

**ANALISIS SENTIMEN TERHADAP KOMENTAR
SARKASME MENGGUNAKAN ALGORITMA
*SUPPORT VECTOR MACHINE (SVM)***

SKRIPSI

Oleh:

ZULNANDI ZIDAN LUBIS

198160072



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

2024

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 20/1/25

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah

3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area
Access From (repository.uma.ac.id)20/1/25

**ANALISIS SENTIMEN TERHADAP KOMENTAR
SARKASME MENGGUNAKAN ALGORITMA
*SUPPORT VECTOR MACHINE (SVM)***

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana di Fakultas Teknik
Universitas Medan Area



Oleh:

ZULNANDI ZIDAN LUBIS

198160072

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA**

2024

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted ⁱⁱ 20/1/25


1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

LEMBAR PENGESAHAN

Judul Skripsi: Analisis sentimen terhadap komentar sarkasme menggunakan algoritma *Support Vector Machine (SVM)*

Nama : Zulnandi Zidan Lubis
NPM : 198160072
Fakultas : Teknik
Prodi : Teknik Informatika

Disetujui Oleh
Pembimbing


Dr. Arnes Sembiring, ST, M.Kom
Mengetahui


Dr. Arnes Sembiring, ST, MT
Fakultas Teknik


Rizki Marlina, S.Kom, M.Kom
Ka. Prodi

Tanggal Lulus: 27 September 2024

iii

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya buat merupakan karya pribadi yang disusun sebagai persyaratan untuk mendapatkan gelar sarjana. Adapun bagian-bagian tertentu yang saya ambil dari karya orang lain telah saya sertakan dengan jelas asal sumbernya, sesuai dengan norma dan etika penulisan ilmiah. Saya siap menerima konsekuensi akademik dan sanksi lainnya sesuai peraturan yang berlaku, jika di masa mendatang terbukti terdapat plagiarisme dalam skripsi ini.

Medan, 30 Agustus 2024
Yang membuat pernyataan



Zulnandi Zidan Lubis
198160072




HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Zulnandi Zidan Lubis
NPM : 198160072
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis karya : Tugas Akhir/Skripsi/Tesis

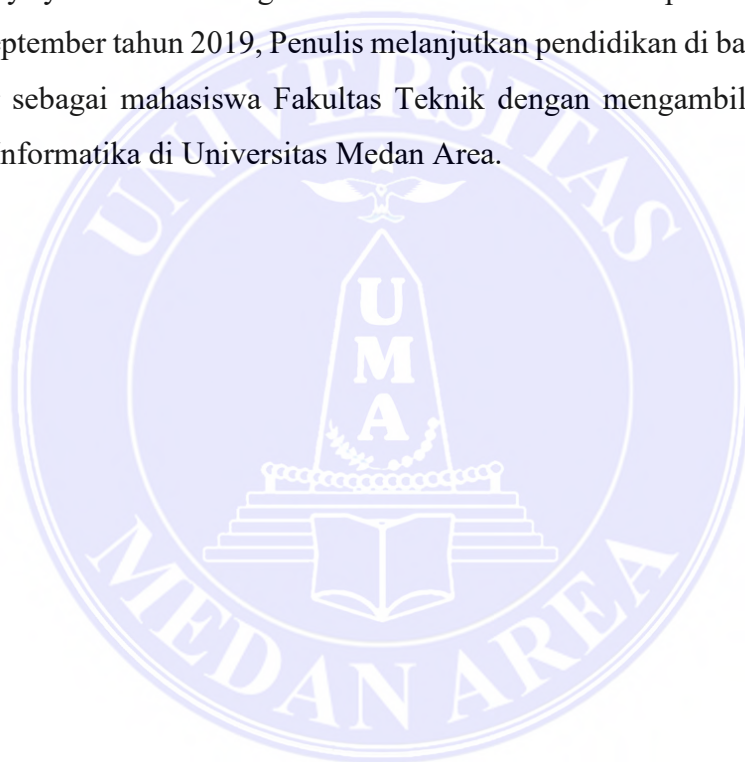
Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul : “**Analisis Sentimen Terhadap Komentar Sarkasme Menggunakan Algoritma *Support Vector Machine (SVM)***” Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir/skripsi/tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan
Pada tanggal : 30 Agustus 2024
Yang menyatakan


Zulnandi Zidan Lubis
198160072

RIWAYAT HIDUP

Penulis dilahirkan di Panyabungan pada tanggal 12 Desember 2000 dari Bapak Zulkarnain Lubis, S.ST dan Ibu Ummi Nadrah Tanjung. Penulis merupakan anak pertama dari tiga bersaudara dan memiliki satu adik laki-laki dan satu adik perempuan. Penulis pertama kali mengenyam pendidikan di bangku SD Negeri 064974 pada tahun 2007 dan lulus pada tahun 2013. Kemudian Penulis melanjutkan pendidikan ke jenjang SMP pada tahun 2013 di SMP Swasta Nur Ihsan dan lulus pada tahun 2016. Pada tahun yang sama, Penulis melanjutkan ke jenjang selanjutnya yaitu di SMK Negeri 1 Percut Sei Tuan dan lulus pada tahun 2019. Pada bulan September tahun 2019, Penulis melanjutkan pendidikan di bangku kuliah dan terdaftar sebagai mahasiswa Fakultas Teknik dengan mengambil Program Studi Teknik Informatika di Universitas Medan Area.



KATA PENGANTAR

Puji Syukur kepada Tuhan Yang Maha Esa, atas berkat dan cinta kasih sehingga penulis mampu menyelesaikan Tugas Akhir yang berjudul “Analisis Sentimen Terhadap Komentar Sarkasme Menggunakan Algoritma *Support Vector Machine* (SVM)”. Tugas Akhir ini adalah salah satu persyaratan yang wajib dipenuhi untuk dapat lulus sarjana Strata Satu (S1) pada Program Studi Teknik Informatika Universitas Medan Area.

Dalam menyelesaikan Tugas Akhir ini tidak lepas dari bantuan dan dukungan, baik secara moril dan materil dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terimakasih kepada :

1. Tuhan Yang Maha Esa, yang telah memberikan kesehatan dan keselamatan selama melakukan studi.
2. Orang tua dan keluarga besar yang telah memberikan doa, semangat, dukungan dan motivasi selama melakukan studi.
3. Bapak Prof. Dr. Dadan Ramdan, M.Eng, M.Sc, selaku Rektor Universitas Medan Area.
4. Bapak Dr. Eng. Supriatno, ST, MT selaku Dekan Fakultas Teknik Universitas Medan Area.
5. Ibu Susilawati, S.Kom, M.Kom, selaku Wakil Bidang Penjamin Mutu Akademik Fakultas Teknik Universitas Medan Area.
6. Bapak Rizki Muliono, S.Kom, M.Kom selaku Kepala Program Studi Teknik Informatika.
7. Bapak Dr. Arnes Sembiring, ST, M.Kom selaku Dosen Pembimbing saya.
8. Bapak/Ibu Dosen dan Staf Teknik Informatika Universitas Medan Area.

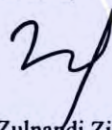
9. Teman-teman Prodi Teknik Informatika Stambuk 2019.

10. Senior yang telah memberi saya banyak ilmu dan pengalaman.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan, oleh karena itu kritik dan saran yang bersifat membangun sangat penulis harapkan demi kesempurnaan skripsi ini. Penulis berharap skripsi ini dapat bermanfaat baik untuk kalangan pendidikan maupun masyarakat. Akhir kata penulis mengucapkan terimakasih.



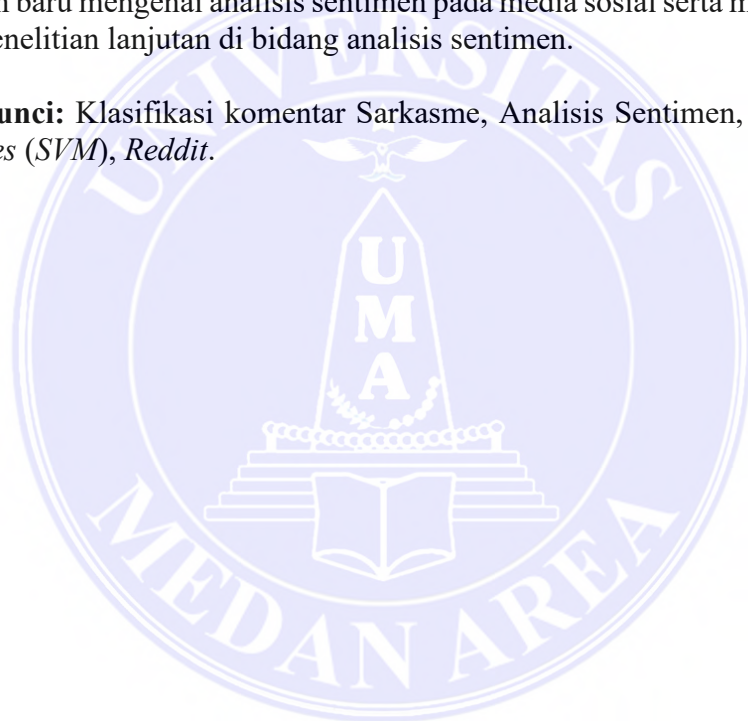
Medan, 21 Juni 2024
Penulis,


Zulnandi Zidan Lubis
198160072

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan komentar sarkasme di media sosial *Reddit* menggunakan algoritma *Support Vector Machines (SVM)*. Komentar sarkasme merupakan bentuk komunikasi yang sering kali sulit diidentifikasi karena menggunakan kata-kata yang bertentangan dengan maksud sebenarnya. Dalam penelitian ini, digunakan data sebanyak 3000 komentar dari *Reddit*, yang melalui tahapan *preprocessing* meliputi *cleaning data*, *case folding*, normalisasi, *stemming*, *filtering*, *tokenizing*, dan *similarity*. Data kemudian dibagi menjadi tiga bagian: 2400 data untuk *training*, 300 data untuk *testing*, dan 300 data untuk *validasi*. Penelitian ini menemukan bahwa penerapan algoritma *SVM* dapat mengklasifikasikan komentar sarkasme dengan tingkat akurasi yang memadai, sehingga dapat meningkatkan efektivitas moderasi konten pada platform daring dan menciptakan lingkungan *online* yang lebih positif. Penelitian ini juga memberikan wawasan baru mengenai analisis sentimen pada media sosial serta menjadi referensi untuk penelitian lanjutan di bidang analisis sentimen.

Kata Kunci: Klasifikasi komentar Sarkasme, Analisis Sentimen, *Support Vector Machines (SVM)*, *Reddit*.



ABSTRACT

This research aimed to classify sarcasm comments on Reddit social media using the Support Vector Machines (SVM) algorithm. Sarcasm comments are a form of communication that is often difficult to identify because it uses words that contradict the actual meaning. In this research, data of 3000 comments from Reddit was used, which went through preprocessing stages including data cleaning, case folding, normalization, stemming, filtering, tokenizing, and similarity. The data was then divided into three parts: 2400 data for training, 300 data for testing, and 300 data for validation. This research found that the application of the SVM algorithm can classify sarcasm comments with a sufficient level of accuracy, thereby increasing the effectiveness of content moderation on online platforms and creating a more positive online environment. This research also provides new insights into sentiment analysis on social media and becomes a reference for further research in the field of sentiment analysis.

Keywords: *Sarcasm comment classification, Sentiment Analysis, Support Vector Machine (SVM), Reddit*



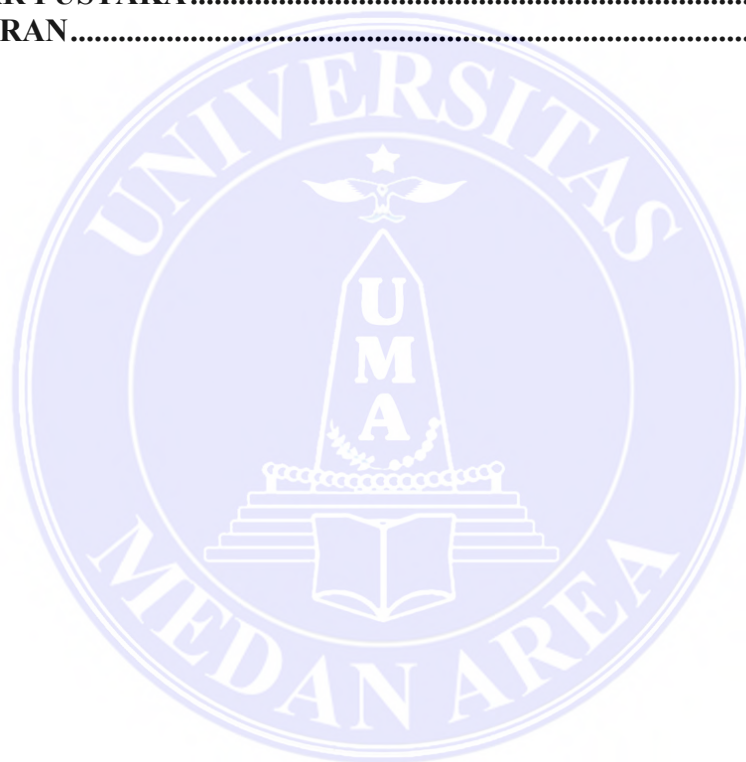
x

DAFTAR ISI

LEMBAR PENGESAHAN	Error! Bookmark not defined.
HALAMAN PERNYATAAN.....	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS	Error! Bookmark not defined.
RIWAYAT HIDUP	vi
KATA PENGANTAR.....	vii
ABSTRAK	ix
ABSTRACT	Error! Bookmark not defined.
DAFTAR ISI.....	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR.....	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1. Komentar	6
2.2. Jenis-jenis Komentar	7
2.3. Analisis Sentimen.....	9
2.4. Metode Pembelajaran Berbasis <i>Kernel</i>	9
2.5. <i>Support Vector Machine (SVM)</i>	10
2.5.1. Prinsip Dasar <i>SVM</i>	11
2.5.2. Penghitungan manual <i>Support Vector Machine (SVM)</i>	13
2.6. <i>Preprocessing</i>	13
2.6.1. <i>Cleaning</i>	14
2.6.2. <i>Case Folding</i>	14
2.6.3. Normalisasi.....	14
2.6.4. <i>Stemming</i>	15
2.6.5. <i>Filtering</i>	15
2.6.6. <i>Tokenizing</i>	15
2.6.7. <i>Cosine Similarity</i>	15
2.7. <i>Reddit</i>	16
2.8. Optimasi Parameter	17
2.9. Penelitian Terkait	18
BAB III METODE PENELITIAN	24
3.1. Metode Penelitian.....	24
3.2. Teknik Pengumpulan Data	27
3.3. Alat dan Bahan Penelitian	27
3.3.1. Perangkat Keras.....	28
3.3.2. Perangkat Lunak.....	28

3.4.	Analisis Kebutuhan Perancangan	28
3.5.	Pelatihan dan Pengujian <i>Support Vector Machines (SVM)</i>	30
3.5.1.	Pemilihan <i>Kernel</i> dan Parameter:	30
3.5.2.	Pemilihan Data Latih:	30
3.5.3.	Representasi Data:	30
3.5.4.	Pelatihan Model:	30
3.5.5.	Optimasi Parameter:	31
3.5.6.	Validasi Model:	31
3.5.7.	Uji Model:	31
3.5.8.	Optimasi Lanjutan:	31
3.5.9.	Pemilihan Dataset Uji	32
3.5.10.	Representasi Data	32
3.5.11.	Uji Model:	32
3.5.12.	Validasi Performa:	32
3.5.13.	<i>Fine-Tuning</i> (Penyetelan Lanjutan):	32
BAB IV HASIL DAN PEMBAHASAN		33
4.1.	Informasi Data	33
4.2.	Hasil	33
4.2.1.	Pemilihan Data	33
4.2.2.	<i>Cleaning</i>	34
4.2.3.	<i>Case Folding</i>	34
4.2.4.	Normalisasi	35
4.2.5.	<i>Stemming</i>	35
4.2.6.	<i>Filtering</i>	35
4.2.7.	<i>Tokenizing</i>	36
4.2.8.	<i>Cosine Similarity</i>	36
4.2.9.	<i>Preprocessing</i>	37
4.2.10.	Klasifikasi Menggunakan <i>SVM Linear Kernel</i>	38
4.2.11.	Klasifikasi Menggunakan <i>Grid Search Linear Kernel</i>	40
4.2.12.	Klasifikasi Menggunakan <i>Random Search Linear Kernel</i>	41
4.2.13.	Klasifikasi Menggunakan <i>Bayesian Optimization Linear Kernel</i>	42
4.2.14.	Hasil Evaluasi <i>SVM Linear Kernel</i>	43
4.2.15.	Klasifikasi Menggunakan <i>SVM Polynomial Kernel</i>	44
4.2.16.	Klasifikasi Menggunakan <i>Grid Search Polynomial Kernel</i>	46
4.2.17.	Klasifikasi Menggunakan <i>Random Search Polynomial Kernel</i> ..	47
4.2.18.	Klasifikasi Menggunakan <i>Bayesian Optimization Polynomial Kernel</i> ..	48
4.2.19.	Hasil Evaluasi <i>SVM Polynomial Kernel</i>	49
4.2.20.	Klasifikasi Menggunakan <i>SVM RBF Kernel</i>	50
4.2.21.	Klasifikasi Menggunakan <i>Grid Search Optimization RBF Kernel</i> ..	52
4.2.22.	Klasifikasi Menggunakan <i>Random Search Optimization RBF Kernel</i> ..	53
4.2.23.	Klasifikasi Menggunakan <i>Bayesian Optimization RBF Kernel</i> ..	54
4.2.24.	Hasil Evaluasi <i>SVM RBF Kernel</i>	55
4.2.25.	Klasifikasi Menggunakan <i>SVM Sigmoid Kernel</i>	56

4.2.26.	Klasifikasi Menggunakan <i>Grid Search Optimazation Sigmoid Kernel</i>	57
4.2.27.	Klasifikasi Menggunakan <i>Random Search Optimazation Sigmoid Kernel</i>	58
4.2.28.	Klasifikasi Menggunakan <i>Bayesian Optimazation Sigmoid Kernel</i>	59
4.2.29.	Hasil Evaluasi <i>SVM Sigmoid Kernel</i>	60
4.3.	Pembahasan	61
BAB V KESIMPULAN DAN SARAN		62
5.1.	Kesimpulan.....	62
5.2.	Saran	62
DAFTAR PUSTAKA		63
LAMPIRAN.....		66



DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	18
Tabel 4. 1 Data Komentar yang akan digunakan	33
Tabel 4. 2 Hasil <i>Cleaning</i>	34
Tabel 4. 3 Hasil <i>Case Folding</i>	34
Tabel 4. 4 Hasil Normalisasi	35
Tabel 4. 5 Hasil <i>Stemming</i>	35
Tabel 4. 6 Hasil <i>Filtering</i>	36
Tabel 4. 7 Hasil <i>Tokenizing</i>	36
Tabel 4. 8 Hasil <i>Cosine Similarity</i>	36
Tabel 4. 9 Hasil Data <i>Preprocessing</i>	37
Tabel 4. 10 Hasil Evaluasi <i>SVM Linear Kernel</i>	43
Tabel 4. 11 Hasil Evaluasi <i>SVM Polynomial Kernel</i>	49
Tabel 4. 12 Hasil Evaluasi <i>SVM RBF Kernel</i>	55
Tabel 4. 13 Hasil Evaluasi <i>SVM Sigmoid Kernel</i>	60
Tabel 4. 14 Hasil Evaluasi Algoritma <i>Support Vector Machine (SVM)</i>	61

DAFTAR GAMBAR

Gambar 3. 1	Diagram Alur Penelitian.....	24
Gambar 4. 1	<i>Confusion Matrix SVM Linear Kernel</i>	38
Gambar 4. 2	<i>Learning Curve SVM Linear Kernel</i>	39
Gambar 4. 3	<i>Confusion Matrix Grid Search Linear Kernel</i>	40
Gambar 4. 4	<i>Confusion Matrix Random Search Linear Kernel</i>	41
Gambar 4. 5	<i>Confusion Matrix Bayesian Optimization Linear Kernel</i>	42
Gambar 4. 6	<i>Confusion Matrix SVM Polynomial Kernel</i>	44
Gambar 4. 7	<i>Learning Curve SVM Polynomial Kernel</i>	45
Gambar 4. 8	<i>Confusion Matrix Grid Search Polynomial Kernel</i>	46
Gambar 4. 9	<i>Confusion Matrix Random Search Polynomial Kernel</i>	47
Gambar 4. 10	<i>Confusion Matrix Bayesian Optimization Polynomial Kernel</i>	48
Gambar 4. 11	<i>Confusion Matrix SVM RBF Kernel</i>	50
Gambar 4. 12	<i>Learning Curve SVM RBF Kernel</i>	51
Gambar 4. 13	<i>Confusion Matrix Grid Search Optimization RBF Kernel</i>	52
Gambar 4. 14	<i>Confusion Matrix Random Search Optimization RBF Kernel</i>	53
Gambar 4. 15	<i>Confusion Matrix Bayesian Optimization RBF Kernel</i>	54
Gambar 4. 16	<i>Confusion Matrix SVM Sigmoid Kernel</i>	56
Gambar 4. 17	<i>Confusion Matrix Grid Search Optimization Sigmoid Kernel</i>	57
Gambar 4. 18	<i>Confusion Matrix Random Search Optimization Sigmoid Kernel</i>	58
Gambar 4. 19	<i>Confusion Matrix Bayesian Optimization Sigmoid Kernel</i>	59

BAB I

PENDAHULUAN

1.1. Latar Belakang

Komentar adalah ungkapan atau pendapat yang diberikan seseorang terhadap suatu hal, topik, atau peristiwa (Mustofa, A., & Suhartono, D, 2021). Komentar bisa berupa respon tertulis atau lisan yang menyatakan pendapat, tanggapan, atau evaluasi terhadap sesuatu. Komentar dapat ditemui dalam berbagai konteks termasuk dalam diskusi *online*, artikel, forum, media sosial, dan banyak lagi.

Sarkasme adalah gaya bahasa yang digunakan untuk menyampaikan sindiran dengan menggunakan kata-kata yang berlawanan dengan maksud sebenarnya, sering kali dengan tujuan mengejek atau mengkritik (Damayanti, E, 2021). Biasanya, sarkasme disampaikan dengan nada atau intonasi tertentu yang menunjukkan ketidaktulusan atau ironi. Misalnya, jika seseorang datang terlambat, seseorang mungkin berkata, "Wah, kamu benar-benar tepat waktu!" untuk menyindir keterlambatan tersebut. Dalam komunikasi tertulis, sarkasme dapat lebih sulit dikenali tanpa petunjuk kontekstual atau penekanan tertentu.

Dalam konteks media sosial atau platform daring lainnya, komentar biasanya ditulis oleh pengguna untuk berbagai pemikiran, menyampaikan pendapat, memberikan dukungan, atau bahkan menyampaikan kritik terhadap suatu konten atau topik. Komentar bisa memiliki berbagai bentuk, mulai dari singkat dan langsung hingga panjang dan mendalam.

Perkembangan pesat teknologi informasi dan penetrasi *internet* secara global telah menciptakan dunia maya yang semakin dinamis. Platform daring, seperti media sosial, forum, dan situs berita, menyediakan ruang bagi pengguna untuk

berpartisipasi dalam diskusi, memberikan tanggapan, dan berbagai informasi. Di tengah interaksi yang begitu intens, muncul tantangan baru terkait dengan pemahaman dan interpretasi konten, terutama ketika berkaitan dengan komentar yang bersifat sarkastik.

Komentar sarkasme adalah bentuk komunikasi yang menyampaikan pesan dengan cara yang bertentangan dengan arti harfiahnya, seringkali untuk menyindir atau mengkritik. Ini dapat terjadi dalam bentuk kata-kata atau nada suara yang bertentangan dengan makna sebenarnya, dan seringkali memerlukan pemahaman konteks untuk diinterpretasikan dengan benar (Mustofa, A., & Suhartono, D, 2021).

Komentar sarkastik seringkali disajikan dengan gaya bahasa yang bertentangan dengan makna literalnya, dan hal ini dapat merugikan dalam konteks *online* (Damayanti, E, 2021). Sarkasme dapat menyebabkan kebingungan, salah paham, atau bahkan konflik di antara pengguna platform. Oleh karena itu, pengembangan sistem otomatis mampu mengenali dan mengklasifikasikan komentar sarkastik menjadi suatu kebutuhan mendesak.

Analisis komentar sarkasme menjadi penting untuk memahami dan mengelola konten *online* dengan lebih efektif. Identifikasi komentar yang bersifat sarkastik membantu dalam meningkatkan moderasi konten, meminimalkan risiko kesalahpahaman, dan menciptakan lingkungan *online* yang lebih positif. Selain itu, analisis ini mendukung pengembangan sistem pemantauan konten yang cerdas dan dapat digunakan untuk meningkatkan akurasi analisis sentimen secara keseluruhan. Penelitian terhadap komentar sarkasme juga memberikan wawasan tambahan dalam pemahaman linguistik, berkontribusi pada pengembangan model kecerdasan buatan yang lebih canggih dalam memproses bahasa alami.

Berdasarkan permasalahan di atas, maka perlu dilakukannya sebuah penelitian untuk melakukan klasifikasi komentar sarkasme dengan metode pembelajaran mesin menggunakan hasil perancangan sistem dan pemanfaatan metode pembelajaran berbasis *kernel*. Metode pembelajaran mesin atau yang biasa disebut metode pembelajaran berbasis *kernel* ialah pendekatan dalam pembelajaran mesin yang menggunakan fungsi *kernel* untuk memproyeksikan data ke dimensi tinggi, memudahkan pemisahan *non-linier*. Salah satu dari metode pembelajaran mesin ini adalah *Support Vector Machines (SVM)* adalah algoritma klasifikasi dengan tujuan untuk menemukan fungsi pemisah yang dapat memisahkan dua set dari dua kelas yang berbeda (Silitonga & Damani, 2021).

Penelitian yang akan dilakukan oleh penulis menggunakan *Support Vector Machines (SVM)* untuk melakukan klasifikasi, dengan tujuan mengetahui komentar yang menjadi data uji apakah termasuk ke dalam jenis komentar sarkasme atau tidak. Berdasarkan uraian latar belakang di atas, maka dirancang sebuah penelitian berjudul “Klasifikasi komentar sarkasme menggunakan algoritma *Support Vector Machines (SVM)*”.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas yang sudah dijelaskan dan di paparkan jadi dibuat sebuah rumusan masalah yaitu bagaimana menerapkan algoritma *Support Vector Machine (SVM)* untuk melakukan analisis sentimen terhadap komentar sarkasme di media sosial *Reddit*.

1.3. Batasan Masalah

Adapun beberapa batasan masalah yang akan dibahas dalam menjalankan penelitian ini adalah:

1. Data dari media sosial/berita yaitu aplikasi *Reddit*.
2. Data yang diambil sebanyak 3000 data.
3. Data yang diambil dalam bentuk teks atau tulisan.
4. Algoritma yang digunakan dalam melakukan klasifikasi komentar sarkasme yaitu algoritma *Support Vector Machine (SVM)*.
5. Proses *preprocessing data* meliputi beberapa tahapan yaitu *cleaning data*, *case folding*, *normalisasi*, *stemming*, *filtering*, *tokenizing*, dan *cosine similarity*.
6. Analisis sentimen terhadap komentar sarkasme menggunakan aplikasi melalui perancangan pada *Google Colab*.
7. Data dibagi menjadi 3 bagian, *training* 80%, *testing* 10%, dan *validasi* 10%. Dengan pembagian *training* 2400 data, *testing* 300 data, dan *validasi* 300 data.

1.4. Tujuan Penelitian

1. Melakukan analisis sentimen terkait komentar sarkasme dengan algoritma *Support Vector Machines (SVM)* untuk menentukan tingkat akurasi dari algoritma tersebut dalam mengolah data.
2. Menerapkan algoritma *Support Vector Machine (SVM)* dalam melakukan analisis sentimen dengan tujuan untuk menguji serta melakukan implementasi algoritma *SVM* ini dalam pembahasan analisis sentimen pada data media sosial *Reddit*.

1.5. Manfaat Penelitian

Diharapkan penelitian ini memiliki manfaat untuk berbagai pihak. Adapun manfaat penelitian ini sebagai berikut:

1. Implementasi Algoritma *Support Vector Machine (SVM)* dalam analisis sentimen sarkasme dapat meningkatkan efektivitas moderasi konten pada platform daring. Hal ini dapat mengurangi dampak konten sarkastik yang mungkin merugikan dan menciptakan lingkungan *online* yang lebih positif.
2. Hasil penelitian ini diharapkan dapat memberikan wawasan dan pengetahuan serta pengertian sentimen komentar sarkasme terhadap media sosial *Reddit*.
3. Diharapkan hasil penelitian ini dapat menjadi referensi dan landasan untuk penelitian selanjutnya terutama pada bidang analisis sentimen.

BAB II

TINJAUAN PUSTAKA

2.1. Komentar

Dalam konteks media sosial, komentar merujuk pada respons atau tanggapan yang diberikan oleh pengguna terhadap suatu konten, seperti postingan, foto, atau video. Komentar berfungsi sebagai cara untuk berinteraksi, memberikan pendapat, atau berbagi pemikiran terhadap konten yang diposting oleh pengguna lain. Komentar dapat berupa *teks*, *emoji*, atau kombinasi keduanya, dan sering digunakan sebagai sarana untuk berkomunikasi, memberikan dukungan, atau memulai percakapan di berbagai platform media sosial. Komentar juga dapat mencerminkan beraneka ragam respon, mulai dari tanggapan positif hingga kritik, sehingga memainkan peran penting dalam membangun komunitas *online* dan menggambarkan beragam pandangan terhadap suatu konten atau topik tertentu. (Mustofa, A., & Suhartono, D. 2021).

Komentar di media sosial tidak hanya merupakan bentuk interaksi, tetapi juga mencerminkan dinamika sosial dan keberagaman pendapat di dalam komunitas *online*. Sebagai saluran komunikasi yang terbuka, komentar memberikan kesempatan kepada pengguna untuk berbagi pengalaman, menyatakan pendapat pribadi, memberikan wawasan tambahan terkait dengan konten yang dibagikan. Komentar juga dapat menjadi tempat untuk membangun koneksi antar pengguna, mendorong diskusi, atau memberikan apresiasi terhadap kreatifitas atau pemikiran orang lain. Di sisi lain, komentar juga dapat menjadi tempat potensial untuk konflik dan kontroversi, dengan adanya perbedaan pendapat yang dapat menciptakan

diskusi yang sengit. Oleh karena itu, komentar memiliki peran penting dalam membentuk atmosfer dan dinamika suatu komunitas *online*.

2.2. Jenis-jenis Komentar

Pada umumnya jenis-jenis komentar dapat bervariasi sesuai konteks dan tujuan komunikasinya. Berikut merupakan jenis komentar umum yang memiliki cirinya masing-masing.

1. Komentar Positif

Komentar positif menyampaikan apresiasi atau dukungan terhadap konten dengan menggunakan kata-kata seperti “keren” atau “bagus”. Biasanya, jenis komentar ini memberikan energi positif dan mendorong pembuat konten untuk lebih kreatif lagi. (Mustofa, A., & Suhartono, D. 2021).

2. Komentar Konstruktif

Komentar konstruktif memberikan masukan atau saran yang bersifat membangun untuk membantu peningkatan. Umumnya, jenis komentar ini diungkapkan dengan cara yang menghargai dan memberikan perspektif yang membantu.

3. Komentar Kritik

Komentar kritik menyampaikan pandangan kritis terhadap konten, baik dalam bentuk evaluasi atau saran perbaikan. Tujuannya adalah memberikan pemikiran mendalam dan konstruktif. (Mustofa, A., & Suhartono, D. 2021).

4. Komentar Humor

Komentar humor menyisipkan unsur lucu atau menghibur sebagai respons terhadap konten. Hal ini dapat menciptakan atmosfer yang santai dan positif di antara pembaca atau penonton.

5. Komentar Bertanya

Komentar bertanya mencakup pertanyaan yang ditujukan kepada pembuat konten atau anggota komunitas untuk mendapatkan informasi tambahan atau memulai interaksi.

6. Komentar Emosional

Komentar emosional mencerminkan perasaan pengguna terhadap konten, seperti kegembiraan, keterkejutan, atau kekecewaan. Komentar ini memberikan dimensi personal pada interaksi *online* (Mustofa, A., & Suhartono, D. 2021).

7. Komentar Spam

Komentar spam berisi informasi tidak relevan atau tautan yang tidak diinginkan. Jenis komentar ini dapat mengganggu dan sering kali dihasilkan oleh *bot* atau akun yang tidak sah.

8. Komentar *Trolling*

Komentar *trolling* ditujukan untuk menciptakan konflik atau memicu reaksi negatif. Tujuannya bisa bermacam-macam, mulai dari hiburan hingga niat merusak atmosfer positif.

9. Komentar Dukungan Sosial

Komentar dukungan sosial menawarkan dukungan emosional, simpati, atau solidaritas terhadap pengguna yang mungkin sedang menghadapi kesulitan atau berbagai pengalaman pribadi.

10. Komentar Menghina atau Pelecehan

Komentar menghina atau pelecehan bersifat merendahkan dan menyerang individu atau kelompok. Jenis komentar ini sering kali tidak etis dapat merusak atmosfer positif komunitas *online* (Mustofa, A., & Suhartono, D. 2021).

2.3. Analisis Sentimen

Analisis sentimen merupakan bidang penelitian berkelanjutan yang berada diantara berbagai bidang seperti *Data Mining*, *Natural Language Processing* (NLP) dan *Machine Learning* yang berfokus pada ekstraksi sentimen dalam sebuah kalimat berdasarkan isinya. (Tuhuteru, 2020).

Sentimen analysis merupakan suatu proses dalam data mining yang digunakan untuk mengidentifikasi dan mengekstrak suatu informasi dari suatu teks yang bertujuan untuk memahami *social sentiment* pada teks tersebut. *Sentiment analysis* juga digunakan untuk mendapatkan informasi mengenai sikap, pendapat dan emosi yang ada pada teks informasi tersebut. *Sentiment analysis* difokuskan untuk *review* klasifikasi berdasarkan polaritas. (Putri et al., 2022)

Sentimen analysis merupakan kajian tentang cara menyelesaikan dan memecahkan masalah dari berdasarkan opini masyarakat, sikap serta emosi suatu entitas, di mana entitas tersebut dapat mewakili individu. *Sentimen analysis* atau juga disebut *opinion mining* merupakan proses memahami, mengekstrak serta mengolah data tekstual secara otomatis guna mendapatkan informasi yang terkandung dalam suatu kalimat opini. Dilakukannya analisis sentimen ini bertujuan untuk melihat pendapat atau kecenderungan opini terhadap suatu masalah ataupun objek oleh seorang, apa memiliki kecenderungan positif, negatif, atau netral. (Fairuz et al., 2021).

2.4. Metode Pembelajaran Berbasis *Kernel*

Metode pembelajaran berbasis *kernel* adalah pendekatan dalam pembelajaran mesin yang menggunakan *kernel*, yaitu fungsi matematis, untuk meningkatkan kinerja algoritma. *Kernel* mengukur kemiripan antara pasangan data dalam ruang

fitur yang mungkin memiliki dimensi tinggi (Aditya, B., & Santoso, F. 2020). Algoritma *Support Vector Machine (SVM)* adalah salah satu algoritma pembelajaran yang umum menggunakan metode *kernel*. Dengan memanfaatkan *kernel*, *SVM* dapat memproyeksikan data ke ruang fitur yang lebih tinggi dan menemukan batas keputusan yang lebih kompleks di sana. Keuntungan utama pembelajaran berbasis *kernel* adalah kemampuannya untuk menangani masalah yang tidak *linear* di ruang fitur tinggi. Ini memungkinkan penanganan pola yang kompleks dan pemisahan data yang tidak *linear*. Beberapa jenis *kernel* yang umum digunakan meliputi *kernel linier*, *kernel polinomial*, *kernel radial basis function (RBF)*, dan *kernel sigmoid*. Meskipun memberikan hasil yang baik, terdapat risiko *overfitting* jika *kernel* digunakan tanpa penanganan yang tepat. *Tuning* parameter, seperti parameter *C* dan *gamma* pada *SVM* dengan *kernel RBF*, juga penting untuk mencapai kinerja optimal dalam metode pembelajaran berbasis *kernel*.

2.5. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang efektif untuk klasifikasi dan regresi. Fokus utama *SVM* adalah menemukan batas keputusan optimal yang memisahkan dua kelas data dalam ruang fitur. Dalam klasifikasi *biner*, *SVM* berupaya memaksimalkan *margin*, yaitu jarak antara batas keputusan dan vektor dukungan, yang merupakan titik-titik terdekat dari masing-masing kelas. Vektor dukungan memainkan peran kunci dalam perhitungan *margin* dan konstruksi batas keputusan, dan keberhasilan *SVM* terletak pada kemampuannya menemukan vektor dukungan yang dapat menggeneralisasi dengan baik terhadap data baru (Silitonga, P. D. ., & Damani, R. 2021).

Salah satu keunggulan utama *SVM* adalah kemampuannya menangani masalah klasifikasi yang kompleks, terutama dengan memanfaatkan teknik *kernel*. *Kernel* memungkinkan *SVM* untuk melakukan transformasi *nonlinier* terhadap data, memproyeksikannya ke ruang fitur yang lebih tinggi. Beberapa jenis *kernel* yang umum digunakan termasuk *kernel linier*, *polinomial*, dan *radial basis function (RBF)*. Meskipun *SVM* sangat efektif, proses penyetelan parameter seperti *C* dan *gamma* dapat menjadi kritis untuk mencapai kinerja optimal. Meskipun dirancang awalnya untuk klasifikasi *biner*, *SVM* dapat diperluas untuk menangani masalah klasifikasi multikelas dengan menggunakan pendekatan *one-vs-one* atau *one-vs-all*. Aplikasi *SVM* melibatkan berbagai bidang, seperti pengenalan wajah, klasifikasi teks, dan bioinformatika, membuatnya menjadi algoritma yang sangat berguna dalam pemecahan masalah dunia nyata.

2.5.1. Prinsip Dasar SVM

Berikut merupakan prinsip dasar dari *Support Vector Machine (SVM)* :

1. Prinsip SVM:

SVM bertujuan untuk membuat *hyperplane* terbaik yang memisahkan dua kelas dengan *margin* maksimum. *Hyperplane* ini harus memiliki jarak maksimum dari titik-titik pelatihan terdekat dari kedua kelas, yang disebut *support vectors*.

2. Hyperlane:

Hyperplane adalah batas keputusan yang memisahkan dua kelas. Pada kasus dua dimensi, *hyperplane* adalah garis lurus; pada kasus tiga dimensi, itu adalah bidang datar, dan pada dimensi yang lebih

tinggi, itu adalah hiperbidang. *SVM* mencari *hyperplane* yang memiliki margin maksimum.

3. *Margin*:

Margin adalah jarak antara *hyperplane* dan titik-titik *support vectors* terdekat dari kedua kelas. *SVM* berusaha untuk memaksimalkan *margin* ini karena hal ini dianggap sebagai indikator keberlanjutan model pada data baru dan mengurangi risiko *overfitting*.

4. *Support Vectors*:

Support vectors adalah titik-titik data pelatihan yang terletak paling dekat dengan *hyperplane*. Mereka adalah kunci dalam menentukan *hyperplane* dan *margin*. Hanya *support vectors* yang mempengaruhi posisi dan orientasi *hyperplane*.

5. Fungsi *Kernel*:

SVM menggunakan fungsi *kernel* untuk mentransformasi data ke dalam ruang fitur yang lebih tinggi, di mana pemisahan *linear* mungkin menjadi lebih mudah. Fungsi *kernel* memungkinkan *SVM* bekerja dengan baik bahkan jika data tidak *linier* terpisah secara *linier* di ruang fitur asli

6. *C*:

Parameter *C* pada *SVM* mengontrol *trade-off* antara menciptakan *margin* yang maksimal dan mengurangi kesalahan klasifikasi pada data pelatihan. Nilai *C* yang lebih tinggi akan membuat model lebih cenderung mengikuti data pelatihan, bahkan jika itu menyebabkan *margin* yang lebih kecil.

7. *Soft Margin SVM*:

SVM dapat diperluas untuk menangani kasus di mana data tidak sepenuhnya terpisah. Dalam hal ini, konsep *soft margin* digunakan, di mana beberapa pelanggaran kelas diizinkan dengan memberikan toleransi terhadap kesalahan klasifikasi pada data pelatihan.

8. *Linear vs. Non-Linear SVM*:

SVM bisa digunakan untuk pemisahan *linear* dan *non-linear*.

Pada kasus pemisahan *non-linear*, fungsi *kernel* digunakan untuk mentransformasikan data ke dalam dimensi yang lebih tinggi di mana pemisahan *linear* mungkin terjadi.

2.5.2. Penghitungan manual *Support Vector Machine (SVM)*

1. Hitung Bobot w :

$$w = \text{Mean (Kelas 1)} - \text{Mean (Kelas -1)}$$

2. Normalisasi Bobot:

$$\text{Normalized } w$$

3. Hitung *Margin*:

$$\text{Margin} = 2/w$$

4. Persamaan *Hyperlane*:

$$\text{Hyperlane: } w \cdot x + b = 0$$

2.6. *Preprocessing*

Preprocessing data adalah tahap pertama dalam *text mining* (Hermawati dkk., 2023). Proses ini bertujuan untuk mempersiapkan data agar dapat diproses selanjutnya dan meningkatkan kualitas data. *Preprocessing* penting dalam kegiatan

text mining karena jika data yang digunakan memiliki kualitas rendah, maka pengetahuan yang dihasilkan juga akan rendah (Hidayah dkk., 2021).

Preprocessing atau prapemrosesan data adalah langkah awal dalam pengolahan data yang bertujuan untuk mempersiapkan dan membersihkan data mentah sebelum dianalisis atau digunakan dalam model *machine learning*.

2.6.1. *Cleaning*

Cleaning atau pembersihan data adalah proses menghapus atau memperbaiki data yang kotor atau tidak valid. Ini termasuk:

- Menghapus duplikat data.
- Menghilangkan atau mengisi nilai-nilai yang hilang (*missing values*).
- Menghapus atau memperbaiki data yang tidak konsisten atau salah format.

2.6.2. *Case Folding*

Case Folding adalah proses mengubah semua huruf dalam teks menjadi huruf kecil (*lowercase*). Ini dilakukan untuk memastikan bahwa kata yang sama dengan huruf kapital berbeda dianggap sebagai entitas yang sama. Misalnya, “Data” dan “data” akan dianggap sama setelah *Case Folding*.

2.6.3. *Normalisasi*

Normalisasi adalah proses standar untuk memastikan data memiliki format dan struktur yang konsisten. Dalam teks, ini bisa termasuk:

- Mengubah teks menjadi bentuk dasar atau standar.
- Menghapus tanda baca, angka, atau karakter khusus yang tidak diperlukan.

2.6.4. *Stemming*

Stemming adalah proses mengurangi kata ke bentuk dasar atau katanya. Misalnya, kata “*running*”, “*runs*”, dan “*ran*” akan dikurangi menjadi “*run*”. Ini membantu dalam mengurangi variasi kata yang memiliki arti yang sama.

2.6.5. *Filtering*

Filtering adalah proses menyaring data untuk menghilangkan kata-kata atau simbol yang tidak diperlukan. Dalam teks, ini biasanya melibatkan penghapusan *stopwords* (kata umum seperti “*and*”, “*the*”, “*is*”) yang tidak memiliki nilai penting dalam analisis.

2.6.6. *Tokenizing*

Tokenizing adalah proses memecah teks menjadi unit-unit yang lebih kecil, seperti kata atau kalimat. Misalnya, kalimat “*I love data science*” akan dipecah menjadi [“*I*”, “*love*”, “*data*”, “*science*”]. Ini adalah langkah penting dalam analisis teks untuk memproses dan memahami teks.

2.6.7. *Cosine Similarity*

Cosine Similarity adalah ukuran untuk menentukan seberapa mirip dua dokumen teks dengan mengukur sudut kosinus antara vektor dalam ruang dimensi. *Cosine Similarity* dihitung sebagai:

$$\text{Cosine Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

Di mana A dan B adalah vektor dari dua dokumen yang dibandingkan. Nilai *Cosine Similarity* berkisar antara -1 dan 1, di mana 1 berarti dokumen sangat mirip, 0 berarti tidak ada kesamaan, dan -1 berarti dokumen sangat berlawanan.

2.7. *Reddit*

Reddit adalah sebuah situs *web* agregasi berita sosial, pemeringkatan konten, dan diskusi asal Amerika Serikat. Pengguna terdaftar mengirimkan konten ke situs tersebut seperti tautan, posting teks, gambar, dan video, yang kemudian di "*upvote*" atau "*downvote*" oleh setiap anggota lainnya (Aditya, B., & Santoso, F. 2020).

Reddit juga terkenal karena bahasa dan budaya *internet* yang khas, termasuk penggunaan istilah-istilah unik dan *meme* yang berkembang di dalam komunitasnya. Meskipun memiliki struktur tampilan sederhana, *Reddit* merupakan sumber informasi yang kuat dan tempat di mana pengguna dapat menemukan berbagai perspektif dan gagasan dari seluruh dunia. Dengan jutaan pengguna aktif setiap hari, *Reddit* menjadi salah satu platform sosial terbesar dan paling berpengaruh di *internet* saat ini. (Aditya, B., & Santoso, F. 2020).

Selain sebagai tempat untuk berbagi informasi dan pendapat, *Reddit* juga dikenal sebagai sumber hiburan yang kreatif dan unik. Banyak *subreddit* yang didedikasikan untuk *meme*, gambar lucu, atau konten kreatif lainnya, menciptakan lingkungan yang menghibur dan ramah. *Reddit* juga menjadi tempat bagi pengguna untuk mendapatkan jawaban atas pertanyaan mereka, dengan *subreddit* khusus yang menyediakan forum untuk konsultasi dan diskusi mengenai berbagai topik, termasuk pertanyaan teknis, saran, dan pengalaman pribadi. Keberagaman isinya, dari serius hingga humoris, membuat *Reddit* menjadi platform yang dinamis dan relevan untuk berbagai kebutuhan dan minat penggunanya. (Saputra, A., & Kurniawan, I. F. 2023).

2.8. Optimasi Parameter

Optimasi parameter adalah proses menyesuaikan parameter dalam suatu model atau sistem untuk mencapai performa terbaik sesuai dengan kriteria yang telah ditentukan. Parameter ini bisa berupa variabel atau pengaturan dalam algoritma, model matematika, atau sistem fisik yang mempengaruhi hasil akhir. Tujuannya adalah menemukan kombinasi parameter yang menghasilkan kinerja optimal, seperti akurasi tertinggi dalam model prediktif, efisiensi maksimum dalam proses manufaktur, atau kinerja terbaik dalam algoritma komputer (Arifianto, R. D., & Prabowo, R. H, 2019).

Proses optimasi parameter melibatkan beberapa langkah penting. Pertama, identifikasi parameter yang akan dioptimasi, di mana kita menentukan parameter mana yang memiliki pengaruh signifikan terhadap kinerja sistem. Kedua, penentuan fungsi objektif, yaitu metrik atau kriteria yang digunakan untuk menilai kinerja model atau sistem. Ketiga, eksplorasi ruang parameter dengan menguji berbagai kombinasi parameter untuk melihat bagaimana perubahan parameter mempengaruhi kinerja. Terakhir, pencarian optimal dengan menggunakan metode optimasi seperti *grid search*, *random search*, atau algoritma yang lebih canggih seperti algoritma genetika dan optimasi *Bayesian*.

Optimasi parameter sangat penting dalam berbagai bidang, termasuk pembelajaran mesin, pemodelan ilmiah, dan pengaturan sistem industri, karena dapat memastikan bahwa sistem berfungsi seefisien dan seefektif mungkin. Dalam pembelajaran mesin, misalnya, optimasi parameter dapat meningkatkan akurasi model prediktif, sementara dalam industri, hal ini dapat meningkatkan efisiensi

proses produksi. Dengan mengoptimalkan parameter, kita dapat mencapai hasil yang lebih baik dan mengurangi pemborosan sumber daya.

2.9. Penelitian Terkait

Untuk melakukan penelitian harus adanya sebuah acuan dari penelitian sebelumnya agar bisa diketahui kontribusi apa saja yang diberikan oleh penelitian tersebut. Berikut merupakan penelitian terdahulu yang terkait pada penelitian penulis:

Tabel 2. 1 Penelitian Terkait

No.	Nama Penulis	Judul Penelitian	Deskripsi
1.	(Kusnia, Kurniawan, 2019)	Analisis Sentimen Review Aplikasi Media Berita Online Pada Google Play Menggunakan Metode Algoritma Support Vector Machines (SVM) dan Naive Bayes	Berdasarkan hasil prosedur pengujian yang dilakukan terhadap <i>Review Aplikasi Media Berita Online Pada Google Play</i> Menggunakan Algoritma <i>Support Vector Machines (SVM)</i> dan Algoritma <i>Naive Bayes</i> . Hasil pengklasifikasian data opini kelas sentimen positif menjadi 5160 dan negatif menjadi 455. Hasil eksperimen <i>machine learning</i> dengan perbandingan data <i>training</i> 80% dan <i>testing</i> 20%, menghasilkan nilai

			<p>akurasi untuk algoritma <i>Support Vector Machines (SVM)</i> 88% dan algoritma <i>Naive Bayes</i> 87%. Untuk data sentimen <i>Review Aplikasi Media berita online</i> membutuhkan upaya yang tidak mudah saat tahap <i>processing</i> di awal yaitu bagian pelabelan.</p>
2.	(Putri, Nama, Sulistiono, 2022)	<p>Analisis Sentimen Kinerja Dewa Perwakilan Rakyat (DPR) Pada <i>Twitter</i> Menggunakan Metode <i>Naive Bayes Classifier</i></p>	<p>Penelitian ini berhasil menciptakan <i>sentiment analyzer</i> untuk mengamati <i>tweet</i> seputar kinerja DPR di <i>Twitter</i>. Hasil analisis, termasuk jumlah klasifikasi dan akurasi, dapat diakses melalui sebuah <i>website</i>. Dengan menggunakan algoritma <i>Naive Bayes</i>, sistem mencapai akurasi sebesar 80%, menunjukkan kemampuan memprediksi data <i>testing</i> secara akurat.</p>

			<p>Analisis juga menunjukkan bahwa hasilnya dapat diterapkan pada dataset baru tanpa perlu <i>labeling</i> terlebih dahulu. Dari 1546 data hasil <i>crawling</i>, sistem mengklasifikasikan 95 <i>tweet</i> sebagai positif, 693 sebagai netral, dan 758 sebagai negatif.</p>
3.	(Tuhuteru, 2020)	<p>Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berskala Besar Menggunakan Algoritma <i>Support Vector Machine</i></p>	<p>Analisis sentimen dengan algoritma <i>SVM</i> berhasil mengklasifikasikan sentimen masyarakat Kota Ambon berdasarkan komentar di <i>Facebook</i> dan <i>Twitter</i>. Hasil klasifikasi menunjukkan sentimen netral lebih mendominasi dari sentimen positif dan negatif, yaitu 45%. Sentimen negatif (27%) yang hampir mendekati sentimen positif (28%) menunjukkan masih ada masyarakat yang</p>

			<p>tidak puas dengan penerapan PSBB di Kota Ambon sehingga perlu diberikan edukasi dari pemerintah daerah. Penelitian selanjutnya masih perlu dilakukan, terutama setelah pandemi ini berakhir agar bisa membandingkan bagaimana sentimen masyarakat pra-PSBB dan pasca-PSBB. Metode yang digunakan juga masih perlu dikembangkan dengan tujuan untuk meningkatkan akurasi hasil klasifikasi, misalnya dengan menggunakan fitur atau operator lain seperti <i>Lexicon Based Features</i>, <i>n-Gram</i> dan <i>Ratio</i>.</p>
4.	(Primandani Arsi, Retno Waluyo, 2021)	Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia	Hasil pengujian pada 1.236 tweets tentang sentimen pemindahan ibu kota (404 positif, 832 negatif)

		<p>Menggunakan Algoritma <i>Support Vector Machine</i> (SVM).</p>	<p>menggunakan <i>SVM</i> menunjukkan hasil lebih baik dibanding metode sebelumnya (BM25 + KNN dan <i>Naive Bayes</i>), dengan akurasi 96,68%, <i>precision</i> 95,82%, <i>recall</i> 94,04%, dan AUC 0,979. Penanganan teks <i>processing</i> awal, khususnya <i>labeling</i>, membutuhkan usaha besar. Sentimen negasi dalam bahasa Indonesia belum optimal ditangani dalam penelitian ini. Penelitian selanjutnya perlu membandingkan analisis sentimen dengan dan tanpa penanganan negasi, menggunakan algoritma seperti <i>First Sentiment Word</i> (FSW), <i>Fixed Window Length</i> (FWL), atau <i>Rest of Sentences</i> (RoS).</p>
--	--	-------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

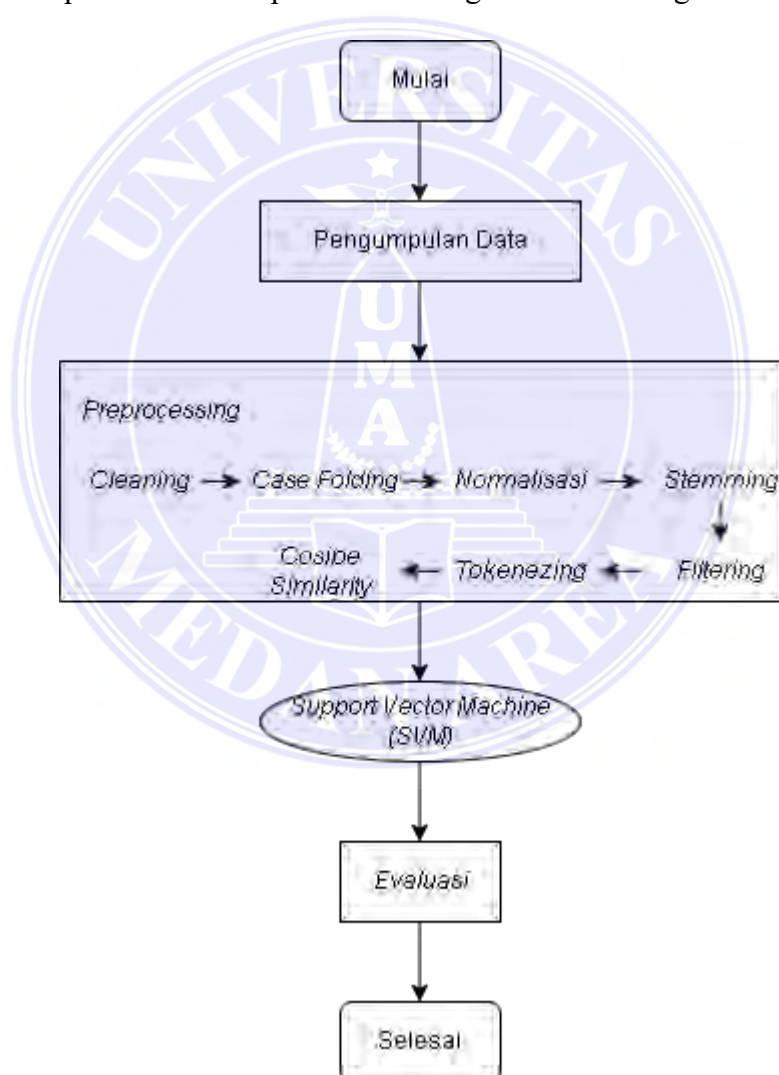
	<p>(Styawati, Nirwana Hendrastuty, Auliya Rahman Isnain, Ari Yanti Rahmadhani, 2021)</p>	<p>Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada <i>Twitter</i> Dengan Metode <i>Support Vector Machine</i></p>	<p>Penelitian ini menyimpulkan bahwa klasifikasi menggunakan metode <i>Support Vector Machine (SVM)</i> menghasilkan kelas netral sebesar 98,34%, kelas negatif 0,99%, dan kelas positif 0,66%. Evaluasi menunjukkan bahwa akurasi <i>kernel linear</i> mencapai 98,67% dengan <i>precision</i> 98%, <i>recall</i> 99%, dan <i>F1-Score</i> 98%, sementara akurasi <i>kernel RBF</i> adalah 98,34% dengan <i>precision</i> 97%, <i>recall</i> 98%, dan <i>F1-Score</i> 98%. Dari hasil ini, dapat disimpulkan bahwa sentimen masyarakat dari pengguna <i>Twitter</i> terhadap program kartu prakerja di masa pandemi lebih condong ke netral sebesar 98,34%.</p>
--	------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

BAB III

METODE PENELITIAN

3.1. Metode Penelitian

Dalam melakukan analisis sentimen terhadap komentar sarkasme menggunakan algoritma *Support Vector Machine (SVM)* akan dibuat beberapa tahapan dalam bentuk diagram alur, dan dapat mempermudah penyusunan dan penyelesaian penelitian ini dapat dilihat dari gambar 3.1 sebagai berikut:



Gambar 3. 1 Diagram Alur Penelitian

Keterangan tahapan pada gambar 3.1 alur penelitian dapat dilihat sebagai berikut:

1. Pengumpulan Data

Data yang sudah dikumpulkan dari media sosial *Reddit* akan dimasukkan pada *microsoft excel* untuk tahap proses selanjutnya.

2. *Preprocessing*

Data yang dikumpulkan merupakan data yang tidak terstruktur sehingga membutuhkan *Pre-Processing*. Proses ini menjadi sangat penting dalam analisis sentimen sehingga tidak bisa dilewatkan. Pada tahap ini juga memiliki beberapa proses yang harus dilakukan secara berurutan, sebagai berikut:

a) *Cleansing*

Pada proses ini, semua karakter di dalam *posts* yang bukan alfabet dihapus sehingga dapat mengurangi karakter yang tidak dikehendaki dan tidak memiliki arti dalam analisis sentimen. Karakter tersebut seperti angka, #, @, emoji, hingga link dari sebuah situs *website* yang ada di dalam suatu *posts*.

b) *Case Folding*

Proses mengubah setiap huruf di dalam *posts* menjadi *lowercase* atau semuanya dijadikan huruf kecil.

c) Normalisasi

Setiap *posts* yang dikirim oleh pengguna bisa dipastikan ada yang menggunakan Bahasa informal, disingkat atau bahkan diperpanjang tiap katanya. Tahap ini menjadi sangat penting dilakukan untuk memperbaiki kata-kata tersebut menjadi sebuah kata baku.

d) *Stemming*

Stemming merupakan proses mengubah kata yang berimbuhan menjadi kata dasar.

e) *Filtering*

Tahap ini dilakukan dengan membuang kata yang tidak memiliki makna atau *stopword*. *Stopword* akan dibuat dan disimpan dalam format teks (.txt) sebelum diinputkan di *Rapid Miner*.

f) *Tokenizing*

Pada tahap ini, sebuah *tweet* atau komentar dari pengguna yang sudah melalui tahap *cleansing* dan *case folding* dipisahkan dari kalimatnya menjadi sebuah kata.

g) *Cosine Similarity*

Cosine similarity adalah metode pengukuran kesamaan antara dua vektor dalam ruang multi-dimensi, khususnya digunakan dalam analisis teks. Nilai *cosine similarity* berkisar antara -1 (vektor arah berlawanan), 0 (vektor saling tegak lurus), hingga 1 (vektor arah sama). Semakin tinggi nilai *cosine similarity*, semakin mirip kedua vektor. Digunakan untuk perbandingan dokumen dan analisis teks.

3. Penerapan *Support Vector Machine (SVM)*

Setelah selesai pada tahap *Pre-Processing* selanjutnya akan dilakukannya penerapan atau implementasi algoritma *Support Vector Machine (SVM)* pada data yang telah dikumpulkan.

4. Evaluasi

Tahap evaluasi pada penerapan algoritma melibatkan penilaian performa dan efektivitasnya. Evaluasi ini mencakup beberapa aspek, seperti mengukur akurasi algoritma, menguji ketahanannya terhadap variasi data, menilai skalabilitasnya, dan memeriksa sejauh mana algoritma bersifat *robust*. Selain itu, evaluasi juga mempertimbangkan interpretabilitas hasil, perbandingan dengan metode lain, dan umpan balik dari pengguna atau pemangku kepentingan. Tujuan utama dari tahap evaluasi ini adalah memastikan bahwa algoritma memberikan hasil yang memuaskan, sesuai dengan harapan, dan dapat diandalkan dalam berbagai situasi penggunaan

3.2. Teknik Pengumpulan Data

Tahap pengumpulan data di sini dijelaskan diambil secara langsung oleh peneliti dari media sosial *Reddit*. Data yang diambil atau dikumpulkan berupa komentar berbahasa Inggris yang diperoleh melalui beberapa tahap pada media sosial *Reddit* tersebut. Data yang diambil atau dikumpulkan sebanyak 3000 komentar dan hanya mengambil data berupa teks tanpa gambar atau lain sebagainya. Data yang sudah dikumpulkan akan menjadi data input sistem untuk proses penelitian selanjutnya.

3.3. Alat dan Bahan Penelitian

Tahapan dalam menyelesaikan perancangan tentu membutuhkan alat maupun sistem pendukung. Berikut merupakan alat atau sistem yang dibutuhkan dalam perancangan penelitian tersebut.

3.3.1. Perangkat Keras

Perangkat keras merupakan komponen-komponen yang akan digunakan untuk penerapan metode dalam melakukan klasifikasi. Berikut ini merupakan perangkat keras yang akan digunakan beserta dengan spesifikasinya, yaitu:

Laptop, spesifikasi:

Processor : *Intel Core i7*

Memory : *Ram 16GB*

SSD : *Penyimpanan 512GB*

3.3.2. Perangkat Lunak

Software merupakan kumpulan perintah yang dijalankan oleh program untuk mengoperasikan *komputer*. Dalam hal ini, *software* yang akan digunakan untuk merancang program klasifikasi, yaitu:

1. *Kaggle*
2. *Google Colab*
3. *Python 3.11*
4. *Windows 11*

3.4. Analisis Kebutuhan Perancangan

Analisis kebutuhan perancangan untuk studi "Analisis Sentimen terhadap Komentar Sarkasme menggunakan Algoritma *Support Vector Machine*" memulai proses dengan mengidentifikasi tujuan bisnis yang ingin dicapai melalui analisis sentimen. Dalam hal ini, perlu dipahami apakah penelitian bertujuan untuk meningkatkan pemahaman respon konsumen, memperbaiki strategi pemasaran, atau mencapai tujuan bisnis lainnya. Setelah itu, fungsionalitas sistem yang diperlukan harus diidentifikasi dengan jelas, termasuk kemampuan sistem dalam

mengenali dan mengklasifikasikan komentar sebagai sarkastik atau tidak, serta bagaimana interaksi dengan pengguna akan diatur.

Tahap berikutnya melibatkan identifikasi kebutuhan *non-fungsional*, seperti kinerja, keamanan, dan skalabilitas, untuk memastikan bahwa sistem dapat beroperasi secara efisien dan aman. Analisis kebutuhan teknis kemudian mencakup pemilihan algoritma, dengan penjelasan mendalam mengenai keputusan untuk menggunakan *Support Vector Machine*, serta konfigurasi vektor fitur dan parameter *SVM* yang relevan. Proses ini juga mencakup definisi dataset dan langkah-langkah *preprocessing*, dengan fokus pada cara sistem mengatasi kompleksitas bahasa alami dan sarkasme dalam data yang diolah.

Desain arsitektur sistem menjadi langkah penting selanjutnya, di mana aspek-aspek seperti penyimpanan data, pelatihan model, dan pengelolaan input komentar dari pengguna atau sumber data lainnya harus direncanakan dengan cermat. Pengujian sistem mencakup pengujian skenario untuk memastikan akurasi dan keandalan analisis sentimen terhadap komentar sarkasme, dan metrik evaluasi seperti *presisi*, *recall*, dan *F1-score* perlu diperhitungkan. Rencana manajemen kesalahan dan penanganan eksepsi dirancang untuk mengatasi situasi di mana sistem mungkin menghadapi kesulitan atau mengalami kegagalan dalam implementasi. Dokumentasi menyeluruh dari seluruh proses ini disiapkan untuk memfasilitasi pemahaman dan pengembangan berkelanjutan. Dengan demikian, analisis kebutuhan perancangan membentuk landasan kokoh bagi pengembangan solusi yang tepat dan efektif dalam konteks penelitian ini.

3.5. Pelatihan dan Pengujian *Support Vector Machines (SVM)*

Proses pelatihan algoritma *Support Vector Machine (SVM)* melibatkan beberapa langkah kunci untuk menghasilkan model yang dapat memisahkan data dengan baik. Berikut merupakan proses pelatihan algoritma *Support Vector Machine (SVM)*:

3.5.1. Pemilihan *Kernel* dan Parameter:

SVM menggunakan *kernel* untuk mengubah data ke dalam dimensi yang lebih tinggi. Pemilihan *kernel* harus disesuaikan dengan karakteristik data, seperti *linear*, *polinomial*, atau fungsi basis *radial/RBF*, dan parameter *kernel* harus disesuaikan dengan baik.

3.5.2. Pemilihan Data Latih:

Dataset yang dipilih mencakup beragam contoh yang mewakili berbagai karakteristik dan variasi data yang diperlukan untuk melatih model, sehingga model dapat mempelajari dengan baik berbagai situasi yang mungkin dihadapi saat digunakan.

3.5.3. Representasi Data:

Dilakukan cara pemilihan untuk mengubah data menjadi vektor fitur agar dapat dimengerti oleh *SVM*. Hal ini melibatkan seleksi fitur serta kemungkinan normalisasi atau penskalaan untuk memastikan konsistensi dalam pengukuran.

3.5.4. Pelatihan Model:

Pada tahap utama proses, *SVM* dilatih untuk menemukan *hyperplane* optimal yang memisahkan kelas-kelas dalam ruang fitur. Tujuannya adalah untuk memaksimalkan *margin*, yaitu jarak antara *hyperplane* dan contoh terdekat dari setiap kelas, sambil juga mengizinkan sejumlah kecil kesalahan klasifikasi.

3.5.5. Optimasi Parameter:

Penyetelan parameter dilakukan untuk memastikan model *SVM* mencapai kinerja yang optimal. Ini melibatkan penyetelan parameter C , yang mengontrol keseimbangan antara penalti kesalahan klasifikasi dan lebar *margin*, serta penyetelan parameter *kernel* jika digunakan.

3.5.6. Validasi Model:

Dataset validasi atau teknik validasi silang (*cross-validation*) dilakukan untuk mengevaluasi kinerja model di luar dataset pelatihan. Hal ini membantu dalam mengidentifikasi apakah model cenderung mengalami *overfitting* atau *underfitting*.

3.5.7. Uji Model:

Setelah melalui tahap pelatihan dan validasi, model dapat diuji menggunakan dataset yang belum pernah dilihat sebelumnya. Proses ini membantu dalam mengevaluasi kemampuan model untuk melakukan klasifikasi pada data yang tidak terlibat dalam proses pelatihan dan validasi sebelumnya.

3.5.8. Optimasi Lanjutan:

Jika hasil evaluasi menunjukkan kebutuhan perbaikan, proses pelatihan dan penyetelan parameter dapat diulang untuk meningkatkan kinerja model.

Proses pengujian algoritma *Support Vector Machine (SVM)* merupakan langkah penting untuk mengukur kinerja model yang telah dilatih. Berikut adalah penjelasan singkat mengenai proses pengujian *SVM*:

3.5.9. Pemilihan Dataset Uji

Dalam memilih dataset uji, dilakukan pemilihan dataset uji yang berbeda dari dataset pelatihan untuk menguji model. Dataset uji sebaiknya mencerminkan situasi dunia nyata yang mungkin dihadapi oleh model saat digunakan.

3.5.10. Representasi Data

Metode yang diterapkan sama untuk mengonversi data uji seperti yang digunakan selama pelatihan. Pastikan bahwa fitur dan skala data konsisten dengan yang telah ditemui selama proses pelatihan.

3.5.11. Uji Model:

Model *SVM* yang digunakan yang sudah terlatih untuk melakukan prediksi pada dataset uji. Lakukan perhitungan metrik evaluasi seperti akurasi, *presisi*, *recall*, *F1-score*, atau metrik lain yang relevan, tergantung pada jenis masalah klasifikasi yang dihadapi.

3.5.12. Validasi Performa:

Proses validasi performa model dilakukan dengan mengukur kinerjanya pada dataset uji. Perhatikan apakah model cenderung mengalami *overfitting* atau *underfitting* dan apakah generalisasinya cukup baik untuk memenuhi tujuan aplikasi.

3.5.13. Fine-Tuning (Penyetelan Lanjutan):

Pada saat yang dibutuhkan, Penyetelan lanjutan dilakukan pada parameter *SVM* atau kumpulan fitur untuk meningkatkan kinerja model pada dataset uji. Tujuannya adalah untuk memperbaiki ketidakcocokan atau ketidakakuratan yang dapat diidentifikasi selama pengujian.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Kesimpulan dari penelitian yang telah dilakukan ini melakukan analisis sentimen terhadap komentar sarkasme menggunakan algoritma *Support Vector Machine (SVM)*. Algoritma ini dapat digunakan sebagai pilihan karena kemampuan dari algoritma ini dalam melakukan analisis serta menyediakan fitur-fiturnya cukup baik. Pada evaluasi model data nya terbagi menjadi 3 bagian yaitu *training*, *testing*, *validation* yang berfokus pada akurasi yang dihasilkan. Hasil dari pengujian yang telah dilakukan menunjukkan bahwa *Support Vector Machine (SVM)* memberikan akurasi yang cukup baik yaitu 65.75%.

5.2. Saran

Beberapa saran yang dapat diberikan oleh peneliti demi kelanjutan dan pengembangan penelitian ini sebagai berikut:

1. Melakukan perbaikan serta pemilihan data yang lebih baik lagi sehingga dapat dibaca atau diproses oleh algoritma dengan proses yang lebih baik sehingga dapat menghasilkan akurasi yang lebih tinggi pula.
2. Melakukan perbaikan fitur dengan menyisipkan proses *Feature Engineer* dan melakukan kombinasi *SVM* dengan model lain sebagai contoh *Random Forest* atau *Neural Networks*.

DAFTAR PUSTAKA

- Aditya, B., & Santoso, F. (2020). "Exploring Sarcasm Detection in Indonesian Tweeters Using Support Vector Machine." *International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 456-461.
- Alfaris, L., Yusuf, J., & Afandi, R.. (2022). Analisis Ragam Gaya Bahasa Sarkasme dalam Tuturan Komentator Sepak Bola Piala AFF 2021 di Instagram Ngapakfootball. *Jurnal Pendidikan Tambusai*, 6(2), 12011–12016. <https://doi.org/10.31004/jptam.v6i2.4360>
- Arifianto, R. D., & Prabowo, R. H. (2019). "Penggunaan Support Vector Machine dalam Analisis Sentimen pada Komentar Media Sosial Berbahasa Indonesia." *Jurnal Ilmiah Informatika*, 7(1), 45-56.
- Damayanti, E. (2021). Ragam Bahasa Sarkasme Pada Percakapan Remaja Di Desa Kepung Kecamatan Kepung Kabupaten Kediri. *Innovative: Journal Of Social Science Research*, 1(1), 27–34. <https://doi.org/10.31004/innovative.v1i1.6>
- Fairuz, A. L., Ramadhani, R. D., & Tanjung, N. A. F. (2021). Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter. *Jurnal Dinda*, 1.
- Ghiffari, Rizaldi Naufal (2022) *Pengaruh Pendeteksian Sarkasme Pada Analisis Sentimen Masyarakat Terhadap Politikus Indonesia Menggunakan Random Forest Classifier*. Other thesis, Universitas Komputer Indonesia.
- Mustofa, A., & Suhartono, D. (2021). "Sarkasme dalam Komentar Media Sosial: Studi Kasus Twitter." *Jurnal Linguistik Terapan*, 8(2), 87-100.
- Putri, D. D., Nama, G. F., & Sulistiono, W. E. (2022). ANALISIS SENTIMEN KINERJA DEWAN PERWAKILAN RAKYAT (DPR) PADA TWITTER

MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER. *Jurnal Informatika Dan Teknik Elektro Terapan (JITET)*, 10.

SAFINA, SAFINA (2020) *IMPLEMENTASI METODE ANALISIS SENTIMEN DAN DETEKSI SARKASME MENGGUNAKAN DATA MEDIA SOSIAL*. Skripsi thesis, Universitas Hasanuddin.

Saputra, A., & Kurniawan, I. F. (2023). "Implementasi Algoritma Support Vector Machine pada Analisis Sentimen Komentar Pengguna Instagram." *Jurnal Sistem Informasi*, 4(1), 15-24.

Silitonga, P. D. ., & Damani, R. (2021). Perbandingan Algoritma k-Nearest Neighbors (k-NN) dan Support Vector Machines (SVM) untuk Klasifikasi Pengenalan Citra Wajah. *Jurnal ICT: Information Communication & Technology*, 20.

Styawati, Andi Nurkholis, Zaenal Abidin, & Heni Sulistiani. (2021). Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(5), 904 - 910. <https://doi.org/10.29207/resti.v5i5.3380>

Tuhuteru, H. (2020). Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine. *INFORMATION SYSTEM DEVELOPMENT (ISD)*, 5.

Ulfatun, U. (2021). Analisis Penggunaan Gaya Bahasa Sarkasme Netizen di Media Sosial Instagram. *Jurnal Onoma: Pendidikan, Bahasa, Dan Sastra*, 7(2), 411-423. <https://doi.org/10.30605/onoma.v7i2.1255>

Wibawa, Gentra Aria (2023) *Pendeteksian Sarkasme Pada Analisis Sentimen Terhadap Selebgram Menggunakan Support Vector Machine Dan K-Nearest Neighbors*. Other thesis, Universitas Komputer Indonesia.



LAMPIRAN

```

import numpy as np
import pandas as pd
import nltk
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize
import re
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
import matplotlib.pyplot as plt
import time

from google.colab import files

# Download NLTK data
nltk.download('punkt')
nltk.download('stopwords')

# Upload data
uploaded = files.upload()

# Membaca file CSV menjadi dataframe
df = pd.read_csv('Sentimen3_cleaned.csv')

# Define preprocessing functions
def cleansing(text):

```

```

if isinstance(text, str):
    text = re.sub(r'http\S+', '', text) # Remove URLs
    text = re.sub(r'^[a-zA-Z\s]', '', text) # Remove non-alphabet characters
    return text

def case_folding(text):
    return text.lower()

def normalisasi(text):
    abbreviation_mapping = {
        "w/": "with",
        "w/o": "without",
        # Add more abbreviations and their replacements
    }
    for abbreviation, full_form in abbreviation_mapping.items():
        text = text.replace(abbreviation, full_form)
    text = re.sub(r'^\w\s', '', text) # Remove special punctuation
    return text

def stemming(text):
    stemmer = PorterStemmer()
    return ' '.join([stemmer.stem(word) for word in word_tokenize(text)])

def filtering(text):
    stop_words = set(stopwords.words('english'))
    return ' '.join([word for word in word_tokenize(text) if word not in stop_words])

def tokenizing(text):
    return word_tokenize(text)

```



```

# Ensure all comments are strings
df['comment'] = df['comment'].astype(str)

# Apply preprocessing steps
# 1. Cleansing
df['cleansed_comment'] = df['comment'].apply(cleansing)
df.to_csv('data_cleansed.csv', index=False)
print("Cleansing done and saved to data_cleansed.csv")
print(df[['comment', 'cleansed_comment']].head())

# 2. Case Folding
df['folded_comment'] = df['cleansed_comment'].apply(case_folding)
df.to_csv('data_folded.csv', index=False)
print("Case folding done and saved to data_folded.csv")
print(df[['cleansed_comment', 'folded_comment']].head())

# 3. Normalisasi
df['normalized_comment'] = df['folded_comment'].apply(normalisasi)
df.to_csv('data_normalized.csv', index=False)
print("Normalization done and saved to data_normalized.csv")
print(df[['folded_comment', 'normalized_comment']].head())

# 4. Stemming
df['stemmed_comment'] = df['normalized_comment'].apply(stemming)
df.to_csv('data_stemmed.csv', index=False)
print("Stemming done and saved to data_stemmed.csv")
print(df[['normalized_comment', 'stemmed_comment']].head())

# 5. Filtering
df['filtered_comment'] = df['stemmed_comment'].apply(filtering)
df.to_csv('data_filtered.csv', index=False)

```

```

print("Filtering done and saved to data_filtered.csv")
print(df[['stemmed_comment', 'filtered_comment']].head())

# 6. Tokenizing
df['tokenized_comment'] = df['filtered_comment'].apply(tokenizing)
df.to_csv('data_tokenized.csv', index=False)
print("Tokenizing done and saved to data_tokenized.csv")
print(df[['filtered_comment', 'tokenized_comment']].head())

# Representasi Data dengan TF-IDF
tfidf_vectorizer = TfidfVectorizer()
tfidf_matrix = tfidf_vectorizer.fit_transform(df['filtered_comment'])

# Menghitung Cosine Similarity
cos_sim = cosine_similarity(tfidf_matrix, tfidf_matrix)
cos_sim_df = pd.DataFrame(cos_sim)
cos_sim_df.to_csv('cosine_similarity_matrix.csv', index=False)
print("Cosine similarity matrix calculated and saved to
cosine_similarity_matrix.csv")
print(cos_sim_df.head())

# Save final preprocessed data
df.to_csv('data_preprocessed.csv', index=False)
print("Final preprocessing done and saved to data_preprocessed.csv")
print(df[['comment', 'filtered_comment']].head())

# Install necessary package
!pip install scikit-optimize
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.SVM import SVC

```

```

from sklearn.metrics import classification_report, accuracy_score,
confusion_matrix

from sklearn.feature_extraction.text import TfidfVectorizer

import seaborn as sns

import matplotlib.pyplot as plt

import numpy as np

from sklearn.model_selection import GridSearchCV, RandomizedSearchCV

from skopt import BayesSearchCV

# Load the preprocessed data
df = pd.read_csv('data_filtered.csv')

# Handle NaN values in 'filtered_comment'
df['filtered_comment'].fillna("", inplace=True)

# Ensure 'filtered_comment' and 'label' columns are available
if 'filtered_comment' not in df.columns or 'label' not in df.columns:
    raise ValueError("DataFrame must contain 'filtered_comment' and 'label'
columns")

# Pemilihan Data Latih dan Data Uji
X_train, X_test, y_train, y_test = train_test_split(df['filtered_comment'],
df['label'], test_size=0.2, random_state=42)

# Representasi Data dengan TF-IDF
tfidf_vectorizer = TfidfVectorizer()
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)

def evaluate_model(model, X_test_tfidf, y_test, title):
    y_pred = model.predict(X_test_tfidf)
    accuracy = accuracy_score(y_test, y_pred)

```

```

print(f"Akurasi Model ({title}):", accuracy)
print(f"\nLaporan Klasifikasi ({title}):")
print(classification_report(y_test, y_pred))
conf_matrix = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(10, 7))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues')
plt.xlabel('Prediksi')
plt.ylabel('Aktual')
plt.title(f'Confusion Matrix ({title})')
plt.show()

# Kernel list
kernels = ['linear', 'poly', 'RBF', 'sigmoid']

for kernel in kernels:
    print(f"\nKernel: {kernel}\n")

    # Train SVM model with default parameters
    SVM_model = SVC(kernel=kernel, C=1.0)
    SVM_model.fit(X_train_tfidf, y_train)
    evaluate_model(SVM_model, X_test_tfidf, y_test, f'Default Parameters,
Kernel={kernel}')

# Grid Search
param_grid = {
    'C': [0.1, 1, 10, 100]
}
if kernel != 'linear':
    param_grid['gamma'] = ['scale', 'auto']
grid_search = GridSearchCV(estimator=SVC(kernel=kernel),
param_grid=param_grid, cv=5, n_jobs=-1, verbose=2)
grid_search.fit(X_train_tfidf, y_train)

```

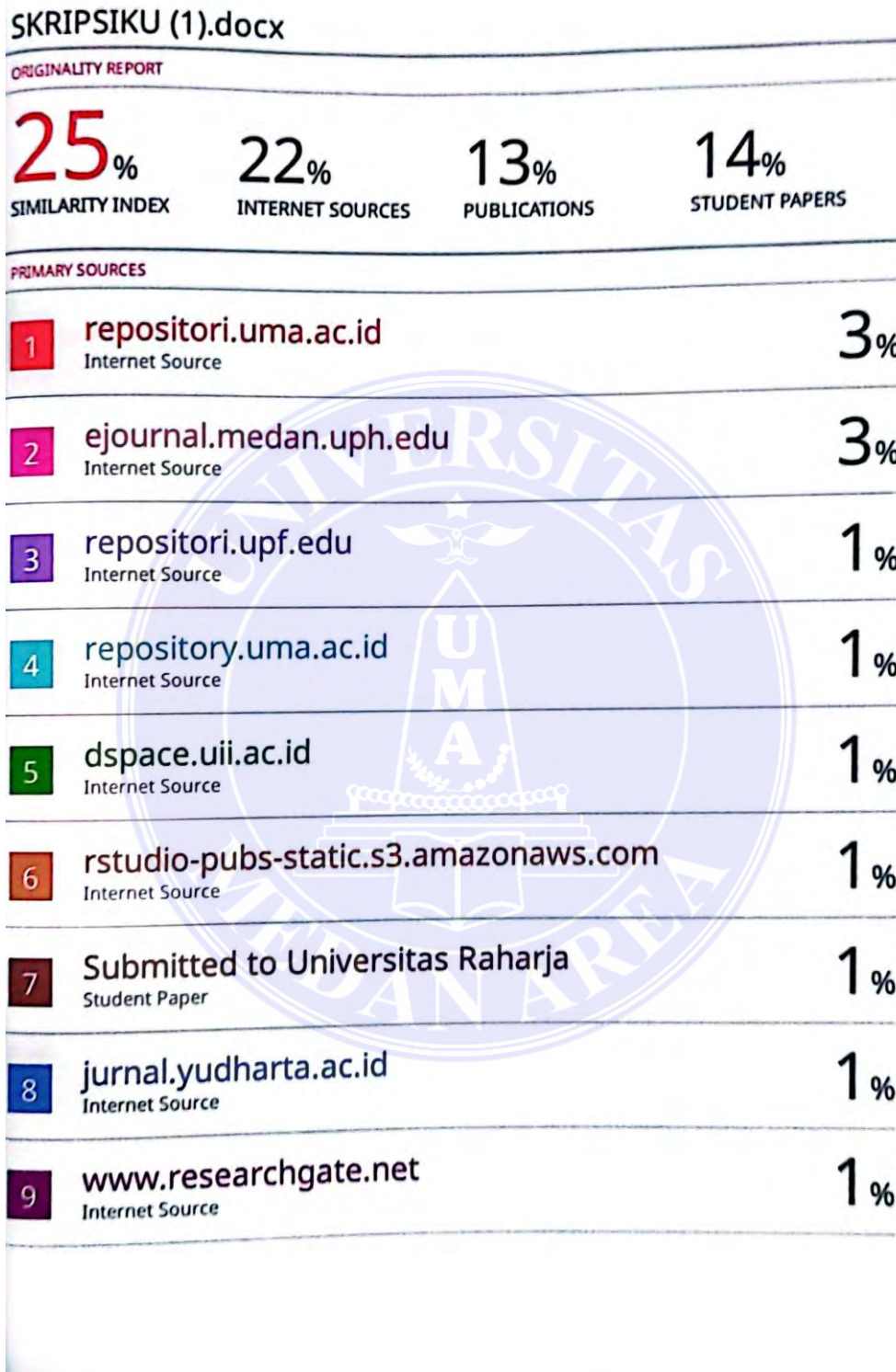
```

best_grid = grid_search.best_estimator_
evaluate_model(best_grid, X_test_tfidf, y_test, f'Grid Search,
Kernel={kernel}')

# Random Search
random_search = RandomizedSearchCV(estimator=SVC(kernel=kernel),
param_distributions=param_grid, n_iter=10, cv=5, n_jobs=-1, verbose=2,
random_state=42)
random_search.fit(X_train_tfidf, y_train)
best_random = random_search.best_estimator_
evaluate_model(best_random, X_test_tfidf, y_test, f'Random Search,
Kernel={kernel}')

# Bayesian Optimization
param_space = {
    'C': (0.1, 100)
}
if kernel != 'linear':
    param_space['gamma'] = ['scale', 'auto']
bayes_search = BayesSearchCV(estimator=SVC(kernel=kernel),
search_spaces=param_space, n_iter=10, cv=5, n_jobs=-1, verbose=2,
random_state=42)
bayes_search.fit(X_train_tfidf, y_train)
best_bayes = bayes_search.best_estimator_
evaluate_model(best_bayes, X_test_tfidf, y_test, f'Bayesian Optimization,
Kernel={kernel}')

```



UNIVERSITAS MEDAN AREA

FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎ (061) 7366878, 7360168, 7364348, 7366781, Fax. (061) 7366998 Medan 20223
Kampus II : Jalan Seiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 46/FT.6/01.10/1/2024
Lamp : -
Hal : **Perubahan Judul Tugas Akhir**

16 Januari 2024

Yth, Pembimbing Tugas Akhir
Dr. Arnes Sembiring, ST, M. Kom
di
Tempat

Dengan hormat, Sehubungan dengan adanya perubahan judul tugas akhir maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa tersebut :

N a m a : Zulnandi Zidan Lubis
N P M : 198160072
Jurusan : Teknik informatika

Maka dengan hormat kami mengharapkan kesediaan saudara :

Dr. Arnes Sembiring, ST, M. Kom (Sebagai Pembimbing)

Adapun Tugas Akhir Skripsi berjudul :

"Analisis Sentimen terhadap komentar Sarkasme menggunakan Algoritma *Support Vector Machine*".

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.


Dekan,
Dr. Eng. Supriatno, ST, MT



UNIVERSITAS MEDAN AREA

FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎ (061) 7366878, 7360168, 7364348, 7366781, Fax (061) 7366998 Medan 2022
 Kampus II : Jalan Setiabudi Nomor 79 / Jalan Set Serayu Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8226331 Medan 20122
 Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 221/FT.6/01.10/IV/2024

29 April 2024

Lamp

: -

Hal

: Penelitian Dan Pengambilan Data Tugas Akhir

Yth. Wakil Rektor Bidang Mutu Sumber Daya Manusia dan Perekonomian

Jln. Kolam No.1

Di

Medan

Dengan hormat, kami mohon kesediaan bapak kiranya berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

NO	N A M A	N P M	PRODI
1	Zulnandi Zidan Lubis	198160072	Teknik informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir di **Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area.**

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan Ilmiah dan Skripsi, yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul :

Analisis Sentimen terhadap Komentar Sarkasme menggunakan Algoritma *Support Vector Machine (SVM)*.

Mohon kiranya tanggal Surat Izin Pengambilan Data Tugas Akhir agar disesuaikan dengan tanggal Terbitnya Surat ini.


Atas perhatian dan kerja sama yang baik diucapkan terima kasih.


 Dekan

 Dr. Eng. Satriatno, ST, MT

Tembusan :

1. Ka. BPMPP
2. Mahasiswa
3. File

**UNIVERSITAS MEDAN AREA**
Kampus I : Jalan Kolam Nomor 1 Medan Estate ☎ (061) 7360168, 7366878, 7364348 📠 (061) 7368012 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 8225602 📠 (061) 8226331 Medan 20122
Website: www.uma.ac.id E-Mail: univ_medanarea@uma.ac.id

SURAT KETERANGAN
Nomor : 946/UMA/B/01.7/VII/2024



Rektor Universitas Medan Area dengan ini menerangkan bahwa :

Nama	: Zulnandi Zidan Lubis
No.Pokok Mahasiswa	: 198160072
Program Studi	: Teknik Informatika

Benar telah selesai Pengambilan Data di Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area dengan Judul Skripsi "**Analisis Sentimen Terhadap Komentar Sarkasme Menggunakan Algoritma Support Vector Machine (SVM)**".

Dan kami harapkan Data tersebut kiranya dapat membantu yang bersangkutan dalam penyusunan Skripsi dan dapat bermanfaat bagi mahasiswa khususnya Fakultas Teknik Universitas Medan Area.

Demikian surat ini diterbitkan untuk dapat digunakan seperlunya

Medan, 03 Juli 2024.
a.n Rektor
Wakil Rektor Bidang Mutu SDM & Ekonomiah,


Dr. Dedi Sahputra, S.Sos, MA

CC :
- Arsip