



TUGAS AKHIR - KS184822

**ANALISIS SENTIMEN PENGGUNA TWITTER  
MENGENAI “SEDOTAN PLASTIK” DENGAN  
METODE *K-NEAREST NEIGHBOR (KNN)* DAN  
*NEIGHBOR-WEIGHTED K-NEAREST NEIGHBOR  
(NWKNN)***

**SHINDY SARI UTAMI  
NRP 062115 4000 0005**

**Dosen Pembimbing  
Pratnya Paramitha Oktaviana, S.Si., M.Si.**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2019**





**TUGAS AKHIR - KS184822**

**ANALISIS SENTIMEN PENGGUNA TWITTER  
MENGENAI “SEDOTAN PLASTIK” DENGAN  
METODE *K-NEAREST NEIGHBOR (KNN)* DAN  
*NEIGHBOR-WEIGHTED K-NEAREST NEIGHBOR  
(NWKNN)***

**SHINDY SARI UTAMI  
NRP 062115 4000 0005**

**Dosen Pembimbing  
Pratnya Paramitha Oktaviana, S.Si., M.Si.**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2019**





**FINAL PROJECT - KS184822**

**TWITTER CENTIMENT ANALYSIS OF “PLASTIC STRAWS” USING K-NEAREST NEIGHBOR (KNN) AND NEIGHBOR-WEIGHTED K-NEAREST NEIGHBOR (NWKNN) METHODS**

**SHINDY SARI UTAMI  
NRP 062115 4000 0005**

**Supervisor  
Pratnya Paramitha Oktaviana, S.Si., M.Si.**

**UNDERGRADUATE PROGRAMME  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2019**



## LEMBAR PENGESAHAN

### ANALISIS SENTIMEN PENGGUNA TWITTER MENGENAI “SEDOTAN PLASTIK” DENGAN METODE *K-NEAREST NEIGHBOR (KNN)* DAN *NEIGHBOR-WEIGHTED K-NEAREST NEIGHBOR* (*NWKNN*)

#### TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Statistika  
pada  
Program Studi Sarjana Departemen Statistika  
Fakultas Matematika, Komputasi, dan Sains Data  
Institut Teknologi Sepuluh Nopember

Oleh :

**Shindy Sari Utami**

NRP. 062115 4000 0005

Disetujui oleh Pembimbing:

**Pratya Paramitha Oktaviana, S.Si., M.Si**

NIP. 1309201405001

( Aut )

Mengetahui,  
Kepala Departemen Statistika



**Dr. Suhartono**  
NIP. 19710929 199512 1 001

SURABAYA, JULI 2019





**ANALISIS SENTIMEN PENGGUNA TWITTER  
MENGENAI “SEDOTAN PLASTIK” DENGAN METODE  
K-NEAREST NEIGHBOR (KNN) DAN NEIGHBOR-  
WEIGHTED K-NEAREST NEIGHBOR (NWKNN)**

**Nama Mahasiswa** : Shindy Sari Utami  
**NRP** : 062115 4000 0005  
**Departemen** : Statistika-FMKSD-ITS  
**Dosen Pembimbing** : Pratnya Paramitha Oktaviana,  
S.Si., M.Si.

**Abstrak**

*Tahun 2017 Divers Clean Action, kelompok pemerhati lingkungan khususnya laut menyebutkan bahwa pemakaian sedotan di Indonesia mencapai 93.244.847 batang setiap harinya. Beberapa waktu terakhir, terdapat gerakan yang dibuat untuk menggugah kepedulian masyarakat akan lingkungan, yaitu gerakan untuk tidak menggunakan sedotan plastik. Gerakan anti sedotan plastik ini telah direspon oleh masyarakat pada media social twitter. Respon masyarakat diklasifikasikan kedalam sentimen positif dan negatif menggunakan metode K-Nearest Neighbor (KNN) serta Neighbor-Weighted K-Nearest Neighbor (NWKNN). Sentimen positif memiliki rasio 81% atau sebesar 2587 tweet, sedangkan sentimen negatif sebesar 19% atau 599 tweet. Metode NWKNN memberikan hasil lebih baik daripada metode KNN ketika nilai K masih bernilai kecil yaitu K=1 sampai K=3. Setelah K=4 hasil AUC metode NWKNN mengalami penurunan lebih banyak dibandingkan dengan metode KNN. Hasil visualisasi wordcloud sentimen positif terdapat kata “besi”, “bambu”, dan “kokop”. Wordcloud sentimen negatif terdapat kata “dingin”, “gigi”, dan “ngilu”.*

**Kata kunci:** Sedotan plastik, klasifikasi, KNN, NWKNN, Twitter

*(Halaman ini sengaja dikosongkan)*

# TWITTER SENTIMENT ANALYSIS OF “PLASTIC STRAWS” USING K-NEAREST NEIGHBOR (KNN) AND NEIGHBOR-WEIGHTED K-NEAREST NEIGHBOR (NWKNN) METHODS

**Name** : Shindy Sari Utami  
**Student Number** : 062115 4000 0005  
**Department** : Statistika-FMKSD-ITS  
**Supervisor** : Pratnya Paramitha Oktaviana,  
S.Si., M.Si.

## **Abstract**

*In 2017 Divers Clean Action, a group of environmentalists, especially the sea, said that the use of straws in Indonesia reached 93,244,847 sticks every day. For the last time, there was a movement that was made to arouse public concern about the environment, namely the movement not to use plastic straws. This action has been responded by the public on twitter social media. Community responses are classified into positive and negative sentiments using the K-Nearest Neighbor (KNN) method and Neighbor-Weighted K-Nearest Neighbor (NWKNN) method. Positive sentiment has a ratio of 81% or 2587 tweets, while the negative sentiment is 19% or 599 tweets. The NWKNN method gives better results than the KNN method when the K value is still small, between  $K = 1$  until  $K = 3$ . After  $K = 4$  the results of the AUC NWKNN method decreased more than the KNN method. The results of the visual wordcloud positive sentiments include the words "iron", "bamboo", and "kokop". The negative sentiments wordcloud has the words "cold", "teeth", and "pain".*

**Keywords:** *Classification, KNN, NWKNN, Plastic straw, Twitter*

*(Halaman ini sengaja dikosongkan)*

## KATA PENGANTAR

Puji syukur penulis panjatkan atas rahmat dan hidayah yang diberikan Allah SWT sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul “Analisis Sentimen Pengguna Twitter Mengenai “Sedotan Plastik” dengan Metode *K-Nearest Neighbor (KNN)* dan *Neighbor-Weighted K-Nearest Neighbor (NWKNN)*” dengan lancar.

Penulis menyadari bahwa Tugas Akhir ini dapat terselesaikan tidak terlepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Kedua orang tua, atas segala do'a, nasehat, kasih sayang, dan dukungan yang diberikan kepada penulis demi kesuksesan dan kebahagiaan penulis.
2. Dr. Suhartono selaku Ketua Departemen Statistika dan Dr. Sutikno, M.Si. selaku Ketua Program Studi Sarjana yang telah memberikan fasilitas, sarana, dan prasarana.
3. MS., Drs. Haryono, MSIE. dan Dr. Drs. Agus Suharsono, MS. selaku dosen-dosen yang menjadi dosen wali selama masa studi yang telah banyak memberikan saran dan arahan dalam proses belajar di Departemen Statistika.
4. Pratnya Paramitha Oktaviana, S.Si., M.Si. selaku dosen pembimbing yang telah meluangkan waktu dan dengan sangat sabar memberikan bimbingan, saran, dukungan serta motivasi selama penyusunan Tugas Akhir.
5. Prof. Nur Iriawan dan Dr. Kartika selaku dosen penguji yang selalu sabar dalam mengomentari serta memberikan masukan dan saran dalam penyelesaian Tugas Akhir.
6. Seluruh dosen Statistika ITS yang telah memberikan ilmu dan pengetahuan yang tak ternilai harganya, serta segenap karyawan Departemen Statistika ITS.
7. Teman-teman Statistika ITS  $\Sigma$ 26 angkatan 2015, yang selalu memberikan dukungan kepada penulis selama ini.
8. Sahabat terdekat Fraschiska, Lianna, Icha, Fonda, Risda, Fiqo yang telah menjadi tempat curhat saya.

9. Teman seperjuangan Excelso Surabaya djohan, yayan, drajad, tama, fredy, zul, april, dan semua cabang Unair juga.
10. Semua teman, relasi dan berbagai pihak yang tidak bisa penulis sebutkan namanya satu persatu yang telah membantu dalam penulisan laporan ini.

Besar harapan penulis untuk mendapatkan kritik dan saran yang membangun sehingga Tugas Akhir ini dapat memberikan manfaat bagi semua pihak yang terkait.

Surabaya, 2019

Penulis

## DAFTAR ISI

	Halaman
<b>LEMBAR PENGESAHAN</b> .....	iii
<b>KATA PENGANTAR</b> .....	iv
<b>DAFTAR ISI</b> .....	xi
<b>DAFTAR GAMBAR</b> .....	xiii
<b>DAFTAR TABEL</b> .....	xv
<b>DAFTAR LAMPIRAN</b> .....	xvii
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	7
2.1 <i>Text Mining</i> .....	7
2.2 Analisis Setimen.....	8
2.3 <i>Preprocessing</i> .....	8
2.4 Pembobotan Kata ( <i>Term Weighting</i> ).....	10
2.5 <i>Crawling</i> .....	13
2.6 Klasifikasi.....	13
2.7 <i>K Nearest Neighbor</i> (KNN).....	14
2.8 Nearest Weighted K Nearest Neighbor (NWKNN).....	16
2.9 <i>Word Cloud</i> .....	18
2.10 <i>k-Fold Cross Validation</i> .....	19
2.11 Pengukuran Kinerja Klasifikasi.....	19
2.12 Twitter.....	22
2.13 Gerakan Anti Sedotan Plastik.....	23
<b>BAB III METODOLOGI PENELITIAN</b> .....	25
3.1 Sumber Data.....	25

3.2	Variabel Penelitian.....	25
3.3	Struktur Data.....	25
3.4	Langkah Analiis.....	26
<b>BAB IV ANALISIS DAN PEMBAHASAN.....</b>		<b>29</b>
4.1	<i>Praproses Data Tweet Masyarakat terhadap Isu “Sedotan Plastik”</i> .....	29
4.2	Karakteristik Data Tweet Masyarakat terhadap Isu “Sedotan Plastik” .....	31
4.3	<i>Term-Weighting</i> .....	36
4.4	Klasifikasi Menggunakan K-Nearest Neighbor.....	37
4.5	Klasifikasi Menggunakan Nearest weighted-KNN .....	38
4.5.1	Penentuan Nilai Eksponen Terbaik.....	39
4.5.2	Klasifikasi Nilai <i>K</i> Terbaik .....	40
4.6	Perbandingan Metode KNN dan NWKNN .....	42
4.7	Visualisasi <i>Word Cloud</i> .....	43
<b>BAB V KESIMPULAN DAN SARAN.....</b>		<b>47</b>
5.1	Kesimpulan.....	47
5.2	Saran .....	47
<b>DAFTAR PUSTAKA .....</b>		<b>49</b>
<b>LAMPIRAN .....</b>		<b>55</b>



## DAFTAR GAMBAR

	Halaman
<b>Gambar 2.1</b> Contoh <i>Word Cloud</i> .....	19
<b>Gambar 2.2</b> Logo Twitter .....	23
<b>Gambar 3.1</b> Diagram Alir .....	28
<b>Gambar 4.1</b> Frekuensi 20 Kata Tertinggi .....	32
<b>Gambar 4.2</b> Presentase Sentimen Positif dan Negatif .....	33
<b>Gambar 4.3</b> AUC Metode KNN .....	37
<b>Gambar 4.4</b> Perbandingan Performa Klasifikasi $K=1$ sampai $K=5$ (a) Akurasi, (b) Presisi, (c) <i>Recall</i> , dan (d) <i>f-</i> <i>measure</i> .....	39
<b>Gambar 4.5</b> AUC Metode NWKNN dengan Eksponen Berbeda .....	40
<b>Gambar 4.6</b> AUC Metode NWKNN $K=1$ sampai $K=10$ .....	41
<b>Gambar 4.7</b> Perbandingan KNN dan NWKNN .....	42
<b>Gambar 4.8</b> <i>Wordcloud</i> Sentimen Positif (a) Sebelum Menghilangkan Kata dengan Frekuensi Besar (b) Setelah Menghilangkan Kata dengan Frekuensi Besar .....	43
<b>Gambar 4.9</b> Frekuensi Kata Sentimen Positif .....	44
<b>Gambar 4.10</b> <i>Wordcloud</i> Sentimen Negatif (a) Sebelum Menghilangkan Kata dengan Frekuensi Besar (b) Setelah Menghilangkan Kata dengan Frekuensi Besar .....	44
<b>Gambar 4.11</b> Frekuensi Kata Sentimen Negatif .....	45

*(Halaman ini sengaja dikosongkan)*

## DAFTAR TABEL

	Halaman
<b>Tabel 2.1</b> Menghitung TF .....	11
<b>Tabel 2.2</b> Menghitung DF.....	12
<b>Tabel 2.3</b> Menghitung IDF .....	12
<b>Tabel 2.4</b> Menghitung TF-IDF .....	13
<b>Tabel 2.5</b> Confusion Matrix.....	17
<b>Tabel 3.1</b> Variabel Penelitian Variabel Keterangan Skala.....	21
<b>Tabel 3.2</b> Struktur Data Penelitian Sebelum Praproses Data.....	22
<b>Tabel 3.3</b> Struktur Data Penelitian Setelah Praproses Data.....	22
<b>Tabel 4.1</b> Praproses Data .....	25
<b>Tabel 4.2</b> Count Vectorizer pada data tweet.....	27
<b>Tabel 4.3</b> Contoh Klasifikasi Tweet Tidak Sesuai .....	29
<b>Tabel 4.4</b> Contoh Tweet Sentimen Positif dan Negatif .....	29
<b>Tabel 4.5</b> Struktur Data IDF .....	32
<b>Tabel 4.6</b> Perfoma Klasifikasi KNN setiap <i>fold</i> $K=1$ .....	33
<b>Tabel 4.7</b> Perfoma Klasifikasi NWKNN setiap <i>fold</i> $K=1$ .....	37

*(Halaman ini sengaja dikosongkan)*

## DAFTAR LAMPIRAN

	Halaman
<b>Lampiran 1.</b> <i>Syntax Crawling</i> .....	49
<b>Lampiran 2.</b> <i>Syntax Preprocessing</i> .....	49
<b>Lampiran 3.</b> <i>Syntax Labelling</i> .....	54
<b>Lampiran 4.</b> <i>Syntax Klasifikasi</i> .....	55
<b>Lampiran 5.</b> <i>Confusion Matrix</i> .....	59
<b>Lampiran 6.</b> <i>Daftar Kata Replace</i> .....	61
<b>Lampiran 7.</b> <i>Daftar Kata Positive Lexicon Based</i> .....	69
<b>Lampiran 8.</b> <i>Daftar Kata Negative Lexicon Based</i> .....	80
<b>Lampiran 9.</b> <i>Surat Keterangan Pengambilan Data</i> .....	97

*(Halaman ini sengaja dikosongkan)*

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Jambeck, dkk (2015) mencatat Indonesia merupakan negara terbesar setelah China yang menghasilkan sampah plastik. Data yang diperoleh dari Asosiasi Industri Plastik Indonesia (INAPLAS) dan Badan Pusat Statistik (BPS) Tahun 2018 menjelaskan bahwa sampah plastik di Indonesia mencapai 64 juta ton/ tahun, dimana sebanyak 3,2 juta tonnya merupakan sampah plastik yang dibuang ke laut. Menteri Kelautan dan Perikanan Susi Pudjiastuti mengatakan bahwa Indonesia merupakan penyumbang sampah plastik terbesar kedua di dunia yang dibuang ke laut. Berita *online* [tribunnews.com](http://tribunnews.com) menyebutkan seekor paus sperma (*Physeter macrocephalus*) tanggal 18 November 2018 terdampar di sekitar Pulau Kapota, Kabupaten Wakatobi, Sulawesi Tenggara. Paus tersebut ditemukan dalam kondisi sudah menjadi bangkai serta perut berisi sampah plastik seberat kurang lebih 6 kilogram. Salah satu produk plastik sekali pakai yang menyumbang polusi laut adalah sedotan plastik. Sedotan plastik memiliki nilai yang rendah dan sulit di daur ulang, maka tidak ada pelaku daur ulang yang bersedia mengambil. Tahun 2017 *Divers Clean Action*, kelompok pemerhati lingkungan khususnya laut menyebutkan bahwa pemakaian sedotan di Indonesia mencapai 93.244.847 batang setiap harinya. Bila 93.244.847 batang sedotan plastik sekali pakai ini direntangkan, maka panjang yang dihasilkan adalah 16.784 km atau sama dengan jarak tempuh Jakarta menuju Meksiko. Bila penggunaan sedotan ini dihitung dalam satu minggu, maka panjang keseluruhan sedotan ini adalah 117.449 km, dan dapat menjadi "sabuk" bagi Bumi. Bahkan sabuk ini dapat tiga kali melilit mengitari Bumi. Jarak satu kali keliling Bumi adalah 40.075 km. Kemudahan dalam mendapatkan sedotan plastik sekali pakai dinilai menjadi penyebab banyaknya sedotan plastik mengotori Bumi. Sedotan plastik mungkin dapat memudahkan kita dalam mengonsumsi minuman, namun dampak dari

ketidakmampuan manusia dalam mengelola sampah membuat sedotan plastik menyumbang polusi di lautan. Walaupun sedotan plastik rata-rata panjangnya hanya sekitar 10 cm, namun perlu 500 tahun lamanya agar sampah plastik tersebut bisa terurai secara alami. Remahan plastik yang terbawa ke laut akan menjadi mikroplastik yang kemudian termakan binatang laut seperti ikan hingga sekecil plankton. Selain merusak ekosistem laut, bila binatang laut tersebut dikonsumsi manusia, efek buruknya pun akan diderita oleh kita sendiri.

Reduksi jumlah sampah plastik oleh beberapa perusahaan dapat mengurangi banyak sampah sedotan plastik sebagai salah satu sampah yang susah terurai. Perubahan untuk beralih ke barang-barang ramah lingkungan dibandingkan plastik sekali pakai dapat menjadi hal yang mungkin untuk dilakukan, dari rencana kecil namun terus termonitor dan evaluasi hingga kedepannya bisa benar-benar berkelanjutan untuk langkah lebih jauh lagi daripada sedotan. Beberapa waktu terakhir ini, ada gerakan yang dibuat untuk menggugah kepedulian masyarakat akan lingkungan, yaitu gerakan untuk tidak menggunakan sedotan plastik. Sejumlah restoran cepat saji hingga *social media influencer* mengatakan bahwa mereka sudah mulai meninggalkan sedotan plastik. Agustin dalam artikel *online* Replubika.co.id menyebutkan bahwa KFC Indonesia mulai berkomitmen tidak lagi menggunakan sedotan plastik sekali pakai sejak tahun 2017. Kampanye dimulai di enam gerai KFC di Jakarta. Berdasarkan data dari KFC, pada akhir 2017 terjadi pengurangan penggunaan sedotan sebanyak 46% di setiap gerai KFC. Kemudian hingga akhir 2018 pengurangan tersebut sudah mencapai 91%. Hal serupa juga dilakukan oleh pesaingnya MCD, McDonald's Indonesia memutuskan menghilangkan dispenser sedotan plastik. Pelanggan tidak akan dapat menggunakan sedotan sekali pakai di 189 gerai yang ada di Indonesia.

Gerakan anti sedotan plastik ini telah direspon oleh masyarakat Indonesia. Mereka dapat mengungkapkan pendapatnya melalui teks pada media sosial. Teks tidak hanya



memuat informasi, teks juga dapat mengekspresikan emosi (Hirat & Mittal, 2015). Salah satu media sosial populer yang digunakan masyarakat Indonesia saat ini adalah Twitter. Berita *online* Kompas Tahun 2019, Pratomo menuliskan bahwa terdapat rata-rata 126 juta pengguna aktif Twitter harian selama kuartal keempat 2018. Tweet memiliki karakteristik khusus, yaitu hanya memiliki panjang maksimal 280 karakter. Hal ini memberikan sebuah tantangan baru pada bidang klasifikasi sentimen dan opini. Tantangan ini berupa bagaimana cara mendapatkan sentimen dan opini dari teks yang pendek dan tidak terstruktur. Kegiatan ekstraksi pengetahuan dan informasi dari pola-pola yang terdapat dalam sekumpulan dokumen teks menggunakan metode analisis tertentu disebut *text mining* (Feldman & Sanger, 2012). *Text mining* dapat diolah untuk berbagai macam keperluan diantaranya adalah untuk *summarization*, pencarian dokumen teks, dan *sentiment analysis*. Sifat data twitter yang berbentuk teks tidak terstruktur dan sulit untuk ditangani sebaiknya diproses terlebih dahulu atau biasa disebut *preprocessing*. Tahap *preprocessing* diantaranya adalah *tokenizing*, *stopwords*, *stemming*, dan normalisasi *text*. Diharapkan melalui *preprocessing* data telah siap untuk diolah lebih lanjut.

Penelitian yang berkaitan dengan sentimen analisis terhadap sampah plastik pernah dilakukan oleh Effendy tahun 2015 dengan judul Analisis Sentimen Berbahasa Indonesia dengan Pendekatan *Lexicon Based*, sedangkan analisis sentimen mengenai sedotan plastik masih sangat minim, sehingga perlu diadakan penelitian lebih lanjut. Penelitian yang menggunakan metode *Neighbor Weighted K-nearest Neighbor* (NWKNN) adalah Indriati dan Achmad Ridok pada tahun 2016 dengan judul *sentiment analysis for review mobile applications*. Berdasarkan hasil pengujian pengaruh data latih tidak seimbang terhadap akurasi metode NWKNN dengan membandingkan metode NWKNN dan KNN, didapatkan bahwa metode NWKNN lebih baik 0.27 dibandingkan metode KNN.

Tugas akhir ini bertujuan untuk melakukan klasifikasi sentimen pengguna media sosial twitter berdasarkan *tweet* yang

berkaitan dengan “sedotan plastik” serta membandingkan hasil klasifikasi antara metode KNN dengan NWKNN. Pemberian label kelas (klasifikasi) merupakan masalah yang cukup sulit pada data tanpa label dengan jumlah yang sangat besar, oleh karena itu digunakan metode pelabelan otomatis yaitu dengan memberikan label secara otomatis berdasarkan penanda tertentu. Beberapa metode klasifikasi teks berbasis statistik adalah *Naïve Bayes Classifier* (NBC), *K-Nearest Neighbour*, dan *Support Vector Machines* (SVM). Metode klasifikasi yang digunakan adalah *K-Nearest Neighbor* (KNN) dan *Neighbor Weighted K-Nearest Neighbor* (NWKNN). NWKNN merupakan pengembangan KNN yang bekerja dengan menerapkan prinsip pembobotan (Tan, 2005). Berdasarkan hasil penelitian yang dilakukan oleh Songbo Tan, NWKNN mampu melakukan klasifikasi dan memiliki kinerja yang baik untuk menyelesaikan masalah data teks yang tidak terdistribusi secara rata (*inbalanced*). Manfaat tugas akhir ini diharapkan dapat membantu pihak-pihak terkait seperti pertimbangan pemerintah untuk menetapkan kebijakan pemakaian sedotan plastik di Indonesia serta masyarakat untuk mengetahui kata kunci yang berkaitan dengan gerakan “sedotan plastik”.

## **1.2 Rumusan Masalah**

Mengetahui respon masyarakat pengguna Twitter terhadap isu “sedotan plastik” merupakan hal yang tidak bisa dikesampingkan. Namun, respon yang diungkapkan oleh masyarakat sangat beragam jenisnya. Mengetahui tanggapan masyarakat membuat proses evaluasi semakin efektif. Diperlukan sebuah penelitian untuk mengetahui opini masyarakat dan mengklasifikasikannya berdasarkan sentimen melalui media sosial. Oleh karena itu dibutuhkan suatu metode klasifikasi, diantaranya *K-Nearest Neighbor* (KNN) dan *Neighbour Weighted K-Nearest Neighbour*. NWKNN merupakan salah satu metode klasifikasi yang digunakan pada data *inbalanced*.

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan yang ingin dicapai dalam penelitian ini adalah

1. Mengetahui hasil klasifikasi sentiment data twitter berbahasa Indonesia yang berkaitan dengan “sedotan plastik” menggunakan metode *K-Nearest Neighbor* dan *Neighbor-Weighted K-Nearest Neighbor*.
2. Mengetahui hasil visualisasi *worldcloud* data twitter berbahasa Indonesia yang berkaitan dengan “sedotan plastik”.

### 1.4 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat bermanfaat untuk beberapa aspek sebagai berikut.

1. Mampu mengetahui proses mengekstrak dan mengolah data teks secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini.
2. Mengetahui metode klasifikasi terbaik.
3. Pertimbangan pemerintah untuk kebijakan pemakaian sedotan plastik di Indonesia.
4. Mengetahui kata kunci yang berkaitan dengan gerakan “sedotan plastik”.

### 1.5 Batasan Masalah

Pada penelitian ini, batasan masalah yang digunakan adalah sebagai berikut.

1. Tweet yang digunakan untuk analisis diambil dengan kata kunci “sedotan plastik”, “sedotan *stainless*”, dan “sedotan”.
2. Pemberian label menggunakan kata kunci yang sudah ditentukan sebelumnya.

*(Halaman ini sengaja dikosongkan)*

## BAB II TINJAUAN PUSTAKA

Bab ini membahas mengenai *text mining*, analisis sentimen, *preprocessing*, pembobotan kata (*term weighting*), *crawling*, klasifikasi, *KNN*, *NWKN*, *k-fold cross validation*, pengukuran kinerja klasifikasi, twitter, dan gerakan anti sedotan plastik.

### 2.1 *Text Mining*

*Text mining* adalah proses penemuan pola yang sebelumnya tidak terlihat pada dokumen atau sumber tertentu menjadi pola yang diinginkan untuk tujuan tertentu (Mustafa, Akbar, & Sultan, 2009). *Text mining* sering digunakan untuk analisis informasi, pengambilan keputusan, dan tugas-tugas manajemen informasi lainnya yang terkait dengan data teks dalam jumlah besar. *Text mining* adalah bagian dari data mining, namun proses *text mining* memerlukan lebih banyak tahapan dibanding data mining karena data teks memiliki karakteristik yang lebih kompleks daripada data biasa bahkan data yang sudah terstruktur. Berdasarkan ketidakteraturan struktur data teks, maka proses *text mining* memerlukan beberapa tahap awal yang pada intinya adalah mempersiapkan agar teks dapat diubah menjadi lebih terstruktur. Data tersebut diolah untuk memenuhi kebutuhan informasi menggunakan berbagai metode klasifikasi, *clustering*, analisis sentimen.

*Text mining* juga dapat digunakan untuk proses penemuan *rule* baru dengan algoritma pengelompokan, asosiasi, dan *ranking*. Beberapa fungsi tersebut, yang paling banyak dilakukan adalah proses pengelompokan. Terdapat dua jenis metode pengelompokan teks, yaitu *text clustering* dan *text classification*. Menurut Darujati dan Gumelar di tahun 2012, *text clustering* berhubungan dengan proses menemukan sebuah struktur kelompok yang belum terlihat (tak terpandu atau *unsupervised*) dari sekumpulan dokumen. Sedangkan *text classification* dapat dianggap proses untuk membentuk golongan (kelas-kelas) dari dokumen berdasarkan

pada kelas kelompok yang sudah diketahui sebelumnya (terpandu atau *supervised*).

## 2.2 Analisis Sentimen

Analisis sentimen atau *sentiment analysis* adalah sebuah proses atau metode memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terdapat pada sebuah kalimat opini (Rozi, Pramono, & Dahlan, 2012). Analisis sentimen memiliki banyak kegunaan, salah satunya untuk mengidentifikasi kecenderungan opini publik terhadap sebuah objek, produk ataupun layanan. Bahkan di Amerika terdapat sekitar 20 – 30 perusahaan memfokuskan pada layanan analisis sentiment (Liu, 2010). Melakukan analisis sentimen dapat menggunakan berbagai macam algoritma klasifikasi seperti Naïve Bayes (NB), *Support Vector Machine* (SVM), dan *Artificial Neural Network* (ANN) (Chandani, Wahono, & Purwanto, 2015). Dalam melakukan analisis sentimen setiap data atau *tweets* akan memiliki label masing masing. *Labelling* data ulasan dalam penelitian ini dilakukan dengan menerapkan kamus *lexicon*, yaitu pemberian label berupa sentimen positif dan negatif ke dalam suatu ulasan berdasarkan frasa yang menyusunnya. Untuk mendapatkan label sentimen, *lexicon* bekerja berdasarkan *rules* pada persamaan (2.1) (Mohammad, Kiritchenko, & Zhu, 2013).

$$\text{skor} = \text{jumlah kata positif} - \text{jumlah kata negatif.} \quad (2.1)$$

Skor merupakan indikator yang digunakan untuk *labelling* sentimen dari suatu ulasan. Pada penelitian ini, daftar kata yang menunjukkan sentimen positif dan negatif diperoleh dari data kamus yang disusun oleh Hu & Liu (2004).

## 2.3 Preprocessing

*Preprocessing* data merupakan langkah penting dalam proses penemuan pengetahuan, karena keputusan-keputusan yang berkualitas harus didasarkan pada data yang berkualitas (Kumar & Cadha, 2012). *Preprocessing* data sering kali digunakan untuk mengurangi kesalahan data. Tujuan *preprocessing* adalah untuk meningkatkan akurasi data (Rifqi, Maharani, & Shaufiah, 2011).

*Preprocessing* dalam *text mining* cukup kompleks bila dibandingkan dengan *preprocessing* untuk data angka atau kategori. Hal ini dikarenakan struktur kalimat atau teks sangatlah rumit. Secara umum hal yang termasuk dalam *preprocessing* adalah *cleaning*, normalisasi, transformasi, pemilihan fitur dan ekstraksi fitur (Kostsiantis, Kanellopoulos, & Pintelas, 2006). Berikut penjelasan masing-masing tahap *preprocessing* yang digunakan.

- a. *Case Folding*, merupakan proses untuk mengubah semua karakter pada teks menjadi huruf kecil. Karakter yang diproses hanya huruf ‘a’ hingga ‘z’ dan selain karakter tersebut akan dihilangkan seperti tanda baca titik (.), koma (,), dan angka (Weiss, 2010).
- b. *Cleansing*, tahap *cleansing* merupakan tahapan untuk menghilangkan kata yang tidak diperlukan misalnya karakter HTML, link URL, username (@username), dan hastag (#). Tahap *cleansing* ini diperlukan karena kata-kata tersebut dianggap *noise* yang tidak diperlukan pada proses data (Buntoro, Adji, & Purnamasari, 2014).
- c. *Punctuation removal*, merupakan proses menghilangkan simbol atau tanda baca di dalam dokumen seperti ! @ # \$ % ^ & \* ( ) { } [ ] \ | : ” < > ? ; ’ , . / ` ~ (Jannah, Fithriasari, & Usugawa, 2018).
- d. *Numbers removal*, merupakan proses menghapuskan angka yang terdapat dalam dokumen seperti 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 (Jannah, Fithriasari, & Usugawa, 2018).
- e. *Stopwords*, merupakan kosakata yang bukan termasuk kata unik atau ciri pada suatu dokumen atau tidak menyampaikan pesan apapun secara signifikan pada teks atau kalimat (Dragut, 2009). Kosakata yang dimaksud yaitu seperti kata penghubung dan kata keterangan yang bukan merupakan kata unik, misalnya “dari”, “akan”, “seorang”, dan sebagainya.
- f. *Whitespace removal*, merupakan suatu proses untuk menghapuskan spasi ganda (Pour, 2008).

- g. *Stemming*, tahap *stemming* ini merupakan tahap mendapatkan kata dasar. Sistem kerja tahap *stemming* ini adalah menghilangkan awalan, akhiran, sisipan, dan *confixes* (kombinasi dari awalan dan akhiran) (Bing, 2010).
- h. *Tokenizing*, tahap *tokenizing* merupakan tahapan memutuskan kata per kata pada kalimat. Tahapan ini bertujuan untuk memecah yang semula berupa kalimat menjadi potongan-potongan kata, sehingga urutan *string* akan terputus menjadi potongan-potongan kata penyusunnya (Ariadi & Fithriasari, 2015).

#### 2.4 Pembobotan Kata (*Term Weighting*)

*Term* merupakan kata atau frase dalam sebuah dokumen yang dapat digunakan untuk mengidentifikasi konteks yang ada pada dokumen itu sendiri. Sering munculnya term dalam dokumen dapat digunakan sebagai proses untuk melakukan perhitungan sehingga diketahui penting tidaknya suatu kalimat. Proses perhitungan term bisa dilakukan setelah dokumen direpresentasikan kedalam bentuk nilai numerik dokumen, salah satu cara merepresentasikan dokumen kedalam bentuk numerik adalah menggunakan *TF-IDF* (*Term Frequency-Inversed Document Frequency*).

*Term frequency* atau biasa disebut *TF* adalah jumlah kemunculan kata pada suatu dokumen. Semakin banyak suatu kata muncul pada dokumen, maka semakin besar kata tersebut berpengaruh pada dokumen tersebut. Sebaliknya, semakin sedikit suatu kata muncul pada dokumen, maka semakin kecil kata tersebut berpengaruh pada dokumen tersebut. Pembobotan menggunakan *TF* dijelaskan pada persamaan (2.2) (Manning, Raghavan, & Schütze, 2009).

$$TF(t, d) = f(t, d), \quad (2.2)$$

keterangan:

$f(t, d)$  = jumlah kemunculan kata  $t$  pada dokumen  $d$ .



*Inverse document frequency (IDF)* adalah pembobotan yang menghitung frekuensi kemunculan sebuah kata pada kumpulan dokumen. Pembobotan menggunakan *IDF* dijelaskan pada persamaan (2.3) (Feldman & Sanger, 2007).

$$IDF(t) = 1 + \log \frac{D}{Dt}, \quad (2.3)$$

keterangan:

$D$  = jumlah seluruh dokumen,

$Dt$  = Jumlah dokumen yang mengandung kata  $t$ .

Pembobotan *TF-IDF* adalah suatu pengukuran statistik untuk mengukur seberapa penting sebuah kata dalam kumpulan dokumen persamaan (2.4).

$$TFIDF(t, d) = TF(t, d_i) \times IDF(t). \quad (2.4)$$

Normalisasi merupakan proses penyetaraan jumlah kata yang berbeda-beda pada setiap dokumen. Normalisasi menggunakan *Euclidean normalized* dapat dilihat pada persamaan (2.5) (Manning, Raghavan, & Schütze, 2009).

$$\vec{v} = \frac{\vec{V}}{\left\| \vec{V} \right\|_2}, \quad (2.5)$$

keterangan:

$\vec{v}$  = vektor *TFIDF* setelah dinormalisasi.

$\vec{V}$  = vektor *TFIDF* sebelum dinormalisasi.

$$\left\| \vec{V} \right\|_2 = \left( \sum_{i=1}^n V_i^2 \right)^{\frac{1}{2}}$$

Ilustrasi sederhana dari *TF-IDF* adalah sebagai berikut:

Dokumen 1 ( $d_1$ ) = kampanye mengurangi sedotan plastik

Dokumen 2 ( $d_2$ ) = butuh sedotan stainless steel

Dokumen 3 ( $d_3$ ) = mahal sedotan stainless

Jadi jumlah dokumen ( $D$ ) = 3 telah dilakukan tahap tokenizing dan proses filtering. Berikut ini adalah tabel perhitungan *TF-IDF*.

- Menghitung *Term Frequency (TF)*, yakni frekuensi kemunculan term ( $t$ ) pada dokumen ( $d$ ). Perhitungan *TF* terdapat pada Tabel 2.1.

**Tabel 2.1** Menghitung *TF*

Term ( $t$ )	D1	D2	D3
kampanye	1	0	0
mengurangi	1	0	0
sedotan	1	1	1
plastik	1	0	0
butuh	0	1	0
stainless	0	1	1
steel	0	1	0
mahal	0	0	1

- Menghitung *Document Frequency (DF)*, yaitu banyaknya dokumen dimana suatu term ( $t$ ) muncul. Contoh berdasarkan soal yang sama pada poin pertama. Perhitungan *DF* terdapat pada Tabel 2.2.

**Tabel 2.2** Menghitung *DF*

Term ( $t$ )	DF
Kampanye	1
mengurangi	1
Sedotan	3
Plastik	1
Butuh	1
Stainless	1
Steel	1
Mahal	1

- Menghitung *Invers Document Frequency (IDF)*. Perhitungan *DF* terdapat pada Tabel 2.3.

**Tabel 2.3** Menghitung *IDF*

Term (t)	DF	IDF
kampanye	1	2,1
mengurangi	1	2,1
sedotan	3	1
plastik	1	2,1
butuh	1	2,1
stainless	1	1,41
steel	1	2,1
mahal	1	2,1

- Hasil perhitungan *TF-IDF* setelah dinormalisasi terdapat pada Tabel 2.4.

**Tabel 2.4** Menghitung *TF-IDF*

Term (t)	TF-IDF		
	D1	D2	D3
kampanye	0,546454	0	0
mengurangi	0,546454	0	0
sedotan	0,322745	0,345205	0,425441
plastik	0,546454	0	0
butuh	0	0,584483	0
stainless	0	0,444514	0,547832
steel	0	0,584483	0
mahal	0	0	0,720333

## 2.5 *Crawling*

*Crawling* merupakan sebuah proses untuk mendapatkan informasi konten atau keseluruhan isi halaman yang terdapat pada suatu halaman *website* dan menyimpannya secara *offline* (Pant, Srinivasan, & Menczer, 2004). Dalam penelitian ini *crawling* dilakukan pada laman media sosial Twitter. *Crawling* bertujuan untuk mendapatkan data berupa *tweet* yang bersumber dari media

sosial Twitter. Data yang diambil adalah data publik berupa *tweet*, waktu, pengguna, dan keterangan lain yang dibutuhkan. Semua informasi ini didapatkan melalui bantuan layanan dari Twitter API.

## 2.6 Klasifikasi

Klasifikasi adalah salah satu metode yang ada di dalam data *mining*. Di dalam klasifikasi, label dari setiap kelas sudah ditentukan terlebih dahulu. Proses klasifikasi sendiri merupakan proses untuk menemukan model atau membedakan kelas atau data yang bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Klasifikasi merupakan bentuk analisis data yang dapat menggambarkan ekstrak model dari suatu data yang penting (Han, Kamber, & Pei, 2012). Proses dalam klasifikasi terdiri atas:

- *Learning* model: di dalam fase ini data *training* akan dianalisa dengan algoritma klasifikasi.
- Klasifikasi: pada fase ini data *testing* digunakan untuk memperkirakan hasil akurasi berdasarkan *rule* klasifikasi.

## 2.7 *K Nearest Neighbor* (KNN)

K-Nearest Neighbor (KNN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari *query distance* yang baru diklasifikan berdasarkan mayoritas dari kategori pada KNN. Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan *training sample*. *Classifier* tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik *query* akan ditemukan sejumlah *K* obyek atau (titik *training*) yang paling dekat dengan titik *query*. Klasifikasi menggunakan *voting* terbanyak diantara klasifikasi dari *K* obyek. Algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru.

Algoritma metode KNN sangatlah sederhana, bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk menentukan KNN-nya. *Training sample* diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-

bagian berdasarkan klasifikasi *training sample*. Sebuah titik pada ruang ini ditandai kelas  $c$  jika kelas  $c$  merupakan klasifikasi yang paling banyak ditemui pada  $K$  buah tetangga terdekat dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan *Euclidean Distance* yang direpresentasikan pada persamaan (2.6) sebagai berikut:

$$d_{ij} = \sqrt{\sum_{i=1}^n (x_{ik} - x_{jk})^2}, \quad (2.6)$$

keterangan :

$d_{ij}$  = jarak antara objek ke- $i$  dan ke- $j$

$x_{ik}$  = nilai pengamatan objek ke- $i$  variabel ke- $k$

$x_{jk}$  = nilai pengamatan objek ke- $j$  variabel ke- $k$

Diberikan ilustrasi perhitungan metode KNN agar lebih mudah dalam memahami cara kerja dari metode ini. Terdapat data *training* sebanyak 5 data dengan dua label *Bad* dan *Good*, serta data baru (*testing*) untuk diklasifikasikan apakah tergolong label *Bad* atau *Good* pada Tabel 2.5.

**Tabel 2.5** Data *Training* dan *Testing*

No	$x_1$	$x_2$	$y$
1	7	8	<i>Good</i>
2	6	6	<i>Bad</i>
3	7	7	<i>Good</i>
4	4	6	<i>Good</i>
5	4	3	<i>Bad</i>
6	4	7	?

Langkah pertama yang dilakukan adalah mencari jarak terdekat melalui *euclidiencie distance*. Hasil perhitungan *euclidiencie distance* terdapat pada Tabel 2.6.

**Tabel 2.6** *Euclidence Distance*

No	<i>Euclidence distance</i>
1	3,16
2	2,24
3	3
4	1
5	4

Mengurutkan *euclidence distance* dari data *testing* dengan data *training* dan menentukan  $K$  tetangga terdekat. Misalkan ditentukan  $K=3$ , maka yang digunakan untuk memprediksi label dari data *testing* adalah 3 tetangga terdekat. Hasil 3 tetangga terdekat dapat dilihat pada Tabel 2.7.

**Tabel 2.7** Pemilihan tetangga terdekat

No	<i>Euclidence distance</i>	Label
1	3,16	<i>Good</i>
2	2,24	<i>Bad</i>
3	3	<i>Good</i>
4	1	<i>Good</i>
5	4	<i>Bad</i>

Didapatkan 3 tetangga terdekat dari data *testing* dengan 2 label *Good* dan 1 label *Bad*. Dari jumlah mayoritas ( $\text{Good} > \text{Bad}$ ) disimpulkan bahwa data *testing* termasuk dalam kategori *Good*.

## 2.8 *Nearest Weighted K Nearest Neighbor (NWKNN)*

Metode NWKNN merupakan pengembangan dari metode KNN. Metode ini menggunakan prinsip pembobotan. Bobot akan diberi lebih sedikit ke jumlah  $K$  tetangga yang berasal dari kelas mayoritas, dan sebaliknya untuk kelas minoritas. Metode NWKNN mampu melakukan klasifikasi dengan baik, karena metode ini cocok untuk diimplementasikan ke data yang tidak terdistribusi secara rata (Indriati & Ridok, 2016). Langkah algoritma pada metode NWKNN tidak jauh berbeda dengan langkah algoritma

KNN, yang membedakan adalah adanya pembobotan untuk setiap jenis/kelas dan proses perhitungan skor untuk menentukan klasifikasi terhadap data *testing* (Faldy, Indrianti, & Novanto, 2015). Langkah-langkah dalam algoritma NWKNN menurut Adeniyi, Wei, & Yongquan tahun 2014 adalah sebagai berikut :

1. Menentukan nilai  $K$ .
2. Menghitung nilai kedekatan ketetanggaan antara data *testing* terhadap data *training* menggunakan Persamaan *Euclidean Distance*. Menghitung ketetanggaan terdekat yaitu dengan menghitung jarak antara data *training* terhadap data *testing* menggunakan rumus *Euclidean Distance* dapat digunakan persamaan (2.6).
3. Mengurutkan hasil perhitungan jarak atau kedekatan kedalam kelompok yang mempunyai kedekatan jarak.
4. Mengumpulkan kategori klasifikasi *nearest neighbor*.
5. Perhitungan bobot digunakan persamaan (2.7).

$$Weight_i = \frac{1}{\left( \frac{Num(C_i^d)}{\text{Min}\{Num(C_j^d) | j = 1, \dots, k\}} \right)^{1/exp}}, \quad (2.7)$$

keterangan :

$Num(C_i^d)$  = banyak data *training*  $d$  pada kelas  $i$

$Num(C_j^d)$  = banyak data *training*  $d$  pada kelas  $j$ , dimana  $j$  terdapat dalam himpunan  $k$  tetangga terdekat

Exp = eksponen (nilai exp lebih dari 1)

Dalam penelitian Achmad Ridok & Retnani Latifa tahun 2015 hasil uji coba klasifikasi dokumen bahasa Indonesia dengan NWKNN memberikan kinerja terbaiknya pada ekponen =2. Setiap data yang telah dihitung nilai bobotnya akan digunakan untuk menghitung nilai skor. Dimana hasil nilai bobot akan dikalikan dengan Persamaan hasil skor. Rumus hasil skor dihitung dengan persamaan (2.8).

$$Skor(X, C_i) = Weight_i * \left( \sum_{d|KNN(x)} \left( \left( \sqrt{\sum_{i=1}^n (x_{ik} - x_{jk})^2} \right) \delta(d_j, C_i) \right) \right), \quad (2.8)$$

keterangan :

$Weight_i$  = bobot jenis/kelas  $i$

$djKNN(x)$  = data *training*  $d_j$  pada kumpulan tetangga terdekat dari data *testing*  $x$

$(d_j, C_i)$  = akan bernilai 1 jika nilai jarak  $\in C_i$  dan bernilai 0 jika nilai jarak  $\notin C_i$ .

$\sqrt{\sum_{i=1}^n (x_{ik} - x_{jk})^2}$  = *Euclidean Distance*

$C_i$  = jenis atau kelas

Dalam mengklasifikasikan data menggunakan metode NWKNN hampir sama dengan metode KNN dimana saat mencari jarak sama menggunakan *euclidienece distance*, hanya saja pada metode NWKNN diberikan bobot. Berikut perhitungan bobot masing-masing label/kelas dengan data yang sama pada Tabel 2.5.

$$Weight_{Good} = \frac{1}{\left(\frac{2}{1}\right)^{\frac{1}{2}}} = 0,707$$

$$Weight_{Bad} = \frac{1}{\left(\frac{1}{2}\right)^{\frac{1}{2}}} = 1,414$$

Setelah didapatkan bobot masing-masing label yang dilakukan adalah menghitung *score*.

$$Score_{Good} = 0,707 \times (1 + 3) = 2,828$$

$$Score_{Bad} = 1,414 \times (2, 24) = 3,162$$

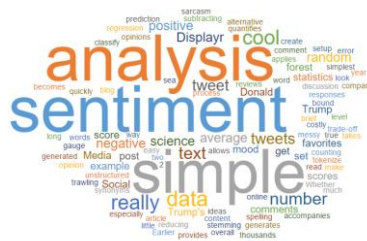
Hasil *score* masing-masing label adalah 2,828 untuk label *Good* dan 3,162 untuk label *Bad*. Dalam menentukan label untuk data *testing* dipilih *score* terkecil ( $Score_{Good} < Score_{Bad}$ ), sehingga data *testing* pada data Tabel 2.5 masuk dalam label *Good*.

## 2.9 Word Cloud

*Word Cloud* adalah salah satu hasil dari metode *text mining*, yang menampilkan kata-kata populer terkait dengan kata kunci internet dan data teks. Schroeder tahun 2002 berargumen bahwa citraan visual memberikan kemudahan dalam melakukan interpretasi dan akses instan terhadap ide yang penting. Terkait hal tersebut, bahasa sebagai citra visual dapat memberikan akses instan



terhadap distribusi ide dalam analisis wacana dibandingkan bahasa dalam bentuk wacana tertulis. *Word Cloud* merupakan sebuah sistem yang memunculkan susunan kata sebagai citra visual terkait frekuensi kemunculan kata dalam suatu teks verbal. Visualisasi *Word Cloud* dari teks akan memudahkan pengamat dalam melihat gagasan dan pendirian sang penulis teks sehingga dapat menjadi alat bantu dalam melakukan analisis terhadap sebuah wacana tertulis. Contoh visualisasi *wordcloud* dapat dilihat pada Gambar 2.1.



Gambar 2.1 Contoh *Word Cloud*

## 2.10 *k-Fold Cross Validation*

Pada *k-fold Cross Validation*, data awal secara acak dibagi menjadi  $k$  himpunan bagian “*folds*” yaitu  $d_1, d_2, \dots, d_k$ , masing-masing berukuran sama dengan pembagian data *training* dan data *testing* dilakukan sebanyak  $k$ . Pada iterasi yang pertama, partisi pada  $d_i$  digunakan sebagai data *testing*, dan partisi yang tersisa digunakan sebagai data *training*. Sehingga, pada iterasi pertama, himpunan bagian  $d_2, \dots, d_k$  secara kolektif digunakan sebagai data *training* untuk mencari model yang pertama dan diujikan pada data *testing* ( $d_i$ ) dan seterusnya. Untuk klasifikasi, taksiran nilai akurasi adalah jumlah keseluruhan klasifikasi yang benar dari iterasi  $k$ , dibagi dengan banyaknya *tuples* pada data awal (Han & Kamber, 2006).

## 2.11 Pengukuran Kinerja Klasifikasi

Pengukuran kinerja didapatkan melalui perbandingan pada *Confusion matrix*, akurasi, presisi, *recall*, dan *F-Measure*. *Confusion matrix* merupakan suatu metode yang biasanya

digunakan untuk melakukan perhitungan akurasi pada konsep penggalian data. *Confusion matrix* berguna untuk menganalisis seberapa baik *classifier* anda dapat mengenali tupel dari kelas yang berbeda. *Confusion matrix* melakukan perhitungan dengan 3 keluaran, yaitu *recall*, presisi, dan akurasi seperti pada Tabel 2.8 (Han & Kamber, 2006).

**Tabel 2.8** *Confusion Matrix*

		Prediksi Kelas	
		Benar	Salah
Kelas Sebenarnya	Benar	TP ( <i>True Positive</i> )	FN ( <i>False Negative</i> )
	Salah	FP ( <i>False Positive</i> )	TN ( <i>True Negative</i> )

Presisi adalah presentase dokumen sebenarnya yang diambil relevan dengan query. Secara formal dapat didefinisikan seperti pada persamaan (2.9). *Recall* adalah presentase dokumen yang relevan dengan query dan pada kenyataannya di ambil. Secara formal dapat dedefinisikan seperti pada persamaan (2.10). Akurasi adalah nilai perbandingan antara nilai data yang diklasifikasikan secara benar dengan seluruh data seperti pada persamaan (2.11). *F-Measure* adalah nilai harmonik atau nilai rata-rata (*mean*) dari nilai presisi dan recal seperti pada persamaan (2.12) (Han & Kamber, 2006). Dalam persama ini, parameter TP, FP, FN, dan TN berturuturut menyatakan jumlah klasifikasi yang benar dari data positif, jumlah klasifikasi yang salah dari data negatif, jumlah klasifikasi yang salah dari data positif, dan jumlah klasifikasi yang benar dari data negatif.

$$precision = \frac{TP}{TP+FP}, \quad (2.9)$$

$$recall = \frac{TP}{TP+FN}, \quad (2.10)$$

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (2.11)$$

$$f - measure = \frac{2 \times precision \times recall}{precision + recall}. \quad (2.12)$$

Akurasi merupakan nilai yang mengukur efektivitas secara keseluruhan dari suatu *classifier*. Sensitivitas (*recall*) menilai kemampuan suatu *classifier* untuk mengklasifikasikan dokumen ke kelas positif dengan tepat. Adapun spesifisitas merupakan ukuran untuk menilai efektivitas suatu *classifier* dalam hal mengidentifikasi label negatif (Sokolova & Lapalme, 2009). Namun akurasi tidak lagi dapat dijadikan sebagai tolok ukur evaluasi ketika ditemui kasus data *imbalance*, karena kelas minoritas tidak memiliki dampak yang berarti pada akurasi (Sun, Wong, & Kamel, 2009). Oleh sebab itu digunakan formula lain yang dapat diterapkan pada data *imbalance*, yakni dengan menggabungkan unsur sensitivitas dan spesifisitas. Salah satu ukuran kinerja klasifikasi yang menggabungkan dua unsur tersebut adalah *Area Under the Curve* (AUC). Nilai AUC dapat diperoleh melalui persamaan (2.13).

$$AUC = \frac{1}{2} \left( \frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right). \quad (2.13)$$

AUC merupakan indikator performansi kurva *Receiver Operating Characteristic* (ROC) yang meringkas kinerja suatu *classifier* menjadi satu nilai. AUC lebih dipertimbangkan dalam penilaian model karena kinerja suatu *classifier* diurutkan berdasarkan perolehan nilai AUC. Umumnya, nilai AUC berkisar antara 0,5 hingga 1,0 dengan interpretasi setiap rentang nilai AUC ditampilkan pada Tabel 2.9 (Bekkar, Djemaa, & Alitouche, 2013).

**Tabel 2.9** Kriteria Nilai AUC

Nilai AUC	Keterangan
0,5 – 0,6	<i>Poor</i>
0,6 – 0,7	<i>Fair</i>
0,7 – 0,8	<i>Good</i>
0,8 – 0,9	<i>Very Good</i>
0,9 – 1,0	<i>Excellent</i>

Keakuratan hasil klasifikasi juga dapat diketahui dengan menghitung Statistik uji Press's  $Q$  (Montgomery & Peck, 1992). Menurut Hair, dkk (2006) Statistik uji *Press's Q* membandingkan antara jumlah ketepatan klasifikasi dengan total pengamatan dan banyaknya kelompok. Jika nilai *Press's Q* lebih besar dari nilai *chi-square* berderajat bebas 1, maka prediksi dalam klasifikasi dianggap signifikan secara statistik.

Hipotesis:

$H_0$  : Pengklasifikasian tidak akurat

$H_1$  : Pengklasifikasian akurat

Taraf signifikansi =  $\alpha$

Uji statistik *Press's Q* didefinisikan pada persamaan (2.14)

$$\text{Press's } Q = \frac{[N - (qC)]^2}{N(C - 1)}, \quad (2.14)$$

dengan:

$N = TP + FP + FN + TN$

$q = TP + TN$

$C =$  banyaknya kelas

Kriteria Uji: tolak  $H_0$  jika  $\text{Press's } Q > \chi_{\alpha,1}^2$ .

## 2.12 Twitter

Twitter adalah jejaring sosial yang memungkinkan penggunanya untuk mengirim pesan teks dengan maksimal 140 karakter, yang dikenal sebagai tweets. Twitter terstruktur dengan follow dan followers, dimana setiap pengguna diberikan kebebasan untuk melakukan *follow* atau *followers* pada akun lainnya. Ada juga kemungkinan untuk mengirim pesan pribadi ke profil lain. Hal ini juga memungkinkan mengirim video, foto, dan mengarahkan pembaca ke halaman web lain melalui *link*. Twitter adalah jejaring sosial yang dinamis yang memungkinkan setiap pengguna untuk memiliki akses ke informasi yang terus diterbitkan (Daniel, Neves, & Horta, 2017).

Twitter menyediakan REST API untuk pengembang yang memungkinkan untuk mengakses data status dan profil pengguna. Twitter juga menyediakan akses pengembang ke sejumlah

informasi secara *real time* melalui *Streaming API*. Tweets dapat dikelompokkan dengan kata-kata *hashtag* didahului oleh karakter “#”, digunakan untuk menandai kata kunci atau topik tweet. Selain itu, pengguna dapat membuat tweet yang diposting oleh pengguna lain. Twitter memiliki logo seperti pada Gambar 2.2.



Gambar 2.2 Logo Twitter

### **2.13 Gerakan Anti Sedotan Plastik**

Permasalahan mengenai sedotan plastik menjadi perhatian khusus masyarakat. Rumah makan sebagai salah satu penyedia sedotan plastik terbesar ikut andil dalam permasalahan ini. Mengetahui akibat yang ditimbulkan dari sedotan plastik yang tidak didaur ulang dan terbuang kelaut, beberapa rumah makan menerapkan kebijakan tidak menyediakan sedotan plastik. Gerakan anti sedotan plastik ramai dibicarakan di media sosial, salah satunya adalah twitter. Muncul beberapa *hashtag* untuk mendukung gerakan anti sedotan plastik, diantaranya *#nostrawmovement* dan *#MulaiTanpaSedotan*.

*(Halaman ini sengaja dikosongkan)*

## BAB III METODOLOGI PENELITIAN

### 3.1 Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah tweet dari pengguna Twitter di Indonesia dengan *keyword* “sedotan plastik”, “sedotan *stainless*”, “sedotan”. Data didapat dari Twitter API (*Application Programming Interface*). Data dari Twitter diambil dengan mengaktifkan/*install* terlebih dahulu paket *twitteR* pada RStudio. Sebelumnya harus dilakukan pendaftaran terlebih dahulu pada Twitter *apps* menggunakan akun pribadi. Setelah terdaftar ditampilkan API *key*, API *secret*, *access tokens*, dan *access tokens secret* yang digunakan untuk menghubungkannya pada pemrograman RStudio. Waktu pengambilan data beberapa hari pada bulan Februari, Maret, April, dan Mei.

### 3.2 Variabel Penelitian

Variabel penelitian pada penelitian ini dapat dilihat pada Tabel 3.1. Variabel penelitian yang digunakan dalam penelitian ini setelah dilakukan praproses pada data teks *tweet* terdiri dari variabel prediktor ( $x$ ) yaitu frekuensi kata setiap tweet dan variabel respon ( $y$ ) yaitu klasifikasi sentimen *tweet*.

**Tabel 3.1** Variabel Penelitian Variabel Keterangan Skala

Variabel	Keterangan	Skala
	Kelas	
$y$	0. Positif 1. Negatif	Nominal
$x$	Frekuensi kata ke- $i$ yang muncul pada objek (twitter)	Rasio

### 3.3 Struktur Data

Struktur data yang diambil dalam penelitian sebelum dan setelah *preprocessing* dapat dilihat pada Tabel 3.2 dan Tabel 3.3. Notasi  $n$  menunjukkan jumlah data yang digunakan dalam penelitian, kolom SN (*Screen Name*) adalah akun yang membuat tweet, serta  $x$  merupakan frekuensi kata ke- $i$ . Berikut merupakan

struktur data penelitian sebelum dan setelah dilakukan praproses data tweet.

**Tabel 3.2** Struktur Data Penelitian Sebelum Praproses Data

No.	Screen Name	Tweet ( $t$ )
1	$account_1$	$t_1$
2	$account_2$	$t_2$
⋮	⋮	⋮
⋮	⋮	⋮
$n$	$account_n$	$t_n$

**Tabel 3.3** Struktur Data Penelitian Setelah Praproses Data

No.	Screen Name	Tweet ( $t$ )	Klasifikasi Sentimen ( $y$ )	Kata Kunci ( $x_1$ )	Kata Kunci ( $x_2$ )	...	Kata Kunci ( $x_k$ )
1	$account_1$	$t_1$	$y_1$	$x_{1,1}$	$x_{1,2}$	...	$x_{1,k}$
2	$account_2$	$t_2$	$y_2$	$x_{2,1}$	$x_{2,2}$	...	$x_{2,k}$
⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
$n$	$account_n$	$t_n$	$y_n$	$x_{n,1}$	$x_{n,2}$	...	$x_{n,k}$

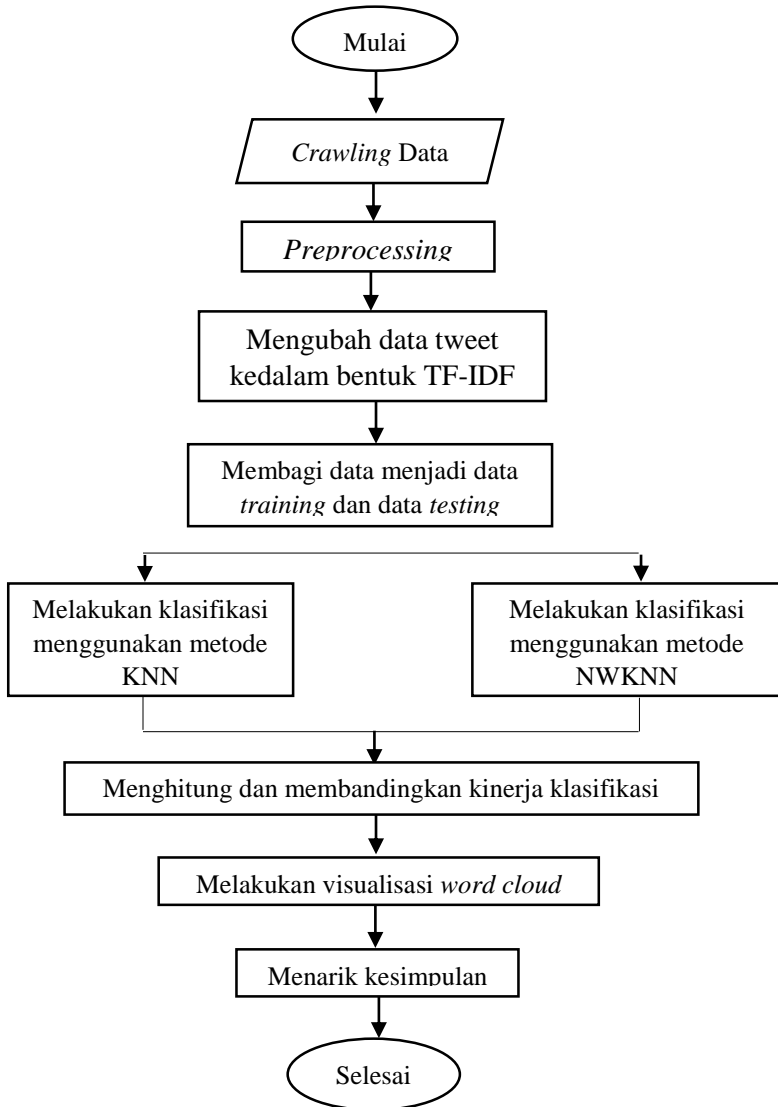
### 3.4 Langkah Analisis

Langkah analisis digunakan untuk menggambarkan langkah-langkah penelitian yang akan dilakukan secara urut. Langkah analisis yang digunakan adalah sebagai berikut.

1. Mengambil data tweet dengan menggunakan Twitter API.
  - a. Memasukkan *keyword* yang berhubungan dengan “sedotan plastik”, “sedotan *stainless*”, dan “sedotan”.
  - b. Menyimpan hasil *crawling* data.
2. Melakukan praproses data.
  - a. Melakukan *case folding*.
  - b. Melakukan *replace* pada kata yang dianggap tidak sesuai EYD.



- c. Melakukan *cleansing*, yaitu proses menghilangkan dalam twitter adalah karakter HTML, hastag (#), username (@username), simbol retweet (response tweet) “RT”, dan link URL, angka, dan spasi yang berlebih.
  - d. Melakukan proses *stopping* berdasarkan *stoplist* yang berisi *stopwords* yang telah ditentukan sebelumnya. Kata-kata yang terdapat pada tweet akan dibandingkan dengan daftar *stopwords*, jika terdapat kata-kata yang terdapat pada *stopwords* maka kata tersebut akan dihapus dari tweet.
  - e. Melakukan *stemming* untuk mendapatkan kata dasar. Penentuan kata dasarnya berdasarkan daftar yang telah disiapkan.
  - f. Melakukan *tokenizing* untuk memecah *tweet* menjadi kata per kata.
  - g. Mengubah data tweet ke dalam bentuk frekuensi kemunculan kata menggunakan *TF-IDF*.
  - h. Membagi data *training* dan data *testing*.
  - i. Menentukan  $k$  (jumlah tetangga paling dekat).
  - j. Menghitung nilai tetangga terdekat dengan *Euclidean distance*.
3. Klasifikasi data menggunakan *K Nearest Neighbor* dan *Neighbor Weighted K Nearest Neighbor*.
  4. Perhitungan nilai skor.
  5. Membandingkan kinerja klasifikasi dari hasil performa klasifikasi yang terbentuk.
  6. Melakukan visualisasi tweet dengan *Word Cloud*.
  7. Membuat kesimpulan dan saran.
- Langkah-langkah analisis secara umum digambarkan pada diagram alir pada Gambar 3.1.



Gambar 3.1 Diagram Alir

## BAB IV ANALISIS DAN PEMBAHASAN

Pada bab ini akan dibahas mengenai hasil analisis data tweet masyarakat terhadap isu “sedotan plastik”. Data yang digunakan merupakan data gabungan dari *keyword* “sedotan”, “sedotan plastik”, dan “sedotan stainless” dengan jumlah data 3186 tweet. Akan dicari metode klasifikasi terbaik antara metode *k-nearest neighbor* (knn) dan *neighbor-weighted k-nearest neighbor* (nwknn). Sebelum menganalisis, dilakukan praproses data terlebih dahulu.

### 4.1 Praproses Data Tweet Masyarakat terhadap Isu “Sedotan Plastik”

Data yang diambil dari twitter masih berupa data kotor. Artinya data tersebut masih memuat duplikat, *username*, *link URL*, tanda baca, dan simbol-simbol lainnya, sehingga perlu dilakukan praproses guna mendapatkan data tweet yang tidak memuat hal-hal tersebut. Data tweet tersebut dilakukan praproses teks yang meliputi *case folding*, *replacing*, *cleansing*, *stopword*, *stemming*, dan *tokenizing*. Beberapa kata yang mengalami *replacing* terdapat pada Lampiran 5. Data hasil simulasi praproses ditunjukkan sebagaimana pada Tabel 4.1 berikut.

**Tabel 4.1** Praproses Data

---

Data Tweet	RT @KementerianLHK: Ayo kurangi sedotan plastik dan tas2 kantong plastik. <U+0001F60A>  @JavaJazzFest  #BeatPlasticPollution <a href="https://t.co/NDeGqXHF8e">https://t.co/NDeGqXHF8e</a> RT @KementerianLHK: Ayo kurangi sedotan plastik dan tas2 kantong plastik. <U+0001F60A>  @JavaJazzFest  #BeatPlasticPollution <a href="https://t.co/NDeGqXHF8e">https://t.co/NDeGqXHF8e</a>
------------	--

---

**Tabel 4.1** Praproses Data (lanjutan)

Menghapus duplikat	RT @KementerianLHK: Ayo kurangi sedotan plastik dan tas2 kantong plastik. <U+0001F60A> @JavaJazzFest
<i>Case Folding</i>	#BeatPlasticPollution <a href="https://t.co/NDeGqXHF8e">https://t.co/NDeGqXHF8e</a> rt @kementerianlhk: ayo kurangi sedotan plastik dan tas2 kantong plastik. <u+0001f60a> @javajazzfest
<i>Replacing</i>	#beatplasticpollution <a href="https://t.co/ndegqxhf8e">https://t.co/ndegqxhf8e</a> rt @kementerianlhk: ayo kurangi sedotan plastik dan tas2 kantong plastik. <u+0001f60a> @javajazzfest
Menghapus <i>Link</i>	#beatplasticpollution <a href="https://t.co/ndegqxhf8e">https://t.co/ndegqxhf8e</a> rt @kementerianlhk: ayo kurangi sedotan plastik dan tas2 kantong plastik. <u+0001f60a> @javajazzfest
Menghapus <i>Emoticon</i>	#beatplasticpollution rt @kementerianlhk: ayo kurangi sedotan plastik dan tas2 kantong plastik. @javajazzfest
Menghapus <i>Username</i>	#beatplasticpollution rt ayo kurangi sedotan plastik dan tas2 kantong plastik.
Menghapus Baris Baru	#beatplasticpollution rt: ayo kurangi sedotan plastik dan tas2 kantong plastik. #beatplasticpollution

**Tabel 4.1** Praproses Data (lanjutan)

Menghapus Angka	rt: ayo kurangi sedotan plastik dan tas kantong plastik. #beatplasticpollution
Menghapus Hashtag	rt: ayo kurangi sedotan plastik dan tas kantong plastik.
Menghapus Punctuation	rt ayo kurangi sedotan plastik dan tas kantong plastik
Menghapus Spasi Berlebih	rt ayo kurangi sedotan plastik dan tas kantong plastik
Stemming	rt ayo kurang sedotan plastik dan tas kantong plastik
Stopward	ayo kurang sedotan plastik tas kantong plastik

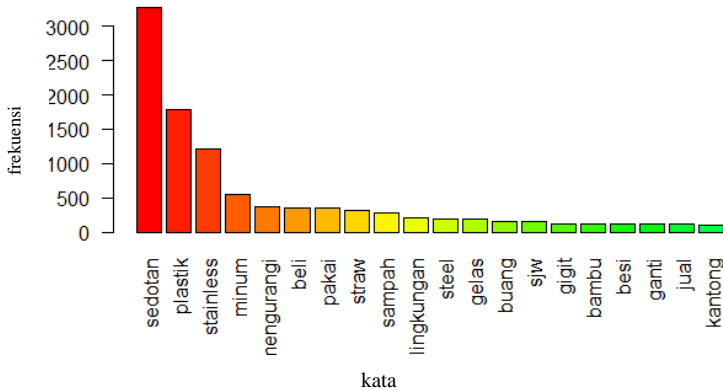
#### 4.2 Karakteristik Data Tweet Masyarakat terhadap Isu “Sedotan Plastik”

Setelah melakukan praprosesing data, proses *stopword* dan *tokenizing* dipakai sebagai kata kunci dari data tweet yang akan dianalisis dengan menggunakan metode KNN dan NWKNN. Tabel 4.2 menunjukkan perhitungan frekuensi tiap kata kunci pada tweet yang berkaitan dengan isu “sedotan plastik”.

**Tabel 4.2** *Count Vectorizer* pada data tweet

Tweet	beli	gelas	lingkungan	pakai	stainless	...
1	0	0	0	1	0	...
2	2	0	1	0	1	...
3	0	0	0	0	0	...
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
3.186	0	0	0	0	0	...

Berdasarkan Tabel 4.2 menunjukkan perhitungan frekuensi pada kata kunci. Kata “beli” disebutkan sebanyak 2 kali pada tweet kedua, tetapi tidak disebutkan pada tweet pertama, sedangkan kata “pakai” disebutkan sebanyak 1 kali pada tweet pertama, tetapi tidak disebutkan pada tweet kedua, dan seterusnya. Pada Gambar 4.1 ditampilkan 20 kata dalam tweet yang memiliki frekuensi kata paling banyak.



**Gambar 4.1.** Frekuensi 20 Kata Tertinggi

Berdasarkan Gambar 4.1 dapat diketahui frekuensi dari 20 kata tertinggi mengenai isu “sedotan plastik”, dimana distribusi kata yang ada mengikuti distribusi diskrit. Kata “sedotan”, “plastik”, dan “*stainless*” merupakan kata yang paling banyak dibicarakan oleh pengguna twitter dalam isu “sedotan plastik”. Hal ini dikarenakan tiga kata tersebut menjadi *keyword* yang digunakan dalam pengambilan data. Kata “*stainless*” dan (sjw) *social justice warrior*, dua kata ini sering disebut dikarenakan untuk mengurangi penggunaan sedotan plastik masyarakat menggantinya dengan sedotan *stainless* atau biasa dikenal sedotan sjw.

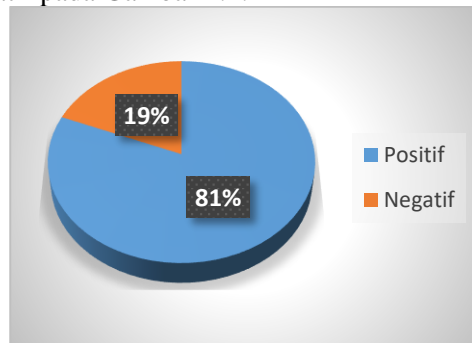
Hal yang dilakukan selanjutnya adalah pelabelan. Pelabelan dilakukan secara otomatis dengan *scoring*. Terdapat daftar kata positif dan negatif untuk membantu menentukan apakah tweet yang ada masuk dalam sentimen positif atau negatif. Proses pelabelan ini dapat memberikan informasi seberapa banyak masyarakat pengguna twitter yang memberikan pendapat positif atau negatif terhadap isu “Sedotan Plastik”. Namun, pelabelan dengan cara *lexicon based* tidak menjamin data akan 100% terklasifikasikan dengan sesuai. Terdapat beberapa tweet yang salah klasifikasi, setelah dilakukan pengecekan ditemukan 70 dari

1.000 tweet masih terklasifikasikan tidak sesuai. Contoh klasifikasi tweet yang tidak sesuai terdapat pada Tabel 4.3.

**Tabel 4.3** Contoh Klasifikasi Tweet Tidak Sesuai

No	Tweet	Sentimen	
		<i>Lexicon based</i>	Manual
1	mengurangi sampah sedotan dengan dikokop	Negatif	Positif
2	pakai stainless straw aneh karena biasa gigitin sedotan	Positif	Negatif

Hal ini dikarenakan masih ada beberapa kata positif maupun negatif yang menjadi kunci *lexicon based* belum masuk ke dalam daftar kata positif dan negatif. Daftar kata positif yang belum masuk adalah “nenggak” dan “dikokop”. Kata “kokop” sudah masuk dalam daftar kata negatif, namun karena kokop belum terdapat dalam Kamus Besar Bahasa Indonesia maka saat *stemming* imbuhan di- tidak dihilangkan. Daftar kata negatif yang belum masuk adalah “cuci”, “karbon”, dan “gigitin”. Imbuhan -in pada kata “gigitin” tidak hilang saat proses *stemming* sehingga masih menjadi kata “gigitin”. Dalam meminimalisir kesalahan klasifikasi dibutuhkan kejelian dalam penentuan kata kunci positif dan negatif yang digunakan untuk *lexicon based*, dimana harus sesuai dengan topik penelitian. Berdasarkan hasil pelabelan menggunakan cara *lexicon based* didapatkan presentase sentimen positif dan negatif pada Gambar 4.2.



**Gambar 4.2.** Presentase Sentimen Positif dan Negatif

Berdasarkan Gambar 4.2 diketahui bahwa tanggapan masyarakat pengguna twitter terhadap isu “sedotan plastik” lebih dominan positif daripada negatif. Sentimen positif memiliki rasio 81% atau sebesar 2.587 tweet, sedangkan sentimen negatif sebesar 19% atau 599 tweet. Contoh tweet yang termasuk kedalam sentimen positif dan negatif dapat dilihat pada Tabel 4.4.

**Tabel 4.4** Contoh Tweet Sentimen Positif dan Negatif

No	Positif	Negatif
1	komitmen tidak sedotan plastik	minum sedotan besi sensasi serasa ngemut antena jaman
2	mengurangi sedotan plastik minum dikokop	pakai stainless straw minum dingin auto gigi ngilu sedotan dingin gabisa gigit
3	konsisten menjaga lingkungan sampah plastik mudah ganti sedotan plastik	mahal sedotan besi

Beberapa contoh tweet menunjukkan bahwa pengguna twitter lebih banyak yang setuju untuk mengurangi ataupun mengganti sedotan plastik. Terdapat beberapa usaha yang dilakukan untuk mengurangi sedotan plastik, seperti minum langsung tanpa sedotan ataupun menggantinya dengan sedotan *stainless* atau yang bersifat ramah lingkungan. Namun dibalik usaha untuk mengurangi sedotan plastik tetap saja ada kekurangan, seperti sedotan besi yang memiliki rasa seperti antena, harga yang lebih mahal, dan terkadang tertinggal.

Twitter sebagai media sosial secara tidak langsung telah mempengaruhi pola hidup masyarakat. Banyaknya tweet yang mendukung untuk mengurangi penggunaan sedotan plastik telah membuat berbagai pihak mulai dari masyarakat, rumah makan, komunitas, hingga pemerintah merespon isu tersebut. Masyarakat mulai sadar untuk mengurangi sedotan plastik dengan membawa sedotan ramah lingkungan sendiri dan berusaha menolak jika akan diberi sedotan.

Beberapa rumah makan ternama yang menjadi salah satu penyumbang pemakaian sedotan plastik dalam jumlah besar juga melakukan aksi. Salah satunya adalah KFC, cara KFC dalam



mendukung pengurangan sedotan plastik dengan tidak menyediakan dispenser sedotan. KFC mencanangkan gerakan #Nostrawmovement di enam gerai KFC pada Mei hingga akhir tahun 2017. Pencanangan #Nostrawmovement atau Gerakan Tanpa Sedotan menjadi gerakan nasional KFC tepat pada hari Selasa, 8 Mei 2018, restoran *fast food Kentucky Fried Chicken* (KFC) resmi meluncurkan gerakan #NoStrawMovement di 630 gerai KFC yang tersebar di seluruh Indonesia. Gerakan ini diinisiasi oleh KFC Indonesia bersama dengan *Divers Clean Action*. Hingga akhir 2018 pengurangan tersebut sudah mencapai 91 persen.

Anthony Cotta selaku Direktur Starbucks Indonesia menjelaskan bahwa sejak bulan Juli 2018 Starbucks berkomitmen untuk tidak lagi menggunakan sedotan plastik di seluruh gerai pada tahun 2020. Program dimulai dengan gerai Starbucks Indonesia yang ada di Bali dan akan berlanjut ke-362 gerai seluruh Nusantara. Penggunaan sedotan plastik akan diganti dengan *sippy cup*, yakni sejenis gelas yang memiliki lubang untuk minum di sisinya. Setelah Starbucks berkomitmen untuk menghentikan penggunaan sedotan plastik pada tahun 2020. Hal ini diikuti oleh perusahaan lain, salah satunya McDonald's. Pada 12 November 2018 McDonald's Indonesia telah memulai gerakan #Mulaitanpasedotan. Larangan ini dilakukan serentak di 189 gerai McDonald's di seluruh Indonesia. McDonald's tidak lagi menyediakan tempat sedotan plastik kecuali pembeli meminta dengan alasan khusus.

Selain perusahaan makanan siap saji, kalangan bisnis pariwisata juga mulai beralih. Alcheringa Holiday, pemilik properti di Yogyakarta, Karimun Jawa, dan Manado, mulai menggunakan sedotan stainless. Pemilik Alcheringa Holiday, membeli 50 buah sedotan stainless dari Zero Waste Warriors untuk digunakan di seluruh propertinya. Banyaknya masyarakat serta perusahaan yang saat ini beralih dari sedotan plastik ke sedotan ramah lingkungan, berdampak pada tingginya permintaan sedotan stainless dan sedotan bambu. Salah satu Mahasiswa Universitas Pendidikan Indonesia, Muhammad Dicky Rifaldi yang menjadi

pengusaha sedotan bambu sejak 17 Januari 2019. Dicky bisa memproduksi 4.000-5.000 sedotan bambu per bulan. Selain Dicky, Alma Mahasiswa Universitas Maranatha juga membentuk gerakan sosial anti sedotan plastik. Aksi Alma diberi nama Sesapan. Dengan memanfaatkan teknologi dalam mengampanyekan Sesapan. Alma memaksimalkan lewat akun media social serta menyediakan lapak dagang sedotan *stainless* sebagai pengganti sedotan plastik. Pada bulan Februari selain berkampanye di media sosial, Alma dan teman kelompoknya mendatangi sekolah di kawasan Cimahi. Di sana menyediakan segudang informasi soal sampah plastik. Selain itu, Sesapan juga mulai menjalin relasi dengan bank sampah di daerah tersebut. Kedepannya Alma akan mengadakan kegiatan menukar sedotan plastik menjadi sedotan bambu.

### 4.3 *Term-Weighting*

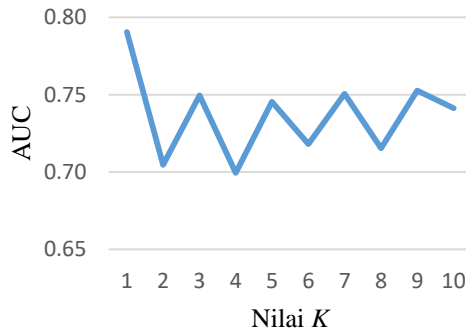
Metode *Term-Weighting* merupakan proses pembobotan yang dilakukan untuk mendapatkan nilai bobot suatu kata di dalam dokumen. Tahap ini juga disebut proses ekstraksi data yang semula merupakan data text kemudian dilakukan perubahan strukturnya menjadi numerik agar dapat dilakukan analisis selanjutnya yaitu klasifikasi. Metode yang digunakan adalah *Term FrequencyInverse Document Frequency (TF-IDF)*. Tabel 4.5 merupakan struktur data hasil ekstraksi *TF-IDF*.

**Tabel 4.5** Struktur Data *TF-IDF*

Tweet	Kata					
	beli	besi	ramah	sjw	Stainless	...
1	0	0	0	0	0	...
2	0,545	0	0,407	0	0,182	...
3	0	0	0	0	0	...
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
3186	0	0	0	0	0	...

#### 4.4 Klasifikasi Menggunakan *K-Nearest Neighbor*

Metode *K-Nearest Neighbor* (KNN) adalah mencari klasifikasi berdasarkan jarak  $K$  terdekat. Data terlebih dahulu dibagi menjadi data *training* dan *testing*. Dalam penelitian ini 90% data menjadi data *training* dan 10% menjadi data *testing*. Usaha untuk meningkatkan akurasi dilakukan dengan cara *k-fold cross validation*,  $k$  yang digunakan sebesar 10. Langkah berikutnya melakukan perhitungan jarak dan diurutkan dari yang terkecil. Penentuan klasifikasi berdasarkan kelas yang dominan pada data latih  $K$  jarak terdekat. Tidak ada ketentuan untuk menentukan nilai  $K$  terbaik. Oleh karena itu, untuk menentukan  $K$  terdekat dengan metode KNN pada tugas akhir ini, akan digunakan beberapa nilai  $K$  yang berbeda untuk melihat pengaruh nilai AUC yang dihasilkan sehingga ditemukan  $K$  yang paling optimum. Nilai  $K$  yang digunakan yaitu mulai dari  $K=1$  hingga  $K=10$ . Gambar 4.3 adalah grafik hasil pengukuran performa klasifikasi metode KNN.



**Gambar 4.3** AUC Metode KNN

Perfoma klasifikasi berupa AUC dari metode KNN memiliki hasil yang tergolong baik. Nilai AUC tertinggi saat  $K=1$  sebesar 0,79 dan cenderung menurun saat nilai  $K$  bertambah. Saat nilai tetangga  $k$  semakin besar dengan menggunakan algoritme KNN, suatu data uji akan memiliki probabilitas kesalahan untuk masuk ke kelas yang memiliki jumlah data latih terbanyak. Ketika  $K$  ganjil memiliki hasil AUC lebih baik dibanding  $K$  genap. Besaran nilai

akurasi, presisi, *recall*, *f-measure*, AUC, dan *Press's Q* pada setiap *fold* dengan  $K=1$  dapat dilihat pada Tabel 4.6.

**Tabel 4.6** Performa Klasifikasi KNN setiap *fold*  $K=1$

K	Akur	Pres	Rec	<i>f-measure</i>	AUC	<i>Press's Q</i>
1	0,88	0,92	0,94	0,93	0,79	189,11
2	0,87	0,92	0,92	0,92	0,79	176,09
3	0,87	0,9	0,94	0,92	0,76	170,19
4	0,86	0,93	0,91	0,92	0,77	167,28
5	0,85	0,92	0,89	0,91	0,76	158,7
6	0,86	0,91	0,91	0,91	0,8	167,28
7	0,86	0,92	0,91	0,91	0,8	169,26
8	0,89	0,93	0,93	0,93	0,84	193,41
9	0,86	0,95	0,89	0,92	0,8	169,26
10	0,85	0,93	0,88	0,9	0,79	152,2

Sebagaimana pada Tabel 4.6 dapat diketahui hasil nilai akurasi, presisi, *recall*, *f-measure*, AUC, dan *Press's Q* dengan metode KNN pada setiap *fold* dengan  $K=1$ . Akurasi pada tiap-tiap *fold* tidak berbeda jauh, hanya 0,85 sampai 0,89, begitu pula pada *f-measure* juga tidak berbeda signifikan. Nilai *recall* atau biasa disebut dengan sensitifitas menunjukkan yang terbaik ada pada *fold* ke-1 dan 3, namun AUC tetap menjadi patokan utama dalam menilai performa klasifikasi. Nilai AUC tertinggi terdapat pada *fold* ke-8 sebesar 0,84. Berdasarkan hasil perhitungan statistik *Press's Q* untuk masing-masing *fold* lebih besar dari  $\chi^2_{0.05,1} = 3,84$  maka dapat dikatakan pengklasifikasian sentimen isu “sedotan plastik” dengan metode *K-Nearest Neighbor* (K-NN) signifikan secara statistik.

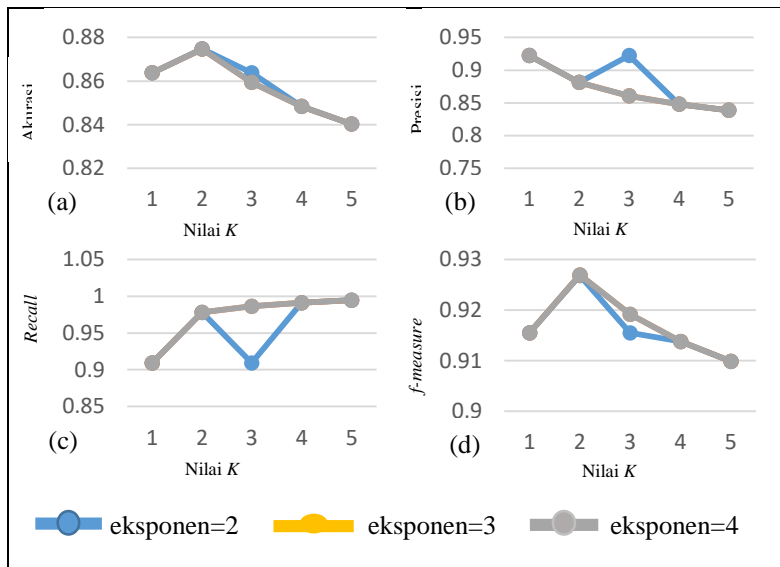
#### 4.5 Klasifikasi Menggunakan *Nearest weighted-KNN*

Metode kedua yaitu *Nearest weighted-KNN* dimana metode ini pengembangan metode KNN, perbedaannya terletak pada pemberian bobot. NWKNN tergolong metode yang masih jarang digunakan, sehingga *package* dalam *software* R atau Python belum

terfasilitasi. Peneliti membuat program sendiri menggunakan Phyton yang terdapat pada Lampiran 4. Jarak terdekat yang digunakan menggunakan cara *euclidiende distane*.

#### 4.5.1 Penentuan Nilai Eksponen Terbaik

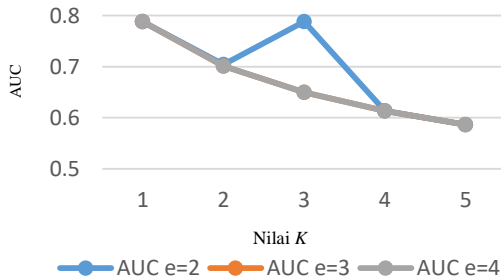
Metode NWKNN memiliki eksponen dalam penentuan bobot, nilai eksponen lebih dari satu. Gambar 4.4. merupakan perbedaan akurasi antara NWKNN dengan eksponen=2, eksponen=3, dan eksponen=4.



**Gambar 4.4** Perbandingan Performa Klasifikasi  $K=1$  sampai  $K=5$  (a) Akurasi, (b) Presisi, (c) *Recall*, dan (d) *f-measure*

Gambar 4.4 menunjukkan hasil performa klasifikasi metode NWKNN mulai dari akurasi, presisi, *recall*, dan *f-measure* menggunakan eksponen 2, 3, dan 4. Nilai akurasi, presisi, *recall*, dan *f-measure* eksponen=3 sama dengan eksponen=4. Namun berbeda saat eksponen=2, hasil akurasi dan presisi lebih tinggi dibandingkan dengan eksponen=3 dan eksponen=4. Berkebalikan dengan akurasi dan presisi, hasil *recall* dan *f-measure* saat eksponen=2 lebih rendah dibandingkan dengan eksponen=3 dan

eksponen=4. Hasil *recall* eksponen=2 yang lebih rendah mempengaruhi hasil *f-measure* juga menjadi lebih kecil. Terdapat satu pengukuran performa klasifikasi lagi, yaitu AUC. AUC lebih merepresentasikan hasil performa klasifikasi untuk data yang tidak setimbang atau *inbalance*. Gambar 4.5 merupakan perbedaan AUC antara NWKNN dengan eksponen=2, eksponen=3, dan eksponen=4.

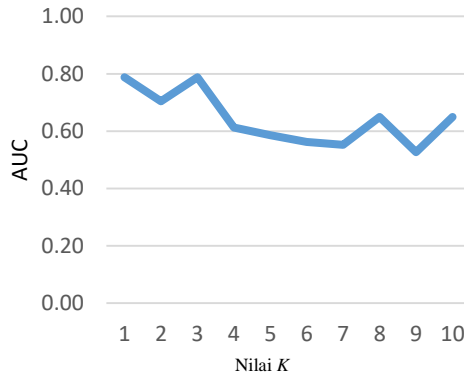


**Gambar 4.5** AUC Metode NWKNN dengan Eksponen Berbeda

Gambar 4.5 menunjukkan hasil AUC metode NWKNN menggunakan eksponen 2, 3, dan 4 dengan  $K=1$  sampai  $K=5$ . Nilai AUC eksponen=3 sama dengan eksponen=4 seperti pengukuran performa klasifikasi sebelumnya yaitu akurasi, presisi, *recall*, dan *f-measure*. Saat eksponen=3, hasil AUC lebih tinggi dibandingkan dengan eksponen=3 dan eksponen=4. Hal ini menandakan bahwa penggunaan eksponen terbaik saat eksponen=2.

#### 4.5.2 Penentuan $K$ Terbaik

Penentuan nilai  $K$  sama dengan metode KNN, tidak ada ketentuan pasti untuk menentukan nilai  $K$  terbaik. Oleh karena itu, nilai  $K$  yang digunakan yaitu mulai dari  $K=1$  hingga  $K=10$  dengan eksponen=2. Penggunaan eksponen=2 karena dalam penentuan nilai eksponen terbaik yang telah dilakukan menunjukkan hasil AUC eksponen=2 lebih baik dibandingkan dengan eksponen=3 dan eksponen=4. Gambar 4.6 adalah grafik hasil AUC metode NWKNN.



**Gambar 4.6** AUC Metode NWKNN  $K=1$  sampai  $K=10$

Nilai AUC tertinggi ketika  $K=1$  dan  $K=3$  sebesar 0,79. Saat nilai  $K$  kecil seperti 1, 2, dan 3 nilai AUC masih tergolong baik, namun saat  $K=4$  sampai  $K=7$  mengalami penurunan dan kembali mengalami kenaikan saat  $K=8$ . Setelah terpilih  $K$  yang optimum akan ditampilkan besaran nilai akurasi, presisi, *recall*, *f-measure*, AUC, dan *Press's Q* pada setiap *fold* dengan  $K=1$  pada Tabel 4.7.

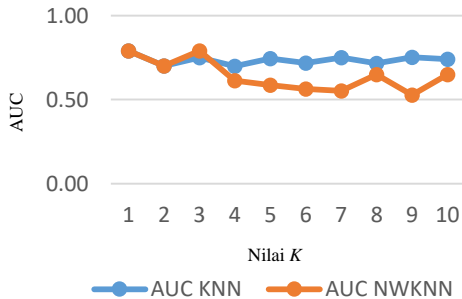
**Tabel 4.7** Perfoma Klasifikasi NWKNN setiap *fold*  $K=1$

K	Akurasi	Presisi	<i>Recall</i>	<i>f-measure</i>	AUC	<i>Press's Q</i>
1	0,87	0,92	0,92	0,92	0,76	170,19
2	0,87	0,92	0,92	0,92	0,79	176,08
3	0,87	0,90	0,94	0,92	0,75	170,19
4	0,86	0,93	0,91	0,92	0,77	167,28
5	0,85	0,92	0,89	0,91	0,79	158,7
6	0,86	0,91	0,91	0,91	0,80	167,28
7	0,86	0,92	0,91	0,91	0,80	169,26
8	0,88	0,93	0,92	0,93	0,83	187,22
9	0,86	0,95	0,89	0,92	0,79	169,26
10	0,85	0,93	0,88	0,90	0,79	152,2

Diketahui hasil nilai akurasi, presisi, *recall*, *f-measure*, AUC, dan *Press's Q* dengan metode NWKNN pada setiap *fold* dengan  $K=1$ . Akurasi pada tiap-tiap *fold* tidak berbeda jauh, akurasi tertinggi pada *fold* ke-8 dan terendah pada *fold* ke-5 dan ke-10. Begitu pula pada *f-measure* juga tidak berbeda signifikan hanya 0,9 sampai 0,93. Hasil presisi tertinggi terdapat pada *fold* ke-9 sebesar 0,95. Nilai *recall* atau biasa disebut dengan sensitifitas menunjukkan yang terbaik ada pada *fold* ke-3, memiliki arti bahwa kemampuan metode *KNN* saat  $K=1$  *fold* ke-3 dapat mengklasifikasikan sentimen positif dengan tepat sebesar 0,94. Namun AUC tetap menjadi patokan utama dalam menilai performa klasifikasi. Hasil AUC tertinggi terdapat pada *fold* ke-8 sebesar 0,83, serta AUC terendah pada *fold* ke-3. Berdasarkan hasil perhitungan statistik *Press's Q* untuk masing-masing *fold* lebih besar dari  $\chi^2_{0,05,1}=3,84$  maka dapat dikatakan pengklasifikasian sentimen isu “sedotan plastik” dengan metode NWKNN signifikan secara statistik.

#### 4.6 Perbandingan Metode KNN dan NWKNN

Gambar 4.3 dan 4.6 telah menunjukkan hasil pengukuran performa klasifikasi AUC dengan nilai  $K=1$  sampai  $K=10$ . Gambar 4.7 merupakan perbandingan nilai AUC secara bersamaan.



**Gambar 4.7** Perbandingan KNN dan NWKNN

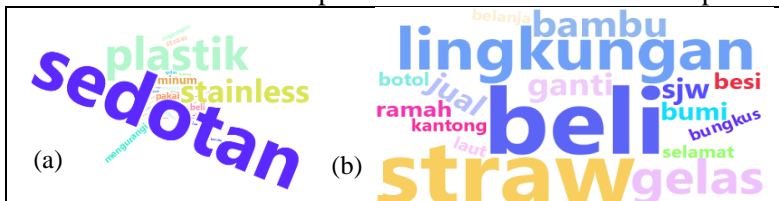
Hasil performa klasifikasi antara metode KNN dan NWKNN dalam mengklasifikasikan data twitter mengenai isu “sedotan



plastik” menunjukkan bahwa saat  $K=1$  dan  $K=2$  nilai AUC yang dihasilkan sama. Namun saat  $K=3$  metode NWKNN lebih baik 0,04 dibandingkan metode KNN. Setelah  $K=4$  hasil AUC metode NWKNN mengalami penurunan, hal ini disebabkan spesifisitas yang rendah.

#### 4.7 Visualisasi Word Cloud

Visualisasi data teks menggunakan *word cloud* digunakan untuk mengetahui kata-kata yang paling sering muncul pada data. Pada penelitian ini, *word cloud* digunakan untuk visualisasi tweet berdasarkan kategori sentimennya sehingga dapat diketahui kata-kata yang sering muncul pada setiap sentimen. Semakin besar ukuran *font* berarti semakin besar frekuensi kemunculan kata tersebut. Gambar 4.8 merupakan *word cloud* dari sentimen positif.

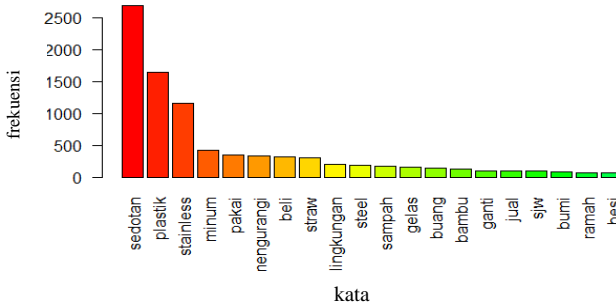


**Gambar 4.8** Wordcloud Sentimen Positif (a) Sebelum Menghilangkan Kata dengan Frekuensi Besar (b) Setelah Menghilangkan Kata dengan Frekuensi Besar

*Wordcloud* sentimen positif pada Gambar 4.8 (a) menunjukkan kata-kata yang sering muncul. Kata “sedotan”, “plastik”, “stainless” menjadi kata dengan ukuran paling besar, karena tiga kata tersebut menjadi *keyword* dalam pengambilan data. Lalu ada kata “pakai” yang ukurannya hampir setara dengan kata “stainless”, pengguna twitter sangat mendukung pemakaian sedotan *stainless* sebagai pengganti sedotan plastik. Kata selanjutnya yaitu “mengurangi”, pengguna twitter juga saling mendukung dan memberikan saran agar mengurangi pemakaian sedotan plastik. Beberapa kata terlihat sangat kecil sehingga kurang bisa dibaca karena frekuensi kata “sedotan”, “plastik”, “stainless” yang berbeda jauh dengan kata yang lain. Kata yang memiliki frekuensi sangat besar akan di hapus agar kata yang lain lebih terlihat jelas. Gambar 4.8 (b) menunjukkan *wordcloud*

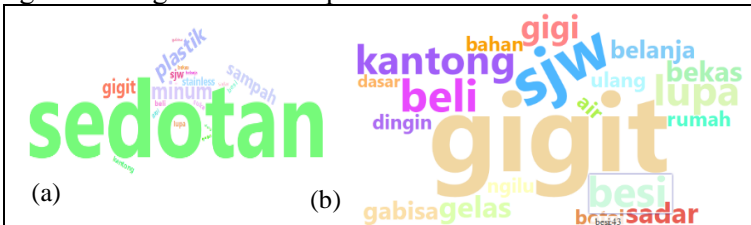
sentimen positif dengan menghilangkan beberapa kata yang memiliki frekuensi sangat besar.

Terdapat kata “besi” dan “bambu” yang merupakan sedotan alternatif untuk mengganti sedotan plastik. Kata “kokop” menandakan selain mengganti sedotan plastik dengan sedotan ramah lingkungan bisa juga dengan cara langsung kokop atau langsung minum. Terdapat pula kata “lingkungan”, “bumi”, dan “laut” yang berarti pengguna twitter peduli dengan keadaan sekitar. Selain mengetahui kata apa yang sering muncul, juga dapat diketahui jumlah katanya. Gambar 4.9 merupakan *Barchart* kata dari sentimen positif yang sering muncul.



**Gambar 4.9** Frekuensi Kata Sentimen Positif

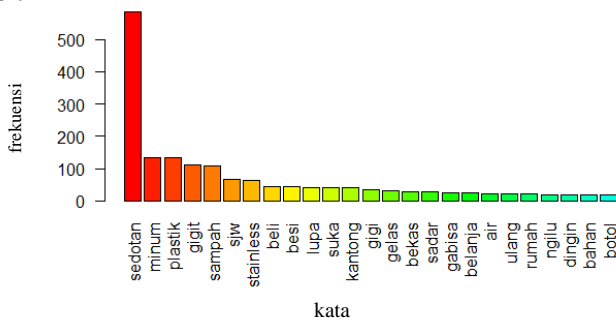
Gambar 4.9 menunjukkan kata yang muncul lebih dari 50 kali dalam tweet yang tergolong sentimen positif isu “sedotan plastik”. Setelah mengetahui kata-kata yang sering muncul dalam sentimen positif, maka Gambar 4.10 menampilkan *wordcloud* dari sentimen negatif tentang isu “sedotan plastik”.



**Gambar 4.10** *Wordcloud* Sentimen Negatif (a) Sebelum Menghilangkan Kata dengan Frekuensi Besar (B) Setelah Menghilangkan Kata dengan Frekuensi Besar

*Wordcloud* sentimen negatif pada Gambar 4.10 (a) menunjukkan kata-kata yang sering muncul. Sama dengan yang terjadi pada *wordcloud* sentiment positif, kata “sedotan”, “plastik”, “*stainless*” menjadi kata dengan ukuran paling besar, karena tiga kata tersebut menjadi *keyword* dalam pengambilan data. Kata yang memiliki frekuensi sangat besar akan di hapus agar kata yang lain lebih terlihat jelas. Gambar 4.10 (b) menunjukkan *wordcloud* sentimen negatif dengan menghilangkan beberapa kata.

Terdapat kata “gigit”, berarti pengguna twitter beropini negatif mengenai isu “sedotan plastik” karena bentuk sedotan *stainless* maupun besi yang kaku membuat masyarakat tidak bisa menggigit sedotan tersebut. Kebiasaan menggigit sedotan plastik bagi beberapa orang belum bisa ditinggalkan, mereka suka menggigit-gigit sedotan hingga gepeng. Kata yang sering muncul selanjutnya adalah “lupa”. Masyarakat yang terbiasa menggunakan sedotan plastik pasti akan langsung pergi setelah minumannya habis, kebiasaan ini terbawa saat mulai beralih ke sedotan *stainless* atau sedotan ramah lingkungan yang lain. Saat membawa sedotan ramah lingkungan sendiri secara tidak sadar masyarakat langsung pergi, sehingga sedotan yang ia miliki tertinggal karena lupa. Ada juga kata “dingin”, “gigi”, dan “ngilu” menandakan saat memakai sedotan *stainless* yang bersifat pengantar yang baik maka ketika minuman yang diminum dingin, sedotan *stainless* juga terasa dingin. Rasa dingin pada sedotan *stainless* membuat beberapa orang yang memiliki gigi sensitif merasakan ngilu.



**Gambar 4.11** Frekuensi Kata Sentimen Negatif

Gambar 4.11 menunjukkan kata yang muncul lebih dari 20 kali dalam tweet yang tergolong sentimen negatif isu “sedotan plastik”.

## BAB V KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan pada bab 4, maka diperoleh kesimpulan sebagai berikut.

1. Metode NWKNN memberikan hasil lebih baik daripada metode KNN ketika nilai  $K$  masih bernilai kecil yaitu  $K=1$  sampai  $K=3$ . Setelah  $K=4$  hasil AUC metode NWKNN mengalami penurunan, hal ini disebabkan spesifisitas yang rendah.
2. Hasil visualisasi *wordcloud* dalam sentimen positif menunjukkan terdapat kata “besi” dan “bambu” yang merupakan sedotan alternatif untuk mengganti sedotan plastik. Kata “kokop” menandakan selain mengganti sedotan plastik dengan sedotan ramah lingkungan bisa juga dengan cara langsung kokop atau langsung minum. *Wordcloud* semtimen negatif terdapat kata “gigit”, berarti pengguna twitter beropini negatif mengenai isu “sedotan plastik” karena bentuk sedotan *stainless* maupun besi yang kaku membuat masyarakat tidak bisa menggigit sedotan tersebut. Kata yang sering muncul selanjutnya adalah “dingin”, “gigi”, dan “ngilu” menandakan saat memakai sedotan *stainless*.

### 5.2 Saran

Berdasarkan kesimpulan yang diperoleh, dapat dirumuskan saran sebagai pertimbangan penelitian selanjutnya adalah sebagai berikut.

1. Lebih teliti dalam pemilihan *lexicon base* yang akan digunakan dalam penentuan label serta penggunaan *stopword*.
2. Lebih banyak mengadakan percobaan dengan mengganti-ganti parameter, sehingga didapatkan hasil yang optimal.
3. Mencoba menggunakan cara *Cosine Similarity* dalam penentuan tetangga terdekat.

4. Kepada masyarakat untuk lebih peduli lagi terhadap lingkungan dengan meminimalisir penggunaan sedotan plastik.
5. Kepada produsen sedotan ramah lingkungan untuk mengembangkan ide dalam menghasilkan produk pengganti sedotan plastik agar masyarakat lebih tertarik menggunakan.

## DAFTAR PUSTAKA

- Ariadi, D., & Fithriasari, K. (2015). Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer. *Jurnal Sains dan Seni ITS*, 4(2).
- Agustin, D. (2018). *McDonald's Indonesia tidak Lagi Sediakan Sedotan*. Retrieved 27, 2019, from <https://www.republika.co.id/berita/gaya-hidup/kuliner/18/11/12/pi2qfp328-mcdonalds-indonesia-tidak-lagi-sediakan-sedotan>.
- Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation Measure for Models Assessment over Imbalanced Data Sets. *Journal of Information Engineering and Applications*, 3(10), 27-38.
- Bing, L. (2010). *Handbook of Natural Language Processing Second Edition*. Boca Raton: CRC Press.
- Buntoro, G. A., Adji, T., & Purnamasari, A. E. (2014). Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation. *The 6th Conference on Information Technology and Electrical Engineering* (pp. 39-43). Yogyakarta: Jurusan Teknik Elektro dan Teknologi Informasi, FT UGM.
- Chandani, V., Wahono, R. S., & Purwanto. (2015). Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film. *Journal of Intelligent Systems*, 1(1), 56-58.
- Daniel, M., Neves, R. F., & Horta, N. (2017). Company event popularity for financial markets using Twitter and sentiment analysis. *Expert Systems with Applications*, 71, 111-124.

- Darujati, C. dan Gumelar, A.B. (2012). Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks Bahasa Indonesia. *Jurnal Bandung Text Mining*, 16(1), 5-1.
- Dragut, E., Fang, F., Sistla, P., Yu, C., & Meng, W. (2009). Stop Word And Related Problem in Web Interface Integration. *Proceedings of the VLDB Endowment* (pp. 349-360). Lyon: ACM.
- Effendy, V. (2015). Analisis Sentimen Berbahasa Indonesia Dengan Pendekatan Lexicon Based (Studi Kasus : Solusi Pengelolaan Sampah). *Jurnal Ilmiah Komputer dan Informatika*, 4, 55-59.
- Faldy, H., Indrianti, & Novanto, Y. (2015). *Analisis Sentimen Opini Film Berbahasa Indonesia Berbasis Kamus Menggunakan Metode NeighborWeighted K-Nearest Neighbor (NWKNN)*. Malang: Universitas Brawijaya.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook*. New York: Cambridge University Press.
- Feldman, R., & Sanger, J. (2012). *The Text Mining Hand Book Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.
- Hair. J. F., Black. W. C., Babin. B. J., Anderson R. E. (2006). *Multivariate Data Analysis. Seventh Edition*. Pearson Education Prentice Hall. Inc.
- Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques* (2 ed.). San Fransisco: Morgan Kaufman Publisher.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concept and Techniques, Second Edition*. Waltham: Morgan Kaufmann Publishers.
- Hirat, R., & Mittal, N. (2015). A Survey On Emotion Detection Techniques using Text Text in Blogposts. *International Bulletin of Mathematical Research*, 2(1), 180-187.



- Hu, M., & Liu, B. (2004). Dipetik February 12, 2019, dari <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>.
- Indriati, & Ridok, A. (2016). Sentiment Analysis For Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (NWKNN). *Journal of Environmental Engineering & Sustainable Technology*, 3(1), 23-32.
- Jambeck, J. R.; Geyer, R.; Wilcox, C.; Perryman, M.; Andrady, A.; Narayan, R.; Siegler, T.; Law, K. L. (2015). *Plastic waste inputs from land into the ocean*, 347(6223), 768-771.
- Jannah, S., Fithriasari, K., & Usugawa, T. (2018). *Clustering and Visualizing Surabaya Citizen Aspirations by Using Text Mining*. Surabaya: Departemen Statistika Institut Teknologi Sepuluh Nopember.
- Kumar, V., & Cadha, A. (2012). Mining Association Rules in Student's Assessment Data. *International Journal of Computer Science Issues*, 9(3), 211-216.
- Kostsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Data Preprocessing for Supervised Learning. *International Journal of Computer Science*, 1(2), 111-117.
- Liu, B. (2010). *Processing, Handbook of Natural Language chapter Sentiment Analysis and Analysis* (2 ed.). Chicago: University of Illinois.
- Manning, D, Christopher. Raghavan, Prabhakar & Schütze, Hinrich. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. *Proceedings of the Seventh International Workshop on Semantic Evaluation Exercises (SemEval-2013)*. Atlanta, USA.

- Montgomery, D. C & Peck, E. A. (1992). *Introduction To Linier Regression Analysis Second Edition*. New York : John Wiley&Sons.
- Mustafa, A., Akbar, A., & Sultan, A. (2009). Knowledge Discovery using Teks Mining: A Programmable Implementation on Information Extraction and Categorization. *International Journal of Multimedia and Ubiquitous Engineering*, 4(2), 183-188.
- Pant, G., Srinivasan, P., & Menczer, F. (2004). *Crawling the Web*. Springer, 153-177.
- Pratomo, Y. (2019). *Untuk Pertama Kali, Twitter Ungkap Jumlah Pengguna Harian*. Retrieved 2 7, 2019, from <https://tekno.kompas.com/read/2019/02/09/11340027/untuk-pertama-kali-twitter-ungkap-jumlah-pengguna-harian>.
- Prayitno, G. (2018). *Seekor Paus Physeter Macrocephalus Ditemukan Mati dengan Sampah Plastik 5,9 kg di Perutnya*. Retrieved 2 7, 2019, from <http://travel.tribunnews.com/2018/11/20/seekor-paus-physeter-macrocephalus-ditemukan-mati-dengan-sampah-plastik-59-kg-di-perutnya>.
- Pour, M. (2008). *Encyclopedia of Information Science and Technology*, 8, 3934-3939. United States: IGI Global.
- Ridok, A. and Latifah, R. (2015). 'Klasifikasi Teks Bahasa Indonesia Pada Corpus Tak Seimbang Menggunakan NWKNN'. *Konferensi Nasional Sistem dan Informatika 2015*, pp. 222–227.
- Rifqi, N., Maharani, W., & Shaufiah. (2011). *Analisis Dan Implementasi Klasifikasi Data Mining Menggunakan Jaringan Syaraf Tiruan Dan Evolution Strategis*. Bali: Konferensi Nasional Sistem dan Informatika.
- Rozi, I. F., Pramono, S. H., & Dahlan, E. A. (2012). *Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data*

- Opini Publik pada Perguruan Tinggi. *Jurnal EECCIS*, 6(1), 37-42.
- Schroeder, J. E. (2002). *Visual Consumption*. London: Routledge.
- Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing and Management*, 45(4), 427-437.
- Sun, Y., Wong, A. K., & Kamel, M. S. (2009). Classification of Imbalanced Data: A Review. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(4), 687-719.
- Tan, S. (2005). Neighbor-weighted Knearest neighbor for unbalanced text corpus. *Expert Systems with Applications*, 28, 667-671.
- Weiss, S. (2010). *Text mining: Predictive Methods for Analyzing Unstructured Information*. New York: Springer.

*(Halaman ini sengaja dikosongkan)*

## LAMPIRAN

### Lampiran 1. Syntax *Crawling*

```
library(twitteR)
options(httr_oauth_cache=T)
rconsumer_key <- 'JgtAKEJUEsGq2B6h36G7bpRYm'
consumer_secret <-
  'v2ZDIy7LuIKW9R1fwhFZFS4TyYbYCURVQWmxMVAwK
  7Fo4KNHQJ'
access_token <- '72809642-
  BmENMhkPCNpZrHv7c9mCSphdHDhJz7Ixb3kbVMDKA'
access_secret <-
  '2OI6M8UUeg7uavSqa4BO08G9Nrtn2Sb1RkG7SewMESXd5'
setup_twitter_oauth(rconsumer_key, consumer_secret, access_token,
  access_secret)
tj <- searchTwitter('sedotan', lang="id", n=100000,
  resultType="recent")
write.csv(twListToDF(tj), file="sedotan.csv")
```

### Lampiran 2. Syntax *Preprocessing*

```
#Remove duplicate
tweets=read.csv("C:/Users/user/Documents/tweets.csv")
sum(duplicated(tweets))
tweets <- tweets[!duplicated(tweets), ]
write.csv(tweets, "C:/Users/user/Documents/tweets.csv")

import pandas as pd
import re
import nltk
from Sastrawi.Stemmer.StemmerFactory import
StemmerFactory
from nltk.tokenize import word_tokenize
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
import string
import csv
```

```

from nltk.stem import PorterStemmer
import numpy as np
from nltk.tokenize import sent_tokenize, word_tokenize
from string import digits
from nltk.stem import WordNetLemmatizer
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn.metrics import f1_score, precision_score,
recall_score, accuracy_score
from sklearn.metrics import classification_report
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix
from sklearn import model_selection
import imblearn
from sklearn import metrics
from sklearn.model_selection import cross_val_score, KFold

#Import Data
data = pd.read_csv(r'C:/Users/user/Documents/testing.csv')
data=data['text']

=====

#ubah huruf kecil
data = data.apply(lambda x: " ".join(x.lower() for x in x.split()))

=====

#replace
kata = {"stenlis":"stainless",
dan seterusnya ada pada Lampiran X.}
def replace_all(text, dic):
    for i, j in dic.items():
        text = text.replace(i, j)
    return text
import collections
from collections import OrderedDict

```

```
dic = OrderedDict(kata)

datachange = []
for line in data:
    result = replace_all(line, dic)
    datachange.append(result)

=====

#ubah ke dataframe
data=pd.DataFrame(datachange,columns=["text"])
data.head(8)

=====

#menyimpan ke csv
data.to_csv('C:/Users/user/Documents/testing1.csv')

=====

#Import Data
data = pd.read_csv(r'C:/Users/user/Documents/testing1.csv')
text = data['text']
text.head(8)

=====

#Menghapus Link
datanolink = []
for line in text:
    result = re.sub(r"http\S+", " ", line)
    datanolink.append(result)

=====

#Menghapus emoticon
datanoemoticon=[]
for line in datanolink :
    result = re.sub(r'<.*?>', "",line)
    datanoemoticon.append(result)
```

```
=====
# Menghapus Simbol Retweet
```

```
datanort = []
```

```
for line in datanoemoticon:
```

```
    result = re.sub(r"rt ", " ", line)
```

```
    datanort.append(result)
```

```
=====
# Menghapus Username
```

```
datanousername = []
```

```
for line in datanort:
```

```
    result = re.sub(r"@S+", " ", line)
```

```
    datanousername.append(result)
```

```
=====
#Menghapus baris baru
```

```
datanoline=[]
```

```
for line in datanousername :
```

```
    result=re.sub("\n", " ",line)
```

```
    datanoline.append(result)
```

```
=====
#Menghapus angka
```

```
datanonum=[]
```

```
for line in datanoline :
```

```
    result=re.sub("\d", " ",line)
```

```
    datanonum.append(result)
```

```
=====
#Menghapus hastag
```

```
datanohastag=[]
```

```
for line in datanonum :
```

```
    result=re.sub(r"#S+", "",line)
```

```
    datanohastag.append(result)
```



```

=====
#Menghapus punctuation
datanopunctuation=[]
for line in datanohastag :
    result=re.sub(r"^[^w\s]", "",line)
    datanopunctuation.append(result)

=====

#Menghapus spasi berlebih
datanodoublespace=[]
for line in datanopunctuation :
    result=re.sub(r"\s+', ' ',line)
    datanodoublespace.append(result)

=====

#Stemming
factory = StemmerFactory()
stemmer = factory.create_stemmer()
data_stemmed = map(lambda x: stemmer.stem(x), data_lower)
datastemmed = list(data_stemmed)

data=pd.DataFrame(datastemmed,columns=["bersih"])

=====

data=pd.DataFrame(datastemmed,columns=["bersih"])
data = pd.read_csv(r'C:/Users/user/Documents/testing2.csv')
data=data['bersih']

=====

stopword =
open('C:/Users/user/Documents/stopwords.txt','r').read()
review = review.apply(lambda x: " ".join(x for x in x.split() if x
not in stopword))

```

**Lampiran 3. Syntax Labelling**

```

library(tm)
library(stringr)

kalimat2<-
read.csv("C:/Users/user/Documents/testing3.csv",header=TRUE)

#Scoring
positif <-
scan("C:/Users/user/Documents/Positif.txt",what="character",comment.char=";")
negatif <-
scan("C:/Users/user/Documents/Negatif.txt",what="character",comment.char=";")
kata.positif = c(positif, "is near to")
kata.negatif = c(negatif, "cant")
score.sentiment = function(kalimat2, kata.positif, kata.negatif, .progress='none') {
  require(plyr)
  require(stringr)
  scores = laply(kalimat2, function(kalimat, kata.positif, kata.negatif)
  {
    kalimat = gsub("[[:punct:]]", "", kalimat)
    kalimat = gsub("[[:cntrl:]]", "", kalimat)
    kalimat = gsub("\\d+", "", kalimat)
    kalimat = tolower(kalimat)
    list.kata = str_split(kalimat, "\\s+")
    kata2 = unlist(list.kata)
    positif.matches = match(kata2, kata.positif)
    negatif.matches = match(kata2, kata.negatif)
    positif.matches = !is.na(positif.matches)
    negatif.matches = !is.na(negatif.matches)
    score = sum(positif.matches) - (sum(negatif.matches))
  }
  )
}

```

```

    return(score)
  }, kata.positif, kata.negatif, .progress=.progress )
  scores.df = data.frame(score=scores, text=kalimat2)
  return(scores.df)
}
hasil = score.sentiment(kalimat2$text, kata.positif, kata.negatif)

#CONVERT SCORE TO SENTIMENT
hasil$klasifikasi<- ifelse(hasil$score<0, "Negatif", "Positif")
hasil$klasifikasi

```

#### **Lampiran 4. Syntax Klasifikasi**

```

vec1 = TfidfVectorizer(min_df=15)
tfidf = vec1.fit_transform(data)
tfidf1 = pd.DataFrame(tfidf.toarray(),
                      columns=vec1.get_feature_names())
tfidf1_=pd.concat([df['sentimen'],tfidf1], axis=1)
tfidf1_.head()

tfidf_new=tfidf1.values

Y=pd.DataFrame(sentimen)
Y1=pd.DataFrame.as_matrix(Y)

Y1_1=np.ravel(Y1)

X_new, Y_new=tfidf_new, Y1_1

#KNN
kfold = KFold(n_splits=10, random_state = 100, shuffle=True)
for train, test in kfold.split(X_new, Y_new):
    (X_new[train], X_new[test])
    (Y_new[train], Y_new[test])
    knn = KNeighborsClassifier(n_neighbors=n).fit(X_new[train],
    Y_new[train])
    prediksi = knn.predict(X_new[test])
    print("Presisi Prediksi = ", precision_score(Y_new[test], prediksi)

```

```

print("Recall Prediksi = ", recall_score(Y_new[test], prediksi)
print("f-measure Prediksi = ", f1_score(Y_new[test], prediksi)

#NWKNN
class DataPoint:
    def __init__(self, features, label = None):
        self.features = features
        self.label = label
        self.match = 0

    def __iter__(self):
        for f in self.features:
            yield f
        yield self.label

def euclideanDistance(instance1,instance2):
    distance=0
    distance=int(distance)
    for i in range(len(instance1.features)):
        m=instance1.features[i]
        m=float(m)
        y=instance2.features[i]
        y=float(y)
        distance=distance + pow((m - y), 2)
    return math.sqrt(distance)

def knn(k,x_i,dataPoints):
    knearest = []

    for x_j in dataPoints:
        d_j = euclideanDistance(x_i, x_j)
        x_j.match=d_j
        #print(d_j)

        if len(knearest) < k:
            knearest.append(x_j)
            knearest.sort(key = lambda x: x.match)
        elif d_j < knearest[k-1].match:
            #print("sebelum sort")

```

```

        #print(d_j)
        #print('Before lists are:', '\n'.join([str(lst.match) for lst in
knearest]), sep='\n')
        knearest[k-1] = x_j
        #print('Append lists are:', '\n'.join([str(lst.match) for lst in
knearest]), sep='\n')
        knearest.sort(key = lambda x: x.match)
        #print('New lists are:', '\n'.join([str(lst.match) for lst in
knearest]), sep='\n')

labels = {}
nearest = (None, 0)
awal = 0

#print("Tetangga")
#vote
for neighbor in knearest:
    #print(neighbor.match)
    if neighbor.label in labels: labels[neighbor.label] += 1
    else: labels[neighbor.label] = 1

    if awal==0:
        nearest = (neighbor.label, labels[neighbor.label])
        awal = 1
    elif labels[neighbor.label] > nearest[1]:
        nearest = (neighbor.label, labels[neighbor.label])
#print(nearest[0].label)
return nearest[0]

##TEST NWKNN
def nwknk(k,e,x_i,dataPoints):
    knearest = []
    #Jarak terdekat diurutkan
    for x_j in dataPoints:
        #d_j = distance(x_i, x_j)
        d_j = euclideanDistance(x_i,x_j)
        x_j.match=d_j
        """"
    if x_j.match==0:

```

```

print(x_j.label)
print(x_j.features)
print(x_i.features)
"""
if len(knearest) < k:
    knearest.append(x_j)
    knearest.sort(key = lambda x: x.match)
elif d_j < knearest[k-1].match:
    knearest[k-1] = x_j
    knearest.sort(key = lambda x: x.match)

labels = { }
sumsDist = { }

#vote
for neighbor in knearest:
    if neighbor.label in labels:
        labels[neighbor.label] += 1
        sumsDist[neighbor.label] += neighbor.match
    else:
        labels[neighbor.label] = 1
        sumsDist[neighbor.label] = neighbor.match

#min=sorted(labels.items(), reverse=True)
minNeighbors=sorted(labels.items())[0][1]
#print(min)
weights={ }
for key in labels:
    #print("key: %s , value: %s" % (key, labels[key]))
    weights[key]=1/pow((labels[key])/minNeighbors , (1/e))

#Score
scores={ }
for key in labels:
    scores[key]=weights[key]*sumsDist[key]

minscores=scores
minscores=sorted(minscores.items())

```

```
#return list(minscores.keys())[0]
return minscores[0][0]

e=2
index=1
for train, test in kfold.split(X_new, Y_new):
    k_range.append(index)
    labels = np.unique(Y_new)
    points = []
    for idx, val in enumerate(X_new[train]):
        points.append(DataPoint(val, Y_new[train][idx]))
    pred = []
    prediksi = []
    for idx, val in enumerate(X_new[test]):
        pred.append(DataPoint(val))
    #print points[0].features
    for idx,p in enumerate(pred):
        p.label= nwknn(n, e , p , points)
        prediksi.append(p.label)
        #print(p.label)
    testing=[]
    for test in Y_new[test]:
        testing.append(test)
    print("K-Fold - ",index)
    print("K-Fold - ",index)
    print(classification_report(testing,prediksi, labels=labels))
    print("Akurasi = ", accuracy_score(testing, prediksi))
    print("confusion_matrix = ", confusion_matrix(testing, prediksi))
    index=index+1
```

**Lampiran 5. Confusion Matrix**

NWKNN			KNN		
<i>Confusion Matrix fold-1</i>			<i>Confusion Matrix fold-1</i>		
	Positif	Negatif		Positif	Negatif
Positif	242	21	Positif	247	16
Negatif	22	34	Negatif	21	36
Total	264	55	Total	268	52
<i>Confusion Matrix fold-2</i>			<i>Confusion Matrix fold-2</i>		
	Positif	Negatif		Positif	Negatif
Positif	238	20	Positif	238	20
Negatif	21	40	Negatif	21	40
Total	259	60	Total	259	60
<i>Confusion Matrix fold-3</i>			<i>Confusion Matrix fold-3</i>		
	Positif	Negatif		Positif	Negatif
Positif	240	16	Positif	240	16
Negatif	27	36	Negatif	27	36
Total	267	52	Total	267	52
<i>Confusion Matrix fold-4</i>			<i>Confusion Matrix fold-4</i>		
	Positif	Negatif		Positif	Negatif
Positif	242	25	Positif	242	25
Negatif	19	33	Negatif	19	33
Total	261	58	Total	261	58



<i>Confusion Matrix fold-5</i>			<i>Confusion Matrix fold-5</i>		
	Positif	Negatif		Positif	Negatif
Positif	232	28	Positif	232	28
Negatif	19	40	Negatif	19	40
Total	251	68	Total	251	68
<i>Confusion Matrix fold-6</i>			<i>Confusion Matrix fold-6</i>		
	Positif	Negatif		Positif	Negatif
Positif	226	22	Positif	226	22
Negatif	22	49	Negatif	22	49
Total	248	71	Total	248	71
<i>Confusion Matrix fold-7</i>			<i>Confusion Matrix fold-7</i>		
	Positif	Negatif		Positif	Negatif
Positif	228	22	Positif	228	22
Negatif	21	47	Negatif	21	47
Total	249	69	Total	249	69
<i>Confusion Matrix fold-8</i>			<i>Confusion Matrix fold-8</i>		
	Positif	Negatif		Positif	Negatif
Positif	231	20	Positif	233	18
Negatif	17	50	Negatif	17	50
Total	248	70	Total	250	68
<i>Confusion Matrix fold-9</i>			<i>Confusion Matrix fold-9</i>		
	Positif	Negatif		Positif	Negatif
Positif	243	29	Positif	243	29
Negatif	14	32	Negatif	14	32
Total	257	61	Total	257	61

<i>Confusion Matrix fold-10</i>			<i>Confusion Matrix fold-10</i>		
	Positif	Negatif		Positif	Negatif
Positif	230	32	Positif	230	32
Negatif	17	39	Negatif	17	39
Total	247	71	Total	247	71

### **Lampiran 6.** Daftar Kata *Replace*

"stenlis": "stainless",  
 "stenl": "stainless",  
 "stenles": "stainless",  
 "stensil": "stainless",  
 "pelastik": "plastik",  
 "plastique": "plastik",  
 "gosah": "tidak usah",  
 "tebel": "tebal",  
 "jgn": "jangan",  
 "team": "tim",  
 "ilang": "hilang",  
 "aq": "saya",  
 "gw": "saya",  
 "guw": "saya",  
 "esjewe": "sjw",  
 "berati": "berarti",  
 "jiji": "jijik",  
 "wesi": "besi",  
 "strau": "straw",  
 "denger": "dengar",  
 " ga ": " tidak ",  
 " ngga ": " tidak ",  
 " gk ": " tidak ",  
 " bgs ": " bagus ",  
 " lbh ": " lebih ",

"kalo":"kalau",  
 "goog":"bagus",  
 "kmrn":"kemarin",  
 "pake":"pakai",  
 " jg ":" juga ",  
 "tetep":"tetap",  
 " kalo ":" kalau ",  
 " ampe ":" sampai",  
 " aer ":"air",  
 " lgsg ":" langsung",  
 " make ":" pakai ",  
 " mekdi ":" mcd ",  
 "plastic":"plastik",  
 "sedotannya":"sedotan",  
 " make ":" pakai "

### Lampiran 7. Daftar Kata *Positive Lexicon Based*

acungan jempol	keinginan	pemberian Tuhan	iba
adaptif	keintiman	pembersih	ideal
adil	kejayaan	pembetulan	idealnya
afinitas	kejelasan	pemecah masalah	idola
afirmasi	kejujuran	pemenang	ikhwan
agilely	kekaguman	pemenuhan	ilahi
agung	kekal	pemulihan	mengilhami
ahli	kekalahan	penampilan terbaik	menginginkan
ahlinya	kekasih	pencerahan	mengisyaratkan
ajaib	kekencangan	pencinta	menguasai
aklamasi	kelancaran	pendewaan	mengucapkan
	berbicara		selamat
	kelangsungan		
akomodatif	hidup	pendukung	mengundang
akurat	kelas atas	penebusan	mengguguli
alam mimpi	kelas satu	penegasan kembali	menguntungkan

alhamdulillah	kelas utama	penentuan nasib sendiri	menikmati
allahu akbar	kelayakan	penerima	menimbulkan perasaan cinta
altruistis	kelezatan	pengagum	menimbulkan rasa antusias
aman	kelimpahan	pengampunan	meninggikan
amanah	kelincahan	pengangkatan	meningkat
amat	kelip redup	pengaruh	meningkatkan
ambisius	kelucuan	pengasih	menitipkan
andal	keluwesan	pengasuhan	menjamin
aneh	kemajuan	pengembalian dana	menjanjikan
anggun	kemakmuran	penggemar	menonjol
angin sepoi-sepoi	kemampuan	penghargaan	mentereng
angkat	kemantapan	Penghargaan	menunjukkan gejala
antusias	kemasyhuran	penghemat	menurut hak
antusiasme	kematangan	penghematan	menyalip
apik	kemauan baik	penghematan biaya	menyambut
apresiasi	kemegahan	penghiburan	menyanjung
asli	kemenangan	penghormatan	menyaring
aspirasi	kemewahan	penguasaan	menyederhanakan
asyik	kemudahan	peninggian derajat	menyegarkan
bagos	kemurahan hati	peningkatan	menyelaraskan
bagus	kenangan	penipu	menyelesaikan
bahagia	kenikmatan	penjualan terbaik	terbaik
baik	kenyamanan	penolong	terbaru
baik diposisikan	kepanjangan akal daya	penting	terbatas
baik sekali	kepentingan	penuh	terbebani
baik-baik	kepercayaan	penuh gaya	terberat
bakat	keperkasaan	penuh harapan	terbersih
bangga	kepincut	penuh kasih	terbesar
bantuan	kepuasan	penuh kebahagiaan	terbukti

banyak	kepuasan diri	penuh kebajikan	tercapai
banyak akal	keramahan	penuh kegembiraan	tercengang
barang baru	keramat	penuh kepercayaan	tercepat
batu permata	keranjingan	penuh pengharapan	tercinta
bebas	kerelaan	penuh perasaan	terdengar
bebas masalah	keren	penuh perhatian	terdepan
bebas pulsa	kerendahan hati	penuh peristiwa	terelakkan
bebas rasa sakit	keriangan	penuh semangat	tergantikan
bebas resiko	keriuhan di pawai	penuh sukacita	terhormat
bekerja	kerubin	penyayang	terima
bekerja keras	kesabaran	penyelamat	terima kasih
belas kasihan	kesalahan	penyuburan	terjamin
benar	kesatuan	perasaan kagum	terjangkau
benar-benar	kesayangan	perasaan suka cita	terkalahkan
bengal	kesehatan	perayaan	terkejar
beradaptasi	kesejahteraan	perbaikan	terkejut
beralasan	kesembronoan	perbedaan	terkekang
berani	kesempurnaan	perdamaian	terkemuka
berapi	kesenangan	perdana menteri	terkenal
berarti	kesetiaan	perhatian	terkenal di dunia
berbaik hati	kesopanan	periang	terkuat
berbakat	kestabilan	perjanjian	ilu
berbesar hati	kesucian	perlawanan	imajinatif
berbudi luhur	kesukaan	perlindungan	iman
bercacat	kesungguhan	permai	imut
bercahaya	ketabahan	permata	indah
bercanda	ketajaman	persahabatan	individual
bercita-cita	ketegasan	persetujuan	infalibilitas
berdaya cipta	keteguhan	persik	inovasi
berdebar	ketekunan	personalisasi	inovatif

berdedikasi	ketenangan	pertama di kelas	inspirasi
berdikari	ketenaran	pertumbuhan tercepat	inspirasional
berempati	keterbukaan	pesona	inspiratif
bergairah	ketercocokan	petualang	instrumental
bergaya	ketidakberpihakan	piala	integral
bergema	keampilan	pikiran-bertiup	intelijen
bergembira	ketulusan	pilihan	intim
bergembira sekali	keuletan	pintar	intrik
bergengsi	keunggulan	Plus	intuitif
bergizi	keuntungan	populer	irama
berguna	kewaspadaan	portabel	iri
berharga	khusus	positif	istimewa
berhasil	kilau	preferensi	isyarat
berhutang	klasik	prestasi	jaminan
berikut	koheren	pro	jaminan mutu
berimbang	koherensi	proaktif	janji
berisi	komitmen	prodigi	jawa
beristirahat	kompak	produktif	menyembuhkan
berjanji	kompetitif	progresif	menyenangkan
berjasa	komplementer	promotor	menyerahkan
berjaya	konsisten	puas	menyeringai
berjenis	konstruktif	puitis	menyetujui
berjuang	kontinuitas	pujian	menyolok
berkapasitas besar	kontras	pukulan yg tdk keras	menyubsidi
berkat	kontribusi	pulih	menyukai
berkeinginan	kooperatif	puncak	menyusul
berkelakuan baik	koperasi	rahmat	merangsang
berkelanjutan	korek api	rajin	merayakan
berkembang	kreatif	rak	merayu

berkeyakinan	Kualitas terbaik	ramah	mereda
berkilau	kualitas tinggi	ramah tamah	mereformasi merekomendasikan
berkilauan	kuasa	rangsang	
berkualitas berlangsung mudah	kuat kue panas	rapi rasa hormat	merelakan meremajakan
berlimpah berlimpah- limpah	kukuh kurus	rasa manis rasional	merendahkan merenung
bermaksud baik	lagu	rata	merevitalisasi
bermanfaat	lancar	realisasi	merevolusi merevolusionerkan
bermanuver	landasan	realistis	
bermartabat	lapang	reformasi	meriangkan
bernilai	laudably	rekomendasi	meringankan
berpendidikan berpendidikan baik	layak lebih baik lebih baik dari perkiraan	rela reputasi	meromantiskan mewah terlalu
berpengalaman	lebih cemerlang dari	reseptif	memanjakan
berpengaruh berpengetahuan luas	lebih cepat	resiko rendah responsif	termotivasi termudah
berpijar	lebih dikenal	restrukturisasi	termurah
bersatu bersedia melakukan	lebih disukai lebih hangat	revolusioner ria gembira	terobosan terompet
bersemangat bersemangat meluap-luap bersenang- senang	lebih kencang lebih keras Lebih memilih	riang ringan ringkas	terpadu terpanas terpelajar
bersenda gurau	lebih memilih	rohani	terpengaruh
berseri	lebih murah	rokstar	terpenting
bersifat dermawan	lebih suka	romantis	terpercaya
bersifat mendamaikan	lebih tenang	rujuk	terpesona

bersih	ledakan	rumit	terpuji
bersinar	legendaris	sabar	terrifically
bersorak	lemah lembut	sah	tersayang
bersubsidi	lembut	saleh	tersedia
bersuka cita	lezat	salut	tersenyum
bersuka ria	licin	sangat	tertarik
bersyukur	lincah	sangat baik	tertib
bertanggung jawab	liris	sangat berharga	tertinggi
bertekun	logis	sangat indah	terus terang
berterimakasih	loyalitas	sangat memuaskan	terutama
beruntung	luar biasa	sangat menarik	tetap
berwarna ria	luas	sangat menyayangi	tiada bandingan
berwawasan	lucu	sangat mudah	tiada duanya
besar	lumayan	sangat penting	jelas
besar sekali	luwes	sangat rajin	jempol
biaya rendah	maaf	sangat sopan	jempolan
bijak	mahal	sangat tepat	jenaka
bijaksana	mahir	sanjungan	jenius
bikin cemburu	main-main	sarankan	juara
bimbingan	maju	sayang	jujur
bintang rock	makanan	sayang anak	kagum
bonus	makmur	secara kuat	kanan
boros	malaikat	secara menyenangkan	karisma
brilian	mampu	secara terbuka	karismatik
cahaya	manfaat	secepatnya	karya
cahaya redup	manis	sederhana	kasih sayang
cakap	manis sekali	segar	kaya
cantik	manja	segera	kaya fitur
cashbacks	manjur	sehat	keadaan yg sebenarnya



cekatan	mantab	seimbang	keadilan
cemas	manusiawi	seksi	keagungan
cemerlang	mapan	Selamat	keajaiban
cepat	martabat	selamat	keanggunan
cerah	master	selayaknya	keangkeran
cerdas	masuk akal	selektif	meyakinkan
cerdik	matang	semakin lembut	minim resiko
cergas	mati-matian	semakin makmur	modern
cermat	mau kalah	semangat	modis
cetar	mawas	sempurna	montok
cetar membahana	megah	semuanya	monumental
cinta	mekar	senang	moralitas
cocok	melampaui	sengaja	muda
contoh	melebihi	sensasi	mudah
cukup	melengkapi	sensasional	mudah digunakan
cukup besar	melindungi	senyum	mujarab
cukuplah	meluruskan	sepanjang masa	mukjizat
damai	memadai	sepantasnya	mulia
dapat diandalkan	memanggil	sepatutnya	mulus
dapat dicapai	memantapkan	sepele	murah
dapat diganti	memastikan	sepenuh hati	murah banget
dapat dipercaya	membandingi keindahan puisi	seperti	murah hati
dapat diraih	membantu	seperti malaikat	murni
dapat disesuaikan	membara	seperti patung	mutakhir
daya tarik	membayar	seperti raja	nasib
dengan mewah	membebaskan	serangan balik	nikmat
dengan senang hati	memberanikan	serba cepat	tidak egois
dengan sopan	memberdayakan	serba guna	tidak ternilai
dermawan	memberkati	sesak napas	tipis

dewasa	membesarkan hati	sesuai sesuai dengan mode terakhir	tokoh TOP
diakses	membesut		
diakui	membiasakan membuat lebih baik dr	sesungguhnya	trendi
diam		setajam silet	tulang punggung
dianggap baik	membudayakan	setia	tulus
dibaca	membujuk	setuju	tumbuh cepat
dibebaskan	membuktikan	sewajarnya	tunggu dulu
diberikan	memelihara	sewenang-wenang	uang kembali
dibersihkan	memeluk	siap	ulung
dibuat dengan baik	memenuhi	sigap	unggul
dicapai	memenuhi syarat	sihir	unggulan
didominasi	memeriahkan	sikap tenang	unik
didukung	memfavoritkan	silakan	untung
diganggu gugat diganti oleh pengguna	memihak memikat	silau simpatisan	untungnya utuh
digunakan	memikat hati	sok aksi	uye
dihargai	memimpin	solidaritas	variasi
diinginkan	memperbaiki memperbaiki	sopan	visioner
diizinkan	akhlak	sorak kegirangan	keaslian
dikelola dikelola dengan baik	mempercantik mempercayakan	sosial spektakuler	kebahagiaan kebaikan
dikembalikan	memperjelas	spontan	kebajikan
dilayari	memperkaya	sporty	kebal
dilengkapi	memperkuat	stabil	kebanggaan
dilepas	mempermanis memperoleh	stabilitas	kebangkitan
dilihat	kembali	sub	kebaruan
dimenangkan	mempesona	subhanallah	kebebasan
dimengerti	memuaskan	subsidi	kebenaran

dinamis	memudahkan	substantif	keberanian
diperbaharui	memuja	subur	keberhasilan
dipercaya	memujanya	suci	keberlanjutan
diperoleh	memuji	sukacita	kebersamaan
diplomatik	memuji-muji memuji-muji	sukses	kebersihan
dipoles	terlebih-lebihan	sulit	keberuntungan
direformasi	memukau	sungguh	kebesaran
direkomendasikan	memuliakan	sungguh-sungguh	kebetulan
diremajakan	memulihkan	super	kebijaksanaan
direstrukturisasi	memurnikan	supremasi	kecakapan
disahkan	menakjubkan	surrealis	kecantikan
disayangi	menambah	surga	kecemerlangan
disederhanakan	menang	surgawi	kecepatan
disubsidi	menantang	swasembada	kecerdasan
ditambah lagi	menarik	swatantra	kecerdikan
ditegakkan diterima dengan baik	menariknya	syukur	keemasan
ditingkatkan	menawan	taat	keenakan
diverifikasi	mencengangkan	taat hukum	kegairahan
dorongan	mencintai	tabah	kegembiraan kegembiraan yg meluap-luap
duanya	mencukupi	tabungan	
Dukung	mendalam	tahan karat	kehangatan
dukungan	mendamaikan	tahan lama	keheranan
durian runtuh	mendapatkan	tajam	keindahan
edukatif	mendebarkan	tak ada salahnya	bambu
efektif	mendewakan	tak ada taranya	logam
efektivitas	mendominasi	tak berbahaya	peduli
efisien	mendorong	tak bercacat	non-kekerasan
egois	mendukung	tak bersuara	nyaman
	menebus	tak bersyarat	nyaring

ekonomis	menegakkan	tak dapat disangsikan	obat untuk segala penyakit
ekstase	menegaskan	tak diragukan lagi	obsesi
elastis	menegaskan lagi	tak disengaja	optimal
elegan	menenangkan	tak kenal takut	optimis
elite	menentukan	tak perlu dipersoalkan	optimisme
emas	menerangi	tak tahu malu	otoriter
empati	menerangkan	tak terbantahkan	padat
enak	menerima	tak terbatas	pahlawan
enchantingly	menerobos	tak tergantikan	pahlawan wanita
energi	mengagetkan	tak terhitung	paling beruntung
ergonomis	mengagumi	tak terkalahkan	paling dikenal
etis	mengagumkan	tak terkendalikan	panas
euforia	mengagungkan	tak terlupakan	pandai berbicara
evaluatif	mengakali	tak ternilai	panorama
exemplar	mengalahkan	tak terpadai	pantas
fajar	menganggap	tak tertahankan	pasien
fantastis	mengangkat	takdir	pasti pasukan penjaga perdamaian
fasih	menganjurkan	takjub	patriot
fav	mengasyikan	takzim	patriotik
fave	mengasyikkan	tampak	patuh
favorit	mengatasi	tampan	pedih
fenomenal	mengecoh	tanda	peka
firdaus	mengejutkan	tangan ke bawah	pelamun
fleksibel	mengerti	tangkas	pelindung
fleksibilitas	mengesahkan	tanpa cela	pelonggaran
futuristik	mengesankan	tanpa dosa	pembaruan
gagah	menggairahkan	tanpa malu-malu	pembebasan
gaib	menggelenyar	tanpa rasa sakit	pembela
gainfully	menggelikan	tanpa rasa takut	

gairah	menggelitik	tanpa tenaga	pemberdayaan
gamblang	menggembirakan	tawar	stainless steel
gembira	menggemparkan	tawaran	melarang
gembira luar biasa	menggetarkan	tdk berbahaya	ganti
gembira sekali	menggiurkan	tegas	wah
gemuk	menggoda	teguh	wahyu
gesit	menggugah	teladan	warna-warni
giat	menghadiahkan	teliti	waspada
gigih	menghaluskan	temanmu	whooa
gokil	menghargai	tenang	Whooooo
gratis	mengharukan	tepat	Wow
gurih	mengherankan	tepat pada waktunya	WOW
habis	menghibur	tepat waktu	yakin
hadiah	menghidupkan	tepek tangan sorak	cucup
hak istimewa	menghormat	terampil	langsung tenggak
halal	menghormati	terang	langsung
halus	mengidealkan	terang benderang	tenggak
handal	mengidolakan	teranyar	lingkungan
handier	harum	heran	nyocop
hangat	hasil karya	heroik	bahaya
harga diri	hasil terbaik	hore	sustainable
harga rendah	hati	horee	dimakan
harmonis	hati-hati	hormat	kokop
harmonis	hebat	hubungan	susup
harta	hemat	hujan es	seruput no straw movement
saveearth	daun pisang mengurangi	zero waste	
save	plastik	kurangi	jijik
stainless	mengurangi	laut	prakarya

**Lampiran 8. Daftar Kata *Negative Lexicon Based***

abnormal	keadaan acuh tak acuh	pencemas	dikorbankan
absurd	kebosanan	pencurian	diktator
acak	kebrutalan	pendapat	diktatoris
acak-acakan	kebuntuan	pendendam	dilapisi gula
acuh	keburukan	penderita	dilecehkan
acuh tak acuh	kecabulan	penderitaan	dilema
adiktif	kecaman	penderitaan mendalam	dilenyapkan
adil	kecanduan	pendosa	dimaafkan
agresi	kecapaian	penekanan	dimarahi
agresif	kecelakaan	pengabaian	dimengerti
agresor	kecelakaan kapal	pengacau	dingin
aib	kecemasan	pengaduan	diperangi
air terjun	kecemburuan	pengakuan	diperkosa diperlakukan dengan buruk
akurat	kecenderungan kecenderungan	pengamuk	
alarm	untuk menurun	penganggur	dipermainkan
alasan	kecerobohan	penganiaya	dipertanyakan
alat permainan	kecewa	penganiayaan	dipikirkan
alergi	kecil	pengap	dirampas
alergik	kecongkakan	pengasingan	diremehkan
amat ketakutan	kecurangan	pengeboman	dirugikan
amat panas	kecurigaan	pengecualian	disalahgunakan
ambigu	kedangkalan	pengecut	disalahpahami
ambivalen	kedengkian	pengemis	disayangkan
ambivalensi	keenggan	pengenaan	disebut-sebut
amoral	kefanatikan	pengempungan	disederhanakan
amoralitas	kegagalan	pengeringan	disengaja
ampun	keganasan	pengganggu	disengketakan
amuk	kegarangan	penggerutu	disiram

anak nakal	kegelapan	penghancuran	diskredit
anak yatim	kegelisahan	penghasut	diskriminasi
anarki	kegemparan	penghinaan	diskriminatif
anarkis	kegemukan	penghindaran	disorient
anarkisme	kegilaan	penghujatan	disproporsional
ancaman	kegoyangan	penghukuman	distorsi
aneh	kegugupan	peringatan	ditakdirkan
aneh lagi	kehabisan	peringkaran	ditinggalkan
anehnya	kehancuran	peringir jenazah	ditipu
angkuh	kehebohan	pengkhianat	ditolak
angriness	Keheranan	pengkhianatan	dogmatis
anjing	kehilangan	pengkritik	dominan
anjlok	kehilangan keimbangan	pengobrol	dongkol
anomali	kehinaan	pengotor	dosa
antagonis	keinginan pribadi	pengrusakan	downgrade
antagonisme	keingkaran	pengumpatan	drastis
antek	keirasionalan	pengunduran diri	drop-out
anti-	kejahatan	pengurangan	dua wajah
anti-Amerika	kejam	penindas	dugaan
anti-Israel	kejang	penindasan	duka
anti-kita	kejanggalan	penipu	dukun
anti-pendudukan	kejangkitan	penipuan	dumping
anti-proliferasi	kejatuhan	penjahat	dungu
anti-putih	kejelekan	penjara	duniawi
anti-Semit	kejengahan	penjara gelap bawah tanah	dupa
antipati	kejengkelan	penjarah	durhaka
antisosial	keji	penolakan	duri
antitesis	kejut	penuh	dusta
apak	kekacauan	penuh benci	dusun

apati	kekakuan	penuh celaan	dwimakna
apatis	kekalahannya	penuh curiga	edan
apek	kekanak-kanakan	penuh dosa	egois
apokaliptik	kekecewaan	penuh sesal	egoisme
apologis	kekecilan	penumpasan	egomania
argumentatif	kekejaman	penundaan	egosentris
artinya jika	kekejamannya	penurunan	ejek
asam	kekeliruan	penurut	ejekan
asap	kekenyangan	penyadap	ekonomis
asem	kekerasan	penyakit	eksploitasi
asing	kekerasan pendirian	penyakit gila	eksploitatif
astaghfirullah	kekeringan	penyalahgunaan	eksplosif
asusila	kekesalan	penyangkalan	ekstremis
awan	kekhawatiran	penyederhanaan berlebihan	enggan
awas	kekhilafan	penyelewengan	enggan membantu
babi	kekosongan	penyembunyian	erosi
badai	kekotoran	penyendiri	fana
bahan tertawaan	kekuatiran	penyerbu	fanatik
bahaya	kekurangan	penyergapan	fanatisme
bajingan	kekurangpekaan	penyesalan	fantastis
baju kotor	kelakuan buruk	penyesat	fasis
balas dendam	kelalaian	penyesatan	fasisme
bandel	kelambanan	penyiksaan	fatal
bandot	kelancangan	penyimpang	fiksi
bangkrut	kelangkaan	penyimpangan	fitnah
bantingan	kelaparan	penyisihan	fitnahan
banyak sekali	kelas dua	penyok	fobia
barang ganjil	kelelahan	penyuapan	friksi
barbar	kelemahan	penyusup	frustasi



basi	kelemahan karena usia tua	peot	frustrasi
bau	kelembutan	peradangan	fundamentalis
bawahan	kelesuan	perampas	fundamentalisme
bebal	keletihan	perampasan	ga jelas
beban	keliru	perangkap	ga karuan
bejat	keluar	perasaan geli	ga peka
bekas luka	keluhan	perasaan waswas	gadungan
bekas roda	Keluhan	perawan tua	gagal
beku	kelupaan	perbedaan pendapat	gagap
belah	kemacetan	perbuatan jahat	gaib
belum dewasa	kemalangan	perbudakan	galak
belum dicoba	kemalasan	perdarahan	ganas
belum dikonfirmasi	kemandekan	perebut perempuan	gangguan
belum pasti	kemarahan	jahanam	ganjil
belum selesai	kemasukan setan	perfidity	ganti rugi
benar-benar	kemasyhuran	pergolakan	gantung
bencana	kematian	perhatian	garang
bencana alam	kembali	perih sekali	garis keras
benci	kemelaratan	peringatan	garu
bengah	kemenduaan	perjuangan	gasang
bengis	kemerosotan	perkelahian	gatal
bengkak	kemiskinan	perlakuan kejam	gegabah
bengkeng	kemunafikan	perlawanan	gejala
bengkok	kemunduran	perlu	Gejala
benjolan	kemurkaan	permohonan	gelandangan
bentrokan	kemurungan	permusuhan	gelap
beracun	kemustahilan	perpecahan	gelisah
beradab	kendor	persahabatan	gelora
berakhir	kendur	persaingan	gembar-gembor

berang	kenekatan	perselisihan	gendut
berani	kental	pertempuran	genit
berantakan	kepahitan	pertengkaran pertengkaran	genosida
berat	kepala batu	sengit	genting
berat sebelah	kepala-sakit	pertentangan	gerah
berawan	kepatahan	pertikaian pertumpahan	geram
berbahaya	kepedaran	darah	gerombolan
berbatu-batu	kependekan	pertunjukkan	gesekan
berbau	kepentingan kepentingan diri	perubahan	gesper
berbeda	sendiri	perusak	getah
berbisa	kepicikan	pesimis	getaran
berbohong	keputusan	pesimisme	giat
berbuat curang	keputusan	pesimistis	gigih
berbuat jahat	keracunan	pesta	gigil
berbuat salah	keraguan	pesta besar	gila
berbuih	kerangka	pesuruh	gila hormat
bercacat	keras	petualangan	gila ketakutan
berdalih	keras hati	PHK	gila-gilaan
berdarah	keras kepala	picik	godaan
berdasar	keras-kapal	picketed	goreng
berdaya	kerasnya	pidana	goresan
berdebar	kerdil	piket	gosip
berdebat	kere	pikun	goyah
berdebu	kerepotan	pilih-pilih	goyangan
berdengung	kereta	pimpinan salah	gua
berdenyut	keretakan	pincang	gugup
berdenyut-denyut	keriangan	pingsan	gurun
berderak	keributan	pisah	gusar
berderit	kerinduan	pisau	habis

berdetak	kering	piuh	hak milik
berdokumen	keriput	plin-plan	hal merendahkan diri
berdosa	keriuhan	polusi	hal tidak dimengerti
berduka	kerlip	prasangka	halangan
berduri	keroncongan	pratfall	hama
berdusta	kerugian	prihatin	hambatan
berebut	Kerugian	primitif	hampir
bergairah	kerusakan	propaganda	hampir mati
bergegas	kerusuhan	protes	hancur
bergelombang	kerut	provokasi	hang
bergeming	keruwetan	provokatif	haram
bergerak	kesal	puas	harga di atas
bergerak lambat	kesalahan	pucat	hasutan
bergerigi	kesalahan besar	pukulan	haus
bergetar	kesalahan hitung	punah	haus darah
bergolak	kesalahpahaman	punk	hedon
bergulat	kesamaan	pura-pura	hedonistik
berhaluan kiri	kesedihan	pusing	hegemoni
berisiko	keseganan	putus	hegemonisme
berita palsu	kesegeraan	putus asa	hemat
berjangkit	kesembronoan	racun	hidung belang
berjuang	kesendirian	radikal	hina
berkarat	kesengitan	Radikal	hingar-bingar
berkata tanpa berpikir	kesengsaraan	radikalisasi	hiruk-pikuk
berkedip	kesepian	ragu	histeri
berkejut	keserakahan	ragu-ragu	histeria
berkenan	keseriusan	rajin	histeris
berkeping-keping	kesesakan	raksasa	hujat
berkeras pendirian	kesilauan	ramai	hukuman

berkeringat	kesombongan	rambut rontok	hukuman penjara
berkerut	kesulitan	rampasan	hukuman setimpal
berkhayal	kesuraman	rantingly	hutang
berkilat	kesyahidan	rants	iblis
berkilau	ketabahan	rapuh	idiot
berkolusi	ketakberanian	rasa bersalah	igauan
berkonflik	ketakutan	rasa cemas	ih
berkubang	ketamakan	rasa gelisah	ikut campur
berkurang	ketat	rasa sakit	ilegal
berlangganan	ketegangan	rasa suram	Iluminati
berlebihan	ketenangan	rasa tidak berterimakasih	ilusi
berlemak	keterbatasan	rasa tidak enak	imajiner
berlengah-lengah	keterbelakangan	rasis	imperialis
berlepotan	keterlaluan	rasisme	impoten
berliku-liku	ketiadaan	reaksioner	impulsif
berlumpur	ketiadaan rasa	realistis	individualis
bermain berlebih-lebihan	hormat	rebusan	indoktrinasi
bermasalah	ketidakabsahan	redundansi	infeksi
berminyak	ketidakadilan	redup	inferioritas
bermoral	ketidakakuratan	relevan	inflamasi
bermuka dua	ketidakamanan	remang	inflasi
bermuram	ketidakbahagiaan	remang-remang	ingusan
bermuram durja	ketidakberdayaan	remeh	inkompeten
bermusuhan	ketidakbijaksanaan	remehkan	inkompetensi
bermutu rendah	ketidakcakapan	rendah	inkonsistensi
bernasib buruk	ketidakcocokan	rendah dinilai	inkonstitusional
bernoda kotor	ketidakcukupan	rendahan	insensitively
berongga	ketidakefektifan	rentan	interupsi
berperang	ketidakefisienan	represi	intimidasi
	ketidakjelasan		

berpura-pura	ketidakjujuran	represif	intrusi
berputus asa	ketidakkasihannya	reruntuhan	invasif
bersaing	ketidaklogisan	resah	irasional
bersakit	ketidakmampuan ketidakmampuan	resesi	irasionalitas
bersanding	menyesuaikan diri	retak	iri
bersehid	ketidakmungkinan	retorik	Iritasi
bersekongkol	ketidakmurnian	retoris	ironi
berselang	ketidakpantasan	rewel	ironis
berselisih	ketidakpedulian	reyot	ironisnya
bersemangat	ketidakpercayaan	ribet	isolasi
bersenandung	ketidakpuasan	ribut	istirahat
berserakan	ketidakrelevanan	rintik	isu
bersetubuh	ketidaksabaran	risiko	jadah
bersisik	ketidaksamaan	riuh	jahanam
bertele-tele	ketidakseimbangan	roboh	jahat
bertemu	ketidaksempurnaan	rongsokan	jalan buntu
bertengkar	ketidaksenangan	ruam	jalan keluar
bertentangan	ketidaksenonohan	rumah kaca	jancuk
berteriak	ketidaksetaraan	rumit	janggal
bertindak tidak pantas	ketidaksetiaan	rumor	jatuh
bertingkah	ketidaksopanan	runtuh	jatuh sakit
bertubuh kecil	ketidakstabilan	rusak	jauh
berukuran terlalu kecil	ketidaktahuan	rusuh	jebakan
berumur pendek	Ketidaktelitian	sabar	jelaga
berwajah dua	ketidaktelitian	sabotase	jelas
biadab	ketidakteraturan	sakit	jelatang
bias	ketidaktertarikan	sakit dirancang	jelek
biasa	ketidaktoleranan	sakit hati	jelu
biasa-biasa saja	ketidaktulusan	sakit kepala	jempol kebawah

bid'ah	ketinggalan	sakit punggung	jenaka
bikinan	ketukan	sakit saraf	jengkel
bimbang	ketus	sakit-dibentuk sakit-	jerat
binasa	kewajiban	didefinisikan	jerawat
bingung	kewalahan	sakit-digunakan	jeritan
Bingung	kezaliman	sakit-dipahami	jeruk nipis
bisa ular	khawatir	sakit-penggunaan	jompo
bising	khayalan	sakitan	jumlah sedikit
blunder	khayalan belaka	sakitnya tuh disini	kabur
bobrok	khayali	salah	kabut
bocor	khianat	Salah	kacau
bodoh	khusus	salah baca	kadaluarasa
bohong	kiamat	salah hitung	kadung
bom	kikir	salah informasi	kafir
boneka	kikuk	salah membaca salah	kagum
boros	kios	menafsirkan	kain kafan
bosan	kisi	salah mengerti salah	kaku
botak	klik	mengucapkan	kalah
brutal	klise	salah mengurus	kalahan
bual	knalpot	salah menilai salah	kambing hitam
buar	kolot	menyebutkan	kambuh
buas	kompatibel	salah paham	kampungana
budak	kompleks	salah pikiran	kandang
bug	komplikasi	salah saji	kandas
bukan kepalang	komplotan	salju longsor	kanibal
buntu	kompong	saluran air	kanker
bunuh diri	konflik	sama sekali	kantong sampah
buram	konfrontasi	samar	kapak
buritan	kongkalikong	samaran	kapalan

buronan	konservatif	sambilan	karat
buruk	konsesi	sandal bakiak	kasar
buruk sekali	konspirasi	sandera	kasihan
busuk	konspiratif	sangat	kata-kata kasar
buta	konspirator	sangat efektif	katastropi
buta huruf	kontaminasi	sangat jahat	limbah
cabul	kontensius	sangat lapar	limbung
cacat	kontol	sangat marah	linglung
cacian	kontra	sangat menakjubkan	lintah
calo	kontra-produktif	sangat sedikit	lirik
cambuk	kontradiksi	sanggahan	longgar
canggung	kontradiktif	sarkasme	luah
cara	kontraproduktif	sarkastik	luar biasa
cari perkara	kontroversi	saru	lubang angin
carut-marut	kontroversial	satir	lucu
cebol	konyol	satiris	luka
cedera	korban	sayangnya sebelum	luka bakar
cekcok	korban kecelakaan	waktunya	luka lecet
cekung	korosi	sedang saja	lumpuh
celaan	korosif	sederhana	lumpur
celah	korup	sedih	lunak
celaka	korupsi	sedikit	lupa
cemas	kotor	segan	lusuh
cemberut	kotoran	segera	Maaf
cemburu	kram	sejajar	mabuk
cemooh	krisis	selai	mabuk kepayang
cemoohan	kritik	selang	mafia
cenderung	kritikus	selendang	mahal
cengeng	kritis	selingan	mahal

cengking	kronis	selokan	main mata
cercaan	kuat	sembarangan	main perempuan mainan
cerdik	kuatir	sembrono	kerincingan
cerewet	kucam	semburan	makian
ceroboh	kucing gemuk	sementara	makin
compang-camping	kudung	sempit	malang
corengan	kukuh	semrawut	malapetaka
cukup	kuning	sendirian	malas
curam	kuno	sengit	malu
curang	kupas	sensasi	malu-malu
curiga	kurang	sensasional	mampat
dancok	kurang ajar	sensasionil	mandek
dangkal	kurang baik	sentakan	manipulasi
deadline	kurang berkembang	sepakan	manipulatif
debu	kurang beruntung	sepele	manipulator
defensif	kurang dikenal	serakah	manja
degenerasi	kurang lengkap	seram	marah
degradasi	kurang menarik	serampangan	marah-marah
dehumanisasi	kurang pengalaman	serangan	marginal
delusi	kurang sehat	Serangan	masalah
demam	kurang sopan	serangan gencar	masam
demoralisasi	kurus	seret	mati
dendam	kusut	serius	mati kelaparan
dengan mencemoohkan	kutukan	serong	mati lemas
dengan mendapat malapetaka	labil	sesat	mati rasa
dengan mengagetkan	labu	setan	mati-matian
dengan menghina dengan	lacur	setengah hati	mau bertobat
menyedihkan	lalai	setia	megah



dengan menyesal	lalim	sewenang-wenang	megap-megap
dengan panas	lama	shamefulness	melalaikan
dengan penuh ketakutan	lamban	shamelessness	melambat
dengan rasa curiga	lambat	sia-sia	melampiaskan
dengan rasa hina	lancang	sial	melancarkan
dengan remeh-temeh	lancar	sialan	melanggar
dengan sedih	lapis kedua	sikap	melanggar hukum
dengan segan	lapuk	siksaan	melanggar susila
dengan sengit	larangan	sindiran	melarang
dengan sia-sia	larut	sindroma	melarikan diri
dengan suara keras	latency	singkat	melawan hukum
dengan sukar	lawan	sinis	melayu
dengan terbahak-bahak	lebam	sinisme	melebih-lebihkan
dengan tidak senang	lebih buruk	sinting	melecehkan
dengki	lebih keras	siput	meledak
depresi	lebih mahal	sirep	melelahkan
derita	lebihhan	sisa-sisa	melemahkan
desis	lecet	sisi bawah	melenceng
destruktif	ledakan	sisi buta	melengking
diam	lekat	siuman	melengkingkan
diam-diam	lekir	skandal	melengkung
dibakar	lekur	skeptis	melenyapkan
dibanjiri	lelah	sobek	melepaskan
dibantai	lelucon	sok	melepuh
dibenarkan	lemah	sok nasehati	meleset
dibenci	lemak	sok ngatur	meletakkan-off
dibesar-besarkan	lemak-kucing	sombong	melibatkan
dibuang	lemas	sporadis	melimpahi
dibuat-buat	lembab	stagnasi	melinglungkan

dicela	lembek	statis	melodramatis
dicerca	lemot	steal	melongo
dicuci	lendir	stereotip	melongsorkan
dicuri	lengkap	stres	melucuti senjata
didanai	lengket	suam	melukai
diganggu	lepas	suar	meluluhkan
dihaluskan	lereng	subversif	melumpuhkan
dihentikan	lesu	suka memandang rendah	melunakkan
dihibur	letih	sulit	meluruhkan
dihukum	letusan	sumbang	memabukkan
dijauhi	liar	sundal	memadamkan
dijelaskan	licik	suram	memakan waktu
dikaburkan	licin	susah	memaksa
dikenakan	lier	syok	memaksakan
memalsukan	menabrak	tak sehat	terkenal
memalukan	menahan	tak sepakat	terkenal jahat
memamerkan	menakuti	tak setuju	terkesima
memanaskan	menakutkan	tak tahu malu	terkulai
memancing	menampar	tak tentu	terkutuk
memandang dengan marah	menanamkan	tak terbayangkan	terlalu
memandang rendah	menangis	tak terbedakan	terlalu banyak
memanjakan	menarik diri	tak terdamaikan	terlalu berani
memar	menarik kembali	tak terduga	terlalu bersemangat
memarahi	menarik perhatian	tak terelakkan	terlalu besar
memarahkan	menaruh simpati	tak terhindarkan	terlalu mahal
memarut	menasihati	tak terhitung	terlalu
mematahkan semangat	mencabut	tak terkatakan	menekankan
	mencabut perlindungan		terlalu tinggi
mematikan	hukum	tak terkendalikan	terlambat

membagi	mencaci	tak terpecahkan	terlantar
membahas	mencairkan	tak terpenuhi	terlarang
membahayakan	mencampuradukkan	tak tertahankan	terlepas
membakar	mencampuri	takhyul	terlibat
membalas	mencap	takut	termiskin
membalikkan	mencekik	tamak	ternoda
membanjiri	mencela	tamparan	teror
membantah	mencemari	tandus	terorirs
Membantah	mencemaskan	tangki	terorisme
membara	mencemoohkan	tanpa	terpencil
membatalkan	mencibir	tanpa alasan	terpengaruh
membatasi	mencicit	tanpa ampun tanpa belas kasihan	terperangkap
membatu	mencium	tanpa berpikir	terperanjat
membayar lebih	menciut-ciut	tanpa berpikir	terputus-putus
membedakan	mencle-mencle	tanpa cinta	tersandung
membeku	mencolok	tanpa dasar	tersangka
membekukan	mencopoti	tanpa diketahui	tersangkut
membela	mencubit	tanpa hasil	tersedak
membelit	mencuci otak	tanpa henti	terselesaikan
membenci	mencuri	tanpa hukum	terselubung
membengisi	mencurigakan	tanpa iman	tersembunyi tersembunyi dan membahayakan
memberatkan	mendadak	tanpa kompromi	tersendat
memberhentikan	mendakwa	tanpa malu tanpa pandang bulu	tersentak-sentak
memberontak	mendatang	tanpa rebewes	tersiksa
membinasakan	mendatangkan	tanpa sadar	tersinggung
membingungkan	mendeklamasikan	tanpa semangat	tersumbat
membohong	menderita khayalan	tanpa tujuan	tertagihnya
memboikot memboikot dr masyarakat	menderita sekali mendesak	tantangan	tertahankan

membolos	mendesis	tawanan	tertanam di hati
membombardir	mendidih	tawar-menawar	tertancap
membongkar	mendistorsi	tebal	tertandingi
membosankan	pembukaan lagu	tegang	tertarik
membuang	pembunuh	tegar	tertekan
membuang waktu	pembunuhan	teguran	tertempa
membuat	pembunuhan besar-	tekanan	tertinggal
membuat bersedih	besaran		
hati	pemecahan	telanjang	tertipu
membuat bingung	pemegatan	teliti	terurai
membuat kasar	pemerasan	tenggelam	tiba-tiba
membuat malu	pemeriksaan	tenggelamnya	tipis
membubuhi gula	pemfitnah	tengkar	tipu
membujuk	pemfitnahan	teralihkan	tipu daya
membunuh	pemilih	teramat	tipu muslihat
memburuk	pemisahan	teramati	tirani
memburukkan	pemotongan	terampil	tolol
memburuknya	pemukulan	terang-terangan	topangan
membusuk	pemusnahan	terangsang	totaliter
mebutakan	pemutarbalikan	terasa tidak enak	tragedi
memecat	penaklukan	terasing	tragis
memekakkan			
telinga	penakut	terbalik	Tunggul
memenjarakan	penalti	terbandingkan	tusukan
mementingkan diri			
sendiri	penangkapan	terbatas	uap
memeras	pencegah	terbawah	ugal-ugalan
memerciki	pencemar	terbayangkan	ulang
memerintahakan	tabir asap	terbelakang	ular berbisa
memfitnah	tabu	terbodoh	ultimatum
memihak	tahayul	terbuang	ultra-garis keras
memikat	tahi	terbujuk	umpan

memisahkan	tai	terbukti	umpatan
memiskinkan	tajam	terbunuh	untuk sementara
memohon	taji	terburu nafsu	usang
memperberat	tak ada artinya	terburuk	usil
memperbudak	tak bahagia	tercela	utuh
memperdaya	tak baik	tercemar	vagina
memperdayakan	tak berarti	terdefinisi	virus
memperingatkan	tak berdaya	terdelusi	volatil
memperkosa	tak bergairah	terduga	vulgar
memperlambat	tak berguna	terencana	wabah
memperlemah	tak beriman	terfragmentasi	waria
mempersiapkan	tak bermutu	tergagap	was-was
mempesona	tak bernyawa	terganggu	waspada
mempolemikkan	tak berperasaan	tergenang	wastafel
memprihatinkan	tak berperikemanusiaan	tergesa-gesa	ya silahkan saja
mempropagandakan	tak berpindah-pindah	terhalang	Yahudi
memprotes	tak bersedia	terhambat	percuma
memprovokasi	tak biasa	terhina	sensitive
memuakkan	tak dihargai	terhukum	males
memuaskan diri	tak henti-hentinya	teriakan	nyuci
memudarnya	tak hormat	terik	hilang
memukul	tak jelas	terinfeksi	ketinggalan
memukul dengan tongkat	tak kelihatan	terinfestasi	gigit
memukul mundur	tak kenal ampun	terjal	ngga enak
memuntahkan	tak kompeten	terjangan	ngilu
memurahkan	tak layak	terjangkau	sensitif
memusnahkan	tak menentu	terjawab	gigi
memutar	tak pernah puas	terjelek	susah
memutarbalikkan	tak populer	terkejut	pelatuk

pointless	gontok	keras	linu
sensitif	tebal	tv	cokot
karbon	aneh	buruk	ribet
antena	nyetrum	bibir	penyok

## Lampiran 9. Surat Keterangan Pengambilan Data

97

### Lampiran 9. Surat Keterangan Pengambilan Data

#### SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen Statistika FMKSD ITS,

Nama : Shindy Sari Utami

NRP : 062115 4000 0005

menyatakan bahwa data yang digunakan dalam Tugas Akhir ini merupakan data sekunder yang diambil dari penelitian/ buku/Tugas Akhir/Thesis/Publikasi/lainnya yaitu :

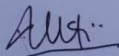
Sumber : Data dari *crawling* twitter

Surat pernyataan ini dibuat dengan sebenarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

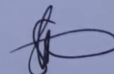
Surabaya, Juli 2019

Mengetahui,  
Pembimbing Tugas Akhir

Mahasiswa



Pratnya Paramitha Oktaviana, S.Si., M.Si  
NIP. 1300201405001



Shindy Sari Utami  
NRP. 062115 4000 0005





## BIODATA PENULIS



Penulis dengan nama lengkap Shindy Sari Utami dilahirkan di Madiun pada 11 Februari 1997. Penulis menempuh pendidikan formal di SDN 02 Madiun Lor, SMPN 2 Madiun, dan SMAN 2 Madiun. Penulis diterima sebagai Mahasiswa Departemen Statistika ITS melalui jalur SNMPTN pada tahun 2015. Pada tahun kedua perkuliahan, penulis mulai mengikuti organisasi dan kepanitiaan di dalam

ITS sebagai staff Tim Penelitian dan Pengembangan (LITBANG) HIMASTA-ITS dan staff di Divisi Logistik UKM KSR PMI ITS. Di tahun ketiga, penulis diamanahi menjadi Ketua Tim LITBANG dan Bendahara I di UKM KSR PMI ITS. Penulis juga menjadi Finalis *Indonesia Research Competition* di ISCO 2018 Semarang. Penulis juga telah mengikuti beberapa kegiatan survey sebagai pengaplikasian ilmu statistika. Segala kritik dan saran serta diskusi lebih lanjut mengenai Tugas Akhir ini, dapat menghubungi penulis melalui email [shindyutami4@gmail.com](mailto:shindyutami4@gmail.com).