

**ANALISIS SENTIMEN PADA REVIEW SESTYC  
MENGUNAKAN METODE SUPPORT VECTOR MACHINE,  
NAIVE BAYES, DAN LOGISTIC REGRESSION**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**MUHAMMAD HASAN S J**

**17.11.1523**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**ANALISIS SENTIMEN PADA REVIEW SESTYC  
MENGUNAKAN METODE SUPPORT VECTOR MACHINE,  
NAIVE BAYES, DAN LOGISTIC REGRESSION**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**MUHAMMAD HASAN S J**

17.11.1523

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**ANALISIS SENTIMEN PADA REVIEW SESTYC  
MENGUNAKAN METODE SUPPORT VECTOR MACHINE,  
NAIVE BAYES, DAN LOGISTIC REGRESSION**

yang disusun dan diajukan oleh

**Muhammad Hasan S J**

**17.11.1523**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 19 April 2022

**Dosen Pembimbing,**

**Barka Satya, M.Kom**

**NIK. 190302126**

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**ANALISIS SENTIMEN PADA REVIEW SESTYC**  
**MENGGUNAKAN METODE SUPPORT VECTOR MACHINE,**  
**NAIVE BAYES, DAN LOGISTIC REGRESSION**

yang disusun dan diajukan oleh

**Muhammad Hasan S J**

**17.11.1523**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 19 April 2022

**Susunan Dewan Penguji**

**Nama Penguji**

**Yoga Prisyanto, S.Kom, M.Eng**  
**NIK. 1903022412**

**Hanif Al Fatta, S.Kom., M.Kom**  
**NIK. 190302096**

**Majid Rahardi, S.kom., M.Eng**  
**NIK. 190302393**

**Tanda Tangan**

---

---

---

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 19 April 2022

**DEKAN FAKULTAS ILMU KOMPUTER**



**Hanif Al Fatta, S.Kom., M.Kom.**  
**NIK. 190302096**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

### HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Muhammad Hasan S J  
NIM : 17.11.1523

Menyatakan bahwa Skripsi dengan judul berikut:

#### Tuliskan Judul Skripsi

Dosen Pembimbing : Barka Satya, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 Januari 2022

Yang Menyatakan,



Muhammad Hasan S J

## HALAMAN PERSEMBAHAN

Dengan mengucapkan syukur Alhamdulillah saya persembahkan skripsi ini kepada semua pihak yang terlibat secara langsung atau tidak langsung dalam proses pembuatan skripsi.

1. Kedua orang tua, kakak dan adik saya, yang selalu mendoakan dan memberikan semangat serta motivasi tiada henti.
2. Dosen pembimbing saya Bapak Barka Satya, M.Kom. Yang telah membimbing saya dari awal sampai akhir pembuatan skripsi
3. Dosen-dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu selama kuliah.
4. Teman-teman kelas 17-IF-09 yang telah menemani dan selalu memberikan semangat untuk menyelesaikan skripsi

## KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya dan shawalat serta salam juga tidak lupa penulis panjatkan kepada junjungan kita Nabi Muhammad SAW yang telah memberikan teladan mulia dalam menuntun umatnya, sehingga penulis dapat menyelesaikan skripsi dengan judul “ANALISIS SENTIMEN PADA REVIEW SESTYC MENGGUNAKAN METODE SUPPORT VECTOR MACHINE, NAIVE BAYES, DAN LOGISTIC REGRESSION” dengan sebaik – baiknya sebagai salah satu syarat utama untuk menyelesaikan program sarjana pada Universitas AMIKOM Yogyakarta

Penyelesaian skripsi ini juga tidak lepas dari bantuan berbagai pihak, karena itu pada kesempatan ini penulis ingin menyampaikan terima kasih kepada:

1. Prof. Dr. M. Suyanto, MM selaku Rektor Universitas AMIKOM Yogyakarta
2. Bapak Barka Satya, M.Kom. selaku dosen pembimbing yang telah banyak memberikan nasehat dan arahan bagi penulis agar penulis menyelesaikan skripsi ini dengan baik.
3. Para Dosen dan Staff Universitas AMIKOM Yogyakarta yang telah banyak memberikan ilmu pengetahuan, pengalaman, dan bantuannya selama penulis kuliah hingga terselesaikannya skripsi ini.
4. Keluarga dan teman-teman yang telah memberikan dukungan dan do'a kepada penulis. Penulis menyadari masih banyak kekurangan yang masih

harus diperbaiki dari penulisan skripsi ini. Penulis menerima kritik dan saran yang membangun

Bagian Penulis Menyadari Masih banyak kekurangan yang harus diperbaiki dari penulis skripsi ini. Penulis menerima kritik dan saran yang membangun untuk kesempurnaan skripsi ini. Semoga tulisan ini bisa bermanfaat dan mendorong kita untuk melakukan penelitian yang lebih baik.

Yogyakarta, 20/1/2023

Penulis





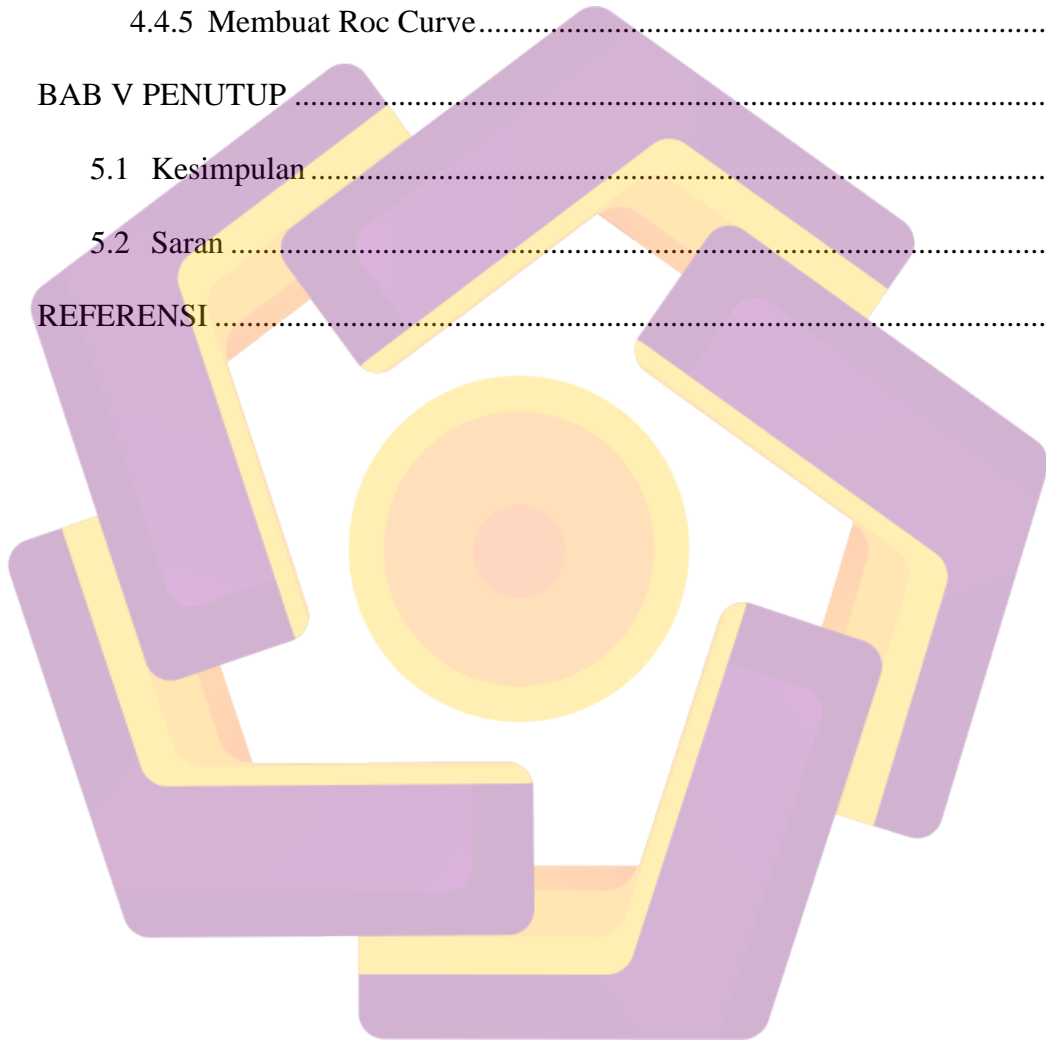
## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	iv
HALAMAN PERSEMBAHAN .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR .....	xiii
INTISARI .....	xiv
ABSTRACT.....	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 .....	4
1.6 Sistematika Penulisan .....	5
BAB II TINJAUAN PUSTAKA .....	6
2.1 Tinjauan Pustaka.....	6
2.2 Media Sosial .....	9
2.3 Sestyc .....	9

2.4	Google Play.....	9
2.5	Analisis Sentimen .....	10
2.6	Online Review .....	10
2.7	Data Mining dan Klasifikasi .....	11
2.8	Text Mining .....	13
2.8.1	Spelling Normalization .....	13
2.8.2	Case Folding.....	13
2.8.3	Tokenizing.....	14
2.8.4	Filterring.....	14
2.9	Term Frequency-Invers Document Frrequency(TF-IDF).....	14
2.10	Evaluasi Sistem Klasifikasi .....	16
2.10.1	K-Flod Cross Validation .....	16
2.10.2	Precision .....	17
2.10.3	Recall.....	17
2.10.4	F-measure .....	17
2.10.4	Accuracy.....	17
2.11	SVM (Support Vector Machine).....	18
2.11.1	SVM Lineary Separabel Data.....	18
2.11.2	SVM Lineary Separabel Data.....	22
2.12	Naïve Bayes .....	25
2.13	Logistic Regression.....	26

2.14 Kurva Roc .....	28
<b>BAB III METODE PENELITIAN .....</b>	<b>30</b>
3.1 Alat Penelitian.....	30
3.2 Tahap Penelitian .....	32
3.3 Objek dan Subjek Penelitian.....	33
3.4 Metode Pengumpulan Data.....	33
3.5 Jenis Dan Sumber Data.....	33
3.6 Pelabelan Data .....	33
3.7 Preprocessing Data .....	34
3.8 Pembobotan TD-IDF .....	34
3.9 Pelatihan Dan Pengujian Model .....	34
3.10 Evaluasi dan Hasil .....	34
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>36</b>
4.1 Pengumpulan Data.....	36
4.2 Pelabelan Data .....	38
4.3 Teks Preprocessing .....	40
4.3.1 Case Folding .....	40
4.3.2 Tokenizing .....	42
4.3.3 Filtering (Stop Removal) .....	42
4.3.4 Spelling Normalization.....	43
4.3.5 Stemming.....	44
4.3.3 Pembobotan TD-IDF .....	44
4.4 Klasifikasi Sentimen.....	45

4.4.1 Data Latih dan Data Uji.....	46
4.4.2 Klasifikasi dengan <i>Support Vector Machine</i> .....	46
4.4.3 Klasifikasi dengan Naïve Bayes .....	49
4.4.4 Klasifikasi dengan Logistic Regression.....	52
4.4.5 Membuat Roc Curve.....	54
<b>BAB V PENUTUP</b> .....	<b>55</b>
5.1 Kesimpulan.....	55
5.2 Saran .....	55
<b>REFERENSI</b> .....	<b>57</b>



## DAFTAR TABEL

<b>Tabel 2. 1</b> Keaslian Penelitian.....	8
<b>Tabel 2. 2</b> Nilai Auc sebagai Kategori Klasifikasi.....	28
<b>Tabel 3. 1</b> Perangkat keras yang digunakan.....	30
<b>Tabel 3. 2</b> Spesifikasi minimal yang dibutuhkan.....	30
<b>Tabel 3. 3</b> Perangkat lunak yang digunakan.....	31
<b>Tabel 4. 1</b> Proses Perubahan Lowercase.....	41
<b>Tabel 4. 2</b> Proses menghapus angka dan tanda baca.....	41
<b>Tabel 4. 3</b> Proses pemisahan teks / <i>tokenizing</i> .....	42
<b>Tabel 4. 4</b> Proses penyaringan kata/ <i>filterring</i> .....	43
<b>Tabel 4. 5</b> Proses <i>Spelling Normalization</i> .....	44
<b>Tabel 4. 6</b> Nilai dari kernel <i>svm</i> .....	47
<b>Tabel 4. 7</b> Hasil dari <i>svm k-flod 5 dan 10</i> .....	48
<b>Tabel 4. 8</b> Hasil akurasi dari kernel <i>naive bayes</i> .....	50
<b>Tabel 4. 9</b> Hasil dari <i>svm k-flod 5 dan 10</i> .....	50
<b>Tabel 4. 10</b> Hasil dari <i>Logistic Regression k-flod 5 dan 10</i> .....	52

## DAFTAR GAMBAR

<b>Gambar 1.1</b> Daftar media sosial yang sering digunakan di Indonesia .....	1
<b>Gambar 2. 1</b> Alternatif Bidang Pemisah Terbaik .....	19
<b>Gambar 2. 2</b> Soft Margin Hyperlane .....	23
<b>Gambar 2. 3</b> Metode Klasifikasi Naïve Bayes Classifier .....	26
<b>Gambar 2. 4</b> Metode Klasifikasi Logistic Regression .....	27
<b>Gambar 3. 1</b> Flowchart Tahap Penelitian .....	32
<b>Gambar 4. 1</b> Code untuk menginstall google-play scraper .....	36
<b>Gambar 4. 2</b> Code untuk melakukan pengambilan data .....	37
<b>Gambar 4. 3</b> Code untuk membentuk data menjadi colom .....	37
<b>Gambar 4. 4</b> Code untuk meyimpan data kedalam csv .....	38
<b>Gambar 4. 5</b> Hasil data yang didapatkan .....	38
<b>Gambar 4. 6</b> Code untuk melakukan pelabelan data .....	39
<b>Gambar 4. 7</b> Jumlah data yang berlabel positif dan negatif .....	39
<b>Gambar 4. 8</b> Code untuk melakukan resampling data .....	40
<b>Gambar 4. 9</b> Code untuk melakukan Vectorizer .....	45
<b>Gambar 4. 10</b> <i>Confusion matrix</i> pada svm .....	49
<b>Gambar 4. 11</b> <i>Confusion matrix</i> pada naive bayes .....	51
<b>Gambar 4. 12</b> <i>Confusion matrix</i> pada logisctic regression .....	53
<b>Gambar 4. 13</b> akurasi dari beberapa metode menggunakan ROC CURVE .....	54

## INTISARI

Pertumbuhan pengguna internet di Indonesia mengalami penambahan yang sangat tinggi setiap tahunnya, dengan meningkatnya pengguna internet di Indonesia jugamengakibatkan banyaknya masyarakat yang menggunakan media sosial. Sestyc merupakan aplikasi media sosial yang dibuat oleh sekelompok anak milineal di Indonesia. penelitian ini dilakukan untuk menganalisis sentimen pengguna sestyc menggunakan data teks berupa *review* yang diperoleh dari google play store. Tujuan dilakukan penelitian ini untuk menganalisis sentimen terhadap aplikasi sestyc serta menemukan algoritma terbaik dalam mengklasifikasi sentimen. Algoritma yang digunakan dalam menganalisis sentimen pada penelitian ini terdiri dari *Support VectorMachine*, *Logistic regression*, dan *Naive Bayes*. Hasil pelabelan kelas sentimen pada data review sestyc diperoleh 8000 *review* dengan jumlah *review* positif sebanyak 4719 dan *review* negatif sebanyak 3281. Hasil dari penelitian ini menunjukkan bahwa algoritma *Support Vector Machine* memiliki nilai akurasi paling tinggi dibandingkan dengan algoritma lainnya, dimana *Support Vector Machine* mendapatkan nilai akurasi sebesar 87,81%.

**Kata kunci:** Analisis Sentimen, Sestyc, *Support Vector Machine*, *Naive Bayes*, *Logistic Regression*

## ABSTRACT

*The growth of internet users in Indonesia experiences a very high increase every year, with the increase in internet users in Indonesia also resulting in many people using social media. Sestyc is a social media application created by a group of millennial children in Indonesia. This research was conducted to analyze the sentiment of users of Sestyc using text data in the form of a review obtained from the Google Play Store. The purpose of this research is to analyze sentiment towards the sestyc application and find the best algorithm for classifying sentiment. The algorithm used in analyzing sentiment in this study consists of Support Vector Machine, Logistic regression, and Naive Bayes. The results of sentiment class labeling on the sestyc review data obtained 8000 reviews with a total of 4719 positive reviews and 3281 negative reviews. The results of this study indicate that the Support Vector Machine algorithm has the highest accuracy value compared to other algorithms, where the Support Vector Machine gets an accuracy value. by 87.81%.*

**Keyword:** *Sentiment Analysis, Sestyc, Support Vector Machine, Naive Bayes, Logistic Regression*