

Analisis Sentimen Menggunakan K-Nearest Neighbors dan Lexicon Based

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



Oleh:

Zikry Kurniawan
09021281621040

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2021

LEMBAR PENGESAHAN TUGAS AKHIR

**ANALISIS SENTIMEN MENGGUNAKAN K-NEAREST
NEIGHBORS DAN LEXICON BASED**

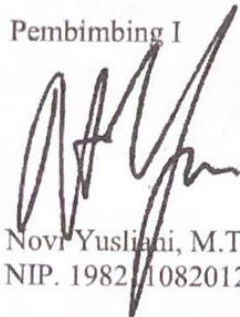
Oleh:

Zikry Kurniawan

NIM: 09021281621040

Palembang, 30 April 2021

Pembimbing I



Novi Yusliani, M.T
NIP. 198211082012122001

Pembimbing II



Muhammad Ali Buchari, S.Kom., M.T.
NIP. 198803302019031007

Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

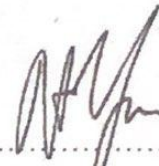
TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Jumat tanggal 16 April 2021 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Zikry Kurniawan
NIM : 09021281621040
Judul : Analisis Sentimen menggunakan *K-Nearest Neighbors* dan *Lexicon Based*

1. Pembimbing I

Novi Yusliani, M.T.
NIP. 198211082012122001



2. Pembimbing II

Muhammad Ali Buchari, S.Kom., M.T.
NIP. 198803302019031007



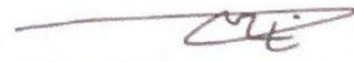
3. Penguji I

Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005



4. Penguji II

Osvari Arsalan, S.Kom., M.T.
NIP. 198806282018031001



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Zikry Kurniawan

NIM : 09021281621040

Program Studi : Teknik Informatika

Judul : Analisis Sentimen Menggunakan *K-Nearest Neighbors* dan
Lexicon Based

Hasil Pengecekan Software *iThenticate/Turnitin*: 16%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek 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, 30 April 2021



Zikry Kurniawan
NIM. 09021281621040

MOTTO DAN PERSEMBAHAN

Motto:

“Jangan berkeluh kesah, apa pun yang hilang darimu, kembali juga padamu
dalam wujud yang lain”

[Maulana Jalaluddin Rumi Muhammad]

“There are no beautiful surfaces without a terrible depth”

[Friedrich Nietzsche]

Kupersembahkan kepada:

Orangtua dan saudaraku

Keluarga besarku

Kawan kawan seperjuangan

Dosen Pembimbing dan Penguji

Tuhan, Bangsa, dan Almamater

ANALISIS SENTIMEN MENGGUNAKAN K-NEAREST NEIGHBORS DAN LEXICON BASED

Oleh:
Zikry Kurniawan
09021281621040

ABSTRAK


Analisis sentimen (*sentiment analysis*) merupakan bagian dari *text mining* yang digunakan untuk klasifikasi polaritas teks. Analisis sentimen dapat melihat bagaimana opini publik terhadap suatu isu atau permasalahan yang terjadi. Sejumlah penelitian tentang analisis sentimen telah dilakukan dengan berbagai pendekatan pembelajaran mesin. Salah satu algoritma dalam pembelajaran mesin adalah *K-Nearest Neighbors* (KNN). KNN merupakan algoritma yang digunakan untuk klasifikasi objek baru berdasarkan data latih. Namun pada sejumlah penelitian analisis sentimen tidak dilakukan evaluasi kata negasi. Kata negasi dapat mengubah sentimen suatu kalimat dari positif menjadi negatif dan sebaliknya. Untuk itu perlu adanya teknik yang mampu melakukan evaluasi kata negasi guna meningkatkan keakuratan hasil sentimen. Salah satu metode yang dapat menangani masalah kata negasi adalah metode *Lexicon Based*. Berdasarkan lima kali pengujian dengan nilai parameter k berbeda, didapatkan nilai pengujian terbaik pada percobaan kedua ($k=3$) dengan nilai akurasi sebesar 0.89, nilai *precision* sebesar 0.88, *recall* sebesar 0.88, dan *f-measure* sebesar 0.88.

Kata Kunci: Analisis sentimen, *K-Nearest Neighbors*, kata negasi, *Lexicon Based*

Pembimbing I


Novy Yusliani, M.T
NIP. 198211082012122001

Palembang, 30 April 2021
Pembimbing II


Muhammad Ali Buchari, S.Kom., M.T.
NIP. 198803302019031007

Mengetahui,
Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

SENTIMENT ANALYSIS USING K-NEAREST NEIGHBORS AND LEXICON BASED

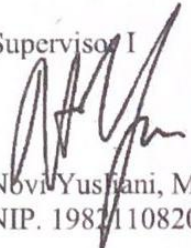
By:
Zikry Kurniawan
09021281621040

ABSTRACT

Sentiment analysis is part of text mining used for the classification of text polarity. Sentiment analysis can see how public opinion on an issue or problem occurs. A number of studies on sentiment analysis have been conducted with various machine learning approaches. One of the algorithms in machine learning is K- Nearest Neighbors (KNN). KNN is an algorithm used to classify new objects based on training data. However, in a number of sentiment analysis studies were not evaluated the word negation. The word negation can change the sentiment of a sentence from positive to negative and vice versa. Therefore, there needs to be a technique that is able to evaluate the word negation in order to improve the accuracy of sentiment results. One method that can handle the problem of the word negation is the Lexicon Based method. Based on five tests with five different parameter k values, the best test value was obtained in the second experiment ($k = 3$) with accuracy value of 0.89, precision value of 0.88, recall of 0.88, and f-measure of 0.88.

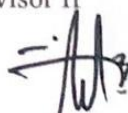
Keywords: Sentiment analysis, *K-Nearest Neighbors*, negation word, *Lexicon Based*

Supervisor I


Novi Yusufjani, M.T
NIP. 198211082012122001

Palembang, 30 April 2021

Supervisor II


Muhammad Ali Buchari, S.Kom., M.T.
NIP. 198803302019031007

Approve,
Head of the Informatics Engineering Department,



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

KATA PENGANTAR

Puji syukur kepada Allah SWT atas rahmat dan karunia-Nya sehingga saya mampu menyelesaikan skripsi ini dengan baik. Skripsi ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam kesempatan ini, penulis ingin menyampaikan kepada seluruh pihak telah memberikan banyak dukungan dan bantuan baik itu berupa doa, tenaga, waktu, dan lainnya dalam proses penyelesaian skripsi ini. Saya ucapkan terima kasih kepada:

1. **Sumarno dan Mardiana**, kedua orang tua saya yang selalu memanjatkan doa sepanjang waktu, ringan tangan membantu ketika saya kesulitan, dan menyediakan telinga dan kasih sayang saat penulis senang maupun sedih.
2. **Mentari Angraeni**, kakak saya yang menjadi sumber inspirasi dan panutan penulis semasa hidup, serta tumbuh bersama dari awal penulis dilahirkan hingga detik tulisan ini dibuat.
3. **Jaidan Jauhari, M.T.**, selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. **Alvi Syahrini Utami, M.Kom.**, selaku Ketua Jurusan Teknik Informatika Ilmu Komputer Universitas Sriwijaya.
5. **Novi Yusliani, M.T.**, selaku Pembimbing Skripsi Pertama yang telah banyak membantu dan membimbing saya dari awal penentuan topik skripsi hingga dinyatakan lulus pada seminar komprehensif.

6. **Muhammad Ali Buchari, M.T.**, selaku Pembimbing Skripsi Kedua yang telah banyak membantu dan membimbing saya dalam pembuatan skripsi ini.
7. **Dr. Abdiansah, S.Kom., M.Cs**, dan **Osvari Arsalan, M.T.**, selaku Penguji Skripsi yang telah membantu menyempurnakan tugas akhir saya dengan saran yang bermanfaat selama proses seminar proposal hingga seminar komprehensif.
8. **Anggina Primanita, M.T.**, dan **Desty Rodiah, M.T.**, selaku Pembimbing Akademik yang telah memberikan semangat dan mengarahkan saya semasa perkuliahan.
9. **Winda Kurnia Sari dan Ricy Firnando**, selaku Admin Teknik Informatika yang membantu administrasi saya selama masa kuliah.
10. **Dosen-dosen Fakultas Ilmu Komputer**, yang tidak bisa saya sebutkan satu persatu, terima kasih telah memberikan ilmu selama masa perkuliahan.
11. **Muhammad Irfan Triyanto Putra**, salah satu manusia terbaik yang pernah penulis temui semasa kuliah dan selalu mengajarkan dan memberikan wejangan baik urusan dunia maupun akhirat sehingga penulis dapat mendapatkan gelar S.Kom.
12. **Muhammad Irsyad Masyhudin dan Mohammad Sulthan Alif Utama**, teman yang mau memberikan saya kesempatan untuk satu tim dan berbagi ilmu pada kerja praktek.
13. **M. Edu Agritama, Muhammad Ghuftron Khaical Alfaris, Muhammad Shafrullah, Reyhan Navind Shaquile Harahap, dan M Ramadhani Sultan Arkas**, teman senasib dan seperantauan di Kota Indralaya yang telah

memberikan bantuan, menjadi saksi hidup dalam berbagi canda dan tawa, selalu sedia memberikan waktunya untuk belajar dan bermain dari awal kuliah hingga saat ini.

14. **Acmad Fadli Aditama, Alif Muhammad, Ahmad Ryadh, Daniel Farhan, M. Farid Landriandani, dan Muhammad Abdi Priyangga**, teman yang selalu meluangkan waktu dan menyediakan tempat tinggal saat penulis berkunjung ke Kota Palembang.
15. **Teman-teman INFORGEN'16**, yang telah berbagi pengalaman, keluh kesah, semangat, canda tawa selama masa perkuliahan.
16. **Seluruh penghuni Pemandokkan Citra dan Kost H Mahmud**, yang telah memberikan suasana seperti penulis berada di rumah sendiri.

Indralaya, 30 April 2021

Zikry Kurniawan

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR.....	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xvi
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan.....	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah.....	I-3
1.4 Tujuan Penelitian	I-4
1.5 Manfaat Penelitian	I-4
1.6 Batasan Masalah	I-4
1.7 Sistematika Penulisan	I-5
1.8 Kesimpulan	I-6
BAB II KAJIAN LITERATUR.....	II-1
2.1 Pendahuluan.....	II-1
2.2. Landasan Teori.....	II-1
2.2.1 Analisis Sentimen	II-1
2.2.2 Praproses Teks	II-3
2.2.3 Pembobotan Kata Term Frequency Inverse Document Frequency (TF-IDF).....	II-6
2.2.4 K-Nearest Neighbors (KNN).....	II-7

2.2.5	Lexicon Based.....	II-8
2.2.6	Coronavirus Disease 2019 (COVID-19).....	II-10
2.2.7	Twitter.....	II-11
2.2.8	<i>Confusion Matrix</i>	II-12
2.2.9	Rational Unified Process (RUP)	II-14
2.3.	Penelitian Lain Yang Relevan	II-15
2.3.1	Penerapan Algoritma <i>K-Nearest Neighbors</i> Pada Analisis Sentimen Review Agen Travel.....	II-16
2.3.2	Analisis Sentimen Wisata Bahari di Sulawesi Tenggara Memanfaatkan Media Sosial Twitter dengan Menggunakan Metode Lexicon Based.....	II-16
2.3.3	Analisis Sentimen Masyarakat terhadap Virus Corona Berdasarkan Opini Dari Twitter Berbasis Web Scrapper	II-17
2.4	Kesimpulan	II-17
BAB III METODOLOGI PENELITIAN		III-1
3.1	Pendahuluan.....	III-1
3.2	Pengumpulan Data	III-1
3.3.	Tahapan Penelitian.....	III-2
3.3.1	Kerangka Kerja	III-2
3.3.2	Kriteria Pengujian	III-7
3.3.3	Format Data Pengujian	III-7
3.3.4	Alat yang digunakan dalam Pelaksanaan Penelitian.....	III-8
3.3.5	Pengujian Penelitian	III-9
3.3.6	Analisis Hasil Pengujian dan Membuat Kesimpulan	III-9
3.4	Metode Pengembangan Perangkat Lunak.....	III-10
3.4.1	Fase Insepsi.....	III-10
3.4.2	Fase Elaborasi	III-10
3.4.3	Fase Kontruksi	III-11
3.5	Kesimpulan	III-11
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-1
4.1	Pendahuluan.....	IV-1
4.2	Fase Insepsi	IV-1

4.2.1	Pemodelan Bisnis.....	IV-1
4.2.2	Kebutuhan Sistem.....	IV-2
4.2.3	Analisis dan Desain.....	IV-3
4.3	Fase Elaborasi.....	IV-27
4.3.1	Pemodelan Bisnis.....	IV-28
4.3.2	Perancangan Data.....	IV-28
4.3.3	Diagram.....	IV-30
4.4	Fase Konstruksi.....	IV-34
4.4.1	Kebutuhan Sistem.....	IV-34
4.4.2	Diagram Kelas.....	IV-35
4.4.3	Implementasi.....	IV-35
4.5	Fase Transisi.....	IV-38
4.5.1	Pemodelan Bisnis.....	IV-38
4.5.2	Rencana Pengujian.....	IV-39
4.5.3	Implementasi.....	IV-39
4.6	Kesimpulan.....	IV-41
BAB V	HASIL DAN ANALISIS PENELITIAN.....	V-1
5.1	Pendahuluan.....	V-1
5.2	Data Hasil Percobaan/Penelitian.....	V-1
5.2.1	Konfigurasi Percobaan.....	V-1
5.2.2	Data Hasil Konfigurasi.....	V-2
5.3	Analisis Hasil Pengujian.....	V-22
5.4	Kesimpulan.....	V-29
BAB VI	KESIMPULAN DAN SARAN.....	VI-1
6.1	Kesimpulan.....	VI-1
6.2	Saran.....	VI-1
DAFTAR PUSTAKA	xvii
LAMPIRAN	xxi

DAFTAR TABEL

Halaman

Tabel II- 1. <i>Confusion Matrix</i> Kelas Positif, Negatif, dan Netral	II-13
Tabel III- 1. Tabel Dataset Twitter	III-1
Tabel III- 2. Tabel <i>Confusion Matrix</i>	III-7
Tabel III- 3. Tabel Nilai Pengujian	III-8
Tabel III- 4. Rancangan Tabel Hasil Analisis Klasifikasi.....	III-9
Tabel IV-1. Kebutuhan Fungsional.....	IV-3
Tabel IV-2. Kebutuhan Nonfungsional.....	IV-3
Tabel IV-3. Contoh Data Tweet.....	IV-6
Tabel IV-4. Contoh Hasil <i>Noise Removal</i>	IV-7
Tabel IV-5. Contoh Hasil <i>Case Folding</i>	IV-7
Tabel IV-6. Contoh Hasil <i>Tokenization</i>	IV-7
Tabel IV-7. Contoh Hasil Normalisasi.....	IV-8
Tabel IV-8. Contoh Hasil <i>Stopword Removal</i>	IV-8
Tabel IV-9. Contoh Hasil <i>Stemming</i>	IV-9
Tabel IV-10. Pembobotan Kata TF-IDF	IV-9
Tabel IV-11. Perkalian Skalar Antara Data Uji dan Data Latih	IV-11
Tabel IV-12. Panjang Setiap Dokumen	IV-12
Tabel IV-13. Identifikasi Kata	IV-14
Tabel IV-14. Identifikasi Kata Negasi	IV-14
Tabel IV-15. Contoh <i>Confusion Matrix</i>	IV-15
Tabel IV-16. Definisi Aktor	IV-17
Tabel IV-17. Definisi Use Case	IV-18
Tabel IV-18. Rancangan Data.....	IV-28
Tabel IV-19. Implementasi Kelas	IV-36
Tabel IV-20. Rencana Pengujian Analisa Sentimen	IV-39
Tabel IV-21. Pengujian Analisa Sentimen.....	IV-40
Tabel V-1. Hasil <i>Confusion Matrix</i> Percobaan Pertama.....	V-2
Tabel V-2. Nilai Hasil Percobaan Pertama KNN dan Lexicon.....	V-3
Tabel V-3. Nilai Hasil Percobaan Pertama KNN.....	V-4
Tabel V-4. Hasil <i>Confusion Matrix</i> Percobaan Kedua.....	V-7
Tabel V-5. Nilai Hasil Percobaan Kedua Kedua KNN dan Lexicon.....	V-8
Tabel V-6. Nilai Hasil Percobaan Kedua KNN	V-8
Tabel V-7. Hasil <i>Confusion Matrix</i> Percobaan Ketiga	V-11
Tabel V-8. Nilai Hasil Percobaan Ketiga Kedua KNN dan Lexicon.....	V-11
Tabel V-9. Nilai Hasil Percobaan Ketiga Kedua KNN.....	V-12
Tabel V-10. Hasil <i>Confusion Matrix</i> Percobaan Keempat KNN dan Lexicon .	V-15
Tabel V-11. Nilai Hasil Percobaan Keempat KNN dan Lexicon	V-15

Tabel V-12. Hasil <i>Confusion Matrix</i> Percobaan Keempat KNN dan Lexicon .	V-16
Tabel V-13. Hasil <i>Confusion Matrix</i> Percobaan Kelima	V-19
Tabel V-14. Hasil <i>Confusion Matrix</i> Percobaan Kelima KNN dan Lexicon....	V-19
Tabel V-15. Hasil <i>Confusion Matrix</i> Percobaan Kelima KNN dan Lexicon....	V-20
Tabel V-16. Analisis Hasil Pengujian.....	V-23

DAFTAR GAMBAR

Halaman

Gambar II- 1. Contoh Kerangka Kerja Analisis Sentimen (Deviyanto & Wahyudi, 2018)	II-2
Gambar II- 2. Ilustrasi Noise Removal	II-3
Gambar II-3. Ilustrasi Case Folding	II-4
Gambar II- 4. Ilustrasi Tokenization	II-4
Gambar II- 5. Ilustrasi Normalisasi	II-5
Gambar II- 6. Ilustrasi Stopword Removal	II-5
Gambar II- 7. Ilustrasi Stemming	II-6
Gambar II- 8. Model Rational Unified Process (Rosa A.S, 2011).....	II-14
Gambar III- 1. Diagram Kerangka Kerja	III-3
Gambar III- 2. Diagram Tahapan Praproses K-Nearest Neighbors	III-4
Gambar III- 3. Diagram Tahapan Praproses Lexicon Based	III-5
Gambar IV-1. Nilai Hasil Evaluasi	IV-16
Gambar IV-2. Diagram Use Case	IV-17
Gambar IV-3. Skenario Use Case kinerja	IV-20
Gambar IV-4. Skenario Use Case Praproses.....	IV-21
Gambar IV-5. Skenario Use Case KNN	IV-22
Gambar IV-6. Skenario Use Case KNN dan Lexicon.....	IV-23
Gambar IV-7. Diagram Aktivitas Kinerja.....	IV-24
Gambar IV-8. Diagram Aktivitas Praproses	IV-25
Gambar IV-9. Diagram Aktivitas KNN	IV-26
Gambar IV-10. Diagram Aktivitas KNN dan Lexicon	IV-27
Gambar IV-11. Diagram Sequence Praproses Dataset	IV-31
Gambar IV-12. Diagram Sequence KNN	IV-31
Gambar IV-13. Diagram Sequence KNN dan Lexicon	IV-32
Gambar IV-14. Diagram Sequence Kinerja	IV-32
Gambar IV-15. Antarmuka Halaman Depan	IV-33
Gambar IV-16. Antarmuka Halaman Hasil	IV-33
Gambar IV-17. Diagram Kelas	IV-35
Gambar IV-18. Antarmuka Halaman Depan	IV-37
Gambar IV-19. Antarmuka Halaman Hasil	IV-38
Gambar V-1. Perbandingan Akurasi pada k=1	V-5
Gambar V-2. Perbandingan Precision pada k=1	V-5
Gambar V-3. Perbandingan Recall pada k=1	V-6
Gambar V-4. Perbandingan F-measure pada k=1	V-6
Gambar V-5. Perbandingan Akurasi pada k=3	V-9
Gambar V-6. Perbandingan Precision pada k=3	V-9
Gambar V-7. Perbandingan Recall pada k=3.....	V-10

Gambar V-8. Perbandingan F-measure pada k=3	V-10
Gambar V-9. Perbandingan Akurasi pada k=5	V-13
Gambar V-10. Perbandingan Precision pada k=5	V-13
Gambar V-11. Perbandingan Recall pada k=5.....	V-14
Gambar V-12. Perbandingan F-measure pada k=5	V-14
Gambar V-13. Perbandingan Akurasi pada k=7	V-17
Gambar V-14. Perbandingan Precision pada k=7	V-17
Gambar V-15. Perbandingan Recall pada k=7.....	V-18
Gambar V-16. Perbandingan F-measure pada k=7	V-18
Gambar V-17. Perbandingan Akurasi pada k=9	V-21
Gambar V-18. Perbandingan Precision pada k=9	V-21
Gambar V-19. Perbandingan Recall pada k=9.....	V-22
Gambar V-20. Perbandingan F-measure pada k=9	V-22
Gambar V-21. Hasil Pengujian Penelitian dengan Accuracy	V-24
Gambar V-22. Hasil Pengujian Penelitian dengan Average Precision	V-25
Gambar V-23. Hasil Pengujian Penelitian dengan Average Recall.....	V-26
Gambar V-24. Hasil Pengujian Penelitian dengan Average F-measure	V-27

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan dibahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan. Bab ini juga akan memberikan penjelasan umum mengenai keseluruhan penelitian.

Bab ini dimulai dengan pembahasan terhadap analisis sentimen serta penelitian yang berkaitan dengan *K-Nearest Neighbors* dan *Lexicon Based*.

1.2 Latar Belakang

Analisis sentimen (*sentiment analysis*) merupakan bagian dari *text mining* yang digunakan untuk klasifikasi polaritas teks. Klasifikasi bertujuan untuk melihat polaritas kalimat bernilai positif, negatif, atau netral. Analisis sentimen dapat melihat bagaimana opini publik terhadap suatu isu atau permasalahan yang terjadi (Fanissa *et al.*, 2018). Analisis sentimen atau *opinion mining* dapat diterapkan ke berbagai bidang, mulai dari produk konsumen, jasa kesehatan, jasa keuangan, peristiwa sosial, dan politik (Zhang *et al.*, 2016).

Sejumlah penelitian tentang analisis sentimen telah dilakukan dengan berbagai pendekatan pembelajaran mesin baik itu *supervised* maupun *unsupervised learning*. Pendekatan dengan *supervised learning* telah diteliti dengan berbagai metode diantaranya *Naïve Bayes* (Pang, B., Lee, L., & Vithyanathan, 2019), *Support Vector Machine* (Pang, B., Lee, L., & Vithyanathan, 2019), dan *lainnya*.

Sementara, pendekatan dengan *unsupervised learning* biasanya menggunakan *Lexicon Based* (Taboada *et al.*, 2011) yang ditandai dengan penggunaan kamus kata positif dan negatif (Zhang *et al.*, 2016).

Salah satu algoritma dalam pembelajaran mesin adalah *K- Nearest Neighbors* (KNN). KNN merupakan algoritma yang digunakan untuk klasifikasi objek baru berdasarkan data latih. Prinsip kerja dari metode KNN adalah klasifikasi yang menggunakan ketetanggan sebagai nilai prediksi dari data uji yang baru (Krisandi *et al.*, 2013). Metode KNN memiliki beberapa kelebihan diantaranya metode ini efektif dalam pengolahan data latih yang besar, tangguh terhadap data latih yang *noise*, dan memiliki konsistensi yang kuat sehingga menghasilkan data yang akurat (Kurniawan, 2012).

Dalam penelitian analisis sentimen yang dilakukan (Ernawati & Wati, 2018) menggunakan *K-Nearest Neighbors* pada *review* agen travel mendapatkan hasil tingkat akurasi keberhasilan sebesar 87%. Pada tahap praproses, penelitian ini melalui beberapa tahapan diantaranya *case folding*, *tokenizing*, *stopword*, dan *stemming*. Penggunaan metode KNN dalam penelitian ini disertai dengan pembobotan kata *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). *Cosine similarity* digunakan pada penelitian ini untuk menentukan kemiripan antara data uji dan latih. Dari hasil penelitian tersebut terlihat *K-Nearest Neighbors* cukup baik diterapkan pada analisis sentimen.

Namun pada sejumlah penelitian analisis sentimen tidak dilakukan evaluasi kata negasi karena telah melewati tahap *stopword removal*. *Stopword removal* adalah tahap mengambil kata-kata penting dari hasil token dengan menggunakan

algoritma *stoplist* (membuang kata tidak penting) atau *wordlist* (menyimpan kata penting). Sejumlah kalimat dapat mengandung kata negasi yang dapat mengubah sentimen suatu kalimat dari positif menjadi negatif dan sebaliknya (Kolchyna *et al.*, 2015). Untuk itu perlu adanya teknik yang mampu melakukan evaluasi kata negasi guna meningkatkan keakuratan hasil sentimen. Salah satu metode yang dapat menangani masalah kata negasi adalah metode *Lexicon Based* (Liu, 2015).

Metode *Lexicon Based* adalah metode yang bersifat *unsupervised learning* dimana prosesnya tidak membutuhkan data latih. Metode ini memerlukan kamus kata yang berisi kata positif dan kata negatif. Untuk menentukan orientasi sentimen dari suatu teks didasarkan atas fungsi positif dan negatif (Pang & Lee, 2008).

Berdasarkan uraian tersebut, maka akan dilakukan penelitian analisis sentimen menggunakan metode *K- Nearest Neighbors* dan *Lexicon Based*.

1.3 Rumusan Masalah

Rumusan masalah pada penelitian ini ialah bagaimana implementasi metode *K- Nearest Neighbors* dan *Lexicon Based* pada sistem analisis sentimen?

Dari masalah tersebut akan diuraikan pertanyaan penelitian sebagai berikut:

1. Bagaimana mengimplementasikan metode *K- Nearest Neighbors* dan *Lexicon Based* ke dalam analisis sentimen?
2. Bagaimana nilai akurasi, *precision*, *recall*, dan *f-measure* pada analisis sentimen menggunakan metode *K- Nearest Neighbors* dan *Lexicon Based*?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah sebelumnya, maka tujuan dalam penelitian ini adalah:

1. Menghasilkan perangkat lunak analisa sentimen menggunakan *K-Nearest Neighbors* dan *Lexicon Based*.
2. Mengetahui nilai akurasi, *precision*, *recall*, dan *f-measure* pada analisis sentimen menggunakan metode *K- Nearest Neighbors* dan *Lexicon Based*.

1.5 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat, yaitu:

- 1 Hasil penelitian dapat digunakan untuk mengetahui sentimen masyarakat terhadap topik COVID-19.
- 2 Hasil penelitian dapat digunakan sebagai rujukan penelitian di bidang analisis sentimen.

1.6 Batasan Masalah

Batasan masalah dari penelitian ini antara lain:

1. Data yang digunakan adalah data *Twitter (tweet)* dalam Bahasa Indonesia yang sesuai dengan pedoman penulisan ejaan Bahasa Indonesia.
2. *Tweet* yang dianalisis hanya bertopik COVID-19.
3. Format data masukkan berupa *file csv*.
4. Penelitian ini mengklasifikasikan tiga jenis sentimen yaitu, positif, negatif, dan netral.

1.7 Sistematika Penulisan

Sistematika penulisan penelitian ini adalah sebagai berikut :

BAB I. PENDAHULUAN

Pada bab ini akan diuraikan mengenai latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan dan juga kesimpulan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian, seperti definisi metode dan algoritma yang digunakan beserta langkah kerja metode dan algoritma serta kajian literatur penelitian terdahulu yang relevan terhadap penelitian ini.

BAB III. METODOLOGI PENELITIAN

Pada bab ini dibahas mengenai tahapan yang akan dilakukan pada penelitian ini. Disetiap tahapan penelitian akan dijelaskan secara terinci berdasarkan pada kerangka kerja dan diteruskan dengan perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab IV berisi pembahasan mengenai proses pengembangan perangkat lunak Analisis Sentimen Menggunakan *K-Nearest Neighbors* dan *Lexicon Based*.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab V berisi hasil pengujian penelitian berdasarkan format yang sudah ditentukan. Selain itu, pada bab ini juga dijelaskan analisis hasil pengujian sebagai dasar dari kesimpulan dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab VI berisi kesimpulan dari semua uraian pada bab-bab sebelumnya dan juga berisi saran yang diharapkan berguna dalam penerapan dan penelitian analisis sentimen selanjutnya.

1.8 Kesimpulan

Berdasarkan masalah yang telah diuraikan diatas maka akan dilakukan penelitian terhadap analisis sentimen menggunakan metode *K-Nearest Neighbors* dan *Lexicon Based*.

DAFTAR PUSTAKA

- Alvianda, F. & Adikara, P.P. 2019. Analisis Sentimen Konten Radikal Di Media Sosial Twitter Menggunakan Metode Support Vector Machine (SVM). Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya, 3(1): 241–246.
- Deviyanto, A. & Wahyudi, M.D.R. 2018. Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor. JISKA (Jurnal Informatika Sunan Kalijaga), 3(1): 1.
- Ding, X., Liu, B. & Yu, P.S. 2008. A holistic lexicon-based approach to opinion mining. WSDM'08 - Proceedings of the 2008 International Conference on Web Search and Data Mining, 231–239.
- Ernawati, S. & Wati, R. 2018. Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. Jurnal Khatulistiwa Informatika, VI(1):64–69.
<https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/3802/26> 26).
- Fanissa, S., Fauzi, M.A. & Adinugroho, S. 2018. Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking | Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2(8): 2766–2770. (<http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1962>).
- Go, A., Bhayani, R. & Huang, L. 2009. Twitter Sentiment Classification using

- Distant Supervision. Processing, 1–6.
- Hall, P., Park, B.U. & Samworth, R.J. 2008. Choice of neighbor order in nearest-neighbor classification. *Annals of Statistics*, 36(5): 2135–2152.
- Kannan, S. 2015. Preprocessing Techniques for Text Mining. (October 2014).
- Kolchyna, O., Treleaven, P. & Aste, T. 2015. Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination. (July).
- Krisandi, N., Helmi & Prihandono, B. 2013. Algoritma K-Nearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada Pt. Minamas Kecamatan Parindu. *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, 2: 33–38.
- Kumar, S., Liu, H. & Morstatter, F. 2013. *Twitter Data Analytics*. Springer-Verlag New York.
- Kurniawan, H. 2012. Sistem Penentuan Tingkat Kualitas Air pada Depot Air Minum Air Minum Menggunakan Metode K-Nearest Neighbor (K-NN).
- Kurniawan, R. & Apriliani, A. 2020. Analisis Sentimen Masyarakat terhadap Virus Corona Berdasarkan Opini Dari Twitter Berbasis Web Scraper. Analisis sentimen masyarakat terhadap virus corona berdasarkan opini dari twitter berbasis web scraper, .
- Liu, B. 2015. Sentiment analysis: Mining opinions, sentiments, and emotions. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, (May): 1–367.
- Liu, B., Hu, M. & Cheng, J. 2005. *Opinion Observer: Analyzing and Comparing*

- Opinions on the Web. Proceedings of the 14th International Conference on World Wide Web, 342–351. (<http://dl.acm.org/citation.cfm?id=1060797>).
- Maarif, A.A. 2015. Penerapan Algoritma TF-IDF untuk Pencarian Karya Ilmiah. Dokumen Karya Ilmiah | Tugas Akhir | Program Studi Teknik Informatika - S1 | Fakultas Ilmu Komputer | Universitas Dian Nuswantoro Semarang, (5): 4. (mahasiswa.dinus.ac.id/docs/skripsi/jurnal/15309.pdf).
- Melita, R., Amrizal, V., Suseno, H.B. & Dirjam, T. 2018. Penerapan Metode Term Frequency Inverse Document Frequency (TF-IDF) dan Cosine Similarity pada Sistem Temu Kembali Informasi untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Syarah Umdatil Ahkam). JTI (Jurnal Teknik Informatika) UIN Syarif Hidayatullah, 11(2): 149–164.
- Pang, B., Lee, L., & Vithyanathan, S. 2019. Thumbs up? Sentiment Classification using Machine Learning Techniques. Proceedings of the Institution of Civil Engineers - Transport.
- Pang, B. & Lee, L. 2008. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2: 1–135.
- Pristiyanti, R.I., Fauzi, M.A. & Muflikhah, L. 2018. Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya, 2(3): 1179–1186. (<http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1140>).
- Putra, R.S. & Adisantoso, J. 2017. Analisis Sentimen Twitter dengan Klasifikasi

Naïve Bayes menggunakan Seleksi Fitur Mutual Information dan Inverse Document Frequency.

Statiswaty, Rusnia & Ransi, N. 2017. Analisis Sentimen Wisata Bahari Di Sulawesi Tenggara Memanfaatkan Media Sosial Twitter Dengan Menggunakan Metode Lexicon-Based. 3(2): 161–168.

Stehman, S. V 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of environment*, 62: 77–89.

Susilo, A., Rumende, C.M., Pitoyo, C.W., Santoso, W.D., Yulianti, M., Herikurniawan, H., Sinto, R., Singh, G., Nainggolan, L., Nelwan, E.J., Chen, L.K., Widhani, A., Wijaya, E., Wicaksana, B., Maksum, M., Annisa, F., Jasirwan, C.O.M. & Yuniastuti, E. 2020. Coronavirus Disease 2019: Tinjauan Literatur Terkini. *Jurnal Penyakit Dalam Indonesia*, 7(1): 45.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K. & Stede, M. 2011. [SO-CAL] [tool] [seminal lexicon] Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*.

Wahid, D.H. & SN, A. 2016. Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 10(2): 207.

Zhang, L., Liu, B. & Francisco, S. 2016. Encyclopedia of Machine Learning and Data Mining. *Encyclopedia of Machine Learning and Data Mining*.