

# **IDENTIFIKASI PELECEHAN SEKSUAL PADA INSTAGRAM DENGAN METODE SUPPORT VECTOR MACHINE**

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



Oleh:

Otto Tirtayasa Akbar

NIM : 09021281722055

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2022

# LEMBAR PENGESAHAN SKRIPSI

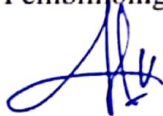
## Identifikasi Pelecehan Seksual Pada Instagram Dengan Metode Support Vector Machine

Oleh :

Otto Tirtayasa Akbar

NIM : 09021281722055

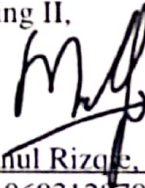
Pembimbing I



Alvi Syahrini Utami, M.Kom  
NIP. 197812222006042003

Palembang, 18 Januari 2023

Pembimbing II,



M. Qurhanul Rizqie, S.Kom., M.T, Ph.D.  
NIP. 1671060312870008

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom  
NIP. 197812222006042003

## TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Kamis tanggal 29 Desember 2023 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Otto Tirtayasa Akbar  
NIM : 09021281722055  
Judul : Identifikasi Pelecehan Seksual Pada Instagram Dengan Metode *Support Vector Machine*

dan dinyatakan **LULUS**.

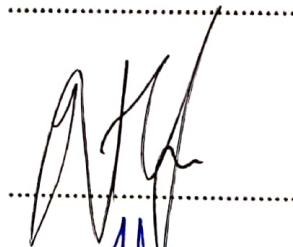
1. Ketua

M. Fachrurrozi, S.Si., M.T.  
NIP. 198005222008121002



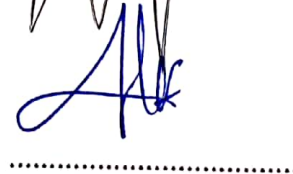
2. Penguji

Novi Yusliani, M.T.  
NIP. 198211082012122001



3. Pembimbing I

Alvi Syahrini Utami, M.Kom  
NIP. 197812222006042003



4. Pembimbing II

M. Qurhanul Rizqie, S.Kom., M.T, Ph.D.  
NIP. 1671060312870008



Mengetahui,  
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom  
NIP. 197812222006042003

## HALAMAN PERNYATAAN

Yang Bertanda Tangan di bawah ini :

Nama : Otto Tirtayasa Akbar  
NIM : 09021281722055  
Program Studi : Teknik Informatika  
Judul Skripsi : Identifikasi Pelecehan Seksual Pada Instagram Dengan Metode *Support Vector Machine*

Hasil Pengecekan Software *ithenticate/Turnitin* : 7%

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

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 18 Januari 2023



Otto Tirtayasa Akbar  
NIM. 09021281722055

## MOTTO DAN PERSEMBAHAN

Motto :

“Sebuah pertandingan belum bisa dianggap berakhir sebelum pertandingan tersebut benar-benar berakhir. Selagi ada peluang sekecil apapun untuk menang, tetap berpeganglah pada peluang kecil itu”

Kupersembahkan karya tulis ini kepada

- Ayah dan Bunda
- Adik-adikku
- Keluarga Besar
- Dosen Pembimbing dan Penguji
- Teman-teman Seperjuangan
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

## **SEXUAL HARASSMENT DETECTION ON INSTAGRAM USING SUPPORT VECTOR MACHINE**

By:

Otto Tirtayasa Akbar(09021281722055)

Informatics Department, Computer Science Faculty, Sriwijaya University

Email : ottotirtayasa16@gmail.com

### **ABSTRACT**

The existence of social media makes many people upload their experiences, either in the form of writing or multimedia. The reactions from these uploads also vary, some respond well in the form of praise, suggestions and positive comments, some respond badly such as giving reproaches, insults and even harassing sentences. This study aims to identify whether a sentence on Instagram includes sexual harassment or not using the Support Vector Machine (SVM) classification method. The SVM method was chosen because this method is often used for classification and has good performance results. In its implementation, the data is pre-processed to make the data cleaner and more structured, then feature extraction is carried out using Term Frequency-Inverse Document Frequency (TF-IDF) and classification is carried out using the SVM method with a linear kernel. Testing this research using the K-Fold Cross Validation method, the data, in total 200 Instagram comments containing 100 harassment comments and 100 non-harassment comments, is divided into 90% training data and 10% test data with a K value of 10. After 10 tests, the SVM classification method with linear kernel produces an accuracy value of 71%, a precision value of 93%, a recall value of 65% and an f-1 value of 76% with an average computation time of 0.035 seconds.

Keyword: Sexual Harassment, Instagram, Social Media, Support Vector Machine (SVM)

# **IDENTIFIKASI PELECEHAN SEKSUAL PADA INSTAGRAM DENGAN METODE SUPPORT VECTOR MACHINE**

Oleh:

Otto Tirtayasa Akbar(09021281722055)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email : ottotirtayasa16@gmail.com

## **ABSTRAK**

Dengan adanya media sosial, masyarakat banyak yang mengunggah pengalaman-pengalamannya, baik dalam bentuk tulisan ataupun multimedia. Reaksi dari unggahan tersebut pun beragam, ada yang merespon dengan baik berupa pujian, saran dan komentar positif, ada juga yang merespon dengan buruk seperti memberikan celaan, hinaan bahkan kalimat yang melecehkan. Penelitian ini bertujuan untuk mengidentifikasi apakah sebuah kalimat pada Instagram termasuk pelecehan seksual atau bukan menggunakan metode klasifikasi *Support Vector Machine* (SVM). Metode SVM dipilih karena metode ini sering digunakan untuk klasifikasi dan memiliki hasil kinerja yang baik. Dalam implementasinya, data dilakukan pra-pengolahan untuk membuat data menjadi lebih bersih dan terstruktur, selanjutnya dilakukan ekstraksi fitur menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF) dan dilakukan klasifikasi menggunakan metode SVM dengan kernel linear. Pengujian penelitian ini menggunakan metode *K-Fold Cross Validation*, data berupa total 200 komentar instagram yang berisi 100 komentar pelecehan dan 100 komentar non-pelecehan dibagi menjadi 90% data latih dan 10% data uji dengan nilai K sebesar 10. Setelah dilakukan 10 kali pengujian, metode klasifikasi SVM kernel linear menghasilkan nilai akurasi sebesar 71%, nilai presisi sebesar 93%, nilai recall sebesar 65% dan nilai f-1 sebesar 76% dengan waktu komputasi rata-rata sebesar 0.035 detik.

Kata Kunci: Pelecehan seksual, Instagram, Media Sosial, *Support Vector Machine* (SVM)

## **KATA PENGANTAR**

Puji syukur kehadirat Allah SWT karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi ini. Dalam menyelesaikan Skripsi ini ada saja kendala dan halangan yang menghampiri, namun atas bantuan dan support dari berbagai pihak, penulis pun mampu mengatasi berbagai kendala dan halangan tersebut. Dalam kesempatan ini pula penulis ingin menyampaikan rasa terima kasih penulis kepada:

1. Kedua orang tua dan adik-adik, serta keluarga besar yang telah memberi doa dan support kepada penulis.
2. Ibu Alvi Syahrini Utami, M.Kom. selaku ketua jurusan Teknik Informatika dan dosen Pembimbing 1 serta Bapak M. Qurhanul Rizqie, S.Kom., M.T, Ph.D. selaku dosen Pembimbing 2 yang telah membimbing penulis dalam pengerjaan skripsi.
3. Ibu Yunita, S.Si, M.Cs. selaku dosen pembimbing akademik yang telah membimbing penulis pada masa perkuliahan.
4. Seluruh dosen bimbingan intensif yang telah membantu penulis dalam pengerjaan skripsi.
5. Seluruh dosen Fakultas Ilmu Komputer yang telah membagikan ilmu pengetahuan kepada penulis.
6. Seluruh staf dan pegawai Fakultas Ilmu Komputer yang juga membantu penulis dalam kelancaran pengerjaan skripsi dan perkuliahan.
7. Kak Dhea selaku partner saya pada saat pengerjaan laporan KP, pengerjaan skripsi beserta pengurusan berkas-berkas di akhir perkuliahan.



8. Iqbal, Arvin, Ikhsan, Dika, Abeng, Tiara, Zaneva, Zhicma, Nur, Aril, Marinah, Naufal, dan Revan yang bersama-sama berjuang di semester akhir ini. Dan juga teman sekelas dan seangkatan yang lainnya.

Dalam pengerjaan skripsi ini penulis menyadari masih banyaknya kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karenanya kritik dan saran sangat diharapkan untuk kemajuan penelitian selanjutnya. Semoga skripsi ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, 28 Desember 2022



Otto Tirtayasa Akbar

## DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN SKRIPSI.....	i
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	ii
HALAMAN PERNYATAAN .....	iii
MOTTO DAN PERSEMBAHAN .....	iv
ABSTRACT .....	v
ABSTRAK .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR .....	xiii
BAB I.....	I-1
1.1 Pendahuluan .....	I-1
1.2 Latar Belakang Masalah.....	I-1
1.3 Rumusan Masalah .....	I-3
1.4 Tujuan Penelitian .....	I-3
1.5 Manfaat Penelitian .....	I-3
1.6 Batasan Masalah .....	I-3
1.7 Sistematika Penulisan .....	I-3
1.8 Kesimpulan .....	I-5
BAB II.....	II-1
2.1 Pendahuluan .....	II-1
2.2 Landasan Teori.....	II-1
2.2.1 Pelecehan Seksual .....	II-1
2.2.2 Natural Language Processing.....	II-2
2.2.3 Support Vector Machine .....	II-3
2.2.4 Term Frequency–Inverse Document Frequency(TF-IDF).....	II-8
2.2.5 Rational Unified Process.....	II-9
2.2.6 Confusion Matrix .....	II-10
2.2.7 K-Fold Cross Validation .....	II-12

2.3 Penelitian Lain yang Relevan .....	II-12
2.4 Kesimpulan .....	II-13
<b>BAB III .....</b>	<b>III-1</b>
3.1 Pendahuluan .....	III-1
3.2 Pengumpulan Data .....	III-1
3.3 Tahapan Penelitian .....	III-1
3.3.1 Kerangka Kerja .....	III-1
3.3.2 Kriteria Pengujian .....	III-4
3.3.3 Format Data Pengujian.....	III-4
3.3.4 Alat yang Digunakan dalam Pelaksanaan Penelitian .....	III-6
3.3.5 Pengujian Penelitian.....	III-6
3.3.6 Analisis Hasil Pengujian dan Membuat Kesimpulan.....	III-6
3.4 Metode Pengembangan Perangkat Lunak.....	III-7
3.4.1 Fase Insepsi .....	III-7
3.4.2 Fase Elaborasi .....	III-8
3.4.3 Fase Konstruksi.....	III-8
3.4.4 Fase Transisi .....	III-8
3.5 Manajemen Proyek Penelitian .....	III-8
<b>BAB IV .....</b>	<b>IV-1</b>
4.1 Pendahuluan .....	IV-1
4.2 Fase Insepsi .....	IV-1
4.2.1 Pemodelan Bisnis.....	IV-1
4.2.2 Kebutuhan Perangkat Lunak .....	IV-2
4.2.3 Analisis Kebutuhan dan Perancangan Perangkat Lunak.....	IV-3
4.2.4 Implementasi .....	IV-20
4.3 Fase Elaborasi .....	IV-23
4.3.1 Implementasi .....	IV-23
4.4 Fase Konstruksi.....	IV-25
4.4.1 Analisis dan perancangan.....	IV-25
4.4.2 Implementasi .....	IV-26
4.5 Fase Transisi .....	IV-27
4.5.1 Rencana Pengujian .....	IV-27
4.5.2 Implementasi .....	IV-29

4.6 Kesimpulan .....	IV-31
<b>BAB V.....</b>	<b>V-1</b>
5.1 Pendahuluan .....	V-1
5.2 Data hasil percobaan .....	V-1
5.2.1 Konfigurasi Percobaan .....	V-1
5.2.2 Data hasil konfigurasi .....	V-1
5.3 Analisis Hasil Penelitian .....	V-3
5.4 Kesimpulan .....	V-5
<b>BAB VI .....</b>	<b>VI-1</b>
6.1 Pendahuluan .....	VI-1
6.2 Kesimpulan .....	VI-1
6.3 Saran.....	VI-2
<b>DAFTAR PUSTAKA .....</b>	<b>xiv</b>
<b>LAMPIRAN</b>	

## DAFTAR TABEL

	Halaman
Tabel III-1. Rancangan tabel hasil deteksi .....	III-5
Tabel III-2. Rancangan tabel hasil pengujian .....	III-5
Tabel III-3. Rancangan Tabel Hasil Anaisis Hasil Pengujian.....	III-7
Tabel III-4. Tabel Perencanaan Kegiatan Penelitian.....	III-8
Tabel IV-1. Kebutuhan Fungsional.....	IV-2
Tabel IV-2. Data Komentar.....	IV-4
Tabel IV-3. Hasil Case Folding .....	IV-5
Tabel IV-4. Hasil cleaning .....	IV-5
Tabel IV-5. Hasil stopword removal.....	IV-6
Tabel IV-6. Hasil stemming .....	IV-6
Tabel IV-7. Hasil tokenizing.....	IV-7
Tabel IV-8. Hasil TF-IDF .....	IV-9
Tabel IV-9. Tabel Definisi Aktor.....	IV-21
Tabel IV-10. Tabel Definisi Use Case .....	IV-21
Tabel IV-11. Skenario Use Case Mengidentifikasi Pelecehan Seksual menggunakan Support Vector Machine.....	IV-22
Tabel IV-12. Tabel Penjelasan implementasi kelas .....	IV-26
Tabel IV-13. Rencana pengujian.....	IV-27
Tabel IV-14. Pengujian perangkat lunak .....	IV-29
Tabel V-1. Data Hasil Testing I .....	V-2
Tabel V-2. Tabel Confusion Matrix hasil pengujian .....	V-3
Tabel V-3. Tabel nilai hasil klasifikasi SVM.....	V-4

## DAFTAR GAMBAR

	Halaman
Gambar II-1. SVM model linier .....	II-4
Gambar II-2. Margin dan vektor pendukung dalam ruang 2 dimensi.....	II-4
Gambar II-3. Fase-fase dari RUP .....	II-9
Gambar III-1. Kerangka Kerja Penelitian .....	III-2
Gambar IV-1. Diagram <i>Use Case</i> Sistem Identifikasi Pelecehan Seksual .....	IV-20
Gambar IV-2. Diagram <i>Activity</i> .....	IV-24
Gambar IV-3. Diagram <i>sequence</i> .....	IV-25
Gambar IV-4. Diagram <i>class</i> perangkat lunak.....	IV-26
Gambar V-1. Grafik nilai hasil klasifikasi SVM .....	V-5

# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab pendahuluan ini menguraikan tentang pokok-pokok pikiran yang melandasi rencana penelitian. Pokok-pokok pikiran yang dimaksud antara lain latar belakang masalah penelitian, perumusan masalah/permasalahan penelitian, tujuan penelitian dan manfaat penelitian.

### **1.2 Latar Belakang Masalah**

Media sosial belakangan ini menjadi sesuatu yang tidak dapat lepas dari kehidupan masyarakat dikarenakan media sosial sekarang ini menjadi alat komunikasi, berekspresi, dan berpendapat dengan sesama pengguna media sosial. Media sosial adalah berbagai layanan yang berbasis internet dan seluler yang memungkinkan pengguna untuk berpartisipasi dalam pertukaran daring, berpartisipasi dalam konten yang dibuat oleh pengguna, atau bergabung dalam komunitas daring (Dewing, 2010). Hingga Januari 2022, menurut data dari Kepios, ada 191,4 juta pengguna media sosial di Indonesia, atau 68,9% keseluruhan penduduk Indonesia. Media sosial yang banyak digunakan di Indonesia yaitu Facebook, Youtube, Instagram, TikTok, dan Twitter. Dengan adanya media sosial, masyarakat pun banyak yang mengunggah pengalaman-pengalamannya, baik dalam bentuk tulisan ataupun multimedia. Reaksi dari dari unggahan tersebut pun beragam, ada yang merespon dengan baik berupa pujian, saran dan komentar

positif, ada juga yang merespon dengan buruk seperti memberikan celaan, hinaan bahkan kalimat yang melecehkan.

Berdasarkan Catatan Tahunan (CATAHU) yang diterbitkan Komnas Perempuan pada tahun 2022, terdapat 338.496 insiden kekerasan berbasis gender terhadap perempuan pada tahun 2021. Kasus pada tahun 2021 meningkat 50% dari tahun sebelumnya.

Salah satu jenis kekerasan berbasis gender yang umum adalah kekerasan siber. CATAHU yang diterbitkan oleh Komnas Perempuan pada tahun 2020 mengungkapkan bahwa terdapat 281 kasus kekerasan siber yang dilaporkan langsung ke Komnas Perempuan, meningkat 300% dibandingkan tahun sebelumnya.

Support Vector Machine merupakan salah satu algoritma *machine learning* yang sering digunakan untuk klasifikasi data. Seperti penelitian yang dilakukan oleh Tineges, Triayudi dan Sholihati(2020) yang melakukan analisis sentiment di Twitter. Penelitian ini menggunakan metode *Support Vector Machine* dan mendapatkan nilai akurasi sebesar 87%, nilai presisi sebesar 86%, nilai *recall* sebesar 95% dan nilai f1 sebesar 90%.

Instagram merupakan salah satu media sosial di mana penggunaanya bebas untuk berekspresi melalui gambar ataupun video, ataupun melalui kata-kata pada komentar dan pesan pribadi. Salah satu yang mungkin untuk dilakukan oleh pengguna adalah kekerasan siber. Oleh karena itu penelitian ini akan membuat program yang mengidentifikasi apakah suatu kalimat yang dikirim oleh pengguna



Instagram pada kolom komentar merupakan sebuah kalimat pelecehan seksual atau bukan menggunakan metode klasifikasi *Support Vector Machine*.

### **1.3 Rumusan Masalah**

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah seberapa akurat metode *Support Vector Machine* dalam mengidentifikasi kalimat pelecehan seksual di Instagram?

### **1.4 Tujuan Penelitian**

Berdasarkan permasalahan yang telah dirumuskan, tujuan dari penelitian ini adalah untuk mengetahui seberapa akurat metode *Support Vector Machine* dalam mengidentifikasi kalimat pelecehan seksual di Instagram.

### **1.5 Manfaat Penelitian**

Penelitian ini diharapkan dapat memberikan kontribusi untuk menambah wawasan berupa pemahaman dan pengetahuan tentang peran NLP dalam membantu mengurangi jumlah kasus pelecehan seksual di media sosial.

### **1.6 Batasan Masalah**

Penelitian ini memiliki batasan masalah sebagai berikut :

1. Objek penelitian merupakan teks berbahasa Indonesia.
2. Data untuk penelitian diambil secara manual di Instagram.

### **1.7 Sistematika Penulisan**

Sistematika penulisan skripsi ini adalah sebagai berikut:

## BAB I . PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan.

## BAB II. KAJIAN LITERATUR

Bab ini menjelaskan tentang dasar-dasar teori yang digunakan dalam penelitian, seperti *Support Vector Machine*, *Machine Learning*, pengolahan teks, vektorisasi.

## BAB III. METODOLOGI PENELITIAN

Bab ini menjelaskan langkah-langkah yang akan dilakukan dalam penelitian ini. Setiap rencana fase penelitian dirinci dengan mengacu pada kerangka kerja. Di akhir bab ini, perancangan manajemen proyek disertakan saat melakukan penelitian.

## BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan dijelaskan proses pengembangan perangkat lunak identifikasi pelecehan seksual pada Instagram menggunakan metode *Support Vector Machine*. Pengembangan perangkat lunak ini akan menggunakan metode *Rational Unified Process*.

## BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini akan dijelaskan mengenai hasil dari penelitian dalam mengidentifikasi kalimat pelecehan seksual pada komentar Instagram menggunakan metode *Support Vector Machine*.

## BAB VI. KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dari penelitian yang telah dilakukan sebelumnya dan juga berisi saran yang dapat dilakukan pada penelitian selanjutnya yang berhubungan dengan Identifikasi Pelecehan Seksual pada Instagram dengan metode Support Vector Machine.

### **1.8 Kesimpulan**

Penelitian ini akan melakukan identifikasi terhadap teks berisi pelecehan seksual di Instagram. Proses identifikasi akan menggunakan metode *Machine Learning* bernama *Support Vector Machine*. Hasil akurasi dari identifikasi menggunakan metode *Support Vector Machine* akan menjadi fokus dalam penelitian ini.

## DAFTAR PUSTAKA

- Alawneh, E., Al-Fawa'reh, M., Jafar, M. T., & Fayoumi, M. A. (2021). Sentiment analysis-based sexual harassment detection using machine learning techniques. *Proceeding - 2021 International Symposium on Electronics and Smart Devices: Intelligent Systems for Present and Future Challenges, ISESD 2021*.
- Barak, A. (2005, 3). Sexual harassment on the internet. *Social Science Computer Review*, 23(1), 77-92.
- Chowdhary, K. (2020). *Natural Language Processing*. New Delhi: Springer India.
- Dewing, M. (2010). *Social media: An introduction*.
- Kohavi, R. (1995). *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*.
- Kruchten, P. (2004). *The Rational Unified Process An Introduction* (Third ed.).
- Mutawalli, L., Taufan, M., Zaen, A., & Bagye, W. (2019). *KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto)*. Diambil kembali dari <http://e-journal.stmiklombok.ac.id/index.php/jire>
- Nugroho, A., Witarto, A., & Handoko, D. (2003). *Support Vector Machine-Teori dan Aplikasinya dalam Bioinformatika*. Diambil kembali dari <http://asnugroho.net>

Raudhoti, L., Herdiani, A., & Romadhony, A. (2020). *Identifikasi Cyberbullying pada Kolom Komentar Instagram dengan Metode Support Vector Machine dan Semantic Similarity (Cyberbullying Identification on Instagram Comment Using Support Vector Machine and Semantic Similarity )*. Diambil kembali dari <http://jcosine.if.unram.ac.id/>

Tineges, R., Triayudi, A., & Sholihati, I. D. (2020, 7 20). Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(3), 650. doi:10.30865/mib.v4i3.2181

Visa, S., Ramsay, B., Ralescu, A., & Knapp, E. v. (2011). Confusion Matrix-based Feature Selection. *Proceedings of the Twenty-second Midwest Artificial Intelligence and Cognitive Science Conference*, 120-127.

Zhang, W., Yoshida, T., & Tang, X. (2011). A comparative study of TF\*IDF, LSI and multi-words for text classification. *Expert Systems with Applications*, 2758-2765.