobot kata menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*).
7. Metode *oversampling* yang digunakan adalah *Borderline Synthetic Minority Over-sampling Technique (Borderline-SMOTE)* dengan tipe 1.

1.7 Sistematika Penulisan

Untuk memudahkan penyusunan dan pemahaman yang lebih baik tentang penelitian ini, setiap isi bab pada laporan akan ditulis secara sistematis sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang, perumusan masalah, tujuan, manfaat, batasan masalah, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi tentang seluruh penjelasan mengenai tinjauan pustaka yang berhubungan dengan permasalahan yang akan dibahas pada penulisan tugas akhir.

BAB III METODOLOGI PENELITIAN

Bab ini berisi penjelasan secara bertahap dan terperinci tentang langkah-langkah (metodologi) yang digunakan untuk kerangka berfikir dan kerangka kerja dalam menyelesaikan tesis.

BAB IV PENGEMBANGAN PERANGKAT LUNAK

Bab ini membahas mengenai pengembangan perangkat lunak berisikan pembahasan mengenai perencanaan dan pembangunan perangkat lunak, mulai dari analisis kebutuhan dan desain, hingga pembuatan dan pengujian perangkat lunak.

BAB V HASIL PENELITIAN DAN PEMBAHASAN

Bab ini menguraikan hasil dan analisis penelitian yang menjadi dasar dalam penarikan kesimpulan dalam penelitian.

BAB VI KESIMPULAN DAN SARAN

Bab ini memaparkan kesimpulan dari hasil penelitian serta saran untuk pengembangan penelitian selanjutnya.

1.8 Kesimpulan

Bab ini menjelaskan latar belakang dan hal-hal terkait dengan penelitian dengan judul klasifikasi iklan lowongan kerja menggunakan *support vector machine* dan *borderline synthetic minority over-sampling technique*.

DAFTAR PUSTAKA

- Anita, C. S., Nagarajan, P., Sairam, G. A., Ganesh, P., & Deepakkumar, G. (2021). Fake Job Detection and Analysis Using Machine Learning and Deep Learning Algorithms. *Revista Gestão Inovação e Tecnologias*, 11(2), 642–650. <https://doi.org/10.47059/revistageintec.v11i2.1701>
- Apriani, I., Sibaroni, Y., & Palupi, I. (2023). Perbandingan Pembobotan Fitur TF-IDF dan TF-ABS Dalam Klasifikasi Berita Online Menggunakan Support Vector Machine (SVM) 2 nd. *e-Proceeding of Engineering*, 10(3), 3652–3663.
- Bordoloi, M., & Biswas, S. K. (2018). Sentiment Analysis of Product using Machine Learning Technique: A Comparison among NB, SVM and MaxEnt. *International Journal of Pure and Applied Mathematics*, 118(19), 71–83. <https://www.researchgate.net/publication/326572767>
- Chrisendo, L., Takariawan, A., & Ramadhani, R. H. (2023). Analisis Yuridis Penegakan Hukum Terhadap Tindak Pidana Penipuan dengan Modus Lowongan Kerja Melalui Media Sosial Berdasarkan Hukum Positif di Indonesia. *COMSERVA: Jurnal Penelitian dan Pengabdian Masyarakat*, 2(12), 3195–3218. <https://doi.org/10.59141/comserva.v2i12.707>
- Dalal, M. K., & Zaveri, M. A. (2011). Automatic Text Classification: A Technical Review. *International Journal of Computer Applications*, 28(2), 37–40. <https://doi.org/10.5120/3358-4633>
- Dutta, S., & Bandyopadhyay, S. K. (2020). Fake Job Recruitment Detection Using Machine Learning Approach. *International Journal of Engineering Trends and Technology*, 68. <http://www.ijettjournal.org>
- 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>
- Fajar, M. M., Putri, A. R., & Holle, K. F. H. (2022). Perbandingan Metode Klasifikasi Data Mining Untuk Deteksi Keaslian Lowongan Pekerjaan di

- Medsos. *Jurnal Ilmiah Informatika*, 7(1), 41–48.
<https://doi.org/10.35316/jimi.v7i1.41-48>
- Feng, L. (2022). Research on Customer Churn Intelligent Prediction Model based on Borderline-SMOTE and Random Forest. *IEEE*, 803–807.
- Fitri, E., Yuliani, Y., Rosyida, S., & Gata, W. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. *TRANSFORMTIKA*, 18(1), 71–80.
www.nusamandiri.ac.id,
- Fitriani, R. D., Yasin, H., & Tarno. (2021). PENANGANAN KLASIFIKASI KELAS DATA TIDAK SEIMBANG DENGAN RANDOM OVERSAMPLING PADA NAIVE BAYES (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal). *JURNAL GAUSSIAN*, 10(1), 11–20.
- HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PLoS ONE*, 15(5), 1–22. <https://doi.org/10.1371/journal.pone.0232525>
- Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. *ICIC*, 878–887.
- Ilmawan, L. B., & Mude, M. A. (2020). Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store. *ILKOM Jurnal Ilmiah*, 12(2), 154–161.
<https://doi.org/10.33096/ilkom.v12i2.597.154-161>
- Indrawati, A. (2021). PENERAPAN TEKNIK KOMBINASI OVERSAMPLING DAN UNDERSAMPLING UNTUK MENGATASI PERMASALAHAN IMBALANCED DATASET. *Jurnal Informatika dan Komputer*, 4(1), 38–43.
<https://doi.org/10.33387/jiko>
- Indrawati, A., Subagyo, H., Sihombing, A., Wagiyah, & Afandi, S. (2020). ANALYZING THE IMPACT OF RESAMPLING METHOD FOR IMBALANCED DATA TEXT IN INDONESIAN SCIENTIFIC ARTICLES CATEGORIZATION. *BACA: JURNAL DOKUMENTASI DAN INFORMASI*, 41(2), 133–141. <https://doi.org/10.14203/j.baca.v41i2.563>

- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(27), 1–54. <https://doi.org/10.1186/s40537-019-0192-5>
- Jwo, D.-J., Wu, J.-C., & Ho, K.-L. (2021). Support vector machine assisted GPS navigation in limited satellite visibility. *Computers, Materials and Continua*, 69(1), 555–574. <https://doi.org/10.32604/cmc.2021.018320>
- Kangwanariyakul, Y., Nantasenamat, C., Tantimongcolwat, T., & Naenna, T. (2010). Data mining of magnetocardiograms for prediction of ischemic heart disease. *EXCLI Journal*, 82–95. <https://doi.org/10.17877/DE290R-15805>
- Lazuardy, A. G., & Setiaji, H. (2019). DATA CLEANSING PADA DATA RUMAH SAKIT. *Proceeding SINTAK*.
- Lee, T., Kim, M., & Sung-Phil, K. (2020). Data Augmentation Effects Using Borderline-SMOTE on Classification of a P300-Based BCI. *2020 8th International Winner Conference on Brain-Computer Interface (BCI)*.
- Maciejewski, T., & Stefanowski, J. (2011). Local Neighbourhood Extension of SMOTE for Mining Imbalanced Data. *IEEE*.
- Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020). Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. *2020 11th International Conference on Information and Communication Systems, ICICS 2020*, 243–248. <https://doi.org/10.1109/ICICS49469.2020.239556>
- Mowafy, M., Rezk, A., & El-Bakry, H. M. (2018). An Efficient Classification Model for Unstructured Text Document. *American Journal of Computer Science and Information Technology*, 06(01), 1–10. <https://doi.org/10.21767/2349-3917.100016>
- Nor, S., Muslim, M. A., & Aswin, M. (2022). Pengenalan Pola Dasar Angka berdasarkan Gerakan Tangan menggunakan Machine Learning. *ELKOMIKA*, 10(3), 596–608. <https://doi.org/10.26760/elkomika.v10i3.596>
- 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 Manajemen*, 7(1), 1–17.

- Peranginangin, R., Harianja, E. J. G., Jaya, I. K., & Rumahorbo, B. (2020). PENERAPAN ALGORITMA SAFE-LEVEL-SMOTE UNTUK PENINGKATAN NILAI G-MEAN DALAM KLASIFIKASI DATA TIDAK SEIMBANG. *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, 4(1), 67–72. <https://doi.org/10.46880/jmika.Vol4No1.pp67-72>
- Perwitasari, R., Afawani, R., & Endang Anjarwani, S. (2020). PENERAPAN METODE RATIONAL UNIFIED PROCESS (RUP) DALAM PENGEMBANGAN SISTEM INFORMASI MEDICAL CHECK UP PADA CITRA MEDICAL CENTRE. *JTIKA*, 2(1), 76–88. <http://jtika.if.unram.ac.id/index.php/JTIKA/>
- Prasetyo, T. A., Widiyanto, D., & Prasvita, D. S. (2022). ANALISIS SENTIMEN TERHADAP VAKSIN NUSANTARA PADA MEDIA SOSIAL YOUTUBE MENGGUNAKAN METODE NAÏVE BAYES DAN SELEKSI FITUR PARTICLE SWARM OPTIMIZATION. *SENAMIKA*, 3(2), 936–945.
- Qadrini, L., Hikmah, & Megasari. (2022). Oversampling, Undersampling, Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejava Timur Tahun 2017. *Journal of Computer System and Informatics (JoSYC)*, 3(4), 386–391. <https://doi.org/10.47065/josyc.v3i4.2154>
- Rachburee, N., & Punlumjeak, W. (2021). Oversampling technique in student performance classification from engineering course. *International Journal of Electrical and Computer Engineering*, 11(4), 3567–3574. <https://doi.org/10.11591/ijece.v11i4.pp3567-3574>
- Reddy, K. S. S., & Reddy, K. L. (2021). FAKE JOB RECRUITMENT DETECTION. *Journal of Emerging Technologies and Innovative Research*, 8(8), 443–448. www.jetir.orgd443
- Romadoni, F., Umaidah, Y., & Sari, B. N. (2020). Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 9(2), 247–253. <https://doi.org/10.32736/sisfokom.v9i2.903>

- Sabita, H., Fitria, & Herwanto, R. (2021). ANALISA DAN PREDIKSI IKLAN LOWONGAN KERJA PALSU DENGAN METODE NATURAL LANGUAGE PROGRAMING DAN MACHINE LEARNING. *Jurnal Informatika*, 21(1), 14–22.
- Saifudin, A., & Wahono, R. S. (2015). Pendekatan Level Data untuk Menangani Ketidakseimbangan Kelas pada Prediksi Cacat Software. *Journal of Software Engineering*, 1(2), 76–85. <http://journal.ilmukomputer.org>
- Smiti, S., & Soui, M. (2020). Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE. *Information Systems Frontiers*, 22(5), 1067–1083. <https://doi.org/10.1007/s10796-020-10031-6>
- Sumendap, R. V., Bagus, I., & Mahendra, M. (2023). Membandingkan Analisis Sentimen Review Pelanggan Shopee Dan Tokopedia Menggunakan Google's NLP API. *Jurnal Elektronik Ilmu Komputer Udayana* , 11(4), 2654–5101.