



**ANALISIS SENTIMEN TERHADAP TWEET KAUM HOMOSEKSUAL
INDONESIA MENGGUNAKAN METODE SUPPORT VECTOR
MACHINE**

SKRIPSI

Oleh:

Umroh Makhmudah

152410101097

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS JEMBER**

2019



**ANALISIS SENTIMEN TERHADAP TWEET KAUM HOMOSEKSUAL
INDONESIA MENGGUNAKAN METODE SUPPORT VECTOR
MACHINE**

SKRIPSI

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat
untuk menyelesaikan Program Studi Sistem Informasi (S1)

Oleh:

Umroh Makhmudah

152410101097

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS JEMBER**

2019

PERSEMBAHAN

Skripsi ini saya persembahkan untuk :

1. Allah SWT yang senantiasa memberikan rahmat dan hidayah-Nya untuk mempermudah dan melancarkan dalam mengerjakan skripsi;
2. Ibunda Lusi Endah Junari;
3. Keluarga Besar;
4. Sahabat – sahabatku dengan dukungan dan doanya;
5. Guru – guruku baik dari Pendidikan formal maupun informal;
6. Almater Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Jember

MOTTO

“But perhaps you hate a thing and it is good for you; and perhaps you love a thing and it is bad for you. And Allah knows while you know not.”

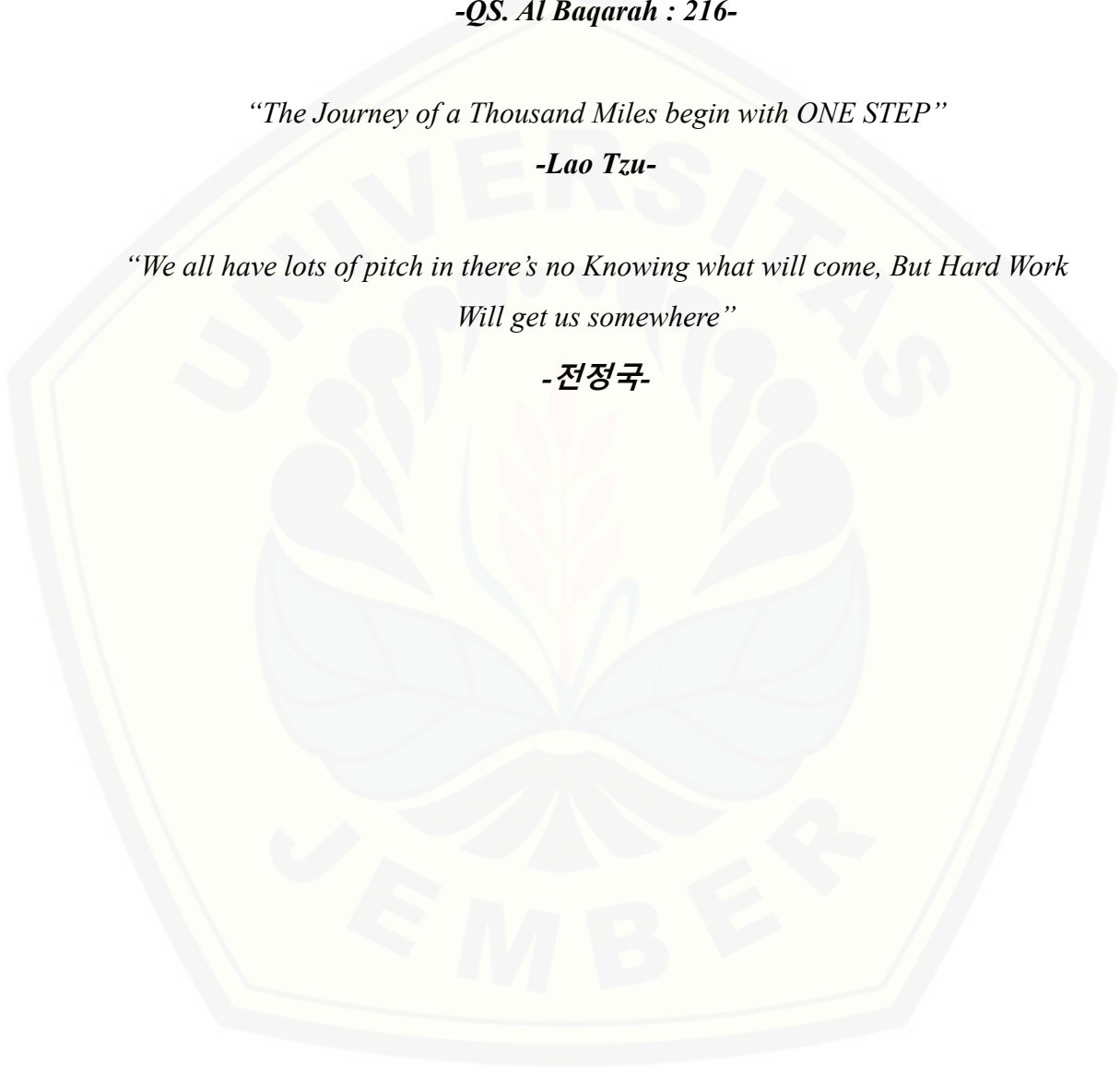
-QS. Al Baqarah : 216-

“The Journey of a Thousand Miles begin with ONE STEP”

-Lao Tzu-

“We all have lots of pitch in there’s no Knowing what will come, But Hard Work Will get us somewhere”

-전정국-



PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Umroh Makhmudah

NIM : 15241010197

Menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “ANALISIS SENTIMEN TERHADAP TWEET KAUM HOMOSEKSUAL INDONESIA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE”, adalah benar – benar hasil karya sendiri, kecuali jika dalam pengutipan substansi disebutkan sembernya, belum pernah diajukan pada institusi mana pun, dan bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa adanya tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika kemudian hari pernyataan ini tidak benar

Jember, 24 Juni 2019

Yang menyatakan,

Umroh Makhmudah

NIM 15241010197

SKRIPSI

**ANALISIS SENTIMEN TERHADAP TWEET KAUM HOMOSEKSUAL
INDONESIA MENGGUNAKAN METODE SUPPORT VECTOR
MACHINE**

Oleh

Umroh Makhmudah

152410101097

Pembimbing

Dosen Pembimbing Utama : Prof. Dr.Saiful Bukhori, ST., M.Kom.

Dosen Pembimbing Pendamping : Januar Adi Putra, S.Kom., M.Kom

PENGESAHAN PENGUJI

Skripsi berjudul "Analisis Sentimen Terhadap Tweet Kaum Homoseksual Indonesia Menggunakan Metode Support Vector Machine" karya Umroh Makhmudah telah diuji dan disahkan pada:

hari, tanggal : Sabtu, 12 Juli 2019

tempat : Fakultas Ilmu Komputer Universitas Jember.

Disetujui Oleh:

Penguji I,

Penguji II,

Prof. Drs. Slamun, M.Comp.Sc., Ph.D

M. Arief Hidayat, S.Kom., M.Kom

NIP 19670420199201101

NIP 198101232010121003

Mengesahkan

a.n Dekan

Wakil Dekan I Fakultas Ilmu Komputer

Drs. Antonius Cahya P, M.App.Sc., Ph.D.

NIP 196909281993021001

RINGKASAN

Analisis Sentimen Terhadap Tweet Kaum Homoseksual Indonesia Menggunakan Metode Support Vector Machine; Umroh Makhmudah, 152410101097; 2019, 74 halaman; Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Jember.

Sentimen analis pada Twitter merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapat informasi sentimen yang terkandung dalam tweet. Berbagai macam konten yang terdapat pada tweet memiliki berbagai macam sentiment dan pengaruh. Pendekatan *Text mining* menjadi alternatif terbaik untuk mengartikan makna dari setiap tweet. Pengklasifikasian konten Positif dan Negatif menjadi sangat penting bagi pengguna twitter untuk menilai seberapa sentimen dari sebuah tweet.

Support Vector Machine secara luas digunakan sebagai garis dasar dalam tugas-tugas yang berhubungan dengan teks tetapi kinerjanya bervariasi secara signifikan diseluruh varian, fitur dan jumlah kumpulan data. Support Vector Machine sangat baik dalam mengklasifikasikan teks dengan jumlah data yang relative banyak atau dokumen yang panjang fulllength. Beberapa scenario data diaplikasikan untuk mendapatkan model yang terbaik sesuai dengan skenario data yang telah ditetapkan.

PRAKATA

Puji syukur kehadirat Allah SWT atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisis Sentimen Terhadap Tweet Kaum Homoseksual Indonesia Menggunakan Metode Support Vector Machine”. Skripsi ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan Strata Satu (S1) pada Program Studi Sistem Informasi Universitas Jember:

Penyusunan skripsi ini tidak lepas dari dukungan berbagai belah pihak. Oleh karena itu penulis menyampaikan terima kasih kepada:

1. Allah SWT yang senantiasa memberikan rahmat dan hidayah-Nya untuk mempermudah dan melancarkan proses pengerjaan skripsi;
2. Prof. Dr.Saiful Bukhori, ST., M.Kom selaku dosen pembimbing utama dan dekan Fakultas Ilmu Komputer;
3. Januar Adi Putra S.Kom., M.Kom selaku dosen pembimbing pendamping yang telah meluangkan waktu, pikiran, dan perhatian dalam penulisan skripsi ini;
4. Seluruh Bapak dan Ibu dosen beserta staf karyawan di Fakultas Ilmu Komputer Universitas Jember;
5. Ibunda tercinta Lusi Endah Junari yang selalu mendukung, mendoakan, memberi kekuatan dan menjadi tujuan hidup terbesar;
6. Om Rudi Sulistyono dan Sugeng Prasetyo serta keluarga besar yang selalu mendukung dan mendoakan;
7. Sofyan Sugianto sebagai mentor yang telah membimbing, membantu dan mendampingi selama proses pengerjaan skripsi;
8. Kepengurusan organisasi Himpunan Mahasiswa Sistem Informasi periode 2016/2017 dan 2017/2018;
9. Kepengurusan Komuitas Developer Students Clubs Chapter Universitas Jember periode 2018/2019;

10. Kepengurusan Komunitas Young On Top Jember Batch 1 – 4;
11. Kepengurusan Asisten Laboratorium Rekayasa Perangkat Lunak;
12. Teman – teman dan karyawan di PT. DOT Indonesia periode Januari – Maret 2019;
13. Sahabat seperjuangan Tim ONE STEP, M.Noor Dwi Eldianto dan Abbi Nizar Muhammad yang telah membantu, memotivasi, menemani dan menjadi salah satu turning point dalam hidup;
14. Sahabat seperjuangan Ela Ayu Ashari dan Aisyah Nur F yang telah membantu, memotivasi, dan menjadi teman bercerita yang pengertian;
15. Tim Lomba HYDRON, M. Zufarulhaq Aleq Insan, Brian Rizqi P.D yang membantu dan menemani;
16. Tim Lomba ZETHO, Vian Elfada dan Ludfi Ika Purwantini yang memberikan pelajaran hidup yang berharga selama menjalani lomba;
17. Keluarga SELECTION atas kekeluargaan yang selalu hangat dan menenangkan;
18. Teman – teman Program Studi Sistem Informasi di semua angkatan;
19. Guru – guru Pendidikan formal maupun non formal;
20. Semua pihak yang tidak dapat disebutkan satu persatu;

Dengan harapan bahwa penelitian ini nantinya terus berlanjut dan berkembang kelak, penulis juga menerima segala kritik dan saran dari semua pihak demi kesempurnaan skripsi ini. Akhirnya penulis berharap, semoga skripsi ini dapat bermanfaat bagi semua pihak.

Jember, 22 Juli 2019

Penulis

DAFTAR ISI

PERSEMBAHAN.....	iii
MOTTO.....	iv
PERNYATAAN.....	v
HALAMAN PEMBIMBINGAN.....	vi
PENGESAHAN Penguji.....	vii
RINGKASAN.....	viii
PRAKATA.....	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL.....	xiv
BAB 1. PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan.....	3
1.4. Manfaat.....	4
1.5. Batasan Masalah.....	4
BAB 2. TINJAUAN PUSTAKA.....	6
2.1. <i>Text mining</i>	6
2.2. <i>Text Preprocessing</i>	6
2.3. <i>Feature Extraction</i>	7
2.4. <i>Learning dan Classification</i>	8
2.5. Orientasi Seksual.....	11
2.6. ANALISIS SENTIMEN.....	12
2.7. Uji Performansi.....	13
2.8. Penelitian Terdahulu.....	14
BAB 3. METODOLOGI PENELITIAN.....	19
3.1. Metode Penelitian.....	19
3.2. Metode Pengumpulan Data.....	19

3.2.1. Studi Literatur.....	19
BAB 4. DESAIN DAN IMPLEMENTASI.....	23
4.1. Analisis Kebutuhan Sistem	23
4.1.1 Kebutuhan Fungsional	23
4.1.2 Kebutuhan Non Fungsional	23
4.2. Desain Sistem	24
4.2.1 Arsitektur Proses Analisis Sentimen.....	24
4.2.2 Business Process Model Notation (BPMN)	25
4.3. Flowchart.....	21
4.3.1 Flowchart Crawling	21
4.3.2 Flowchart Preprocessing.....	23
4.4. Implementasi Sistem	25
4.4.1 Implementasi Crawling.....	25
4.4.2 Implementasi Text Preprocessing	28
4.4.3 Implementasi Feature Extraction.....	30
4.4.4 Implementasi Klasifikasi	32
4.5. Pengujian Sistem	33
BAB 5. HASIL DAN IMPLEMENTASI.....	34
5.1. Hasil Penerapan Crawling pada Twitter	34
5.2. Hasil Dataset.....	34
5.3. Hasil Implementasi Preprocessing	36
5.4. Hasil Perhitungan TF IDF	40
5.5. Hasil Testing dengan Skenario Data.....	42
5.6. Tampilan Halaman Search Tweet	45
5.7. Tampilan Halaman Hasil Crawling	46
5.8. Tampilan Halaman Hasil Analisis Sentimen	46
DAFTAR PUSTAKA	48

DAFTAR GAMBAR

Gambar 1. Contoh <i>Hyperplane</i>	9
Gambar 2. <i>Hyperplane</i> dengan margin kecil	9
Gambar 3. Proyeksi Vektor untuk margin	10
Gambar 4. Tahap - tahap Sentimen Analisis	19
Gambar 5. Arsitektur Proses Analisis Sentimen	24
Gambar 6. BPMN Sistem Analisis Sentimen	20
Gambar 7. Flowchart Proses Crawling	22
Gambar 8. Flowchart <i>Preprocessing</i>	23
Gambar 9. Flowchart Proses Klasifikasi	24
Gambar 10. Kode Program Inisialisasi Variabel Crawling	25
Gambar 11. Kode Program Fungsi Main()	26
Gambar 12. Kode Program Autentikasi	27
Gambar 13. Kode Program <i>Tweet Gathering</i>	27
Gambar 14. Kode Program <i>Preprocessing</i>	28
Gambar 15. Kode Program <i>stop word removal</i>	29
Gambar 16. Kode Program <i>Stemming</i>	30
Gambar 17. Kode Program TF IDF	31
Gambar 18. Kode Program <i>Oversampling</i>	31
Gambar 19. Kode Program Klasifikasi	32
Gambar 20. Tampilan Pencarian Tweet	45
Gambar 21. Tampilan Hasil <i>Crawling</i>	46
Gambar 22. Tampilan Hasil Analisis Sentimen	46

DAFTAR TABEL

Tabel 1. Penjelasan Tahap Sentimen Analisis	21
Tabel 2. Hasil Pengujian Sistem	33
Tabel 3. Dataset.....	35
Tabel 4. Hasil Casefolding	36
Tabel 5. Hasil <i>Stop word</i>	38
Tabel 6. Hasil <i>Stemming</i>	39
Tabel 7. Contoh Data TF IDF	40
Tabel 8. Perhitungan TF	41
Tabel 9. Perhitungan W.....	42
Tabel 10. CM 60:40	43
Tabel 11. CM 70:30.....	43
Tabel 12. CM 80:20	44
Tabel 13. Hasil Performa Keseluruhan	45

BAB 1. PENDAHULUAN

1.1. Latar Belakang

Masyarakat Indonesia sedang mengalami era pergeseran budaya. Pergeseran budaya terjadi di dalam kehidupan sosial masyarakat, yang membawa dampak positif dan negatif. Dampak negatif yang sedang dirasakan di kehidupan sosial dan budaya adalah tergerusnya nilai – nilai serta norma kesopanan, kesusilaan, serta agama di masyarakat. Berkembangnya kaum homoseksual di Indonesia merupakan cerminan dari pergeseran budaya yang terjadi dan menjadi kekhawatiran masyarakat. Menurut Laporan Perkembangan HIV-AIDS dan Penyakit Infeksi Menular Seksual (PIMS) Triwulan I Tahun 2017 dari Kementerian Kesehatan Republik Indonesia, jumlah infeksi HIV yang dilaporkan sebanyak 10.376 orang. Total penderita HIV/AIDS didominasi oleh kelompok laki – laki suka laki – laki (LSL) atau homoseksual sebesar 28 persen (Kementerian Kesehatan RI, 2017). Hal tersebut merupakan bukti bahwa kaum homoseksual memiliki kerentanan yang sangat tinggi dalam kesehatan dan kehidupan sosial.

Berkembangnya kaum homoseksual di Indonesia ditandai dengan berbagai macam peristiwa seperti pernikahan sesama jenis pasangan homoseksual yang terjadi di Kecamatan Ajung, Kabupaten Jember pada tahun 2017. Kaum homoseksual lebih mudah mengekspresikan diri kepada masyarakat luas. Kaum homoseksual juga semakin impulsif mengekspresikan dan memperkenalkan diri mereka melalui media sosial. Terdapat beberapa komunitas serta grup komunitas Homoseksual di Jember dengan anggota berjumlah lebih dari 1000 orang. Peristiwa ini juga terjadi di daerah kota Garut dimana terdapat sebuah grup homoseksual di media sosial dengan anggota pelajar SMP dan SMA yang berjumlah lebih dari 2500 anggota (Ghani, 2018).

Kebebasan berekspresi kaum homoseksual di media sosial menimbulkan berbagai macam dampak bagi masyarakat luas. Melalui media sosial, kaum homoseksual dapat dengan mudah berinteraksi secara bebas dengan sesama homoseksual. Penggunaan kata – kata yang tidak senonoh dan mengandung unsur

pornografi juga sangat sering digunakan oleh kaum homoseksual di media sosial. Konten – konten tersebut dapat menjadi konsumsi masyarakat luas dan dapat memberikan berbagai dampak terutama bagi anak di bawah umur. Lebih lanjut dijelaskan bahwa homoseksual terjadi bukan karena kelainan genetik, ketidakseimbangan hormon, sakit mental ataupun merupakan hasil dari kejahatan. Homoseksual terjadi disebabkan karena pengaruh lingkungan dan pengalaman yang terjadi pada seorang individu (Kaplan, 1997). Sehingga, seseorang yang sering berhubungan dengan berbagai macam jenis interaksi dengan homoseksual dapat memiliki kecenderungan untuk menjadi homoseksual.

Pendampingan terhadap kaum homoseksual sangat diperlukan agar pemerintah serta Lembaga – Lembaga terkait dapat memberikan bantuan bagi kaum homoseksual untuk dapat menghadapi kondisi yang sedang dialami. Pengawasan terhadap berbagai konten yang tersebar di media sosial juga sangat diperlukan bagi orang tua untuk dapat menjaga putra putrinya agar dapat terhindar dari konten – konten tentang homoseksual yang berdampak buruk.

Analisis sentimen adalah Natural Language Processing berupa pemrosesan terhadap sentimen dari kata maupun kalimat yang digunakan untuk mengetahui polaritas dari kata atau kalimat tersebut. Metode tersebut membantu untuk mempelajari sentimen dari berbagai macam konten – konten media sosial dalam hal ini berupa tweet yang dibuat oleh kaum homoseksual. Penelitian ini berfokus pada twitter sebagai objek analisis sentimen karena pertumbuhan pengguna aktif twitter di Indonesia memiliki grafik yang baik dari total 12 juta pengguna pada 2014 menjadi 22,8 juta pada 2019 (Statista, 2019). Menurut Kepala Biro Humas Kementrian Komunikasi dan Informasi Fernandus Setu, twitter menjadi media sosial dengan penyebaran konten pornografi tertinggi (JPNN, 2019). Hal ini menjadikan twitter sebagai objek yang sesuai untuk proses analisis sentimen *tweet* kaum homoseksual yang salah satu fokusnya adalah sentimen negatif yang bermuatan pornografi. Pada penelitian ini digunakan metode *Support Vector Machine Classifier* untuk mengklasifikasikan sentimen berdasarkan penelitian yang dilakukan oleh Lutfi Budi Ilmawan dan Edi Winarko pada tahun 2015 yang

menyatakan bahwa metode *Support Vector Machine* mengungguli performa dari *naïve bayes* dalam penggunaannya untuk analisis sentimen (Budi Ilmawan and Winarko, 2015). Hasil Klasifikasi sentimen akan dikelompokkan ke dalam dua jenis sentimen yaitu sentimen positif dan sentimen negatif. Sentimen negatif diperuntukkan untuk *tweet* yang memiliki konten pornografi dan konten impulsif dari kaum homoseksual. Sedangkan sentimen positif berupa konten – konten yang tidak mengandung unsur pornografi serta memiliki sentimen yang kontra dengan homoseksual. Uji Performansi juga dilaksanakan untuk menguji model klasifikasi yang telah dihasilkan. Penelitian ini diharapkan dapat memberikan kontribusi bagi pemerintah untuk mengetahui tentang konten dan akun yang bersifat pornografi dan negatif, memberikan informasi bagi masyarakat agar tidak terpapar dengan konten – konten negatif tentang homoseksual yang dapat ditemukan di media sosial *twitter*.

1.2. Rumusan Masalah.

Dari uraian latar belakang diatas dapat diidentifikasi masalah-masalah yang harus ditentukan untuk membantu menentukan fokus penelitian. Rumusan Masalah pada penelitian sebagai berikut:

1. Bagaimana menganalisis sentimen *tweet* kaum homoseksual Indonesia dengan menggunakan metode klasifikasi *support vector machine*?
2. Bagaimana mengukur akurasi dari metode klasifikasi *support vector machine* dalam analisis sentimen *tweet* kaum homoseksual Indonesia?

1.3. Tujuan

Tujuan penelitian ini ditetapkan untuk menjawab rumusan masalah yang telah ditetapkan, sehingga penelitian dapat dilaksanakan berdasarkan tujuan yang telah ditetapkan. Tujuan yang ingin di capai dalam penelitian ini adalah:

1. Menerapkan metode klasifikasi *support vector machine* dalam mengklasifikasikan *tweet* kaum homoseksual Indonesia

2. Menguji Performansi klasifikasi dalam penerapan metode *support vector machine*

1.4. Manfaat

Manfaat yang dipaparkan merupakan dampak yang ingin dihasilkan dari penelitian ini. Manfaat yang didapatkan dari penelitian ini adalah:

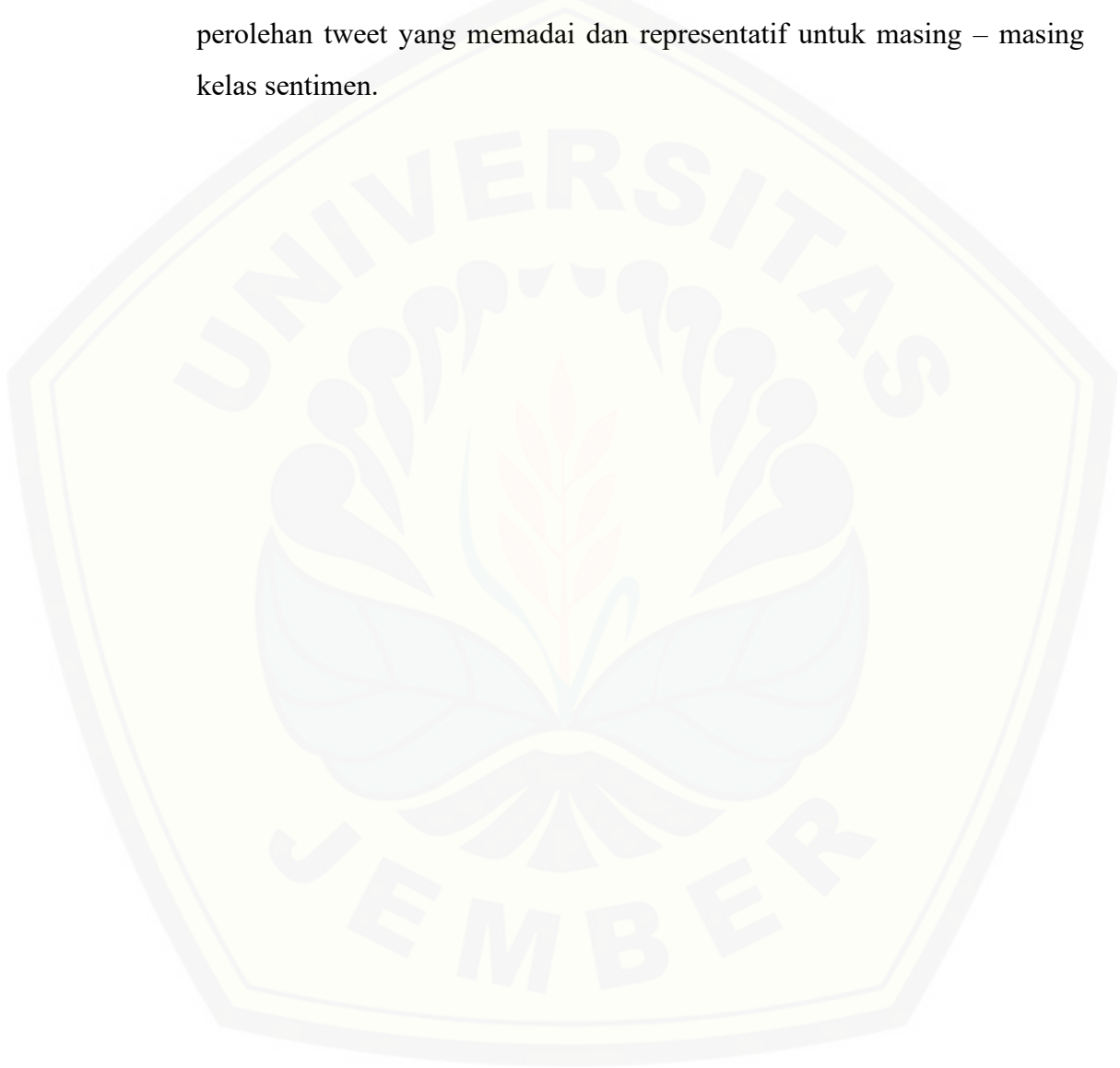
1. Bagi Pemerintah dan Lembaga terkait
Pemerintah dan Lembaga terkait dapat menjadikan penelitian ini sebagai sarana untuk mengetahui tentang konten dan akun yang bersifat negatif dan positif yang berhubungan dengan konten tweet homoseksual.
2. Bagi Penulis
Penelitian ini dapat memberikan wawasan serta pengembangan keilmuan bagi penulis dalam bidang *Natural Language Processing* dan Metode Klasifikasi.

1.5. Batasan Masalah

Terdapat beberapa batasan masalah yang diangkat sebagai parameter dalam melakukan penelitian ini, antara lain:

1. Data diambil menggunakan Application Programming Interface (API) Twitter dengan API KEY penulis
2. Penelitian ini berfokus pada analisis aktivitas media sosial Twitter yaitu tweet dari pengguna twitter berbahasa Indonesia tanpa menganalisa gambar, video, dan pranala
3. Algoritma yang digunakan adalah support vector machine tanpa membandingkan dengan metode lain
4. Tweet yang terkumpul akan dikelompokkan ke dalam dua analisis sentimen yaitu positif dan negatif
5. Sistem ini akan berhenti menjalankan proses crawling twitter jika API Twitter sedang down
6. Menggunakan konsep Natural Language Processing dalam Preprocessing data sebelum proses klasifikasi sentimen

7. Jumlah data yang akan digunakan dalam proses analisis adalah sebanyak 549 data tweet
8. Hashtag yang dipilih adalah #gaybrondong, #gaybocah, #pijatgay, #gaylokal, #gayremaja, #gayindo, #tolakLGBT, #awasLGBT, #lawanLGBT. Hashtag – hashtag tersebut dipilih berdasarkan jumlah perolehan tweet yang memadai dan representatif untuk masing – masing kelas sentimen.



BAB 2. TINJAUAN PUSTAKA

2.1. *Text mining*

Text mining merupakan penerapan konsep dan teknik data mining untuk mencari pola dalam teks. Proses penganalisisan teks guna menyarikan informasi yang bermanfaat untuk tujuan tertentu. Proses data mining untuk data dokumen atau teks memerlukan lebih banyak tahapan, mengingat data teks memiliki karakteristik yang lebih kompleks daripada data biasa. Berdasarkan ketidakteraturan struktur data teks, maka proses *Text mining* memerlukan beberapa tahap awal yang pada intinya adalah mempersiapkan agar teks dapat diubah menjadi lebih terstruktur. Bentuk perubahan yang dilakukan adalah ke dalam *spreadsheet*, kolom menunjuk dokumen dan baris menunjuk kata, sedangkan selnya menunjuk frekuensi kata dalam dokumen (Chrystal and Joseph, 2015). Tahap – tahap *Text mining* adalah sebagai berikut :

2.2. *Text Preprocessing*

Tahapan awal ini adalah mempersiapkan data agar dapat digunakan pada proses selanjutnya. Proses ini diawali dengan proses *tokenization* yaitu membagi kalimat menjadi kata – kata individual dengan memberikan space atau pembatas untuk setiap katanya dengan tetap mempertahankan arti dari kata tersebut. Proses selanjutnya adalah proses *stop word removal* atau proses menghilangkan kata – kata yang tidak memiliki pengaruh terhadap makna dan sentimen dari sebuah kalimat. Proses ini mempermudah untuk mendeteksi kata kunci di setiap kalimat dan frekuensi distribusi data. Proses yang terakhir adalah Lemmatization adalah proses mengubah kata – kata individual yang telah terbentuk menjadi bentuk kata dasar. Pembentukan kata dasar akan membantu untuk menentukan kata – kata yang paling digunakan dari sebuah korpus atau kalimat (Chrystal and Joseph, 2015). Proses untuk melakukan *lemmatization* atau *stemming* adalah dengan menggunakan Algoritma Nazief Andriani, algoritma tersebut melakukan proses *stemming* dengan tahapan berikut :

1. Algoritma memeriksa kata yang dianalisis adalah merupakan kata dasar yang sudah terdapat pada daftar kata dasar dari algoritma. Ketika kata tersebut merupakan kata dasar maka proses akan dihentikan pada tahap pertama
2. Algoritma menghilangkan *inflection suffix* atau imbuhan, Proses ini akan dilakukan kembali jika terdapat *particles* seperti (-lah, -kah, -tah, atau -pun) atau akhiran untuk menghilangkan *possessive pronouns* atau keterangan kepemilikan seperti (-ku, -mu, atau -nya).
3. Proses selanjutnya adalah menghilangkan imbuhan turunan seperti (-i, -kan, -an).
4. Dan proses *stemming* selanjutnya adalah proses untuk menghilangkan awalan turunan seperti (be-, di-, ke-, me- dan te-).
5. Proses tersebut akan dilanjutkan ke proses analisis untuk memasukkan kata – kata yang tidak berhasil diproses ke table ambiguitas.
6. Bila semua proses yang dilakukan gagal dan tidak berhasil menemukan kata dasar dari suatu kata yang telah dilakukan proses *stemming*, maka kata akan dikembalikan ke bentuk asli sebelum diproses.

2.3. Feature Extraction

Tahap ini menggunakan metode *Term Frequency – Inverse Document Frequency (TF-IDF)*. Metode ini adalah metode dalam proses feature extraction untuk dokumen tunggal tiap kalimat dianggap sebagai dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu *Term Frequency (TF)* merupakan frekuensi kemunculan kata (t) pada kalimat (d). *Document Frequency (DF)* adalah banyaknya kalimat dimana suatu kata (t) muncul. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Bobot kata

semakin besar jika sering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam banyak dokumen (Dadgar, Araghi and Farahani, 2016). Nilai IDF sebuah term dihitung menggunakan persamaan 1 :

$$W_{d,t} = TF(d,t) * IDF(t) \dots \dots \dots (1)$$

Keterangan :

W : bobot kalimat ke-d terhadap kata (term) ke-t

d : kalimat ke-d

t : kata (term) ke-t

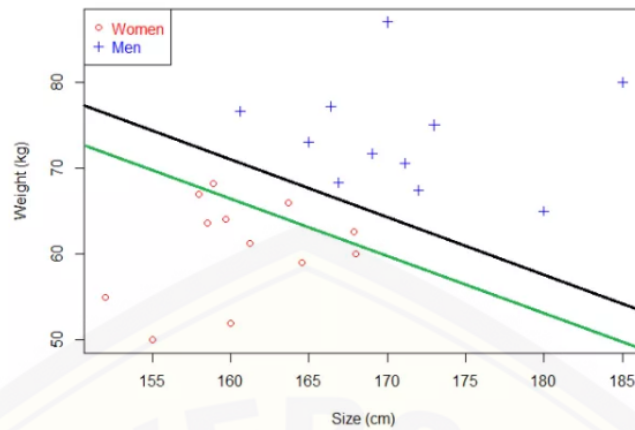
TF : *term frequency*

IDF : *inverse document frequency*

2.4. *Learning dan Classification*

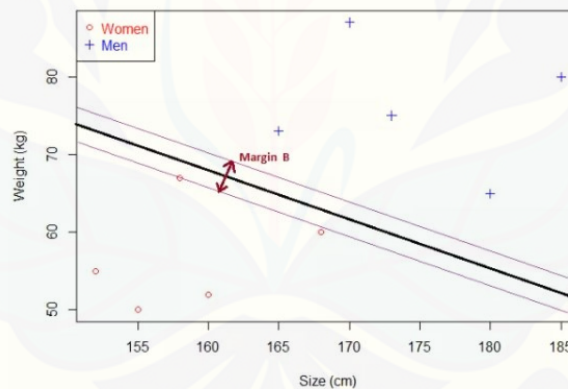
Tahap ini adalah tahap yang paling terpenting dari keseluruhan tahapan *Text mining*. Tahap ini adalah tahap untuk menemukan pola dari keseluruhan teks. Pada penelitian ini digunakan Teknik *Supervised Learning* pada metode yang diterapkan yaitu Support Vector Machine. Support Vector Machine adalah algoritma klasifikasi yang dapat digunakan untuk memprediksi suatu data baru berada pada kelas tertentu. Teknik *Supervised Learning* adalah Teknik pembelajaran dimana terdapat label atau nama kelas pada data latih dan data baru diklasifikasikan berdasarkan data latih. (Chrystal and Joseph, 2015)

Support *Vector Machine* memiliki tujuan untuk mendapatkan *hyperplane* paling optimal yang dapat memaksimalkan jarak margin dari data training. *Hyperplane* yang paling optimal adalah *hyperplane* yang memiliki jarak sejauh mungkin dengan setiap poin data yang terdapat pada setiap kelas. Jarak yang jauh dari setiap poin data diperlukan karena jika *hyperplane* berada terlalu dekat dengan data maka margin atau jarak dari *hyperplane* terhadap titik data menjadi sangat kecil. Jarak yang jauh memberikan keuntungan agar *hyperplane* dapat mengklasifikasikan data dengan baik dan jika terdapat poin data yang masuk ke dalam area margin masih dapat diklasifikasikan dengan baik karena masih berada dalam daerah yang benar.



Gambar 1. Contoh *Hyperplane*

Pada gambar 1 menunjukkan *hyperplane* (garis hijau) yang tidak membagi secara baik karena berada lebih dominan pada bagian poin data tertentu sehingga dapat mengklasifikasikan dengan salah beberapa data.



Gambar 2. *Hyperplane* dengan margin kecil

Gambar diatas menggambarkan *hyperplane* yang kecil dan juga memiliki kecenderungan untuk condong ke suatu kelas. Hal ini juga dapat menyebabkan klasifikasi yang dihasilkan oleh *hyperplane* yang telah terbentuk juga tidak optimal karena poin – poin data baru yang akan diklasifikasikan bisa masuk ke dalam kelas yang tidak sesuai.

Support Vector Machine adalah metode yan bekerja dengan menemukan hyperplane terbaik untuk memisahkan dua buah kelas. Perhitungan Support Vector Machine menggunakan persamaan 2 :

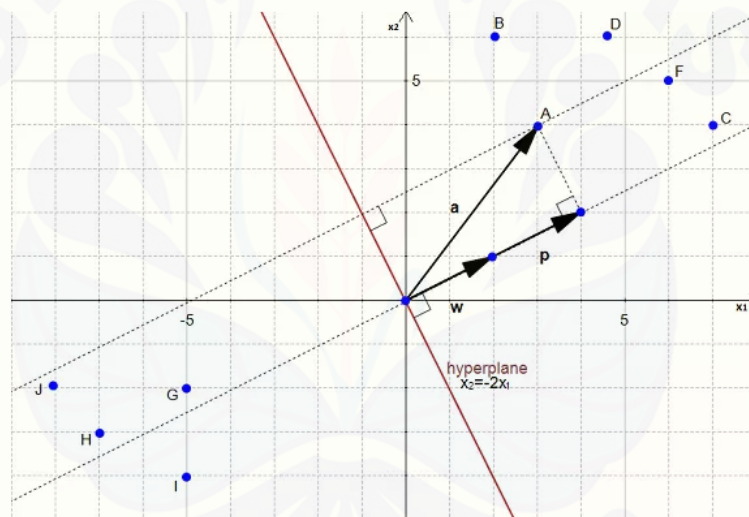
$$f(x) = \sum_{i=1}^m \alpha_i y_i K(X, X_i) + b \dots \dots \dots (2)$$

Keterangan :

- n : dimensi data atau banyaknya fitur
- α : nilai bobot setiap titik data
- m : jumlah *support vector* atau titik data yang memiliki $\alpha_i > 0$
- $K(X, X_i)$: fungsi kernel

Data yang akan diklasifikasikan dapat dibagi menggunakan persamaan diatas, dengan nilai yang tidak terbatas untuk K dan b. Persamaan diatas menunjukkan model atau garis yang akan membagi kelas data.

Support Vector Machine adalah proses perhitungan yang sangat berdekatan dengan konsep vector seperti pada gambar



Gambar 3. Proyeksi Vektor untuk margin

Pada gambar 3 ditunjukkan jika kita telah memiliki sebuah persamaan *hyperplane*. Persamaan *hyperplane* tersebut ekuivalen dengan vector $w \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Sedangkan vector w bukan merupakan poin data. Untuk mengetahui margin dari *hyperplane*, perlu dilakukan proses perhitungan jarak antara titik A(3,4) dengan *hyperplane*. Titik A jika diproyeksikan ke titik normal vector w maka akan terbentuk gambar seperti diatas. Melalui perhitungan proyeksi vector maka dapat ditemukan jika Panjang $2\|p\|$ adalah jarak margin dari *hyperplane*.

Proses yang dilakukan saat melakukan perhitungan SVM adalah menentukan dataset lalu memilih dua *hyperplane* dan yang terakhir adalah melakukan optimalisasi terhadap margin dari *hyperplane* tersebut. Proses perhitungan untuk menentukan dua *hyperplane* yang dapat memenuhi ketentuan yaitu tidak ada poin data yang berada diantara kedua *hyperplane* tersebut adalah sebagai berikut:

$$y_i(w \cdot X_i + b) \geq 1 \text{ untuk } 1 \leq i \leq n \dots\dots\dots(3)$$

Persamaan 3 adalah persamaan yang dapat digunakan untuk menentukan dua *hyperplane* awal yang akan digunakan sebelum melakukan proses optimalisasi margin *hyperplane*.

Optimalisasi *hyperplane* dapat dilakukan dengan meminimalisasi nilai $\|w\|$ dalam perhitungan margin pada persamaan berikut:

$$m = \frac{2}{\|w\|} \dots\dots\dots(4)$$

Sehingga proses optimalisasi dapat dilaksanakan dengan meminimilisasasi nilai $\|w\|$ pada persamaan 4 dengan bergantung pada persamaan 3 untuk setiap nilai i .

2.5. Orientasi Seksual

Orientasi seksual atau kecenderungan seksual adalah pola ketertarikan seksual, romantis, atau emosional (atau kombinasi dari keseluruhan) kepada orang-orang dari lawan jenis atau gender, jenis kelamin yang sama atau gender, atau untuk kedua jenis kelamin atau lebih dari satu gender. Ini umumnya digolongkan dalam heteroseksual, homoseksual, dan biseksual, sementara aseksual (kurangnya ketertarikan seksual kepada orang lain) kadang-kadang diidentifikasi sebagai kategori keempat.

Kategori-kategori ini adalah aspek sifat lebih bernuansa identitas seksual dan terminologi. Misalnya, orang dapat menggunakan label lain, seperti pansexual atau polysexual, atau tidak sama sekali. Menurut American Psychological Association, orientasi seksual juga mengacu pada pengertian identitas seseorang berdasarkan pada atraksi, perilaku terkait, dan keanggotaan dalam komunitas lain yang berbagi atraksi mereka. Androphilia dan gynephilia adalah istilah yang

digunakan dalam ilmu perilaku untuk menggambarkan orientasi seksual sebagai alternatif gender konseptualisasi biner. Androphilia menggambarkan daya tarik seksual untuk maskulinitas; gynephilia menggambarkan daya tarik seksual untuk feminitas. Istilah preferensi seksual sebagian besar tumpang tindih dengan orientasi seksual, tetapi pada umumnya dibedakan dalam penelitian psikologis. Seseorang yang mengidentifikasi sebagai biseksual, misalnya, dapat secara seksual lebih memilih satu jenis kelamin daripada yang lain. Preferensi seksual mungkin juga menyarankan gelar pilihan secara sukarela, sedangkan konsensus ilmiah adalah bahwa orientasi seksual adalah bukan pilihan.

Tidak ada konsensus di antara para ilmuwan tentang mengapa seseorang mengembangkan orientasi seksual tertentu. Banyak ilmuwan berpikir bahwa nature dan nurture - kombinasi genetik, hormonal, dan pengaruh lingkungan - faktor dalam penyebab orientasi seksual. Mereka mendukung teori berbasis biologis, yang menunjuk ke faktor genetik, lingkungan rahim awal, keduanya, atau dimasukkannya faktor genetik dan keadaan sosial. Tidak ada bukti substantif yang menunjukkan pengasuhan anak usia dini atau pengalaman berperan dalam orientasi seksual. Berkenaan dengan perilaku seksual sesama jenis, berbagi atau lingkungan keluarga memainkan peran untuk pria dan peran kecil bagi perempuan. Penelitian selama beberapa dekade telah menunjukkan bahwa orientasi seksual berkisar sepanjang kontinum, dari atraksi eksklusif untuk lawan jenis ke atraktif eksklusif untuk jenis kelamin yang sama.

Orientasi seksual dilaporkan terutama dalam ilmu biologi dan psikologi (termasuk seksologi), tetapi juga suatu subyek dalam antropologi, sejarah (termasuk konstruksionisme sosial), dan hukum, dan ada penjelasan lain yang berhubungan dengan orientasi seksual dan budaya.

2.6. ANALISIS SENTIMEN

Analisis Sentimen adalah proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Proses utama dalam analisis sentimen yaitu

mengelompokkan teks yang terdapat dalam sebuah kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan tersebut apakah bersifat positif, negatif, atau netral. Analisis sentimen dapat digunakan untuk mencari pendapat tentang produk, merk atau tokoh publik dan menentukan apakah mereka dilihat positif atau negatif (Saraswati, 2011).

Berdasarkan sumber datanya, analisis sentimen dibedakan menjadi dua kategori, yaitu :

1. Coarse Grained Sentimen Analysis

Analisis yang dilakukan pada keseluruhan dokumen sehingga analisis yang dilakukan menganggap bahwa keseluruhan isi dokumen sebagai sentimen negative dan positif.

2. Fined Grained Sentimen Analysis

Analisis yang dilakukan pada kategori ini adalah analisis pada tingkat kalimat. Pada kategori ini dikatakan bahwa setiap kalimat di dalam suatu dokumen memiliki sentimen yang berbeda.

2.7. Uji Performansi

Pengujian performa model dilakukan dengan pengujian terhadap nilai yang dihasilkan dari kelas hasil prediksi dan nilai asli dari suatu kelas. Uji performansi mencocokkan antara kedua nilai tersebut. Beberapa jenis nilai yang akan digunakan untuk mengukur performa dari model yang dibuat adalah nilai akurasi, precision, recall dan nilai F.

Akurasi adalah bentuk pengukuran yang baik untuk kelas dengan distribusi yang seragam. Akurasi juga dapat digunakan untuk kelas yang memiliki ketidakseimbangan distribusi dan jumlah kelas. Akurasi memberikan rasio prediksi benar untuk kelas positif dan negative terhadap keseluruhan data. Perhitungan akurasi menggunakan persamaan 3 berikut:

$$akurasi = \frac{(True\ Positive\ (TP) + True\ Negative\ (TN))}{(TP + False\ Positive + False\ Negative + TN)} \dots\dots\dots(5)$$

Precision merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil prediksi positif. Precision memberikan informasi tentang seberapa akurat model yang telah dibuat dari keseluruhan kelas positif yang telah diprediksi. Perhitungan precision menggunakan persamaan 4 sebagai berikut :

$$precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \dots\dots\dots(6)$$

Recall menjelaskan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Perhitungan recall menggunakan persamaan 5 sebagai berikut :

$$recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \dots\dots\dots(7)$$

F1 Score adalah nilai perbandingan rata = rata presisi dan recall yang dibobotkan. F1 dibutuhkan jika ingin mendapatkan nilai yang seimbang antara precision dan recall. Kelas dengan distribusi kelas yang tidak seimbang seperti nilai kelas asli untuk kelas negative lebih dominan seperti yang terjadi pada penelitian ini. Nilai harmonisasi baik jika sama dengan atau mendekati satu dan nilai harmonisasi buruk jika sama dengan dan mendekati nol. Perhitungan F1 Score menggunakan persamaan 6 sebagai berikut:

$$F1\ Score = \frac{2(Recall \times Precision)}{Recall + Precision} \dots\dots\dots(8)$$

2.8. Penelitian Terdahulu

Penelitian terdahulu digunakan sebagai bahan pertimbangan dalam penelitian yang sedang dilakukan. Berikut penelitian yang di gunakan:

1. Penelitian terdahulu yang dilakukan dengan menerapkan metode *Support Vector Machine* adalah penelitian "*Predicting Sexual Orientation based on Facebook Status*" oleh Aaron Loh di Stanford University pada tahun 2016 (Loh *et al.*, 2016). Penelitian ini membahas tentang penentuan orientasi social seseorang melalui status *facebook* yang dimiliki. Penelitian ini meneliti 560 ribu data set dan membagi keseluruhan data set tersebut

manjadi dua yaitu status yang dimiliki oleh laki – laki dan perempuan. Penelitian ini menggunakan beberapa *learning algorithm* diantaranya adalah *Multinomial Naïve Bayes* yang menghasilkan probabilitas tentang kecenderungan seseorang adalah heteroseksual atau homoseksual. *learning algorithm* yang kedua adalah *Logistic Regression*, algoritma ini membuat asumsi probabilitas yang tidak sekuat *Multinomial Naïve Bayes* sehingga lebih fleksibel, selanjutnya adalah *Support Vector Machine*, Algoritma ini berjalan dengan baik jika kelas dipisahkan dengan baik dan algoritma ini berjalan lebih lambat dari algoritma *Multinomial Naïve Bayes* dan *Logistic Regression*. Hasil dari penelitian ini adalah SVM memiliki hasil yang terbaik dalam skor presisi. Model yang digunakan dalam penelitian untuk memprediksi heteroseksual, dan tidak bekerja dengan baik untuk memprediksi homoseksual. Model ini juga bekerja dengan baik pada perempuan. Seseorang pria yang homoseksual cenderung memiliki status yang cenderung seperti perempuan sehingga model salah mengklasifikasikan laki – laki yang homoseksual sebagai seorang perempuan. Model ini juga tidak bisa membaca konteks penggunaan dari suatu kata, sehingga interpretasi yang dihasilkan tidak sesuai dengan konteks asli dari sebuah status karena status tersebut menggunakan kata yang dapat membuat sistem mengindikasikan bahwa orang tersebut adalah seorang homoseksual.

2. Penelitian kedua adalah “*Machine “Gaydar” : Using Facebook Profiles to Predict Sexual Orientation*” oleh Nikhol Bhattasali dan Esha Maiti dari Stanford University pada tahun 2015 (Bhattasali and Maiti, 2015). Penelitian ini menggunakan informasi yang dibagi di *facebook* seperti informasi umum seseorang, foto, dan teks – teks yang dipublikasikan di *facebook* untuk memprediksikan orientasi seksual dari laki – laki menggunakan algoritma *machine learning*. Model yang dibangun menggunakan *Naïve Bayes Classifier* dan *Support Vector Machine*. Penentuan datasetnya secara binary dengan melambangkan heteroseksual sebagai 0 dan non heteroseksual sebagai 1 untuk menyederhakan

klasifikasi. Data set diambil dari mahasiswa S1 Stanford University, dengan 62% data set adalah *LGBTQ* dan sisanya adalah *non LGBTQ* sejumlah 167 profil facebook. Model ini menganalisa informasi pribadi pada bagian tentang, teman, foto profil, dan timeline. Penggunaan metode *Naïve Bayes Classifier* pada model adalah untuk menentukan orientasi seksual dari timeline, metode ini juga memudahkan kita untuk mengetahui kata – kata yang biasa digunakan oleh LGBT. Metode ini juga membantu dalam menganalisa atribut di dalam bagian tentang dan foto. Hasil dari Analisa tersebut menghasilkan skor yang dikombinasikan dengan *Support Vector Machine* dan menghasilkan klasifikasi. Kedepannya penelitian ini akan menggunakan metode *Neural Network* yang digunakan untuk mendeskripsikan kompleksitas dari identitas.

3. Penelitian ketiga adalah “*Social Networks’ Text mining for Sentimen Classification: The case of Facebook statuses updates in the “Arabic Spring” Era*” adalah penelitian yang dilakukan oleh Safa Ben Hamouda dan Jalel Akaichi pada *International Journal of Application or Innovation in Engineering & Management (IJAIEEM)* di tahun 2013 (Hamouda and Akaichi, 2013). Jurnal ini membahas tentang penggunaan *Text mining* dalam *sentimen classification* atau untuk mengklasifikasikan sentimen atau keadaan seseorang dari postingan facebook yang dibuat. Ilustrasi yang digunakan adalah postingan *facebook* dari warga Tunisia pada era *Arabic Spring*. Tujuan dari penelitian ini adalah untuk mengekstrak informasi penting dari sentimen dan perilaku pengguna pada waktu yang sensitif. Penelitian ini menggunakan metode *Naïve Bayes Classifier* dan *Support Vector Machine*. Penelitian juga melakukan beberapa perbandingan tentang kosakata seperti emotikon, akronim dan lain – lain dari status yang telah di ekstrak. Data yang digunakan adalah 260 status dari warga Tunisia. Hasil yang didapatkan sangat memuaskan karena postingan yang dibuat oleh objek adalah postingan – postingan yang unik karena waktu pengambilan data merupakan waktu yang sensitive dan juga penting. Penggunaan metode *SVM* dan *Naïve Bayes* mempermudah proses

karena kedua metode tersebut adalah metode machine learning yang sangat umum digunakan untuk pemrosesan *Text mining*.

4. Penelitian keempat adalah “*A Review: Text Classification on Social Media*” Data oleh Priyanka Patel dan Khushali Mistry, penelitian ini dipublikasikan pada *IOSR Journal of Computer Engineering (IOSR-JCE)* pada tahun 2015 (Patel and Mistry, 2015). Penelitian ini bertujuan untuk mengklasifikasikan tipe – tipe dari pengguna media sosial berdasarkan postingan yang dibuat. *Machine Learning* digunakan untuk melakukan klasifikasi terhadap teks. Jurnal ini menunjukkan kelebihan dan kekurangan dari metode – metode yang biasa digunakan. Kelebihan dan kekurangan dari masing – masing metode adalah sebagai berikut: *Bayessian Classifier* memiliki kelebihan mudah diimplementasikan, mudah dalam proses perhitungan dan bekerja dengan baik pada data numerik dan tekstual dan memiliki kekurangan bekerja dengan tidak baik. *Decision Tree* memiliki kelebihan yaitu mudah dimengerti, mudah dalam membuat aturan, dan dan mengurangi kompleksitas dan memiliki kelemahan yaitu ketika kesalahan dilakukan di level yang lebih tinggi maka level selanjutnya akan salah juga. *KNN* memiliki kelebihan dalam efektifitas dan memiliki kekurangan dalam waktu pengklafikisian yang Panjang. *SVM* memiliki kelebihan dalam menangkap karakteristik dari data dan memiliki kekurangan dalam hal penyesuaian parameter. *Neural Network* memiliki kelebihan dalam menghasilkan hasil yang lebih baik dan proses testing yang cepat dan kekurangannya adalah proses training data yang lamban. *Rocchio’s* memiliki kelebihan yaitu mudah dalam proses implementasi dan memiliki kemampuan yang cepat dalam belajar dan memiliki kekurangan dalam akurasi.

Perbedaan yang ada antara penelitian terdahulu seperti telah dijelaskan sebelumnya dengan penelitian ini adalah proses untuk melakukan Natural Language Processing diterapkan pada media sosial twitter dengan menggunakan metode Support Vector Machine. Pengaplikasian klasifikasi

menggunakan beberapa skenario data untuk mengetahui pengaruh berbagai skenario data terhadap akurasi dari model yang dihasilkan.



BAB 3. METODOLOGI PENELITIAN

3.1. Metode Penelitian

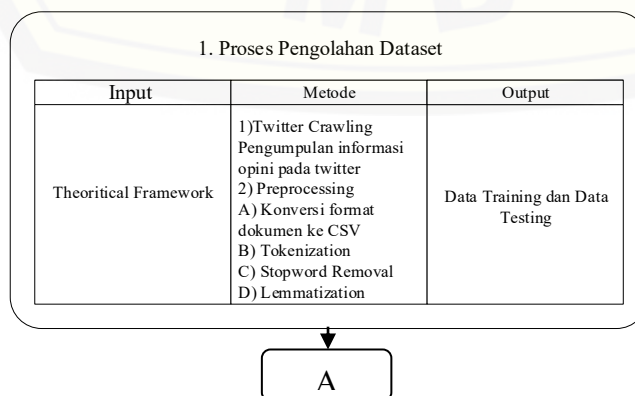
Metode penelitian yang dilakukan merupakan penelitian pengembangan. Penelitian pengembangan bertujuan untuk mengembangkan produk tertentu dan menguji keefektifan produk tersebut. Penelitian ini mengembangkan sentimen analisis yang untuk mengklasifikasikan tweet yang dibuat oleh kaum homoseksual bersentimen positif atau negatif. Metode yang digunakan untuk mengklasifikasikan sentimen tweet kaum homoseksual adalah support vector machine.

3.2. Metode Pengumpulan Data

Data diperoleh untuk melakukan penelitian ini adalah data yang didapat langsung dari status Twitter menggunakan Twitter API. Pengumpulan data dilakukan terhadap beberapa hashtag yaitu #gaybrondong, #gaybocah, #pijatgay,,#gaylokal,, #gayremaja, #gayindo, #tolakLGBT, #awasLGBT, #lawanLGBT.. Jumlah keseluruhan data yang akan diambil adalah 600 data dengan jumlah untuk masing – masing hashtag adalah 100 data.

3.2.1. Studi Literatur

Berikut merupakan tahapan implementasi dari Twitter Sentimen Analysis terhadap Tweet Kaum Homoseksual Indonesia Menggunakan Metode Support Vector Machine Classifier:



A

2. Feature Extraction

Input	Metode	Output
Data Training dan Data Testing	<i>Term Frequency – Inverse Document Frequency (TF-IDF)</i>	Data Training dan Data Testing

3. Proses pengolahan data *training*

Input	Metode	Output
Data Training	Pembuatan model dengan algoritma klasifikasi : <i>support vector machine</i>	Model Hasil Klasifikasi

4. Proses Kategorisasi data *testing*

Input	Metode	Output
Model Hasil Klasifikasi, Data Testing	1. <i>Data testing</i> diterapkan ke model klasifikasi 2. Klasifikasi <i>data testing</i>	Hasil Kategorisasi sentimen

5. Uji Performansi

Input	Metode	Output
Hasil Kategori Sentimen , Model Hasil Klasifikasi	Confusion Matrix	Hasil Uji Performansi Model Klasifikasi

Gambar 4. Tahap - tahap Sentimen Analisis

Tabel 1. Penjelasan Tahap Sentimen Analisis

NO	Tahap	Input	Proses	Output
1.	<i>Crawling</i> Data	Query, <i>Twitter</i> API	<i>Crawling</i> Data dengan menggunakan <i>Twitter</i> API	<i>Dataset</i>
2.	<i>Text</i> <i>Processing</i>	<i>Dataset</i>	<ol style="list-style-type: none"> 1. Konversi format dokumen ke CSV 2. <i>Tokenization</i> 3. <i>Stopword Removal</i> 4. <i>Lemmatization</i> Algoritma yang digunakan untuk proses <i>stemming</i> berbahasa Indonesia adalah Nazief dan Andriani 	Data <i>Training</i> dan Data <i>Testing</i>
3.	<i>Feature</i> <i>Extraction</i>	Data <i>Training</i> dan Data <i>Testing</i>	Pemrosesan dokumen melalui proses <i>Feature Extraction</i> menggunakan metode <i>Term Frequency – Inverse Document Frequency (TF-IDF)</i>	Data <i>Training</i> dan Data <i>Testing</i>
5.	Proses Pengolahan Data <i>Training</i>	Data <i>Training</i>	Penghitungan klasifikasi sentimen dari Data yang telah terbentuk menggunakan <i>support vector machine</i>	Model Hasil Klasifikasi
6.	Proses Kategorisasi Data <i>Testing</i>	Model Hasil Klasifikasi, Data <i>Testing</i>	Penerapan Model Klasifikasi yang telah terbentuk dari proses <i>training</i> data terhadap <i>data testing</i> .	Hasil Kategori Sentimen

7.	Uji Performansi	Model Klasifikasi Sentimen	Penghitungan performa model menggunakan <i>Confusion Matrix</i> yaitu tingkat akurasi	Hasil Uji Performansi model klasifikasi
----	-----------------	----------------------------	---	---



BAB 5. HASIL DAN IMPLEMENTASI

5.1. Hasil Penerapan Crawling pada Twitter

Pada penelitian ini, penerapan crawling berada pada proses pencarian dataset di twitter. Dataset yang didapatkan dari Youtube dengan cara menginputkan query pencarian pada proses crawling data. Twitter API yang digunakan dapat diakses melalui <https://dev.twitter.com/oauth>. Untuk dapat menggunakan Twitter API, pengguna diwajibkan login di alamat <https://dev.twitter.com> untuk mendapatkan 4 buah key berupa consumer key, consumer secret, access token dan access token secret yang akan digunakan sebagai syarat authentication. Twitter Search API berfungsi berfungsi untuk mencari kata kunci tertentu dalam tweet, menemukan tweet referensi pengguna tertentu, atau mencari tweets dari pengguna tertentu. Dengan API ini pengguna dapat memiliki akses ke data tren. Twitter Search API hanya dapat digunakan untuk mengambil data tweet. Hashtag yang dipilih adalah #gaybrondong, #gaybocah, #pijatgay, #gaylokal, #gayremaja, #gayindo, #tolakLGBT, #awasLGBT, #lawanLGBT. Hashtag – hashtag tersebut dipilih berdasarkan jumlah perolehan tweet yang memadai dan representatif untuk masing – masing kelas sentiment.

5.2. Hasil Dataset

Pada tahap klasifikasi dibutuhkan dataset yang digunakan sebagai data training. Data training digunakan sebagai bahan pembelajaran pada tahap klasifikasi sehingga model mampu memprediksi klasifikasi baik. Jumlah dataset yang digunakan sangat mempengaruhi pada nilai akurasi pada klasifikasi. Data training diperoleh dari database yang sudah diberikan label secara manual. Labeling manual dilakukan oleh Mahasiswa jurusan Psikologi Fakultas Psikologi dengan peminatan Psikologi Pendidikan dan Pengembangan. Pemilihan

Mahasiswa Jurusan Psikologi sebagai pemberi label manual adalah karena memiliki kesesuaian dengan peminatan ilmu Pendidikan dan Pengembangan sehingga dapat memberikan penilaian yang sesuai tentang Tweet yang sesuai untuk konsumsi anak pada konten tweet kaum homoseksual. Pada penelitian ini pembagian rasio penggunaan data dibagi menjadi beberapa skenario data dengan rasio sebagai berikut yaitu 70 : 30, 80 : 20 dan 80 : 20 untuk perbandingan data training dan data testing. Total Dataset yang digunakan adalah 549 dataset yang memiliki sentimen negatif dan positif. Dataset tersebut didapat dari beberapa hashtag yang telah ditetapkan, lalu dilakukan proses pemilihan dataset yang representatif terhadap sentimen negatif dan positif.

Tabel 3. Dataset

No	Tweet	Sentimen
1	Pemerintahan Mahathir Mohamad menghadapi skandal politik besar pertamanya, menyusul beredarnya video, diduga menterinya, berhubungan seks dengan seorang pria. https://t.co/DiuZyJgo7g #gayIndo	1
2	RT @detikcom: Proses sanksi disiplin mantan anggota Polda Jateng yang dipecat karena orientasi seksnya yang suka sesama jenis, berlanjut ke 1/2 #gayIndo	1
3	Yang sange malam ini coba DM #gaysmp #gaysma #gaysmk #gayremaja #gayIndo	0
4	Cari temen gay under 18 daerah semarang, gw umur 18 Kitaa sange coli colii bareng kuyy #kontolmp #kontolsma #gaysmp #gaysma #kontolbrondong #seponkontol #sepong #vcs	0

#GayRemaja #kontolsmk #bokep #gayIndo

5	Bingung sama kontol sendiri. Panjang sih lumayan. Tapi kok gemuk banget?? mau pict follow+Dm ya sayang,?? #bigdick #gayindo #kontolgede #kontolgemuk #Gayhot	0
---	---	---

5.3. Hasil Implementasi Preprocessing

Pada tahap text preprocessing, dataset diambil dengan melakukan perulangan hingga seluruh dataset selesai di Analisa dengan 4 tahapan yaitu casefolding, penghilangan URL, RT dan Hashtag tokenization, filtering dan stemming. Pada tahap casefolding data komentar yang telah diambil diubah menjadi lowercase dan melakukan penghapusan terhadap tanda baca yang terdapat dalam dataset tweet serta menghilangkan komponen lain yang tidak memberikan pengaruh pada feature dan tidak menjadi komponen dalam proses analisis.

Tabel 4. Hasil Casefolding

No	Tweet	Hasil Casefolding
1	Pemerintahan Mahathir Mohamad menghadapi skandal politik besar pertamanya, menyusul beredarnya video, menteri, berhubungan seks dengan pria seorang pria. https://t.co/DiuZyJgo7g	pemerintahan mahathir mohamad menghadapi skandal politik besar pertamanya, menyusul beredarnya video, menteri, berhubungan seks dengan pria seorang pria.
2	RT @detikcom: Proses sanksi disiplin mantan anggota Polda	proses sanksi disiplin mantan anggota polda jateng yang dipecat

	Jateng yang dipecat karena karena orientasi seksnya yang orientasi seksnya yang suka suka sesama jenis, berlanjut sesama jenis, berlanjut ke;½
3	Yang sange malam ini coba yang sange malam ini coba DM #gaysmp #gaysma #gaysmk dm #gayremaja
4	Cari temen gay under 18 daerah cari temen gay under daerah semarang, gw umur 18 semarang gw umur Kitaa sange coli colii bareng kitaa sange coli colii bareng kuyy kuyy #kontolmp #kontolsma #gaysmp #gaysma #kontolbrondong #sepongkontol #sepong #vcs #GayRemaja #kontolsmk #bokep
5	Bingung sama kontol sendiri. bingung sama kontol sendiri Panjang sih lumayan. Tapi kok panjang sih lumayan tapi kok gemuk banget?? mau pict gemuk banget mau pict follow dm follow+Dm ya sayang,?? ya sayang #bigdick #gayindo #kontolgede #kontolgemuk #Gayhot

Dataset yang sudah melalui pemrosesan awal akan masuk ke pemrosesan kedua yaitu menghilangkan stopwords atau kata penutup yang tidak memiliki andil terhadap sentimen dari sebuah tweet. Hasil penghilangan stopword atau stopword removal dapat dilihat pada tabel berikut:

Tabel 5. Hasil *Stop word*

No	Tweet	Hasil Stopword
1	Pemerintahan Mahathir Mohamad menghadapi skandal politik besar pertamanya, menyusul beredarnya video, diduga menterinya, berhubungan seks dengan seorang pria. https://t.co/DiuZyJgo7g	pemerintahan mahathir mohamad menghadapi skandal politik menyusul beredarnya video diduga menterinya seks pria berhubungan seks dengan seorang pria.
2	RT @detikcom: Proses sanksi disiplin mantan anggota Polda Jateng yang dipecat karena orientasi seksnya yang sama jenis, berlanjut ke 1/2	proses sanksi disiplin mantan anggota polda jateng dipecat orientasi seksnya suka jenis berlanjut
3	Yang sange malam ini coba DM #gaysmp #gaysma #gaysmk #gayremaja	sange malam coba dm
4	Cari temen gay under 18 daerah semarang, gw umur 18 Kita sange coli colii bareng kuyy #kontolmp #kontolsma #gaysmp #gaysma #kontolbrondong #seponkontol #sepong #vcs #GayRemaja #kontolsmk #bokep	cari temen gay under daerah semarang gw umur kita sange coli colii bareng kuyy
5	Bingung sama kontol sendiri.	bingun kontol panjang lumayan

Panjang sih lumayan. Tapi kok gemuk pict follow dm sayang
 gemuk banget?? mau pict
 follow+Dm ya sayang,??
 #bigdick #gayindo #kontolgede
 #kontolgemuk #Gayhot

Dataset yang sudah melalui pemrosesan stopword removal akan masuk ke pemrosesan ketiga yaitu melakukan proses *stemming* untuk menjadikan keseluruhan kata dalam setiap tweet menjadi kata dasar. Hasil proses *stemming* dapat dilihat pada tabel berikut:

Tabel 6. Hasil *Stemming*

No	Tweet	Hasil <i>Stemming</i>
1	Pemerintahan Mahathir Mohamad menghadapi skandal politik besar pertamanya, menyusul beredarnya video, diduga menterinya, berhubungan seks dengan seorang pria. https://t.co/DiuZyJgo7g	pemerintah mahathir mohamad hadap skandal politik susul beredar video duga menteri hubungan seks pria diduga menterinya, berhubungan seks dengan seorang pria.
2	RT @detikcom: Proses sanksi disiplin mantan anggota Polda Jateng yang dipecat karena orientasi seksnya yang sama jenis, berlanjut ke 1/2	proses sanksi disiplin mantan anggota polda jateng dipecat orientasi seksnya suka jenis berlanjut
3	Yang sange malam ini coba DM #gaysmp #gaysma #gaysmk	sange malam coba dm

#gayremaja

4 Cari temen gay under 18 daerah semarang, gw umur 18
 Kitaa sange coli colii bareng kuyy
 #kontolmp #kontolsma #gaysmp
 #gaysma #kontolbrondong
 #sepongkontol #sepong #vcs
 #GayRemaja #kontolsmk
 #bokep

5 Bingung sama kontol sendiri. Panjang sih lumayan. Tapi kok gemuk banget?? mau pict follow+Dm ya sayang,??
 #bigdick #gayindo #kontolgede
 #kontolgemuk #Gayhot

5.4. Hasil Perhitungan TF IDF

Proses perhitungan TF IDF digunakan untuk memberikan pembobotan terhadap kata yang akan dilakukan proses analisis sentimen. Proses perhitungan akan dilakukan pada contoh data sebagai berikut :

Tabel 7. Contoh Data TF IDF

NO	Teks	Kelas
1	I love gay couple	-
2	straight is best	+
3	need gay sex	-
4	I love straight	+
5	poor sex massage	-

Contoh data tersebut menghasilkan sepuluh kata untuk proses perhitungan TF IDF sebagai berikut:

Setiap kata yang telah terbentuk dinilai bobotnya berdasarkan masing – masing dokumen atau kalimat. Kata mendapat bobot satu jika muncul pada sebuah kalimat dan menambah nilai D dari kata tersebut. Setelah semua kata telah dibobotkan, maka akan masuk ke proses penghitungan DF yang merupakan jumlah dari bobot suatu kata di keseluruhan dokumen. DF akan menjadi pembagi untuk total dokumen yang dianalisis. Untuk perhitungan ini, total dokumen yang digunakan adalah lima dokumen maka D/DF adalah 5 dibagi dengan nilai DF.

Tabel 8. Perhitungan TF

Kata	TF	TF	TF	TF	TF	DF	D/DF
	D1	D2	D3	D4	D5		
I	1			1		2	2.5
love	1			1		2	2.5
gay	1		1			2	2.5
straight		1		1		2	2.5
good		1				1	5
need			1			1	5
sex			1		1	2	2.5
poor					1	1	5
couple	1					1	5
massage		1			1	2	2.5

Setelah mendapatkan nilai D/DF maka masuk ke proses selanjutnya yaitu menghitung nilai IDF. Nilai IDF didapatkan dari log nilai D/DF. Setelah didapatkan nilai IDF, maka nilai tersebut ditambahkan satu. Proses terakhir adalah menghitung nilai W atau bobot dari setiap dokumen atau kalimat. Nilai W didapatkan dari melakukan perkalian nilai dari TF dari masing – masing kata untuk setiap dokumen dengan nilai IDF+1. Total keseluruhan dari nilai W untuk setiap dokumen merupakan W untuk setiap dokumen. Proses perhitungan ini

sudah dapat mempersiapkan dataset untuk diproses dalam algoritma klasifikasi SVM.

Tabel 9. Perhitungan W

IDF	IDF+1	W=TF*(IDF+1)				
		D1	D2	D3	D4	D5
0.39794	1.397940009	1.39794	0	1.397940009	1.39794	0
0.39794	1.397940009	1.39794	0	1.397940009	1.39794	0
0.39794	1.397940009	1.39794	0	0	0	0
0.39794	1.397940009	0	1.39794	1.397940009	1.39794	0
0.69897	1.698970004	0	1.69897	0	0	0
0.69897	1.698970004	0	0	0	0	0
0.39794	1.397940009	0	0	0	0	1.39794
0.69897	1.698970004	0	0	0	0	1.69897
0.69897	1.698970004	1.69897	0	0	0	0
0.39794	1.397940009	0	1.39794	0	0	1.39794
Hasil TF IDF per Kalimat		5.89279	4.49485	4.193820026	4.19382	4.49485

Hasil perhitungan bobot yang dihasilkan oleh setiap kata untuk setiap kalimat menjadi dasar proses perhitungan sebagai nilai X untuk proses klasifikasi menggunakan SVM.

5.5. Hasil Testing dengan Skenario Data

Proses testing dengan data baru dilakukan agar dapat mengetahui akurasi model yang sebenarnya, pada proses ini dilakukan testing dengan menggunakan beberapa skenario data. Berikut adalah hasil testing di berbagai model melalui beberapa skenario data.

- a) Hasil Testing dengan skenario data 60:40

Tabel 10. CM 60:40

CONFUSION MATRIX		Predicted Class	
		Negatif	Positif
Actual	Negatif	206	2
Class	positif	1	199

Tabel diatas menunjukkan hasil pengujian menggunakan model yang telah dihasilkan dengan skenario data 40:60 menunjukkan bahwa 206 data diprediksi secara akurat untuk kelas negatif dan 199 data diprediksi secara akurat untuk kelas positif. Dan 8 data diprediksi secara tidak akurat. Akurasi untuk skenario data ini adalah sebesar 0,9926470588235294 atau 99,26 %. Sedangkan untuk nilai precision dari model adalah sebesar 0,9900497512437811 atau 99 %, recall sebesar 0,995 dan f1 score sebesar 0,9925187032418954

b) Hasil testing dengan skenario data 70:30

Tabel 11. CM 70:30

CONFUSION MATRIX		Predicted Class	
		Negatif	Positif
Actual	Negatif	144	2
Class	positif	0	160

Tabel diatas menunjukkan hasil pengujian menggunakan model yang telah diproduksi dengan skenario data 30:70 menunjukkan bahwa 144 data diprediksi secara akurat untuk kelas negatif dan 160 data diprediksi secara akurat untuk kelas positif. Dan 2 data diprediksi secara tidak akurat. Akurasi untuk skenario data ini adalah sebesar 0,9934640522875817 atau 99,34%. Precision sebesar 0,9876543209876543 atau 98,6% , recall sebesar 1,0 dan f1 score sebesar 0,9937888198757764

c) Hasil Testing Skenario Data 80:20

Tabel 12. CM 80:20

CONFUSION MATRIX		Predicted Class	
		Negatif	Positif
Actual Class	Negatif	92	1
	positif	0	111

Skenario data 20:80 memiliki akurasi 0.9950980392156863 atau 99,50 %. Berdasarkan Tabel diatas, hasil pengujian menggunakan model yang telah diproduksi dengan skenario data 20:80 menunjukkan bahwa 92 data diprediksi secara akurat untuk kelas negatif dan 111 data diprediksi secara akurat untuk kelas positif. Dan 1 data diprediksi secara tidak akurat. Sedangkan untuk precision sebesar 0,9910714285714286 atau 99,10% ,recall sebesar 1,0 , dan f1 score sebesar 0,9955156950672646

Berikut adalah daftar semua hasil pengujian performa berdasarkan pada skenario data yang diimplementasikan pada set tes. Untuk membuat skor akhir akurasi, precision, recall dan F1 Score kita perlu menghitung data dari *confusion matriks*. Ambil contoh dari Tabel untuk menghitung nilai akurasi dengan perhitungan sebagai berikut :

Proses perhitungan akurasi :

$$akurasi = \frac{405}{408} = 0,99264705 = 99,26\%$$

Proses perhitungan precision :

$$precision = \frac{199}{201} = 0,9900497 = 99\%$$

Proses perhitungan recall :

$$recall = \frac{199}{200} = 0,995$$

Proses perhitungan f1 Score:

$$f1\ score = \frac{2 \cdot (0,9900497512437811 \cdot 0,995)}{(0,9900497512437811 + 0,995)} = 0,9925187032418954$$

Tabel 13. Hasil Performa Keseluruhan

No	Data Scenario	Accuracy	Precision	Recall	F1 Score
1	40 : 60	99,26 %	99 %	0,995	0,9925187032418
2	30 : 70	99,34%	98,6%	1,0	0,9937888198757
3	20 : 80	99,50 %	99,10%	1,0	0,9955156950672

5.6. Tampilan Halaman Search Tweet

Tampilan awal dalam sistem ini adalah tampilan search box untuk mencari frasa untuk melakukan proses crawling tweet. Pengguna dapat memasukkan frasa yang diinginkan dalam kolom pencarian dan memilih ikon cari. Sistem akan melakukan proses crawling sehingga tweet yang diinginkan dapat ditampilkan seperti pada gambar.



Analisis Sentimen Tweet Kaum Homoseksual

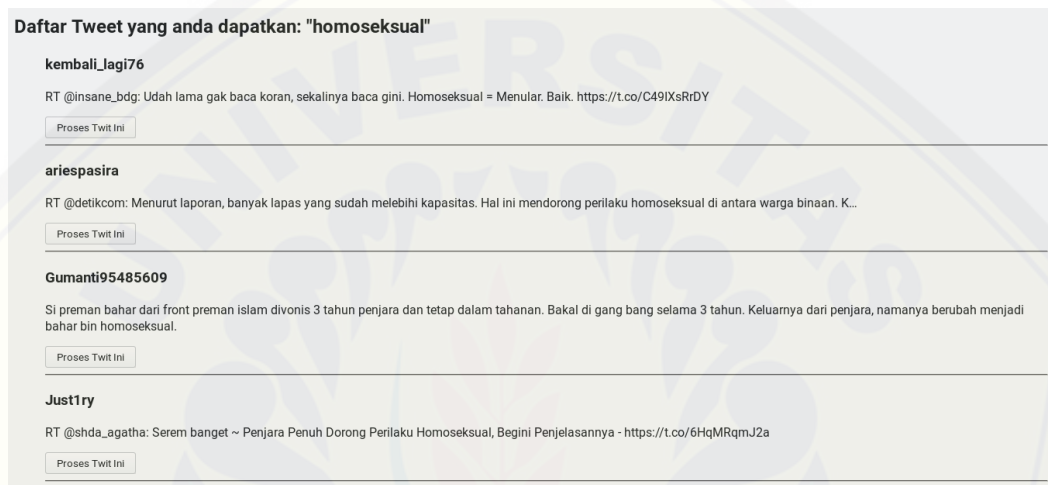
Isikan Tweet yang ingin anda Analisis:

homoseksual

Gambar 20. Tampilan Pencarian Tweet

5.7. Tampilan Halaman Hasil Crawling

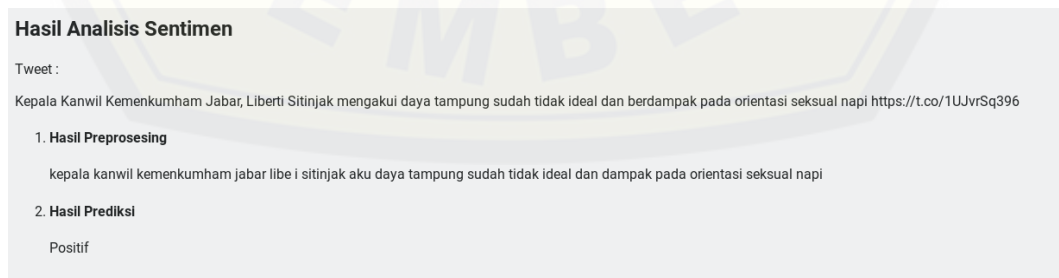
Tampilan hasil crawling ketika proses crawling telah selesai dilaksanakan. Crawling hanya akan menampilkan sepuluh tweet yang dapat dipilih pengguna sesuai dengan frasa yang diinginkan. Pengguna dapat memulai proses analisis sentimen pada tweet dengan memilih tombol Proses Twit Ini seperti pada gambar.



Gambar 21. Tampilan Hasil *Crawling*

5.8. Tampilan Halaman Hasil Analisis Sentimen

Tweet yang sudah dipilih akan diproses dan setelah proses selesai akan memberikan hasil. Hasil yang ditampilkan adalah hasil Preprocessing dan juga hasil prediksi sentimen dari tweet tersebut.



Gambar 22. Tampilan Hasil Analisis Sentimen

BAB 6 KESIMPULAN DAN SARAN

Kesimpulan

1. Sistem analisis sentimen ini dapat mengklasifikasikan tweet yang diperoleh dari proses perayapan twitter dengan akurasi yang cukup baik. Sistem ini mampu mengklasifikasikan sentimen menjadi dua jenis yaitu positif dan negatif.
2. Berdasarkan hasil pengujian, model yang dihasilkan mencapai akurasi tertinggi yaitu 99,50% saat menggunakan data skenario 80:20, sedangkan akurasi terendah 99,26% dicapai ketika menggunakan skenario data 60:40
3. Skenario data dapat memengaruhi tingkat akurasi semakin banyak jumlah data pelatihan yang diberikan, semakin tinggi akurasi yang akan diperoleh, sedangkan jika Anda menggunakan lebih sedikit data pelatihan, hasilnya akan sedikit tidak stabil.
4. Keseimbangan rasio dataset dapat mempengaruhi nilai Precision, recall, dan F1 Score. Dataset yang *imbalanced* pada penelitian melewati proses *oversampling* agar performa yang dihasilkan model dapat menjadi lebih maksimal.

Saran

1. Daftar kata berhenti perlu ditingkatkan ke daftar kata berhenti terbaru yang digunakan. *Stemming* factory dictionary perlu ditingkatkan sehingga proses *stemming* dapat bekerja secara efisien pada kumpulan dan data yang digunakan oleh orang-orang saat ini.
2. Jumlah kelas positif dan negatif yang tidak seimbang dalam dataset juga merupakan salah satu kendala dalam penelitian ini. Jumlah yang tidak seimbang itu tidak akan memberi tahu kita kinerja model dalam memprediksi kelas positif. Ini terjadi karena jumlah tweet positif dalam tweet homoseksual tidak sebesar tweet negatif.

DAFTAR PUSTAKA

- Bhattasali, N. and Maiti, E. (2015) 'Machine "Gaydar": Using Facebook Profiles to Predict Sexual Orientation CS 229 : Machine Learning', p. 2015.
- Budi Ilmawan, L. and Winarko, E. (2015) 'Aplikasi Mobile untuk Analisis Sentimen pada Google Play', *Ijccs*, 9(1), pp. 53–64.
- Chrystal, J. B. and Joseph, S. (2015) '*Text mining* and Classification of Product Reviews Using Structured Support Vector Machine', *The International Journal of Multimedia & Its Applications (IJMA)*, 5(4), pp. 21–31.
- Dadgar, S. M. H., Araghi, M. S. and Farahani, M. M. (2016) 'A novel *Text mining* approach based on TF-IDF and support vector machine for news classification', *Proceedings of 2nd IEEE International Conference on Engineering and Technology, ICETECH 2016*, (March), pp. 112–116. doi: 10.1109/ICETECH.2016.7569223.
- Ghani, H. (2018) Muncul Grup Gay Pelajar SMP di Garut, Wabup: Jumlahnya Ribuan, Muncul Grup Gay Pelajar SMP di Garut, Wabup: Jumlahnya Ribuan. Available at: <https://news.detik.com/berita-jawa-barat/d-4247306/muncul-grup-gay-pelajar-smp-di-garut-wabup-jumlahnya-ribuan>.
- Hamouda, S. Ben and Akaichi, J. (2013) 'Social Networks' *Text mining* for Sentiment Classification: The case of Facebook' statuses updates in the "Arabic Spring" Era', *International Journal of Application or Innovation in Engineering & Management*, 2(5), pp. 470–478. Available at: <http://www.ijaiem.org/Volume2Issue5/IJAIEM-2013-05-26-063.pdf>.
- JPNN (2019) Twitter Jadi Aplikasi Paling Disukai Menyebarkan Konten Pornografi. Available at: <https://www.jpnn.com/news/twitter-jadi-aplikasi-paling-disukai-menyebarkan-konten-pornografi?page=1>.
- Kaplan, M. B. (1997) *Sexual Justice: Democratic Citizenship and the Politics of*

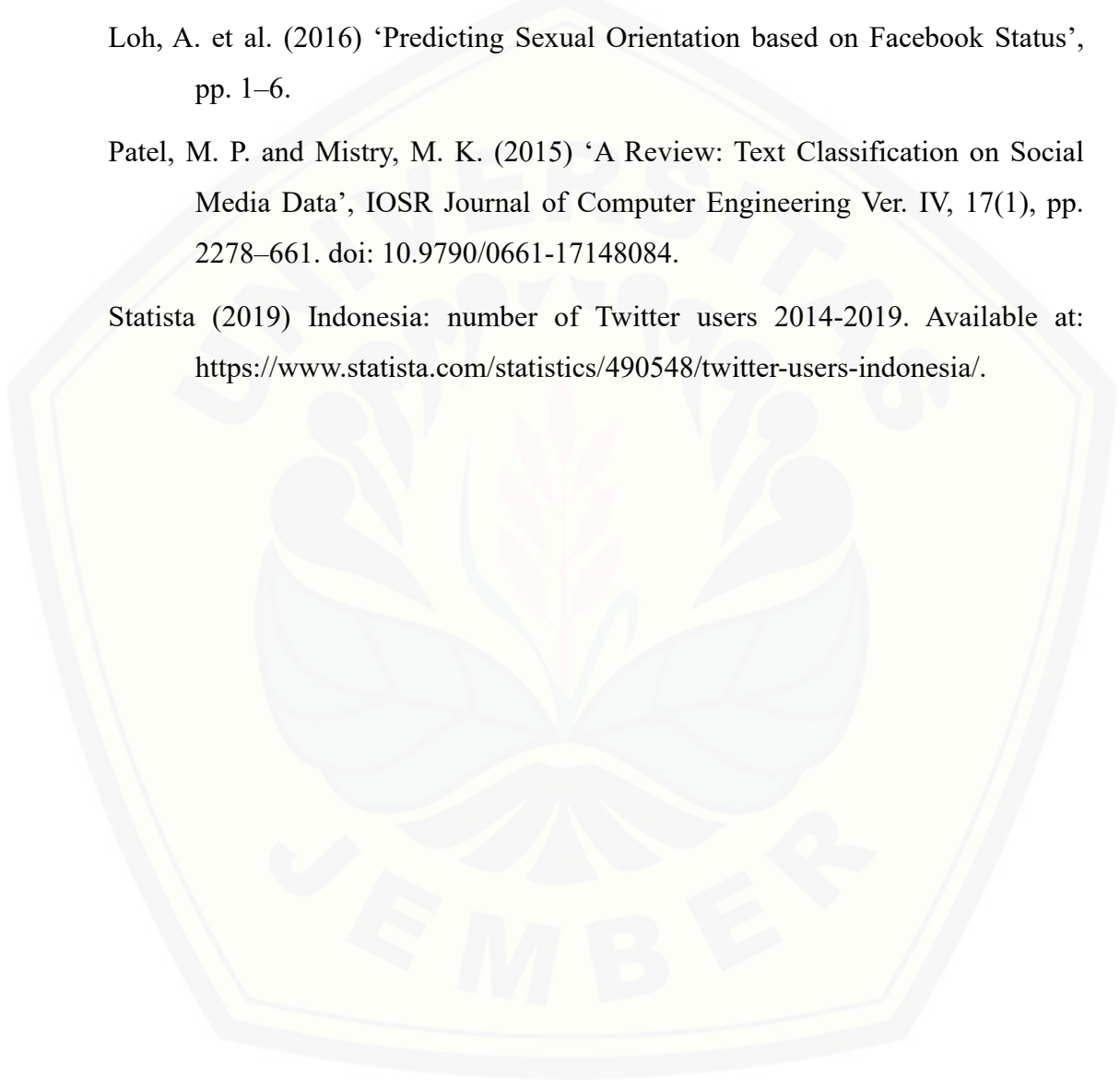
Desire. 1st Editio. New York. doi: <https://doi.org/10.4324/9780203610046>.

Kementrian Kesehatan RI, K. R. (2017) Laporan Perkembangan HIV-AIDS dan Penyakit Menular Seksual (PIMS) Triwulan I Tahun 2017.

Loh, A. et al. (2016) 'Predicting Sexual Orientation based on Facebook Status', pp. 1–6.

Patel, M. P. and Mistry, M. K. (2015) 'A Review: Text Classification on Social Media Data', IOSR Journal of Computer Engineering Ver. IV, 17(1), pp. 2278–661. doi: 10.9790/0661-17148084.

Statista (2019) Indonesia: number of Twitter users 2014-2019. Available at: <https://www.statista.com/statistics/490548/twitter-users-indonesia/>.



LAMPIRAN

NO	sentimen	teks
1	0	twitter ngulah lagi gak bisa post foto/video jadi gak bisa up testi MAU VCS? DM AJA YA #memekbecek #memekabg #memekmulus #vcscrot #vcswhatsapp #vcsreal #vcsyuk #bobandung #vcstrusted #spermanya #abnggentot #abgmesum #abgsange #gaylokal #gay #bokepgay #ceweksange #vcs
2	0	Pngen gitu yah punya pcr brondong biar bisa sepong"an?? pengen coba di anal juga klo brondong'y mau wkwk #gaylokal #gayjakarta #gayjkt #gaytangerang #gaybrondong #gayjogja #GayMalang #gaysmk #GayRemaja #gayColi #Crossdresser #ladyboyjakarta #ladyboy #wariajakarta #shemalejakarta
3	0	AVAIL VCS NOW! LANGSUNG DM TESTI> #testicindy OPEN GRUP #VVIP 100K #memekbecek #memekabg #memekmulus #vcscrot #vcswhatsapp...
4	0	Sumpah pengen banget gw diewe sama model kek gini Part 1 #gay #gaylokal #kuligay https://t.co/WjxqSRgvAh
5	0	Open Booking area Jogja, therapist umur 25th, yang minat dan Yg serius DM. Selama bulan puasa open malam saja. #gaylokal #gayjogja #JogjaBO #pijatsensual #pijatjogja #pijatenak #gayindonesia #pijatpanggilan #gaypijatenak #massageplus #readyjogja #jogjacrot #jogjanikmat
6	0	Ada yang di alun"Sukowati ga..?@Birudongker16 #gaysragen #gaylokal #gayindonesia #gay
7	0	Yang Top temenin chat yu, wa gw ada di bio tuh! #gayindonesia #gaylokal #gaybrondong #GayRemaja #gaytop #gaymalang #gayjember #gayjatim
8	0	ngewe sama tukang somay udh punya istri. yg mau order full video nya+bonus wa nya bs lgsg order ke 081398866172 #gayvid #gaylokal #gayindo #gayjkt #escortjkt #vidgay #bokepindo #bokepgay #videogay #gayindonesia #gayjakarta #gaybandung #gaymedan #gayescort #moneyboy https://t.co/IOSaZOB4rM
9	0	Ada T surabaya? Need fun/3s DM me #gaylokal#gaysurabaya#gaychinese
10	0	Colbar yu #gaylokal #gaybrondong #gayindonesia

11	0	Yuk yg mau vcs bisa add saya. Bot dan sma boleh Wa 08973902961#gaylokal #GayRemaja #gaybrondong #ngewe #ngentot #vcs #kontolsma #Sma
12	0	Yukk gue siap di genjot sampai lo puas 088211447875 #gay #gayindonesia #gaythailand #gaytaiwan #gaysiangapore #gaymalaisia...
13	0	Yukk gue siap di genjot sampai lo puas 088211447875 #gay #gayindonesia #gaythailand #gaytaiwan #gaysiangapore #gaymalaisia #gayamerika #gayarab #gayjakarta #gaytangerang #gaybali #gayporn #gaytop #gaybot #gayjabodatabek #gaybo #gaymedan #gaybandung #GayRemaja #gaysex #gaykorea
14	0	Yang jauh mendekat yang dekat merapat 08977029728 Monggo di chat @GayINDONESIA @gaybandung @gayjakarta #brondongs...
15	0	Disini top yang 17 keatas ada nggak ya?#gayindoesia #gaylokal #GayRemaja #GayMalang #gaybrondong #gayjatim #gayjember
16	0	Area Bali kuta dan sekitarnya wa:081333280463@PijatBali4 @Massage_bali_ #bo#ESCORT#pijatpria#pijatenak#GayRemaja @jack19680@luckypoin7@bintaro_boi@Candupijjat@DMassageJkt@GudangPijat2@fjr_pijadict1 #gaybrondong @Player05625885#gaybali @Rifaiju1 @riky21713617 https://t.co/6YqDuUgfMX
17	0	Yang jauh mendekat yang dekat merapat 08977029728 Monggo di chat @GayINDONESIA @gaybandung @gayjakarta #brondongsma#brondongjakarta#gaymalaysia#gaysingapore #GayRemaja #gaybali
18	0	ME : Bot,16th Stay Surabaya, yang ada tempat, Dm cari under 25#GaySby#gaybrondong #gaySmk#Gaysby #gaylokal #GayRemaja #gaysurabaya #Gay
19	0	Good morning...para suhu ??Tetap semangat ya puasanyalzin promo...pijat pull body Anda minat silahkan Wa 082112236525In call/out call#ESCORT #gaybandung #gayjakarta#GayRemaja@MasTri1103@jack19680@luckypoin7 @bintaro boi @pondok pria@Candupijja@DMassageJkt @GudangPija https://t.co/mA8R6FhD0v
20	0	RT @Pijat1Surabaya: Out luar kota dalam kota #Jatim Wa +6283831389000@pijatprofesonal@CanduPijat @PijatPriaSby1 @Jasapijatsby1 @pijats...
21	0	Aduhhh, kalo udah ketemu laki2 yang begini, pikir sange deh ?????? #gay #gaybrondong#GayRemaja #gaybandung #GayChines...
22	0	16 tahun. #gayklaten #gaylokal #gayjogja #gaysma #gaysmk #GayRemaja #gayindonesia #gaybrondong

23	0	MB_Big?? Malam?? "Izin update zuhuWa 087848345687Incall/Outcall #ESCORT #gaybandung #gayjakarta #GayRemaja@MasTri1103...
24	0	Good night....para suhu ?? Izin promo.... pijat pull body ++ Anda minat silahkan Wa 082112236525 In call/out call#ESCORT #ga...
25	0	Fun malem ini yg punya tempat daerah buah batu bandung #gaybandung #gaybdg #bandungbelok #gbpromote #gaysma #gayremaja
26	0	Pngen gitu yah punya pcr brondong biar bisa sepong"an?? pengen coba di anal juga klo brondong'y mau wkwk #gaylokal #gayjakarta #gayjkt #gaytangerang #gaybrondong #gayjogja #GayMalang #gaysmk #GayRemaja #gayColi #Crossdresser #ladyboyjakarta #ladyboy #wariajakarta #shemalejakarta
27	0	#gaybrondong #gayasurabaya #GayRemaja #kontrolbrondong #kontolsma sempak nya siapa ini? Wkwkwkwk https://t.co/R4c5mHNoKS
28	0	Good night....para suhu ??Izin promo.... pijat pull body ++Anda minat silahkanWa 082112236525In call/out call#ESCORT#gaybandung #gayjakarta #GayRemaja @MasTri1103 @jack19680 @luckypoin7 @bintaro_boi @pondok_pria @Candupijjat @DMassageJkt @GudangPijat2 @fjr pijaddict1 https://t.co/jbJZVL8AiK
29	0	Open BO area sidoarjo dan surabaya .slahkan DM atau wa 088217234436 /CALL: 089687758225 #gaysurabaya #gaysidoarjo #gaysex #gayhot #gaytop #gaybot #GayRemaja https://t.co/Ao7zvdlfWd
30	0	Open BO area sidoarjo dan surabaya .slahkan DM atau wa 088217234436 /CALL: 089687758225 #gaysurabaya #gaysidoarjo #gaysex #gayhot #gaytop #gaybot #GayRemaja https://t.co/MFiFEo4s7K
31	0	Yang Top temenin chat yu, wa gw ada di bio tuh! #gayindonesia #gaylokal #gaybrondong #GayRemaja #gaytop #gaymalang #gayjember #gayjatim
32	0	Open vidcol buat siapa aja 085325105018 privasi aman, berbayar #gaylokal #gayjakarta #gaybrondong #gayjogja #GayRemaja
33	0	Selamat siang...all ?? Tetap semangat ya puasanyaIzin promo.... pijat pull body Anda minat silahkan Wa 082112236525 In call/out call #ESCORT#gaybandung #gayjakarta #GayRemaja @MasTri1103 @jack19680 @luckypoin7 @bintaro_boi @pondok_pria @Candupijjat @DMassageJkt @GudangPija https://t.co/nYjftCo2NO

34	0	Area Bali kuta dan sekitarnya wa:089671780209 @PijatBali4 @Blueskymassage2 #ESCORT #pijatpria #pijatenak #GayRemaja @jack19680 @luckypoin7 @bintaro_boi @Candupijjat @DMassageJkt@GudangPijat @fjr_pijaddict1 #gaybrondong @Player05625885 #gaybali @Rifaiju1 @riky21713617 https://t.co/ZEWpyRVGNF
35	0	Di Barat sendiri, notabene Kiblatnya pelegalan LGBT, ilmunya terpecah menjadi dua kubu dalam hal Gay-Biseksual. *Pro Gay-Biseksual tidak dapat dirubah, pendapatnya APA *Pro Gay-Biseksual dapat dirubah, pendapatnya NART #gaylokal #gayindonesia #gayjogja #gaybrondong #GayRemaja
36	0	Pengen punyo kenalan wong palembang #gaylokal #GayRemaja #gay #gayindonesia #gaypalembang #gaybocah
37	0	Available..Big dick 20cm jakbarWA : 087848345687 Izin suhu the best #ESCORT #gaybandung #gayjakarta #GayRemaja @MasTr...
38	0	Gua nyari temen gay gua 16 yo kita sange sangean bareng vcs coli skuy?? #GayRemaja #kontolbrondong #gay #gayjakarta #kontolsm...
39	0	Ada ngga nih anak klaten umur 16 kebawah klo gk 16 mpe 18? #gaybocah #GayRemaja #gaysurakarta #gayklaten #gaysmp #gaysma #gaysmk
40	0	18 thn boy, coli bareng atao sm cwe jg bisa lgsg dm aja ya mksih #kontol #MEMEKBASAH #ayamkampusjogja #sange #toketabg #toketgede #toketbesar #GayRemaja #gaybrondong #kontolbrondong #kontolsma #tantegirang #gayjakarta
41	0	MB_Big?? Malam??"Izin update zuhu Wa 087848345687 Incall/Outcall #ESCORT#gaybandung#gayjakarta #GayRemaja @MasTri1103 @jack19680 @luckypoin7 @bintaro_boi @pondok_pria @Candupijjat @DMassageJkT @GudangPijat2 @fjr pijaddict1 #gaybrondong #bokepgay #GayMalang #gaydepok https://t.co/iHkUpk0y0R
42	0	Kepo pengen nyobain anjir #gaybandung #gayremaja
43	0	yang di solo mana nih?? sini ke kost kita berzina???? #gaysolo #gaysurakarta #gaykaranganyar #gaysukoharjo #GayIndonesia #GayR...
44	0	Good night....para suhu Izin promoallI am massage boy.....Anda minat silahkanWa 082112236525In call/out call (Jakarta) #ESCORT #gaybandung #gayjakarta #GayRemaja @MasTri1103 @jack19680 @luckypoin7 @bintaro_boi @pondok_pria @Candupijjat @DMassageJkt @GudangPijat2 https://t.co/X0qlF4FgQv
45	0	Yang sange malam ini coba DM #gaysmp #gaysma #gaysmk #gayremaja

46	0	Sepi nihhhhhhhhhh yuk seneng seneng #gayindonesia #GayRemaja #gaybrondong #kontolsma #kontolsmp #bokepindo #bokep #sange #sepong #vcs #cs #Sangebanget #gaysalatiga #KontolGededanPanjang #kontolbrondong #gaysmk #gaysemarang #gayjakarta #gayColi #gaysemarang
47	0	Realita TDK membohongi fakta Selesai eksekusi customer di cilandak #gayjakarta #GayRemaja @MasTri1103 @jack19680 @luckypoin7

