



**KLASIFIKASI SENTIMEN *TWITTER* MENGENAI PERSEPSI MASYARAKAT  
TERHADAP PRESIDENSI G20 MENGGUNAKAN *SUPPORT VECTOR MACHINE***

**SKRIPSI**

**YOHANES BILLY PUTERA**

**1810511105**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**2023**



**KLASIFIKASI SENTIMEN *TWITTER* MENGENAI PERSEPSI MASYARAKAT  
TERHADAP PRESIDENSI G20 MENGGUNAKAN *SUPPORT VECTOR MACHINE***

**SKRIPSI**

**Diajukan sebagai Salah Satu Syarat untuk Mendapatkan Gelar Sarjana Komputer**

**YOHANES BILLY PUTERA**

**1810511105**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**2023**

## LEMBAR PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Yohanes Billy Putera  
NIM : 1810511105  
Program Studi : Informatika  
Judul Tugas Akhir : Klasifikasi Sentimen Twitter mengenai Persepsi Masyarakat terhadap Presidensi G20 Menggunakan *Support Vector Machine*

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Dr. Widya Cholil, S.Kom., M.I.T.  
Penguji I



Henki Bayu Seta, S.Kom., MTI  
Penguji II



Yuni Widiastawi, S.Kom., M.Si  
Pembimbing I



Ika Nurlaili Isnainiyah, S.Kom., M.Sc.  
Pembimbing II



Ermatita, M.Kom.  
Dekan



Dr. Widya Cholil, S.Kom., M.I.T.  
Kepala Program Studi

Ditetapkan di : Jakarta  
Tanggal Ujian : 20 Januari 2023



Dipindai dengan CamScanner

## PERNYATAAN ORISINALITAS

Tugas skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Yohanes Billy Putera

NIM : 1810511105

Tanggal : 18 Januari 2023

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 18 Januari 2023



(Yohanes Billy Putera)

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN  
AKADEMIS**

Sebagai civitas akademika Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini :

Nama : Yohanes Billy Putera

NIM : 1810511105

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non Eksklusif(*Non-Exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

**KLASIFIKASI SENTIMEN *TWITTER* MENGENAI PERSEPSI MASYARAKAT  
TERHADAP PRESIDENSI G20 MENGGUNAKAN *SUPPORT VECTOR MACHINE***

Dengan hak bebas royalti ini, Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/ formatkan, mengelola dalam bentuk pangkalan kata (basis data), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik hak cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 18 Januari 2023



(Yohanes Billy Putera)

# KLASIFIKASI SENTIMEN *TWITTER* MENGENAI PERSEPSI MASYARAKAT TERHADAP PRESIDENSI G20 MENGGUNAKAN *SUPPORT VECTOR MACHINE*

YOHANES BILLY PUTERA

## ABSTRAK

Pada tahun ini, sesuai penetapan *Riyadh Summit 2020*, Indonesia terpilih untuk mengemban fungsi presidensi G20 tahun 2022. Adapun KTT G20 tahun 2022 nanti akan diselenggarakan pada 30-31 Oktober 2022 di Bali dengan Presiden Joko Widodo sebagai ketua konferensi. Hal ini tentunya menjadi suatu topik yang ramai diperbincangkan oleh masyarakat Indonesia, khususnya para pengguna media sosial. Masyarakat Indonesia pengguna media sosial sangat aktif dalam membagikan pandangan mereka mengenai suatu hal yang viral, seperti Presidensi G20 tahun 2022. Berbagai pendapat sering terlihat pada kolom komentar di berbagai platform media sosial, salah satunya pada *Twitter*. Komentar yang ada bukan hanya komentar yang bersifat positif, melainkan terdapat yang bersifat netral dan negatif juga. Komentar-komentar dalam *Twitter* tersebut dapat diolah menjadi informasi dengan cara tertentu, yaitu klasifikasi sentimen. Klasifikasi dilakukan menggunakan algoritma *Support Vector Machine*. Dari data yang digali dengan *Application Programming Interface (API)* dan praproses data, didapat *tweet* sebanyak 268 data yang akan dibagi dalam tiga rasio data latih dengan data uji, yaitu 60:40, 70:30, dan 80:20 yang dijadikan sebagai data untuk klasifikasi dengan algoritma *Support Vector Machine*. Hasil akurasi terbaik dari klasifikasi SVM dimiliki oleh model SVM dengan kernel linear untuk pembagian rasio 60:40 dan 70:30 (berturut-turut sebesar 77,78% dan 81,48%) dan kernel sigmoid untuk pembagian rasio 70:30 dan 80:20 (berturut-turut sebesar 81,48% dan 77,78%). Besarnya akurasi pada kernel linear dan sigmoid menunjukkan bahwa *classifier SVM* bisa dipakai dalam klasifikasi sentimen mengenai persepsi masyarakat terhadap Presidensi G20.

**Kata kunci** : *Support Vector Machine*, *tweet*, klasifikasi.

# **KLASIFIKASI SENTIMEN *TWITTER* MENGENAI PERSEPSI MASYARAKAT TERHADAP PRESIDENSI G20 MENGGUNAKAN *SUPPORT VECTOR MACHINE***

**YOHANES BILLY PUTERA**

## ***ABSTRACT***

*This year, according to the stipulation of the 2020 Riyadh Summit, Indonesia was chosen to carry out the functions of the G20 presidency in 2022. The 2022 G20 Summit will be held on 30-31 October 2022 in Bali with President Joko Widodo as chairman of the conference. This is of course a topic that is widely discussed by Indonesian people, especially social media users. Indonesians who use social media are very active in sharing their views on something that is viral, such as the G20 Presidency in 2022. Various opinions can often be seen in the comments column on various social media platforms, one of which is Twitter. The comments that are there are not only positive comments, but there are also neutral and negative ones. Comments on Twitter can be processed into information in a certain way, namely sentiment classification. The classification is performed using the Support Vector Machine algorithm. From the data extracted with the Application Programming Interface (API) and data pre-processing, a total of 268 tweets were obtained which would be divided into three ratios of training data to test data, namely 60:40, 70:30 and 80:20 which were used as data for classification with the Support Vector Machine algorithm. The best accuracy results from the SVM classification are owned by the SVM model with a linear kernel for distribution ratios of 60:40 and 70:30 (77.78% and 81.48% respectively) and sigmoid kernels for distribution ratios of 70:30 and 80: 20 (respectively of 81.48% and 77.78%). The high accuracy of the linear and sigmoid kernels indicates that the SVM classifier can be used to classify sentiments about people's perceptions of the G20 Presidency.*

**Keywords :** *Support Vector Machine, tweet, classification.*

## KATA PENGANTAR

Puji dan syukur dihaturkan kepada Tuhan Yang Maha Esa yang telah melimpahkan berkat dan rahmat-Nya sehingga penulis sanggup menyelesaikan skripsi berjudul “Klasifikasi Sentimen *Twitter* Mengenai Persepsi Masyarakat terhadap Presidensi G20 Menggunakan *Support Vector Machine*.” Penulis juga berterima kasih kepada :

1. Ibu Yuni Widiastiwi, S.Kom., M.Si. selaku dosen pembimbing I,
2. Ibu Ika Nurlaili Isnainiyah, S.Kom., M.Sc. selaku dosen pembimbing II,
3. Ibu Dr. Ermatita, M.Kom. selaku Dekan Fakultas Ilmu Komputer,
4. Ibu Dr. Widya Cholil, S.Kom., M.I.T. selaku Kepala Program Studi Informatika,
5. Ibu Iin Ernawati, S.Kom., M.Si. selaku Dosen Pembimbing Akademik,
6. Segenap dosen dan pegawai di Fakultas Ilmu Komputer,
7. Ghazy, Hasbi, Annisa, dan Anggitha selaku teman-teman Informatika Angkatan 2018,
8. Rian selaku teman Angkatan 2018 UKM PSM Gita Advayatva,
9. Keluarga penulis yang tidak pernah berhenti memotivasi penulis untuk menuntaskan skripsi.

Penulis juga menghaturkan terima kasih kepada segenap pihak yang membantu penulisan skripsi ini.

Jakarta, 18 Januari 2023



(Yohanes Billy Putera)



## DAFTAR ISI

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	iv
ABSTRAK.....	vi
KATA PENGANTAR .....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xii
BAB 1 PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Ruang Lingkup Penelitian.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Luaran yang Diharapkan .....	4
1.7 Sistematika Penulisan.....	4
BAB 2 TINJAUAN PUSTAKA .....	6
2.1 <i>Text Mining</i> (Penambangan Teks).....	6
2.2 <i>Text Preprocessing</i> .....	6
2.3 <i>Term Frequency – Inverse Document Frequency (TF-IDF)</i> .....	7
2.4 <i>Synthetic Minority Oversampling Technique (SMOTE)</i> .....	7
2.5 <i>Support Vector Machine (SVM)</i> .....	7
2.6 <i>Confussion Matrix</i> .....	10
2.7 Penelitian Terkait .....	12
BAB 3 METODE PENELITIAN .....	15
3.1 Tahapan Penelitian .....	15

3.1.3 <i>Text Mining</i> .....	16
3.1.4 <i>Preprocessing</i> .....	16
3.1.5 Pembobotan Kata.....	17
3.1.6 Pelatihan dan Pengujian Model .....	18
3.1.7 Evaluasi Performa Model .....	18
3.1.8 Dokumentasi.....	18
3.2 Perangkat Penelitian.....	18
3.2.1 Perangkat Keras.....	18
3.2.2 Perangkat Lunak.....	19
BAB 4 HASIL DAN PEMBAHASAN .....	20
4.1 <i>Text Mining</i> .....	20
4.2 <i>Preprocessing</i> .....	21
4.3 Pembobotan Kata .....	31
4.4 Klasifikasi.....	34
4.4.1 Pembagian Data Latih dan Data Uji .....	34
4.4.2 SMOTE ( <i>Synthetic Minority Oversampling Technique</i> ).....	35
4.4.3 Pengujian Model SVM .....	36
4.5 Evaluasi .....	42
BAB 5 KESIMPULAN DAN SARAN .....	44
5.1 Kesimpulan.....	44
5.2 Saran.....	44
DAFTAR PUSTAKA .....	45
RIWAYAT HIDUP .....	46
LAMPIRAN.....	47

## DAFTAR GAMBAR

Gambar 2. 1 Hyperplane memisah kedua kelas .....	8
Gambar 2. 2 Transformasi data ke klasifikasi non-linear .....	10
Gambar 3. 1 Tahapan Penelitian .....	15
Gambar 4. 1 Kode Program Crawling Data .....	20
Gambar 4. 2 Hasil Crawling Data pada file '.csv' .....	21

## DAFTAR TABEL

Tabel 2. 1 Confussion Matrix .....	11
Tabel 2. 2 Penelitian Terkait.....	12
Tabel 4. 1 Data duplikat dengan isi tweet yang sama.....	21
Tabel 4. 2 Data dengan isi tweet hampir sama .....	22
Tabel 4. 3 Data yang tidak konsisten .....	23
Tabel 4. 4 Tampilan Data Tweet.....	23
Tabel 4. 5 Contoh hasil tweet data labelling.....	25
Tabel 4. 6 Tampilan data dalam proses Cleansing .....	26
Tabel 4. 7 Tampilan data dalam proses Case Folding .....	27
Tabel 4. 8 Tampilan data dalam proses tokenisasi .....	28
Tabel 4. 9 Tampilan data dalam proses normalisasi .....	29
Tabel 4. 10 Tampilan data dalam proses filtering.....	30
Tabel 4. 11 Tampilan data diproses stemming .....	31
Tabel 4. 12 Sampel Dokumen TF-IDF .....	32
Tabel 4. 13 Contoh Implementasi TF-IDF .....	32
Tabel 4. 14 Jumlah pembagian data latih dan data uji tiap kelas dengan rasio 60:40 .....	34
Tabel 4. 15 Jumlah pembagian data latih dan data uji tiap kelas dengan rasio 70:30 .....	34
Tabel 4. 16 Jumlah pembagian data latih dan data uji tiap kelas dengan rasio 80:20 .....	35
Tabel 4. 17 Jumlah data latih setiap kelas sebelum dan setelah proses SMOTE.....	35
Tabel 4. 18 Confussion Matrix Klasifikasi SVM kernel linear (rasio pembagian 60:40) .....	36
Tabel 4. 19 Confussion Matrix Klasifikasi SVM kernel RBF (rasio pembagian 60:40).....	37
Tabel 4. 20 Confussion Matrix Klasifikasi SVM kernel sigmoid (rasio pembagian 60:40) ...	37
Tabel 4. 21 Confussion Matrix Klasifikasi SVM kernel Polinomial (rasio pembagian 60:40) .....	37
Tabel 4. 22 Tabel akurasi SVM pembagian rasio 60:40.....	38
Tabel 4. 23 Confussion Matrix Klasifikasi SVM kernel linear (rasio pembagian 70:30) .....	38
Tabel 4. 24 Confussion Matrix Klasifikasi SVM kernel RBF (rasio pembagian 70:30).....	39
Tabel 4. 25 Confussion Matrix Klasifikasi SVM kernel sigmoid (rasio pembagian 70:30) ...	39
Tabel 4. 26 Confussion Matrix Klasifikasi SVM kernel Polinomial (rasio pembagian 70:30) .....	40
Tabel 4. 27 Tabel akurasi SVM pembagian rasio 70:30.....	40
Tabel 4. 28 Confussion Matrix Klasifikasi SVM kernel linear (rasio pembagian 80:20) .....	40

Tabel 4. 29 Confussion Matrix Klasifikasi SVM kernel RBF (rasio pembagian 80:20).....	41
Tabel 4. 30 Confussion Matrix Klasifikasi SVM kernel sigmoid (rasio pembagian 80:20) ...	41
Tabel 4. 31 Confussion Matrix Klasifikasi SVM kernel Polinomial (rasio pembagian 80:20) .....	42
Tabel 4. 32 Tabel akurasi SVM pembagian rasio 80:20.....	42
Tabel 4. 33 Akurasi dari setiap kernel klasifikasi SVM .....	43