



***SENTIMENT ANALYSIS OPINI PELANTIKAN  
KABINET PEMERINTAH INDONESIA TAHUN 2019  
MENGUNAKAN VADER DAN RANDOM FOREST***

Skripsi

disusun sebagai salah satu syarat  
untuk memperoleh gelar Sarjana Komputer  
Program Studi Teknik Informatika

Oleh

Tanzilal Mustaqim  
4611416055

**JURUSAN ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI SEMARANG**

**2020**

## PERNYATAAN

Saya menyatakan bahwa skripsi saya yang berjudul “*Sentiment analysis* Opini Pelantikan Kabinet Pemerintah Indonesia Tahun 2019 Menggunakan *Vader* dan *Random forest*” disusun atas dasar penelitian saya dengan arahan dosen pembimbing. Sumber informasi atau kutipan yang berasal dari karya yang diterbitkan telah disebutkan dalam teks dan dicantumkan dalam daftar pustaka di bagian akhir skripsi ini. Dan saya menyatakan bahwa skripsi ini bebas plagiat dan apabila di kemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan perundang-undangan.

Semarang, 15 September 2020



Tanzilal Mustaqim  
4611416055

## PERSETUJUAN PEMBIMBING

Nama : Tanzilal Mustaqim  
NIM : 4611416055  
Program Studi : Teknik Informatika S1  
Judul Skripsi : *Sentiment Analysis* Opini Pelantikan Kabinet Pemerintah  
Indonesia Tahun 2019 Menggunakan *Vader* dan *Random  
forest*

Skripsi ini telah disetujui oleh pembimbing untuk diajukan ke sidang panitia ujian skripsi Program Studi Teknik Informatika FMIPA UNNES.

Semarang, 15 September 2020

Pembimbing



Much Aziz Muslim, S.Kom., M.Kom.  
NIP. 197404202008121001

## PENGESAHAN

Skripsi yang berjudul

*Sentiment Analysis* Opini Pelantikan Kabinet Pemerintah Indonesia Tahun  
2019 Menggunakan *Vader* dan *Random Forest*

disusun oleh

Tanzilal Mustaqim  
4611416055

Telah dipertahankan di hadapan sidang Panitia Ujian Skripsi FMIPA UNNES pada  
tanggal 17 September 2020.

Panitia:



Sekretaris



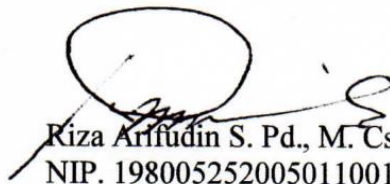
Dr. Alamsyah, S.Si., M.Kom.  
NIP. 197405172006041001

Penguji 1



Dr. Alamsyah, S.Si., M.Kom.  
NIP. 197405172006041001

Penguji 2



Riza Arifudin S. Pd., M. Cs.  
NIP. 198005252005011001

Anggota Penguji/  
Pembimbing



Much Aziz Muslim, S.Kom., M.Kom.  
NIP. 197404202008121001

## **MOTTO DAN PERSEMBAHAN**

### **MOTTO**

- Biarkanlah dirimu dibentuk oleh tarikan yang kuat dari sesuatu yang kamu cintai (Jalaluddin Rumi)
- Kalau ada hal yang mudah, kenapa harus mulai dari yang susah (Aprilia Dewi Ardiyanti)
- Fokus kepada satu hal kecil, niscaya suatu pekerjaan akan terlihat lebih mudah dikerjakan (Tanzilal Mustaqim)

### **PERSEMBAHAN**

Skripsi ini saya persembahkan kepada:

- Allah Subhanahu Wa Ta'ala, yang telah memberikan rahmat dan hidayah -Nya dalam menyelesaikan setiap masalah dalam proses pengerjaan skripsi dan masalah kehidupan.
- Ibu saya, Partini yang telah mencurahkan keringatnya untuk membiayai pendidikan saya, yang selalu memberikan kasih sayang, doa, dan dukungannya.
- Almarhum bapak saya yang telah memberikan pengertian tentang kehidupan dan proses pengajaran ilmu-ilmu kebermanfaatan manusia.
- kakak saya, Atina Khasanah dan Inna Ana Khasanah yang telah memberikan dukungan serta doa yang terus dipanjatkan.
- Sudara saya, Rafliq Ramadhan, Arrafin Kahfi, Zakiya Farhana, Fatih, Supriyono dan Mas Ayis yang telah memberikan dukungan moral dalam proses pengerjaan skripsi.

- Partner saya sesama agen neptunus dan yang menjadi penyedia berbagai macam suplai ilmu pengetahuan, percikan api inspirasi dan sumber semangat yakni Aprilia Dewi Ardiyanti.
- Teman-teman saya di jurusan Ilmu Komputer, Fakultas MIPA, serta teman-teman di Universitas Negeri Semarang.
- Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu hingga terselesaikannya penulisan skripsi ini.
- Almamater, Universitas Negeri Semarang.

## **PRAKATA**

Puji syukur penulis panjatkan kepada Allah *Subhanahu wa ta'ala* atas berkat rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi yang berjudul **“SENTIMENT ANALYSIS OPINI PELANTIKAN KABINET PEMERINTAH INDONESIA TAHUN 2019 MENGGUNAKAN VADER DAN RANDOM FOREST”**.

Penulis menyadari bahwa penulisan skripsi ini tidak akan selesai tanpa adanya dukungan serta bantuan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Prof. Dr. Fathur Rokhman, M.Hum., Rektor Universitas Negeri Semarang.
2. Dr. Sugianto M.Si., Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.
3. Dr. Alamsyah S.Si., M.Kom., Ketua Jurusan Ilmu Komputer FMIPA Universitas Negeri Semarang yang telah meluangkan waktu, membantu, membimbing, mengarahkan dan memberikan saran sehingga penulis dapat menyelesaikan skripsi ini.
4. Assoc Prof. Much Aziz Muslim, S.Kom., M.Kom., Dosen Pembimbing yang telah meluangkan waktu, membantu, membimbing, mengarahkan dan memberikan saran sehingga penulis dapat menyelesaikan skripsi ini.
5. Bapak dan Ibu Dosen Jurusan Ilmu Komputer yang telah memberikan bekal kepada penulis dalam penyusunan skripsi ini.

6. Ibu saya, Partini yang telah mencurahkan keringatnya untuk membiayai pendidikan saya, yang selalu memberikan kasih sayang, doa, dan dukungannya.
7. Almarhum bapak saya yang telah memberikan pengertian tentang kehidupan dan proses pengajaran ilmu-ilmu kebermanfaatan manusia.
8. kakak saya, Atina Khasanah dan Inna Ana Khasanah yang telah memberikan dukungan serta doa yang terus dipanjatkan.
9. Sudara saya, Rafliq Ramadhan, Arrafin Kahfi, Zakiya Farhana, Fatih, Supriyono dan Mas Ayis yang telah memberikan dukungan moral dalam proses pengerjaan skripsi.
10. Partner saya sesama agen neptunus dan yang menjadi penyedia berbagai macam suplai ilmu pengetahuan, percikan api inspirasi, kenyamanan dalam berjuang dan sumber semangat yakni Aprilia Dewi Ardiyanti.
11. Teman-teman seperjuangan dan satu bimbingan dari Assoc Prof. Much Aziz Muslim, S.Kom., M.Kom., Ryo Pambudi, Novi, Ilham Esa Tiffani, Anisa Falasari, Diah Alifia E, dan Sulistiana sebagai teman diskusi yang banyak memberikan dukungan dan motivasi.
12. Teman-teman saya di jurusan Ilmu Komputer, terutama teman-teman ilkom angkatan 2016 dan teman-teman kontrakan saya yang telah memberikan semangat dan dukungannya.
13. Teman-teman satu kontrakan Toples Space yang telah menemani selama masa kuliah dan memberikan banyak warna dalam proses perkuliahan dan pengerjaan skripsi.



14. Semua pihak yang telah membantu terselesaikannya skripsi ini yang tidak dapat penulis sebutkan satu persatu, terimakasih atas bantuannya.

Semoga skripsi ini dapat memberikan manfaat bagi pembaca di masa yang akan datang.

Semarang, 15 September 2020

Penulis



Tanzilal Mustaqim

4611416055

## ABSTRAK

Tanzilal Mustaqim. 2020. *Sentiment analysis* Opini Pelantikan Kabinet Pemerintah Indonesia Tahun 2019 Menggunakan *Vader* dan *Random forest*. Skripsi, Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Much Aziz Muslim, S.Kom., M.Kom.

Kata kunci: *Sentiment analysis, Vader, Random forest.*

Proses pelantikan kabinet Menteri baru di Indonesia tahun 2019 memantik banyak dukungan maupun penolakan serta banyak ungkapan opini masyarakat yang beragam. Salah satu media pengungkapan pendapat masyarakat yakni media sosial *twitter*. Media sosial *twitter* dapat memberikan kumpulan beragam data opini masyarakat sesuai dengan kebutuhan penelitian. Salah satu yang dapat dilakukan dengan banyaknya opini masyarakat yakni analisis *sentiment*. Pada penelitian ini dilakukan *sentiment analysis* terhadap opini masyarakat terkait pelantikan kabinet Menteri baru tahun 2019 menggunakan *vader* dan *random forest*. Analisa dilakukan pada 8223 dataset *tweet* yang diambil pada rentang waktu 1 Oktober 2019 sampai 31 Desember 2019 pada media sosial *twitter*. Tahapan dalam penelitian ini yakni data *collecting*, data *preprocessing*, data *labelling* dan data *modelling*. Data *collecting* yakni proses mengumpulkan data *tweet*. Data *preprocessing* yakni mempersiapkan data seperti melakukan proses pembersihan dan penormalan data. Data *labelling* yakni memberikan label *sentiment* terhadap masing-masing data *tweet* berdasarkan algoritma *vader*. Data *modelling* yakni melakukan uji akurasi *sentiment analysis* terhadap model penelitian menggunakan algoritma *machine learning Random forest* dibantu dengan *feature extraction TF-IDF* dan *Count vectorizer* serta *N-gram*. Hasil pengujian menunjukkan tingkat akurasi tertinggi diperoleh oleh kombinasi model algoritma dari *random forest, count vectorizer* dan *n-gram* berjumlah 3 gram mencapai 89,19%.

# DAFTAR ISI

	Halaman
PERNYATAAN .....	ii
PERSETUJUAN PEMBIMBING .....	iii
PENGESAHAN.....	iv
MOTTO DAN PERSEMBAHAN.....	v
PRAKATA .....	vii
ABSTRAK .....	x
DAFTAR ISI .....	xi
DAFTAR TABEL .....	xiv
DAFTAR GAMBAR .....	xv
DAFTAR LAMPIRAN.....	xvi
BAB 1 .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	4
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	5
1.6. Sistematika Penulisan .....	5
1.6.1 Bagian Awal Skripsi .....	5

1.6.2 Bagian Isi Skripsi.....	5
1.6.3 Bagian Akhir Skripsi.....	6
BAB 2 .....	7
BAB 3 .....	23
3.1. Studi Literatur .....	23
3.2. Analisis Metode.....	23
3.3. Analisis Hasil .....	28
3.3.1. Tahap pengambilan data.....	29
3.3.2. Tahap <i>preprocessing</i> .....	29
3.3.2.1. <i>Case folding</i> .....	30
3.3.2. <i>Data cleaning</i> .....	30
3.3.3. <i>Lemmatization</i> .....	31
3.3.4. <i>Remove stopwords</i> .....	31
3.3.5. Stemming .....	31
3.3.3. Tahap <i>Labelling</i> .....	32
3.3.4. Tahap <i>Modelling</i> .....	32
3.3.5. Tahap Implementasi .....	33
3.4. Penarikan Kesimpulan.....	33
BAB 4 .....	35
4.1. Hasil Penelitian .....	35

4.1.1. Hasil pengambilan data .....	35
4.1.2. Hasil <i>Preprocessing</i> .....	37
4.1.2.1 <i>Case folding</i> .....	37
4.1.2.2 <i>Data cleaning</i> .....	38
4.1.2.3 <i>Lemmatization</i> .....	39
4.1.2.4 <i>Remove stopwords</i> .....	40
4.1.2.5 <i>Stemming</i> .....	40
4.1.3. Hasil <i>Labelling</i> .....	41
4.1.1. Tahap <i>Modelling</i> .....	42
4.1.5. Hasil Implementasi Sistem .....	42
4.1. Beranda .....	43
4.2. <i>Collection</i> .....	43
4.3. <i>Cleaning</i> .....	44
4.4. <i>Labelling</i> .....	45
4.1.5.5. <i>Modelling</i> .....	46
4.2. Pembahasan .....	47
BAB 5 .....	51
5.1. Simpulan .....	51
5.2. Saran .....	51
DAFTAR PUSTAKA .....	52
LAMPIRAN .....	56

## DAFTAR TABEL

Tabel	Halaman
Tabel 4.1. Bagian Hasil Pengambilan Data <i>Tweet</i> 1.....	35
Tabel 4.2. Bagian Hasil Pengambilan Data <i>Tweet</i> 2.....	35
Tabel 4.3. Bagian Hasil Pengambilan Data <i>Tweet</i> 3.....	36
Tabel 4.5. Tabel Hasil <i>Data cleaning</i> .....	38
Tabel 4.6. Hasil <i>Lemmatization</i> .....	39
Tabel 4.1.2.4. Hasil <i>Remove stopwords</i> .....	40
Tabel 4.1.2.5 Hasil <i>Stemming</i> .....	40
Tabel 4.9. Hasil <i>Labelling</i> .....	41
Tabel 4.10. Tabel Hasil Kumulatif <i>Sentiment</i> .....	42
Tabel 4.11. Hasil <i>Modelling</i> .....	42
Tabel 4.12. Hasil <i>modelling</i> kombinasi algoritma.....	48
Tabel 4.13. Tabel Perbandingan .....	49

## DAFTAR GAMBAR

Gambar	Halaman
Gambar 2.2.1. Sistem Arsitektur <i>Text mining</i> (Feldman dan Sanger, 2007) .....	11
Gambar 3.2.1. <i>Flowchart Data Collection</i> .....	24
Gambar 3.2. <i>Flowchart Data Pre-processing</i> .....	25
Gambar 3.3. <i>Flowchart Data Labelling</i> .....	26
Gambar 3.4. <i>Flowchart Data Modelling</i> .....	27
Gambar 3.3. Tahap Penelitian .....	28
Gambar 4.1. Tampilan <i>Graphical user interface Beranda</i> .....	43
Gambar 4.2. Tampilan <i>Graphical user interface Collection</i> .....	44
Gambar 4.3. Tampilan <i>Graphical user interface Cleaning</i> .....	45
Gambar 4.4. Tampilan <i>Graphical user interface Labelling</i> .....	46
Gambar 4.4. Tampilan <i>Graphical user interface Modelling</i> .....	47

## DAFTAR LAMPIRAN

Lampiran	Halaman
Lampiran 1	
Dataset yang berhasil di kumpulkan dari <i>Twitter</i> .....	57
Lampiran 2	
Desain Tampilan Sistem .....	62
Lampiran 3	
<i>Source code</i> Algoritma Penelitian .....	64
Lampiran 4	
<i>Source code</i> Tampilan <i>Graphical user interface</i> .....	72



# **BAB 1**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Aktivitas politik di Indonesia membuat dinamika kehidupan masyarakat menjadi lebih dinamis terutama dari berbagai macam opini yang muncul ditujukan kepada pemerintah (Pangestu *et al.*, 2019). Dunia politik di Indonesia pada tahun 2019 sedang dalam masa penyusunan kabinet. Kabinet yang terdiri dari para menteri, tugas utamanya adalah untuk membantu tugas presiden dalam mengatur dan mengkondisikan kegiatan bernegara dalam segala lini mulai dari segi sosial, pendidikan, wisata, ekonomi dan keamanan (Ulya, 2016). Masyarakat mempunyai harapan yang besar kepada pemerintah untuk mampu merubah taraf kehidupan sebelumnya menjadi lebih baik dan mampu membuat berkembang lingkungan sekitar. Pola kerja pemerintah dalam membentuk kabinet baru secara langsung memantik dukungan maupun penolakan dari masyarakat yang diwakilinya (Furqon *et al.*, 2018).

Dukungan dari masyarakat dapat berupa opini positif dan saran bagi para menteri yang akan dilantik. Dukungan positif masyarakat adalah hasil dari buah kerja para menteri dalam periode waktu sebelumnya atau karya yang dihasilkan untuk membangun negara Indonesia (Moento *et al.*, 2019), selain dukungan positif terdapat pula penolakan atau sanggahan dari kalangan masyarakat tertentu. Penolakan dari masyarakat dihasilkan dari pola kerja buruk dari periode

sebelumnya atau perilaku yang cenderung membuat masyarakat tidak puas apabila pejabat tersebut dilantik (Dewi, 2019).

Opini masyarakat baik dari dukungan maupun penolakan dapat disampaikan kepada *public* dengan tujuan sampai kepada pihak terkait di pemerintahan dengan menggunakan media sosial (Izzati *et al.*, 2018). Media sosial yang sering digunakan masyarakat Indonesia yakni media sosial dalam bentuk digital. Ada beberapa jenis media sosial digital populer seperti *twitter*, *instagram*, *whatsapp* dan lain-lain (Sholihin *et al.*, 2018). *Twitter* adalah salah satu media sosial populer yang sering digunakan masyarakat Indonesia untuk mengungkapkan berbagai macam opini terkait kehidupan pribadi, sosial maupun terkait politik. *Twitter* dapat digunakan sebagai media penelitian yang bertujuan untuk menganalisis data-data tersembunyi menggunakan *text mining* karena penyampaian opini pengguna yang terbatas 140 karakter sehingga memudahkan proses ekstraksi data (Yang & Zhang, 2018). *Text mining* dalam *twitter* melalui beberapa proses yakni *text preprocessing*, *labelling* dan *modelling* (Pinto & Murari, 2019).

*Text preprocessing* terdiri dari *data collection* dan *data cleaning*. *Data collection* adalah proses pengumpulan data dari *twitter* secara *scraping* atau secara langsung dipindahkan ke *local harddrive* untuk keperluan penelitian (Fitriasih & Kusumaningrum, 2019). *Data cleaning* adalah proses pembersihan dan normalisasi data dari beberapa hal yang tidak berguna seperti simbol tanda baca dan karakter yang tidak mempunyai kontribusi dalam proses analisis data (Indriani & Muslim, 2019).

Proses *sentiment analysis* dalam *twitter* sering dilakukan untuk mengetahui opini positif, netral atau negatif dari masyarakat terkait suatu hal tertentu. Contoh

penelitian *sentiment analysis* di *twitter* yakni *sentiment analysis* tentang kinerja gubernur DKI Jakarta tahun 2017-2018 lalu dan *sentiment analysis* tentang kepuasan konsumen (Julian *et al.*, 2019; Kumar *et al.*, 2019).

*Vader lexicon polarity detection* adalah salah satu contoh dari metode *sentiment analysis* secara leksikal (Hutto & Gilbert, 2014). Metode leksikal adalah metode yang tidak memerlukan *data training* atau data yang telah dilabeli namun sudah tersedia dalam kamus lengkap dengan kepolaritasan sentimennya (Garay *et al.*, 2019). *Vader sentiment polarity detection* digunakan untuk proses *sentiment analysis* karena proses penentuan *sentiment* yang mudah bersumber dari *dictionary* yang tersedia sehingga bisa mendekati penilaian manusia (Garay *et al.*, 2019). *Vader* digunakan pada penelitian ini untuk memberikan label pada data *tweet*. Tujuan dari *labelling* yakni mempersiapkan data *tweet* sebagai data *training* sebelum masuk ke proses *modelling* menggunakan *random forest* (Davidson *et al.*, 2017).

Proses *modelling* yakni proses uji akurasi *sentiment analysis* menggunakan algoritma *machine learning random forest*. Proses pengujian ditambah dengan penggunaan *feature extraction TF-IDF*, *count vectorizer* dan *N-gram* untuk meningkatkan tingkat akurasi dari proses pendeteksian. *N-gram* yang digunakan berkisar 1 sampai 3 gram kata. Hasil akhir dari *modelling* berupa tingkat akurasi dari data *input tweet* yang telah dilabeli sebelumnya menggunakan *vader*.

Analisis data *twitter* dengan memanfaatkan *vader* dan *random forest* dapat berguna untuk mengambil data penting lengkap dengan polaritas sentimen emosi dari data *twitter* dan tingkat akurasi yang dihasilkan. Penelitian ini bermaksud melakukan proses analisis data *twitter* untuk mengetahui polaritas sentimen emosi

dan tingkat akurasi dari opini masyarakat terkait dengan pelantikan kabinet baru Indonesia tahun 2019 dan mengambil judul “*SENTIMENT ANALYSIS OPINI PELANTIKAN KABINET PEMERINTAH INDONESIA TAHUN 2019 MENGGUNAKAN VADER DAN RANDOM FOREST*”.

## **1.2. Rumusan Masalah**

Berdasarkan uraian latar belakang diatas, maka rumusan masalah dalam penelitian ini adalah bagaimana proses *labelling sentiment* dan *modelling* terhadap opini masyarakat terkait pelantikan kabinet baru pemerintah Indonesia tahun 2019 dengan menggunakan *Vader* dan *Random forest*?

## **1.3. Batasan Masalah**

Pada penelitian ini diperlukan batasan-batasan agar tujuan penelitian dapat tercapai. Adapun batasan masalah yang dibahas pada penelitian ini adalah:

- a. Data yang digunakan dalam penelitian ini adalah data hasil *data collection* dari *Twitter* menggunakan *library GetOldTweet3* bahasa pemrograman *python* dengan kata kunci "menteri", "pelantikan", "kabinet".
- b. Proses *labelling sentiment tweet* menggunakan kamus *lexicon* dari *Vader*.
- c. Proses *modelling* untuk uji akurasi *sentiment* menggunakan algoritma *machine learning Random forest*.

## **1.4. Tujuan Penelitian**

Tujuan penelitian ini adalah untuk memberikan *label sentiment* pada data *tweet* menggunakan *Vader* dan melakukan uji akurasi menggunakan algoritma

*machine learning Random forest* dari opini masyarakat terkait dengan pelantikan kabinet baru pemerintah Indonesia tahun 2019.

### **1.5. Manfaat Penelitian**

Manfaat penelitian ini adalah sebagai berikut:

1. Dapat mengetahui performa akurasi *Random forest* pada data *tweet* yang telah dilabeli menggunakan *Vader*.
2. Dapat memberikan wawasan bagi penulis dan pengetahuan bagi dunia pendidikan terkait performa akurasi *Random forest* pada data *tweet* yang telah dilabeli menggunakan *Vader*

### **1.6. Sistematika Penulisan**

Sistematika penulisan berguna untuk memudahkan dalam memahami jalan pemikiran secara keseluruhan skripsi. Penulisan skripsi ini secara garis besar dibagi menjadi tiga bagian, yaitu sebagai berikut.

#### **1.6.1 Bagian Awal Skripsi**

Bagian awal skripsi terdiri dari halaman judul, halaman pengesahan, halaman pernyataan, halaman motto dan persembahan, abstrak, kata pengantar, daftar isi, daftar Gambar, daftar table, dan daftar lampiran.

#### **1.6.2 Bagian Isi Skripsi**

Bagian isi skripsi terdiri dari lima bab, yaitu sebagai berikut.

##### **a. BAB 1: PENDAHULUAN**

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

b. **BAB 2: TINJAUAN PUSTAKA**

Bab ini berisi penjelasan mengenai definisi atau landasan teori maupun pemikiran-pemikiran yang dijadikan kerangka teoritis yang menyangkut masalah dalam skripsi ini serta penelitian terkait.

c. **BAB 3: METODE PENELITIAN**

Bab ini berisi penjelasan mengenai studi literatur, identifikasi alat, eksperimen, analisis hasil dan penarikan kesimpulan.

d. **BAB 4: HASIL DAN PEMBAHASAN**

Bab ini berisi hasil penelitian berserta pembahasannya.

e. **BAB 5: PENUTUP**

Bab ini berisi simpulan dari penulisan skripsi dan saran yang diberikan penulis untuk mengembangkan skripsi ini.

**1.6.3 Bagian Akhir Skripsi**

Bagian akhir skripsi ini berisi daftar pustaka yang merupakan informasi mengenai buku-buku, sumber-sumber dan referensi yang digunakan penulis serta lampiran-lampiran yang mendukung dalam penulisan skripsi

## BAB 2

### TINJAUAN PUSTAKA

#### 2.1. Tinjauan Pustaka

Penelitian ini dikembangkan dari beberapa referensi yang mempunyai keterkaitan dengan metode dan objek penelitian (Irza et al., 2017; Zhang & Zhu, 2014). Penggunaan referensi ini ditujukan untuk memberikan batasan-batasan terhadap metode dan sistem yang nantinya akan dikembangkan lebih lanjut. Berikut adalah beberapa penelitian yang terkait dengan penelitian yang diusulkan.

Pinto & Murari, (2019) dalam penelitiannya yang berjudul “*Real Time Sentiment analysis of Political Twitter Data Using Machine learning Approach*” berhasil melakukan proses *sentiment analysis* menggunakan *lexicon based* yakni *vader sentiment analysis polarity detection*. Penelitiannya mengambil data dari *twitter* dan kemudian melakukan proses analisis polaritas sentimen setiap *tweet* yang telah dipersiapkan sebelumnya melalui *text preprocessing*. Penelitian ini menggunakan *feature selection* berupa *TF-IDF* dan algoritma *machine learning support vector machine*. Hasil penelitian ini adalah mendapatkan akurasi sebesar 82% dari 800 *tweet* dengan rincian 760 diprediksi negative dan 40 lainnya positif sesuai dengan label dari *vader*.

Parmar et al., (2014) melakukan penelitian *sentiment analysis* terhadap *dataset review film* menggunakan *random forest*. Judul penelitian mereka adalah “*Sentiment Mining of Movie Reviews using Random forest with Tuned Hyperparameters*”. Hasil penelitian berupa tingkat akurasi yang didapatkan

mencapai 91%. Metode yang digunakan yakni algoritma *random forest* yang di ubah pada hyperparameter-nya.

Al Amrani et al., (2018) dalam penelitiannya yang berjudul “*Random forest and support vector machine based hybrid approach to sentiment analysis*” berhasil melakukan proses *sentiment analysis* pada *dataset* review produk di Amazon. Penelitian mereka menggunakan perpaduan support vector machine dan *random forest* yang menghasilkan akurasi senilai 83,4%.

Wan & Gao, 2016 melakukan penelitian *sentiment analysis* terhadap pelayanan maskapai penerbangan dengan *dataset* diambil dari media sosial *twitter*. Judul dalam penelitian ini adalah “*An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis*” dengan hasil berupa akurasi dari metode *random forest* mencapai 82.4%.

## **2.2. Landasan Teori**

### **2.2.1 Text Mining**

*Text mining* adalah proses penambangan atau proses ekstraksi data penting tersembunyi dari kumpulan teks (Ramanathan & Meyyappan, 2019). Data tersembunyi dari teks biasa disebut dengan pengetahuan implisit. Pengetahuan implisit tidak ditulis secara gamblang dalam teks dan cenderung perlu teknik pemahaman manusia dan algoritma komputer tertentu untuk mengetahuinya (Thu & New, 2017). Proses ekstraksi pengetahuan implisit secara tradisional memerlukan proses pemahaman yang lama dan panjang seperti membaca dan memahami dengan seksama.



Text mining adalah bagian dari data mining. Data mining bekerja secara luas untuk ekstraksi informasi dari banyak macam jenis data (Safri et al., 2018). Jenis data mining yang lain yakni relational data mining, web mining dan big data mining. Text mining sendiri jenis data mining yang khusus bekerja untuk menganalisis teks. Penggunaan text mining untuk penelitian telah digunakan secara luas menganalisis berbagai macam bentuk text seperti sentiment analysis yang digunakan dalam penilaian kinerja gubernur DKI Jakarta tahun 2017-2018 dan sentiment analysis tentang kepuasan konsumen (Julian et al., 2019; Kumar et al., 2019).

*Text mining* dapat diartikan sebagai penemuan informasi yang baru dan tidak diketahui sebelumnya oleh komputer, dengan secara otomatis mengekstrak informasi dari sumber-sumber yang berbeda. Kunci dari proses ini adalah menggabungkan informasi yang berhasil diekstraksi dari berbagai sumber (Hearst, 2003). Sedangkan menurut (Harlian, 2006) *text mining* memiliki definisi menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. *Text mining* mengacu pada proses mengambil informasi berkualitas tinggi dari teks.

Informasi berkualitas tinggi biasanya diperoleh melalui peramalan pola dan kecenderungan melalui sarana seperti pembelajaran pola statistik. *Text mining* biasanya melibatkan proses penataan teks *input* (biasanya parsing, bersama dengan penambahan beberapa fitur linguistik turunan dan penghilangan beberapa diantaranya, dan penyisipan subsequent ke dalam database), menentukan pola

dalam data terstruktur, dan akhirnya mengevaluasi dan menginterpretasi output. Berkualitas tinggi di bidang *text mining* biasanya mengacu ke beberapa kombinasi relevansi, kebaruan, dan interestingness.

Proses *text mining* yang khas meliputi kategorisasi teks, *text clustering*, ekstraksi konsep/entitas, produksi taksonomi granular, *sentiment analysis*, penyimpulan dokumen, dan pemodelan relasi entitas (yaitu, pembelajaran hubungan antara entitas bernama). Pendekatan manual *text mining* secara intensif dalam laboratorium pertama muncul pada pertengahan 1980-an, namun kemajuan teknologi telah memungkinkan ranah tersebut untuk berkembang selama dekade terakhir.

*Text mining* adalah bidang interdisipliner yang mengacu pada pencarian informasi, pertambangan data, pembelajaran mesin, statistik, dan komputasi linguistik. Dikarenakan kebanyakan informasi (perkiraan umum mengatakan lebih dari 80%) saat ini disimpan sebagai teks, *text mining* diyakini memiliki potensi nilai komersial tinggi (Bridge, 2011).

Menurut Feldman dan Sanger (Feldman dan Sanger, 2007), *text mining* dapat didefinisikan secara luas sebagai proses pengetahuan intensif yang memungkinkan pengguna berinteraksi dengan koleksi dokumen dari waktu ke waktu menggunakan berbagai macam analisis. Dalam cara yang sejalan dengan *data mining*, *text mining* berusaha mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi patterns. *Text mining* menjadi menarik karena sumber data koleksi dokumen dan pola yang menarik tidak

ditemukan dari database formal namun ditemukan dalam data tekstual yang tidak terstruktur pada kumpulan dokumen.

*Text mining* dapat didefinisikan secara luas sebagai proses pengetahuan intensif di mana pengguna berinteraksi dengan koleksi dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis. *Text mining* berusaha untuk mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi pola yang menarik. *Text mining* banyak mengarah pada bidang penelitian *data mining*. Oleh karena itu, tidak mengherankan bahwa *text mining* dan *data mining* akan berada pada tingkat arsitektur yang sama (Feldman, dkk. 2007)

Berikut Gambaran sistem arsitektur *text mining* yang dicantumkan pada buku (Feldman, dkk. 2007) Gambar 2.1.



Gambar 2.1. Sistem Arsitektur *Text mining* (Feldman dan Sanger, 2007)

Selain itu, Feldman dan Sanger (Feldman dan Sanger, 2007) juga berpendapat bahwa *text mining* juga merupakan bidang baru dalam cabang ilmu komputer yang berupaya untuk mengatasi krisis informasi yang berlebihan dengan cara menggabungkan beberapa teknik dari *data mining*, mesin pembelajaran (*machine learning*), pengolahan bahasa alami (*natural language processing*),

*information retrieval* dan pengelolaan ilmu pengetahuan (*knowledge management*).

Franke dalam Langgeni dkk. (Langgeni dkk., 2010) menjelaskan bahwa *text mining* didefinisikan sebagai menambang data berupa teks yang bersumber dari dokumen. *Text mining* bertujuan untuk mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen. *Text mining* juga dapat diartikan sebagai sebuah proses untuk menemukan suatu informasi atau tren baru yang sebelumnya tidak terungkap dengan memroses dan menganalisis data dalam jumlah besar (Feldman dan Sanger, 2007).

Sebelumnya, *website* paling sering menggunakan pencarian berbasis teks, yang hanya menemukan dokumen yang berisi kata-kata atau frase spesifik yang ditentukan oleh pengguna. Sekarang, melalui penggunaan web semantik, *text mining* dapat menemukan konten berdasarkan makna dan konteks (daripada hanya dengan kata tertentu). *Text mining* juga digunakan dalam beberapa filter email spam sebagai cara untuk menentukan karakteristik pesan yang mungkin berupa iklan atau materi yang tidak diinginkan lainnya.

Penelitian dibidang *text mining* menangani masalah yang berkaitan dengan representasi teks, klasifikasi, clustering, ekstraksi informasi atau pencarian dan pemodelan pola. Dalam hal ini pemilihan karakteristik, juga domain penelitian dan prosedur penelitian menjadi peran penting. Oleh karena itu, adaptasi dari algoritma *data mining* dari teks yang diketahui sangat diperlukan. Maka dari itu untuk mencapai hal ini seringkali berdasarkan penelitian sebelumnya *text mining* bergantung pada *information retrieval*, *natural language processing* dan

*information extraction*. Selain itu juga penerapan metode *data mining* dan statistik juga diterapkan untuk menangani masalah ini (Hotho, 2005).

*Information retrieval* (IR) adalah menemukan bahan (biasanya dokumen) dari suatu keadaan yang tidak terstruktur (biasanya teks) yang memenuhi kebutuhan informasi dari dalam kumpulan data yang besar (biasanya disimpan didalam komputer) (Manning, dkk. 2009). *Natural language processing* (NLP) bertujuan untuk mencapai hasil yang lebih baik dalam pemahaman bahasa alami dengan menggunakan komputer. Sedangkan Ekstraksi Informasi (IE). Bertujuan untuk menemukan informasi tertentu dari dokumen teks yang kemudian Ini disimpan dalam basis data seperti pola sehingga dapat digunakan dan dimanfaatkan (Hotho, 2005).

Hotho (2005). juga mengatakan bahwa pada penelitian *text mining* diperlukan tahapan *text preprocessing* pada koleksi dokumen dan menyimpan informasi tersebut dalam struktur data. Pendekatan *text mining* didasarkan pada pemikiran bahwa dokumen teks dapat diwakili oleh satu set kata-kata, yaitu dokumen teks digambarkan berdasarkan pada set kata-kata yang terkandung di dalamnya.

### **2.2.2 Data Collection**

*Data collection* adalah proses pengumpulan data yang digunakan untuk penelitian (Kotu & Deshpande, 2019). Data penelitian diperoleh melalui media sosial *twitter*. Metode pengambilan data dari *twitter* secara *real time* yakni menggunakan *twitter streaming API* yang memberikan akses untuk transfer data sesuai permintaan dengan syarat tertentu (Gabarron *et al.*, 2019). *Twitter* memiliki

banyak API seperti *public streaming API* dan *REST API* yang memudahkan proses pengambilan data menjadi mudah.

Penelitian ini menggunakan *twitter API* dari paket program *GetOldTweet3* yang merupakan paket program tambahan pada bahasa *python* khusus untuk menangani pengolahan data *twitter*. Paket program *GetOldTweet3* mendukung akses *twitter* melalui *basic authentication* dan yang terbaru yakni *OAuth method* (Singh *et al.*, 2016). Metode yang digunakan *GetOldTweet3* adalah *OAuth method* dan untuk metode *basic authentication* sudah diberhentikan. Untuk mulai menggunakan layanan *GetOldTweet3* dan memulai proses analisis, diharuskan untuk mendaftarkan aplikasi klien dengan *twitter* untuk mendapatkan akses API. Setelah mendaftar kita akan mendapatkan *consumer token*, *consumer secret key*, *access token* dan akses *secret key* yang semuanya digunakan dalam paket program *GetOldTweet3* (Al-sultany & Aleqabie, 2019).

### **2.2.3 Text preprocessing**

Tahap *text preprocessing* merupakan tahap awal dari *text mining*. *Text preprocessing* merupakan proses menggali, mengolah dan mengatur informasi dengan cara menganalisis hubungannya dengan aturan-aturan yang ada di data tekstual semi terstruktur atau tidak terstruktur (Luhulima, Marji, dan Muflikhah, 2013). Untuk lebih efektif dalam proses *text preprocessing*, dilakukan langkah transformasi data ke dalam suatu format yang memudahkan untuk kebutuhan pemakai. Proses ini disebut *text preprocessing*. Setelah dalam bentuk yang lebih terstruktur dengan adanya proses di atas, data dapat dijadikan sumber data yang

dapat diolah lebih lanjut. *Text preprocessing* adalah proses pembersihan dan penormalan data sebelum masuk ke proses data *analysis*.

#### **2.2.4 Case folding**

*Case folding* adalah metode untuk mengubah semua huruf dalam *dataset* menjadi kapital atau kecil semua (Romadloni et al., 2019). Hal ini dilakukan untuk memudahkan proses analisis *dataset* dan mengurangi jumlah penggunaan memori. Contoh *case folding* yakni mengubah kalimat “Pemerintah Indonesia akan Maju tahun 2019 ini!” menjadi “pemerintah indonesia akan maju tahun 2019 ini!”. *Case folding* membantu proses *lemmatization* dan *stemming* untuk mencari kecocokan tiap data pada *dictionary*.

#### **2.2.5 Lemmatization**

Kata-kata tidak baku seringkali digunakan dalam berkomunikasi dan berinteraksi dengan orang lain (Ravi et al., 2019). Kata-kata tidak baku banyak terbentuk hasil dari interaksi manusia itu sendiri dan kadang jauh dari aturan baku dan kamus bahasa aslinya. Dalam *sentiment analysis* kata-kata tidak baku sangat berpengaruh terhadap hasil kalkulasi analisis data. Untuk meningkatkan hasil kualitas sentimen diperlukan perbaikan atau perubahan kata-kata tidak baku menjadi kata baku. *Lemmatization* adalah proses mengubah kata-kata tidak baku menjadi bahasa aslinya. Contoh lemmatization yakni dari kalimat “Aku sdng tdk di kantor hari ini”, dari kalimat tersebut ditemukan 3 kata tidak baku yakni “Aku”, “sdng” dan “tdk”. Dalam proses lemmatization kata-kata tidak baku yang ditemukan tadi akan diubah menjadi “saya”, “sedang” dan “tidak”.

### 2.2.6 *Remove stopwords*

Dokumen teks di dalamnya terdapat kata-kata yang tidak terlalu berguna seperti kata depan, kata sambung, kata sifat, kata slank, kata ganti dan masih banyak lagi (Rahutomo et al., 2019). Kata-kata tersebut biasanya sering muncul bersamaan dengan kata utama sehingga menjadikannya tidak unik dan tidak memiliki sebuah arti tertentu. Daftar kata yang tidak terlalu berkontribusi dalam teks analisis biasa disebut dengan *stopwords* atau *stoplist*. *Stopwords* tidak mempunyai potensi untuk dijadikan indeks dokumen. *Stoplist* bersifat unik karena tiap bahasa mempunyai *stoplist* masing-masing. Dengan menghapus *stopwords* dalam *dataset* dapat meningkatkan kualitas *sentiment analysis*.

### 2.2.7 **Stemming**

*Stemming* adalah proses penghapusan atribut tambahan pada kata seperti penghapusan “me-“ dan “-kan” dari “membuatkan” menjadi “buat” (Julian et al., 2019). Secara sederhana *stemming* adalah membuat kata berimbuhan menjadi kata dasarnya. *Stemming* biasanya digunakan untuk sistem pengambilan informasi seperti mesin pencari dan teks analisis lainnya. Algoritma *stemming* biasanya berbasis aturan tertentu dan merupakan proses heuristik untuk memotong atau menghapus karakter tambahan di ujung kata. Proses penghapusannya dimulai dengan melihat sebuah kata kemudian apabila memenuhi kriteria algoritma tambahan kata tersebut akan dihapus. Dalam *stemming* diketahui ada 2 masalah yang sering muncul yakni *over-stemming* dan *under-stemming*. *Over-stemming* terjadi karena penghapusan kata yang berlebihan kemudian *under-stemming* terjadi karena proses penghapusan yang tidak sesuai sehingga mengakibatkan arti kata



berubah. Dalam penelitian ini proses *stemming* dioptimalkan dengan bantuan paket program Sastrawi khusus untuk analisa bahasa Indonesia.

### **2.2.8 Vader**

*Vader* adalah akronim untuk *Valence Aware Dictionary for Social Reasoning* yang digunakan sebagai model untuk analisis sentimen dan mampu menentukan keragaman data melalui intensitas kekuatan emosional yang ada sesuai dengan kamus data *Lexicon* yang tersedia (Elbagir & Yang, 2019). *Vader* diperkenalkan pada tahun 2014 oleh C.J Hutto dan Eric Gilbert yang metode pembentukannya didasarkan pada pendekatan *human-centric*, menggabungkan analisis kualitatif dan validasi empiris menggunakan kebijaksanaan dan penilaian manusia (Hutto & Gilbert, 2014). *Vader* mampu memberikan polaritas yang berbeda antara "Aku suka kamu" dan "Aku tidak suka kamu".

Penilaian polaritas menggabungkan fitur kamus leksikal dengan skor sentimen dari 5 kriteria tambahan yaitu tanda seru, huruf besar, tingkat susunan kata, pergeseran polaritas karena kata "tapi" dan menggunakan fitur *tri-gram* untuk memeriksa keberadaan negasi (Alaei et al., 2019). Pendekatan leksikal bertujuan untuk memetakan kata menjadi sentimen dengan membangun leksikon atau 'kamus sentimen.'

Kamus *lexicon* dapat digunakan untuk menilai sentimen frasa dan kalimat, tanpa perlu melihat yang lain. Sentimen dapat dikategorikan - seperti {negatif, netral, positif} - atau dapat numerik - seperti kisaran intensitas atau skor. Pendekatan leksikal melihat kategori sentimen atau skor setiap kata dalam sebuah kalimat dan memutuskan kategori atau skor sentimen keseluruhan kalimat itu (Saif

et al., 2016). Kekuatan dari pendekatan leksikal terletak pada kenyataan bahwa tidak diperlukan melatih model menggunakan data berlabel.

*Vader* adalah contoh dari metode leksikal. Keuntungan dari menggunakan *vader* polarity detection adalah sudah tersedia kamus yang berisi nilai dari setiap kata. Proses penentuan polaritas kalimat didapatkan dari penyatuan attribute “compound” dari setiap kata yang tersedia (Ghiassi & Lee, 2018). Kriteria pengelompokan positif, netral dan negative yakni jika hasil compound lebih dari 0,05 maka dimasukkan kategori positif yang diwakilkan dengan angka 1 lalu jika hasil compound terletak diantara -0,05 dan 0,05 maka termasuk kategori netral yang diwakilkan dengan angka 0 dan yang terakhir jika hasil compound dibawah -0,05 maka termasuk kategori negative yang diwakilkan dengan angka -1.

### **2.2.9 Random forest**

*Random forest* dimulai dengan teknik penambangan data dasar, *decision tree*. Di *decision tree*, *input* dimasukkan di bagian atas (root) lalu turun ke bawah (daun) untuk menentukan data, termasuk jenis urutan kelas (Liaw & Wiener, 2002). *Random forest* adalah penggolong yang terdiri dari kumpulan pengklasifikasi pohon terstruktur di mana setiap pohon mengeluarkan unit suara untuk kelas paling populer di *input x*.

*Random forest* terdiri dari kumpulan *decision tree*, di mana kumpulan *decision tree* digunakan untuk mengklasifikasikan data ke suatu kelas (Parmar et al., 2014). *Random forest* adalah salah satu dari beberapa algoritma *machine learning* yang dapat digunakan untuk melakukan *sentiment analysis*. *Random forest* merupakan algoritma *machine learning* yang dasar pembentukannya dari algoritma *decision tree*.

Dasar algoritma *random forest* adalah menggabungkan dari beberapa algoritma *decision tree* (Oshiro et al., 2012). Proses penggabungan disebut dengan ensemble learning dengan mengambil metode *vote* (Pham et al., 2015). Metode *vote* berarti memilih hasil yang paling banyak dikeluarkan oleh masing-masing algoritma *decision tree*.

Proses *input variable* dari beberapa algoritma *decision tree* dilakukan secara acak. Proses acak dilakukan untuk membuat masing-masing algoritma *decision tree* tidak saling tersambung dan mempengaruhi satu sama lainnya (Del Río et al., 2014). Proses pengacakan menggunakan metode bagging( *bootsrrap aggregating*) yang memungkinkan mengambil *variable input* secara acak dan menggantinya pada proses *modelling* masing-masing algoritma *decision tree* .

Contohnya variabel *input random forest* berupa list dari 4,3,2,8,3,1,8 maka Ketika dilakukan bagging *variable* yang diambil dari menjadi 2,3,3,8,1,6,4. Ada duplikasi dari *variable* angka 3 yang diambil. Hal itu diperlukan untuk membuat masing-masing algoritma *decision tree* tidak saling terhubung dan mempengaruhi satu sama lain.

Proses pengacakan juga berlaku pada fitur yang digunakan pada algoritma *random forest*. Contohnya Ketika ada fitur 1, fitur 2 dan fitur 3. Pada salah satu algoritma *decision tree* hanya menggunakan fitur 2 dan fitur 3 kemudian pada algoritma *decision tree* lainnya yang digunakan adalah fitur 1 dan fitur 2 begitu seterusnya.

Fungsi lain dari pengacakan yakni untuk menghindari *error* dan kesalahan perhitungan akibat dari data yang homogen. Hasil akhir dari algoritma *random*

*forest* berupa perhitungan *vote* dominan dari seluruh algoritma *decision tree* yang digunakan (Kullarni & Sinha, 2013). Proses kalkulasi algoritma *random forest* mirip dengan penilaian manusia yang dalam hal ini dari beberapa algoritma *decision tree* dengan tujuan menghindari kesalahan perhitungan dari salah satu individu algoritma *decision tree*.

### **2.2.10 Sentiment analysis**

Analisis sentimen adalah salah satu metode pemrosesan bahasa alami yang tujuannya adalah untuk mengetahui sentimen emosional dari teks yang dianalisis (Nimesh et al., 2019). Analisis sentimen bekerja secara sistematis untuk mengidentifikasi, mengekstrak, dan mempelajari kondisi dan informasi subjektif. Analisis sentimen secara luas diterapkan pada analisis pendapat konsumen, ulasan, tanggapan survei, dan media sosial. Tugas utama analisis sentimen adalah untuk mengklasifikasikan polaritas informasi dalam dokumen baik kalimat maupun kata-kata (K & F, 2016). Polaritas dokumen dikelompokkan menjadi positif, negatif, dan netral.

*Sentiment analysis* adalah proses analisa teks yang dilihat dari sudut pandang polaritas *sentiment* yang dimiliki. Setiap teks mempunyai arti dan maksud tertentu sesuai dengan sumber penulisannya yang menunjukkan kesubjektifitasan masing-masing penulis (Saif et al., 2016). Opini dari subjektifitas mempengaruhi hasil penilaian dari orang atau alat yang memproses dan membaca teks tersebut. Polaritas emosi ditimbulkan dari pengalaman yang diperoleh kemudian dirasakan lalu dicurahkan dalam bentuk tulisan.

Penunjukkan polaritas emosi yang terkandung dalam teks dapat ditemukan pada pemilihan kata yang digunakan seperti “kecewa” mempunyai arti penulis teks mempunyai polaritas emosi negative. Penunjukkan polaritas emosi juga ditunjukkan pada penggunaan emoji teks (Gabarron et al., 2019). Penggunaan emoji teks mempunyai arti sebagai interpretasi penunjukkan emosi penulis dalam teks contohnya emoji “😊” menunjukkan wajah tersenyum sehingga membuat teks mengandung arti polaritas emoji positif.

Sentimen analisis bertujuan menemukan polaritas-polaritas emosi yang terkandung dalam teks untuk mengetahui makna sesungguhnya pengalaman emosi yang dirasakan oleh penulis teks sehingga ditemukan insight tertentu yang dapat digunakan dalam banyak aspek seperti peningkatan kualitas pelayanan perusahaan, peristiwa bencana alam, peristiwa penting sosial dan respon masyarakat terhadap kebijakan pemerintah (Ansari et al., 2020; Ruz et al., 2020; Terán & Mancera, 2019; Tiwari et al., 2019; Wan & Gao, 2016).

Proses *sentiment analysis* dapat dilakukan dalam beberapa cara yakni menggunakan lexicon-based atau *machine learning*-based (A. & Sonawane, 2016). Lexicon based yakni menggunakan penilaian polaritas setiap kata berdasarkan nilai kata pada kamus yang sudah ditentukan sebelumnya. *Sentiment analysis* melalui pendekatan *machine learning* berarti melakukan analisa *sentiment* berdasarkan algoritma *machine learning* melalui *training* data yang telah dilabeli kemudian dilakukan *testing* data dengan tujuan akhir mengetahui hasil akurasi prediksi sentimen yang dilakukan. Pada metode *machine learning* dapat pula ditambahkan pengestrakan fitur teks yang biasa digunakan adalah *bag-of-words* yakni proses

pengelempokan kata dalam beberapa jumlah token yang disesuaikan dengan penelitian.

## **BAB 5**

### **PENUTUP**

#### **5.1. Simpulan**

Penelitian ini melakukan uji *sentiment analysis* pada *dataset twitter* dengan detail topik yakni yang berkaitan dengan pelantikan kabinet menteri Indonesia tahun 2019. Hasil penelitian ini berupa tingkat akurasi dari kombinasi masing-masing *feature extraction* dan *N-gram*. Kombinasi terbaik dengan tingkat akurasi tertinggi mencapai 89,19% terdapat pada penggunaan *feature extraction count vectorizer* dan *N-gram* sejumlah 3 gram.

#### **5.2. Saran**

Saran yang diberikan untuk penelitian selanjutnya adalah sebagai berikut:

1. Diharapkan untuk menggunakan metode lain seperti algoritma *machine learning* yang berbeda dan tambahan penggunaan algoritma lain seperti *feature extraction* dan *feature selection*.
2. Ditambahkan validasi labelling dataset
3. Diharapkan untuk melakukan pengujian dengan *dataset* lain baik bersumber dari media sosial atau media teks lain.

## DAFTAR PUSTAKA

- A., V., & Sonawane, S. S. (2016). *Sentiment analysis of Twitter Data: A Survey of Techniques. International Journal of Computer Applications, 139*(11), 5–15. <https://doi.org/10.5120/ijca2016908625>
- Al-sultany, G. A., & Aleqabie, H. J. (2019). Events Tagging in Twitter Using Twitter Latent Dirichlet Allocation. *International Journal of Engineering & Technology, 8*(1.5), 503–508.
- Al Amrani, Y., Lazaar, M., & El Kadirp, K. E. (2018). *Random forest and support vector machine based hybrid approach to sentiment analysis. Procedia Computer Science, 127*, 511–520. <https://doi.org/10.1016/j.procs.2018.01.150>
- Alaei, A. R., Becken, S., & Stantic, B. (2019). *Sentiment analysis in Tourism: Capitalizing on Big Data. Journal of Travel Research, 58*(2), 175–191. <https://doi.org/10.1177/0047287517747753>
- Ansari, M. Z., Aziz, M. B., Siddiqui, M. O., Mehra, H., & Singh, K. P. (2020). *Analysis of Political Sentiment Orientations on Twitter. Procedia Computer Science, 167*, 1821–1828. <https://doi.org/10.1016/j.procs.2020.03.201>
- Creswell, J. W. (2015). Qualitative Inquiry and Research Design: Choosing Among Five Approaches. In *Health Promotion Practice* (Vol. 16, Issue 4). <https://doi.org/10.1177/1524839915580941>
- Davidson, T., Warmusley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017*, 512–515.
- Del Río, S., López, V., Benítez, J. M., & Herrera, F. (2014). On the use of MapReduce for imbalanced big data using *Random forest. Information Sciences, 285*(1), 112–137. <https://doi.org/10.1016/j.ins.2014.03.043>
- Dewi, I. G. S. (2019). Penolakan masyarakat terhadap reklamasi teluk benoa provinsi bali. *Diponegoro Private Law Review, 4*(1), 390–400.
- Elbagir, S., & Yang, J. (2019). Twitter *sentiment analysis* using natural language toolkit and *Vader sentiment. Lecture Notes in Engineering and Computer Science, 2239*, 12–16.
- Fitriasih, M., & Kusumaningrum, R. (2019). Analisis Klasifikasi Opini Tweet Pada Media Sosial Twitter Menggunakan Latent Dirichlet Allocation (LDA). *Seminar Nasional Teknologi Informasi Dan Komunikasi 2019 (SENTIKA 2019)*, 177–186.
- Furqon, M. A., Hermansyah, D., Sari, S., Sukma, A., Akbar, Y., & Rakhmawati, N. A. (2018). Analisis Sosial Media Pemerintah Daerah di Indonesia Berdasarkan Respons Warganet. *Jurnal Sositologi, 17*(2), 2–4.
- Gabarron, E., Dorrnzoro, E., Rivera-Romero, O., & Wynn, R. (2019). Diabetes on Twitter: A *Sentiment analysis. Journal of Diabetes Science and Technology, 13*(3), 439–444. <https://doi.org/10.1177/1932296818811679>
- Garay, J., Yap, R., & Sabellano, M. J. (2019). An *analysis* on the insights of the anti-vaccine movement from social media posts using k-means clustering algorithm and *VADER sentiment analyzer. IOP Conference Series: Materials Science and Engineering, 482*(1). [52](https://doi.org/10.1088/1757-</a></p></div><div data-bbox=)



899X/482/1/012043

- Ghiassi, M., & Lee, S. (2018). A domain transferable lexicon set for Twitter *sentiment analysis* using a supervised machine learning approach. *Expert Systems with Applications*, *106*, 197–216. <https://doi.org/10.1016/j.eswa.2018.04.006>
- Hutto, C. J., & Gilbert, E. E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment analysis of Social Media Text*. Eighth International Conference on Weblogs and Social Media (ICWSM-14).". *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*. <http://sentiment.net/>
- Indriani, A. F., & Muslim, M. A. (2019). SVM Optimization Based on PSO and AdaBoost to Increasing Accuracy of CKD Diagnosis. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, *10*(2), 119. <https://doi.org/10.24843/lkjiti.2019.v10.i02.p06>
- Irza, I. F., Zuhendra, & Efrizon. (2017). Analisis Perbandingan Kinerja Web Server Apache dan Nginx Menggunakan Httperf Pada Portal Berita (Studi Kasus beritalinux.com). *Jurnal Vokasional Teknik Elektronika & Informatika (VOTENIKA)*, *5*(2), 1–8.
- Izzati, A. N., Pratama, A., Aristamy, I. M., Najwa, N. F., & Rakhmawati, N. A. (2018). Kategorisasi Jenis Interaksi Pemerintah dan Masyarakat serta Popularitas Media Sosial Pemerintah Daerah. *Jurnal Sistem Informasi (Journal of Information System)*, *14*(1), 1–8.
- Julian, B. G., Budi, I., & Tanaya, D. (2019). Performance of DKI Jakarta Governor and Vice Governor on 2017-2018 based on *Sentiment analysis* using Twitter and Instagram Data. *Proceedings of 2019 2nd International Conference on Data Science and Information Technology (DSIT'19)*. Seoul, Republic of Korea, 122–127. <https://doi.org/10.1145/3352411.3352431>
- K, S., & F, F. (2016). Survey on aspect-level *sentiment analysis*. *IEEE Transactions on Knowledge and Data Engineering*, *28*(3), 813–830.
- Kotu, V., & Deshpande, B. (2019). *Text mining*. In Springer. <https://doi.org/10.1016/b978-0-12-814761-0.00009-5>
- Kullarni, V. Y., & Sinha, P. K. (2013). *Random forest Classifier: A Survey and Future Research Directions*. *International Journal of Advanced Computing*, *36*(1), 1144–1156.
- Kumar, S., Yadava, M., & Roy, P. P. (2019). Fusion of EEG response and *sentiment analysis* of products review to predict customer satisfaction. *Information Fusion*, *52*, 41–52. <https://doi.org/10.1016/j.inffus.2018.11.001>
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22. <https://doi.org/10.1177/154405910408300516>
- Moento, P. A., Firman, & Yusuf, A. P. (2019). Good Governance Dalam Pemerintahan. *Musamus Journal of Public Administration*, *1*(2), 10–16.
- Nimesh, R., Veera Raghava, P., Prince Mary, S., & Bharathi, B. (2019). A Survey on Opinion Mining and *Sentiment analysis*. In *IOP Conference Series: Materials Science and Engineering* (Vol. 590, Issue 1). <https://doi.org/10.1088/1757-899X/590/1/012003>
- Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a

- random forest? Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7376 LNAI, 154–168. [https://doi.org/10.1007/978-3-642-31537-4\\_13](https://doi.org/10.1007/978-3-642-31537-4_13)
- Pangestu, S. Y., Astuti, Y., & Farida, L. D. (2019). Algoritma Support Vector Machine untuk Klasifikasi Sikap Politik terhadap Partai Politik Indonesia. *Jurnal Mantik Penusa*, 3(1), 236–241.
- Parmar, H., Bhandaru, S., & Shah, G. (2014). *Sentiment Mining of Movie Reviews using Random forest with Tuned Hyperparameters. International Conference on Information Science.*
- Pham, V. Q., Kozakaya, T., Yamaguchi, O., & Okada, R. (2015). COUNT forest: Co-voting uncertain number of targets using *random forest* for crowd density estimation. *Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter*, 3253–3261. <https://doi.org/10.1109/ICCV.2015.372>
- Pinto, J. P., & Murari, V. (2019). Real Time *Sentiment analysis* of Political Twitter Data Using Machine Learning Approach. *International Research Journal of Engineering and Technology (IRJET)*, 6(4), 4124–4129.
- Rahutomo, F., Retno, A., & Ririd, T. H. (2019). Evaluasi Daftar *Stopwords* Bahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 6(1), 41. <https://doi.org/10.25126/jtiik.2019611226>
- Ramanathan, V., & Meyyappan, T. (2019). *Twitter text mining for sentiment analysis* on people's feedback about Oman tourism. *2019 4th MEC International Conference on Big Data and Smart City, ICBDS 2019*, 1–5. <https://doi.org/10.1109/ICBDSC.2019.8645596>
- Ravi, A., Khettry, A. R., & Yelandur Sethumadhavachar, S. (2019). Amazon reviews as corpus for *sentiment analysis* using machine learning. In *Communications in Computer and Information Science* (Vol. 1045). Springer Singapore. [https://doi.org/10.1007/978-981-13-9939-8\\_36](https://doi.org/10.1007/978-981-13-9939-8_36)
- Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019). Perbandingan Metode Naive Bayes , Knn Dan *Decision tree* Terhadap Analisis Sentimen Transportasi Krl. *Jurnal IKRA-ITH Informatika*, 3(2), 1–9.
- Ruz, G. A., Henríquez, P. A., & Mascareño, A. (2020). *Sentiment analysis* of Twitter data during critical events through Bayesian networks classifiers. *Future Generation Computer Systems*, 106, 92–104. <https://doi.org/10.1016/j.future.2020.01.005>
- Safri, Y. F., Arifudin, R., & Muslim, M. A. (2018). K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor. *Scientific Journal of Informatics*, 5(1), 18. <https://doi.org/10.15294/sji.v5i1.12057>
- Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for *sentiment analysis* of Twitter. *Information Processing and Management*, 52(1), 5–19. <https://doi.org/10.1016/j.ipm.2015.01.005>
- Sholihin, M. R., Arianto, W., & Khasanah, D. F. (2018). Keunggulan Sosial Media dalam Perkembangan Ekonomi Kreatif Era Digital di Indonesia. *Prosiding 4th Seminar Nasional Fakultas Ekonomi Universitas Muhammadiyah Jember*, 149–160.
- Singh, K., Shakya, H. K., & Biswas, B. (2016). Clustering of people in social

- network based on *textual similarity*. *Perspectives in Science*, 8, 570–573. <https://doi.org/10.1016/j.pisc.2016.06.023>
- Terán, L., & Mancera, J. (2019). Dynamic profiles using *sentiment analysis* and twitter data for voting advice applications. *Government Information Quarterly*, 36(3), 520–535. <https://doi.org/10.1016/j.giq.2019.03.003>
- Thu, P. P., & New, N. (2017). Impact *analysis* of emotion in figurative language. *Proceedings - 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017, 2010*, 209–214. <https://doi.org/10.1109/ICIS.2017.7959995>
- Tiwari, P., Yadav, P., Kumar, S., Mishra, B. K., Nguyen, G. N., Gochhayat, S. P., Singh, J., & Prasad, M. (2019). *Sentiment analysis* for Airlines Services Based on Twitter Dataset. In *Social Network Analytics*. Elsevier Inc. <https://doi.org/10.1016/b978-0-12-815458-8.00008-6>
- Ulya, Z. (2016). Kedudukan Wakil Menteri Dalam Sistem Penyelenggaraan Pemerintahan Negara Menurut Undang-Undang Dasar Tahun 1945 ( Position of Deputy Minister of State in the System Operation of Government Under the 1945 Constitution of the Republic of. *Jurnal Legislasi Indonesia, 1945*(39), 213–219.
- Wan, Y., & Gao, Q. (2016). An Ensemble *Sentiment Classification* System of Twitter Data for Airline Services *Analysis*. *Proceedings - 15th IEEE International Conference on Data mining Workshop, ICDMW 2015, March*, 1318–1325. <https://doi.org/10.1109/ICDMW.2015.7>
- Yang, S., & Zhang, H. (2018). *Text mining* of Twitter Data Using a Latent Dirichlet Allocation Topic Model and *Sentiment analysis*. *International Journal of Computer and Information Engineering*, 12(7), 525–529. <https://doi.org/10.5281/zenodo.1317350>
- Zhang, L., & Zhu, Q. (2014). The overtime waiting model for web server performance evaluation. *Proceedings - 2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, CyberC 2014*, 229–232. <https://doi.org/10.1109/CyberC.2014.48>