

**KLASIFIKASI SMS MENGGUNAKAN METODE SELEKSI
FITUR MUTUAL INFORMATION (MI), ALGORITMA
SUPPORT VECTOR MACHINE (SVM) DAN SMOTE**

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



Oleh:

Nur Zahira Indrayati Pakpahan

NIM : 09021382025114

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

LEMBAR PENGESAHAN SKRIPSI

KLASIFIKASI SMS MENGGUNAKAN METODE SELEKSI
FITUR MUTUAL INFORMATION (MI), ALGORITMA SUPPORT
VECTOR MACHINE (SVM) DAN SMOTE

Oleh:

NUR ZAHIRA INDRAYATI PAKPAHAN
NIM: 09021382025114

Palembang, 23 April 2024

Pembimbing I

Alvi Svahrini Utami, M.Kom.
NIP. 197812222006042003

Pembimbing II

Desty Rodiah, M.T.
NIP. 198912212020122011

Mengetahui,



Ketua Jurusan Teknik Informatika

DR. M. Fachrurrozi, S.Si., M.T.
NIP. 198005222008121002

TANDA LULUS UJIAN KOMPRESIENSIF SKRIPSI

Pada hari Jumat tanggal 5 April 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

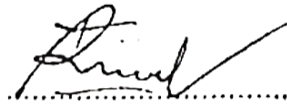
Nama : Nur Zahira Indrayati Pakpahan

NIM : 09021382025114

Judul : KLASIFIKASI SMS MENGGUNAKAN METODE SELEKSI FITUR MUTUAL INFORMATION (MI), ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN SMOTE dan dinyatakan LULUS

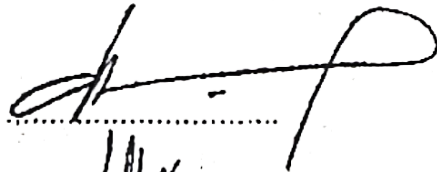
1. Ketua Penguji

Mastura Diana Marieska, M.T.
NIP. 198603212018032001



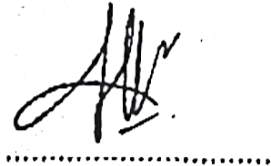
2. Penguji I

Julian Supardi, S.Pd., M.T., Ph.D.
NIP. 197207102010121001



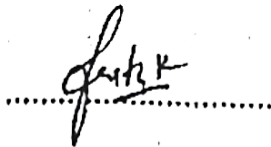
3. Pembimbing I

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



4. Pembimbing II

Desty Rodiah, M.T.
NIP. 198912212020122011



Mengetahui,
Ketua Jurusan Teknik Informatika



DR. M. Fachrurrozi, S.Si., M.T.
NIP. 198005222008121002



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Nur Zahira Indrayati Pakpahan

NIM : 09021382025114

Program Studi : Teknik Informatika Bilingual

Judul : Klasifikasi SMS Menggunakan Metode Seleksi Fitur Mutual Information (MI), Algoritma Support Vector Machine (SVM) dan SMOTE

Hasil Pengecekan Software *iThenticate/Turnitin*: 12%

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, 22 April 2024



Nur Zahira Indrayati Pakpahan

NIM. 09021382025114

MOTO DAN PERSEMBAHAN

“It’s fine to fake it until you make it, until you do, until it true” – Taylor Swift

Kupersembahkan karya tulis ini kepada:

- Orang Tua, Saudari, dan Keluargaku
- Dosen Pembimbing
- Teman-teman seperjuangan
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

Communication media such as SMS (Short Message Service) is now rarely used, but SMS is still needed because it can still be useful for users in receiving information. However, problems arise due to fraudulent messages. Thus, SMS classification becomes important by using the Synthetic Minority Oversampling Technique (SMOTE) method which can balance unbalanced data, Mutual Information (MI) which is used to measure the relationship between words and classes, and Support Vector Machine (SVM) for classifiers that can separate classes. The data used in this research is divided into promo, fraud and normal. Test modelling was performed with a comparison of parameter C. The results showed that the use of the SMOTE + SVM model produced a higher accuracy value than the addition of MI, but the SMOTE + MI + SVM method produced a more stable performance than without the use of MI.

Keywords: SMS, Synthetic Minority Oversampling, Mutual Information, Support Vector Machine

ABSTRAK

Media komunikasi seperti SMS (*Short Message Service*) sekarang sudah jarang digunakan, tetapi SMS tersebut tetap dibutuhkan karena masih dapat bermanfaat bagi pengguna dalam menerima informasi. Namun, masalah muncul karena terdapat pesan penipuan. Sehingga, pengklasifikasian SMS menjadi penting dengan menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE) yang dapat menyeimbangkan data yang tidak seimbang, *Mutual Information* (MI) yang digunakan untuk mengukur hubungan kata dengan kelas, dan *Support Vector Machine* (SVM) untuk pengklasifikasian yang dapat memisahkan kelas. Data yang digunakan pada penelitian ini dibagi menjadi promo, penipuan dan normal. Pembuatan model pengujian dilakukan dengan perbandingan parameter C. Hasil penelitian menunjukkan bahwa pada penggunaan model SMOTE + SVM menghasilkan nilai *accuracy* yang lebih tinggi dibandingkan dengan penambahan MI, tetapi Metode SMOTE + MI + SVM menghasilkan kinerja yang lebih stabil dibandingkan tanpa penggunaan MI.

Kata kunci: SMS, *Synthetic Minority Oversampling*, *Mutual Information*, *Support Vector Machine*

KATA PENGANTAR

Segala puji dan syukur atas kehadiran Allah SWT atas segala rahmat dan hidayahnya sehingga dapat menyelesaikan Skripsi dengan judul “Klasifikasi SMS Menggunakan Metode Seleksi Fitur *Mutual Information* (MI), Algoritma *Support Vector Machine* (SVM) dan SMOTE”. 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 Allah SWT dan penulis mengucapkan banyak terima kasih kepada :

1. Keluarga saya yaitu Indra Jaya Pakpahan, Yahmawati Simatupang dan kakak saya Bella Sabrina Indrayati Pakpahan. Terima kasih sudah membantu mendoakan dan memberikan semangat kepada penulis.
2. Prof. Dr. Erwin, S.Si., M.Si., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Dr. M. Fachrurrozi, S.Si., M. T. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Ibu Alvi Syahrini Utami, M.Kom. selaku Dosen Pembimbing Akademik di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Desty Rodiah, S.Kom., M.T. selaku Dosen Pembimbing Akademik di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Mba Wiwin dan Mba Rika selaku admin Jurusan Teknik Informatika yang telah membantu mengurus berkas administrasi penulis.

7. Seluruh dosen dan staff Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Kak Raisha Fatiya yang telah membantu dalam pengerjaan Skripsi ini.
9. Alia Rizkya, Daffa Nabila Andrianov, Fransiska Kristina Setyawan, Nazifah Suci Aqillah, Shabrina Ramadhanti, Muhammad Bil Hakki, Faiq Fadlurrahman, Muhammad Ganda Rizky, Fauzan Aziman Zohro dan teman-teman TI Bilingual A serta seluruh teman-teman seperjuangan Angkatan 2020 Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
10. Sindi Nurisa Isman, Salshabilla Aulya Putri, Vanya Nasywa Alfiyyah Putri, Febby Thasya Arini, Syahri Muthia Hilda Farhanie, Dede Oktavia, Muhammad Fiqih Julian Rizki selaku teman edcy dan teman-teman saya till jannah yang membantu menyemangati saat proses pembuatan skripsi.

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, 23 April 2024



Nur Zahira Indrayati Pakpahan

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN SKRIPSI.....	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	iii
HALAMAN PERNYATAAN.....	iv
MOTO DAN PERSEMBAHAN.....	v
ABSTRACT.....	vi
ABSTRAK.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan.....	I-1
1.2 Latar Belakang.....	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
BAB I. PENDAHULUAN.....	I-5
BAB II. KAJIAN LITERATUR.....	I-6
BAB III. METODOLOGI PENELITIAN.....	I-6
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 SMS.....	II-1
2.2.2 Klasifikasi Teks.....	II-1
2.2.3 <i>Mutual Information (MI)</i>	II-2
2.2.4 SMOTE (<i>Synthetic Minority Oversampling Technique</i>).....	II-3
2.2.5 <i>Support Vector Machine (SVM)</i>	II-4
2.2.6 <i>Confusion matrix</i>	II-8

2.2.7	RUP (<i>Rational Unified Process</i>).....	II-9
2.3	Penelitian lain yang relevan.....	II-11
2.4	Kesimpulan.....	II-13
BAB III METODOLOGI PENELITIAN.....		III-1
3.1	Pendahuluan.....	III-1
3.2	Unit Penelitian.....	III-1
3.3	Pengumpulan Data.....	III-1
3.4	Tahapan Penelitian.....	III-3
3.4.1	Membuat Kerangka Kerja Penelitian.....	III-4
3.4.2	Menentukan Alat yang Digunakan Pada Penelitian.....	III-6
3.4.3	Menentukan Kriteria Pengujian.....	III-7
3.4.4	Menentukan Format Data Pengujian.....	III-7
3.4.5	Melakukan Pengujian Penelitian.....	III-8
3.4.6	Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan.....	III-8
3.5	Metode Pengembangan Perangkat Lunak.....	III-8
3.5.1	Fase Insepsi.....	III-9
3.5.2	Fase Elaborasi.....	III-9
3.5.3	Fase Konstruksi.....	III-9
3.5.4	Fase Transisi.....	III-9
3.6	Manajemen Proyek Penelitian.....	III-10
3.7	Kesimpulan.....	III-13
BAB IV PENGEMBANGAN PERANGKAT LUNAK.....		IV-1
4.3	Pendahuluan.....	IV-1
4.2	Fase <i>Inception</i>	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 <i>Elaboration</i>	IV-46
4.3.1	Pemodelan Bisnis.....	IV-46
4.3.2	Perancangan Data.....	IV-46
4.3.3	Perancangan <i>interface</i>	IV-47
4.3.4	Kebutuhan Sistem.....	IV-48
4.3.5	<i>Activity Diagram</i>	IV-49
4.3.6	<i>Sequence Diagram</i>	IV-53
4.4	Fase <i>Construction</i>	IV-56
4.4.1	Kebutuhan Sistem.....	IV-56
4.4.2	<i>Class Diagram</i>	IV-57
4.4.3	Implementasi.....	IV-58
4.5	Fase <i>Transition</i>	IV-61

4.5.1	Pemodelan Bisnis.....	IV-61
4.5.2	Rencana Pengujian.....	IV-61
4.5.3	Implementasi.....	IV-62
4.6	Kesimpulan.....	IV-64
BAB V HASIL DAN ANALISIS PENELITIAN.....		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 Pengujian Klasifikasi.....	V-8
5.3	Kesimpulan.....	V-14
BAB VI KESIMPULAN DAN SARAN.....		V-1
6.1	Pendahuluan.....	V-1
6.2	Kesimpulan.....	V-1
DAFTAR PUSTAKA.....		xi
LAMPIRAN.....		xiv

DAFTAR TABEL

Tabel II-1. Contoh Penerapan <i>Preprocessing</i>	II-1
Tabel III-1. Contoh Data Penelitian.....	III-2
Tabel III-2. Model <i>Confusion Matrix</i>	III-7
Tabel III-3. Performa <i>Confusion Matrix</i>	III-8
Tabel III-4. Tabel Manajemen Proyek Penelitian.....	III-10
Tabel IV-1. Kebutuhan Fungsional.....	IV-2
Tabel IV-2. Kebutuhan Non-Fungsional.....	IV-2
Tabel IV-3. Contoh Data SMS.....	IV-4
Tabel IV-4. Hasil <i>Cleaning</i>	IV-5
Tabel IV-5. Hasil Proses <i>Case Folding</i>	IV-5
Tabel IV-6. Hasil Proses <i>Normalization</i>	IV-6
Tabel IV-7. Hasil Proses <i>Tokenization</i>	IV-6
Tabel IV-8. Hasil <i>Stopword Removal</i>	IV-7
Tabel IV-9. Hasil <i>Stemming</i>	IV-7
Tabel IV-10. Hasil Pembobotan TF-IDF.....	IV-8
Tabel IV-11. Hasil Pembobotan Kata TF-IDF.....	IV-11
Tabel IV-12. Hasil Jarak Terdekat Pada Kelas Minoritas.....	IV-14
Tabel IV-13. Hasil Pembentukan Data Sintesis Pada Kelas Minoritas.....	IV-19
Tabel IV-14. Term pada Data Sintesis.....	IV-23
Tabel IV-15. Perhitungan bobot nilai <i>Mutual Information</i>	IV-26
Tabel IV-16. Hasil Seleksi Fitur <i>Mutual Information</i>	IV-36
Tabel V-1. <i>Confusion Matrix</i> SMOTE + SVM.....	V-2
Tabel V-2. <i>Confusion Matrix</i> SMOTE +MI + SVM.....	V-2
Tabel V-3. Hasil Pelatihan Model Pertama.....	V-3
Tabel V-4. <i>Confusion Matrix</i> SMOTE + SVM.....	V-3
Tabel V-5. <i>Confusion Matrix</i> SMOTE +MI + SVM.....	V-4
Tabel V-6. Hasil Pelatihan Model Kedua.....	V-4

Tabel V-7. <i>Confusion Matrix</i> SMOTE + SVM.....	V-5
Tabel V-8. <i>Confusion Matrix</i> SMOTE +MI + SVM.....	V-5
Tabel V-9. Hasil Pelatihan Model Ketiga.....	V-6
Tabel V-10. <i>Confusion Matrix</i> SMOTE + SVM.....	V-6
Tabel V-11. <i>Confusion Matrix</i> SMOTE +MI + SVM.....	V-6
Tabel V-12. Hasil Pelatihan Model Keempat.....	V-7
Tabel V-13. <i>Confusion Matrix</i> SMOTE + SVM.....	V-7
Tabel V-14. <i>Confusion Matrix</i> SMOTE +MI + SVM.....	V-8
Tabel V-15. Hasil Pelatihan Model Kelima.....	V-8
Tabel V-16. Hasil Pengujian <i>Accuracy</i> Pada Klasifikasi.....	V-9
Tabel V-17. Perbandingan <i>Accuracy</i> Data <i>Train</i> dan Data <i>Testing</i>	V-13

DAFTAR GAMBAR

Gambar II-1. <i>Hyperplane</i> pada SVM (Rana & Singh, 2016).....	II-5
Gambar II-2. Model <i>Rational Unified Process</i> (Kruchten, 2004).....	II-10
Gambar III-1. Tahapan Penelitian.....	III-3
Gambar III-2. Diagram Kerangka Kerja.....	III-4
Gambar IV-1. Diagram <i>Use Case</i>	IV-41
Gambar IV-2. Rancangan <i>interface</i> Halaman Utama.....	IV-47
Gambar IV-3. Rancangan <i>Interface</i> Halaman Hasil <i>Preprocessing</i>	IV-47
Gambar IV-4. Rancangan <i>Interface</i> Halaman Evaluasi Model, <i>Confusion Matrix</i> dan Hasil Klasifikasi.....	IV-48
Gambar IV-5. <i>Activity diagram</i> Menginput Data SMS.....	IV-49
Gambar IV-6. <i>Activity diagram</i> Melakukan <i>Preprocessing</i> Pada Dataset.....	IV-50
Gambar IV-7. <i>Activity diagram</i> Melakukan Klasifikasi SMOTE dan SVM...IV-51	
Gambar IV-8. <i>Activity diagram</i> melakukan Klasifikasi SMOTE, MI dan SVM.....	IV-52
Gambar IV-9. <i>Sequence Diagram</i> Menginput Data SMS.....	IV-53
Gambar IV-10. <i>Sequence Diagram</i> Melakukan <i>Preprocessing</i> Data.....	IV-54
Gambar IV-11. <i>Sequence Diagram</i> Melakukan Klasifikasi Menggunakan SMOTE dan SVM.....	IV-55
Gambar IV-12. <i>Sequence Diagram</i> Melakukan Klasifikasi Menggunakan SMOTE, MI dan SVM.....	IV-56
Gambar IV-13. <i>Class Diagram</i>	IV-57
Gambar IV-14. Implementasi <i>Interface</i> Halaman Utama.....	IV-59
Gambar IV-15. Implementasi <i>Interface</i> Halaman Hasil <i>Preprocessing</i>	IV-60
Gambar IV-16. Implementasi <i>Interface</i> Halaman Evaluasi Model, <i>Confusion Matrix</i> dan Hasil Klasifikasi.....	IV-60
Gambar V-1. Perbandingan <i>Accuracy</i> Model.....	V-9
Gambar V-2. Perbandingan <i>Precision</i> Model.....	V-10
Gambar V-3. Perbandingan <i>Recall</i> Model.....	V-11
Gambar V-4. Perbandingan <i>F1-Score</i> Model.....	V-12

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan menguraikan latar belakang masalah, rumusan masalah, tujuan penelitian dan juga menjabarkan sistematika penulisan. Selain itu, bab ini juga memberikan gambaran umum dari semua kegiatan penelitian yang telah dilaksanakan.

1.2 Latar Belakang

Kemunculan media komunikasi saat ini banyak memberikan dampak yang besar bagi pengguna, karena seseorang dapat dengan mudah berkomunikasi dengan teman, keluarga dekat maupun keluarga jauh. Contoh dari media komunikasi tersebut yaitu SMS. SMS atau *Short Message Service* merupakan sebuah media komunikasi berbentuk teks yang memungkinkan pengguna ponsel untuk saling berbagi teks (Irawan et al., 2021). Walaupun SMS sekarang sudah jarang digunakan, tetapi fasilitas tersebut tetap dibutuhkan bagi pengguna yang masih menggunakan telepon genggam karena SMS masih dapat memberikan informasi. Seseorang dapat menerima SMS dari orang yang dikenal maupun yang tidak dikenal, karena kemudahan dalam berkomunikasi tersebut maka dapat merugikan seseorang. Contoh dari kerugian tersebut yaitu banyaknya pesan yang masuk dari sejenis promo ataupun penipuan, Sehingga SMS dari orang yang dikenal akan tertumpuk oleh SMS promo ataupun penipuan. Terdapat kerugian lainnya yaitu seseorang dapat mengaku sebagai kerabat penerima SMS untuk mengisikan pulsa, meminta uang ataupun memberikan informasi yang palsu.

Dikarenakan banyaknya SMS yang diterima, maka dibutuhkannya pengklasifikasian untuk pengkategorian setiap SMS. Klasifikasi adalah metode yang digunakan untuk mengevaluasi data dan mengelompokkan data tersebut berdasarkan karakteristik atau atribut yang dimiliki ke dalam kategori (Herwanto et al., 2021). Dalam pengklasifikasian, *Support Vector Machine* (SVM) merupakan metode *machine learning* yang cocok digunakan, karena memiliki tujuan untuk menemukan *hyperplane* terbaik yang dapat memisahkan kelas-kelas (Nugroho et al., 2003). Namun, SVM memiliki kekurangan pada saat memilih parameter yang sesuai (Basari et al., 2013).

Untuk menutupi kekurangan dari SVM tersebut, maka dapat menggunakan seleksi fitur seperti *Mutual Information* (MI). MI memiliki kelebihan yaitu dapat membantu dalam menyeleksi fitur yang tidak informatif sehingga klasifikasi hanya menggunakan fitur yang relevan (Irham et al., 2019) dan terdapat pula kekurangan dari MI yaitu pada saat menangani fitur-fitur yang ambigu atau memiliki banyak makna akan mempengaruhi akurasi model (Edwar et al., 2023).

Setelah mengidentifikasi fitur-fitur yang informatif, terdapat pula tantangan lainnya yaitu pada banyaknya kategori pada SMS sehingga dapat menyebabkan permasalahan yang biasanya dialami yaitu berupa data yang tidak seimbang (*imbalance data*) antara kategori satu dengan lainnya. Sehingga untuk mengatasi *imbalance data* tersebut dapat menggunakan SMOTE (*Synthetic Minority Oversampling Technique*). Teknik SMOTE mampu membuat model lebih stabil dan memiliki kinerja yang lebih baik dari sisi pengukuran (Feng et al., 2021). Sedangkan kekurangan dari SMOTE ini ialah data sampel yang dibentuk tidak

memperhatikan kelas mayoritas sehingga dapat menyebabkan adanya tumpang tindih pada data (Sabilla & Vista, 2021).

Penelitian sebelumnya dilakukan oleh (Qodrin et al., 2022) yang mengklasifikasikan pertanyaan bahasa Indonesia dengan menggunakan *Support Vector Machine* dan *Mutual Information*. Pada SVM tanpa perlu seleksi fitur menimbulkan hasil kerja yang mencukupi hanya saja rentan pada parameter tertentu. Model tersebut menyebabkan nilai akurasi sebesar 91% dan waktu eksekusi sebesar 19 s pada parameter kernel *linear* dan nilai $C : 1$. Sebaliknya, penerapan model SVM dengan penambahan metode seleksi fitur *Mutual Information* bisa memperoleh hasil kerja paling baik pada kernel *linear* dengan nilai akurasi 92 % dan waktu eksekusi sebesar 7 s pada kernel *linear* dan nilai $C : 1$. Pada SVM tanpa menggunakan seleksi fitur, memberikan hasil yang lumayan hanya saja kurang stabil pada parameter tertentu. Model tersebut membentuk akurasi 91 % dan waktu komputasi 19 detik dengan menggunakan kernel *linear* dan nilai C sebesar 1. Sementara itu, penerapan metode seleksi fitur *mutual information* di model SVM mampu mendapatkan hasil paling baik dengan akurasi 92% dan waktu komputasi 7 detik menggunakan kernel *linear* dan nilai C sebesar 1. Jadi metode seleksi fitur *mutual information* pada SVM meningkatkan akurasi dan mengurangi waktu komputasi dibandingkan SVM tanpa seleksi fitur. Sehingga dapat dilihat bahwa penggunaan *mutual information* lebih baik. Sedangkan penelitian lainnya yang dilakukan oleh (Febriansyah et al., 2021) melakukan pengujian sentimen analisis pada ulasan film menggunakan seleksi fitur *mutual information* dan *k-nearest neighbour classifier*. Pada pengujian tersebut menghasilkan nilai yang lebih rendah

pada penggunaan seleksi fitur *mutual information* yang menghasilkan akurasi sebesar 73,32% dibandingkan tanpa penggunaan penggunaan seleksi fitur *mutual information* yang menghasilkan akurasi sebesar 77,06%. Model tersebut menandakan *overfitting*.

Berdasarkan latar belakang yang telah dijabarkan di atas, maka dilakukan penelitian tentang klasifikasi SMS dengan judul “klasifikasi SMS Menggunakan Metode Seleksi Fitur *Mutual Information* (MI), Algoritma *Support Vector Machine* (SVM) dan SMOTE (*Synthetic Minority Oversampling Technique*)”.

1.3 Rumusan Masalah

Sesuai dari latar belakang penelitian ini, maka permasalahannya yaitu :

1. Bagaimana penerapan metode seleksi fitur *Mutual Information* dan Algoritma *Support Vector Machine* dalam pengklasifikasian SMS.
2. Bagaimana pengaruh *Mutual Information* yang dikombinasikan dengan *Support Vector Machine* dalam pengklasifikasian SMS.

1.4 Tujuan Penelitian

1. Menerapkan metode *Mutual Information* dan *Support Vector Machine* untuk pengklasifikasian SMS.
2. Melihat pengaruh *Mutual Information* yang digabungkan bersama *Support Vector Machine* dalam pengklasifikasian SMS.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini yaitu sebagai berikut :

1. Memudahkan pengguna dalam mengelompokkan antara SMS promo, penipuan dan normal.
2. Diharapkan penelitian ini dapat dijadikan sebagai referensi, sehingga dapat menghasilkan penelitian yang lebih bagus dan optimal.

1.6 Batasan Masalah

Batasan yang diterapkan yaitu :

1. Dataset yang digunakan adalah kumpulan SMS yang didapat dari Blog Yudi Wibisono.
2. Data yang digunakan adalah 1143 dalam Bahasa Indonesia.
3. Klasifikasi SMS dibagi menjadi 3 yaitu SMS promo, penipuan dan normal.
4. Metode yang digunakan yaitu metode *Mutual Information* (MI) serta metode *Support Vector Machine* (SVM) dengan kernel linear.
5. Menggunakan Teknik SMOTE pada *imbalance data*.

1.7 Sistematika Penulisan

Adapun sistematika penulisan pada penelitian ini yaitu sebagai berikut :

BAB I. PENDAHULUAN

Bab pendahuluan ini akan menguraikan mengenai latar belakang dari penelitian, permasalahan yang diangkat, tujuan pelaksanaan penelitian, manfaat yang akan dicapai dari penelitian, batasan masalah serta sistematika penyusunan laporan penelitian.

BAB II. KAJIAN LITERATUR

Bab ini menguraikan tentang konsep dasar teori tentang klasifikasi, SMS, *Mutual Information* (MI) , *Support Vector Machine* (SVM) dan *Synthetic Minority Oversampling Technique* (SMOTE). Bab ini juga menguraikan tentang penelitian-penelitian terdahulu yang terkait dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini berisi tentang metodologi dan tahapan perancangan penelitian seperti pengumpulan data, tahapan penelitian, metode pengembangan perangkat lunak dan perancangan klasifikasi SMS.

1.8 Kesimpulan

Berdasarkan informasi latar belakang penelitian yang sudah dijabarkan, penelitian ini akan mencoba menerapkan seleksi fitur *Mutual Information* (MI) Algoritma *Support Vector Machine* (SVM) dan Teknik *Synthetic Minority Oversampling Technique* (SMOTE) untuk klasifikasi SMS.

DAFTAR PUSTAKA

- Apandi, T. H., & Sugianto, C. A. 2018. Analisis Komparasi Machine Learning Pada Data Spam SMS. *Jurnal TEDC* 12(1): 58-62.
- Apriyani, H., & Kurniati. 2020. Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus. *Journal of Information Technology Ampera* 1(3):133-143.
- Basari, Abd. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. 2013. Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Engineering* 53:453–462.
- Deolika, A., Kusriani, K., & Luthfi, E. T. 2019. Analisis Pembobotan Kata Pada Klasifikasi Text Mining. (*JURTI*) *Jurnal Teknologi Informasi* 3(2):179-184.
- Edwar, E., Semadi, I.G., Samsudin, M., & Dharmendra, I. K. 2023. Perbandingan Metode Seleksi Fitur Pada Analisis Sentimen (Studi Kasus Opini Pilkada DKI 2017). *Informatics for Educators And Professionals : Journal of Informatics* 8(1):11–18.
- Feng, S., Keung, J., Yu, X., Xiao, Y., & Zhang, M. 2021. Investigation on The Stability of SMOTE-Based Oversampling Techniques in Software Defect Prediction. *Information and Software Technology* 139:106662.
- Fikry, M., & Yusra. 2015. Pemodelan Pengguna berdasarkan Klasifikasi SMS Menggunakan Support Vector Machine pada Perangkat Bergerak Android. *Jurnal Sains, Teknologi Dan Industri* 12(2):286–291.
- Herwanto, H., Chusna, N. L., & Arif, M. S. 2021. Klasifikasi SMS Spam Berbahasa Indonesia Menggunakan Algoritma Multinomial Naïve Bayes. *Jurnal Media Informatika Budidarma* 5(4): 1316-1325.
- Irawan, D., Perkasa, E. B., Yurindra, Y., Wahyuningsih, D., & Helmud, E. 2021. Perbandingan Klassifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)* 10(3):432–437.
- Irham, L. G., Adiwijaya, A., & Wisesty, U. N. 2019. Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine. *Jurnal Media Informatika Budidarma* 3(4):284-292.

- Kasanah, A.N, Muladi, & Pujiyanto, U. 2019. Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. RESTI (Rekayasa Sistem Dan Teknologi Informasi) 3(2):196–201.
- Khang, B., 2002. Trik pemrograman aplikasi berbasis SMS. Jakarta: Elex Media Komputindo.
- Kristianti, D., & Hariyadi, M.A. 2022. Support Vector Machine (SVM) dan Algoritma Naïve Bayes (NB) untuk Mengklasifikasi Keterlambatan Pembayaran Sumbangan Pendidikan di Madrasah Ibtidaiyah. Jurnal Pendidikan Tambusai 6(2):13468-13477.
- Kruchten, P. 2004 . The Rational Unified Process an Introduction Third Edition. Addison-Wesley Professional .
- Maulida, I., Suyatno, A., & Hatta, H. R. 2016. Seleksi Fitur Pada Dokumen Abstrak Teks Bahasa Indonesia Menggunakan Metode Information Gain. Jurnal SIFO Mikroskil 17(2):249–258.
- Nurfikri., Salman, F. & Mubarak, M. S. 2018. News topic classification using mutual information and bayesian network. In 2018 6th International Conference on Information and Communication Technology (ICoICT). IEEE, pp. 162-166.
- Nugroho, A. S., Witarto, A. B., & Handoko, D. 2003. Support Vector Machine. Proceeding Indonesian Scientific Meeting in Central Japan.
- Nurdian, R. A., Ridwan, M., & Yusuf, A. 2022. Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru. Jurnal Teknik Informatika dan Sistem Informasi 8(1):24–32.
- Qodrin, S. A, Yusliani, N., & Syahrini, A. 2022. Klasifikasi Pertanyaan Berbahasa Indonesia Menggunakan Algoritma Support Vector Machine dan Seleksi Fitur Mutual Information. JUPITER (Jurnal Penelitian Ilmu dan Teknik Komputer) 14(2):44–52.
- Rana, S. & Singh, A., 2016. Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques. In 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), pp. 106-111.
- Reviantika, F., Azhar, Y., & Marthasari, G. I. 2021. Analisis Klasifikasi SMS Spam Menggunakan Logistic Regression. Jurnal Repositor 3(4).
- Ricoida, D. I. 2013. Rancang Bangun Sistem Informasi Manajemen Hubungan Pelanggan Pada PT. ABC Menggunakan SMS Gateway. Jurnal Informatika Global 4(1).

- Sabilla, W. I., & Vista, C. B. 2021. Implementasi SMOTE dan Under Sampling pada Imbalanced Dataset untuk Prediksi Kebangkrutan Perusahaan. *Jurnal Komputer Terapan* 7(2):329–339.
- Sari, R., & Mandiri, S. N. 2017. Komparasi Algoritma Support Vector Machine, Naïve Bayes Dan C4.5 Untuk Klasifikasi SMS. *IJCIT (Indonesian Journal on Computer and Information Technology* 2(2):7–13.
- Siringoringo, R. 2018. Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan K-Nearest Neighbor. *Journal Information System Development (ISD)* 3(1):44-49.
- Trisnadi, M.F., Al Faraby, S., & Dwifebri, M. 2021. Sentiment Analysis Pada Movie Review Menggunakan Feature Selection Mutual Information dan K-Nearest Neighbour Classifier. *eProceedings of Engineering* 8(5).
- Febriansyah, M.T., Al Faraby, S., & Dwifebri, M. 2021. Sentiment Analysis Pada Movie Review Menggunakan Feature Selection Mutual Information dan K-Nearest Neighbour Classifier. *eProceedings of Engineering* 8(5).
- Wibawa, A. P., Guntur, M., Purnama, A., Fathony Akbar, M., & Dwiyanto, F. A. 2018. Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi* 3(1).
- Zheng, Z., Cai, Y., & Li, Y. 2015. Oversampling Method For Imbalanced Classification. *Computing and Informatics* 34(5):1017–1037.