

**ANALISIS EMOSI *TWITTER* MENGGUNAKAN METODE SELEKSI  
FITUR *MUTUAL INFORMATION* DAN ALGORITMA *SUPPORT VECTOR*  
*MACHINE (SVM)***

Diajukan sebagai Syarat untuk Menyelesaikan  
Pendidikan Program Strata-1 pada  
Jurusan Teknik Informatika



Oleh :

Ahmad Marzuqi Yasykur Luthfi

NIM : 09021381823092

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2023**

**LEMBAR PENGESAHAN SKRIPSI**

**ANALISIS EMOSI PADA *TWITTER* MENGGUNAKAN METODE  
SELEKSI FITUR *MUTUAL INFORMATION* DAN ALGORITMA  
*SUPPORT VECTOR MACHINE (SVM)***

Oleh :

Ahmad Marzuqi Yasykur Luthfi  
NIM: 09021381823092

Palembang, 14 Juni 2023

Pembimbing I,



Samsuryadi, M.Kom., Ph.D.  
NIP. 197102041997021003

Pembimbing II,



Rizki Kurniati, M.T.  
NIP. 199107122019032016

Mengetahui,  
Ketua Jurusan Teknik Informatika



Akiz Syahrini Utami, M.Kom.  
NIP. 197812222006042003

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Jum'at tanggal 14 April 2023 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Ahmad Marzuqi Yasykur Luthfi

NIM : 09021381823092


Judul : Analisis Emosi *Twitter* Menggunakan Metode Seleksi Fitur

*Mutual Information* dan Algoritma *Support Vector Machine* (SVM)

dan dinyatakan LULUS.

1. Ketua

Osvari Arsalan, M.T.  
NIP. 198806282018031001



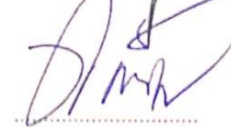
2. Penguji

Novi Yusliani, M.T.  
NIP. 198211082012122001



3. Pembimbing I

Samsuryadi, M.Kom., Ph.D.  
NIP. 197102041997021003



4. Pembimbing II

Rizki Kurniati, M.T.  
NIP. 199107122019032016



Mengetahui,  
Ketua Jurusan Teknik Informatika

Alw Syahrini Utami, M.Kom.  
NIP. 197812222006042003

## STATEMENT PAGE

The undersigned below:

Name : Ahmad Marzuqi Yasykur Luthfi  
NIM : 09021381823092  
Study Program : Informatics Engineering  
Title : Emotional Analysis on Twitter Using the Mutual Information  
Feature Selection Method and the Support Vector Machine  
(SVM) Algorithm

Checking result *Software Ithenticate/Turnitin* : 17%

Stating that my final project report is the result of my own work and not the result of plagiarism/plagiarism. If plagiarism or plagiarism is found in this final report, then I am willing to accept academic sanctions from Sriwijaya University.

Thus, I make this statement truthfully and without coercion from anyone.



Palembang, 7 Maret 2023



Ahmad Marzuqi Yasykur Luthfi  
NIM. 09021381823092

Motto :

- “Kamu tidak boleh melakukan sesuatu secara berbeda hanya karena berbeda. Sesuatu itu harus lebih baik.” – Elon Musk
- “Usaha dan doa tergantung pada cita-cita. Manusia tiada memperoleh selain apa yang telah diusahakannya.” – Jalaludin Rumi
- “Jangan jadikan ujian sebagai penghambat untuk maju. Nikmati dan jalani saja.”

Kupersembahkan karya tulis ini kepada :

- Allah SWT.
- Kedua Orang Tua dan Keluarga
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

## ABSTRACT

*Most texts have a large number of features. However, the features contained in the text mostly have a low level of relevance and even contain noise which can later reduce the accuracy of the results. Feature selection is used to reduce the dimensions of feature space by weighting all features then features with lower weights than threshold will be eliminated. It aims to improve the accuracy and efficiency of computational time in the text classification process. In this research, selection method Mutual Information were used in the text classification process in the form of Twitter emotion sentiment using the Support Vector Machine (SVM) algorithm. Then, a comparative analysis will be carried out on each classification model based on the evaluation results obtained. The results showed that the use of the feature selection method was able to increase accuracy and reduce computation time. The use of the Mutual Information feature selection method on the SVM algorithm with a linear kernel give the best performance with average of accuracy 0.71, precision 0.72, recall 0.71, f-measure 0.71 and computation time 1.17 seconds.*

*Key Word : Text Classification, Feature Selection, Support Vector Machine (SVM)*

## ABSTRAK

Emosi adalah hal yang bersifat umum dan penting dalam kehidupan yang membentuk perilaku manusia. Emosi seseorang bisa diidentifikasi menggunakan data tekstual yaitu berupa teks, karena teks dapat digunakan untuk berkomunikasi dan menyampaikan informasi. Seleksi fitur digunakan untuk mengurangi dimensi ruang fitur dengan cara melakukan pembobotan pada semua fitur kemudian fitur dengan bobot yang kurang dari ambang batas akan dieliminasi. Hal ini bertujuan untuk meningkatkan akurasi serta efisiensi waktu komputasi pada proses klasifikasi teks. Pada penelitian ini, metode *Mutual Information* digunakan pada proses klasifikasi teks berupa sentimen emosi pada *Twitter* menggunakan algoritma *Support Vector Machine* (SVM). Kemudian, akan dilakukan analisis perbandingan pada setiap model klasifikasi berdasarkan hasil evaluasi yang didapat. Hasil penelitian menunjukkan penggunaan metode seleksi fitur mampu memberikan peningkatan akurasi serta mengurangi waktu komputasi. Penggunaan metode seleksi fitur *Mutual Information* pada algoritma SVM dengan kernel *linear* menghasilkan kinerja terbaik dengan rata-rata *accuracy* 0.71, *precision* 0.72, *recall* 0.71, *f-measure* 0.71 dan waktu komputasi 1.16 detik.

Kata Kunci : Klasifikasi Teks, Seleksi Fitur, *Support Vector Machine* (SVM)

## KATA PENGANTAR

Segala puji bagi Allah Subhanahu Wa Ta'ala karna atas karunia dan rahmat-nya sehingga penulis dapat menyelesaikan Skripsi dengan judul “**Analisis Sentimen Emosi *Twitter* Menggunakan Metode Seleksi Fitur *Mutual Information* dan Algoritma *Support Vector Machine* (SVM)**”. Penulisan Skripsi ini ditunjukkan untuk melengkapi salah satu syarat dalam menyelesaikan pendidikan dan mendapatkan gelar Sarjana Komputer di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Atas selesainya skripsi ini, penulis mengucapkan rasa syukur kepada Allah SWT. Dan penulis menyampaikan rasa terima kasih kepada yang terhormat :

1. Kedua Orang Tua saya yaitu Eko Soerjanto dan Evi Fitriyanti, saudara dan saudari saya yaitu M.Thoriq Abrar, M.Naufal Sabiq dan Syafina Aureli Kusumaningtyas, dan semua keluarga besar penulis yang sangat saya cintai. Terima kasih untuk semua doa yang telah dipanjatkan dan terimakasih juga untuk semua dukungan dan bantuan yang telah diberikan.
2. Bapak Jaidan Jauhari, S.Pd., M.T. (alm), selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Ibu Alvi Syahrini Utami, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak M. Fachrurrozi, M.T. selaku Dosen Pembimbing Akademik di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Bapak Samsuryadi, M.Kom., Ph.D. dan Ibu Rizki Kurniati, M.T. Selaku



Dosen Pembimbing Skripsi di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

6. Ibu Lestari, S.SI., M.T. selaku ketua lab Basis Data dan Big Data dan kak Abdi selaku admin lab fasikom unsri bukit yang telah memberi saya fasilitas untuk membantu pengerjaan skripsi saya.
7. Mba Wiwin selaku admin Jurusan Teknik Informatika yang telah membantu mengurus berkas administrasi penulis.
8. Seluruh dosen dan staff Fakultas Ilmu Komputer Universitas Sriwijaya.
9. Syecky Al-Qodrin, Rafliandi Ardana, Sultan Al-farid, Altundri Wahyu Hidayatullah, Annisa Aulia, Indra Gifari selaku teman kelas saat SMP, Julia Shakira Hutaaruk, komunitas *discord* OMC yang telah menemani hari saya, serta seluruh teman-teman seperjuangan angkatan 2018 Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
10. Teman-teman Lab Bawah Palugada Jeas, Icha, Ikik, Zulpa, Indra, Dani, Manda yang telah menjaga ketertiban dan merawat lab tercinta.

Palembang, 14 April 2023  
Penyusun,

Ahmad Marzuqi Yasykur Luthfi  
NIM. 09021281924053

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	iii
STATEMENT PAGE .....	iii
MOTTO .....	v
ABSTRACT.....	vi
ABSTRAK .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR .....	xv
BAB I PENDAHULUAN .....	I-1
1.1    Pendahuluan .....	I-1
1.2    Latar Belakang .....	I-1
1.3    Rumusan Masalah .....	I-3
1.4    Tujuan Penelitian .....	I-3
1.5    Manfaat Penelitian .....	I-3
1.6    Batasan Masalah.....	I-4
1.7    Sistematika Penulisan .....	I-4
1.8    Kesimpulan .....	I-5
BAB II KAJIAN LITERATUR .....	II-1
2.1    Pendahuluan .....	II-1
2.2    Landasan Teori.....	II-1
2.2.1 Emosi .....	II-1
2.2.2 <i>Preprocessing</i> .....	II-2
2.2.3 Klasifikasi Teks.....	II-3
2.2.4 Seleksi Fitur .....	II-4
2.2.5 <i>Mutual Information</i> .....	II-5
2.2.6 <i>Support Vector Machine (SVM)</i> .....	II-6
2.2.7 <i>Multiclass SVM</i> .....	II-10
2.2.8 <i>K-Fold Cross Validation</i> .....	II-11
2.2.9 <i>Confusion Matrix</i> .....	II-12
2.2.10 <i>Rational Unified Process</i> .....	II-15
2.3    Penelitian Lain yang Relevan.....	II-16
2.4    Kesimpulan .....	II-17
BAB III METODOLOGI PENELITIAN .....	III-1
3.1    Pendahuluan .....	III-1
3.2    Pengumpulan Data .....	III-1

3.2.1	Jenis dan Sumber Data .....	III-1
3.3	Tahapan Penelitian .....	III-3
3.3.1	Kerangka Kerja .....	III-5
3.3.2	Kriteria Pengujian .....	III-7
3.3.3	Format Data Pengujian.....	III-8
3.3.4	Alat yang Digunakan dalam Pelaksanaan Penelitian .....	III-8
3.3.5	Pengujian Penelitian.....	III-9
3.3.6	Analisis Hasil Pengujian dan Membuat Kesimpulan.....	III-10
3.3.7	Metode Pengembangan Perangkat Lunak.....	III-10
3.3.8	Fase Insepsi .....	III-10
3.3.9	Fase Elaborasi .....	III-11
3.3.10	Fase Konstruksi.....	III-11
3.3.11	Fase Transisi .....	III-11
3.4	Kesimpulan .....	III-12
BAB IV PENGEMBANGAN PERANGKAT LUNAK .....		IV-1
4.1	Pendahuluan .....	IV-1
4.2	Fase Insepsi .....	IV-1
4.2.1	Pemodelan Bisnis .....	IV-1
4.2.2	Kebutuhan Sistem .....	IV-2
4.2.3	Analisis dan Desain.....	IV-3
4.2.4	Implementasi .....	IV-11
4.3	Fase Elaborasi .....	IV-20
4.3.1	Pemodelan Bisnis .....	IV-20
4.3.2	Perancangan Data.....	IV-20
4.3.3	Perancangan Antarmuka .....	IV-20
4.3.4	Kebutuhan Sistem .....	IV-21
4.3.5	Diagram Aktivitas .....	IV-22
4.3.6	Diagram Sequence .....	IV-26
4.4	Fase Konstruksi.....	IV-31
4.4.1	Kebutuhan Sistem .....	IV-31
4.4.2	Diagram Kelas.....	IV-31
4.4.3	Implementasi .....	IV-33
4.5	Fase Transisi .....	IV-35
4.5.1	Pemodelan Bisnis .....	IV-35
4.5.2	Rencana Pengujian .....	IV-36
4.5.3	Implementasi .....	IV-38
4.6	Kesimpulan .....	IV-41
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
5.3	Analisis Hasil Penelitian .....	V-8
5.3.1	Analisis Kernel, Nilai C dan <i>Threshold</i> .....	V-8
5.3.2	Analisis Jumlah Fitur .....	V-19

5.3.3 Analisis Hasil Kinerja Metode Seleksi Fitur <i>Mutual Information</i> ..	V-20
5.3.4 Analisis Hasil Prediksi Pengujian Klasifikasi.....	V-20
5.4 Kesimpulan .....	V-21
BAB VI KESIMPULAN DAN SARAN .....	VI-1
6.1 Pendahuluan .....	VI-1
6.2 Kesimpulan .....	VI-1
6.3 Saran.....	VI-1
DAFTAR PUSTAKA .....	xvi
Lampiran 1. Proses Perhitungan Algoritma SVM Tanpa Seleksi Fitur <i>Mutual Information</i> .....	VII-20
Lampiran 2. Proses Perhitungan Algoritma SVM Menggunakan Metode Seleksi Fitur <i>Mutual Information</i> .....	VIII-33

## DAFTAR TABEL

Tabel II-1 Jenis Emosi (Riana, Mashar 2011) .....	II-2
Tabel II-2 Model Confusion Matrix.....	II-12
Tabel III-1 Tabel Jumlah Data .....	III-1
Tabel III-2 Contoh Tabel Pelabelan pada Emosi .....	III-2
Tabel III-3 Rancangan Tabel Hasil Pengujian .....	III-8
Tabel III-4 Rancangan Tabel Hasil Analisis Klasifikasi.....	III-10
Tabel IV-1 Kebutuhan Fungsional.....	IV-2
Tabel IV-2 Kebutuhan Non-Fungsional .....	IV-3
Tabel IV-3 Contoh Data Emosi .....	IV-5
Tabel IV-4 Hasil Noise Removal.....	IV-6
Tabel IV-5 Hasil Proses Case Folding .....	IV-6
Tabel IV-6 Hasil Proses Tokenizing .....	IV-7
Tabel IV-7 Definisi Actor .....	IV-13
Tabel IV-8 Definisi Use Case .....	IV-13
Tabel IV-9 Skenario Melakukan Memasukkan Data.....	IV-14
Tabel IV-10 Skenario Melakukan Praproses Data.....	IV-15
Tabel IV-11 Skenario Melakukan Proses Klasifikasi Menggunakan Algoritma SVM .....	IV-16
Tabel IV-12 Skenario Melakukan Proses Klasifikasi Menggunakan Algoritma SVM dan Seleksi Fitur Mutual Information .....	IV-17
Tabel IV-13 Skenario Melakukan Proses Pengujian Klasifikasi Sentimen Emosi .....	IV-19
Tabel IV-14 Implementasi Kelas .....	IV-33
Tabel IV-15 Rencana Pengujian Use Case Melakukan Praproses Data .....	IV-36
Tabel IV-16 Rencana Pengujian Use Case Melakukan Klasifikasi Menggunakan Algoritma SVM.....	IV-36
Tabel IV-17 Rencana Pengujian Use Case Melakukan Klasifikasi Menggunakan Algoritma SVM dan Seleksi Fitur Mutual Information.....	IV-37
Tabel IV-18 Rencana Pengujian Use Case Melakukan Proses Pengujian Klasifikasi Sentimen Emosi .....	IV-37
Tabel IV-19 Pengujian Use Case Memasukkan Data .....	IV-38
Tabel IV-20 Pengujian Use Case Melakukan Praproses Data .....	IV-38
Tabel IV-21 Pengujian Proses Klasifikasi Menggunakan Algoritma SVM ....	IV-39
Tabel IV-22 Pengujian Proses Klasifikasi Menggunakan Algoritma SVM dan Seleksi Fitur Mutual Information .....	IV-40
Tabel IV-23 Melakukan Proses Pengujian Klasifikasi Sentimen Emosi .....	IV-41
Tabel V-1 Hasil Evaluasi Pada Klasifikasi SVM Tanpa Seleksi Fitur pada Kernel Linear .....	V-2
Tabel V-2 Hasil Evaluasi Pada Klasifikasi SVM untuk Kernel RBF .....	V-2
Tabel V-3 Hasil Evaluasi Pada Klasifikasi SVM untuk Kernel Polynomial .....	V-2

Tabel V-4 Hasil Evaluasi Pada Klasifikasi SVM+MI untuk Kernel Linear .....	V-3
Tabel V-5 Hasil Evaluasi Pada Klasifikasi SVM+MI untuk Kernel RBF .....	V-4
Tabel V-6 Hasil Evaluasi Pada Klasifikasi SVM+MI untuk Kernel Polynomial	V-4
Tabel V-7 Hasil Evaluasi Pada Klasifikasi Model SVM untuk Kernel Linear...	V-5
Tabel V-8 Hasil Evaluasi Pada Klasifikasi Model SVM untuk Kernel RBF .....	V-5
Tabel V-9 Hasil Evaluasi Pada Klasifikasi Model SVM untuk Kernel Polynomial .....	V-6
Tabel V-10 Data Hasil Pengujian Prediksi Klasifikasi Sentimen Emosi menggunakan Kernel Linear .....	V-7
Tabel V-11 Data Hasil Pengujian Prediksi Klasifikasi Sentimen Emosi menggunakan Kernel RBF .....	V-7
Tabel V-12 Data Hasil Pengujian Prediksi Klasifikasi Sentimen Emosi menggunakan Kernel Polynomial .....	V-8
Tabel L- 1 Hasil Pembobotan TF- IDF .....	VII-20
Tabel L- 2 Hasil Pembobotan Kata TF-IDF .....	VII-23
Tabel L- 3 Klasifikasi Multiclass SVM one-against-one.....	VII-30
Tabel L- 4 Perhitungan Bobot Nilai Mutual Information .....	VIII-33
Tabel L- 5 Hasil Seleksi Fitur Mutual Information .....	VIII-37
Tabel L- 6 Hasil Ekstraksi Fitur TF-IDF .....	VIII-38
Tabel L- 7 Hasil Ekstraksi Fitur TF-IDF .....	VIII-39
Tabel L- 8 Klasifikasi Multiclass SVM one-against-one.....	VIII-3

## DAFTAR GAMBAR

Gambar II-1 Ilustrasi Pola SVM .....	II-7
Gambar II-2 Ilustrasi <i>K-Fold Cross Validation</i> .....	II-12
Gambar II-3 Arsitektur RUP .....	II-15
Gambar III-1 Tahapan Penelitian .....	III-4
Gambar III-2 Diagram Kerangka Kerja .....	III-5
Gambar IV-1 Diagram <i>Use Case</i> .....	IV-12
Gambar IV-2 Rancangan Antarmuka Perangkat Lunak .....	IV-21
Gambar IV-3 Diagram Aktivitas Memasukkan Data dan Melakukan <i>Preprocessing</i> .....	IV-23
Gambar IV-4 Diagram Aktivitas Melakukan Proses Klasifikasi menggunakan Algoritma SVM .....	IV-24
Gambar IV-5 Diagram Aktivitas Melakukan Proses Klasifikasi menggunakan Algoritma SVM dan Seleksi Fitur <i>Mutual Information</i> .....	IV-25
Gambar IV-6 Diagram Aktivitas Melakukan Proses Pengujian Klasifikasi Sentimen Emosi .....	IV-26
Gambar IV-7 Diagram <i>Sequence</i> Memilih Dataset dan Melakukan <i>Preprocessing</i> .....	IV-27
Gambar IV-8 Diagram <i>Sequence</i> Melakukan Proses Klasifikasi Menggunakan SVM .....	IV-28
Gambar IV-9 Diagram <i>Sequence</i> Melakukan Proses Klasifikasi Menggunakan SVM dan Seleksi Fitur <i>Mutual Information</i> .....	IV-29
Gambar IV-10 Diagram <i>Sequence</i> Melakukan Proses Pengujian Klasifikasi Sentimen Emosi .....	IV-30
Gambar IV-11 Diagram Kelas .....	IV-32
Gambar IV-12 Implementasi Tampilan Antarmuka Perangkat Lunak .....	IV-35
Gambar V-1 Grafik Data Perbandingan Hasil Model Klasifikasi Kernel <i>Linear</i> & $C = 0.1$ .....	V-9
Gambar V-2 Grafik Data Perbandingan Hasil Model Klasifikasi Kernel <i>Linear</i> & $C = 1$ .....	V-9
Gambar V-3 Grafik Data Perbandingan Hasil Model Klasifikasi Kernel <i>Linear</i> & $C = 10$ .....	V-10
Gambar V-4 Grafik Data Perbandingan Hasil Model Klasifikasi Kernel RBF & $C = 0.1$ .....	V-12
Gambar V-5 Grafik Data Perbandingan Hasil Model Klasifikasi Kernel RBF & $C = 1$ .....	V-13
Gambar V-6 Grafik Data Perbandingan Hasil Model Klasifikasi Kernel RBF & $C = 10$ .....	V-13
Gambar V-7 Grafik Data Perbandingan Hasil Model Klasifikasi Kernel <i>Polynomial</i> & $C = 0.1$ .....	V-16

Gambar V-8 Grafik Data Perbandingan Hasil Model Klasifikasi Kernel <i>Polynomial</i> & $C = 1$ .....	V-16
Gambar V-9 Grafik Data Perbandingan Hasil Model Klasifikasi Kernel <i>Polynomial</i> & $C = 10$ .....	V-17
Gambar V-10 Grafik Data Perbandingan Jumlah Fitur Metode <i>Mutual Information</i> .....	V-19



# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab pendahuluan akan membahas tentang latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah penelitian, dan sistematika penulisan. Bab ini akan memberikan penjelasan umum mengenai pokok pikiran dari keseluruhan penelitian.

### **1.2 Latar Belakang**

Emosi adalah hal yang bersifat umum dan penting dalam kehidupan yang membentuk perilaku manusia. Emosi seseorang bisa diidentifikasi menggunakan data tekstual yaitu berupa teks, karena teks dapat digunakan untuk berkomunikasi dan menyampaikan informasi. Emosi sering diutarakan seseorang melalui media sosial untuk mencurahkan emosi mereka baik itu sedih, marah, takut, maupun senang. Salah satu media sosial yang digunakan untuk bertukar informasi adalah *Twitter*. *Twitter* berisi informasi tentang sikap serta keadaan emosi seseorang. Akan tetapi, emosi yang diutarakan seseorang melalui platform *Twitter* terkadang menimbulkan pro dan kontra kepada pengguna *Twitter* terkait jenis emosi yang diutarakan. Hal tersebut terjadi karena komunikasi yang dilakukan pada media sosial ini adalah komunikasi yang bersifat secara tidak langsung, sehingga sulit untuk mengetahui perasaan atau kondisi orang tersebut. Oleh karena itu, untuk menentukan emosi seseorang atau informasi yang tepat, maka diperlukan suatu analisa dalam memahami emosi tersebut. Proses klasifikasi emosi secara otomatis dapat dipergunakan untuk mempermudah melakukan analisa terkait kategori suatu emosi. Berdasarkan pembagian pada analisa emosi,

emosi sendiri terbagi menjadi 5 jenis emosi dasar yang dirasakan manusia, yaitu *anger*, *happy*, *sadness*, *fear*, dan *love*.

Masalah utama pada klasifikasi teks adalah sebagian besar teks memiliki dimensi fitur yang banyak, sedangkan sebagian besar fitur ini tidak relevan dan bahkan mengandung *noise* yang dapat mengurangi tingkat akurasi klasifikasi (Chandra, 2019). Oleh karena itu *feature selection* umumnya digunakan dalam klasifikasi teks untuk mengurangi dimensi ruang fitur dan meningkatkan efisiensi dan akurasi pengklasifikasi. Terdapat beberapa metrik seleksi fitur yang terkenal seperti *Information Gain (IG)*, *Chi-Square*, dan *Mutual Information* (Khan et al., 2010).

Irham et al. (2019) dalam penelitiannya menjelaskan metode *Mutual Information* membantu mengeliminasi fitur yang tidak menginterpretasikan sebuah kelas. Sehingga metode ini dapat meningkatkan kecepatan dan efektifitas hasil kinerja dalam klasifikasi.

Terdapat beberapa penelitian yang telah dilakukan sebelumnya berkaitan dengan komparasi fitur seleksi pada klasifikasi dokumen. Dalam penelitian Chandani. (2015), peneliti melakukan perbandingan algoritma *machine learning* (*Artificial Neural Network (ANN)*, *Support Vector Machine (SVM)*, *Naïve Bayes (NB)*) dan juga membandingkan seleksi fitur (*Information Gain*, *Chi-Square*, *Forward Selection* dan *Backward Selection*). Hasil dari penelitian tersebut didapati bahwa SVM sebagai algoritma *machine learning* terbaik.

Berdasarkan uraian dan referensi penelitian sebelumnya, maka penelitian ini akan melakukan kombinasi terhadap metode seleksi fitur yaitu *Mutual Information* pada klasifikasi emosi pada *Twitter* dan menggunakan algoritma

*Support Vector Machine* untuk mengetahui tingkat akurasi hasil kinerja metode seleksi fitur *Mutual Information* tersebut. Hasil penelitian ini diharapkan dapat menjadi rujukan bagi pengembang sistem analisa emosi dan memberikan solusi dalam melakukan optimalisasi hasil kinerja pada metode seleksi fitur *Mutual Information* dan algoritma *Support Vector Machine*.

### **1.3 Rumusan Masalah**

Berdasarkan penjelasan dari latar belakang, maka rumusan masalah dari penelitian ini sebagai berikut:

1. Bagaimana mengklasifikasi emosi pada *Twitter* menggunakan metode seleksi fitur *Mutual Information* dan algoritma *Support Vector Machine* (SVM)?
2. Apakah penggunaan seleksi fitur *Mutual Information* dalam melakukan pemilihan fitur pada sistem klasifikasi emosi pada *Twitter* dapat meningkatkan akurasi algoritma *Support Vector Machine* (SVM)?

### **1.4 Tujuan Penelitian**

Tujuan penelitian ini adalah sebagai berikut:

1. Mengembangkan perangkat lunak untuk mengklasifikasi emosi pada *Twitter* menggunakan metode seleksi fitur *Mutual Information* dan algoritma *Support Vector Machine* (SVM).
2. Menunjukkan pengaruh *Mutual Information* dalam melakukan seleksi fitur pada sistem klasifikasi emosi pada *Twitter* menggunakan algoritma *Support Vector Machine* (SVM).

### **1.5 Manfaat Penelitian**

Manfaat penelitian ini adalah sebagai berikut:

1. Perangkat lunak dapat digunakan klasifikasi emosi pada *Twitter*.
2. Mengetahui hasil dari penggunaan *Mutual Information* dalam melakukan seleksi fitur pada sistem analisis emosi pada *Twitter* menggunakan algoritma *Support Vector Machine* (SVM).

### **1.6 Batasan Masalah**

Batasan masalah penelitian ini adalah sebagai berikut :

1. Data komentar pada *Twitter* yang telah dikumpulkan berjumlah 3000 komentar yang terbagi menjadi 600 data berlabel '*anger*', 600 data berlabel '*happy*', 600 data berlabel '*sadness*', 600 data berlabel '*fear*', dan 600 data berlabel '*love*'.
2. Klasifikasi terdiri dari 5 kelas, yaitu kelas *anger*, *happy*, *sadness*, *fear*, dan *love*.
3. Kernel yang digunakan adalah RBF, *linear*, dan *polynomial*.

### **1.7 Sistematika Penulisan**

Sistematika penulisan pada penelitian ini adalah sebagai berikut.

## **BAB I. PENDAHULUAN**

Bab ini menguraikan mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan pada penelitian ini.

## **BAB II. KAJIAN LITERATUR**

Bab ini membahas mengenai landasan teori yang digunakan dalam penelitian ini, seperti klasifikasi emosi pada *Twitter*, jenis-jenis emosi, seleksi fitur *Mutual Information* dan algoritma *Support Vector Machine*, serta membahas beberapa

penelitian yang relevan.

### **BAB III. METODOLOGI PENELITIAN**

Bab ini berisi pembahasan mengenai metodologi dan tahapan perancangan penelitian seperti pengumpulan data, metode pengembangan perangkat lunak, dan manajemen proyek penelitian.

### **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Bab ini membahas mengenai analisa dan rancangan pengembangan sistem perangkat lunak. Diawali dengan kebutuhan analisis, perancangan dan konstruksi, kemudian diakhiri dengan melakukan pengujian.

### **BAB V. HASIL DAN ANALISA SARAN**

Bab ini menguraikan hasil pengujian berdasarkan perancangan. Tabel hasil evaluasi pengujian dan analisis serta grafik menjadi patokan dari kesimpulan yang akan diambil dalam penelitian

### **BAB VI. KESIMPULAN DAN SARAN**

Bab ini membahas mengenai kesimpulan berdasarkan semua uraian pada bab sebelumnya dan juga saran yang diberikan dari hasil penelitian.

#### **1.8 Kesimpulan**

Berdasarkan latar belakang penelitian yang telah diuraikan maka penelitian ini akan mengembangkan sistem analisis sentimen emosi *Twitter* menggunakan metode seleksi fitur *Mutual Information* dan algoritma *Support Vector Machine*.

**DAFTAR PUSTAKA**

- Alita, D., Fernando, Y., & Sulistiani, H. (2020). Implementasi Algoritma Multiclass Svm Pada Opini Publik Berbahasa Indonesia Di Twitter. *Jurnal Tekno Kompak*, 14(2), 86. <https://doi.org/10.33365/jtk.v14i2.792>
- Amrullah, A. Z., Anas, S. A., Hidayat, & Muh. Adrian. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal*, 2(1), 40–44. <https://doi.org/10.30812/bite.v2i1.804>
- Bahassine, S., Madani, A., Al-sarem, M., & Kissi, M. (2020). Feature selection using an improved Chi-square for Arabic text classification. *Journal of King Saud University - Computer and Information Sciences*, 32(2), 225–231. <https://doi.org/10.1016/j.jksuci.2018.05.010>
- Bramer, M. (2007). Principles of Data Mining. In *Principles of Data Mining* (Issue January 2007). <https://doi.org/10.1007/978-1-84628-766-4>
- Buani, D. (2021). Penerapan Algoritma Naïve Bayes dengan Seleksi Fitur Algoritma Genetika Untuk Prediksi Gagal Jantung. 9(2), 43–48.
- Chandani, V. (2015). *Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film*. 1(1), 56–60.
- Chandra, A. (2019). Comparison of Feature Selection for Imbalance Text Datasets. *Proceedings of 2019 International Conference on Information Management and Technology, ICIMTech 2019, August*, 68–72. <https://doi.org/10.1109/ICIMTech.2019.8843773>
- Februariyanti, H., & Zuliarso, E. (2012). Klasifikasi Dokumen Berita Teks Bahasa Indonesia menggunakan Ontologi. *Teknologi Informasi DINAMIK*, 17(1), 14–23. <http://www.unisbank.ac.id/ojs/index.php/fti1/article/view/1612/594>

- Hanafi, A., Adiwijaya, A., & Astuti, W. (2020). Klasifikasi Multi Label pada Hadis Bukhari Terjemahan Bahasa Indonesia Menggunakan Mutual Information dan k-Nearest Neighbor. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 9(3), 357–364. <https://doi.org/10.32736/sisfokom.v9i3.980>
- Hanati, H., & Sari, K. (2021). *Perbandingan Metode Support Vector Machine ( SVM ) dan Artificial Neural Network ( ANN ) pada Klasifikasi Gizi Balita*. 1036–1043.
- 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. <https://doi.org/10.30865/mib.v3i4.1410>
- Irmanda, H., & Astriratma, R. (2020). Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(5), 915–922. <https://doi.org/10.29207/resti.v4i5.2313>
- Khan, A., Baharudin, B., & Lee, L. H. (2010). *A Review of Machine Learning Algorithms for Text-Documents Classification*. May 2014. <https://doi.org/10.4304/jait.1.1.4-20>
- Lucia, G., & Londo, Y. (2019). *A Study of Text Classification for Indonesian News Article*. 2019–2022.
- Luthfiana, L., Young, J. C., & Rusli, A. (2020). *Implementasi Algoritma Support Vector Machine dan Chi Square untuk Analisis Sentimen User Feedback Aplikasi*. XII(2), 125–128.
- Made, N., Dwi, G., Fauzi, M. A., & Dewi, L. S. (2018). *Identifikasi Tweet*

- Cyberbullying pada Aplikasi Twitter menggunakan Metode Support Vector Machine ( SVM ) dan Information Gain ( IG ) sebagai Seleksi Fitur. 2(11), 5326–5332.*
- Maulida, I., Suyatno, A., Rahmania Hatta, H., & Mulawarman, U. (2016). Seleksi Fitur Pada Dokumen Abstrak Teks Bahasa Indonesia Menggunakan Metode Information Gain. *JSM STMIK Mikroskil, 17(2), 249–258.*
- Mutawalli, L., Zaen, M. T. A., & Bagye, W. (2019). KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto). *Jurnal Informatika Dan Rekayasa Elektronik, 2(2), 43.* <https://doi.org/10.36595/jire.v2i2.117>
- Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *Jurnal Informatika, 6(2), 226–235.* <https://doi.org/10.31311/ji.v6i2.5129>
- Pessoa, M., & Anwar, A. (2014). *A Review of RUP ( Rational Unified Process ) - 2014.*
- Prakoso, B. S., Rosiyadi, D., Aridarma, D., Utama, H. S., Fauzi, F., & Qhomar, M. A. N. (2019). Optimalisasi Klasifikasi Berita Menggunakan Feature Information Gain Untuk Algoritma Naive Bayes Terhubung Random Forest. *Jurnal Pilar Nusa Mandiri, 15(2), 211–218.* <https://doi.org/10.33480/pilar.v15i2.684>
- Rahmad, A., & Pribadi, F. (2015). *Edu Komputika Journal. 2(1), 13–21.*
- Rahman, O. H., Abdillah, G., & Komarudin, A. (2021). Classification of Hate Speech on Social Media Twitter Using Support Vector Machine. *RESTI Journal (Systems Engineering and Information Technology), 5(1), 17–23.*



- Ridok, A., & Latifah, R. (2015). *Klasifikasi Teks Bahasa Indonesia Pada Corpus Tak Seimbang Menggunakan NWKNN*. 9–10.
- Sahuri, G. (2018). *Studi Perbandingan Penggabungan Metode Pemilihan Fitur dengan Metode Klasifikasi dalam Klasifikasi Teks*. 01(02), 1–5.
- Saniyah. (2019). *Named Entity Recognition pada Teks Berita menggunakan Support Vector Machine*.
- Sudin, S., Junaedi, H., & Santosa, J. (2019). *Analisis Jenis Pertanyaan Berbahasa Indonesia pada Question and Answering System Menggunakan Metode Support Vector Machine (SVM)*. 12, 72–80.
- Suharno, C., Fauzi, A., & Perdana, R. (2017). *KLASIFIKASI TEKS BAHASA INDONESIA PADA DOKUMEN PENGADUAN SAMBAT ONLINE MENGGUNAKAN METODE K*. 03(01), 25–32.
- SUPARTINI, I. A. M., SUKARSA, I. K. G., & SRINADI, I. G. A. M. (2017). *Analisis Diskriminan Pada Klasifikasi Desa Di Kabupaten Tabanan Menggunakan Metode K-Fold Cross Validation*. *E-Jurnal Matematika*, 6(2), 106. <https://doi.org/10.24843/mtk.2017.v06.i02.p154>
- Tanti, Sirait, P., & Andri. (2021). *Optimalisasi Kinerja Klasifikasi Melalui Seleksi Fitur dan AdaBoost dalam Penanganan Ketidakseimbangan Kelas*. 5, 1377–1385. <https://doi.org/10.30865/mib.v5i4.3280>
- Tuhenay, D. (2021). *Perbandingan Klasifikasi Bahasa Menggunakan Metode Naïve Bayes Classifier (NBC) Dan Support Vector Machine (SVM)*. *JIKO (Jurnal Informatika Dan Komputer)*, 4(2), 105–111. <https://doi.org/10.33387/jiko.v4i2.2958>