

**ANALISIS SENTIMEN PADA KOMENTAR VIDEO ULASAN  
*SMARTPHONE* DI YOUTUBE**

Skripsi



Diajukan oleh:

KEVIN GIOVANNI

71150046

PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA  
YOGYAKARTA

2019

**ANALISIS SENTIMEN PADA KOMENTAR VIDEO ULASAN  
SMARTPHONE DI YOUTUBE**

Skripsi



Diajukan kepada Fakultas Teknologi Informasi Program Studi Informatika  
Universitas Kristen Duta Wacana

Sebagai salah satu syarat dalam memperoleh gelar Sarjana Komputer

Diajukan oleh:

**KEVIN GIOVANNI**

**71150046**

PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA  
YOGYAKARTA

2019

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

### ANALISIS SENTIMEN PADA KOMENTAR VIDEO ULASAN SMARTPHONE DI YOUTUBE

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi keserjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar keserjanaan saya.

Yogyakarta, 14 Juni 2019



KEVIN GIOVANNI  
71150046

## HALAMAN PERSETUJUAN

Judul Skripsi : ANALISIS SENTIMEN PADA KOMENTAR  
VIDEO ULASAN SMARTPHONE DI YOUTUBE  
Nama Mahasiswa : KEVIN GIOVANNI  
NIM : 71150046  
Matakuliah : Skripsi (Tugas Akhir)  
Kode : TIW276  
Semester : Genap  
Tahun Akademik : 2018/2019

Telah diperiksa dan disetujui di  
Yogyakarta,  
Pada tanggal 14 Juni 2019

Dosen Pembimbing I



Lucia Dwi Krisnawati, Dr. Phil.

Dosen Pembimbing II



Antonius Rachmat C., S.Kom., M.Cs.



## HALAMAN PENGESAHAN

### ANALISIS SENTIMEN PADA KOMENTAR VIDEO ULASAN SMARTPHONE DI YOUTUBE

Oleh: KEVIN GIOVANNI / 71150046

Dipertahankan di depan Dewan Penguji Skripsi  
Program Studi Informatika Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana - Yogyakarta  
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar  
Sarjana Komputer  
pada tanggal 12 Juni 2019


Yogyakarta, 14 Juni 2019  
Mengesahkan,

Dewan Penguji:

1. Lucia Dwi Krisnawati, Dr. Phil.
2. Antonius Rachmat C., S.Kom., M.Cs.
3. R. Gunawan Santosa, Drs. M.Si.
4. Lukas Chrisantyo, S.Kom., M.Eng.



Dekan  
  
(Buch Susanto, S.Kom., M.T.)

Ketua Program Studi  
  
(Gloria Virginia, Ph.D.)

## DAFTAR ISI

<b>PERNYATAAN KEASLIAN SKRIPSI</b> .....	Error! Bookmark not defined.
<b>HALAMAN PERSETUJUAN</b> .....	Error! Bookmark not defined.
<b>HALAMAN PENGESAHAN</b> .....	Error! Bookmark not defined.
<b>DAFTAR ISI</b> .....	<b>iv</b>
<b>DAFTAR TABEL</b> .....	<b>vi</b>
<b>DAFTAR GAMBAR</b> .....	<b>viii</b>
<b>UCAPAN TERIMA KASIH</b> .....	<b>x</b>
<b>INTISARI</b> .....	<b>xi</b>
<b>BAB I</b> .....	<b>1</b>
<b>PENDAHULUAN</b> .....	<b>1</b>
1.1. Latar Belakang Masalah .....	1
1.2. Perumusan Masalah .....	2
1.3. Batasan Masalah .....	3
1.4. Tujuan Penelitian .....	3
1.5. Metodologi Penelitian.....	4
1.6. Sistematika Penulisan .....	5
<b>BAB II</b> .....	<b>6</b>
<b>LANDASAN TEORI</b> .....	<b>6</b>
2.1. Tinjauan Pustaka .....	6
2.2. Landasan Teori .....	8
2.2.1. <i>Text Mining</i> dan Analisis Sentimen .....	8
2.2.2. Ekstraksi Fitur .....	10
2.2.3. Credibility Adjusted Term Frequency.....	11
2.2.4. <i>Support Vector Machine</i> .....	13
2.2.5. <i>Hyperplane</i> .....	13
2.2.6. <i>Kernel Trick</i> .....	14
2.2.7. Evaluasi Sistem .....	20
<b>BAB III</b> .....	<b>22</b>
<b>PERANCANGAN SISTEM</b> .....	<b>22</b>
3.1. Perancangan Kebutuhan Sistem .....	22
3.1.1. Kebutuhan Non-Fungsional .....	22
3.1.2. Kebutuhan Fungsional .....	22
3.2. Blok Diagram Sistem.....	23
3.3. Perancangan Pengolahan Data.....	26
3.3.1. Struktur Data .....	26
3.3.2. Penyimpanan Data .....	27
3.4. Perancangan Antarmuka .....	27
3.4.1. Program Pengumpulan Data.....	27
3.4.2. Program Analisis Sentimen Komentar .....	28
3.5. Perancangan Evaluasi Sistem .....	30

3.6.	Perancangan Pengujian Sistem.....	30
<b>BAB IV</b>	.....	<b>31</b>
<b>IMPLEMENTASI DAN ANALISIS SISTEM</b>	.....	<b>31</b>
4.1.	Implementasi Antarmuka Sistem.....	31
4.1.1.	Halaman <i>Home</i> .....	31
4.1.2.	Halaman <i>Crawling</i> .....	31
4.1.3.	Halaman <i>Overview</i> .....	32
4.1.4.	Halaman Korpus.....	33
4.2.	Implementasi Sistem.....	34
4.2.1.	Implementasi Pengumpulan Data.....	35
4.2.2.	Implementasi Prapemrosesan .....	36
4.2.3.	Implementasi Pembobotan Kata.....	37
4.2.4.	Implementasi Pembagian <i>Dataset</i> .....	39
4.2.5.	Implementasi Klasifikasi SVM .....	40
4.2.6.	Implementasi Pengujian Sistem .....	41
4.3.	Hasil Pengujian.....	42
4.3.1.	Hasil Pengujian Menggunakan Data Tidak Seimbang.....	42
4.3.2.	Hasil Pengujian Menggunakan Data Seimbang .....	48
<b>BAB V</b>	.....	<b>55</b>
<b>KESIMPULAN DAN SARAN</b>	.....	<b>55</b>
5.1.	Kesimpulan.....	55
5.2.	Saran .....	57
<b>DAFTAR PUSTAKA</b>	.....	<b>58</b>

## DAFTAR TABEL

2.1.	Perbandingan Rata-rata Tingkat Akurasi	
	Hasil Pengujian .....	8
2.2.	Contoh Perhitungan TF .....	11
2.3.	Contoh Data Latih SVM <i>Linear</i> .....	15
2.4.	Proses Klasifikasi Menggunakan	
	Persamaan <i>Hyperplane</i> .....	18
2.5.	<i>Confusion Matrix</i> .....	20
2.6.	Contoh <i>Confusion Matrix</i> .....	21
3.1.	Contoh Tahap <i>Case Folding</i> .....	24
3.2.	Contoh Kata Tidak Baku dan Kata Baku .....	24
3.3.	Contoh <i>Stopwords</i> dalam Bahasa Indonesia .....	25
3.4.	Perancangan Skenario Pengujian Sistem .....	30
4.1.	Hasil Pengujian Pertama Pada Data Tidak	
	Seimbang dengan <i>Kernel Linear</i> dan RBF .....	43
4.2.	Hasil Pengujian Kedua Pada Data Tidak	
	Seimbang dengan <i>Kernel Linear</i> dan RBF .....	43
4.3.	Hasil Pengujian Ketiga Pada Data Tidak	
	Seimbang dengan <i>Kernel Linear</i> dan RBF .....	44
4.4.	Hasil Pengujian Keempat Pada Data Tidak	
	Seimbang dengan <i>Kernel Linear</i> dan RBF .....	44
4.5.	Hasil Pengujian Kelima Pada Data Tidak	
	Seimbang dengan <i>Kernel Linear</i> dan RBF .....	45
4.6.	Rata-rata Seluruh Hasil Pengujian Pada	
	Data Tidak Seimbang .....	46
4.7.	Hasil Pengujian Pertama Pada Data	
	Seimbang dengan <i>Kernel Linear</i> dan RBF .....	49
4.8.	Hasil Pengujian Kedua Pada Data	
	Seimbang dengan <i>Kernel Linear</i> dan RBF .....	50
4.9.	Hasil Pengujian Ketiga Pada Data	



	Seimbang dengan <i>Kernel Linear</i> dan RBF .....	50
4.10.	Hasil Pengujian Keempat Pada Data	
	Seimbang dengan <i>Kernel Linear</i> dan RBF .....	51
4.11.	Hasil Pengujian Kelima Pada Data	
	Seimbang dengan <i>Kernel Linear</i> dan RBF .....	51
4.12.	Rata-rata Seluruh Hasil Pengujian	
	Pada Data Seimbang .....	52

©UKDW

## DAFTAR GAMBAR

2.1.	Tahapan Pada Proses <i>Text Mining</i> .....	8
2.2.	Contoh <i>Hyperplane</i> .....	13
2.3.	Data dengan Persebaran yang Tidak <i>Linear</i> .....	14
2.4.	Data pada Gambar 2.3. Dipisahkan oleh Sebuah <i>Hyperplane</i> .....	14
2.5.	Visualisasi Data pada Tabel 2.3. ....	16
2.6.	Visualisasi <i>Hyperplane</i> untuk Data Uji .....	18
2.7.	Data yang Sulit Dipisahkan oleh <i>Polynomial Kernel</i> .....	19
2.8.	RBF <i>Kernel</i> Mampu Memisahkan Kelas Bintang dan Segitiga dengan Benar .....	19
3.1.	Diagram Alur Sistem yang akan Dibangun .....	23
3.2.	Perancangan Arsitektur Sistem yang akan Dibangun .....	26
3.3.	<i>Entity Relationship Diagram</i> (ERD) Sistem yang akan dibangun .....	27
3.4.	Perancangan Antarmuka Sistem Pengumpulan Data .....	28
3.5.	Perancangan Antarmuka Sistem Pengumpulan Data (2) .....	28
3.6.	Perancangan Antarmuka Sistem Klasifikasi Komentar .....	29
3.7.	Perancangan Antarmuka Sistem Klasifikasi Komentar (2) .....	29
3.8.	Perancangan Antarmuka Sistem Klasifikasi Komentar (3) .....	29
4.1.	Halaman <i>Home</i> .....	31
4.2.	Halaman <i>Crawling</i> .....	32
4.3.	Halaman <i>Overview</i> .....	33
4.4.	Halaman Korpus .....	34

4.5.	Contoh Komentar yang Tidak Relevan dengan Konten Video .....	35
4.6.	Gambaran Proses Prapemrosesan Komentar .....	36
4.7.	Contoh Format Data Hasil Proses <i>Binary Classification</i> .....	37
4.8.	<i>Pseudocode</i> Pembobotan Kata .....	38
4.9.	<i>Pseudocode Feature Selection</i> .....	39
4.10.	Kode Program untuk Membagi Data Latih dan Data Uji .....	39
4.11.	Kode Program untuk Membuat Objek SVC .....	40
4.12.	Kode Program untuk Melakukan Proses Pelatihan .....	41
4.13.	Kode Program untuk Menyimpan Objek SVC ke dalam <i>File</i> .....	41
4.14.	Kode Program untuk Melakukan Pengujian Sistem Klasifikasi .....	42
4.15.	Grafik Performa Sistem Berdasarkan <i>Feature Selection</i> untuk Data Tidak Seimbang .....	47
4.16.	Grafik Perbandingan Jumlah Data Komentar Tidak Seimbang .....	48
4.17.	Grafik Performa Sistem Berdasarkan <i>Feature Selection</i> untuk Data Seimbang .....	53
4.18.	Grafik Perbandingan Jumlah Data Komentar untuk Data Tidak Seimbang .....	54

## UCAPAN TERIMA KASIH

Puji syukur dipanjatkan kepada Tuhan Yang Maha Esa atas rahmat dan penyertaan-Nya sehingga proses penelitian dengan judul “Analisis Sentimen Pada Komentar Video Ulasan *Smartphone* di YouTube” dapat terlaksana dengan lancar.

Laporan ini dibuat untuk memenuhi persyaratan tugas Mata Kuliah Skripsi di Program Studi Informatika, Fakultas Teknologi Informasi Universitas Kristen Duta Wacana Yogyakarta. Tujuan dibuatnya laporan skripsi ini yaitu untuk menjabarkan proses penelitian yang dilakukan oleh penulis dari awal hingga akhir.

Dalam penyusunan laporan skripsi, tentu tak lepas dari pengarahan dan bimbingan dari berbagai pihak. Maka penulis mengucapkan terima kasih sebesar-besarnya kepada semua pihak yang membantu, di antaranya sebagai berikut:

1. Ibu Lucia Dwi Krisnawati, Dr. Phil., dan Bapak Antonius Rachmat C., S.Kom., M.Cs., selaku dosen pembimbing Mata Kuliah Skripsi.
2. Bapak Budi Susanto, S.Kom., M.T. selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
3. Ibu Gloria Virginia, Ph.D. selaku Kepala Program Studi Informatika Universitas Kristen Duta Wacana.
4. Kepada orang tua atas doa dan dukungannya sehingga proses penelitian dapat berjalan dengan lancar.

Berkat kerjasama semua pihak yang terlibat dalam proses penelitian ini maka penulis bisa menyelesaikan laporan skripsi dengan sebaik-baiknya. Laporan skripsi ini memang masih jauh dari kesempurnaan, tapi penulis sudah berusaha sebaik mungkin. Semoga laporan ini bermanfaat bagi semua pihak.

## INTISARI

YouTube merupakan salah satu media sosial yang memiliki daya tarik yang tinggi di masyarakat. YouTube memiliki konten yang sangat beragam, penelitian ini menitikberatkan analisis sentimen komentar mengenai video ulasan *smartphone* di YouTube. Adapun *channel* YouTube yang digunakan sebagai sumber data adalah “Sobat HAPE” dan “DroidLime” yang merupakan *channel* pengulas *smartphone* asal Indonesia yang sudah diverifikasi oleh YouTube. Sistem klasifikasi sentimen yang dibangun bertujuan untuk membantu pengguna dalam mengetahui bagaimana sentimen masyarakat terhadap suatu video ulasan *smartphone*.

Peneliti ini mencoba untuk menerapkan metode pembobotan *term Credibility Adjusted Term Frequency* (CATF) pada model klasifikasi *Support Vector Machine* (SVM) dengan menggunakan *kernel linear* dan RBF. Pada proses pengujian, peneliti membandingkan performa sistem klasifikasi untuk *dataset* dengan jumlah data komentar yang seimbang pada setiap kelasnya dan yang tidak seimbang. Selain itu, peneliti juga membandingkan performa *kernel linear* dan RBF. Parameter evaluasi yang digunakan pada penelitian ini adalah akurasi dan presisi yang didapatkan dari *confusion matrix*.

CATF dapat digunakan sebagai metode pembobotan pada model klasifikasi SVM untuk klasifikasi sentimen komentar ulasan *smartphone* di YouTube. Sistem ini menghasilkan nilai akurasi rata-rata tertinggi 79,2% pada data seimbang, dan 65,8% pada data tidak seimbang. *Kernel linear* menghasilkan sistem klasifikasi yang optimal dengan akurasi yang tinggi, meskipun begitu sistem klasifikasi yang seimbang diraih menggunakan *kernel RBF*.

**Kata kunci:** Analisis Sentimen, Sistem Klasifikasi, *Support Vector Machine*, *Credibility Adjusted Term Frequency*

## INTISARI

YouTube merupakan salah satu media sosial yang memiliki daya tarik yang tinggi di masyarakat. YouTube memiliki konten yang sangat beragam, penelitian ini menitikberatkan analisis sentimen komentar mengenai video ulasan *smartphone* di YouTube. Adapun *channel* YouTube yang digunakan sebagai sumber data adalah “Sobat HAPE” dan “DroidLime” yang merupakan *channel* pengulas *smartphone* asal Indonesia yang sudah diverifikasi oleh YouTube. Sistem klasifikasi sentimen yang dibangun bertujuan untuk membantu pengguna dalam mengetahui bagaimana sentimen masyarakat terhadap suatu video ulasan *smartphone*.

Peneliti ini mencoba untuk menerapkan metode pembobotan *term Credibility Adjusted Term Frequency* (CATF) pada model klasifikasi *Support Vector Machine* (SVM) dengan menggunakan *kernel linear* dan RBF. Pada proses pengujian, peneliti membandingkan performa sistem klasifikasi untuk *dataset* dengan jumlah data komentar yang seimbang pada setiap kelasnya dan yang tidak seimbang. Selain itu, peneliti juga membandingkan performa *kernel linear* dan RBF. Parameter evaluasi yang digunakan pada penelitian ini adalah akurasi dan presisi yang didapatkan dari *confusion matrix*.

CATF dapat digunakan sebagai metode pembobotan pada model klasifikasi SVM untuk klasifikasi sentimen komentar ulasan *smartphone* di YouTube. Sistem ini menghasilkan nilai akurasi rata-rata tertinggi 79,2% pada data seimbang, dan 65,8% pada data tidak seimbang. *Kernel linear* menghasilkan sistem klasifikasi yang optimal dengan akurasi yang tinggi, meskipun begitu sistem klasifikasi yang seimbang diraih menggunakan *kernel RBF*.

**Kata kunci:** Analisis Sentimen, Sistem Klasifikasi, *Support Vector Machine*, *Credibility Adjusted Term Frequency*

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Internet menyediakan informasi yang tidak terbatas, tidak terkecuali dengan informasi mengenai *smartphone*. Ketika ingin membeli *smartphone*, sering kali YouTube digunakan sebagai bahan pertimbangan sebelum memilih *smartphone* yang sesuai dengan kriteria yang ingin dibeli. Tidak hanya melihat video ulasan *smartphone* yang ingin dibeli, pembeli juga tentu melihat komentar orang lain terhadap video tersebut untuk mengetahui opini orang lain terhadap video tersebut. Hal ini dilakukan secara manual dengan cara membaca komentar pada suatu video satu demi satu dan menyimpulkan sendiri apakah produk tersebut layak atau tidak untuk dibeli. Namun, proses tersebut tentu saja membutuhkan banyak waktu dan tenaga karena dilakukan secara manual. Untuk mengatasi masalah ini, dibutuhkan sistem yang mampu menilai secara otomatis apakah *smartphone* tersebut layak untuk dibeli berdasarkan sentimen komentar yang terhadap video ulasannya. Sentimen positif menunjukkan *smartphone* layak untuk dibeli, sedangkan sentimen negatif menunjukkan *smartphone* tidak layak untuk dibeli.

Pada penelitian ini, peneliti akan meneliti sentimen komentar pada video ulasan *smartphone* di YouTube. YouTube adalah sebuah situs web berbagi video yang dibuat pada Februari 2005 oleh tiga mantan karyawan PayPal, yaitu Jawed Karim, dan Steve Chen. Situs ini memungkinkan pengguna mengunggah, menonton, dan berbagi video ke media sosial lainnya. Menurut Smith (2018), YouTube diakses 1 miliar jam sehari secara akumulasi seluruh dunia. Hal ini menunjukkan bagaimana YouTube sudah menjadi gaya hidup bagi sebagian orang di dunia. Orang akan cenderung mencari informasi apapun melalui YouTube.

Selain Youtube, media sosial lainnya yang kerap dijadikan sumber untuk analisis sentimen adalah Facebook. Menurut Donnelly (2018), Facebook memiliki user aktif sebanyak 2,01 miliar pada 30 Juni 2017. Dengan jumlah user aktif sebanyak ini tentu saja Facebook memiliki data yang sangat banyak yang dapat dijadikan sebagai sumber untuk penelitian seperti pada penelitian yang dilakukan

oleh Kurniawan (2017). Kurniawan (2017) yang meneliti opini di media sosial Facebook terhadap Pemilu Presiden 2014 dengan fitur *Term Frequency-Inverse Document* (TF-IDF). Akurasi yang dihasilkan oleh sistem ini sangat tinggi, yaitu 93, 84%. Kurniawan (2017) menggunakan pendekatan *Support Vector Machine* untuk membuat sistem tersebut dengan *kernel* RBF. Penelitian lainnya dilakukan oleh Ling, Kencana, dan Oka (2014) yang menggunakan algoritme *Naïve Bayes Classifier* mampu menghasilkan akurasi sentimen sebesar 83% dengan ekstraksi fitur *Chi Square*. Akurasi tersebut sebenarnya dapat ditingkatkan dengan menerapkan *n-gram*.

Peneliti akan menggunakan pendekatan model *Support Vector Machine* (SVM) untuk menganalisis sentimen pada komentar video ulasan *smartphone* di Youtube dengan alasan bahwa penelitian sebelumnya menyatakan bahwa metode ini memiliki akurasi tertinggi dibandingkan dengan metode lainnya. Diharapkan pada sistem ini juga menghasilkan analisis sentimen yang memiliki akurasi dan efektifitas yang tinggi.

Perbedaan penelitian ini dengan penelitian yang dilakukan oleh Kurniawan (2017) adalah fitur yang digunakan, Kurniawan (2017) menggunakan TF-IDF sedangkan peneliti menggunakan *Credibility Adjusted Term Frequency* (CATF) dan *n-gram* serta data yang digunakan adalah objek yang berbeda. Penelitian Kurniawan (2017) menggunakan data pemilu presiden 2014, sedangkan peneliti menggunakan data komentar video ulasan *smartphone*. Adapun *kernel* yang akan digunakan oleh peneliti adalah *linear* dan RBF.

## 1.2. Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, peneliti menemukan permasalahan yang dihadapi, yaitu:

- a. Bagaimana implementasi pembobotan term dengan metode CATF menggunakan metode klasifikasi SVM terhadap analisis sentimen komentar video ulasan *smartphone* di Youtube?
- b. Bagaimana kinerja analisis sentimen dengan pembobotan CATF dan metode klasifikasi SVM dilihat dari tingkat presisi dan akurasinya?



### 1.3. Batasan Masalah

Dalam melakukan penelitian ini tentunya perlu ditetapkan batasan-batasan untuk mempertahankan fokus penelitian. Penelitian ini menitikberatkan pada analisis sentimen komentar pada video ulasan *smartphone* yang memberikan opini terhadap produk yang sedang diulas, bukan komentar yang memberikan opini terhadap kualitas video ataupun pengulas yang memberikan ulasan pada video tersebut. Data komentar yang akan digunakan didapatkan dari situs berbagi web milik Google yaitu YouTube pada kategori teknologi.

Adapun *smartphone* yang akan diteliti komentarnya adalah *smartphone* yang dirilis pada tahun Januari 2016 sampai Juli 2018 yang diunggah oleh *channel* YouTube pengulas Indonesia yang sudah *channel*-nya sudah terverifikasi oleh YouTube seperti “Sobat HAPE”, dan “DroidLime”. Peneliti akan mengumpulkan sekitar 2000 data komentar dari masing-masing *channel* YouTube.

Komentar yang akan diteliti hanya komentar Bahasa Indonesia. Pemberian label pada data akan dilakukan secara manual. Adapun label yang akan digunakan adalah positif, dan negatif. Pada penelitian ini tidak akan dilakukan proses menghilangkan imbuhan baik awalan dan akhiran (*stemming*).

### 1.4. Tujuan Penelitian

Mengembangkan sebuah sistem analisis sentimen dengan menggunakan metode pembobotan CATF dan klasifikasi SVM untuk digunakan sebagai bahan pertimbangan dalam membeli sebuah *smartphone*. Diharapkan dengan penerapan metode pembobotan CATF, sistem ini dapat menghasilkan tingkat presisi dan akurasi yang tinggi.

## 1.5. Metodologi Penelitian

Metode disini merujuk pada suatu kumpulan prinsip atau teknik yang digunakan oleh peneliti untuk memecahkan masalah dalam membangun sistem.

Secara umum, penelitian ini terdiri dari 4 tahap, yaitu:

### 1. *Requirement gathering*

Tahap ini adalah tahap pengumpulan kebutuhan, dan batasan dari sistem yang akan dibangun. Yang dimaksud dengan kebutuhan adalah data komentar video ulasan *smartphone* di Youtube. Adapun batasan yang dimaksud adalah ruang lingkup penelitian yang akan dibangun sebagaimana dijelaskan pada batasan masalah.

Selain itu penelitian-penelitian yang berkaitan dengan penelitian ini digunakan sebagai referensi dalam membangun sistem. Hasil evaluasi dari penelitian sebelumnya akan dijadikan pertimbangan dalam membangun sistem ini. Data komentar didapatkan dengan menggunakan YouTube *Application Programming Interface* (API). Kemudian data komentar yang akan dipilah secara manual untuk membuang komentar yang termasuk ke dalam iklan atau *spam*.

### 2. *System development*

Ini adalah tahap pembangunan sistem yang akan dilakukan oleh peneliti. Tahap ini meliputi prapemrosesan, ekstraksi fitur, klasifikasi, pengujian sistem serta perbaikan masalah pengkodean.

### 3. *Evaluation*

Tahap evaluasi adalah tahap dimana peneliti melakukan evaluasi pada sistem yang telah dibangun. Parameter yang akan dijadikan evaluasi adalah tingkat akurasi dan nilai presisi sistem yang telah dibangun.

### 4. *Final report*

Pada laporan ini peneliti akan menuliskan hasil dari sistem yang telah dibangun, terdapat juga evaluasi dari sistem yang telah dibuat untuk rencana pengembangan sistem selanjutnya.

## 1.6. Sistematika Penulisan

Sistematika penulisan yang dipergunakan pada tugas akhir ini disesuaikan dengan ketentuan yang telah diatur pada tata acara penulisan tugas akhir Program Studi Informatika Universitas Kristen Duta Wacana.

Pada bab pertama, yakni pendahuluan berisikan latar belakang permasalahan, rumusan masalah, tujuan penelitian, metodologi penelitian, dan sistematika penulisan.

Mengenai teori-teori yang mendasari penelitian ini tertulis pada bab kedua yang berjudul landasan teori.

Sedangkan pada bab ketiga yang berjudul analisis dan perancangan sistem akan dijelaskan mengenai rancangan system yang dibuat baik desain basis data secara dimensional maupun kerangka tampilan program di dalam pola sistem masukan dan keluaran.

Hasil akhir dari analisis dan perancangan system akhirnya akan diimplementasikan dalam bentuk program. Sistem yang sudah jadi berikut penjelasan *form-form* yang digunakan di dalamnya dapat dilihat pada bab keempat yang berjudul implementasi sistem.

Bab kelima yang berjudul kesimpulan dan saran berisikan kesimpulan dari proses pemnbuatan sistem berikut hasil yang dicapai yang nantinya akan berguna sebagai saran pengembangan selanjutnya dari program ini.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berdasarkan implementasi dan analisis hasil pengujian sistem, maka penulis membuat kesimpulan sebagai berikut:

1. Pembobotan dengan menggunakan metode CATF berhasil diterapkan pada model klasifikasi SVM untuk klasifikasi sentimen menggunakan *dataset* komentar ulasan *smartphone* di YouTube dengan nilai akurasi rata-rata tertinggi dicapai pada angka 79,2% pada data tidak seimbang dan 65,8% pada data seimbang.
2. Sistem klasifikasi dengan menggunakan *kernel linear* menghasilkan performa klasifikasi yang lebih baik daripada *kernel* RBF untuk data seimbang seperti pada Tabel 4.6, sedangkan untuk data tidak seimbang *kernel* RBF menghasilkan performa klasifikasi yang lebih baik seperti pada Tabel 4.12.
3. Penggunaan nilai  $C = 1$  dan  $C = 100$  menghasilkan performa yang lebih baik pada kedua *kernel*. Dari total 20 pengujian menggunakan *kernel linear* dan RBF pada data seimbang dan tidak seimbang, 11 diantaranya mencapai nilai akurasi rata-rata tertinggi dengan nilai  $C = 1$ , dan 7 pengujian lainnya mencapai nilai akurasi rata-rata tertinggi dengan nilai  $C = 100$ , sedangkan 2 pengujian sisanya mencapai nilai akurasi rata-rata tertinggi dengan nilai  $C = 1000000$  (1 juta).
4. Penggunaan nilai  $C = 1$  dengan *kernel* RBF mampu menghasilkan sistem klasifikasi yang terbaik dengan nilai akurasi rata-rata dan presisi rata-rata yang seimbang, yaitu 77,9% dengan nilai presisi rata-rata 100%. Figur ini dicapai dengan *feature selection* 50% dan pengambilan 5 *term* secara acak pada *dataset* yang memiliki jumlah data tidak seimbang pada setiap kelas.
5. Penggunaan *feature selection* 25%, 50% dan 100% tidak menghasilkan perbedaan performa yang signifikan pada seluruh pengujian. Meskipun

begitu, pengambilan 6 *term* secara acak pada setiap komentar menghasilkan performa yang lebih rendah dibandingkan dengan pengambilan 4 atau 5 *term*.

6. Hasil pengujian untuk data seimbang menghasilkan nilai akurasi rata-rata yang lebih kecil daripada data tidak seimbang karena jumlah data pada dataset komentar lebih sedikit sehingga banyak yang *term* uji yang tidak terdapat pada *term* latih.

©UKDW

## 5.2. Saran

Berikut adalah saran-saran yang dapat digunakan untuk penelitian mengenai klasifikasi sentimen selanjutnya:

1. Melakukan prapemrosesan yang lebih kompleks seperti membuang angka dan *term* setelah proses normalisasi yang memiliki panjang kurang dari 3 huruf, menambah korpus kata tidak baku, dan *stopword*.
2. Menghitung nilai presisi setiap kelas agar butiran evaluasinya lebih halus.
3. Menerapkan metode validasi K-Fold Cross Validation.
4. Menambahkan analisis semantik sebelum proses pembobotan *term*.
5. Melakukan eksplorasi pengujian nilai *gamma* pada persamaan *Credibility Adjustment*.
6. Pengambilan *term* pada setiap komentar berjumlah di bawah rata-rata jumlah *term* pada komentar yang telah melalui tahap prapemrosesan dan diurutkan berdasarkan bobot tertinggi.

## DAFTAR PUSTAKA

- Al-Tamimi, A.-K., Satnawi, A. M., & Issa, B. E. (2018, January 31). *Arabic Sentiment Analysis For YouTube Comment*. Retrieved November 29, 2018, from Research Gate: <http://www.researchgate.net>
- Bernstein, M. N. (n.d.). *RBFKernel*. Retrieved from University of Wisconsin Computer Science User Page: <http://pages.cs.wisc.edu/>
- Blanco, E., & Moldovan, D. (2011). Some Issues on Detecting Negation from Text. *Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference* (pp. 228-233). Dallas: Human Language Technology Research Institute The University of Texas.
- Bühlmann, H., & Gisler, A. (2005). *A Course in Credibility Theory and its Applications*. Berlin: Springer.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (n.d.). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
- Cholissodin, I. (2015, Februari). *Pengenalan Pola/Pattern Recognition – Support Vector Machine*. Retrieved Oktober 17, 2018, from Imam Cholissodin: [http://imamcs.lecture.ub.ac.id/files/2015/02/Support-Vector-Machines\\_EJ\\_v5.06.pptx](http://imamcs.lecture.ub.ac.id/files/2015/02/Support-Vector-Machines_EJ_v5.06.pptx)
- Croft, W. B., Metzler, D., & Strohman, T. (2010). *Search Engines Information Retrieval in Practice*. Pearson Education.
- Donnelly, G. (2018, September 7). *75 Super-Useful Facebook Statistics for 2018 / WordStream*. Retrieved November 12, 2018, from WordStream: <https://www.wordstream.com/blog/ws/2017/11/07/facebook-statistics>
- Feizar, F. H., Indrianti, & Yudistira, N. (2017, Desember). Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(12), 1733-1741.
- Kim, Y., & Zhang, O. (2014). Credibility Adjusted Term Frequency: A Supervised Term Weighting Scheme for Sentiment Analysis and Text Classification.

*Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 5, pp. 79-83. Baltimore.

Kurniawan, J. (2017). *Klasifikasi Sentimen Terhadap Dataset Komentar Pemilu Presiden 2014 Pada Facebook Menggunakan Support Vector Machine*. Retrieved September 5, 2018, from <http://sinta.ukdw.ac.id>

Ling, J., Kencana, I. P., & Oka, T. B. (2014, August). Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, 3(3), 92-99. Retrieved September 2018, from E-Jurnal Matematika Universitas Udayana: <https://ojs.unud.ac.id/index.php/mtk>

Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge, England: Cambridge University Press.

Novantirani, A. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. 2. Bandung, Jawa Barat, Indonesia: Universitas Telkom.

Nugroho, E. (2011). Sistem Deteksi Plagiarisme Dokumen Tesks Dengan Menggunakan Algoritma Rabin-Karpi. Malang: Universitas Brawijaya.

Rinaldi, E., & Musdholifah, A. (2017). FVEC-SVM for opinion mining on Indonesian comments of youtube video. *2017 International Conference on Data and Software Engineering (ICoDSE)* (pp. 1-5). Palembang: IEEE.

Saraswati, N. W. (2011). Text Mining Dengan Metode Naiye Baiyes Classifier dan Support Vector Machine untuk Sentiment Analysis. Denpasar: Universitas Udayana.

Saskara, I. W. (2014). *Klasifikasi Tweet Menggunakan Algoritma Naive Bayes*. Retrieved September 20, 2018, from <http://sinta.ukdw.ac.id>

Shodiq, M. D. (2017). Implementasi Sentiment Analysis Pada Data Komentar Video Review Produk Handphone di Youtube Menggunakan Algoritma Support Vector Machine. Malang, Jawa Timur, Indonesia.



- Smith, C. (2018, August). *160+ YouTube Stats and Facts (August 2018) | By the Numbers*. Retrieved September 5, 2018, from DMR: <https://expandedramblings.com/index.php/youtube-statistics/>
- Yamamoto, M., & Church, K. W. (2001). Using Suffix Arrays to Compute Term Frequency and Document Frequency for All Substrings in a Corpus. *Journal Computational Linguistics*, 27(1), 1-30.
- Zainuddin, N., & Selamat, A. (2014). Sentiment Analysis Using Support Vector Machine. *International Conference of Computer, Communications, and Control Technology* (pp. 333-337). Langkawi: IEEE.

©UKDW