

***SENTIMENT ANALYSIS* LAYANAN PESAN ANTAR MAKANAN PADA  
TWITTER MENGGUNAKAN *INSET LEXICON*  
(STUDI KASUS: GOFOOD DAN GRABFOOD)**

(Skripsi)

Oleh :

**YULIVIA ANNISA PUTRI**

**NPM. 1857051008**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2023**

## ABSTRAK

### ***SENTIMENT ANALYSIS LAYANAN PESAN ANTAR MAKANAN PADA TWITTER MENGGUNAKAN INSET LEXICON (STUDI KASUS: GOFOOD DAN GRABFOOD)***

Oleh

**YULIVIA ANNISA PUTRI**

Perkembangan ilmu pengetahuan dan teknologi yang pesat banyak menghadirkan teknologi baru. Teknologi yang dimaksud yakni GoFood dan GrabFood yang merupakan layanan pesan antar makanan terbesar di Indonesia yang dapat dipesan secara *online* melalui sebuah aplikasi *smartphone*. Sebagian pengguna GoFood dan GrabFood menyampaikan opini terhadap layanan GoFood dan GrabFood melalui *Twitter*. Opini tersebut dapat digunakan sebagai evaluasi bagi kedua perusahaan. Banyaknya opini pengguna GoFood dan GrabFood yang ditemukan pada media sosial *Twitter*. Karenanya dilakukan penelitian mengenai analisis sentimen terhadap pengguna GoFood dan GrabFood pada media sosial *Twitter* pada periode 01 Februari 2022 – 21 Februari 2022. Penelitian ini dilakukan dengan metode *Lexicon Based* menggunakan *corpus InSet Lexicon*. Data diklasifikasikan ke dalam sentimen positif dan sentimen negatif. Hasil klasifikasi layanan GoFood pada sentimen positif sebesar 40% dan sentimen negatif sebesar 60%. Sedangkan layanan GrabFood memperoleh sentimen positif sebesar 47% dan sentimen negatif sebesar 53%.

Kata kunci: *InSet Lexicon*, GoFood, GrabFood, *tweet*, analisis sentimen.

## **ABSTRACT**

### ***SENTIMENT ANALYSIS LAYANAN PESAN ANTAR MAKANAN PADA TWITTER MENGGUNAKAN INSET LEXICON (STUDI KASUS: GOFOOD DAN GRABFOOD)***

**By**

**YULIVIA ANNISA PUTRI**

The rapid development of science and technology presents many new technologies. The technologies in question are GoFood and GrabFood which are the largest food delivery services in Indonesia and can be ordered online via a smartphone application. Some GoFood and GrabFood users expressed their opinions on GoFood and GrabFood services via Twitter. This opinion can be used as an evaluation for both companies. Many opinions of GoFood and GrabFood users were found on Twitter social media. Therefore, research was conducted on sentiment analysis of GoFood and GrabFood users on social media Twitter in the period 01 February 2022 – 21 February 2022. This research was conducted using the Lexicon Based method using the InSet Lexicon corpus. Data is classified into positive sentiment and negative sentiment. GoFood service classification results in a positive sentiment of 40% and negative sentiment of 60%. Meanwhile, GrabFood services received positive sentiment of 47% and negative sentiment of 53%.

**Key word:** InSet Lexicon, GoFood, GrabFood, tweet, sentiment analysis.

***SENTIMENT ANALYSIS* LAYANAN PESAN ANTAR MAKANAN PADA  
TWITTER MENGGUNAKAN *INSET LEXICON*  
(STUDI KASUS: GOFOOD DAN GRABFOOD)**

Oleh

**YULIVIA ANNISA PUTRI**

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar  
**SARJANA ILMU KOMPUTER**

Pada

**Jurusan Ilmu Komputer  
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2023**

Judul Skripsi : **SENTIMENT ANALYSIS LAYANAN PESAN ANTAR MAKANAN PADA TWITTER MENGGUNAKAN INSET LEXICON (STUDI KASUS: GOFOOD DAN GRABFOOD)**

Nama Mahasiswa : **Yulivia Annisa Putri**

Nomor Induk Mahasiswa : **1857051008**

Program Studi : **Ilmu Komputer**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**



1. Komisi Pembimbing

  
**Dr. Ir. Kurnia Muludi, M.S.Sc.**  
NIP. 19640616 198902 1 001

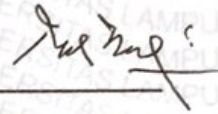
2. Ketua Jurusan Ilmu Komputer

  
**Didik Kurniawan, S.Si., M.T.**  
NIP 19800419 200501 1 004

**MENGESAHKAN**

**1. Tim Penguji**

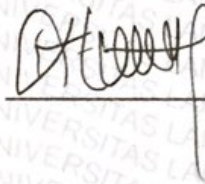
Ketua : **Dr. Ir. Kurnia Muludi, M.S.Sc**



Penguji I : **Favorisen R. Lumbanraja, Ph.D**



Penguji II : **Tristiyanto, S.Kom., M.I.S., Ph.D**



**2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam**



**Dr Eng. Sripito Dwi Yuwono, S.Si., M.T.**

NID. 19740705 200003 1 001



**Tanggal Lulus Ujian Skripsi : 03 Februari 2023**

## PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Yulivia Annisa Putri

NPM : 1857051008

menyatakan bahwa skripsi saya yang berjudul "***Sentiment Analysis Layanan Pesan Antar Makanan Pada Twitter Menggunakan InSet Lexicon (Studi Kasus: GoFood dan GrabFood)***" merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila dikemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung, 10 Februari 2023



Yulivia Annisa Putri

NPM. 1857051008

## RIWAYAT HIDUP



Penulis bernama Yulivia Annisa Putri bertempat lahir di Bandar Lampung pada tanggal 28 Juli 2000, sebagai anak ketiga dari empat bersaudara. Penulis menyelesaikan pendidikan formal di SD Negeri 2 Rawa Laut, Bandar Lampung dan selesai pada tahun 2006-2012. Kemudian melanjutkan pendidikan menengah pertama di SMP Negeri 4 Bandar Lampung yang diselesaikan pada tahun 2015, lalu melanjutkan ke pendidikan menengah atas di SMA YP Unila Bandar Lampung yang diselesaikan pada tahun 2018.

Pada tahun 2018 penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SMMPTN BARAT. Selama menjadi mahasiswa, penulis melakukan beberapa kegiatan antara lain.

1. Menjadi anggota Printer Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2018/2019.
2. Menjadi anggota Bidang Eksternal pada Himpunan Mahasiswa Jurusan Ilmu Komputer periode 2018/2019.



3. Mengikuti Karya Wisata Ilmiah (KWI) di Desa Tanjung Titro, Kec Way Bungur, Kabupaten Lampung Timur pada tahun 2018.
4. Melaksanakan Kerja Praktik di Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu Provinsi Lampung pada tahun 2021.
5. Melaksanakan Kuliah Kerja Nyata (KKN) di Kelurahan Bumi Waras, Kecamatan Bumi Waras, Kota Bandar Lampung pada tahun 2021.

## **PERSEMBAHAN**

Alhamdulillahirobbilalamin

Puji dan syukur tercurahkan kepada Allah Subhanahu Wa Ta'ala atas segala Rahmat dan Karunia-Nya sehingga saya dapat menyelesaikan skripsi ini. Shalawat serta salam selalu tercurahkan kepada Nabi Muhammad Shallallahu Alaihi Wasallam.

Kupersembahkan karya ini kepada:

### **Kedua Orang Tuaku Tercinta**

Yang senantiasa memberikan yang terbaik, dan melantunkan do'a yang selalu menyertaiku. Kuucapkan pula terima kasih sebesar-besarnya karena telah mendidik dan membesarkanku dengan cara yang dipenuhi kasih sayang, dukungan, dan pengorbanan yang belum bisa terbalaskan.

### **Seluruh Keluarga Besar Ilmu Komputer 2018**

Yang selalu memberikan semangat dan dukungan.

### **Almamater Tercinta, Universitas Lampung dan Jurusan Ilmu Komputer**

Tempat bernaung mengemban semua ilmu untuk menjadi bekal hidup.

## MOTTO

*“... Boleh jadi kamu membenci sesuatu padahal ia amat baik bagimu, dan boleh jadi pula kamu menyukai sesuatu padahal ia amat buruk bagimu, Allah mengetahui sedang kamu tidak mengetahui.”*

*(QS Al-Baqarah: 216)*

*“Manusia tidak memiliki kuasa untuk memiliki apa pun yang dia mau, tetapi dia memiliki kuasa untuk tidak mengingini apa yang dia belum miliki, dan dengan gembira memaksimalkan apa yang dia terima”*

*(Seneca)*

*“Ada hal - hal yang berada di bawah kendali kita, ada hal – hal yang tidak berada di bawah kendali kita”*

*( Epictetus (Enchiridion) )*

## SANWACANA

Puji syukur kehadirat Allah *Subhanahu Wa Ta'ala* atas berkat rahmat dan hidayah-Nya, serta petunjuk dan pedoman dari Rasulullah Nabi Muhammad *Shollallahu Alaihi Wasallam* penulis dapat menyelesaikan skripsi yang berjudul **“Sentiment Analysis Layanan Pesan Antar Makanan Pada Twitter Menggunakan InSet Lexicon (Studi Kasus: GoFood dan GrabFood)”** dengan baik dan lancar.

Selesainya skripsi ini tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, ucapan terima kasih ditujukan kepada:

1. Allah *Subhanahu Wa Ta'ala* yang telah memberikan rahmat dan karunia-Nya selama menjalani penelitian hingga laporan dapat diselesaikan dengan baik.
2. Alm Bapak yang lebih dulu dipanggil oleh Allah SWT sebelum penulis menyelesaikan pendidikan ini dan Ibu, terima kasih selalu memberikan doa, kasih sayang, dan dukungannya sehingga penelitian ini terselesaikan.
3. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T., selaku dekan Fakultas Matematika dan Ilmu Pengetahuan Alam.
4. Bapak Didik Kurniawan, S.Si., M.T., selaku Ketua Jurusan Ilmu Komputer Universitas Lampung.
5. Bapak Dr. Ir. Kurnia Muludi, M. S.Sc. selaku dosen pembimbing utama dalam penelitian ini yang selalu memberikan bimbingan, ilmu, dan saran.
6. Bapak Favorisen R. Lumbanraja, Ph.D selaku dosen pembahas satu, yang telah memberikan ilmu dan saran dalam penelitian ini.
7. Bapak Tristiyanto, S.Kom., M.I.S., Ph.D selaku dosen pembahas dua, yang telah memberikan ilmu dan saran dalam penelitian ini.

8. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc. selaku sekretaris Jurusan Ilmu Komputer Universitas Lampung dan koordinator skripsi.
9. Seluruh Bapak dan Ibu dosen Jurusan Ilmu Komputer yang telah memberikan ilmu nya selama masa perkuliahan.
10. Ibu Ade Nora Maela dan Mas Naufal yang sudah membantu dalam urusan administrasi penulis di Jurusan Ilmu Komputer dengan sikap kerja yang kooperatif.
11. Afifah, Agnes Pramudani, Gema Annisa Hermastuti, Evania Intiha, dan Azzah Roudhoh selaku sahabat sejak awal kuliah yang selalu memberikan semangat dan membantu mencari jalan keluar setiap ada masalah.
12. Mutiara Widdi Ayunintyas dan Muhammad Arsyi Syobirin selaku teman seperbimbingan yang selalu menyemangati, membantu, dan saling mengingatkan terkait bimbingan dan laporan penelitian.
13. Inthan Noer Afrida, Nindya Swastika, dan Ria Charoline sahabat yang selalu mendengarkan keluh kesah dan memberi semangat selama penelitian.
14. Kak Namjoon, kak seokjin, kak yoongi, kak hobi, kak jimin, kak taehyung, kak jungkook, kak yeonjun, soobin, beomgyu, taehyun, dan hyuka yang sudah menemani dan menghibur selama melakukan penelitian hingga terselesainya penulisan laporan ini.
15. Teman-teman seangkatan Jurusan Ilmu Komputer 2018 di Universitas Lampung.

Disadari masih banyak kekurangan dalam penulisan laporan skripsi ini yang disebabkan terbatasnya kemampuan, pengetahuan, dan pengalaman. Tetapi, semoga skripsi ini memberikan manfaat bagi yang pihak yang membaca.

Bandar Lampung, 10 Februari 2023

Yulivia Annisa Putri

NPM. 1857051008

## DAFTAR ISI

	Halaman
<b>DAFTAR ISI</b> .....	<b>iii</b>
<b>DAFTAR TABEL</b> .....	<b>v</b>
<b>DAFTAR GAMBAR</b> .....	<b>viii</b>
<b>DAFTAR PSEUDOCODE</b> .....	<b>viii</b>
<b>I. PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian.....	5
<b>II. TINJAUAN PUSTAKA</b> .....	<b>6</b>
2.1 Penelitian Sebelumnya .....	6
2.2 Landasan Teori .....	10
2.2.1 <i>Natural Language Processing</i> .....	10
2.2.2 <i>Sentiment Analysis</i> .....	10
2.2.3 GoFood.....	11
2.2.4 GrabFood .....	11
2.2.5 Twitter .....	12
2.2.6 <i>InSet Lexicon</i> .....	12
2.2.7 <i>Pemodelan Bahasa Term Frequency-Inverse Document Frequency</i> .....	13
2.2.8 <i>K-Nearest Neighbors</i> .....	14
2.2.9 <i>Grid Search Cross Validation</i> .....	14
2.2.10 <i>Confusion Matrix</i> .....	15

<b>III. METODE PENELITIAN .....</b>	<b>17</b>
3.1 Waktu dan Tempat Penelitian .....	17
3.2 Alat Penelitian .....	17
3.2.1 Perangkat Lunak.....	17
3.2.2 Perangkat Keras .....	18
3.3 Alur Penelitian.....	19
3.3.1 Identifikasi Masalah .....	19
3.3.2 Studi Literatur .....	20
3.3.3 Membuat Akun Twitter Developer .....	20
3.3.4 Crawling Data Tweet .....	20
3.3.5 <i>Preprocessing</i> data .....	20
3.3.6 Klasifikasi Sentimen .....	23
a. Pelabelan Manual .....	23
b. <i>InSet Lexicon</i> .....	23
3.3.7 Pengujian Hasil Klasifikasi .....	25
<b>IV. HASIL DAN PEMBAHASAN.....</b>	<b>26</b>
4.1 <i>Crawling</i> Data Tweet .....	26
4.2 <i>Preprocessing</i> Data .....	29
4.3 Klasifikasi Sentimen.....	33
4.3.1 Pelabelan Manual .....	33
4.3.2 <i>InSet Lexicon</i> .....	36
4.4 Pengujian Hasil Klasifikasi .....	42
4.4.1 Model 1 .....	42
4.4.2 Model 2 .....	45
4.4.3 Keterbatasan Penelitian .....	50
<b>V. PENUTUP .....</b>	<b>51</b>
5.1 Kesimpulan.....	51
5.2 Saran .....	52
<b>DAFTAR PUSTAKA .....</b>	<b>53</b>
<b>LAMPIRAN.....</b>	<b>56</b>

## DAFTAR TABEL

	Halaman
Tabel 1. Penelitian terdahulu.....	7
Tabel 2. Contoh daftar kata dalam kamus InSet Lexicon .....	13
Tabel 3. Contoh perhitungan <i>polarity score</i> .....	13
Tabel 4. <i>Confusion matrix</i> .....	15
Tabel 5. Proses <i>cleaning</i> .....	21
Tabel 6. Proses <i>case folding</i> .....	21
Tabel 7. Proses <i>tokenizing</i> .....	22
Tabel 8. Proses <i>stopword removal</i> .....	22
Tabel 9. Proses normalisasi kalimat.....	22
Tabel 10. Identitas mahasiswa pelabelan manual .....	23
Tabel 11. Rincian data mentah.....	29
Tabel 12. Jumlah data setelah proses <i>cleaning</i> .....	30
Tabel 13. Hasil klasifikasi dari pelabelan manual .....	34
Tabel 14. Proses pelabelan manual .....	35
Tabel 15. Hasil klasifikasi dari <i>InSet Lexicon</i> .....	37
Tabel 16. Perbandingan hasil klasifikasi pada dataset GoFood.....	38
Tabel 17. Perbandingan hasil klasifikasi pada dataset GrabFood.....	39
Tabel 18. <i>Confusion matrix</i> 3 kelas dari model 1 pada dataset GoFood.....	43
Tabel 19. <i>Confusion matrix</i> 2 kelas dari model 1 pada dataset GoFood.....	43
Tabel 20. Hasil pengujian model 1 pada dataset GoFood.....	44
Tabel 21. <i>Confusion matrix</i> 3 kelas dari model 1 pada dataset GrabFood .....	44
Tabel 22. <i>Confusion matrix</i> 2 kelas dari model 1 pada dataset GoFood.....	44
Tabel 23. Hasil pengujian model 1 pada dataset GrabFood.....	45
Tabel 24. Contoh dokumen pembobotan TF-IDF.....	45
Tabel 25. Contoh pembobotan kata menggunakan TF-IDF.....	46



Tabel 26. Rasio pembagian data pada pengujian model 2.....	46
Tabel 27. Hasil pengujian model 2 pada dataset GoFood.....	47
Tabel 28. <i>Confusion matrix</i> model 2 pada dataset GoFood.....	47
Tabel 29. Hasil pengujian model 2 pada dataset GrabFood.....	48
Tabel 30. <i>Confusion matrix</i> model 2 pada dataset GoFood.....	49
Tabel 31. Perbandingan hasil pengujian model 1 dan model 2 pada dataset GoFood.....	49
Tabel 32. Perbandingan hasil pengujian model 1 dan model 2 pada dataset GoFood.....	49

## DAFTAR GAMBAR

	Halaman
Gambar 1. Tahapan Penelitian .....	19
Gambar 2. Tahapan <i>preprocessing</i> .....	21
Gambar 3. <i>Flowchart</i> klasifikasi menggunakan <i>InSet Lexicon</i> .....	24
Gambar 4. <i>API Key Twitter</i> .....	26
Gambar 5. Tampilan Anaconda .....	27
Gambar 6. Tampilan Jupyter Notebook .....	27
Gambar 7. File excel hasil <i>crawling</i> data.....	29
Gambar 8. Contoh <i>tweet</i> duplikat pada dataset.....	31
Gambar 9. Hasil akhir pelabelan manual .....	34
Gambar 10. Persentase klasifikasi pelabelan manual .....	34
Gambar 11. <i>Output</i> klasifikasi sentimen.....	37
Gambar 12. Persentase klasifikasi <i>InSet Lexicon</i> .....	38
Gambar 13. <i>Wordcloud</i> sentimen negatif pada GoFood.....	41
Gambar 14. <i>Wordcloud</i> sentimen positif pada GoFood.....	41
Gambar 15. <i>Wordcloud</i> sentimen negatif pada GrabFood.....	42
Gambar 16. <i>Wordcloud</i> sentimen positif pada GrabFood.....	42

**DAFTAR PSEUDOCODE**

	Halaman
<i>Pseudocode 1. Kode program crawling data</i> .....	28
<i>Pseudocode 2. Kode program tahapan cleaning</i> .....	30
<i>Pseudocode 3. Kode program tahapan case folding</i> .....	31
<i>Pseudocode 4. Kode program tahapan tokenizing</i> .....	32
<i>Pseudocode 5. Kode program tahapan normalisasi kalimat</i> .....	32
<i>Pseudocode 6. Kode program tahapan stopwords removal</i> .....	33
<i>Pseudocode 7. Kode program klasifikasi sentimen dengan corpus</i> <i>InSet Lexicon</i> .....	36

## I. PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan ilmu pengetahuan dan teknologi yang pesat banyak menghadirkan teknologi baru salah satunya teknologi pada bidang komunikasi dan informasi. Dampak dari perkembangan tersebut memberikan kemudahan mengakses informasi secara mudah dan cepat ke arah serba digital dengan memanfaatkan internet untuk mengakses situs media sosial yang banyak memberikan informasi terbaru. Berdasarkan data laporan dari *We are Social Hootsuite*, pengguna internet di Indonesia pada Januari 2021 mencapai 202.6 juta pengguna. Hal tersebut menunjukkan persentase kenaikan sebesar 15,5% pengguna dari tahun sebelumnya. Sedangkan pengguna media sosial di Indonesia pada Januari 2021 mencapai 170 juta pengguna dengan kenaikan sebesar 6.3% pengguna dari tahun sebelumnya.

Twitter merupakan salah satu media sosial yang dapat mengakses dan menyebarkan informasi. Menurut data laporan dari *Statista* pengguna twitter di Indonesia pada Januari 2022 mencapai 18.45 juta pengguna yang menjadikan Indonesia sebagai negara ke-5 pengguna Twitter terbanyak di dunia. Tidak hanya dapat mengakses dan menyebarkan informasi, Twitter sering digunakan sebagai sarana untuk pembelajaran, promosi, bahkan menyampaikan opini. Hal tersebut dapat dimanfaatkan oleh perusahaan, penjual maupun organisasi sebagai sarana untuk meninjau langsung opini pengguna terkait suatu layanan, produk, ataupun isu yang sedang terjadi guna meningkatkan atau memperbaiki hal tersebut.

Selain perkembangan teknologi pada bidang komunikasi dan informasi, pada bidang transportasi pun mengalami perkembangan dengan hadirnya layanan pesan antar makanan secara *online* yang dapat dipesan melalui sebuah aplikasi. Pada

penelitian yang berjudul *Anteseden Peningkatan Penggunaan Online Food Delivery*. Pada Masa Pandemi COVID-19 (Novita & Ari Wijaya, 2021) menjelaskan bahwa terjadinya peningkatan transaksi di layanan pesan antar sejak pandemi Covid-19 berlangsung dikarenakan berlakunya pembatasan aktivitas masyarakat di luar ruangan. Layanan pesan antar makanan yang dimaksud yakni GoFood dan GrabFood. GoFood merupakan salah satu layanan pesan antar makanan yang terdapat pada aplikasi Gojek sedangkan GrabFood merupakan salah satu layanan pesan antar makanan yang terdapat pada aplikasi Grab. GoFood dan GrabFood memiliki kelebihan dan kekurangan pada masing-masing layanan yang diberikan sehingga pengalaman yang dirasakan oleh tiap pengguna tentunya berbeda. Sebagian pengguna layanan GoFood dan GrabFood menyampaikan pendapat serta pengalamannya melalui media sosial, salah satunya Twitter.

*Sentiment Analysis* adalah riset komputasional dalam mengekstrak informasi yang bertujuan untuk menganalisa opini, sentimen, sikap, evaluasi, emosi, dan penilaian seseorang yang diekspresikan secara tekstual untuk melihat pendapat terhadap sebuah masalah pada suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu. Menurut Bing Liu (2015) pada bukunya yang berjudul *Sentiment Analysis Mining Opinions, Sentiments, and Emotions* terdapat tiga pendekatan dalam melakukan *sentiment analysis*, yaitu *machine learning approach*, *lexical based approach* dan *hybrid based learning*. Terdapat penelitian mengenai topik yang serupa menggunakan metode yang berbeda dengan judul Analisis Sentimen Terkait Layanan Gofood dan Grabfood pada Media Sosial Twitter Menggunakan Algoritma *Support Vector Machine* (SVM) (Agsar, et al., 2021). Penelitian ini menghasilkan informasi bahwa kedua layanan tersebut, Gofood dan Grabfood memiliki pelayanan yang baik menurut penggunanya karena didominasi oleh sentimen positif. Untuk nilai *accuracy* yang didapatkan sebesar 79,26%, *recall* sebesar 86,23%, dan *specificity* sebesar 56%. Adapun penelitian sebelumnya yang berjudul Klasifikasi Tweets Masyarakat yang Membicarakan Layanan GoFood dan GoRide pada GoJek Dimedia Sosial Twitter Selama Masa Kenormalan Baru (*New Normal*) dengan Metode Naïve Bayes

(Nabhan, et al., 2021) menghasilkan nilai *accuracy* sebesar 94%, *precision* sebesar 73%, *recall* sebesar 72%, dan *f-measure* sebesar 72%. *Lexicon based* merupakan salah satu pendekatan yang digunakan dalam penelitian analisis sentimen yang didasarkan pada sebuah kamus kata atau *corpus* yang dilengkapi dengan bobot pada tiap kata. Kualitas hasil yang didapatkan menggunakan pendekatan ini tergantung pada kamus kata yang digunakan. Terdapat beberapa kamus kata pada *Lexicon based* diantaranya SentiWordNet, Liu Lexicon, AFINN Lexicon, Pencari Opini, Senti-Strength, HBE Lexicon, dan juga NRC Emotion Lexicon. Namun kamus kata tersebut masih berbentuk dalam Bahasa Inggris. Adapun kamus kata yang sudah berbentuk Bahasa Indonesia yakni *Vania Lexicon* dan *InSet Lexicon*. Pada penelitian yang berjudul *InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs* (Koto & Rahmaningtyas, 2017) membangun sebuah kamus kata baru ke dalam Bahasa Indonesia yang diberi nama *InSet Lexicon*. Penelitian ini juga membandingkan hasil akurasi yang diperoleh *InSet Lexicon* dengan *Vania Lexicon*, *SentiWordNet*, *Liu Lexicon*, dan *AFINN Lexicon*. Hasil akurasi tertinggi didapatkan dengan menggunakan kamus kata *InSet Lexicon* dengan perolehan akurasi sebesar 65,78%.

Berdasarkan uraian latar belakang di atas, penelitian ini berfokus pada *Sentiment Analysis* Layanan Pesan Antar Makanan Pada Twitter Menggunakan *InSet Lexicon* (Studi Kasus: GoFood dan GrabFood). Data yang digunakan pada penelitian ini berasal dari Twitter dalam Bahasa Indonesia dengan memanfaatkan Twitter API. Kamus kata yang akan digunakan yaitu *InSet Lexicon*.

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, permasalahan yang akan diuraikan dan diteliti sebagai berikut :

1. Bagaimana cara menentukan klasifikasi sentimen dari dataset menggunakan *InSet Lexicon*?
2. Bagaimana hasil pengujian klasifikasi dari *InSet Lexicon* dan hasil pengujian dari metode *K-Nearest Neighbors* pada dataset.

### 1.3 Tujuan Penelitian

Tujuan dari penelitian ini berdasarkan pemaparan rumusan masalah di atas sebagai berikut :

1. Mengimplementasikan metode *Lexicon Based* menggunakan *corpus InSet Lexicon* pada *sentiment analysis*.
2. Menganalisis dan mengidentifikasi opini dari dataset dengan menggunakan *corpus InSet Lexicon*
3. Mengetahui hasil pengujian klasifikasi dari perbandingan *gold standard data* dengan klasifikasi *InSet Lexicon*.
4. Mengetahui hasil pengujian klasifikasi terbaik pada nilai K terbaik dari *gold standard data* dengan *K-Nearest Neighbors* pada *sentiment analysis* terhadap dataset.
5. Membandingkan evaluasi model dari hasil pengujian menggunakan *gold standard data* dan *InSet Lexicon* dengan hasil pengujian menggunakan *gold standard data* dan *K-Nearest Neighbors*.

### 1.4 Batasan Masalah

Dalam penelitian ini perlu adanya batasan masalah supaya penelitian lebih terarah dan sesuai dengan tujuan yang diinginkan. Adapun batasan masalah yang diberikan sebagai berikut :

1. Data yang akan dianalisis merupakan *tweet* berbahasa Indonesia yang diperoleh dari situs media sosial Twitter dengan menggunakan API Twitter dalam kurung waktu 1 Februari – 21 Februari 2022.
2. Pengujian hasil klasifikasi menggunakan metode *Lexicon Based* dengan *corpus InSet Lexicon* dan metode *K-Nearest Neighbors* dengan variasi *grid search cross validation*.
3. *Corpus* atau kamus kata yang digunakan yakni *InSet Lexicon*.
4. Klasifikasi opini hanya berupa sentimen positif dan sentimen negatif. .
5. Data akan dianalisis menggunakan bahasa pemrograman *python*.
6. Data yang sudah dianalisis tidak diimplementasikan pada suatu aplikasi.

7. *Query* atau kata kunci pada pencarian *tweet* yaitu gofood, grabfood, goput, grabput, gopud, grabpud, gofut, grabfut, gofud, dan grabfud.

### **1.5 Manfaat Penelitian**

Manfaat yang diharapkan pada penelitian ini yaitu :

1. Dapat mengetahui perbandingan kualitas layanan yang diberikan dari GoFood dan GrabFood.
2. Mengetahui kinerja dari tiap metode yang diujikan.
3. Diharapkan mampu menjadi acuan pembelajaran dan pengembangan penelitian selanjutnya mengenai *sentiment analysis*.



## II. TINJAUAN PUSTAKA

### 2.1 Penelitian Sebelumnya

Pada Tabel 1 yang merupakan beberapa penelitian terdahulu terkait *sentiment analysis* yang dijadikan referensi dalam penelitian yang akan dilakukan. Dari penelitian tersebut dapat disimpulkan bahwa pada penelitian ini yang berjudul *Sentiment Analysis Layanan Pesan Antar Makanan Pada Twitter Menggunakan InSet Lexicon (Studi Kasus : GoFood dan GrabFood)* memiliki kekuatan pada *corpus* yang akan digunakan. InSet Lexicon merupakan kamus kata yang sudah berbentuk dalam Bahasa Indonesia dan tiap katanya sudah memiliki *polarity score*, hal tersebut membuat pengklasifikasian lebih efektif. Namun kelemahan pada penelitian ini juga terletak pada *corpus* yang akan digunakan. Kamus yang digunakan sangat berpengaruh pada hasil klasifikasi maupun pengujian. Untuk mengetahui seberapa baik hasil pengujian yang diperoleh dari klasifikasi *InSet Lexicon* maka dilakukan perbandingan hasil pengujian dari pengujian menggunakan metode *K-Nearest Neighbors*.

Tabel 1. Penelitian terdahulu.

<b>Penulis</b>	<b>Judul</b>	<b>Metode</b>	<b>Output</b>
Ipmawati , et al., (2017)	Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen	SVM, Naïve Bayes, KNN	<p><i>Accuracy</i> yang diperoleh pada tiap metode dari dataset Twitter yakni</p> <ol style="list-style-type: none"> <li>1. SVM : 72%</li> <li>2. Naïve Bayes : 67,33%</li> <li>3. KNN : 56,83%</li> </ol> <p><i>Accuracy</i> yang diperoleh dari dataset IMDB review film yakni</p> <ol style="list-style-type: none"> <li>1. SVM : 78,55%</li> <li>2. Naïve Bayes : 78,55%</li> <li>3. KNN : 56,7%</li> </ol>
Septian, A, et al., (2019)	Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor	KNN	<p>Dataset : 2.000 tweet</p> <p><i>Accuracy</i> yang diperoleh sebesar 80% yang didapat dari nilai <math>K = 23</math> yang diuji dari range <math>K = 1</math> hingga <math>K = 30</math>.</p>

Penulis	Judul	Metode	Output
Koto, F., & Rahmaningtyas, G. Y. (2017)	InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs	<i>Lexicon based</i> dengan <i>corpus</i> InSet Lexicon	InSet adalah sentiment bahasa Indonesia dengan 10.218 kata yang terdiri dari 6.609 kata negatif dan 3.609 kata positif. Untuk menguji InSet, dilakukan pengujian dengan mengumpulkan 1.259 tweet positif dan 1.371 tweet negative. Pengujian dilakukan pada beberapa leksikon bahasa Inggris yang diterjemahkan yakni SentiWordNet, Liu Lexicon, dan AFINN Lexicon dan juga leksikon berbahasa Indonesia yaitu Vania Lexicon. Akurasi tertinggi dimiliki oleh InSet Lexicon dengan akurasi sebesar 65,78% lebih unggul dari Vania Lexicon yang memiliki akurasi sebesar 61,48%.
Musfiroh et.al., (2021)	Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon	Lexicon Based menggunakan InSet Lexicon.	<p>Dataset : 6.000 tweet.</p> <p>Data latih : data uji = 80% : 20%, dengan hasil klasifikasi yang didapat,</p> <ul style="list-style-type: none"> <li>• Sentimen negatif sebesar 63,6%</li> <li>• Sentimen positif sebesar 27,6%</li> <li>• Sentimen netral sebesar 8,9%</li> </ul> <p>Hasil pengujian yang didapat:</p> <ul style="list-style-type: none"> <li>• Akurasi sebesar 79,2%</li> <li>• <i>Precision</i> sebesar 72,9%</li> <li>• <i>Recall</i> sebesar 62,8%</li> <li>• <i>F-measure</i> sebesar 67,4%</li> </ul>

<b>Penulis</b>	<b>Judul</b>	<b>Metode</b>	<b>Output</b>
Undap, et al., (2021)	Analisis Sentimen Situs Pembajak Artikel Penelitian Menggunakan Metode <i>Lexicon-Based</i>	Lexicon based dengan kamus InSet Lexicon	Dataset : 37.754 tweet Data latih : data uji = 90%:10% Hasil klasifikasi yang didapatkan : <ul style="list-style-type: none"> <li>• Sentimen positif sebesar 59,51%</li> <li>• Sentimen negatif sebesar 26,42%</li> <li>• Sentimen netral 14,07%</li> </ul> Hasil pengujian yang didapatkan : <ul style="list-style-type: none"> <li>• Akurasi sebesar 55,6% dengan 1.730 data diprediksi benar oleh <i>classifier</i> dari 3.108 data</li> </ul>

## 2.2 Landasan Teori

### 2.2.1 *Natural Language Processing*

*Natural language processing* atau biasa disebut NLP merupakan teknik untuk menganalisis dan menggambarkan teks dalam satu atau lebih tingkat analisis linguistik yang bertujuan untuk mencapai manusia seperti pengolahan bahasa untuk berbagai tugas dan aplikasi. Selain itu NLP juga disebut sebagai salah satu pengembangan teknik komputasi bahasa alami dalam menganalisis dan merepresentasikan teks ataupun lisan untuk mencapai pemahaman bahasa seperti bahasa manusia. NLP yang memiliki tujuan untuk memahami bahasa manusia telah diaplikasikan pada beberapa bidang (Pustejovsky dan Stubbs, 2012) diantaranya sebagai berikut:

1. *Question Answering System (QAS)*, kemampuan komputer untuk menjawab pertanyaan yang diajukan oleh pengguna. Pengguna dapat langsung bertanya dalam bahasa yang digunakan tanpa memasukkan kata kunci ke dalam mesin pencarian.
2. *Summarization*, pembuatan rangkuman dari sekumpulan dokumen dengan mengkonversikan dokumen teks berukuran besar ke bentuk *slide* presentasi.
3. *Machine translation*, kemampuan memahami bahasa manusia dan menterjemahkannya ke dalam bahasa lain. Contohnya pada Google Translate.
4. *Speech recognition*, kemampuan mengenali bahasa yang diucapkan berupa pertanyaan atau perintah. *Speech recognition* merupakan cabang ilmu NLP yang cukup sulit.
5. *Document classification*, kemampuan mengidentifikasi kategori dokumen yang harus dimasukkan. Hal ini berguna pada aplikasi penyaringan spam, artikel berita klasifikasi, dan ulasan film.

### 2.2.2 *Sentiment Analysis*

*Sentiment analysis* disebut juga sebagai *opinion mining* merupakan kegiatan menganalisa topik atau tren tertentu berupa teks terhadap entitas seperti layanan,

produk, dan organisasi, agar mendapatkan suatu informasi. *Sentiment analysis* merupakan topik penelitian pada *Natural Language Processing* yang bertujuan untuk membangun metode yang dapat digunakan untuk mengekstraksi informasi subjektif berupa opini dalam sebuah data teks. Analisis sentiment cenderung fokus pada sentiment yang memiliki nilai positif dan negatif (Liu, 2016). Tujuan lain dilakukannya *sentiment analysis* untuk mengetahui opini seseorang terhadap permasalahan ataupun isu terkait suatu topik tertentu apakah sebagai opini atau sentimen positif, negatif, atau netral yang dapat dijadikan sebagai acuan dalam meningkatkan suatu layanan ataupun kualitas produk. Terdapat tiga pendekatan dalam melakukan *sentiment analysis* diantaranya sebagai berikut.

### **2.2.3 GoFood**

GoFood merupakan salah satu layanan yang terdapat di dalam aplikasi GoJek yakni layanan pesan antar makanan dari Indonesia yang bekerja sama dengan lebih dari 125.000 restoran di beberapa kota di Indonesia yang telah secara resmi menjadi *Partner* GoFood. GoFood beroperasi selama 24 jam dalam 7 hari. GoFood menerima berbagai macam metode pembayaran baik berupa tunai, GoPay, PayLater, maupun LinkAja. Banyak promosi yang diberikan oleh GoFood, baik berupa voucher potongan harga dengan minimum pembelian yang didapat tiap minggunya maupun didapat langsung dari *Partner* GoFood.

### **2.2.4 GrabFood**

GrabFood merupakan salah satu layanan pesan antar makanan yang terdapat di dalam aplikasi Grab. GrabFood beroperasi selama 24 jam dalam 7 hari. GrabFood menerima berbagai macam metode pembayaran untuk pemesanan makanan baik berupa tunai, OVO, kartu kredit, kartu debit, maupun PayPal. Banyak promosi yang diberikan oleh GrabFood, baik berupa voucher gratis ongkir, potongan harga dengan minimum pembelian, maupun potongan harga pada jam – jam tertentu. Hal tersebut menjadi daya tarik bagi pengguna untuk menggunakan layanan tersebut tanpa harus keluar rumah makanan pun sampai.

### 2.2.5 *Twitter*

Twitter merupakan situs jejaring sosial dan mikroblog daring yang dioperasikan oleh Twitter, Inc sehingga memungkinkan pengguna untuk mengirim dan membaca *tweet*. Mikroblog adalah salah satu jenis alat komunikasi online yang dapat memperbarui sebuah status dan mempublikasikannya untuk dilihat semua orang atau beberapa orang tertentu yang dipilih pengguna. *Tweet* merupakan pesan singkat yang terdiri dari 280 karakter. Melalui Twitter API (*Application Programming Interface*) keterbukaan terhadap data yang dimiliki oleh Twitter memungkinkan pengembang pihak ketiga untuk dapat mengambil data dan mengolahnya tanpa sepengetahuan pengguna. API menyamai *Software as a Service (SaaS)* karena perangkat lunak ini memudahkan pengembang jika ingin menggunakannya dan tidak perlu memulai dari awal untuk menulis program..

### 2.2.6 *InSet Lexicon*

*InSet Lexicon (Indonesian Sentimen Lexicon)* merupakan kamus kata atau *corpus* yang disusun oleh Fajri Koto dan Gemal Y. Rahamaningtyas pada penelitiannya pada tahun 2017 dengan judul *InSet Lexicon: Evaluation of a word list for Indonesian Sentiment Analysis in microblogs* dengan menggunakan kata – kata yang dikumpulkan dari Twitter sebagai media sosial yang umum digunakan di Indonesia. *InSet Lexicon* terdiri dari 3.609 kata positif dan 6.609 kata negatif berbahasa Indonesia yang memiliki bobot nilai pada tiap katanya dengan kisaran nilai antara -5 sampai +5. *Polarity score* digunakan untuk mengklasifikasikan jenis sentimen. Tiap *tweet* yang terdapat pada dataset akan dicocokkan dengan kamus kata *InSet Lexicon* untuk dilakukan perhitungan *polarity score* pada setiap *tweet*. Proses perhitungan *polarity score* dengan menjumlahkan keseluruhan bobot dari kata yang terdeteksi pada tiap *tweet* dan hasil akhir dari pembobotan akan diklasifikasikan ke dalam sentimen positif maupun sentimen negatif. Jika bobot yang didapatkan lebih besar dari 0 maka *tweet* tersebut masuk ke dalam kategori sentimen positif. Jika bobot yang diperoleh kurang dari 0 maka *tweet* tersebut tergolong kategori sentimen negatif. Tabel 2 di bawah ini merupakan beberapa contoh kata beserta bobotnya yang terdapat dalam *InSet Lexicon*.

Tabel 2. Contoh daftar kata dalam kamus InSet Lexicon.

<b>Kata</b>	<b>Bobot</b>
Maksa	-5
Jauh	-4
Doktrin	-3
Asam	-2
Mencair	-1
Mahal	1
Irit	2
Murah	3
Ayo	4
Ramah	5

Adapun cara kerja dari metode *Lexicon Based* dengan *corpus InSet Lexicon* dengan menjumlahkan bobot pada tiap kata yang terdapat dalam *tweet*. Tabel 3 merupakan contoh perhitungan *polarity score* pada salah satu *tweet* menggunakan *corpus InSet Lexicon*.

Tabel 3. Contoh perhitungan *polarity score* menggunakan *InSet Lexicon*.

<i>Tweet</i>	<i>Polarity score</i>	<i>polarity</i>
'pensiun', 'gofood', 'diskon', 'tidak', 'ada', 'jelas' (-3 + 0 + 2 + (-5) + (-3) + 1)	- 8	<i>negative</i>
'kenapa', 'starbucks', 'grabfood', 'lebih', 'murah', 'banding', 'langsung', 'beli', 'toko' (-3 + 0 + 0 + 1 + 3 + 3 + 3 + 2 + 0)	9	<i>positive</i>

### 2.2.7 Pemodelan Bahasa *Term Frequency-Inverse Document Frequency*

*Term Frequency-Inverse Document Frequency* adalah salah satu metode pembobotan kata dalam tiap dokumen teks yang menggabungkan *term frequency* dengan *inverse document frequency* (B. Trstenjak, et. al, 2014). *Term frequency* didefinisikan dengan kemunculan kata dalam suatu dokumen sedangkan *inverse document frequency* merupakan jumlah seluruh dokumen yang berisikan kata



tertentu. Persamaan (2.1) merupakan perhitungan total bobot dari kata pada tiap dokumen dalam TF-IDF.

$$tf_{td}idf_t = tf_{td} * \log\left(\frac{N}{df_t}\right) \quad (2.1)$$

Keterangan

$tf_{td}$  = Jumlah kemunculan kata t dalam suatu dokumen

N = Total dokumen

$df_t$  = Jumlah dari seluruh dokumen yang mengandung kata t

$tf * idf$  = Bobot total dari kata t

### 2.2.8 K-Nearest Neighbors

*K-Nearest Neighbors* (KNN) merupakan algoritma klasifikasi *supervised learning* yang dilakukan dengan melihat k tetangga terdekatnya, dengan k merupakan banyaknya tetangga (B, Trstenjak et al., 2014). Metode ini bekerja dengan hanya menghafal data *training* yang tersedia selama proses pelatihan. KNN sendiri tergolong metode yang sederhana untuk pengklasifikasian tanpa harus melakukan perhitungan secara kompleks, karena itu KNN biasa disebut *lazy learning* (S, Jeremy Andre, et al., 2019). Persamaan (2.2) merupakan perhitungan jarak antar titik, dengan titik yang dimaksud yakni data *train* dan data *test* digunakan rumus *eucliden distance* (Kursini & Luthfi, 2009).

$$d = \sqrt{\sum_{i=1}^N (X_{2i} - x_{1i})^2} \quad (2.2)$$

Keterangan

d = jarak antar titik

i = variable

$x_1$  = data *training*

$x_2$  = data *testing*

### 2.2.9 Grid Search Cross Validation

*Grid search cross validation* merupakan proses menentukan *hyperparameter* terbaik pada suatu model dengan melakukan kombinasi *hyperparameter*

yang telah ditentukan dan menghitung *average* dari nilai *cross validation*. *Hyperparameter* sendiri didefinisikan sebagai variabel yang nilainya tidak ditentukan oleh mesin dengan implementasi nilai terbaiknya akan digunakan pada model (Ailiyya, 2020).

### 2.2.10 Confusion Matrix

Untuk menghitung performa pengujian hasil klasifikasi akan digunakan *confusion matrix*. *Confusion matrix* didefinisikan sebagai metode yang digunakan untuk melakukan perhitungan nilai performansi dalam suatu model pada konsep data mining (Rosandy, 2016).

Tabel 4. *Confusion Matrix*.

		Prediksi	
		Positif	Negatif
Aktual	Positif	TP ( <i>True Positive</i> )	FN ( <i>False Negative</i> )
	Negatif	FP ( <i>False Positive</i> )	TN ( <i>True Negative</i> )

*True Positive* merupakan jumlah *tweet* positif yang dideteksi benar oleh sistem. *True Negative* merupakan jumlah *tweet* negatif yang dideteksi benar oleh sistem. *False Positive* merupakan jumlah *tweet* negatif yang dideteksi sebagai *tweet positif* oleh sistem. Sedangkan *false negative* merupakan jumlah *tweet* positif yang dideteksi sebagai *tweet* negatif oleh sistem. Hasil yang didapat pada kolom matriks di atas maka akan dapat diketahui nilai *accuracy*, *precision*, *recall/sensitivity*, maupun *f1-score*

- a. *Accuracy*, diperoleh dari perbandingan antara data yang berhasil diklasifikasikan secara benar dengan keseluruhan data. Akurasi menunjukkan kedekatan hasil klasifikasi dengan nilai sesungguhnya. Persamaan (2.3) merupakan persamaan untuk mencari nilai *accuracy* berdasarkan hasil *confusion matrix* yang diperoleh.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.3)$$

- b. *Precision*, hasil klasifikasi positif yang benar dengan keseluruhan data yang diprediksikan sebagai kelas positif. Nilai *precision* dapat mengetahui pemrosesan yang relevan terhadap informasi yang ingin dicari dan kedekatan hasil antara informasi yang diminta dengan jawaban yang diberikan. Persamaan (2.4) merupakan persamaan untuk menghitung nilai *precision* berdasarkan hasil *confusion matrix* yang diperoleh.

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2.4)$$

- c. *Recall* atau *sensitivity*, jumlah dokumen yang memiliki klasifikasi positif yang benar dari semua dokumen yang benar-benar positif. Persamaan (2.5) merupakan persamaan untuk menghitung nilai *recall* berdasarkan hasil *confusion matrix* yang diperoleh.

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2.5)$$

- d. *F1-score*, perhitungan yang mendapatkan informasi dengan memadukan nilai *recall* dan *precision*. Persamaan (2.6) merupakan persamaan untuk menghitung *f1-score* berdasarkan hasil *confusion matrix* yang diperoleh.

$$F1-score = \frac{2 \times recall \times precision}{recall + precision} \times 100\% \quad (2.6)$$

### III. METODE PENELITIAN

#### 3.1 Waktu dan Tempat Penelitian

Penelitian dilaksanakan di Jurusan Ilmu Komputer Universitas Lampung yang berlokasi di Jl. Prof. Dr. Ir. Sumantri Brojonegoro Nomor 1, Gedong Meneng, Kecamatan Rajabasa, Kota Bandar Lampung. Penelitian dimulai pada bulan Januari 2022 dan diperkirakan selesai pada bulan Desember 2022.

#### 3.2 Alat Penelitian

Terdapat dua jenis alat yang digunakan pada penelitian ini yaitu perangkat lunak dan perangkat keras. Adapun spesifikasi alat yang digunakan selama penelitian sebagai berikut :

##### 3.2.1 Perangkat Lunak

Perangkat lunak yang digunakan dalam penelitian ini antara lain :

- a. Sistem operasi Windows 10
- b. Microsoft Word
- c. Microsoft Excel
- d. Google Chrome
- e. Anaconda3
- f. *Developer Twitter*
- g. Jupyter Notebook 6.4.5
- h. *Library python*
  1. *Tweepy* : *Library* untuk mengakses dan mengambil data dari API *Twitter*.
  2. *Scikit-learn* : *Library* untuk pembangunan model pembelajaran mesin.
  3. *Regular Expression* : *Library* untuk melakukan *preprocessing* data.

4. NLTK : *Library* untuk melakukan tahapan *tokenizing*.
5. Sastrawi : *Library* untuk melakukan tahapan *stopword removal*.
6. *Pandas* : *Library* untuk memproses dataset.
7. *Numpy* : *Library* untuk memproses komputasi numerik.
8. *Matplotlib.pyplot* : *Library* untuk melakukan visualisasi hasil data analisis.
9. *TfidfVectorizer* : *Library* untuk memproses pemodelan bahasa pada tiap kata yang muncul pada dokumen.
10. *KNeighborsClassifier* : *Library* untuk meingplementasikan metode *K-Nearest Neighbors* pada analisis.

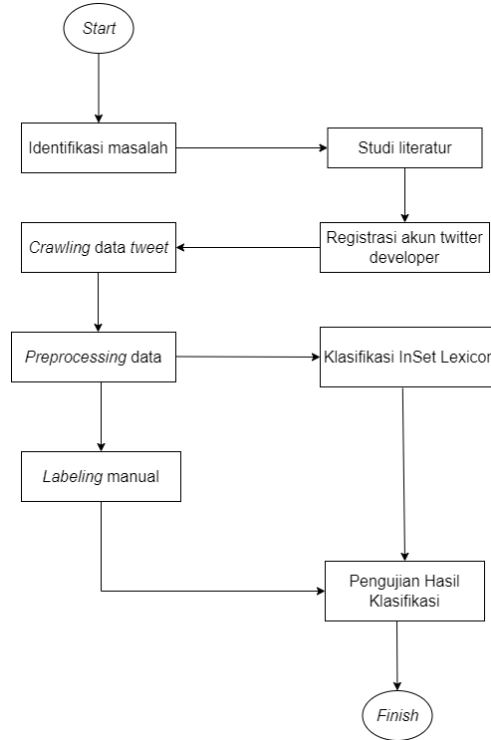
### 3.2.2 Perangkat Keras

Perangkat keras yang digunakan dalam penelitian ini antara lain :

- a. Laptop asus Vivobook
  1. *Processor* : Intel® Core™ i5-10210U CPU @1.60GHz 2.11GHz
  2. RAM : 8,00 GB

### 3.3 Alur Penelitian

Pada Gambar 1 merupakan tahapan penelitian yang akan dilakukan untuk memperoleh hasil analisis yang diinginkan.



Gambar 1. Tahapan Penelitian.

Berikut penjelasan dari masing – masing tahapan yang akan dilakukan pada penelitian ini.

#### 3.3.1 Identifikasi Masalah

Pada tahap awal penelitian yaitu mengidentifikasi masalah. Tahapan ini menganalisis metode dan studi kasus yang akan digunakan pada penelitian ini. Metode yang digunakan merupakan Lexicon Based dengan InSet Lexicon sebagai *corpus* atau kamus kata karena InSet Lexicon merupakan kamus kata yang sudah berbentuk ke dalam Bahasa Indonesia selain itu menurut Fajri Koto dan Gemala Y. Rahmaningtyas pada tahun 2017 dengan penelitiannya yang berjudul *InSet Lexicon: Evaluation of a Word List For Indonesian Sentiment Analysis in Microblogs* menunjukkan bahwa InSet Lexicon menghasilkan tingkat akurasi tertinggi sebesar 65,78% dibandingkan *corpus* yang lain, yakni Vania Lexicon, SentiWordNet, Liu Lexicon, dan AFINN

Lexicon. Studi kasus pada penelitian ini membahas tentang layanan pesan antar makanan terbesar di Indonesia yakni GoFood dan GrabFood.

### **3.3.2 Studi Literatur**

Tahapan ini merupakan tahap dalam mengumpulkan dan mempelajari informasi dari penelitian terdahulu yang memiliki kaitannya dengan topik yang akan diteliti. Informasi yang berhasil dikumpulkan dapat berupa buku, jurnal, maupun situs laporan yang berhubungan dengan topik penelitian.

### **3.3.3 Membuat Akun Twitter Developer**

Pada tahap ini melakukan registrasi akun pada Twitter developer supaya dapat melanjutkan tahapan *crawling* data *tweet* atau pengumpulan data. *Crawling* data membutuhkan *access tokens* berupa *api\_key*, *api\_secret\_key*, *access\_token*, dan *access\_token\_secret* untuk dapat mengakses API Twitter Search.

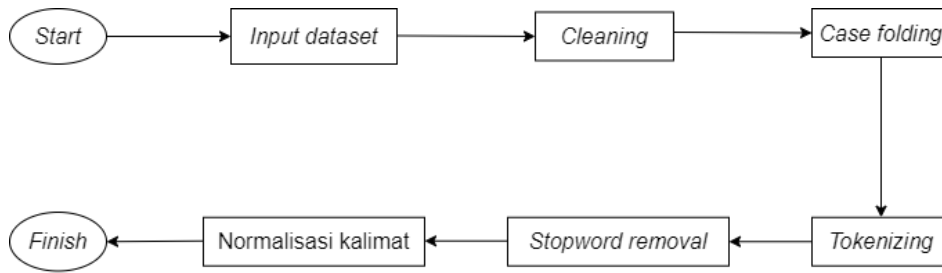
### **3.3.4 Crawling Data Tweet**

Tahapan *crawling* data *tweet* dari media sosial Twitter menggunakan *library tweepy* dengan bahasa pemrograman *python*. *Query* atau kata kunci yang digunakan dalam proses *crawling* data *tweet* berupa *gofood*, *goput*, *gopud*, *gofut*, *gofud*, *grabfood*, *grabput*, *grabpud*, *grabfut*, dan *grabfut*. Data yang akan dianalisis berupa teks yang berisikan *tweet* pengguna terkait pengalaman menggunakan layanan GoFood dan GrabFood. *Tweet* diambil pada rentang waktu 01 Februari – 21 Februari 2022 dikarenakan pada waktu tersebut terjadi gelombang COVID-19 kedua dengan munculnya varian *omicron* di Indonesia.

### **3.3.5 Preprocessing data**

Pada tahapan ini data yang berhasil dicrawling merupakan data mentah yang banyak mengandung *noise*. Data mentah tersebut harus dilakukan pembersihan melalui tahapan *preprocessing*, jika tidak melakukan tahapan ini maka akan menurunkan tingkat klasifikasi dan performansi dari sistem yang

dihasilkan. Adapun Gambar 2 merupakan alur dari tahapan *preprocessing* pada penelitian yang akan dilakukan.



Gambar 2. Tahapan *preprocessing*.

### a. *Cleaning*

Data mentah yang diperoleh masih berupa *tweet* dengan kalimat yang belum terstruktur. Proses *cleaning* dilakukan dengan cara membersihkan *tweet* dari karakter yang tidak diperlukan dan menghapus data yang duplikat. Karakter yang dimaksud berupa *mention username* (@username), kata kunci, Retweet, *hashtag* Twitter, *link* tautan, emotikon, simbol, angka, dan tanda baca. Pada Tabel 5 merupakan proses *cleaning* pada dataset.

Tabel 5. Proses *cleaning*.

<i>Input</i>	<i>Output</i>
RT FESS lagi promo nih cm 12rban aja aku beli pake gofood 2 cm 19rb kalo pake promo gopay lebih murah lagi kali	RT FESS lagi promo nih cm rban aja aku beli pake gofood cm rb kalo pake promo gopay lebih murah lagi kali

### b. *Case folding*

Proses *case folding* merupakan proses mengubah keseluruhan *tweet* menjadi *lower case*. Tabel 6 merupakan proses *case folding* pada dataset.

Tabel 6. Proses *case folding*.

<i>Input</i>	<i>Output</i>
RT FESS lagi promo nih cm rban aja aku beli pake gofood cm rb kalo pake promo gopay lebih murah lagi kali	rt fess lagi promo nih cm rban aja aku beli pake gofood cm rb kalo pake promo gopay lebih murah lagi kali



### c. Tokenizing

Proses *tokenizing* merupakan tahap pemenggalan kalimat menjadi kata per kata pada *tweet*. Proses ini dilakukan supaya memudahkan dalam menghitung frekuensi kemunculan kata dalam *corpus*. Tabel 7 merupakan proses *tokenizing* pada dataset.

Tabel 7. Proses *tokenizing*.

<i>Input</i>	<i>Output</i>
rt fess lagi promo nih cm rban aja	'rt', 'fess', 'lagi', 'promo',
aku beli pake gofood cm rb kalo	'nih', 'cm', 'rban', 'aja',
pake promo gopay lebih murah	'aku', 'beli', 'pake', 'gofood',
lagi kali	'cm', 'rb', 'kalo', 'pake',
	'gopay', 'lebih',
	'murah', 'lagi', 'kali'

### d. Stopword removal

Pada tahapan ini juga *stopword* akan dihilangkan karena terdiri dari daftar kata umum yang dianggap tidak memiliki makna seperti “yang, ke, di, dsb”.

Tabel 8 merupakan proses dari *stopword removal* pada dataset.

Tabel 8. Proses *stopword removal*.

<i>Input</i>	<i>Output</i>
'rt', 'fess', 'lagi', 'promo',	'lagi', 'promo',
'nih', 'cm', 'rban', 'aja',	'nih', 'cm', 'rban',
'aku', 'beli', 'pake', 'gofood',	'aku', 'beli', 'pake', 'gofood',
'cm', 'rb', 'kalo', 'pake',	'cm', 'rb', 'kalo', 'pake',
'gopay', 'lebih',	'gopay', 'lebih',
'murah', 'lagi', 'kali'	'murah', 'lagi', 'kali'

### e. Normalisasi Kalimat

Normalisasi kalimat merupakan proses untuk memperbaiki kata ejaan yang disingkat supaya sesuai dengan Kamus Besar Bahasa Indonesia (KBBI).

Tabel 9 merupakan proses normalisasi kalimat pada dataset

Tabel 9. Proses normalisasi kalimat.

<i>Input</i>	<i>Output</i>
'lagi', 'promo',	'lagi', 'promo',
'nih', 'cm', 'rban',	'nih', 'hanya', 'ribuan',

'aku', 'beli', 'pake', 'gofood', 'cm', 'rb', 'kalo', 'pake', 'gopay', 'lebih', 'murah', 'lagi', 'kali'	'saya', 'beli', 'pakai', 'gofood', 'hanya', 'ribu', 'kalau', 'pakai', 'gopay', 'lebih', 'murah', 'lagi', 'kali'
---	--

### 3.3.6 Klasifikasi Sentimen

#### a. Pelabelan Manual

Pada tahap ini dilakukan pelabelan jenis sentimen secara manual pada tiap *tweet* yang sudah melewati tahapan *preprocessing*. Data yang akan dilakukan pelabelan manual berasal dari data bersih yang sudah melewati tahapan *preprocessing*. Tujuan dilakukannya pelabelan manual untuk mendapatkan *gold standar dataset* yang nantinya dijadikan acuan saat dilakukannya pengujian hasil klasifikasi. Nilai sentimen yang ditentukan pada tahapan ini terdiri dari sentimen positif, negatif, dan netral. Akan tetapi, hanya data *tweet* yang memiliki sentimen positif dan negatif saja yang akan dilanjutkan penggunaannya pada proses selanjutnya. Pelabelan secara manual dilakukan oleh tiga orang yang merupakan mahasiswa Universitas Lampung dengan sistem *vote* atau memilih jenis sentimen terbanyak jika terjadi pelabelan yang berbeda. Sentimen dengan nilai terbanyak maka menjadi nilai akhir dari pelabelan manual pada tiap *tweet*. Tabel 10 merupakan identitas dari tiga orang mahasiswa yang melakukan pelabelan manual pada dataset.

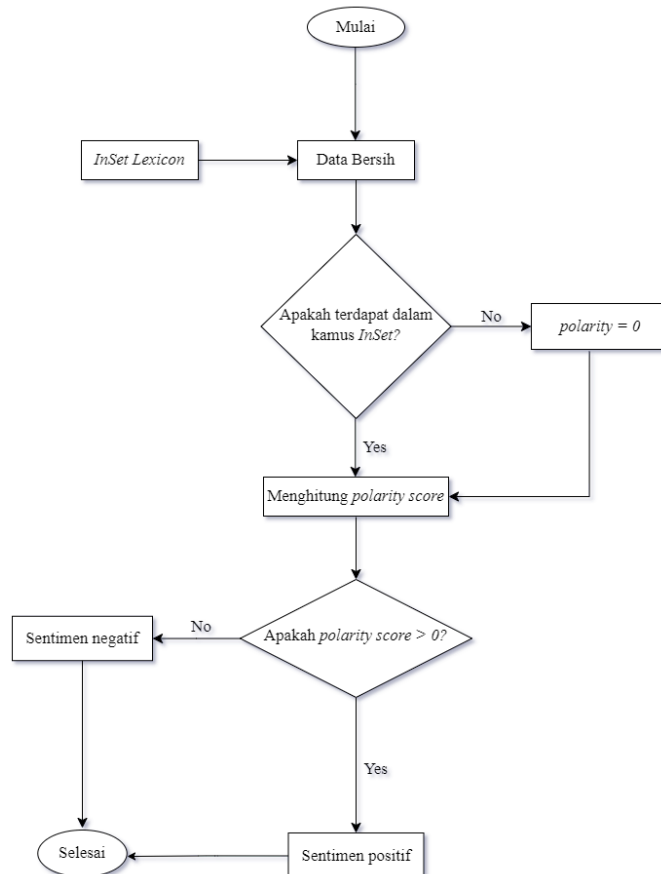
Tabel 10. Identitas mahasiswa pelabelan manual.

Nama	Instansi	Jurusan
Agnes Pramudani	Universitas Lampung	Ilmu Komputer
Evania Intiha	Universitas Lampung	Ilmu Komputer
Gema Annisa Hermastuti	Universitas Lampung	Ilmu Komputer

#### b. InSet Lexicon

Tidak hanya dilakukan klasifikasi secara manual, data bersih juga akan dilakukan klasifikasi menggunakan *InSet Lexicon*. *InSet Lexicon* memiliki kamus kata yang bersifat positif dan negatif dalam bentuk Bahasa Indonesia dengan disertai bobot dari tiap kata tersebut. Bobot dari tiap kata pada Inset

Lexicon berkisar pada nilai antara -5 sampai +5. Bobot atau *polarity score* tersebut digunakan untuk mengklasifikasikan jenis sentimen pada tiap *tweet*. Gambar 3 merupakan *flowchart* dari klasifikasi pada dataset menggunakan *InSet Lexicon*.



Gambar 3. *Flowchart* klasifikasi menggunakan *InSet Lexicon*.

Proses perhitungan *polarity score* dengan menjumlahkan keseluruhan bobot dari tiap kata yang terdeteksi pada *tweet* dan hasil akhir dari *polarity score* akan diklasifikasikan ke dalam sentimen positif maupun sentimen negatif. Secara umum dapat dinyatakan dengan algoritma sebagai berikut :

$$\text{If } \textit{polarity score} > 0 \text{ then Sentimen Positif} \quad (3.1)$$

$$\text{If } \textit{polarity score} < 0 \text{ then Sentimen Negatif} \quad (3.2)$$

Menentukan klasifikasi *tweet* ke dalam salah satu jenis sentimen ditentukan berdasarkan hasil akhir dari *polarity score* (*sentiment score*) yang diperoleh. Jika *sentiment score* yang diperoleh lebih dari 0 maka *tweet* tersebut tergolong sebagai sentimen positif dan jika *sentiment score* yang diperoleh kurang dari 0 maka *tweet* tersebut tergolong sebagai sentimen negatif (Musfiroh et.al., 2021).

### 3.3.7 Pengujian Hasil Klasifikasi

Pengujian hasil klasifikasi dibutuhkan untuk mengetahui tingkatan performa dari analisis penelitian ini yang menggunakan metode dari *Lexicon Based* dengan *corpus InSet Lexicon* dan *K-Nearest Neighbors*. Pengujian hasil klasifikasi pada penelitian ini menggunakan dua model pengujian, adapun dua model tersebut sebagai berikut.

#### a. Model 1

Pada model 1, dilakukan perhitungan evaluasi model dari hasil klasifikasi yang didapatkan dengan menggunakan *corpus InSet Lexicon* dengan hasil klasifikasi dataset secara manual. Perhitungan dilakukan dengan cara menghitung banyaknya data yang berhasil diprediksikan sebagai *true positive*, *true negative*, *false positive*, dan *false negative* oleh *InSet Lexicon* yang selanjutnya akan dihitung performa yang didapatkan menggunakan *confusion matrix*.

#### b. Model 2

Pengujian pada model 2 dilakukan dengan menggunakan hasil klasifikasi secara manual yang akan diuji pada metode *K-Nearest Neighbors* (KNN). Metode KNN digunakan untuk testing data dengan bantuan *grid search cross validation* guna mencari nilai hasil pengujian dari parameter terbaik dan menggunakan *Euclidian Distance* untuk menghitung jarak antar titik pada data *training* dan data *testing*. Selain itu, sebelum mencari nilai hasil pengujian dilakukan pembobotan kata terlebih dahulu dengan menggunakan TF-IDF.

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai analisis sentimen layanan pesan antar makanan GoFood dan GrabFood pada Twitter menggunakan *corpus InSet Lexicon*, maka berikut adalah kesimpulan pada penelitian ini.

1. *Corpus InSet Lexicon* dapat digunakan dalam menentukan klasifikasi teks pada *sentiment analysis*.
2. Hasil klasifikasi yang diperoleh secara pelabelan manual maupun menggunakan *corpus InSet Lexicon* memiliki perbedaan jumlah pada sentimen netral dan sentimen negatif. Hal tersebut terjadi dikarenakan cukup banyak *tweet* berbahasa Melayu yang diklasifikasikan menjadi sentimen negatif oleh *InSet Lexicon*. Sedangkan pada pelabelan manual *tweet* yang berbahasa Melayu diklasifikasikan menjadi sentimen netral karena pada penelitian yang dilakukan hanya menganalisis *tweet* yang berbahasa Indonesia.
3. Hasil pengujian menggunakan *InSet Lexicon* pada dataset GoFood diperoleh *accuracy* sebesar 71%, *precision* sebesar 81%, *recall* atau *sensitivity* sebesar 60%, dan *f1-score* sebesar 69%. Sedangkan untuk dataset GrabFood diperoleh *accuracy* sebesar 72%, *precision* sebesar 81%, *recall* atau *sensitivity* sebesar 640%, dan *f1-score* sebesar 72%.
4. Dari hasil klasifikasi yang diperoleh, selama bulan Februari 2022 GrabFood memberikan kualitas layanan yang lebih baik dibandingkan GoFood dikarenakan persentase dari sentimen positif pada GrabFood lebih unggul.
5. Hasil pengujian pada *gold standard data* dengan *K – Nearest Neighbors* pada dataset GoFood diperoleh hasil terbaik pada rasio pembagian data sebesar

80% data *training* dan 20% data *testing* dengan hasil *accuracy* sebesar 61%, *precision* sebesar 60%, *recall* atau *sensitivity* sebesar 56%, dan *f1-score* sebesar 54%. Hasil tersebut didapatkan dari nilai  $k = 1$ .

6. Hasil pengujian pada *gold standard data* dengan *K – Nearest Neighbors* pada dataset GrabFood diperoleh hasil terbaik pada rasio pembagian data sebesar 90% data *training* dan 10% data *testing* dengan hasil *accuracy* sebesar 65%, *precision* sebesar 67%, *recall* atau *sensitivity* sebesar 62%, dan *f1-score* sebesar 61%. Hasil tersebut didapatkan dari nilai  $k = 44$ .
7. *InSet Lexicon* memberikan hasil yang cukup baik dalam pengklasifikasian teks karena hasil yang diperoleh lebih unggul jika dibandingkan dengan salah satu metode *supervised learning* yang diujikan yakni *K – Nearest Neighbors* (KNN).

## 5.2 Saran

Berdasarkan dari hasil penelitian yang dilakukan, peneliti memiliki beberapa saran yang bisa menjadi masukan dan bahan pertimbangan untuk penelitian selanjutnya. Adapun saran tersebut sebagai berikut.

1. Penelitian selanjutnya dapat dikembangkan dengan menggunakan metode yang berbeda.
2. Objek pada penelitian selanjutnya diharapkan diambil pada waktu dan fenomena yang berbeda dari penelitian yang sudah dilakukan.
3. Dapat dilakukan pengembangan lebih lanjut dalam bentuk implementasi program.

## DAFTAR PUSTAKA

- Agsar, R., Anggoro, D., Habsi, R. Al, Valio, M. A., Widiastiwi, Y., Kom, S., Si, M., Chamidah, N., Kom, S., & Kom, M. (2021). *Analisis Sentimen Terkait Layanan Gofood dan Grabfood pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine ( SVM ).* September, 249–259.
- Azhar, Y. (2018). Metode Lexicon-Learning Based Untuk Identifikasi Tweet Opini Berbahasa Indonesia. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 6(3), 237.
- Asghar MZ, Kundi FM, Khan A, Ahmad S. Lexicon-Based Sentiment Analysis in the Social Web. *J Basic Appl Sci Res.* 2014;4(6):238-248.
- Hernikawati Balai Pengembangan Sumber Daya Manusia dan Penelitian Kominfo Jakarta, D., Pegangsaan Timur No, J., & Pusat, J. (2021). Kecenderungan Tanggapan Masyarakat Terhadap Vaksin Sinovac Berdasarkan Lexicon Based Sentiment Analysis The Trend of Public Response to Sinovac Vaccine Based on Lexicon Based Sentiment Analysis. *Jurnal Ilmu Pengetahuan Dan Teknologi Komunikasi*, 23(1), 21–31.
- Ipmawati, J., Kusriani, & Taufiq Luthfi, E. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. *Indonesian Journal on Networking and Security*, 6(1), 28–36.
- Koto, F., & Rahmaningtyas, G. Y. (2017). Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. *Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017, 2018-January(December)*, 391–394.

- Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. In *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*.
- Mahendrajaya, R., Buntoro, G. A., & Setyawan, M. B. (2019). *Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode SVM*.
- Musfiroh, D., Khaira, U., Eko, P., Utomo, P., & Suratno, T. (2021). Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(April), 24–33.
- Nabhan, A. A., Rahayudi, B., & Ratnawati, D. E. (2021). *Klasifikasi Tweets Masyarakat yang Membicarakan Layanan GoFood dan GoRide pada GoJek Dimedia Sosial Twitter Selama Masa Kenormalan Baru ( New Normal ) dengan Metode Naïve Bayes*. 5(7), 3018–3025.
- Nurjanah, Winda Estu., Perdana Rizal Setya., & Fauzi, M. (2017). Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN 2548(12)*.
- Novita., & Wijaya, Ari. (2021). Antecedent Peningkatan Penggunaan *Online Food Delivery* Pada Masa Pandemi COVID-19, 16(2), 441–452.
- Salim, Siti Saidah., & Mayary, Joanna. (2020). Analisis Sentimen Pengguna Twitter Terhadap Dompot Elektronik Dengan Metode Lexicon Based Dan K-Nearest Neighbor. *Jurnal Ilmiah Informatika Komputer* 25(1).
- Sumitro, P. A., Mulyana, D. I., & Saputro, W. (2021). Analisis Sentimen Terhadap Vaksin Covid-19 di Indonesia pada Twitter Menggunakan Metode Lexicon Based. *J-ICOM-Jurnal Informatika ...*, 02(02), 50–56.
- Undap, M. G., Rantung, V. P., & Rompas, P. T. D. (2021). *Analisis Sentimen Situs Pembajak Artikel Penelitian Menggunakan Metode Lexicon-Based*.
- We Are Social. (2021). Digital 2021. *Global Digital Insights*, 103.



Yerzi, F. S., & Sibaroni, Y. (2021). *Analisis Sentimen Terhadap Kebijakan Pemerintah Dalam Menangani Covid-19 Dengan Pendekatan Lexicon Based*. 8(5), 11354–11366.

Mti.binus.ac.id (2017) Sentiment Analysis Approaches and Methods  
<https://mti.binus.ac.id/2017/10/04/1900/> diakses pada 16 Februari 2021  
pukul 22.23 WIB

Statista.com (2022). Leading countries based on number of Twitter users as of Januari 2022 <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/> diakses pada 29 Maret 2022 pukul 21.21  
WIB