

**Analisis Sentimen Komentar Pos Endorsemen di Instagram
menggunakan Algoritma *Support Vector Machine* dan *Chi-Square***

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



Oleh:

Nurul Hannah Djati
NIM : 09021181621021

**JURUSAN TEKNIK INFORMATIKA FAKULTAS ILMU
KOMPUTER UNIVERSITAS SRIWIJAYA
2022**

LEMBAR PENGESAHAN SKRIPSI

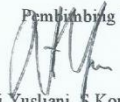
ANALISIS SENTIMEN KOMENTAR POS ENDORSEMENT DI INSTAGRAM
MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* DAN *CHI-SQUARE*

Oleh :

Nurul Hannah Djati
NIM : 09021181621021

Palembang, 29 Juli 2022


Pembimbing I


Novi Yuslimi, S.Kom., M.T.
NIP. 198211082012122001

Pembimbing II


Desty Hadiah, S.Kom., M.T.
NIP. 1989122120201220211

Mengetahui,
Ketua Jurusan Teknik Informatika,


Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003



TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Rabu tanggal 27 Juli 2022 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Nurul Hannah Djati
NIM : 09021181621021
Judul : Analisis Sentimen Komentar Pos Endorsemen di Instagram menggunakan Algoritma *Support Vector Machine* dan *Chi-Square*

1 Ketua

Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003



2 Penguji I

Mastura Diana Marieska, M. T NIP
198603212018032001



3 Penguji I

M. Naufal Rachmatullah, M.T
NIP 199212012022031008



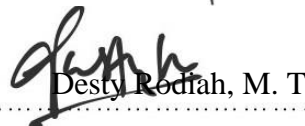
4 Pembimbing I

Novi Yusliani, M.T.
NIP 198211082012122001




5 Pembimbing II

Desty Rodiah
NIP 1989122120201220211



Desty Rodiah, M. T

Mengetahui
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Nurul Hannah Djati
NIM : 09021181621021
Program Studi : Teknik Informatika
Judul Skripsi : Analisis Sentimen Komentar Pos Endorsemen di
Instagram menggunakan Algoritma *Support Vector
Machine* dan *Chi-Square*

Hasil Pengecekan Software *iThenticate/Turnitin* : 7%

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

Demikian pernyataan ini saya buat dengan sebenarnya dan tanpa paksaan dari pihak manapun.



Palembang, 29 Juli 2022



Nurul Hannah Djati

NIM. 09021181621021

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya.” (QS. Al-Baqarah, 2:286)

Kupersembahkan karya tulis ini kepada :

- **Almarhum dan Almarhumah Orang Tuaku**
- **Dosen Pembimbing**
- **Fakultas Ilmu Komputer Universitas Sriwijaya**

ABSTRACT

Instagram is the fourth most used social media with 1,38 billion active users worldwide. The rapid development of the number of Instagram users turns Instagram as one of the options to market products. Recently, Instagram celebrities emerge as they promote products and services of a business or widely known as endorsemen. Sentiment analysis is required to be done to determine the opinions of public or to discover the demeanor of consumers towards the marketed products or services. Support Vector Machine is one of the methods used to solve linier and non-linier problems. This method transforms every word that has been processed to features which turns the computation time longer. In order to solve the long computation time problem, a seleksi fitur feature is needed to reduce the number of features and improve the performance of SVM method. The objective of this research is to determine the performance of SVM method with *Chi-Square* seleksi fitur and the performance of SVM method without the seleksi fitur method in analysing the sentiments of endorsemen posts comments in Instagram. The results of the test will be compared to learn the impact of using feature selection method in order to improve the performance of SVM. The test is done using three SVM kernels, which are Linier, Polynomial, and RBF with C parameter values are 0,1, 1, and 10. The results of the test reveal that using seleksi fitur reduces the performance of SVM compared to using SVM method without seleksi fitur with polynomial kernel generates the best result. The drop of parameter values in confusion matrix are 3,6% in accuration, 18% in precision, 8% in recall, and 15% in f-measure.

Keywords : Instagram, Analisis Sentimen, *Support Vector Machine*, *Chi-Square*

Palembang, 25 Juli 2022

Pembimbing I,



Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001

Pembimbing II,



Desty Rodiah, S.Kom., M.T.
NIP. 1989122120201220211

Mengetahui,
Ketua Jurusan Teknik Informatika,



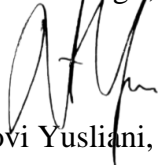
Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

ABSTRAK

Dengan sekitar 1,38 miliar pengguna aktif di dunia, Instagram adalah media sosial terbesar ke-4 di dunia saat ini. Pesatnya perkembangan jumlah pengguna Instagram menjadikan aplikasi media sosial ini salah satu pilihan dalam memasarkan produk. Belakangan ini, banyak ditemukan selebritis di Instagram yang mempromosikan produk dan jasa sebuah bisnis atau juga dikenal dengan istilah endorsemen. Analisis sentimen penting dilakukan untuk menentukan opini terhadap produk dan jasa, atau mengetahui sikap konsumen terhadap produk dan jasa yang dipasarkan. *Support Vector Machine* merupakan metode yang dapat menyelesaikan permasalahan secara linier maupun non-linier. Dalam metode *Support Vector Machine*, semua kata hasil pemrosesan teks akan menjadi fitur, sehingga komputasinya akan memakan waktu yang lama. Untuk mengatasi masalah tersebut dibutuhkan sebuah metode seleksi fitur untuk mengurangi jumlah fitur sehingga dapat meningkatkan kinerja dan mengurangi waktu komputasi pada proses pelatihan. Tujuan dari penelitian ini adalah mengetahui kinerja penggunaan metode SVM dengan seleksi fitur *Chi-Square* dan penggunaan metode SVM tanpa seleksi fitur dalam analisis sentimen komentar pos endorsemen di Instagram. Hasil dari penelitian ini akan dibandingkan untuk mengetahui pengaruh seleksi fitur yang digunakan dalam meningkatkan kinerja metode SVM. Pengujian dilakukan menggunakan tiga kernel SVM, yaitu kernel Linier, Polinomial, dan RBF dengan nilai parameter C 0,1, 1, dan 10. Hasil pengujian menunjukkan bahwa pada ketiga kernel, penggunaan seleksi fitur *Chi-Square* memberikan hasil pengujian yang lebih rendah daripada penggunaan metode SVM tanpa seleksi fitur. Hasil terbaik dihasilkan oleh kernel polinomial tanpa seleksi fitur. Penurunan nilai parameter *confusion matrix* adalah 3,6% pada akurasi, 18% pada presisi, 8% pada *recall*, dan 15% pada *f-measure*.

Kata kunci : Instagram, Analisis Sentimen, *Support Vector Machine*, *Chi-Square*
Palembang, 25 Juli 2022

Pembimbing I,



Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001

Pembimbing II,



Desty Rodiah, S.Kom., M.T.
NIP. 1989122120201220211

Mengetahui,

Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

KATA PENGANTAR

Puji syukur kepada Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini yang disusun sebagai salah satu syarat dalam menyelesaikan pendidikan program Strata-1 Program Studi Teknik Informatika di Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis ingin menyampaikan terima kasih kepada seluruh pihak yang telah memberikan bantuan dan dukungan selama proses pembuatan Tugas Akhir ini. Adapun penulis merujuk secara khusus pihak-pihak tersebut sebagai berikut :

1. Alm. Djamaluddin Anthoni, Almh. Titin Orini, dan Nabel Albar selaku orang tua dan adik saya yang menjadi motivasi terbesar saya dalam menyelesaikan Tugas Akhir.
2. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer
3. Ibu Alvi Syahrini Utami, M.Kom selaku Ketua Jurusan Teknik Informatika dan pembimbing akademik yang telah memberikan arahan dalam urusan akademik.
4. Ibu Novi Yusliani, M.T. selaku pembimbing I yang telah dengan sabar memberikan arahan, bimbingan, motivasi, kritik dan saran kepada penulis untuk menyelesaikan Tugas Akhir.
5. Ibu Desty Rodiah, M.T. selaku pembimbing II yang selalu memberikan dukungan untuk menyelesaikan Tugas Akhir.
6. Ibu Alvi Syahrini Utami, M.Kom, Ibu Mastura Diana Marieska, M.T, dan Pak Muhammad Naufal Rachmatullah, M.T selaku Ketua Penguji, Dosen Penguji I, dan Dosen Penguji II yang telah memberikan masukan dan arahan untuk menyempurnakan Tugas Akhir kepada penulis.
7. Seluruh Bapak dan Ibu Dosen Program Studi Teknik Informatika yang telah memberikan ilmu kepada penulis selama masa perkuliahan.
8. Staff Jurusan Teknik Informatika dan staff Fakultas Ilmu Komputer yang telah membantu melancarkan urusan administrasi dan akademik selama masa perkuliahan.
9. Cesil Oktavia Hersa, Desi Indriani, M. Abdul Aziz, M. Reza Perdana, Friska, dan Veti Vera yang telah memberikan semangat serta motivasi dalam proses penulisan Tugas Akhir.
10. Semua pihak lainnya yang telah banyak membantu penulis dalam menyelesaikan Tugas Akhir ini dan tidak dapat disebutkan satu persatu, terima kasih banyak semuanya.

Semoga tugas akhir ini dapat bermanfaat bagi penulis dan orang banyak, serta menjadi referensi dan rujukan untuk hal-hal bermanfaat. Penulis menyadari dalam masih banyak ketidaksempurnaan dalam proses penulisan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun agar tugas akhir ini menjadi karya tulis yang sempurna.

Palembang, 29 Juli 2022

Nurul Hannah Djati

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN SKRIPSI	ii
HALAMAN TANDA LULUS UJIAN SIDANG SKRIPSI	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vii
ABSTRAK	viii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvii
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian	I-4
1.5 Manfaat Penelitian	I-5
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan	I-6
1.7.1 Bab I. Pendahuluan	I-6
1.7.2 Bab II. Kajian Literatur	I-6
1.7.3 Bab III. Metode Penelitian	I-7
1.8 Kesimpulan	I-8
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Instagram.....	II-1
2.2.2 Endorsemen Selebritas	II-3
2.2.3 Analisis Sentimen	II-3

2.2.4	Pra-pengolahan	II-4
2.2.5	Pembobotan TF-IDF	II-7
2.2.6	<i>Support Vector Machine</i>	II-8
2.2.7	<i>Chi-Square</i>	II-16
2.2.8	Confusion Matrix.....	II-18
2.2.9	<i>Rational Unified Process</i> ...	II-20
2.3	Penelitian Lain yang Relevan	II-22
2.4	Kesimpulan ...	II-23
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
3.2.2	Metode Pengumpulan Data	III-3
3.3	Tahapan Penelitian	III-5
3.3.1	Pengumpulan Data	III-6
3.3.2	Menentukan Kerangka Kerja Penelitian	III-6
3.3.3	Menentukan Kriteria Pengujian.....	III-11
3.3.4	Menetapkan Format Data Pengujian	III-11
3.3.5	Menentukan Alat Bantu Penelitian	III-12
3.3.6	Melakukan Pengujian Penelitian	III-13
3.3.7	Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan Penelitian	III-13
3.4	Metode Pengembangan Perangkat Lunak	III-14
3.4.1	Fase Insepsi	III-14
3.4.2	Fase Elaborasi	III-15
3.4.3	Fase Konstruksi	III-15
3.4.4	Fase Transisi	III-15
3.5	Kesimpulan ...	III-16
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	IV-2
4.2.3	Analisis dan Perancangan	IV-4
4.2.4	Implementasi	IV-25
4.3	Fase Elaborasi.....	IV-32
4.3.1	Pemodelan Bisnis	IV-32
4.3.2	Kebutuhan	IV-33
4.3.3	Analisis dan Perancangan	IV-34
4.4	Fase Konstruksi	IV-41
4.4.1	Kebutuhan Sistem	IV-42
4.4.2	Diagram Kelas	IV-42
4.4.3	Implementasi	IV-43
4.5	Fase Transisi	IV-45
4.5.1	Pemodelan Bisnis	IV-45
4.5.2	Kebutuhan ..	IV-45
4.5.3	Analisis dan Perancangan	IV-46
4.5.4	Implementasi ..	IV-48
4.6	Kesimpulan	IV-51
 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-45
5.3.1	Analisis Kernel	V-45
5.3.2	Kernel Linier	V-45
5.3.3	Kernel Polinomial	V-47
5.3.4	Analisis Waktu Komputasi.....	V-51
5.3.5	Analisis Jumlah Fitur	V-52
5.6	Kesimpulan	V-53
 BAB VI KESIMPULAN DAN SARAN		VI-1

6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran	VI-2
DAFTAR PUSTAKA	xviii
LAMPIRAN	L-1

DAFTAR TABEL

	Halaman
II-1. Tabel tahapan pra-pengolahan.....	II-6
II-2. Tabel hubungan antara kata terhadap kelas	II-17
II-3. Tabel Nilai kritis χ^2 pada taraf nyata α	II-18
II-4. <i>Confusion Matrix</i>	II-19
III-1. Contoh data yang telah dikumpulkan.....	III-1
III-2. Rancangan Tabel <i>Confusion Matrix</i> Hasil Klasifikasi	III-11
III-3. Rancangan Tabel Hasil Pengujian.....	III-12
III-4. Tabel Hasil Analisis Klasifikasi	III-14
III-5. Tabel Rencana Jadwal Penelitian	III-16
IV- 1. Kebutuhan Fungsional Perangkat Lunak	IV-2
IV- 2. Kebutuhan Non-Fungsional Perangkat Lunak	IV-2
IV- 3. Contoh Data KomentarI	IV-5
IV- 4. Hasil <i>Cleaning</i> dari Data Komentar	IV-6
IV- 5. Hasil <i>Case Folding</i> dari Data Komentar	IV-7
IV- 6. Hasil Normalisasi dari Data Komentar	IV-8
IV- 7. Hasil <i>Tokenizing</i> dari Data Komentar	IV-8
IV- 8. Hasil <i>Stopword Removal</i> dari Data Komentar	IV-9
IV- 9. Hasil <i>Stemming</i> dari Data Komentar.....	IV-10
IV- 10. Perhitungan nilai TF dan IDF dari data hasil pra-pengolahan	IV-11
IV- 11. Perhitungan nilai TF-IDF dari data hasil pra-pengolahan	IV-12
IV- 12. Perhitungan Nilai <i>Chi-Square</i>	IV-18
IV- 13. Data hasil seleksi fitur	IV-20
IV- 14. Perhitungan nilai TF dan IDF fitur hasil pra-pengolahan	IV-20
IV- 15. Perhitungan nilai TF-IDF fitur hasil pra-pengolahan	IV-21
IV- 16. Tabel Definisi Aktor	IV-26
IV- 17. Tabel Definisi <i>Use Case</i>	IV-27
IV- 18. Skenario <i>Use Case</i> Memproses Data	IV-28
IV- 19. Skenario Melakukan Klasifikasi Menggunakan Algoritma SVM	IV-31
IV- 20. Skenario Melakukan Klasifikasi Menggunakan Algoritma SVM dan seleksi fitur <i>Chi-Square</i>	IV-30
IV- 21. Tabel Implementasi Kelas.....	IV-44
IV- 22. Rencana Pengujian <i>Use Case</i> Memproses Data	IV-47
IV- 23. Rencana Pengujian <i>Use Case</i> Klasifikasi Menggunakan SVM	IV-47
IV- 24. Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi Menggunakan SVM dan Seleksi Fitur <i>Chi-Square</i>	IV-47
IV- 25. Pengujian <i>Use Case</i> Memproses Data	IV-48
IV- 26. Pengujian <i>Use Case</i> Melakukan Klasifikasi Menggunakan SVM	IV-49
IV- 27. Pengujian <i>Use Case</i> Melakukan Klasifikasi Menggunakan SVM dan Seleksi Fitur <i>Chi-Square</i>	IV-50
V-1. <i>Confusion Matrix</i> Klasifikasi SVM menggunakan Kernel Linier dan $C=0,1$	V-2
V- 2. Data Hasil Evaluasi Klasifikasi SVM menggunakan Kernel Linier dan $C=0,1$	V-3
V- 3. <i>Confusion Matrix</i> Klasifikasi SVM menggunakan Kernel Linier dan $C=1$	V-4

V-4. Data Hasil Evaluasi Klasifikasi SVM menggunakan Kernel Linier dan C=4.....	V-5
V-5. <i>Confusion Matrix</i> Klasifikasi SVM menggunakan Kernel Linier dan C=10.....	V-7
V-6. Data Hasil Evaluasi Klasifikasi SVM menggunakan Kernel Linier dan C=10.....	V-8
V-7. <i>Confusion Matrix</i> Klasifikasi SVM Kernel Polinomial dan C=0,1.....	V-9
V-8. Data Hasil Evaluasi Klasifikasi SVM Kernel Polinomial dan C=0,1	V-10
V-9. <i>Confusion Matrix</i> Klasifikasi SVM Kernel Polinomial dan C=1.....	V-11
V-10. Data Hasil Evaluasi Klasifikasi SVM Kernel Polinomial dan C=1	V-12
V-11. <i>Confusion Matrix</i> Klasifikasi SVM Kernel Polinomial dan C=10	V-13
V-12. Data Hasil Evaluasi Klasifikasi SVM Kernel Polinomial dan C=10	V-14
V-13. <i>Confusion Matrix</i> Klasifikasi SVM menggunakan Kernel RBF dan C=0,1	V-15
V-14. Data Hasil Evaluasi Klasifikasi SVM Kernel RBF dan C=0,1	V-15
V-15. <i>Confusion Matrix</i> Klasifikasi SVM menggunakan Kernel RBF dan C=1	V-17
V-16. Data Hasil Evaluasi Klasifikasi SVM menggunakan Kernel RBF dan C=1	V-18
V-17. <i>Confusion Matrix</i> Klasifikasi SVM menggunakan Kernel RBF dan C=10	V-19
V-18. Data Hasil Evaluasi Klasifikasi SVM menggunakan Kernel RBF dan C=10.....	V-20
V- 19. <i>Confusion Matrix</i> Klasifikasi SVM+CS Kernel Linier dan C=0,1	V-21
V- 20. Data Hasil Evaluasi Klasifikasi SVM+CS Kernel Linier dan C=0,1	V-22
V- 21. <i>Confusion Matrix</i> Klasifikasi SVM+CS Kernel Linier dan C=1	V-23
V- 22. Data Hasil Evaluasi Klasifikasi SVM+CS Kernel Linier dan C=1	V-24
V- 23. <i>Confusion Matrix</i> Klasifikasi SVM+CS Kernel Linier dan C=10	V-25
V- 24. Data Hasil Evaluasi Klasifikasi SVM+CS Kernel Linier dan C=10.....	V-26
V- 25. <i>Confusion Matrix</i> Klasifikasi SVM+CS Kernel Polinomial dan C=0,1	V-27
V- 26. Data Hasil Evaluasi Klasifikasi SVM+CS Kernel Polinomial dan C=0,1	V-28
V- 27. <i>Confusion Matrix</i> Klasifikasi SVM+CS Kernel Polinomial dan C=1	V-29
V- 28. Data Hasil Evaluasi Klasifikasi SVM+CS Kernel Polinomial dan C=1	V-30
V- 29. <i>Confusion Matrix</i> Klasifikasi SVM+CS Kernel Polinomial dan C=10	V-31
V- 30. Data Hasil Evaluasi Klasifikasi SVM+CS Kernel Polinomial dan C=10	V-32
V- 31. <i>Confusion Matrix</i> Klasifikasi SVM+CS Kernel RBF dan C=0,1	V-33
V- 32. Data Hasil Evaluasi Klasifikasi SVM+CS Kernel RBF dan C=0,1	V-34
V- 33. <i>Confusion Matrix</i> Klasifikasi SVM+CS Kernel RBF dan C=1	V-35
V- 34. Data Hasil Evaluasi Klasifikasi SVM+CS Kernel RBF dan C=1	V-36
V- 35. <i>Confusion Matrix</i> Klasifikasi SVM+CS Kernel RBF dan C=10	V-37
V- 36. Data Hasil Evaluasi Klasifikasi SVM+CS Kernel RBF dan C=10.....	V-38
V- 37. Perbandingan Rata-Rata Hasil Klasifikasi Menggunakan Kernel Linier.....	V-40
V- 38. Perbandingan Rata-Rata Hasil Klasifikasi Menggunakan Kernel Polinomial	V-40
V- 39. Perbandingan Rata-Rata Hasil Klasifikasi Menggunakan Kernel RBF	V-41

DAFTAR GAMBAR

	Halaman
II-1. Letak bidang pemisah antarkelas	II-9
II-2. Bidang pemisah dengan margin terbesar	II-9
II-3. Pemetaan data ke ruang vektor yang berdimensi lebih tinggi	II-14
II-4. Arsitektur <i>Rational Unified Process</i>	II-21
III-1. Proses pengambilan data komentar Instagram	III-4
III-2. Komentar yang telah dilabeli	III-5
III-3. Alur tahapan penelitian	III-5
III-4. Diagram Alur Proses Umum Perangkat Lunak tanpa Proses Seleksi Fitur.....	III-7
III-5. Diagram Alur Proses Umum Perangkat Lunak dengan Proses Seleksi Fitur.....	III-8
IV- 1. Diagram <i>Use Case</i>	IV-26
IV- 2. Rancangan Antarmuka Perangkat Lunak.....	IV-33
IV- 3. Diagram Aktivitas Memproses Data.....	IV-34
IV- 4. Diagram Aktivitas Melakukan Klasifikasi Menggunakan Algoritma SVM.....	IV-35
IV- 5. Diagram Aktivitas Melakukan Klasifikasi Menggunakan Algoritma SVM dan seleksi fitur <i>Chi-Square</i>	IV-36
IV- 6. <i>Sequence Diagram</i> Memproses Data	IV-37
IV- 7. <i>Sequence Diagram</i> Melakukan Klasifikasi Menggunakan Algoritma SVM.....	IV-38
IV- 8. <i>Sequence Diagram</i> Melakukan Klasifikasi Menggunakan Algoritma SVM+CS...	IV-40
IV- 9. Diagram Kelas Perangkat Lunak	IV-43
IV-10. Implementasi Antarmuka Perangkat Lunak.....	IV-45
V-1. Grafik Hasil Pengujian 10 Fold menggunakan Kernel Linier dan $C=0,1$	V-3
V-2. Grafik Hasil Pengujian 10 Fold menggunakan Kernel Linier dan $C = 1$	V-6
V-3. Grafik Hasil Pengujian 10 Fold menggunakan Kernel Linier dan $C=10$	V-8
V-4. Grafik Hasil Pengujian 10 Fold Kernel Polinomial dan $C=0,1$	V-10
V-5. Grafik Hasil Pengujian 10 Fold Kernel Polinomial dan $C = 1$	V-12
V-6. Grafik Hasil Pengujian 10 Fold menggunakan Kernel Polinomial dan $C=10$	V-14
V-7 Grafik Hasil Pengujian 10 Fold menggunakan Kernel RBF dan $C = 0,1$	V-16
V-8 Grafik Hasil Pengujian 10 Fold menggunakan Kernel RBF dan $C=1$	V-18
V-9. Grafik Hasil Pengujian 10 Fold Kernel RBF dengan $C=10$	V-20
V-10. Grafik Hasil Pengujian 10 Fold Kernel Linier dan $C = 0,1$	V-22
V- 11. Grafik Hasil Pengujian 10 Fold Kernel Linier dan $C = 1$	V-24
V- 12. Grafik Hasil Pengujian 10 Fold Kernel Linier dan $C = 10$	V-26
V- 13. Grafik Hasil Pengujian 10 Fold Kernel Polinomial dan $C=0,1$	V-28
V- 14. Grafik Hasil Pengujian 10 Fold Kernel Polinomial dan $C = 1$	V-30
V- 15. Grafik Hasil Pengujian 10 Fold Kernel Polinomial dan $C=10$	V-33
V- 16. Grafik Hasil Pengujian 10 Fold Kernel RBF dan $C = 0,1$	V-35
V- 17. Grafik Hasil Pengujian 10 Fold Kernel RBF dan $C = 1$	V-37
V- 18. Grafik hasil Pengujian 10 Fold Kernel RBF dengan $C = 10$	V-39
V- 19. Perbandingan Hasil Evaluasi Menggunakan Kernel Linier dan $C=0,1$	V-42
V- 20. Perbandingan Hasil Evaluasi Menggunakan Kernel Linier dan $C=1$	V-43
V- 21. Perbandingan Hasil Evaluasi Menggunakan Kernel Linier dan $C=10$	V-43

V- 22. Perbandingan Hasil Evaluasi Menggunakan Kernel Polinomial dan $C=0,1$	V-44
V- 23. Perbandingan Hasil Evaluasi Menggunakan Kernel Polinomial dan $C=1$	V-45
V- 24. Perbandingan Hasil Evaluasi Menggunakan Kernel Polinomial dan $C=1$	V-45
V- 25. Perbandingan Hasil Evaluasi Menggunakan Kernel RBF dan $C=0,1$	V-46
V- 26. Perbandingan Hasil Evaluasi Menggunakan Kernel RBF dan $C=1$	V-47
V- 27. Perbandingan Hasil Evaluasi Menggunakan Kernel RBF dan $C=10$	V-47
V- 28. Diagram Perbandingan Waktu Komputasi Metode Klasifikasi	V-48
V- 29. Perbandingan Jumlah Fitur yang Dihasilkan Setiap Metode Klasifikasi	V-50

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini terdapat uraian mengenai latar belakang, rumusan dan tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan, dan kesimpulan dari penelitian.

1.2 Latar Belakang

Media sosial sekarang bukan hanya sebuah situs web biasa. Seperti Instagram, media sosial telah menjadi sarana komunikasi bagi pengguna internet (Shrivatava and Pant, 2012). Pada Januari 2021, terlihat bahwa dari 4,66 miliar pengguna internet aktif, 4,2 miliar di antaranya adalah pengguna media sosial (Johnson, 2021). Dengan sekitar 1,38 miliar pengguna aktif di seluruh dunia, Instagram adalah media sosial terbesar ke-4 di dunia saat ini. Jumlah pengguna Instagram dan internet yang selalu meningkat dari tahun ke tahun menunjukkan potensi besar bahwa jumlah tersebut akan tetap bertambah seiring waktu.

Dari banyaknya media sosial yang saat ini digunakan, Instagram memberikan ruang bagi penggunanya untuk berbagi konten berupa video dan foto secara bebas hingga bertukar pesan pribadi dengan pengguna lainnya. Hingga Juli 2021, tercatat dari 91,77 juta pengguna Instagram di Indonesia, pengguna berusia 18 hingga 24 tahun menempati kelompok usia terbanyak yang menggunakan Instagram (Rizaty, 2021). Pesatnya perkembangan jumlah pengguna Instagram menjadikan aplikasi media sosial ini sebagai salah satu pilihan dalam memasarkan

produk. Belakangan ini, banyak ditemukan selebritis di Instagram yang mempromosikan produk dan jasa sebuah bisnis atau juga dikenal dengan istilah endorsemen. Dalam berinteraksi di Instagram, pengguna dapat meninggalkan komentar dan *like* pada foto atau video yang diunggah. Semakin banyak pengikut yang mengikuti suatu laman, maka semakin tinggi pula interaksi di laman tersebut, sehingga akan meningkatkan *engagement rate* atau keterlibatan pengikut di suatu laman. Dalam kata lain, komentar-komentar yang ditinggalkan pengguna merupakan hal penting dalam pemasaran produk sehingga bisnis dapat berkembang (Nurfitriyani, 2019). Analisis sentimen penting dilakukan untuk menentukan opini terhadap produk dan jasa, atau mengetahui sikap konsumen terhadap produk dan jasa tersebut (Neri, et al, 2012). Dengan analisa komentar pada pos endorsemen, pemilik bisnis dapat mengetahui respon publik secara pasti terhadap produk atau jasa yang diiklankan.

Klasifikasi sentimen adalah salah satu metode untuk mengolah dan mengekstrak data secara otomatis sehingga menjadi informasi yang dapat dipahami dan bermanfaat (Rozi et al., 2013). Metode yang lazim digunakan untuk proses klasifikasi yaitu *Naïve Bayes* (NB), *K-Nearest Neighbor* (K-NN), dan *Support Vector Machine* (SVM). Pada sebuah penelitian yang dilakukan oleh Ilmawan dan

Mude (2020) mengenai perbandingan metode klasifikasi *Naive Bayes* dan *Support Vector Machine* pada kasus analisis sentimen ulasan di *Google Play Store*, metode *Support Vector Machine* memiliki akurasi lebih tinggi dari *Naive Bayes* dengan nilai 81,46% dibanding 75,41%. Pada penelitian Watmah, Suryanto, dan

Matias (2021), yaitu perbandingan akurasi metode klasifikasi *K-Nearest Neighbor* (KNN), *Support Vector Machine*, dan *Random Forest* pada penelitian kepuasan pelanggan terhadap *E-Commerce* Shopee di *Google Play Store*, *Support Vector Machine* memiliki akurasi 89,4%, lebih tinggi dari metode *Random Forest* dengan akurasi 83% dan *K-Nearest Neighbor* (K-NN) dengan akurasi 89%. Dari kedua penelitian tersebut, metode *Support Vector Machine* memberikan hasil akurasi lebih tinggi dibandingkan metode *Naive Bayes*, *K-Nearest Neighbor* (K-NN), dan *Random Forest*.

Permasalahan linier dan non-linier dapat diselesaikan dengan metode SVM. Dalam penggunaan metode SVM, semua kata hasil pemrosesan teks akan menjadi fitur, sehingga komputasinya akan memakan waktu yang lama (Indriati, Purnamasari, dan Fauzi, 2018). Dalam mengurangi waktu komputasi, sebuah metode *seleksi fitur* akan digunakan untuk mengurangi jumlah fitur.

Pada penyeleksian fitur data, terdapat metode *Chi-Square* yang tergolong ke dalam kategori seleksi fitur *supervised* selain *Information Gain* dan *Mutual Information*. *Chi-Square* berguna untuk mengukur keterkaitan sebuah kata pada kelas data tertentu. Penelitian yang digarap oleh Tsani, et al (2020) menampilkan bahwa *Chi-Square* mampu mengurangi jumlah fitur atau atribut tidak relevan pada dataset sehingga kinerja metode klasifikasi yang diaplikasikan dapat ditingkatkan. Pada penelitian lainnya yang dilakukan oleh Putra (2018), penggunaan *seleksi fitur Chi-Square* meningkatkan nilai presisi dan *recall* secara signifikan.

Berlandaskan hasil penelitian yang dilakukan oleh Ilmawan dan Mude (2020), Watmah, Suryanto, dan Matias (2021), dan Putra (2018) sebelumnya, peneliti akan mengembangkan sistem analisis sentimen menggunakan algoritma *Support Vector Machine* dengan seleksi fitur *Chi-Square* dalam penelitian analisis sentimen pada kolom komentar pos endorsemen di Instagram.

1.3 Rumusan Masalah

Berdasarkan penjelasan masalah yang telah dijabarkan di latar belakang sebelumnya, berikut adalah rumusan masalah pada penelitian ini :

1. Bagaimana mengembangkan sistem analisis sentimen pada komentar pos endorsemen di Instagram menggunakan metode *Support Vector Machine*?
2. Bagaimana kinerja algoritma *Support Vector Machine* dalam system analisis sentiment pada komentar pos endorsemen di Instagram?
3. Bagaimana pengaruh seleksi fitur *Chi-Square* terhadap hasil analisis sentimen menggunakan metode *Support Vector Machine*?

1.4 Tujuan Penelitian

Tujuan penelitian ini berdasarkan rumusan masalah yang telah dijelaskan di atas adalah :

1. Menghasilkan sistem yang dapat melakukan analisis sentimen pada kolom komentar Instagram menggunakan metode *Support Vector Machine* tanpa seleksi fitur *Chi-Square* dan metode *Support Vector*

Machine dengan seleksi fitur *Chi-Square*.

2. Mengetahui kinerja dari metode *Support Vector Machine* dalam melakukan analisis sentimen pada komentar Instagram.
3. Mengetahui pengaruh seleksi fitur *Chi-Square* terhadap analisis sentimen menggunakan metode *Support Vector Machine*.

1.5 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut :

1. Memperoleh informasi mengenai kinerja metode *Support Vector Machine* tanpa seleksi fitur *Chi-Square* dan metode *Support Vector Machine* dengan seleksi fitur *Chi-Square* guna mendapatkan informasi mengenai metode yang memberikan hasil terbaik dalam melakukan analisis sentimen komentar di Instagram.
2. Hasil penelitian dapat dijadikan sebagai referensi untuk penelitian terkait serta dapat menjadi referensi bagi pemilik bisnis untuk mengetahui mayoritas respon publik terhadap produk dan jasa yang diiklankan.

1.6 Batasan Masalah

Beberapa batasan masalah yang telah ditetapkan untuk penelitian ini adalah :

1. Data masukan yang diolah merupakan gabungan komentar yang dikumpulkan secara langsung dari akun resmi *public figure* Indonesia di Instagram dengan pengikut lebih dari 500.000 pengguna.

2. Data diambil dari pos endorsemen dengan *likes* lebih dari 10.000 dan komentar lebih dari 300.
3. Data komentar yang digunakan adalah sebanyak 500 komentar berbahasa Indonesia.
4. Komentar dengan emoji atau emotikon tidak diikutsertakan.
5. Klasifikasi data dilakukan menjadi 2 kelas, yaitu positif dan negatif.
6. Fungsi kernel SVM yang digunakan yaitu kernel Linier, Polinomial dan RBF.
7. Masing-masing kernel SVM akan menggunakan nilai C 0,1 , 1, juga 10.
8. Seleksi fitur menggunakan metode *Chi-Square* dan menggunakan nilai taraf nyata α 0,05 dengan nilai kritis sebesar 3,84.

1.7 Sistematika Penulisan

Di bawah ini adalah sistematika penyusunan penelitian :

1.7.1 Bab I. Pendahuluan

Bab ini berisi tentang pokok pikiran yang menjadi landasan penelitian yang mencakup latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan.

1.7.2 Bab II. Kajian Literatur

Pada bab berikut, terdapat uraian meliputi dasar-dasar teori yang menjadi fundamental pembahasan penelitian secara terperinci, seperti definisi Instagram, endorsemen, analisis sentimen, algoritma *Support Vector Machine*,

algoritma *Chi-Square*, dan berbagai literatur ilmiah tentang penelitian lainnya yang berkaitan dengan topik penelitian.

1.7.3 Bab III. Metode Penelitian

Pada bab berikut, terdapat uraian yang berisi tentang tahapan pengembangan metodologi penelitian, seperti pengumpulan data yang digunakan, analisis data, mekanisme pengujian, dan desain sistem yang kemudian dikembangkan. Setiap tahap dalam penelitian diuraikan secara rinci dan dibuat berdasarkan suatu kerangka kerja yang telah ditentukan.

1.7.4 Bab IV. Pengembangan Perangkat Lunak

Bab ini mencakup desain dan analisis perangkat lunak yang akan digunakan sebagai alat penelitian. Proses ini dimulai dengan pengumpulan dan analisis kebutuhan, desain dan pengembangan perangkat lunak, hingga pengujian dengan tujuan untuk memastikan semua persyaratan pengembangan perangkat lunak terpenuhi.

1.7.5 Bab V. Hasil dan Analisa Penelitian

Bab ini memberikan uraian mengenai hasil pengujian yang dilakukan peneliti. Setiap hasil yang diperoleh akan disusun di dalam tabel dan dianalisis untuk mempermudah proses pengambilan kesimpulan penelitian tersebut.

1.7.6 Bab VI. Kesimpulan dan Saran

Bagian ini menyajikan ringkasan dari detail yang dibahas pada bab sebelumnya, serta memberikan rekomendasi berdasarkan hasil penelitian yang telah dilakukan.

1.8 Kesimpulan

Pada bab ini telah diuraikan secara rinci mengenai pokok pikiran yang menjadi landasan penelitian.

DAFTAR PUSTAKA

- Shrivatava, A., & Pant, B. (2012). *Opinion Extraction and Classification of Real Time Facebook Status*. *Opinion Extraction and Classification of Real Time Facebook Status*. <https://computerresearch.org/index.php/computer/article/view/544/544>
- Johnson, Joseph. (2021). *Worldwide Digital Population as of January 2021*. <https://www.statista.com/statistics/617136/digital-populationworldwide/#professional>
- Rizaty, M. A. (2021). *Inilah Negara Pengguna Instagram Terbanyak, Indonesia Urutan Berapa?* <https://databoks.katadata.co.id/datapublish/2021/08/03/inilah-negarapengguna-instagram-terbanyak-indonesia-urutan-berapa>
- Neri, F., Aliprandi, C., Capeci, F., Cuadros, M., & By, T. (2012). Sentiment analysis on social media. *Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012*, 919–926. <https://doi.org/10.1109/ASONAM.2012.164>
- Rozi, I. F., Pramono, S. H., & Dahlan, E. A. (2012). *Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi*. <https://jurnaleccis.ub.ac.id/index.php/eccis/article/view/164>
- Ilmawan, L. B., & Mude, M. A. (2020). Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store. *ILKOM Jurnal Ilmiah*, 12(2), 154–161. <https://doi.org/10.33096/ilkom.v12i2.597.154-161>
- Watmah, S. (2021). Komparasi Metode K-NN, Support Vector Machine, Dan Random Forest Pada E-Commerce Shopee. In *Sains Teknik* (Vol. 2, Issue 1). <http://jurnal.bsi.ac.id/index.php/insantek>
- Purnamasari, N. M. G. D., Ali Fauzi, M., Indriati, & Dewi, L. S. (2020). *Cyberbullying identification in twitter using support vector machine and information gain based seleksi fitur*. *Indonesian Journal of Electrical Engineering and Computer Science*, 18(3), 1494–1500. <https://doi.org/10.11591/ijeecs.v18.i3.pp1494-1500>
- Tsani, M., Rupaka, A., Asmoro, L., & Pradana, B. (2020). Analisis Sentimen Review Transportasi Menggunakan Algoritma Support Vector Machine

Berbasis *Chi-Square Smart Comp :Jurnalnya Orang Pintar Komputer*,
9(1), 35–39. <https://doi.org/10.30591/smartcomp.v9i1.1817>

Putra, Armanda Eka (2018). Pengaruh Seleksi Fitur *Chi-Square* Terhadap Kinerja Algoritma Naïve Bayes Classifier pada Analisis Sentimen Dokumen.
<https://repository.uinjkt.ac.id/dspace/handle/123456789/54952>

Azwar, M., & Sulthonah, S. (2018). *The Utilization of Instagram as a Media Promotion: the Case Study of Library in Indonesia*.
<https://journal.uinjkt.ac.id/index.php/insaniyat/article/view/7320/pdf>

Cahyani, Riska Indah. (2018). Peran Media Sosial Instagram Akun Akhyar TV dalam Penyebaran Dakwah.
<http://repository.radenintan.ac.id/3980/1/SKRIPSI.pdf>

Ipmawati, J., Kusriani, & Luthfi, E. T. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen.
<http://ijns.org/journal/index.php/ijns/article/view/1444>

Imron, A. (2019). Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naïve Bayes Classifier.
<https://dspace.uui.ac.id/handle/123456789/14268>

Ramya, M., & Pinakas, J. A. (2014). *Different Type of Seleksi fitur for Text Classification. International Journal of Computer Trends and Technology*, 10(2). <http://www.ijcttjournal.org>

Februariyanti, H., & Zuliarso, E. (2012). Klasifikasi Dokumen Berita Teks Bahasa Indonesia menggunakan Ontologi. *Jurnal Teknologi Informasi DINAMIK*, 17(1), 14–23. <https://doi.org/10.35315/dinamik.v17i1.1612>

Deolika, A., & Taufiq Luthfi, E. (2019). Analisis Pembobotan Kata pada Klasifikasi *Text Mining. Teknologi Informasi*, 3(2).
<http://jurnal.una.ac.id/index.php/jurti/article/view/1077>

Nikhath, A. K., Subrahmanyam, K., & Vasavi, R. (2016). *Building a K-Nearest Neighbor Classifier for Text Categorization*.
<https://ijcsit.com/docs/Volume%207/vol7issue1/ijcsit2016070156.pdf>

- Indra, M., Arsyah, H., Akbari, D., Novianty, A., & Setianingsih, C. (2017). Analisis Sentimen Menggunakan Metode *Learning Vector Quantization*.
https://openlibrary.telkomuniversity.ac.id/pustaka/files/135356/jurnal_eproc/analisis-sentimen-menggunakan-metode-learning-vectorquantization.pdf
- Monika Parapat, I., & Tanzil Furqon, M. (2018). Penerapan Metode *Support Vector Machine* (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak (Vol. 2, Issue 10). <http://j-ptiik.ub.ac.id>
- Munawarah, R., Soesanto, O., Reza Faisal, M., Yani Km, J. A., & selatan, K. (2016). Penerapan Metode *Support Vector Machine* pada Diagnosa Hepatitis. <http://klik.ulm.ac.id/index.php/klik/article/view/39>
- Saniyah (2019). *Named Entity Recognition* pada Teks Berita Menggunakan *Support Vector Machine*
- Ling, J., Putu, I., Kencana, E. N., & Oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier dengan Seleksi Fitur *Chi-Square*. <https://ojs.unud.ac.id/index.php/mtk/article/view/11992>
- Arum Sari, Y. (2018). Analisis Sentimen Pada Review Konsumen Menggunakan Metode Naive Bayes Dengan Seleksi Fitur Chi Square Untuk Rekomendasi Lokasi Makanan Tradisional. <http://j-ptiik.ub.ac.id>
- Putri, A. D. (2013). Klasifikasi Dokumen Teks Menggunakan Metode *Support Vector Machine* dengan Pemilihan Fitur *Chi-Square*.
- Fitria, A., & Widowati, H. (2017). Implementasi Metode *Rational Unified Process* dalam Pengembangan Sistem Administrasi Kependudukan.
<https://ejournal.gunadarma.ac.id/index.php/tekno/article/view/1604>
- Hidayatullah (2015). Analisis Sentimen dan Klasifikasi Kategori terhadap Tokoh Publik pada Twitter.
<http://jurnal.upnyk.ac.id/index.php/semnasif/article/view/1016>