

**IMPLEMENTASI NAÏVE BAYES CLASSIFIER DAN ASOSIASI UNTUK  
ANALISIS SENTIMEN DATA ULASAN APLIKASI *E-COMMERCE*  
SHOPEE PADA SITUS GOOGLE PLAY**

**TUGAS AKHIR**

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana  
Program Studi Statistika



Disusun oleh:

Zia Ayu Nuansa Gumilang  
14611215

**PROGRAM STUDI STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2018**

**HALAMAN PERSETUJUAN PEMBIMBING  
TUGAS AKHIR**


Judul : Implementasi Naïve Bayes Classifier dan Asosiasi untuk Analisis Sentimen Data Ulasan Aplikasi *E-Commerce* Shopee pada Situs Google Play

Nama Mahasiswa : Zia Ayu Nuansa Gumilang

Nomor Mahasiswa : 14 611 215

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK  
DIUJIKAN**

Yogyakarta, 28 April 2018

Pembimbing  
الإمامة الإسلامية  


(Ayundyah Kesumawati, S.Si., M.Si.)

**HALAMAN PENGESAHAN**

**TUGAS AKHIR**

**IMPLEMENTASI NAÏVE BAYES CLASSIFIER DAN ASOSIASI UNTUK  
ANALISIS SENTIMEN DATA ULASAN APLIKASI E-COMMERCE  
SHOPEE PADA SITUS GOOGLE PLAY**

**Nama Mahasiswa : Zia Ayu Nuansa Gumilang**

**Nomor Mahasiswa : 14 611 215**

**TUGAS AKHIR INI TELAH DIUJIKAN  
PADA TANGGAL 11 MEI 2018**

**Nama Penguji**

**Tanda tangan**

1. Andrie Pasca Hendradewa, S. T., M.T
2. Tuti Purwaningsih, S.Stat., M.Si.
3. Ayundyah Kesumawati, S.Si., M.Si.

.....  
.....  
.....

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



**Drs. Allwar, M.Sc., Ph.D**

## KATA PENGANTAR



*Assalamu'alaikum Wr. Wb.*

Puji dan syukur penulis haturkan kepada Allah SWT Tuhan Yang Maha Esa karena atas berkat dan karunia-Nya peneliti dapat menyelesaikan Laporan Tugas Akhir / Skripsi yang berjudul “Implementasi Naïve Bayes Classifier dan Asosiasi untuk Analisis Sentimen Data Ulasan Aplikasi *E-Commerce* Shopee pada Situs Google Play” ini sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika (S.Stat) di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.

Tugas Akhir (TA) / Skripsi dapat terselesaikan hingga tersusunnya laporan ini, tidak lepas dari bimbingan dan pengarahan berbagai pihak. Untuk itu, dengan segala kerendahan hati, pada kesempatan ini penulis mengucapkan terima kasih kepada,

1. Bapak Drs. Allwar, M.Sc., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
2. Bapak Dr. RB. Fajriya Hakim, S.Si, M. Si, selaku Ketua Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
3. Bapak Edy Widodo, S.Si., M.Si, selaku Dosen Pembimbing Akademik, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.



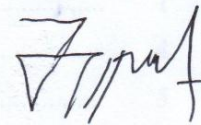
4. Ibu Ayundyah Kesumawati, S.Si, M.Si, selaku dosen Pembimbing yang selalu membantu, membimbing dan memotivasi selama penulisan Tugas Akhir ini.
5. Seluruh Dosen, dan Staff Administrasi yang telah banyak membantu dan memberikan kemudahan, terima kasih atas semua ilmu dan bimbingannya.
6. Kedua orang tua tercinta dan tersayang, Bapak Zulkipli dan Ibu Aina Astati yang selalu memberikan semangat, dukungan serta doa disetiap langkah peneliti.
7. Kakak dan Adik penulis, Zia Betros Malikul Amjad dan Zia Sri Ghulam Agong yang selalu memberi dukungan serta semangat, serta untuk Keluarga Besar yang selalu setia mendoakan yang terbaik untuk peneliti.
8. Sahabat-Sahabat Pokemon seperjuangan Sari, Edwika, Maulida, Reny, Zarina, Rabi, Nisa, Ditia, Dhea dan Khusnul yang senantiasa memberi semangat, doa, dorongan dan nasihat.
9. Teman-teman satu bimbingan tugas akhir (bimbingan ibu Ayun) Reny, Erdwika, Suci, Sari, Maulida, Molyda, Riza, Achmad, Afifah, Syauqi, Gustiara, Herlin, Afifah, Elsa, Dian yang selalu berbagi pengalaman, cerita dan ilmu.
10. Teman-teman Kontrakan Boki, Nur Jannah, Nisa dan Lia yang selalu berbagagi cerita dan pengalaman.
11. Teman-teman KKN Unit 116 Desa Cacaban Kdiul, kecamatan Bener, Purworejo ,Fitri, Yeni, Fira, Helda, Reza, Bagus, Didik dan Akbar

12. Serta pihak – pihak lainnya yang tidak dapat disebutkan yang membantu saya selama ini, saya mengucapkan terima kasih yang sebesar – besarnya.

Peneliti menyadari bahwa masih terdapat banyak kesalahan dalam penulisan dan penyusunan laporan ini, peneliti sangat mengharapkan kritik dan saran yang membangun berkaitan dengan perbaikan baik isi maupun teknis pada Laporan Tugas Akhir. Semoga laporan ini dapat memberikan manfaat yang kepada pembaca.

***Wassalamu'alaikum, wr.wb.***

Yogyakarta, 28 April 2018



Zia Ayu Nuansa Gumilang

BAB I	PENDAHULUAN	
1.1.	Latar Belakang Masalah	1
1.2.	Rumusan Masalah	2
1.3.	Batasan Masalah	3
1.4.	Tujuan Penelitian	4
1.5.	Manfaat Penelitian	5
1.6.	Sistematika Penulisan	6
BAB II	TINJAUAN PUSTAKA	8
BAB III	LANDASAN TEORI	12
3.1.	E-Commerce	12
3.2.	Shopee	13
3.3.	Analisa Sentimen	14
3.4.	Data Mining	15
3.5.	Preprocessing	15
3.5.1.	Correcting Wrong Word	16
3.5.2.	Filteasi	16
3.5.3.	Stemming	16
3.5.4.	Tokenisasi	17

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PERSETUJUAN PEMBIMBING .....</b>	<b>ii</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>iii</b>
<b>KATA PENGANTAR .....</b>	<b>iv</b>
<b>DAFTAR ISI .....</b>	<b>vii</b>
<b>DAFTAR TABEL .....</b>	<b>x</b>
<b>DAFTAR GAMBAR .....</b>	<b>xi</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xii</b>
<b>PERNYATAAN .....</b>	<b>xiii</b>
<b>INTISARI .....</b>	<b>xiv</b>
<b>ABSTRACT .....</b>	<b>xv</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1. Latar Belakang Masalah .....	1
1.2. Rumusan Masalah .....	4
1.3. Batasan Masalah .....	5
1.4. Tujuan Penelitian .....	5
1.5. Manfaat Penelitian .....	5
1.6. Sistematika Penulisan .....	6
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>8</b>
<b>BAB III LANDASAN TEORI .....</b>	<b>12</b>
3.1. <i>E-Commerce</i> .....	12
3.2. <i>Shopee</i> .....	13
3.3. Analisa Sentimen .....	14
3.4. Data Mining .....	15
3.5. <i>Preprocessing</i> .....	15
3.5.1. <i>Correcting Slang Word</i> .....	16
3.5.2. Filtrasi .....	16
3.5.3. <i>Stemming</i> .....	16
3.5.4. Tokenisasi .....	17

3.6.	Fitur dan Pembobotan .....	17
3.7.	Klasifikasi .....	19
3.8.	Teorema Bayes .....	20
3.9.	<i>Naïve Bayes Classifier</i> .....	21
3.10.	Asosiasi Teks .....	23
<b>BAB IV</b>	<b>METODOLOGI PENELITIAN</b> .....	24
4.1.	Populasi dan Sampel Penelitian .....	24
4.2.	Jenis dan Sumber Data .....	24
4.3.	Variabel Penelitian .....	24
4.4.	Metode Pengumpulan Data .....	25
4.5.	Metode Analisis Data .....	25
4.6.	Tahapan Penelitian .....	26
<b>BAB V</b>	<b>ANALISIS DAN PEMBAHASAN</b> .....	27
5.1.	Pengumpulan Data .....	27
5.2.	Analisis Deskriptif .....	29
5.3.	<i>Preprocessing</i> atau Prapemrosesan Data .....	30
5.3.1.	<i>Translate</i> .....	31
5.3.2.	<i>Spelling Normalization</i> .....	32
5.3.3.	<i>Case Folding</i> .....	32
5.3.4.	<i>Tokenizing</i> .....	32
5.3.5.	<i>Filtering</i> .....	33
5.4.	Pelabelan Kelas Sentimen .....	34
5.5.	Pembuatan Data Latih dan Data Uji .....	38
5.6.	Klasifikasi dengan Metode <i>Naïve Bayes Classifier</i> .....	39
5.7.	Visualisasi dan Asosiasi .....	41
5.7.1.	Ulasan Positif .....	42
5.7.2.	Ulasan Negatif .....	45
5.7.3.	Ulasan Bulanan .....	49
5.8.	Diagram <i>Fishbone</i> .....	51



<b>BAB VI PENUTUP .....</b>	<b>55</b>
6.1. Kesimpulan.....	55
6.2. Saran .....	56
<b>DAFTAR PUSTAKA .....</b>	<b>57</b>
<b>LAMPIRAN.....</b>	<b>61</b>

## DAFTAR TABEL

<b>Tabel 2.1</b>	Kelebihan <i>Naïve Bayes Classifier</i> pada penelitian-penelitian sebelumnya .....	10
<b>Tabel 5.1</b>	Perbandingan jumlah data pada kelas sentimen .....	35
<b>Tabel 5.2</b>	Hasil pelabelan menggunakan kamus <i>lexicon</i> dan proses manual.....	36
<b>Tabel 5.3</b>	Simulasi perhitungan skor sentimen .....	37
<b>Tabel 5.4</b>	Jumlah ulasan pada kelas sentimen .....	37
<b>Tabel 5.5</b>	Perbandingan data latih dan data uji .....	39
<b>Tabel 5.6</b>	Hasil <i>confusion matrix</i> .....	41
<b>Tabel 5.7</b>	Asosiasi kata pada kelas sentimen positif.....	44
<b>Tabel 5.8</b>	Asosiasi kata pada kelas sentimen negatif.....	47
<b>Tabel 5.8</b>	Rencana pemecahan masalah <i>e-commerce</i> Shopee .....	52

## DAFTAR GAMBAR

<b>Gambar 3.1</b>	Bagan Proses Klasifikasi (Han dan Kamber, 2016) .....	19
<b>Gambar 4.1</b>	Diagram alir penelitian .....	25
<b>Gambar 5.1</b>	Pemblokiran data ulasan <i>Shopee</i> di <i>google play</i> .....	27
<b>Gambar 5.2</b>	Proses <i>scraping</i> data.....	27
<b>Gambar 5.3</b>	Proses <i>Input</i> data di <i>Ms.Excel</i> .....	28
<b>Gambar 5.4</b>	Proses perubahan isi ulasan di <i>Ms.Excel</i> .....	29
<b>Gambar 5.5</b>	Grafik jumlah ulasan pengguna aplikasi <i>Shopee</i> Januari 2018 – Maret 2018.....	29
<b>Gambar 5.6</b>	Grafik <i>Rating</i> yang diberikan pengguna aplikasi <i>Shopee</i> Januari 2018 – Maret 2018.....	30
<b>Gambar 5.7</b>	Proses <i>Translate</i> .....	31
<b>Gambar 5.8</b>	Proses <i>spelling</i> .....	32
<b>Gambar 5.9</b>	Proses <i>case folding</i> .....	32
<b>Gambar 5.10</b>	Proses <i>tokenizing</i> .....	33
<b>Gambar 5.11</b>	Proses <i>filtering</i> .....	34
<b>Gambar 5.12</b>	Pembagian Kelas Sentimen .....	36
<b>Gambar 5.13</b>	Grafik jumlah kelas sentimen berdasarkan urutan waktu....	38
<b>Gambar 5.14</b>	Grafik nilai akurasi, presisi, dan <i>recall</i> dengan <i>Naïve</i> <i>Bayes Classifier</i> .....	40
<b>Gambar 5.15</b>	Kata yang paling banyak muncul pada keseluruhan data....	42
<b>Gambar 5.16</b>	Kata yang paling banyak muncul pada kelas positif .....	43
<b>Gambar 5.17</b>	<i>Wordcloud</i> ulasan positif .....	44
<b>Gambar 5.18</b>	Kata yang paling banyak muncul pada kelas negatif .....	46
<b>Gambar 5.19</b>	<i>Wordcloud</i> ulasan negatif .....	47
<b>Gambar 5.20</b>	Kata yang paling banyak muncul pada bulan Januari 2018	49
<b>Gambar 5.21</b>	Kata yang paling banyak muncul pada bulan Februari 2018 .....	50
<b>Gambar 5.22</b>	Kata yang paling banyak muncul pada bulan Maret 2018 ..	51
<b>Gambar 5.23</b>	Diagram <i>fishbone</i> komplain pengguna <i>Shopee</i> .....	52

## DAFTAR LAMPIRAN

<b>Lampiran 1</b>	<i>Script R preprocessing data dengan text mining .....</i>	61
<b>Lampiran 2</b>	<i>Script R pelabelan dan pembobotan .....</i>	63
<b>Lampiran 3</b>	<i>Script R klasifikasi dengan menggunakan metode Naïve Bayes Classifier.....</i>	64
<b>Lampiran 4</b>	<i>Script R visualisasi dan asosiasi teks (Untuk ulasan positif).....</i>	67
<b>Lampiran 5</b>	<i>Script R visualisasi dan asosiasi teks (Untuk ulasan negatif).....</i>	69
<b>Lampiran 6</b>	<i>Stopwords berbahasa Indonsia .....</i>	71
<b>Lampiran 7</b>	<i>Hasil iterasi pada proses klasifikasi dengan Naïve Bayes Classifier.....</i>	75

## IMPLEMENTASI NAIVE PERNYATAAN DAN ASOSIASI UNTUK ANALISIS SENTIMEN DATA UJIAN APLIKASI E-COMMERCE

Dengan ini penulis menyatakan bahwa dalam Tugas Akhir tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar sarjana di suatu perguruan tinggi dan sepanjang pengetahuan penulis juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

### INTISARI

Salah satu pengaruh dari peningkatan jumlah pengguna internet di Indonesia telah membuat meningkatnya kegiatan berbisnis melalui media internet. Hal ini membuat pengguna internet menginginkan segala hal yang bersifat instan dan tidak menyipitkan. Mengurangi akan hal ini akan dimanfaatkan oleh beberapa bisnis online untuk menawarkan produknya terutama dibidang barang dan jasa. Salah satu contohnya adalah e-commerce dapat diakses dengan menggunakan ponsel pintar. Salah satu contoh untuk aplikasi yang dapat diunduh dengan mudah berbasis android adalah aplikasi untuk menawarkan layanan e-commerce. Salah satu contoh aplikasi yang dapat diunduh di Indonesia ialah Shopee. Hal inilah yang membuat Shopee yang mendapatkan datanya dari situs web dan aplikasi diunduh oleh pelanggan dan diunduh menggunakan media sosial untuk mengklasifikasi alasan berdasarkan kategori kelas atau jenis produk yang negatif. Dari hasil penelitian kemudian dilihat melalui gambar menunjukkan sebuah informasi yang dianggap penting dan dapat menjadi dasar untuk pengambilan keputusan. Hasil klasifikasi sentiment menggunakan model Naive Bayes Classifier didapatkan tingkat akurasi sebesar 97,5%. Selanjutnya, pada proses analisis teks diperoleh informasi bahwa pengguna aplikasi Shopee Indonesia menggunakan aplikasi, ongkir, harga dan pajak untuk kelas positif dan untuk kelas negatif diperoleh informasi pengguna banyak menggunakan aplikasi ongkir, barang, merek, gambar, pengiriman, dan pajak. Hasil alasan negatif tersebut dibuat dalam diagram fishbone untuk pemecahan masalah.

Kata Kunci : E-Commerce, Google Play, Shopee, Analisis Sentimen, Naive Bayes Classifier, Analisis Teks, Diagram Fishbone

Yogyakarta, 28 April 2018



Zia Ayu Nuansa Gumilang



# IMPLEMENTASI NAÏVE BAYES CLASSIFIER DAN ASOSIASI UNTUK ANALISIS SENTIMEN DATA ULASAN APLIKASI E-COMMERCE SHOPEE PADA SITUS GOOGLE PLAY

Zia Ayu Nuansa Gumilang

Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Islam Indonesia

## INTISARI

Salah satu pengaruh dari peningkatan jumlah pengguna internet di Indonesia ialah semakin menjamurnya kegiatan berbelanja melalui media internet. Hal ini membuat pengguna internet menginginkan segala hal yang bersifat instan dan tidak merepotkan. Mengetahui akan hal itu maka dimanfaatkan oleh sebagian besar pembisnis untuk memasarkan produknya terutama dibidang barang dan jasa. Seiring perkembangannya, kini *e-commerce* dapat diakses dengan menggunakan perangkat *mobile phone* dalam bentuk aplikasi yang dapat diunduh dengan mudah berbagai macam aplikasi yang dapat menawarkan layanan *e-commerce*. Salah satu contoh aplikasi *e-commerce* yang memiliki reputasi baik di Indonesia ialah Shopee. Hal inilah yang mendorong peneliti untuk meneliti Shopee yang didapatkan datanya dari situs website google play. Selanjutnya dilakukan pelabelan dan dianalisis menggunakan metode *Naïve Bayes Classifier* untuk mengklasifikasikan ulasan berdasarkan kategori kelas sentimen positif dan kelas sentimen negatif. Dari hasil pelabelan kemudian dilihat asosiasi teks pada setiap kelas sentimen guna menemukan sebuah informasi yang dianggap penting dan dapat menjadi berguna untuk pengambilan keputusan. Hasil klasifikasi sentimen menggunakan model *Naïve Bayes Classifier* didapatkan tingkat akurasi sebesar 97,4%. Selanjutnya, pada proses asosiasi teks diperoleh informasi bahwa pengguna aplikasi *Shopee* Indonesia membicarakan aplikasi, ongkir, harga dan puas untuk kelas positif dan untuk kelas negatif diperoleh informasi pengguna banyak membicarakan aplikasi ongkir, buruk, susah, gambar, pengiriman, dan penjual. Hasil ulasan negatif tersebut dibuat dalam diagram fishbone untuk pemecahan masalah.

**Kata kunci :** *E-Commerce*, Google Play, Shopee, Analisis Sentimen, *Naïve Bayes Classifier*, Asosiasi Teks, Diagram *Fishbone*

# IMPLEMENTATION OF NAÏVE BAYES CLASSIFIER METHOD AND ASSOCIATION FOR SENTIMENT ANALYSIS OF DATA REVIEW APLIKASI E-COMMERCE SHOPEE ON GOOGLE PLAY SITE

Zia Ayu Nuansa Gumilang

Department of Statistics, Faculty of Mathematics and Natural Sciences

Islamic University of Indonesia

## ABSTRACT

*One of the influence of the increasing number of internet users in Indonesia is the increasingly mushrooming of shopping activities through the internet media. This makes internet users want everything that is instant and not troublesome. Knowing that it is then used by most of the business to market their products, especially in the field of goods and services. Along with its development, now e-commerce can be accessed by using mobile phone device in the form of application which can be downloaded easily various application which can offer ecommerce service. One example of a reputable e-commerce application in Indonesia is Shopee. This is what encourages researchers to examine the Shopee data obtained from the website google play website. Labeling and then analyzed using the Naïve Bayes Classifier method to classify the reviews based on positive class sentiment category and negative sentiment class. From labeling results it is seen by text associations in each sentiment class to find an information that is considered important and can be useful for decision making. The result of classification of sentiments using Naïve Bayes Classifier model obtained an accuracy of 97.4%. Furthermore, in the text association process obtained information that Shopee Indonesia application users discuss the application, postage, price and satisfied for the positive class and for the negative class of information obtained the user a lot of talking about the application of postage, bad, hard, pictures, delivery, and seller. The results of these negative reviews are made in the fishbone diagram for troubleshooting.*

**Keywords :** *E-Commerce, Google Play, Shopee, Sentiment Analysis, Naïve Bayes Classifier, Text Association, Fishbone Diagram*

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Perkembangan *e-commerce* di Indonesia setiap tahunnya mengalami pertumbuhan dan perkembangan yang sangat cepat. Hal ini tak terlepas dari semakin banyaknya jumlah pengguna internet di Indonesia yang menginginkan segala hal yang bersifat instan dan tidak merepotkan. Selain itu potensi lainnya yang mendorong berkembangnya bisnis *e-commerce* di Indonesia disebabkan oleh jumlah masyarakat yang tergolong besar dan letak geografis Indonesia sebagai negara kepulauan sehingga bisnis *e-commerce* dapat menjembatani antara para pelaku bisnis dan konsumennya.

Berdasarkan data *survey* yang diselenggarakan oleh Asosiasi Penyelenggaraan Jasa Internet Indonesia (APJII) menunjukkan bahwa jumlah pengguna internet di Indonesia pada tahun 2017 tercatat sebanyak 143,26 juta jiwa dari total keseluruhan penduduk Indonesia yaitu sebanyak 262 juta orang. Angka tersebut tercatat mengalami peningkatan jumlah dibandingkan pada tahun sebelumnya, yakni tahun 2016 yang tercatat mencapai 132,7 juta jiwa. Pada beberapa tahun kedepan diperkirakan jumlah pengguna internet di Indonesia akan terus meningkat (APJII, 2017).

Salah satu pengaruh dari peningkatan jumlah pengguna internet di Indonesia ialah semakin menjamurnya kegiatan berbelanja melalui media internet. Hal ini dimanfaatkan oleh sebagian besar pembisnis untuk memasarkan produknya terutama dibidang barang dan jasa. Berdasarkan data *survey* yang dilakukan lembaga riset ICD memperkirakan bahwa adanya potensi pertumbuhan *e-commerce* di Indonesia yang diduga akan tumbuh 42% pada 2012-2015, yang akan menempatkan indonesia berada di posisi pertama dengan pertumbuhan *e-commerce* terbesar untuk wilayah Asia-Pasifik (Saragih dan Ramadhani, 2012). Sementara itu, berdasarkan data *survey* yang dilakukan oleh APJII (2017)

memaparkan bahwa jual dan beli secara online masuk dalam 15 besar layanan yang diakses oleh pengguna internet di Indonesia. Hal ini dimanfaatkan oleh sebagian besar perusahaan-perusahaan di Indonesia yang bergerak dibidang *e-commerce* untuk terus berinovasi dalam meningkatkan pelayanannya. Seiring perkembangannya, kini *e-commerce* dapat diakses dengan menggunakan perangkat *mobile phone* dalam bentuk aplikasi yang dapat diunduh dengan mudah.

Pada saat ini terdapat berbagai macam aplikasi yang dapat menawarkan layanan *e-commerce*. Salah satu contoh aplikasi *e-commerce* yang memiliki reputasi baik di Indonesia ialah *Shopee*. Berdasarkan data riset yang dilakukan oleh *iprice* pada tahun 2017 bahwa *Shopee* masuk dalam 10 besar *e-commerce* yang paling sering dikunjungi dan diakses dengan data kunjungan sebanyak 25.736.356 oleh pengguna internet (Iprice, 2017). Perusahaan ini memiliki konsep dan ciri khas yaitu *consumer to consumer* (Reza, 2015). Konsep ini dapat diartikan bahwa transaksi jual-beli dilakukan oleh konsumen ke konsumen lainnya. Perbedaan *e-commerce* ini dengan yang lainnya ialah pengguna aplikasi tersebut dapat menjadi pihak penjual atas nama pribadi tanpa harus menjadi pelaku bisnis besar serta merupakan *marketplace* sosial di Indonesia yang tidak hanya fokus jual-beli, akan tetapi dapat berinteraksi sesama pengguna melalui fitur pesan instan yang disediakan secara langsung. *Shopee* pertama kali didirikan yaitu pada tahun 2015. Meskipun *e-commerce* ini masih tergolong baru, akan tetapi pengguna aplikasi *e-commerce* ini semakin meningkat. Berdasarkan data *Shopee* mengalami pertumbuhan tertinggi yakni 76,7 persen dan mencapai angka 10,8 juta pengguna. Selain itu, saat ini *Shopee* telah diunduh lebih dari 10 ribu kali dan telah menggandeng 1,3 juta pengguna aktif (Indra, 2017).

Aplikasi *Shopee* yang terdapat disitus *Google Play* tercatat sempat berada pada posisi pertama sebagai aplikasi *e-commerce* yang menawarkan layanan gratis ongkos kirim keseluruh Indonesia, akan tetapi pada bulan Maret 2018 posisi tersebut tergeser oleh salah satu aplikasi *e-commerce* lainnya yaitu *Lazada*. Hal ini disebabkan karena adanya perubahan sistem *Shopee* yang menyebabkan adanya pembatasan biaya gratis ongkos kirim. *Google Play* itu sendiri merupakan suatu

layanan dengan fasilitas digital yang diperkasai oleh perusahaan *Google* yang didalamnya mencakup sebuah toko atau media yang digunakan untuk memasarkan produk-produk seperti aplikasi, permainan, buku, musik/lagu dan sebagainya. Layanan tersebut dapat diakses melalui website, aplikasi *android* (*Play Store*), ataupun *Google TV*. Hal yang menarik dari aplikasi *google play* ialah adanya sebuah fitur yang dapat berisi ulasan (*review*) dari para penggunanya. Hal ini dapat dimanfaatkan untuk melihat ulasan atau *review* dari pengguna aplikasi tertentu, salah satunya ialah *Shopee*. *Review* dari pengguna tersebut umumnya dapat digunakan sebagai suatu tolak ukur yang efektif dan efisien untuk menemukan informasi terhadap suatu produk.

*Review* atau ulasan dari pengguna aplikasi seperti halnya *Shopee* umumnya berisi saran yang bersifat positif maupun keluhan yang bersifat negatif yang dituliskan dapat secara tidak langsung baik sedikit atau pun banyak, akan berpengaruh pada calon pelanggan. Untuk mensortir dan memantau ulasan tersebut bukan hal yang ringan karena jumlah ulasan yang dimuat dalam media sosial umumnya jumlahnya sangat banyak apabila diproses secara manual. Oleh karena itu dibutuhkan suatu metode yang dapat mensortir dan memantau ulasan tersebut secara cepat dan otomatis dalam mengkategorikan ulasan-ulasan tersebut baik yang positif maupun yang bersifat negatif (Wijayanti, 2014). Pada umumnya jumlah data ulasan pengguna aplikasi *Shopee* yang masuk ke situs *Google Play* akan terus bertambah, hal ini menyebabkan sulitnya pihak perusahaan untuk memperoleh data informasi secara menyeluruh dari semua ulasan yang ada. Oleh sebab itu diperlukan suatu metode atau teknik khusus untuk mengumpulkan data informasi dalam jumlah besar. Menurut Marres, Joan dan Wilson (2013) metode yang cocok untuk mengumpulkan data informasi tersebut ialah dengan teknik *scraping*. *Scraping* merupakan suatu teknik yang memberikan kesempatan untuk mengembangkan bentuk baru dalam pengumpulan data, analisis dan visualisasi.

Pada penelitian ini, pengumpulan data ulasan atau *review* yang diberikan oleh pengguna aplikasi *Shopee* dilakukan dengan teknik *scraping*. Kemudian pada proses analisisnya, penulis mencoba untuk melakukan klasifikasi teks ulasan pengguna guna mengidentifikasi mana ulasan yang berbentuk positif dan negatif.



Setelah melakukan klasifikasi, penulis mencoba mengekstrak dan mengeksplorasi seluas-luasnya informasi apa yang ada pada ulasan-ulasan tersebut yang sekiranya dianggap penting untuk digunakan pada berbagai keperluan.

Dalam penelitian ini proses klasifikasi sentimen akan dilakukan dengan menggunakan metode *Naïve Bayes Classifier*. Kemudian untuk proses ekstraksi dan eksplorasi penulis menggunakan statistik deskriptif dan asosiasi antar *terms* (kata atau topik yang sering dibicarakan) yang saling berkaitan. Metode *Naïve Bayes Classifier* ini adalah metode klasifikasi biner yang memanfaatkan probabilitas statistika sederhana dengan menerapkan aturan *Bayesian* menggunakan asumsi independen yang kuat. Metode *Bayesian* merupakan metode analisis berdasar informasi sampel dan informasi prior. Gabungan dari informasi sampel dengan informasi prior tersebut dinamakan peluang posterior. Penerapan metode *Naïve Bayes Classifier* adalah dengan memanfaatkan data *training* untuk menguji data *testing*.

Harapannya dengan penelitian ini mampu mengklasifikasikan teks dengan baik sehingga nantinya informasi yang ada di dalamnya dapat diekstraksi dengan baik serta penyajian informasi dari data yang diamati dapat memberikan informasi yang berguna bagi berbagai pihak yang membutuhkannya.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah diatas, maka dalam penelitian ini permasalahan yang akan dirumuskan yaitu sebagai berikut:

1. Bagaimana gambaran umum mengenai persepsi pengguna *Shopee* pada *website Google Play*?
2. Bagaimana hasil dari penerapan metode *Naïve Bayes Classifier* dalam mengklasifikasikan data ulasan pengguna *Shopee* menjadi kelas positif dan negatif?
3. Informasi apa yang diperoleh dalam setiap klasifikasi dan asosiasi teks yang telah dilakukan?
4. Faktor-faktor apa saja yang harus dilakukan untuk memperbaiki dari hasil ulasan negatif yang didapat?

### 1.3 Batasan Masalah

Adapun batasan masalah yang ditentukan dalam penelitian ini yaitu sebagai berikut:

1. Pada penelitian ini menggunakan data ulasan mengenai aplikasi *E-commerce Shopee* pada *website Google Play* yang diambil pada *upgrade* pertengahan tanggal 17 Januari 2018 – akhir Maret 2018.
2. Ulasan yang diambil adalah ulasan dengan menggunakan bahasa Indonesia.

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini yaitu sebagai berikut :

1. Mengetahui gambaran umum mengenai persepsi pengguna *Shopee* pada *website Google Play*.
2. Hasil dari penerapan metode *Naïve Bayes Classifier* dalam mengklasifikasikan data ulasan pengguna *Shopee* menjadi kelas positif dan negatif.
3. Memperoleh informasi yang penting dan berguna dalam setiap klasifikasi dan asosiasi teks yang dilakukan.
4. Mengetahui faktor-faktor yang harus dilakukan untuk memperbaiki dari hasil ulasan negatif yang didapat.

### 1.5 Manfaat Penelitian

1. Mengetahui gambaran umum mengenai persepsi pengguna *Shopee* berdasarkan data ulasan pada *website Google Play*.
2. Pihak yang memiliki kepentingan dapat melihat informasi yang ada dalam kumpulan ulasan atau komentar yang sangat banyak, sehingga dapat fokus melakukan penanganan serta evaluasi kearah yang lebih baik.

### 1.6 Sistematika Penulisan

Sistematika penulisan yang dipergunakan dalam penulisan tugas akhir ini dapat diuraikan sebagai berikut:

## **BAB I PENDAHULUAN**

Pada bab pendahuluan ini akan dibahas tentang latar belakang masalah penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

## **BAB II TINJAUAN PUSTAKA**

Dalam bab ini memaparkan penelitian-penelitian terdahulu yang berhubungan dengan topik penelitian yaitu mengenai metode dengan Teks Mining yaitu metode *Naïve Bayes Classifier* dan Asosiasi.

## **BAB III LANDASAN TEORI**

Pada bab ini menjelaskan teori-teori dasar yang berhubungan dengan penelitian sebagai dasar teori untuk memecahkan permasalahan pada penelitian ini.

## **BAB IV METODOLOGI PENELITIAN**

Bab ini memaparkan populasi dan sampel penelitian, jenis dan sumber data, variabel penelitian, metode pengumpulan data, metode analisis data, dan tahapan penelitian.

## **BAB V ANALISIS DAN PEMBAHASAN**

Pada bab ini akan dibahas mengenai analisa yang dilakukan terhadap hasil pengumpulan data dan penguraian hasil dari pengolahan dan analisa dimana hasil penelitian ini untuk mengetahui hasil keakuratan dari hasil pengklasifikasian dengan metode *naïve bayes classifier* dan hasil dari asosiasi berdasarkan kelas pengklasifikasian.

## **BAB VI PENUTUP**

Pada bab ini akan memuat kesimpulan mengenai hasil yang telah dilakukan pada bab sebelumnya berdasarkan dari rumusan masalah serta serta pemberian saran-saran yang dapat diimplementasikan dari hasil pengolahan data menjadi masukan yang berguna untuk kedepannya.

## **BAB II**

### **TINJAUAN PUSTAKA**

Berdasarkan hasil peninjauan dari berbagai sumber karya ilmiah, terdapat beberapa penelitian sebelumnya yang membahas berkaitan dengan penerapan metode *naïve bayes classifier* dan metode algoritma apriori dalam pensortiran ulasan atau *review* pengguna atau pelanggan suatu aplikasi. Seperti halnya penelitian yang dilakukan oleh Purwanto dan Santoso (2015) dengan menggunakan metode *naïve bayes classifier* untuk menganalisis mengenai sentimen *review* pelanggan situs penjualan online. Penelitian ini bertujuan untuk menentukan sentimen *review* yang diberikan pelanggan terhadap suatu produk apakah positif atau negatif. Hasil penelitian memperoleh akurasi yang cukup baik, yaitu sebesar 90,2%.

Dalam penelitiannya, Hayatin, Mentari dan Izzah. (2014) menggunakan metode *Naive Bayes Classifier (NBC)* dalam menyelesaikan masalah mengenai ekstraksi opini terhadap publik figur menggunakan data tweet pada twitter. Penelitian ini menggunakan tambahan fitur untuk melakukan sentimen analisis. Fitur yang digunakan adalah jumlah followers lover dan akun hater. Pada penelitiannya, peneliti menggunakan 600 tweet terhadap 6 tokoh politik yang diklasifikasikan ke dalam tiga kategori, yaitu netral, positif dan negatif. Dari hasil penelitiannya mendapatkan nilai presisi 99% untuk positif dan 28% untuk negatif sedangkan nilai recall 75% untuk positif dan 95% untuk negatif.

Hamzah (2012) melakukan penelitian pada level dokumen, yaitu mengklasifikasi teks berita dan abstrak akademis dari berbagai jurusan menggunakan metode *Naïve Bayes* dengan seleksi fitur kata yang digunakan sebagai dasar klasifikasi. Jumlah data yang digunakan sebanyak 1000 dokumen berita dan 450 dokumen abstrak. Hasil pengujian menunjukkan pada dokumen berita hasil akurasi maksimal 91%, sedangkan pada dokumen abstrak 82%.



Xhemali, D., Hinde Christopher J., Stone, Roger J.(2009) dalam penelitian mereka dengan judul *Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages*. Dalam penelitian ini melakukan perbandingan metode untuk mendapatkan hasil yang paling baik pada pengklasifikasian halaman website (opini/*review*). Hasil penelitian menunjukkan tingkat akurasi paling baik adalah dengan menggunakan *Naive Bayes* dengan akurasi sebesar 97.

Dalam penelitian Tugas Akhir yang dilakukan oleh Ulwan (2016) dengan judul *Pattern Recognition Pada Unstructured Data Teks Menggunakan Support Vector Machine Dan Association* (Studi Kasus: Portal Layanan Aspirasi dan Pengaduan Online Rakyat) dengan menggunakan *Text Mining* metode *Support Vector Machine* dan metode asosiasi didapatkan nilai akurasi sebesar 96,7% yang didalamnya terdapat kelas keluhan masyarakat tentang mengeluhkan masalah macet, busway yang sering bermasalah, pembagian BLSM atau KPS yang tidak adil serta Telkom yang buruk. Sedangkan untuk kelas pertanyaan yang sering didatnyakan yaitu masalah seputar informasi mengenai agama, masalah BLSM dan KPS, sertifikasi, BPJS, tunjangan serta beasiswa.

Arifin (2016) dalam penelitiannya menyebutkan penentuan prioritas pemasangan internet untuk pelanggan baru dapat diimplementasikan dengan menggunakan algoritma *naive bayes*. Pengujian akurasi model dari sistem yang dikembangkan dengan menggunakan metode *10-fold cross validation* menghasilkan nilai akurasi sebesar 90% dengan sampel sebanyak 200 data. Sedangkan hasil pengujian akurasi model dari aplikasi *Rapidminer 7.1* menggunakan algoritma *support vector machine* dengan kernel *Radial Basis Function* (RBF) diperoleh akurasi sebesar 88% dengan sampel sebanyak 200 data. Pengujian dengan data testing menghasilkan akurasi 88.89% dengan sampel sebanyak 18 data sehingga untuk program bantu dapat dilakukan.

Pada penelitian yang dilakukan oleh Ramadhani (2015), dengan menggunakan *teks mining* metode *support vector machine* dan *naive bayes classifier* dilakukan pengklasifikasian untuk menghitung akurasi kebenaran dari

data didapatkan akurasi paling tinggi yaitu pada metode *naïve bayes classifier* data yang telah diklasifikasi kemudian dihitung tingkat akurasi kebenarannya menggunakan metode dengan menentukan fungsi kernel serta proporsi untuk data *training* dan data *testing* yang sesuai. Dari perbandingan nilai akurasi klasifikasi dengan menggunakan kedua metode diatas, didapatkan nilai akurasi paling tinggi adalah dengan menggunakan metode *naïve bayes classifier* (NBC) dengan akurasi sebesar 62,6295%. Dari 5412 dokumen *training* dan 2701 dokumen *testing* yang telah diklasifikasi, didapatkan proporsi opini untuk kelas positif sebesar 44,46501% dan opini untuk kelas negatif sebesar 55,53499%.

Dalam Tugas Akhir yang dilakukan oleh Mulajati (2017) dengan judul Implementasi Teknik *Web Scraping* dan Klasifikasi Sentimen Menggunakan Metode *Naïve Bayes Classifier* dan Asosiasi Teks (Studi Kasus : Data Ulasan Penumpang Maskapai Penerbangan Garuda Indonesia Pada Situs TripAdvisor) dengan menggunakan metode *Naïve Bayes Classifier* (NBC) diperoleh tingkat akurasi sebesar 82,02%, Pada hasil asosiasi teks untuk pengklasifikasi kelas positif memberikan informasi diantaranya terkait *service, time, food, seat, time, entertrainment, staff, check-in*, dan *cabin*, dan untuk pengklasifikasi kelas negatif diperoleh yang sering dikeluhkan diantaranya *service, seat, staff, food, hour, luggage, check-in*, dan *boarding*.

Dari penelitian-penelitian sebelumnya dapat diambil beberapa kesimpulan kelebihan *Naïve Bayes Classifier* dapat dilihat pada **Tabel 2.1** berikut.

**Tabel 2.1** Kelebihan *Naïve Bayes Classifier* pada penelitian-penelitian sebelumnya

Penulis	Fokus Penelitian	Metode	Hasil Penelitian
Xhemali (2009)	Perbandingan metode klasifikasi pada halaman website	<i>Naïve Bayes</i> vs. <i>Decision Trees</i> vs. <i>Neural Networks</i>	Akurasi paling tinggi dengan metode NBC sebesar 97,26%

Ramadhani (2015)	Klasifikasi Opini Publik (Review Movie)	<i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine</i>	Akurasi paling tinggi dengan metode NBC sebesar 62,6295%
Arifin (2016)	Klasifikasi prioritas data pelanggan PT. Time Excelindo	<i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine</i>	Akurasi paling tinggi dengan metode NBC sebesar 90%

Berdasarkan dari penelitian-penelitian sebelumnya dan **Tabel 2.1** penulis dapat mengimplementasikan *Naïve Bayes Classifier* dan Asosiasi untuk Sentimen Analisis Data Ulasan Aplikasi *E-Commerce Shopee* pada Situs *Google Play* menggunakan Teks Mining dengan metode *Naïve Bayes Classifier (NBC)* pada data ulasan sebanyak 3.840 ulasan pada situs *website playstore* tentang aplikasi *e-commerce Shopee*. Dalam penelitian ini dilakukan pengklasifikasian ke dalam dua kategori kelas yaitu kelas positif dan kelas negatif. Dari hasil penelitian ini didapatkan nilai akurasi sebesar 97,4%, dengan hasil asosiasi pada kelas positif yaitu pada kata “Aplikasi”, “ongkir”, “harga”, dan “puas”. Untuk hasil asosiasi kelas negatif yaitu pada kata “aplikasi”, “ongkir”, “buruk”, “susah”, “gambar”, “pengiriman”, dan “penjual”. Hasil ulasan negatif tersebut dibuat dalam diagram *fishbone* untuk pemecahan masalah.

## **BAB III**

### **LANDASAN TEORI**

#### **3.1 *E-Commerce***

*E-Commerce* merupakan suatu media transaksi barang dan jasa melalui sistem informasi yang memanfaatkan teknologi informasi. Menurut Sutabri (2012) *E-Commerce* itu ialah penjualan, pembelian, penyebaran, pemasaran barang dan jasa yang melalui sistem elektronik seperti halnya internet, televisi berbayar, *website* atau jaringan komputer lainnya, yang dapat melibatkan transfer dana elektronik, penukaran data elektronik dan sistem pengumpulan data otomatis.

Menurut Turban, King, Lee, Liang dan Turban D. (2012) perdagangan elektronik (*electronic commerce* atau *e-commerce*) merupakan suatu cakupan proses yang meliputi penjualan, transfer, pembelian, atau pertukaran produk, layanan dan informasi melalui jaringan komputer, termasuk internet. Sebagian besar orang memandang istilah dari perdagangan elektronik (*e-commerce*) hanya dapat menjelaskan transaksi yang dilakukan antar mitra bisnis saja. Sehingga, kebanyakan orang yang menggunakan istilah *e-bussines* sebagai istilah penggantinya. Bisnis elektronik (*electronic bussines* atau *e-bussines*) mengarah pada definisi *e-commerce* yang lebih luas, tidak adanya pembelian dan penjualan barang saja. Akan tetapi juga pada layanan pelanggan, kolaborasi dengan mitra bisnis dan lain-lainnya.

##### **3.1.1 Jenis-Jenis *E-Commerce***

Menurut Turban, King, Lee, Liang dan Turban D. (2012) *E-Commerce* dapat dibedakan berdasarkan jenisnya secara umum ialah sebagai berikut.

##### **1. Bisnis ke Konsumen (*bussines to consumer*)**

Dalam bisnis ini, penjual ialah sebagai perusahaan. sedangkan pembeli ialah sebagai perorangan, atau bisa disebut juga *e-tailing*.

2. Bisnis ke Bisnis (*bussines to bussines*)

Dalam transaksi bisnis ke bisnis ini, baik penjual maupun pembeli ialah organisasi bisnis.

3. Konsumen ke konsumen (*consumer to consumer*)

Dalam transaksi ini, seorang menjual produknya ke orang lain. Orang yang menjual produk dan jasa ke satu sama lainnya.

4. Perdagangan kolaborasi (*collaborative c commerce*)

Dalam *c-commerce* ini umumnya para mitra bisnis saling berkolaborasi secara elektronik. Kolaborasi ini seringkali terjadi antara mitra bisnis dalam memenuhi pasokan barang dan jasanya.

5. Konsumen ke bisnis (*consumer to bussines*)

Dalam transaksi ini, konsumen memberitahukan kebutuhan atas produk atau jasa tertentu, dan para pemasok bersaing untuk dapat menyediakan produk atau jasa tersebut ke konsumen.

6. Perdagangan mobile (*mobile commerce atau m-commerce*)

Transaksi ini tidak hanya dilakukan dalam lingkungan nirkabel saja, akan tetapi dapat dengan menggunakan telepon seluler untuk mengakses internet dan berbelanja, maka hal ini disebut *m-commerce*.

### 3.2 Shopee

*Shopee* merupakan salah satu perusahaan yang bergerak dibidang aplikasi *e-commerce* dan website secara *online*. *Shopee* mulai masuk ke pasar Indonesia pada akhir bulan Mei 2015 dan *Shopee* baru mulai beroperasi pada akhir Juni 2015 di Indonesia untuk memenuhi permintaan dan gaya hidup masyarakat di Indonesia. *Shopee* merupakan anak perusahaan dari Garena yang berbasis di Singapura yang dipimpin oleh *CEO* bernama Chris Feng. *Shopee* telah hadir di beberapa negara di kawasan Asia Tenggara seperti Singapura, Malaysia, Vietnam, Thailand, Filipina, dan Indonesia (Margaretha, 2017).

*Shopee* merupakan *e-commerce* yang menawarkan berbagai produk barang seperti halnya pakaian pria, pakaian wanita, alat rumah tangga, barang elektronik dan kebutuhan olahraga dan sebagainya. *Shopee* diketahui merupakan *mobile*



*marketplace* pertama di Indonesia yang menawarkan biaya gratis ongkos kirim *se-Indonesia*. *Shopee* hadir di Indonesia dengan membawa konsep dan pengalaman berbelanja baru. Aplikasi *Shopee* ini dapat memfasilitasi penjual untuk berjualan dengan mudah dan aman dalam proses pembayaran serta pengaturan logistik yang terintegrasi. Saat ini aplikasi *Shopee* tidak hanya tersedia hanya di *website* resminya saja, akan tetapi sudah dapat tersedia untuk perangkat *mobile* dengan sistem operasi Android dan iOS.

Meskipun *Shopee* merupakan aplikasi belanja online yang terbilang baru diantara aplikasi online jenis lainnya, hingga saat ini *Shopee* telah di download sebanyak 10 juta kali melalui aplikasi *Google Play*. Tidak kalah dengan aplikasi populer lainnya seperti Lazada yang berdiri di tahun 2012 dengan hasil download sebanyak 50 ribu kali, Tokopedia yang berdiri sejak 2009, Bukalapak yang berdiri sejak 2010 dan OLX yang berdiri dari tahun 2005 yang masing-masing dengan jumlah perolehan download sebanyak 10 ribu kali download (Margaretha, 2017). Di tahun 2016 ini *Shopee* berhasil membukukan total nilai transaksi sebesar USD 1.8 miliar untuk kawasan Asia Tenggara dan Taiwan. Indonesia muncul sebagai salah satu pasar terbesar dan potensial bagi *Shopee*, dengan kontribusi mencapai 43 persen dari keseluruhan bisnis perusahaan (<https://www.google.co.id/amp/s/telset.id/153102/shopee-kembali-pasang-target-double-digit-di-2018/amp/>, diakses pada 28 Maret 2018).

### 3.3 Analisa Sentimen

Menurut Kristiyanti (2015), analisa sentimen atau *opinion mining* adalah studi komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas. Entitas tersebut dapat menggambarkan individu, kejadian atau topik. Analisa sentimen juga terlibat dalam pembuatan sistem untuk mengoleksi dan memeriksa pendapat tentang suatu produk yang dibuat pada suatu postingan *blog*, *review* atau *tweet*. Pada analisa sentimen terdapat beberapa tantangan dalam menganalisanya, misalnya suatu opini dapat dikatakan sebagai opini positif pada satu situasi tertentu namun juga bisa dikatakan negatif di lain situasi, orang-orang

tidak menyampaikan opini-opini dengan cara yang sama sehingga akan susah dalam menganalisanya.

### **3.4 *Data Mining***

#### **3.4.1 *Pengertian Teks Mining***

Data mining merupakan suatu konsep yang umumnya digunakan untuk menemukan pengetahuan yang tersembunyi dalam basis data. Data mining secara umum merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang tersimpan dalam basis data (Turban Aronson, Liang dan McCarthy, 2006).

Saat ini istilah data mining sudah berkembang jauh dalam mengadaptasi setiap bentuk analisa data. Dimana data mining ini berhubungan dengan analisa data dan penggunaan teknik-teknik perangkat lunak untuk mencari keteraturan dan pola dalam himpunan data yang bersifat tersembunyi. Teknik-teknik ini banyak variasainya dan disesuaikan dengan kebutuhan analisis datanya (Turban Aronson, Liang dan McCarthy, 2006).

Menurut Deepa dan Geetha (2013) data mining adalah serangkaian aktivitas yang digunakan dalam menemukan pola baru yang bersifat tersembunyi atau pola yang tidak terduga sebelumnya dalam data. Lalu menurut Sadiku, Shadare dan Musa. (2015) data mining ialah menemukan pola yang berarti pada jumlah data yang banyak. Sedangkan menurut Fayyad, Piatetsky dan Smyth. (1996) data mining adalah aplikasi dari algoritma yang spesifik untuk mengekstrak pola dari data. Pola yang dihasilkan dari data mining dapat digunakan dalam melakukan prediksi terhadap data baru berdasarkan pola itu tersebut. Pola tersebut direpresentasikan dalam suatu bentuk struktur yang dapat dilakukan analisa, dapat dan mudah dimengerti serta dapat digunakan dalam pembuatan keputusan (Witten Frank dan Hall, 2011).

### 3.5 *Preprocessing*

Tahapan *preprocessing* adalah dimana dilakukannya penyeleksian data yang akan diproses. Setiap kata akan dipecah menjadi beberapa bagian yang lebih kecil sehingga mempunyai arti yang lebih sempit. Ada beberapa tahapan dalam proses *preprocessing* yaitu *correcting slang word*, filtrasi, *stemming*, dan tokenisasi.

#### 3.5.1 *Correcting Slang Word*

*Slang word* merupakan penggunaan kata-kata informal dan ekspresi yang tidak baku atau dianggap tidak standar dalam bahasa tertentu tapi mungkin bisa diterima jika digunakan dalam pergaulan. *Slang word* perlu diperbaiki guna untuk menjadikannya informasi yang dapat diproses. Contoh dari *slang word* adalah “yg”, “smga”, “jhat”, “sgera” dan lain sebagainya.

#### 3.5.2 Filtrasi

Tahap filtrasi merupakan tahap pengambilan kata-kata yang penting dari hasil token. Pada tahap ini digunakan Algoritma *stoplist* (membuang kata yang dirasa kurang penting dan tidak ada pengaruhnya jika kata itu dibuang) atau *wordlist* (menyimpan kata yang dirasa penting). Sedangkan untuk *Stopword* adalah kata-kata yang kurang penting dan tidak bersifat deskriptif dari suatu dokumen sehingga dapat dilakukan pembuangan kata tersebut. Contohnya adalah “dan”, “saya”, “kamu” dan seterusnya (Putri, 2016).

Dalam filtrasi ini menggunakan *stopword/stoplist* supaya kata-kata yang sering muncul yang bukan merupakan kata penting dalam sebuah dokumen dapat dibuang tanpa mempengaruhi makna atau arti dari dokumen yang akan diproses ke tahap selanjutnya.

#### 3.5.3 *Stemming*

*Stemming* dilakukan guna dalam pengolahan kata untuk mendapatkan kata dasar dari sebuah kata yang telah mengalami imbuhan dengan asumsi bahwa kata-kata tersebut sebenarnya memiliki makna dan arti yang sama. Algoritma ini bekerja berdasarkan struktural *morfologi* dalam kalimat bahasa Indonesia, yang

terdiri atas awalan, akhiran, sisipan, dan awalan+akhiran. Inti dari tahap ini memiliki tujuan (1) dalam perkara keefisiensian, pada *stemming* dilakukan pengurangan jumlah kata dalam dokumen agar mengurangi kebutuhan dalam ruang penyimpanan dan mempercepat dalam melakukan pencarian. (2) dalam perkara keefektifan, *stemming* dilakukan untuk mengurangi *recall* dengan pengurangan bentuk-bentuk kata ke dalam bentuk dasarnya (Putri, 2016). Sebagai contoh adalah kata “duduk-lah”, “minum-lah”, “jika-pun” dan sebagainya.

### 3.5.4 Tokenisasi

Tokenisasi merupakan proses memecah kalimat menjadi kata-kata menjadi lebih berarti dan bermakna. Tahapan pertama yang dilakukan yaitu dengan normalisasi kata dengan mengubah semua karakter huruf menjadi *toLowerCase* atau huruf kecil. Proses ini diawali dengan menghilangkan delimiter-delimiter seperti tanda baca dan simbol yang ada pada dokumen tersebut seperti tanda (,),\$,\*,@, !, ?, / dan sebagainya. Tahap selanjutnya yaitu proses penguraian teks yang semula berupa kalimat-kalimat yang berisi kata-kata. Proses pemotongan *string* berdasarkan tiap kata yang menyusunnya, seperti halnya spasi, proses tokenisasi ini mengandalkan karakter spasi pada dokumen teks guna melakukan pemisahan. Hasil dari proses ini hanyalah kumpulan kata saja (Putri, 2016).

### 3.6 Fitur dan Pembobotan

Pembobotan merupakan metode untuk mengubah *input* data menjadi suatu fitur vektor. Metode pembobotan yang umum digunakan adalah *bag-of-feature*. Sebagai contoh terdapat sederetan fitur seperti pada vector  $\{f1, f2, ..., fn\}$  dimana yaitu sekumpulan fitur-fitur sebanyak  $n$  yang sudah ditentukan sebelumnya. Misalkan kata “puas” maka fitur vektor dari data adalah vektor.

#### a. *Term Presence*

*Term Presence* (TP) ialah metode pembobotan pada suatu dokumen teks yang melihat keberadaan daftar kata-kata (*term*) atau fitur yang ada pada *corpus* terhadap suatu dokumen. Jika suatu fitur yang ada pada daftar fitur acuan terdapat pada dokumen yang sedang diboboti maka nilai fitur tersebut pada

*feature vector* akan diberi nilai 1 dan tidak menghiraukan jumlah kemunculan fitur tersebut. Jika fitur tersebut tidak ada pada dokumen maka diberi nilai 0 pada *feature space* (O’Keefe dan Koprinska, 2009). Rumus yang dipakai untuk menghitung *Term Presence* (TP) dari fitur  $t_i$ , pada dokumen  $d_j$  ditulis dengan notasi 3.1.

$$tp(t_i, d_j) = \begin{cases} 1 & \text{Jika terdapat } t_i \text{ pada } d_j \\ 0 & \text{Jika tidak terdapat } t_i \text{ pada } d_j \end{cases} \quad (3.1)$$

**b. Term Frequency**

*Term Frequency* (TF) memiliki kesamaan dengan TP yang sudah dijelaskan sebelumnya, tapi yang membedakan adalah TF menghitung jumlah kemunculan fitur acuan pada suatu dokumen bukan hanya keberadaan fitur tersebut (O’Keefe dan Koprinska, 2009). Rumus TF dapat ditulis dalam persamaan 3.2 dengan  $\#(t_i, d_j)$  mempunyai arti jumlah kemunculan fitur  $t_i$  pada dokumen  $d_j$ . Misalkan suatu fitur berupa kata “bagus” muncul sebanyak 10 kali maka nilai fitur tersebut pada *feature vector* adalah 10.

$$tf(t_i, d_j) = \#(t_i, d_j) \quad (3.2)$$

**c. Term Frequency – Inverse Document Frequency (TF-IDF)**

*Term Frequency – Inverse Document Frequency* (TF-IDF) merupakan algoritma pembobotan tersusun dari dua nilai yang berasal dari dua algoritma dengan pembobotan yang berbeda, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Rumus 3.3 menunjukkan formula perhitungan IDF pada suatu kumpulan dokumen D dengan  $|D|$  merupakan jumlah dokumen dan  $\#d(t_i)$  merupakan banyaknya dokumen dimana suatu kata ( $t_i$ ).

$$idf(t_i, d_j) = \log \frac{|D|}{\#d(t_i)} \quad (3.3)$$

$$tfidf(t_i, d_j) = tf(t_i, d_j) \times idf(t_i, d_j) \quad (3.4)$$

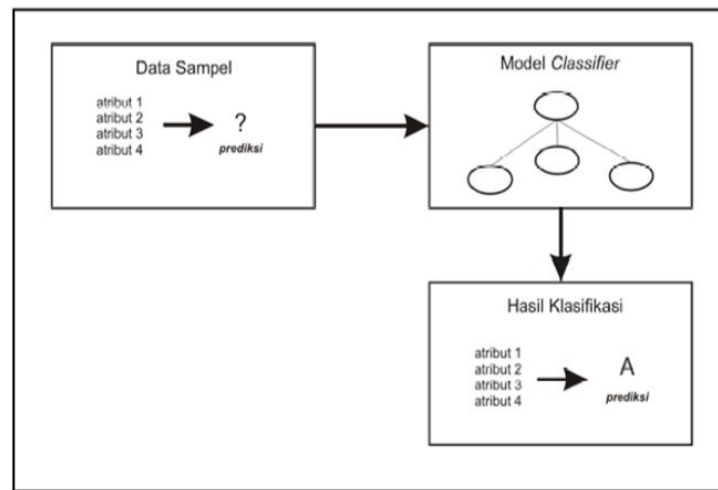
Keluaran dengan fitur/*term* tersebut dengan kata yang sering muncul pada dokumen akan menghasilkan nilai TF-IDF yang tinggi. Sementara, fitur yang sering muncul pada dokumen akan menghasilkan nilai rendah. Dengan menggunakan metode ini *terms* atau fitur-fitur yang penting akan memiliki nilai tinggi dan sebaliknya untuk *term* atau fitur yang tidak penting akan memiliki nilai yang rendah (O’Keefe dan Koprinska, 2009).

### 3.7 Klasifikasi

Klasifikasi merupakan suatu pekerjaan yang melakukan penilaian terhadap suatu objek data untuk masuk dalam suatu kelas tertentu dari sejumlah kelas yang tersedia (Prasetyo, 2012). Teknik klasifikasi merupakan teknik yang paling cocok untuk memprediksi atau menggambarkan kumpulan data dengan kategori biner atau nominal namun kurang efektif untuk kategori ordinal (misalkan, untuk mengklasifikasikan seseorang dengan kriteria tinggi, menengah dan rendah untuk tingkat penghasilan) (Tan, Michael dan Vipin. 2006).

Menurut Han dan Kamber (2006) data *classification* memiliki dua tahap proses. Tahap pertama adalah membangun suatu model yang berdasarkan serangkaian data *class* yang disebut *learned model*. Model tersebut dibangun dengan menganalisa *record database*. Setiap *record* diasumsikan menjadi *predefined class* yang ditentukan oleh suatu atribut yang disebut *class label* atribut. Akibat terdapat *class label* maka tahap ini juga dikenal dengan *supervised learning*. Berbeda dengan *unsupervised learning* atau dikenal dengan *clustering* yang tidak memerlukan *class label*. Tahap pertama ini juga disebut sebagai tahap pembelajaran. Sebuah algoritma klasifikasi akan membangun sebuah model klasifikasi dengan cara menganalisis data *training*. Tahap pembelajaran dapat juga dipandang sebagai tahap pembentukan fungsi atau pemetaan  $y = f(x)$  dimana  $y$  adalah kelas hasil prediksi dan  $x$  adalah *record* yang ingin diprediksi *class*-nya.

Bagan proses klasifikasi data sampel menggunakan model *classifier* untuk mendapatkan hasil prediksi (Han dan Kamber, 2006) dapat dilihat pada **Gambar 3.1** berikut :



**Gambar 3.1** Bagan proses klasifikasi (Han dan Kamber, 2006)

Beberapa persiapan yang dilakukan untuk mendapatkan hasil klasifikasi yang baik diantaranya adalah (Han dan Kamber, 2006):

1. Pembersihan Data

Pembersihan data ini dilakukan untuk mengurangi kecacatan data didalam data pelatihan, beberapa metode yang digunakan diataranya dengan teknik *smoothing* untuk menghilangkan noise data, melengkapi data yang hilang dan sebagainya.

2. Analisis Relevansi

Dari beberapa atribut yang akan digunakan untuk proses klasifikasi mungkin saja terdapat atribut yang sangat berhubungan kuat satu sama lain, kedua atribut ini memiliki kemiripan sehingga menyebabkan proses klasifikasi menjadi tidak optimal, maka salah satu dari atribut ini dapat dibuang.

Hasil klasifikasi dan prediksi dapat dievaluasi menggunakan beberapa kriteria (Han dan Kamber 2006):

1. Akurasi

Akurasi digunakan untuk mengetahui kemampuan model klasifikasi untuk dapat memberikan ketepatan hasil prediksi.

## 2. Kecepatan

Mengetahui kecepatan iterasi untuk mendapatkan model klasifikasi dan iterasi mendapatkan hasil prediksi.

### 3.8 Teorema Bayes

Teorema bayes merupakan teorema yang mengacu pada konsep probabilitas bersyarat (Tan, Michael dan Vipin. 2006). Metode ini merupakan pendekatan statistik untuk melakukan inferensi induksi pada persoalan klasifikasi. Misalkan A dan B adalah kejadian dalam ruang sampel. Larose (2006) menyatakan probabilitas bersyarat dalam persamaan (3.5).

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (3.5)$$

Dimana  $P(A \cap B)$  adalah probabilitas interaksi A dan B dan  $P(B)$  adalah probabilitas B. Demikian pula  $P(B|A) = \frac{P(A \cap B)}{P(A)}$ , sehingga nilai  $P(A \cap B) = P(B|A)P(A)$ . Nilai  $P(A \cap B)$  kemudian disubstitusikan ke dalam persamaan (3.5), maka diperoleh persamaan (3.6).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3.6)$$

### 3.9 Naïve Bayes Classifier

*Naïve Bayes Classifier* merupakan salah satu algoritma yang sederhana namun memiliki kemampuan dan akurasi yang tinggi dan termasuk dalam metode machine learning (Rish, 2001). Berdasarkan kompleksitasnya *Naïve Bayes* merupakan salah satu algoritma paling sederhana dalam menerapkan aturan *Bayes* disertai beberapa keunggulan yaitu sangat efisien, membutuhkan data uji yang sedikit, mudah untuk diimplementasikan dan memiliki akurasi relatif tinggi. Selain itu, *Naïve Bayes* juga memiliki kekurangan diantaranya akurasi akan berkurang jika atribut yang menjadi parameter dalam klasifikasi tidak independen atau nonparametric-continuous (Rianto, 2016).

Metode klasifikasi *Naïve Bayes* menggunakan konsep peluang dalam menentukan kelas dokumen. Metode ini menggunakan asumsi bahwa dalam



sebuah dokumen kemunculan kata tidak mempengaruhi kemunculan kata yang lain dan ketidak munculan kata tidak mempengaruhi ketidakmunculan kata yang lain (Rianto, 2016). *Naïve Bayes* merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan teorema *Bayes* dengan asumsi independensi yang kuat (Suryadi, 2017). *Naïve Bayes classifier* mengasumsikan ada atau tidaknya suatu fitur tertentu pada sebuah kelas tidak mempengaruhi keberadaan fitur lainnya.

#### a. Model Probabilitas untuk *Naïve Bayes*

Secara mendasar, model probabilitas untuk sebuah *classifier* adalah model peluang bersyarat.

$$P(C|F_1, \dots, F_n) \quad (3.7)$$

Dengan menggunakan teorema *Bayes* :

$$P(C|F_1, \dots, F_n) = \frac{P(C)P(F_1, F_2, \dots, F_n|C)}{P(F_1, F_2, \dots, F_n)} \quad (3.8)$$

Dimana Variabel *C* merepresentasikan kelas, sementara variabel  $F_1 \dots F_n$  merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas *C* (Posterior) adalah peluang munculnya kelas *C* (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas *C* (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{Prior \times Likelihood}{Evidence} \quad (3.9)$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan.

### 3.10 Asosiasi Teks

Dalam (Ulwan, 2016) disebutkan bahwa istilah korelasi sering sekali digunakan untuk menyatakan suatu hubungan dua atau lebih variabel yang bersifat kuantitatif, sedangkan istilah asosiasi sering diartikan keeratan hubungan antara dua atau lebih variabel yang bersifat kualitatif. Tujuan analisis korelasi adalah untuk mencari hubungan variabel bebas ( $X$ ) dengan variabel terikat ( $Y$ ), dengan ketentuan data memiliki syarat-syarat tertentu (Fadlisyah, 2014). Dapat dilihat pada persamaan (3.10) merupakan persamaan untuk menentukan nilai korelasi, pada persamaan (3.11) merupakan persamaan untuk mendapatkan *R Square*.

Pada penelitian ini digunakan pendekatan asosiasi untuk menemukan hubungan antar ulasan pengguna sehingga mendapatkan informasi yang dapat dijadikan bahan rujukan dalam pengembangan serta untuk meningkatkan kualitas perusahaan.

$$r = \frac{n \cdot (\sum XY) - (\sum X) \cdot (\sum Y)}{\sqrt{(n \cdot \sum X^2 - (\sum X)^2)(n \cdot \sum Y^2 - (\sum Y)^2)}} \quad (3.10)$$

$$R = r^2 \quad (3.11)$$

## **BAB IV**

### **METODOLOGI PENELITIAN**

#### **4.1 Populasi dan Sampel Penelitian**

Populasi dalam penelitian ini adalah *database website Google Play*, yaitu semua data ulasan *Shopee* pada bulan Januari 2018 hingga Maret 2018. Sedangkan untuk sampel yang digunakan adalah ulasan *Shopee* dari sejak pertengahan *upgrade* aplikasi *Shopee* yaitu tanggal 17 Januari 2018 hingga terakhir *upgrade* pada akhir Maret 2018.

#### **4.2 Jenis dan Sumber Data**

Jenis data yang digunakan pada penelitian ini ialah data primer. Data primer itu sendiri menurut Sekaran (2011) merupakan data yang mengacu pada suatu informasi yang diperoleh dari tangan pertama oleh peneliti yang dapat berkaitan dengan variabel minat untuk tujuan spesifik studi. Data pada penelitian ini diperoleh dengan menggunakan teknik *scraping* dari halaman situs *website shopee* dengan menggunakan aplikasi *Scraper* bawaan dari *Google Chrome* dengan alamat *website* yaitu <https://play.google.com/store/apps/details?id=com.shopee.id>). Data yang diperoleh berupa ulasan atau *review* pengguna aplikasi *Shopee* yang diambil sejak tanggal *upgrade* mulai dari 17 Januari 2018 sampai terakhir *upgrade* tanggal 31 Maret 2018 sebanyak 3.840 ulasan.

#### **4.3 Variabel Penelitian**

Variabel penelitian merupakan segala sesuatu yang dapat berbentuk apa saja dan dapat ditetapkan oleh penulis untuk dipelajari sehingga dapat memperoleh informasi yang kemudian dapat ditarik menjadi kesimpulan (Sugiyono, 2011). Adapun variabel yang digunakan dalam penelitian ini hanya ada dua macam, yaitu:

- a. *Date* merupakan tanggal dibuatnya komentar
- b. *Rating* merupakan tingkat kepuasan pengguna

- c. *Review* merupakan ulasan atau isi komentar pengguna

#### 4.4 Metode Pengumpulan Data

Pada penelitian ini, menggunakan metode pengumpulan data dengan menggunakan teknik *web scraping*. Menurut Vargiu (2012) *web Scraping* merupakan suatu gabungan teknik yang digunakan untuk mendapatkan informasi dari sebuah *website* secara otomatis tanpa harus menyalinnya secara manual. Adapun alat yang digunakan untuk *web scraping* ialah *software Scraper* yang ada di *Google Chrome*.

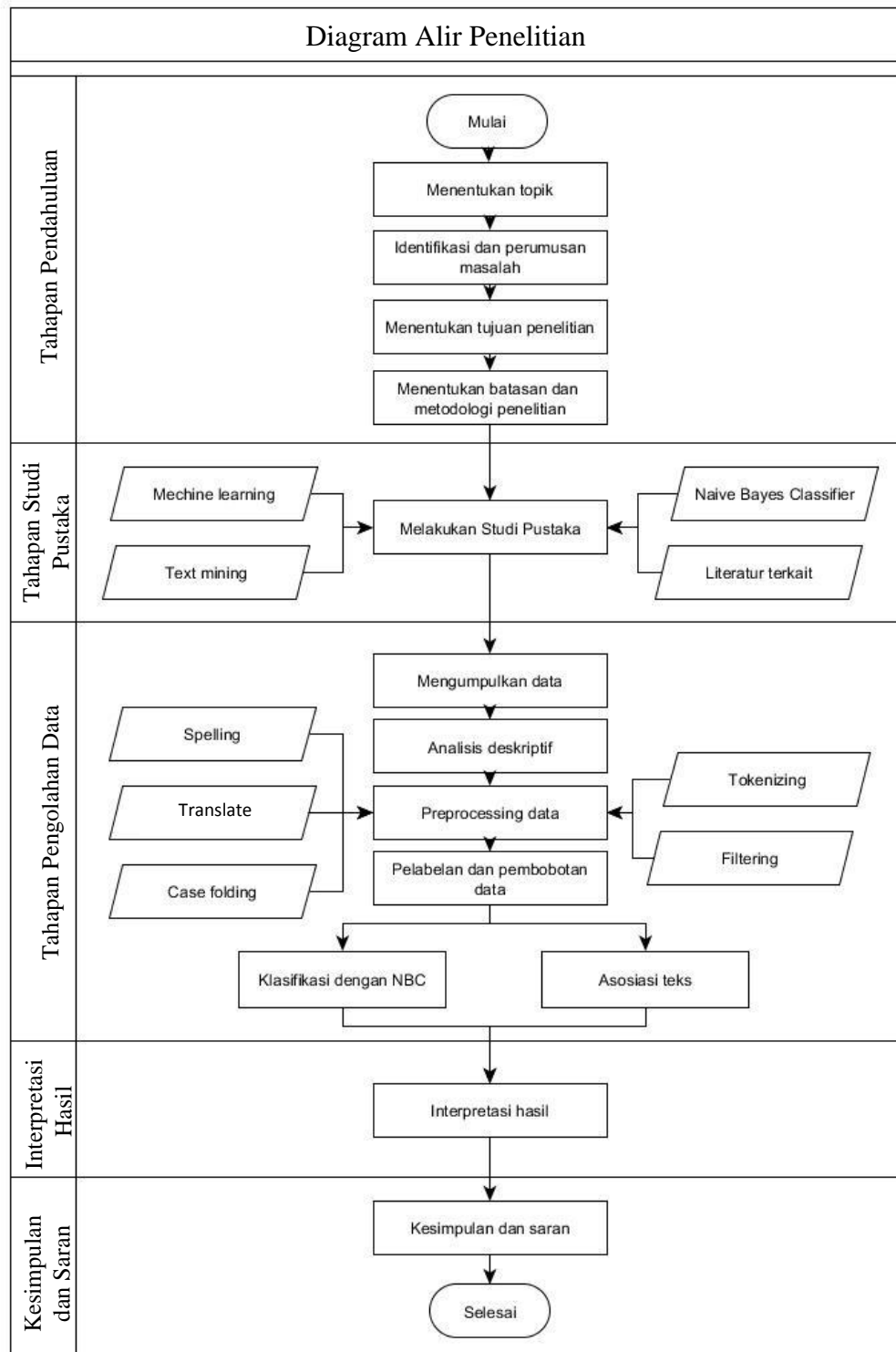
#### 4.5 Metode Analisis Data

Adapun beberapa metode analisis data yang digunakan dalam penelitian ini yang dapat membantu dalam menganalisis data ialah, antara lain:

1. Metode Analisis Deskriptif, yang digunakan dalam menggambarkan secara umum ulasan *Shopee* yang terdapat dalam situs *website Google Play*
2. Metode *Maching Learning* yaitu *Naive Bayes Classifier*, yang dalam penelitian ini digunakan untuk mengklasifikasikan ulasan atau *review* yang dapat berbentuk kelas positif maupun kelas negatif.
3. *Association*, yang digunakan dalam penelitian ini untuk mengidentifikasi dan membentuk pola kata yang dapat berasosiasi dengan kata lainnya untuk mendapatkan informasi yang dianggap penting.

#### 4.6 Tahapan Penelitian

Tahapan yang dilakukan pada penelitian ini digambarkan sebagai berikut :



**Gambar 4.1** Diagram alir penelitian

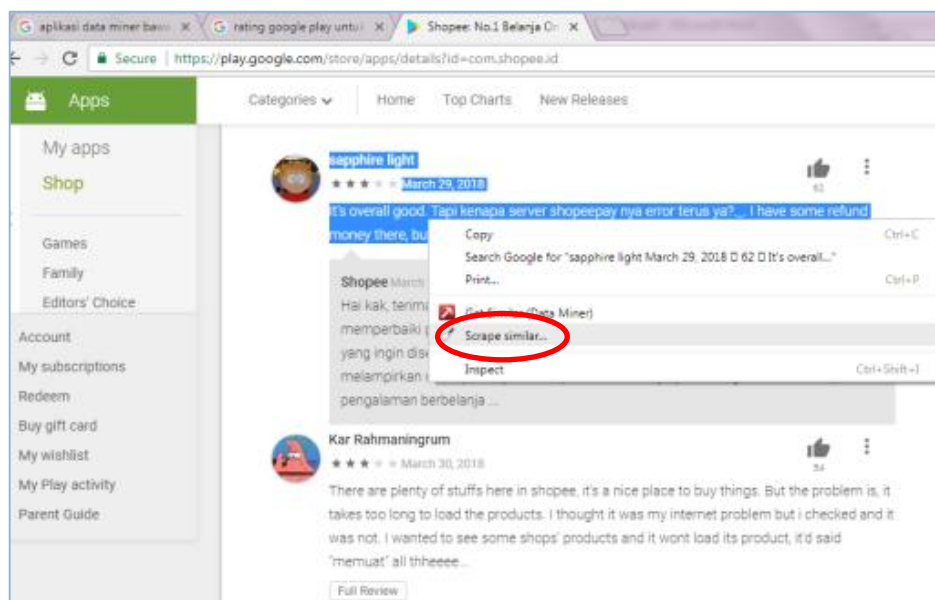
## BAB V

### ANALISIS DAN PEMBAHASAN

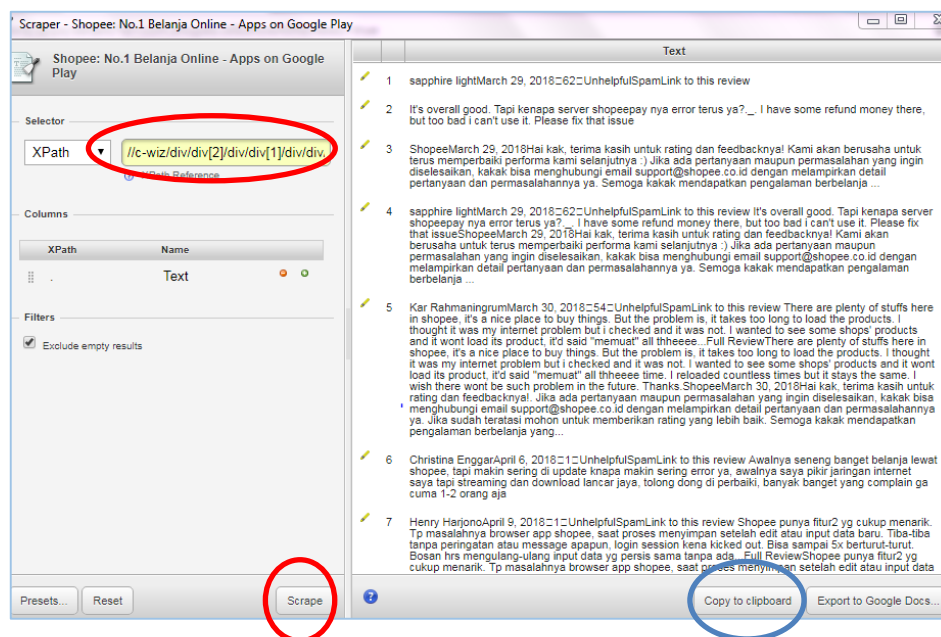
#### 5.1 Pengumpulan data

Pengambilan data ulasan aplikasi *shopee* yang ada di *google play* dengan alamat <https://play.google.com/store/apps/details?id=com.shopee.id>. pada pengumpulan data ini hanya dengan memblok tanggal, rating dan kalimat ulasan pertama yang dapat dilihat pada **Gambar 5.1**, selanjutnya dengan klik kanan lalu pilih *Scrape similar*, maka pada bagian *selector XPath* akan muncul sendiri dengan *XPath* `//c-wiz/div/div[2]/div/div[1]/div/div/div/div[2]/div/div/div` kemudian klik pada bagian *Scrape* yang dapat dilihat pada **Gambar 5.2**.

*XPath* (the *XML Path language*) adalah bahasa atau *script* yang digunakan untuk menemukan informasi dalam dokumen *XML*. *XML* (e*Xtensible Markup Language*) ini sendiri merupakan bahasa *markup* yang menyediakan kata-kata dan *tag* yang menggambarkan dokumen dan mengidentifikasi bagian-bagiannya. pengambilan keseluruhan ulasan dapat dilihat pada gambar 4.1 dan 4.2 berikut.



**Gambar 5.1.** Pemblokian data ulasan *Shopee* situs *google play*



**Gambar 5.2** Proses data scraping

Setelahnya untuk mengekstrak data ulasan dengan memilih *Copy to clipboard* (lingkaran berwarna biru **Gambar 5.2**), selanjutnya dengan membuka *ms. excel* dan menekan tombol **Ctrl + V** maka akan muncul tampilan seperti **Gambar 5.3** berikut.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	30 Maret 2018;3	Sekarang shopee appnya makin berat, sering ngelag. Tolong diperbaiki bugnya. Kalu sudah fix saya ganti ke bintang 5 lagi.												
2	30 Maret 2018;1	Yang bener aja. Di update terbaru malah home nggak ke-load sama sekali, mau akses lewat fitur apa pum gabisa, padahal ko												
3	29 Maret 2018;4	Bagus apk nya □□ aku suka												
4	29 Maret 2018;5	Love this app												
5	29 Maret 2018;3	jadi lemot parah nih. padahal wifi udh kenceng, mau ganti profile lemotnya ga nahan												
6	29 Maret 2018;5	Sangat membantu mempermudah berbelanja online... terima kasih shopee semoga semakin banyak discount" nya dan gratis												
7	29 Maret 2018;4	Akhir2 ini aplikasinya agak lolak.. Tp overall I LIKED THIS APP.												
8	29 Maret 2018;3	It's overall good. Tapi kenapa server shopeepay nya error terus ya? _ _ _ I have some refund money there, but too bad i can't use												
9	29 Maret 2018;1	Waktu pembayaran dengan CC dengan metode cicilan 6 bulan, ternyata sama pihak bank bisa. Padahal di apk bilang sudah t												
10	29 Maret 2018;1	Gratis ongkir penipuan... mesti ada limitnya baru gratis ongkir sama aja bohong, bilang aja setiap belanja diatas 100 rb gratis c												
1	29 Maret 2018;2	Input resi J&T tidak bisa manual												
2	29 Maret 2018;3	Sehari ini susah bgt buka halaman shoope.loading terus.mohon diperbaiki												
3	29 Maret 2018;3	Always force close												

**Gambar 5.3** Proses Input data di Ms.Excel

Setelah melakukan *input* data di *ms.excel* dilanjutkan dengan merapikan data dan melakukan edit kata-kata ulasan dimana sebenarnya untuk isi ulasan dengan bahasa Indonesia pada aplikasi Shopee banyak pengguna yang melakukan

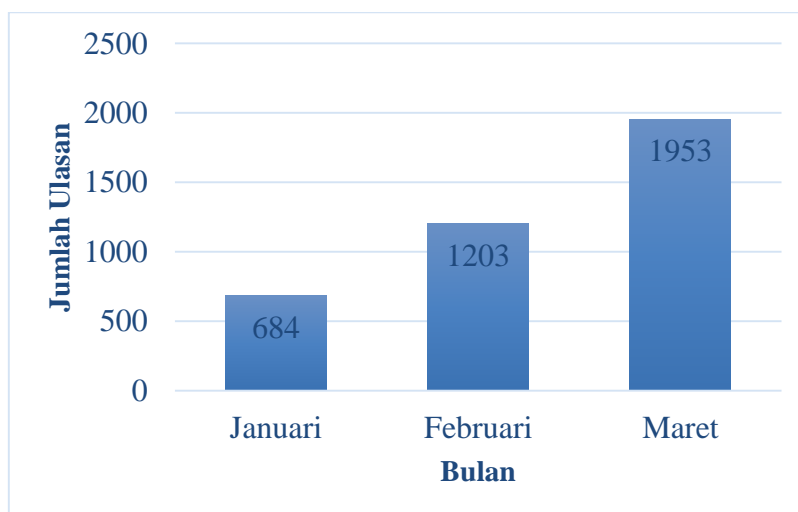
pemborosan kata (banyak kata-kata kurang penting/ curhat berkepanjangan), maka didapatkan setelah melakukan edit ulasan pada **Gambar 5.4** berikut.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Tanggal	Rating	Ulasan														
2	30 Maret 2018	2	3 Aplikasi shopee sekarang berat dan lambat, tolong perbaiki														
3	30 Maret 2018	1	1 Setelah di update aplikasi semakin lambat, tidak bisa akses fitur apapun meskipun koneksi cepat														
4	29 Maret 2018	4	4 aplikasi bagus dan saya suka sekali														
5	29 Maret 2018	5	5 suka aplikasi ini														
6	29 Maret 2018	3	3 koneksi wifi cepat tapi aplikasi lambat dan berat, gagal ganti profil														
7	29 Maret 2018	5	5 Sangat membantu mempermudah berbelanja online, terima kasih shopee semoga semakin banyak diskonnya dan gratis ongkir semakin banyak														
8	29 Maret 2018	4	4 akhir ini aplikasi lambat tapi untuk keseluruhan saya suka aplikasi ini														
9	29 Maret 2018	3	3 semua bagus tapi server shopeepay sering eror terus, saya sudah mendapatkan pengembalian dana akan tetapi tidak bisa digunakan														

**Gambar 5.4** Proses perubahan isi ulasan di *Ms.Excel*

## 5.2 Analisis Deskriptif

Analisis deskriptif dalam penelitian ini bertujuan untuk melihat gambaran secara umum informasi tentang *Shopee* berdasarkan data ulasan pengguna aplikasi *Shopee* dari situs *Google Play* yang sebelumnya diperoleh dengan teknik *scraping*. Dari data tersebut didapatkan gambaran tentang jumlah ulasan berdasarkan urutan waktu yang diambil dan *rating*.

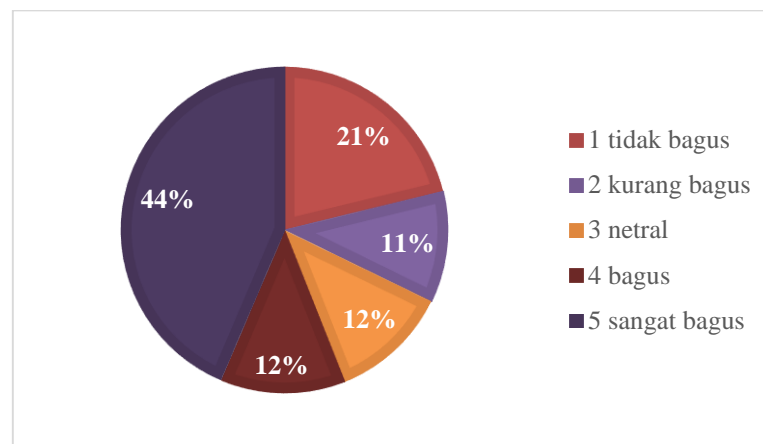


**Gambar 5.5** Grafik jumlah ulasan pengguna aplikasi *Shopee* Januari 2018 – Maret 2018

Berdasarkan **Gambar 5.5**, pada pengambilan data ulasan 3 bulan yaitu Januari 2018 hingga Maret 2018 diperoleh hasil meningkat jika dilihat dari grafik, hal ini sebenarnya untuk ulasan pengguna pada bulan Januari sebanyak 684 ulasan



dikarenakan pengambilan data dimulai pertengahan bulan Januari sesuai dengan tanggal *upgrade* aplikasi itu sendiri. Bulan Februari sebanyak 1.203 ulasan untuk normalnya dan terjadi peningkatan pada bulan Maret sebanyak 1.953 ulasan, hal ini diperkirakan karena sistem *Shopee* mengalami perubahan pada bulan tersebut.



**Gambar 5.6** Grafik Rating yang diberikan pengguna aplikasi *Shopee* Januari 2018 – Maret 2018

Berdasarkan **Gambar 5.6** ulasan berdasarkan *rating* terdapat 5 nilai kategori yaitu 1-5 dimana nilai 1 dikategorikan “tidak bagus”, nilai 2 dikategorikan “kurang bagus”, nilai 3 dikategorikan “netral”, nilai 4 dikategorikan “bagus” dan terakhir nilai 5 dikategorikan “sangat bagus. Dari grafik *pie chart* tersebut dapat dilihat bahwa dengan total 3.840 pengguna untuk aplikasi *Shopee* paling banyak memilih nilai 5 dengan persentase 44% atau sebanyak 1.675 pengguna, selanjutnya kedua terbanyak memilih nilai 1 dengan persentase 21% atau sebanyak 815 pengguna, untuk yang ketiga dan keempat tertinggi memilih nilai 4 dan 3 pada persentase 12% dengan 475 pengguna dan 454 pengguna, untuk terakhir paling sedikit memilih nilai 2 dengan persentase 11% atau sebanyak 421 pengguna.

### 5.3 Preprocessing Data

Data ulasan *Shopee* yang didapatkan dari proses *scraping* pada situs *google play* belum dapat digunakan karena memiliki bentuk kalimat atau teks tidak berstruktur dimana memiliki banyak *noise* , terdapat ulasan berbahasa asing

(Inggris) yang belum diterjemahkan ke dalam bahasa Indonesia dimana informasi yang didapatkan akan sangat susah diterjemahkan, maka dari itu data tersebut perlu melakukan proses pembersihan terlebih dahulu. Adapun langkah-langkah prapemrosesan yang akan digunakan adalah :

1. *Translate* ulasan bahasa Asing
2. *Spelling Normalization*
3. *Case Folding*
4. *Tokenizing*
5. *Filtering*

### 5.3.1 *Translate* ulasan bahasa asing

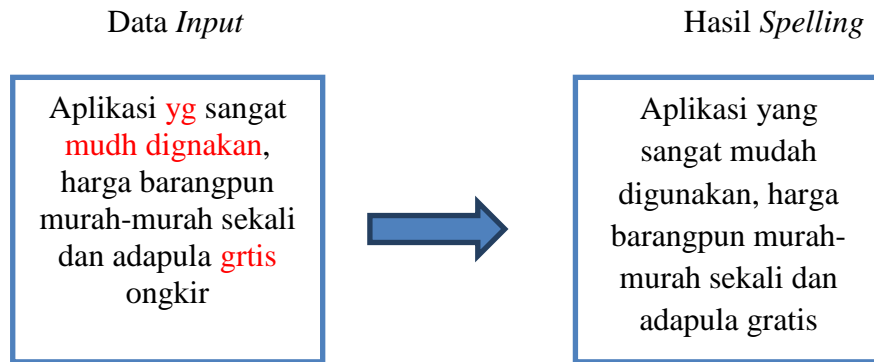
Tahap *translate* merupakan tahap dilakukannya pengubahan ulasan yang berbahasa Inggris maupun ulasan campuran (Inggris dan Indonesia) akan di ubah menjadi bahasa Indonesia, hal ini dikarenakan ada beberapa ulasan dalam bahasa Inggris yang dianggap penting. Adapun proses *translate* dapat dilihat pada **Gambar 5.7** berikut:

<i>Input</i>	<i>Output</i>
<i>Still cant load all items I searched for, omg really buy item cant see anything</i>	masih tidak bisa membuka semua barang yang saya cari, benar-benar tidak bisa melihat barang yang saya mau beli

**Gambar 5.7** Proses *Translate*

### 5.3.2 *Spelling Normalization*

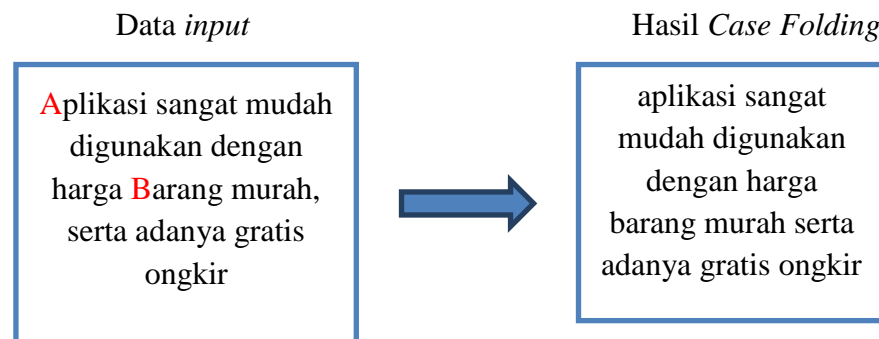
Pada bagian *spelling normalization* berguna untuk melakukan perbaikan kata-kata yang disingkat maupun salah ejaan dengan bentuk tertentu dengan maksud yang sama, sebagai contoh pada kata “tidak” memiliki banyak bentuk penulisan yaitu tak, enggak, tdak, tdk, dan sebagainya, untuk salah ejaan sebagai contoh kata “tersedia” biasa salah penulisan dengan trsedia, trsdia, tersdia dan lain sebagainya. Adapun contoh dalam kalimat ulasan **Gambar 5.8** :



**Gambar 5.8** Proses *Spelling*

### 5.3.3 *Case Folding*

*Case folding* merupakan proses untuk penyeragaman bentuk huruf atau kata kedalam bentuk huruf kecil. Selain itu *case folding* juga bisa untuk menghilangkan tanda baca dan menghapus karakter spasi yang berlebihan. Pada proses ini berguna agar huruf besar dan huruf kecil tidak terdeteksi memiliki perbedaan arti. Adapun sebagai contoh kalimat ulasan pada **Gambar 5.9** berikut.

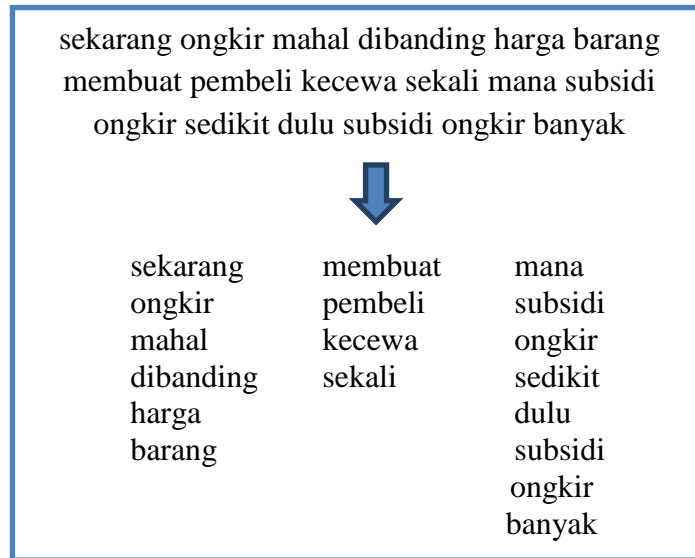


**Gambar 5.9** Proses *Case Folding*

### 5.3.4 *Tokenizing*

*Tokenizing* atau tokenisasi merupakan proses memisahkan teks dalam dokumen menjadi potongan kata yang tidak saling berpengaruh atau independen yang disebut dengan token. *Tokenizing* digunakan untuk mendapatkan potongan kata atau token yang akan menjadi entitas yang memiliki nilai dalam penyusunan matriks dokumen pada proses selanjutnya. Pada proses ini memudahkan proses perhitungan keberadaan kata tersebut dalam dokumen maupun dalam menghitung

frekuensi kemunculan kata tersebut dalam *corpus*. Adapun sebagai contoh kalimat ulasan pada **Gambar 5.10** berikut.



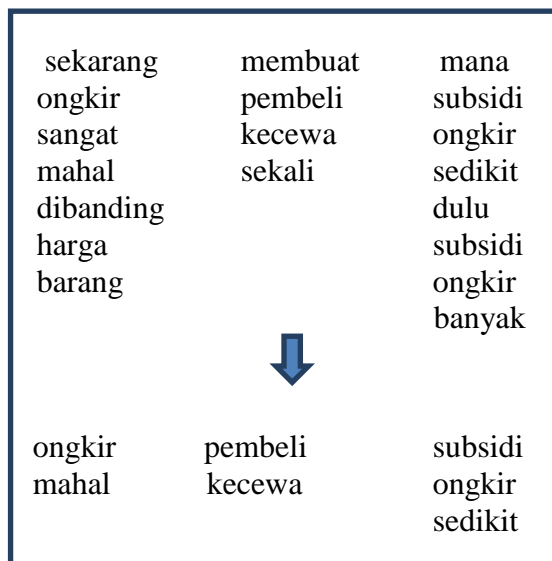
**Gambar 5.10** Proses *Tokenizing*

Setelah tokenisasi selesai, dokumen akan dilanjutkan ke tahap *stemming* yaitu proses merubah kata yang telah ditokenisasi menjadi kata dasarnya. Namun tahap *stemming* tidak sering digunakan karena mengakibatkan kerancuan dan menjadi tidak spesifik dalam merepresentasikan arti yang sebenarnya dari kata hasil *stemming*. Oleh sebab itu, pada penelitian ini juga tidak akan menggunakan tahap *stemming*.

### 5.3.5 Filtering

Tahap *filtering* merupakan tahap dilakukannya penghilangan kata di dalam dokumen atau pengurangan dimensi kata di dalam *corpus* yang disebut *stopwords*. Pada penelitian ini, penulis menggunakan *stopword* berbahasa Indonesia yang telah disusun oleh (Bouge, Kevin., 2011) dan dengan dilakukan secara manual karena ada beberapa kata yang tidak masuk dalam *stopword*. *Stopwords* digunakan untuk menghilangkan kata-kata yang tidak ada berpengaruh atau tidak mengurangi informasi didalam dokumen tersebut namun keberadaannya sering muncul. Kata-kata tersebut seperti kata ganti orang, kata penghubung, kata seruan

dan kata-kata lainnya. Pada penelitian ini kata-kata *stopwords* akan didata didalam *stoplist* (kata yang kurang penting). Contoh isi *stoplist* dalam bahasa indonesia berdasarkan data penelitian yaitu “saya”, “kamu”, “itu”, “untuk”, “tolong”, “mohon”, “segera”, “sangat” dan lain sebagainya. Adapun proses *filtering* dapat dilihat pada **Gambar 5.11** berikut:



**Gambar 5.11** Proses *Filtering*

#### 5.4 Pelabelan Kelas Sentimen

Setelah melakukan proses *preprocessing*, maka dilanjutkan dengan melakukan pelabelan kelas sentimen. Pada bagian ini juga merupakan salah satu proses untuk mendapatkan hasil representasi *corpus* yang diharapkan. Proses pelabelan dilakukan secara otomatis dengan cara menghitung nilai pelabelan sentimen menggunakan kamus *lexicon* dan manual.

Pada dasarnya, proses pelabelan dibagi menjadi tiga kelas sentimen, yaitu sentimen positif, sentimen negatif dan sentimen netral dengan cara melakukan skoring. Penilaian dokumen masuk kategori kelas segmentasi positif atau negatif ditentukan dengan memanfaatkan kumpulan kata dengan bahasa Indonesia yang terdiri dari kumpulan kata-kata positif dan kumpulan kata-kata negatif. Berdasarkan kumpulan dengan kata bahasa Indonesia tersebut kemudian akan dilakukan pelabelan otomatis oleh *aplikasi R* dengan cara menghitung skor jumlah

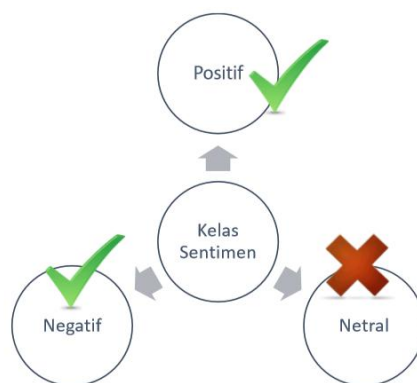
kata positif dikurangi dengan skor jumlah kata negatif dalam suatu kalimat ulasan. Jika suatu kalimat memiliki skor  $> 0$  akan diklasifikasikan dalam kelas positif, jika kalimat memiliki skor  $= 0$  akan diklasifikasikan dalam kelas netral, sedangkan jika kalimat memiliki skor  $< 0$  diklasifikasikan dalam kelas negatif. Berikut merupakan hasil perbandingan jumlah data dari pelabelan kelas sentimen yang diperoleh:

**Tabel 5.1** Perbandingan jumlah data pada kelas sentimen

Sentimen Sementara	Jumlah Ulasan
Positif	1.715
Negatif	1.867
Netral	258

Klasifikasi data pada penelitian ini dibagi menjadi sentimen positif, sentimen negatif dan sentimen netral, akan tetapi data yang digunakan hanya menggunakan data dengan sentimen positif dan sentiment negatif (**Gambar 5.12**). Hal ini dikira untuk kelas sentimen netral kurang memberikan masukan dan manfaat bagi pihak *Shopee*. Pada ulasan klasifikasian yang mengandung pernyataan positif seperti membanggakan, ungkapan terimakasih maupun kata-kata pujian dan lain sebagainya dimasukkan kedalam kategori sentimen positif. Untuk ulasan klasifikasikan yang mengandung pernyataan negatif seperti penghinaan, ketidakpuasan dan sebagainya dimasukkan kedalam kategori sentimen negatif. Untuk yang terakhir, pada ulasan yang tidak mengandung kata pernyataan positif ataupun negatif serta terdapat ulasan yang mengandung kategori negatif dan kategori positif yang seimbang maka masuk dalam ketegori sentimen netral seperti pertanyaan tanpa sentimen, iklan dan sebagainya.

Pada penelitian ini dilakukan reduksi kelas dengan mengkategorikan kelas sentimen netral ke dalam kelas sentiment positif atau negatif yang dilakukan secara manual



**Gambar 5.12** *Pembagian Kelas Sentimen*

Apabila kelas sentimen netral tidak teridentifikasi kata sentiment positif ataupun negatif maka akan masuk dalam pengklasifikasian kelas sentimen positif dan apabila sentiment netral terdapat kata sentiment positif dan negatif yang seimbang maka akan masuk dalam pengklasifikasian kelas sentimen negatif. Sebelumnya hal ini telah dilakukan dengan pertimbangan bahwa informasi negatif dapat diekstraksi dengan lebih mudah untuk diterjemahkan sebagai keluhan atau ketidakpuasan pengguna, untuk itu diharapkan untuk pihak *Shopee* dapat melakukan perbaikan kearah yang lebih baik tentunya. Adapun hasil pelabelan data ulasan sebagai berikut.

**Tabel 5.2** Hasil pelabelan menggunakan kamus *lexicon* dan proses manual

Kelas Sentimen	Skor	Ulasan
Positif	4	Aplikasi <i>mudah</i> digunakan harga barang <i>murah</i> kualitas <i>bagus</i> tersedia <i>gratis</i> ongkir
Negatif	-4	Aplikasi <i>lambat</i> gambar sering <i>hilang</i> pelayanan cs <i>buruk</i> pengembalian dana <i>lama</i>

#### 5.4.1 Simulasi Perhitungan Skor Sentimen

Berdasarkan teks ulasan “aplikasi lambat digunakan harga barang murah status pengiriman lambat update, terdapat 2 kata positif dan 1 kata negatif yang

terdeteksi pada kamus *lexicon*, yaitu “*mudah*” dan “*murah*” sebagai kata positif, untuk kata negatif yaitu “*lambat*” sebagai kata negatif. Adapun rumus yang digunakan dalam proses perhitungan skor sentimen adalah sebagai berikut:

$$\text{Skor} = (\text{Jumlah kata positif}) - (\text{Jumlah kata negatif}) \quad (5.1)$$

**Tabel 5.3** Simulasi perhitungan skor sentimen

Teks ulasan	Kata positif	Kata negatif
aplikasi mudah digunakan harga barang murah status pengiriman lambat update	<b>mudah</b> <b>murah</b>	<b>lambat</b>
<b>Jumlah</b>	<b>2</b>	<b>1</b>

Sehingga didapatkan hasil perhitungan sebagai berikut :

$$\text{Skor} = \text{Jumlah kata positif} - \text{Jumlah kata negatif}$$

$$\text{Skor} = 2 - 1$$

$$\text{Skor} = 1$$

Skor akhir yang didapatkan dari simulasi perhitungan bernilai 1 atau  $> 0$ , maka hasil klasifikasi ulasan adalah positif.

Untuk hasil pelabelan kelas sentimen positif dan kelas sentimen negatif dapat dilihat sebagai berikut :

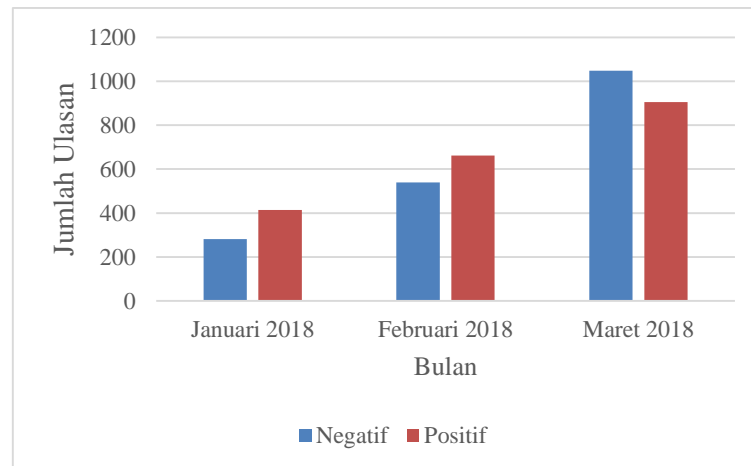
**Tabel 5.4** Jumlah ulasan pada kelas sentimen

Kelas Sentimen	Jumlah Ulasan
<b>Positif</b>	1.782
<b>Negatif</b>	2.058
<b>Total</b>	<b>3.840</b>

Berdasarkan **Tabel 5.4**, diperoleh hasil pelabelan kelas sentimen dengan jumlah ulasan negatif memiliki frekuensi lebih banyak dibandingkan jumlah ulasan positif. Dari total 3.840 ulasan, jumlah ulasan positif adalah sebanyak



1.781 ulasan, dan ulasan negatif adalah sebanyak 2.059 ulasan. Berikut adalah grafik perbandingan jumlah kelas sentimen berdasarkan waktu :



**Gambar 5.13** Grafik perbandingan jumlah kelas sentimen berdasarkan urutan waktu

Berdasarkan **Gambar 5.13** diketahui bahwa pada bulan Januari 2018 dan Februari 2018 sentimen positif lebih tinggi dibandingkan sentiment negatif hal ini menunjukkan bahwa pengguna memiliki persepsi yang baik terhadap *Shopee* pada bulan-bulan tersebut, akan tetapi pada bulan Maret 2018 sentimen negatif lebih tinggi dibandingkan sentimen positif, hal ini bisa jadi karena berkaitan dengan sistem *Shopee* yang telah berubah.

## 5.5 Pembuatan Data Latih dan Data Uji

Data latih digunakan oleh algoritma klasifikasi untuk membentuk sebuah model *classifier*, model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada, semakin besar data latih yang digunakan, maka akan semakin baik *machine* dalam memahami pola data. Data uji digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Data yang digunakan untuk data latih dan data uji adalah data yang telah memiliki label kelas, dengan jumlah data latih dan data uji memiliki perbandingan 80% : 20%. Berdasarkan *Pareto Principle*, Rasio

yang umum digunakan adalah 80:20 untuk *data sets training* dan *testing*. Perbandingan jumlah data latih dan data uji dapat dilihat pada **Tabel 5.5** berikut :

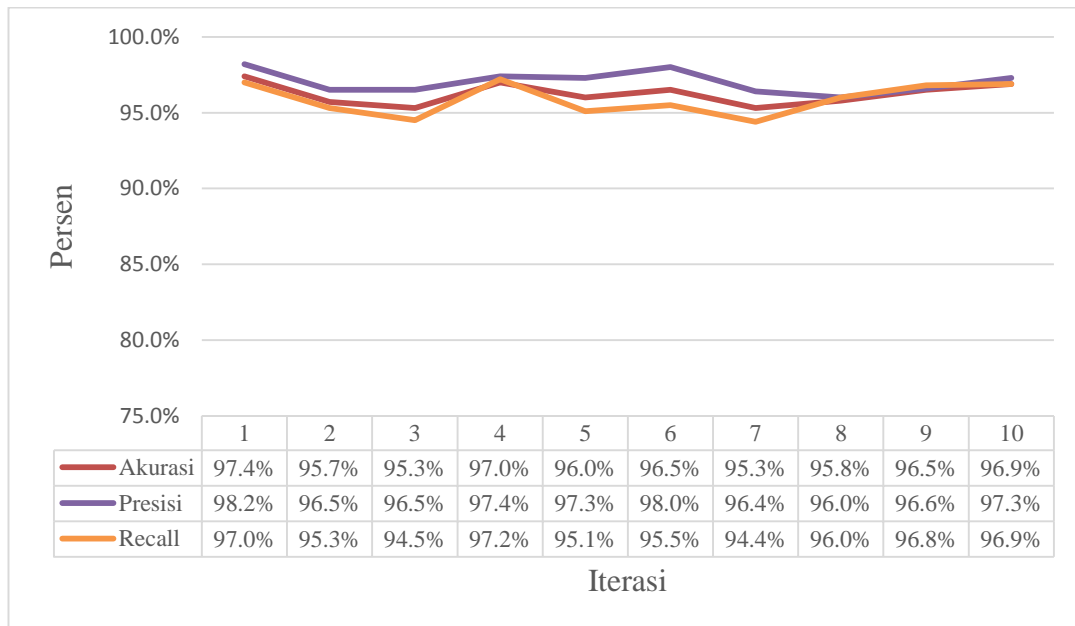
**Tabel 5.5** Perbandingan data latih dan data uji

Klasifikasi	Jumlah	Data Latih (80%)	Data uji (20%)
Positif	1.782	1.424,8 $\approx$ 1.425	356,2 $\approx$ 357
Negatif	2.058	1.647,2 $\approx$ 1.647	411,8 $\approx$ 411
<b>Total</b>	<b>3.840</b>	<b>3.073</b>	<b>768</b>

Berdasarkan **Tabel 5.5** dengan perbandingan data latih dan data uji sebesar 80% : 20%, dari total 3.840 data ulasan berbahasa Indonesia tanpa pembuangan data kembar, digunakan sebanyak 3.072 data sebagai data latih dan 768 data sebagai data uji.

## 5.6 Klasifikasi dengan Metode *Naïve Bayes Classifier*

Penelitian ini menggunakan metode *confusion matrix* dalam proses evaluasi untuk mengetahui hasil akurasi klasifikasi. *Confusion matrix* merupakan salah satu *tools* penting dalam metode evaluasi yang digunakan pada *machine learning* yang biasanya memuat dua kategori atau lebih (Manning, Prabhakar dan Hinrich, 2009). Setiap unsur matiks menunjukkan jumlah contoh data uji untuk kelas sebenarnya yang digambarkan dalam bentuk baris sedangkan kolom menggambarkan kelas yang diprediksi. Pada proses klasifikasi dilakukan dengan cara membuat *machine learning* menggunakan data latih dan data uji secara acak dengan percobaan dilakukan iterasi sebanyak 10 kali pada *dataset* sebagai *cross validation* untuk mendapatkan nilai akurasi prediksi terbaik. Berikut grafik nilai akurasi dari 10 kali iterasi :



**Gambar 5.14** Grafik nilai akurasi, presisi, dan *Recall* dengan metode *Naïve Bayes Classifier*

Dapat dilihat pada **Gambar 5.14** hasil akurasi menggunakan metode *Naïve Bayes Classifier* (NBC) memiliki variasi kenaikan akurasi dengan selisih yang sedikit, dengan nilai rata-rata akurasinya adalah 96,2%. Dengan nilai rata-rata akurasi 96,2% dapat dikatakan bahwa metode *Naïve Bayes Classifier* (NBC) ini baik digunakan dalam pengklasifikasian ulasan dengan teks bahasa Indonesia dengan topik *multiclass* pada *Shopee*.

Dari hasil yang didapatkan dengan melakukan iterasi sebanyak 10 kali menggunakan metode *Naïve Bayes Classifier* (NBC) maka diambil model dengan klasifikasi nilai akurasi, presisi dan *recall* tertinggi yaitu pada iterasi 1 dengan nilai masing-masing sebesar 97,4%, 98,2% dan 97%. Selain akurasi, *recall*, dan presisi, *confusion matrix* juga bisa digunakan untuk melakukan evaluasi kinerja dari *classifier*. Untuk perhitungan mencari tingkat akurasi ini dari jumlah data uji yang terklasifikasi dengan benar dibandingkan dengan total semua data yang di uji.

Disini penulis menggunakan metode *confusion matrix* dalam proses evaluasi mencari nilai akurasi, nilai presisi dan nilai *recall*. *Confusion matrix* merupakan salah satu *tools* penting dalam metode visualisasi yang digunakan

pada mesin pembelajaran yang biasanya memuat dua kategori atau lebih (Manning, Prabhakar dan Hinrich, 2009). Pada matriks bagian baris merupakan contoh jumlah data uji untuk kelas sebenarnya sedangkan untuk bagian kolom merupakan kelas yang diprediksi. Dapat dilihat pada **Tabel 5.6** merupakan hasil *confusion matrix* prediksi dua kelas sentimen yaitu kelas positif dan kelas negatif dengan nilai akurasi, presisi, dan *recall* terbaik yang didapatkan dari iterasi.

**Tabel 5.6** Hasil *confusion matrix*

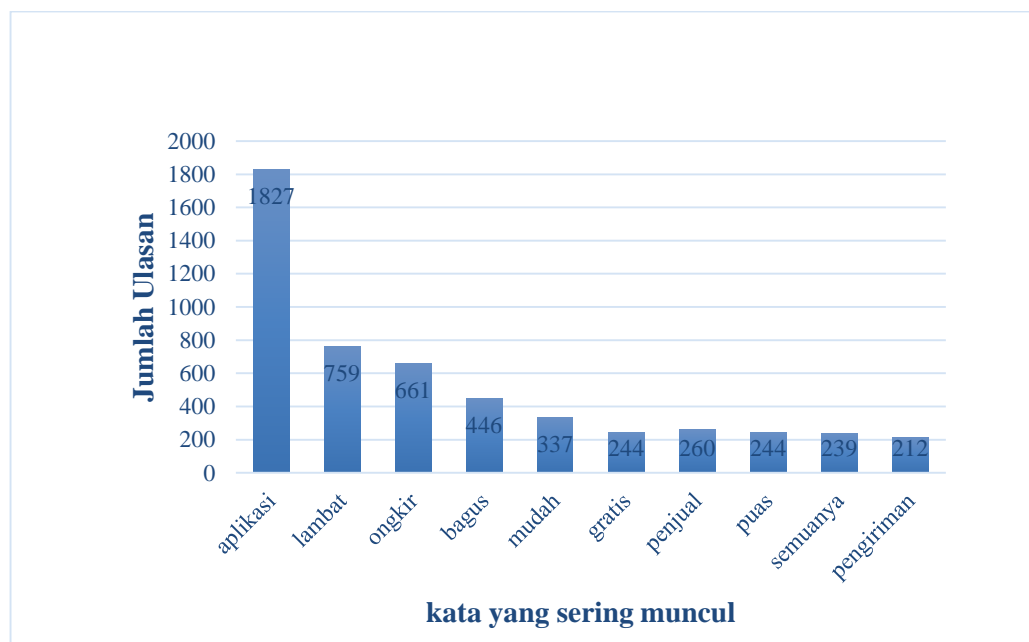
Prediksi	Aktual		Class
	Negatif	Positif	Precision
Negatif	398	7	98,2%
Positif	13	350	96,4%
Class Recall	97,0%	98,0%	
Akurasi			
97,4%			

Dapat dilihat pada **Tabel 5.6** dengan menggunakan metode *Naïve Bayes Classifier* didapatkan hasil prediksi dari 411 ulasan yang masuk dalam kelas negatif terdapat 398 ulasan yang telah terklasifikasi dengan benar dan terdapat kesalahan prediksi sebanyak 13 ulasan yang masuk dalam ulasan positif sehingga didapatkan nilai presisi untuk kelas negatif sebesar 98,2%. Sedangkan dari 357 ulasan yang masuk dalam kelas positif terdapat 350 ulasan terklasifikasi dengan benar dan terdapat kesalahan prediksi sebanyak 7 yang masuk dalam ulasan negatif sehingga didapatkan nilai presisi untuk kelas positif sebesar 96,4%. Selanjutnya dari hasil *confusion matrix* tersebut maka diperoleh tingkat akurasi sebesar 97,4%, yang artinya dari 768 data ulasan yang diuji, terdapat 748 ulasan yang benar diklasifikasikan oleh model *Naïve Bayes Classifier*.

## 5.7 Visualisasi dan Asosiasi

Dalam penelitian ini dilakukan visualisasi pada semua data yang ada baik itu masuk kelas sentimen negatif maupun positif hal ini bertujuan untuk mengekstrasi informasi secara keseluruhan tentang topik atau bahasan yang sering

diulas para pengguna *Shopee*. Selain itu juga dilakukan visualisasi pada masing-masing kelas sentimen untuk diambil informasi yang dianggap penting atau berguna serta dicari asosiasi antar kata pada masing-masing kelas sentiment yang paling sering muncul secara bersamaan guna memperkuat pencarian informasi tersebut. Adapun visualisasi untuk keseluruhan data sebagai berikut:



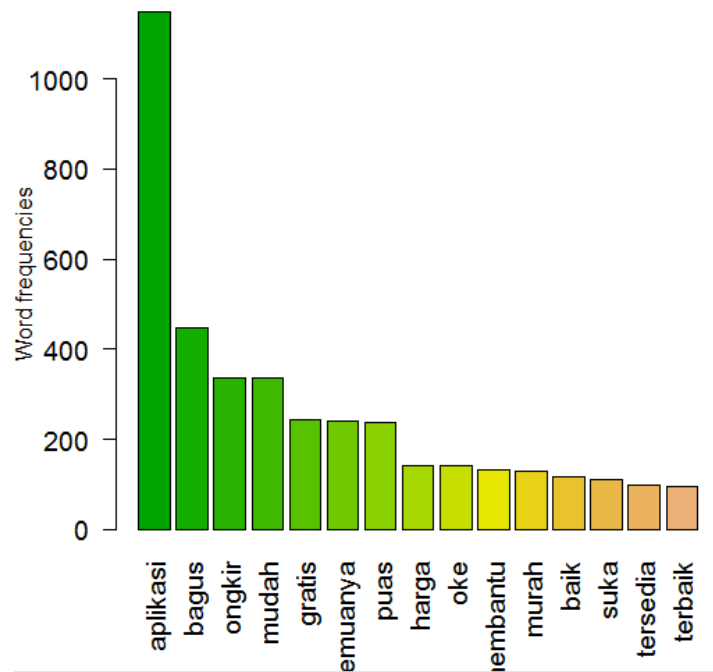
**Gambar 5.15** Kata yang paling banyak muncul pada keseluruhan data

Dapat dilihat pada **Gambar 5.15** pada hasil klasifikasi ulasan keseluruhan data sebanyak 3.840 data didapatkan beberapa kata yang paling banyak muncul yaitu kata “aplikasi” sebanyak 1.827 kali, kata “lambat” sebanyak 759 kali, kata “ongkir” sebanyak 661 kali, kata “bagus” sebanyak 446 kali dan seterusnya.

### 5.7.1 Ulasan Positif

Data ulasan positif yang digunakan adalah data hasil pelabelan yang dilakukan dengan penggunaan kamus *lexicon* dan secara manual. Ekstraksi informasi pada ulasan positif dilakukan secara berulang-ulang hingga mendapatkan informasi ulasan positif pengguna *Shopee* yang paling sering diulas. Ulasan positif tersebut diidentifikasi berdasarkan frekuensi kata dalam ulasan,

berikut merupakan visualisasi hasil ekstraksi informasi yang didapatkan dari ulasan pengunjung dengan klasifikasi ulasan positif.



**Gambar 5.16** Kata yang paling banyak muncul pada kelas positif

Dapat dilihat pada **Gambar 5.16** pada hasil klasifikasi ulasan positif dari jumlah ulasan sebanyak 1.781 didapatkan beberapa kata yang paling banyak muncul yaitu kata “aplikasi” sebanyak 1.148 kali, kata “bagus” sebanyak 446 kali, kata “ongkir” sebanyak 337 kali, kata “mudah” sebanyak 337 kali, kata “gratis” sebanyak 244 kali dan seterusnya. Kata-kata yang muncul seperti pada **Gambar 5.16** merupakan kata yang memiliki sentimen positif serta merupakan topik pembicaraan atau bahasan positif yang paling banyak diulas oleh pengguna *Shopee*. Kata-kata tersebut selanjutnya digunakan sebagai dasar untuk menemukan asosiasi dengan kata lainnya, untuk mendapatkan informasi yang lebih baik. Kumpulan kata-kata yang sering muncul tersebut juga dapat ditampilkan dalam bentuk *wordcloud* seperti terlihat pada **Gambar 5.17**.



**Gambar 5.17** Wordcloud ulasan positif

Berdasarkan **Gambar 5.17** visualisasi *wordcloud* dapat di lihat dengan lebih jelas kata-kata positif yang sering digunakan pengguna dalam memberikan ulasan. Semakin besar ukuran kata pada *wordcloud* maka semakin tinggi pula frekuensi kata tersebut, artinya pengguna sering menggunakan kata tersebut sebagai topik pembicaraan dengan penilaian positif dalam ulasan. Selanjutnya, dilakukan pencarian asosiasi antar kata yang berhubungan dengan kata yang sering muncul secara bersamaan dan didapatkan hasil sebagai berikut :

**Tabel 5.7** Asosiasi kata pada kelas sentimen positif

aplikasi		ongkir		Harga		Puas	
mudah	0,48	gratis	0,79	Murah	0,65	Semua	0,99
bagus	0,47			Bersahabat	0,30		
membantu	0,31			terjangkau	0,22		
simpel	0,28						

Berdasarkan hasil pada **Tabel 5.7**, diperoleh beberapa asosiasi kata pada klasifikasi kelas positif yaitu, jika dilihat asosiasi kata yang berkaitan dengan kata “aplikasi”, maka didapatkan informasi tentang penggunaan aplikasi itu dengan mudah, bagus secara keseluruhan tampilan aplikasi, membantu dalam berbelanja

secara *online* dan simpel atau tidak rumit dalam menggunakan aplikasi tersebut dalam semua hal seperti melakukan belanja, dan lainnya.

Kata-kata yang berasosiasi dengan kata “ongkir” juga memberikan informasi bagaimana *Shopee* sendiri memberikan keringanan pada pembeli dalam berbelanja dengan biaya ongkir digratiskan.

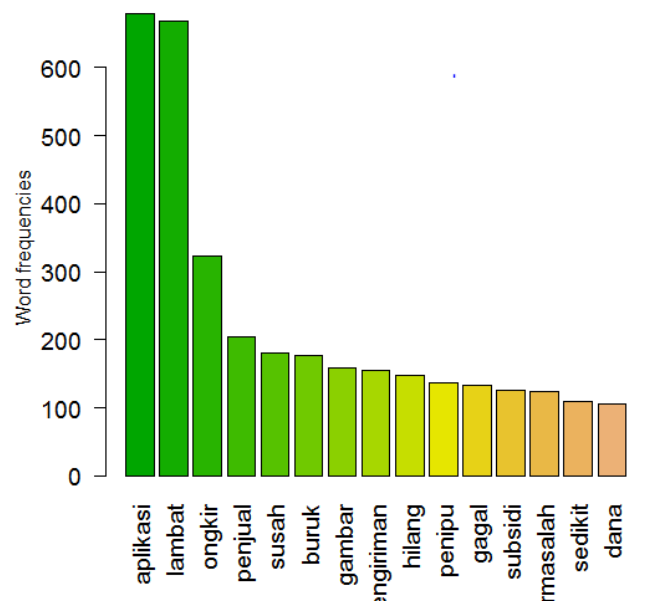
Kata-kata yang berasosiasi dengan kata “Harga” memberikan informasi bahwasannya harga barang yang dijual didalam *shopee* rata-rata semuanya murah, terjangkau dan bersahabat dengan artian pembeli bisa membeli barang yang mereka inginkan tanpa mengeluarkan biaya yang besar.

Kata-kata yang berasosiasi dengan kata “puas” memberikan informasi tentang bagaimana pengguna *Shopee* merasa puas akan seluruh aspek yang ada didalam sistem *shopee*, mulai dari aplikasi, gratis ongkir, berbelanja mudah dan sebagainya.

### 5.7.2 Ulasan Negatif

Ekstraksi informasi pada ulasan negatif dilakukan secara berulang-ulang hingga mendapatkan informasi ulasan negatif pengguna *Shopee* yang paling sering diulas. Berdasarkan hasil pelabelan, ulasan negatif pada pengguna *Shopee* lebih banyak dibandingkan dengan jumlah ulasan positif, dari total ulasan sebanyak 3.840 terdapat 2.059 ulasan negatif. Hal ini menunjukkan bahwa banyak pengguna *Shopee* mempunyai persepsi yang kurang baik. Berikut adalah visualisasi hasil ekstraksi informasi yang didapatkan dari ulasan pengguna dengan klasifikasi ulasan negatif.





**Gambar 5.18** Kata yang paling banyak muncul pada kelas negatif

Dapat dilihat pada **Gambar 5.18** kata-kata negatif yang sering digunakan pengunjung dalam memberikan ulasan diantaranya adalah kata “aplikasi” sebanyak 679 kali, kata “lambat” sebanyak 669 kali, kata “ongkir” sebanyak 324 kali, kata “penjual” sebanyak 205 kali dan seterusnya. Kata-kata yang muncul seperti pada **Gambar 5.18** merupakan kata yang memiliki sentimen negatif serta merupakan topik pembicaraan atau bahasan negatif yang paling banyak diulas oleh pengguna *Shopee*. Kata-kata tersebut selanjutnya digunakan sebagai dasar untuk menemukan asosiasi dengan kata lainnya, untuk mendapatkan informasi yang lebih baik. Kumpulan kata-kata yang sering muncul tersebut juga dapat ditampilkan dalam bentuk *wordcloud* seperti terlihat pada **Gambar 5.19** berikut.



**Gambar 5.19** *Wordcloud* ulasan negatif

Berdasarkan **Gambar 5.19** visualisasi *wordcloud* dapat di lihat dengan lebih jelas kata-kata negatif yang sering digunakan pengguna dalam memberikan ulasan. Semakin besar ukuran kata pada *wordcloud* maka semakin tinggi pula frekuensi kata tersebut, artinya pengguna sering menggunakan kata tersebut sebagai topik pembicaraan dengan penilaian negatif dalam ulasan. Selanjutnya, dilakukan pencarian asosiasi antar kata yang berhubungan dengan kata yang sering muncul secara bersamaan dan didapatkan hasil sebagai berikut :

**Tabel 5.8** Asosiasi kata pada kelas sentimen negatif

aplikasi		ongkir		Buruk		Susah	
lambat	0,56	mahal	0,64	pelayanan	0,46	Pengembalian	0,16
bermasalah	0,19	sedikit	0,57	komplain	0,16	Puzzle	0,26
berat	0,16	berbayar	0,50			Penjualan	0,19
logout	0,17	dibatas	0,19			upload	0,32
gambar		pengiriman		penjual			
hilang	0,49	lambat	0,28	penipu	0,70		
		lambatupdate	0,40	rugi	0,26		
				seleksi	0,18		

**Tabel 5.8** menunjukkan asosiasi antar kata pada ulasan negatif, kata-kata tersebut merupakan topik yang paling sering dibicarakan pengguna dalam ulasannya. Berdasarkan tabel tersebut dapat diperoleh beberapa informasi berikut.

Kata-kata yang berasosiasi dengan kata “aplikasi” pada ulasan negatif memberikan informasi tentang keluhan pengguna terhadap aplikasi *Shopee* yang berat karena aplikasi ini memerlukan memori yang besar, aplikasi lambat jika dibuka meskipun internet lancar dirasa karena masih berhubungan dengan memori tadi, serta aplikasi bermasalah dan sering *logout* dirasa karena terdapat masalah dari sistem aplikasi *Shopee* itu sendiri.

Kata-kata yang berasosiasi dengan kata “ongkir” pada ulasan negatif memberikan informasi tentang keluhan pengguna terhadap mahalnnya ongkos kirim jika berbelanja di *Shopee*, sedikitnya subsidi ongkir yang diberikan, iklan *Shopee* yang menyatakan gratis ongkir ternyata hanya bohongan karena kenyataannya pembeli masih harus membayar biaya dari pengiriman barang yang mereka beli, serta pembelanjaan yang dibatasi untuk mendapatkan subsidi ongkos kirim.

Kata-kata yang berasosiasi dengan kata “susah” pada ulasan negatif memberikan informasi tentang pengajuan pengembalian dana yang susah oleh pihak *Shopee*, mengadakan promo *daily prize* dengan sistem harus mengumpulkan semua *puzzle* baru bisa mendapatkan hadiah, akan tetapi pada kenyataannya banyak pengguna mengeluh karena terdapat salah satu *puzzle* yang tidak pernah muncul atau susahnya mendapatkan *puzzle* itu, selanjutnya juga terdapat keluhan penjual tentang sistem *Shopee* memberatkan dan banyak aturan dalam berjualan, serta untuk penjual susah dalam melakukan *upload* gambar barang meskipun peletakan gambar sudah benar.

Kata-kata yang berasosiasi dengan kata “gambar” pada ulasan negatif memberikan informasi tentang ketika membuka *Shopee* gambar tidak bisa dilihat atau hilang meskipun koneksi pada jaringan lancar.

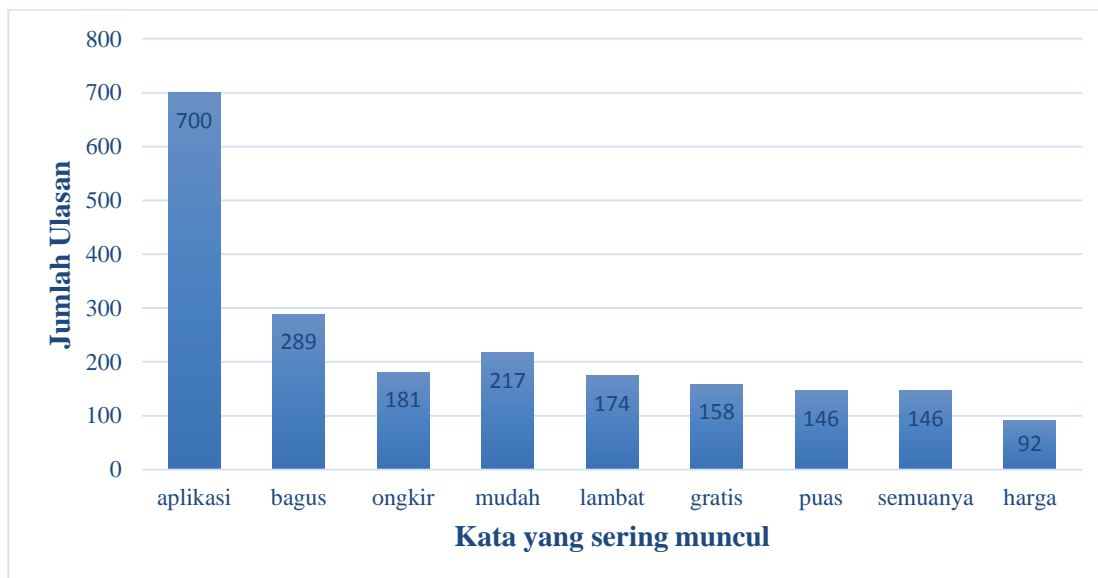
Kata-kata yang berasosiasi dengan kata “pengiriman” memberikan informasi tentang keluhan pengguna tentang masalah pengiriman barang yang lambat, karena harus melakukan beberapa tahap dalam proses sebelum melakukan

pengiriman serta pada sistem *Shopee* pada status pengiriman yang ada tidak *update*, misal seperti barang sudah diterima oleh pembeli akan tetapi status pengiriman masih dalam keadaan pengemasan.

Kata-kata yang berasosiasi dengan kata “penjual” memberikan informasi tentang keluhan pembeli tentang banyaknya penipuan yang mereka alami, pihak *Shopee* tidak ada melakukan seleksi penjual dengan ketat serta ada juga dari penjual menyatakan mereka rugi karena dalam proses pencairan dana oleh pihak *Shopee* diperlama.

#### 5.4.2 Ulasan Bulanan

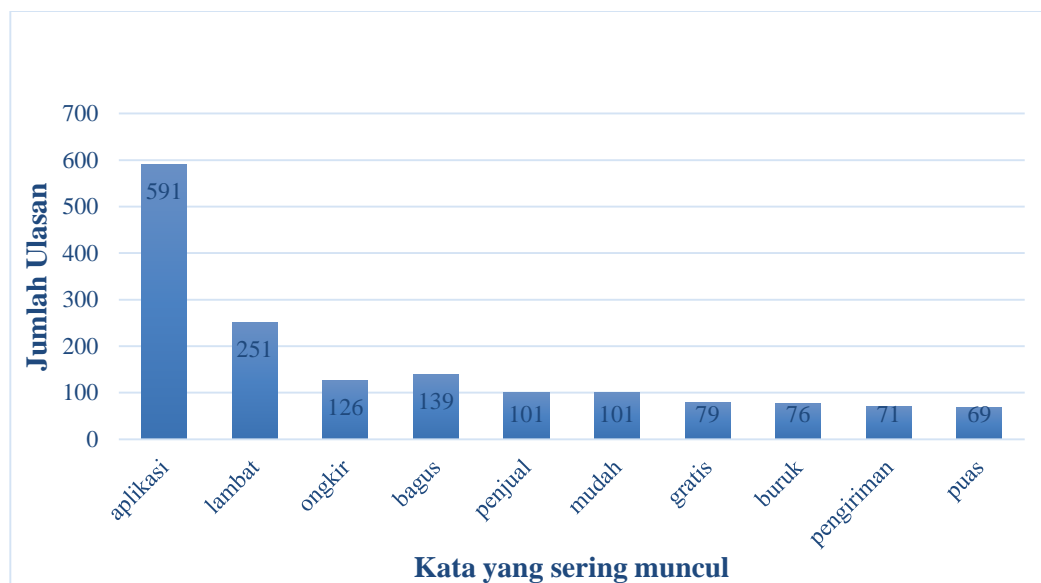
Ekstraksi informasi pada ulasan bulanan dilakukan untuk mengetahui informasi semua ulasan pada bulan Januari, hal ini bertujuan untuk melihat secara keseluruhan pada bulan Januari pengguna banyak mengulas kata negatif atau positif.



**Gambar 5.20** Kata yang paling banyak muncul pada bulan Januari 2018

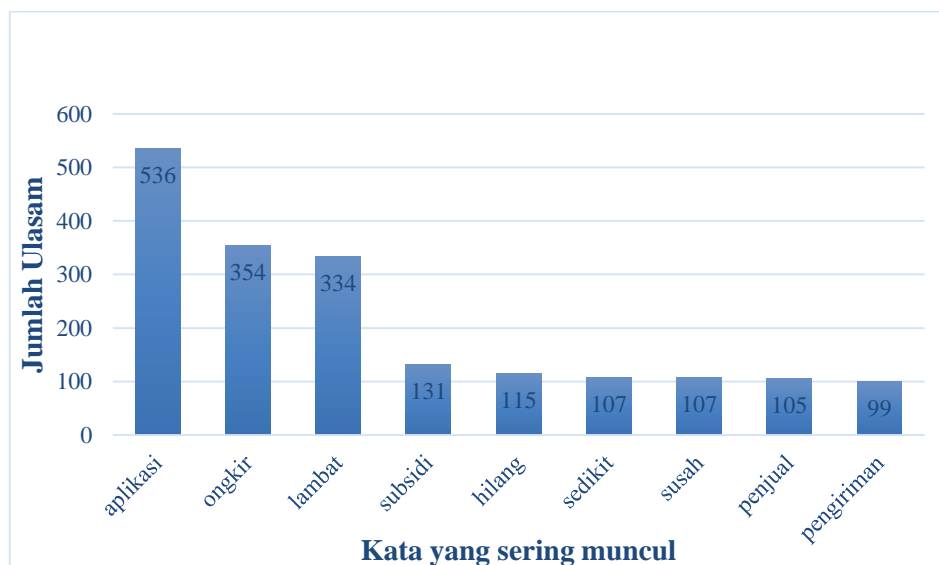
Dapat dilihat pada **Gambar 5.20** pada bulan Januari 2018 kata-kata yang sering digunakan pengguna dalam memberikan ulasan diantaranya kata “aplikasi” sebanyak 700 kali, kata “bagus” sebanyak 289 kali, kata “ongkir” sebanyak 181 kali, kata “mudah” sebanyak 217 kali, kata “lambat” sebanyak 174 kali, dan

seterusnya. Dilihat pada bulan Januari 2018 respon pengguna banyak mengeluarkan kata-kata positif dibanding kata-kata negatif seperti kata “bagus”, “mudah”, “gratis” dan “puas”. Untuk kata negatif yaitu pada kata “lambat”.



**Gambar 5.21** Kata yang paling banyak muncul pada bulan Februari 2018

Dapat dilihat pada **Gambar 5.21** pada bulan Februari 2018 kata-kata yang sering digunakan pengguna dalam memberikan ulasan diantaranya kata “aplikasi” sebanyak 591 kali, kata “lambat” sebanyak 251 kali, kata “ongkir” sebanyak 126 kali, kata “bagus” sebanyak 139 kali, dan seterusnya. Dilihat pada bulan Februari 2018 respon pengguna banyak mengeluarkan kata-kata positif dibanding kata-kata negatif seperti kata “bagus”, “mudah”, “gratis” dan “puas”. Untuk kata negatif yaitu pada kata “lambat” dan “buruk”.

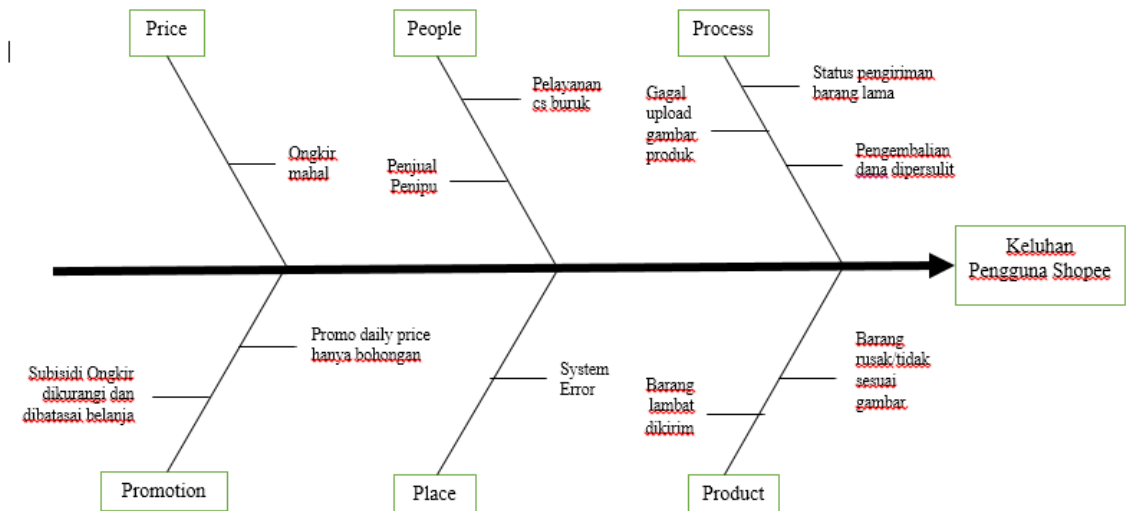


**Gambar 5.22** Kata yang paling banyak muncul pada bulan Maret 2018

Dapat dilihat pada **Gambar 5.22** pada bulan Maret 2018 kata-kata yang sering digunakan pengguna dalam memberikan ulasan diantaranya kata “aplikasi” sebanyak 536 kali, kata “ongkir” sebanyak 354 kali, kata “lambat” sebanyak 334 kali, kata “subsidi” sebanyak 131 kali, kata “hilang” sebanyak 115 kali, dan seterusnya. Dilihat pada bulan Maret 2018 respon pengguna banyak mengeluarkan kata-kata negatif seperti kata “lambat”, “hilang”, “sedikit” dan “susah” dan tidak didapatkan banyaknya kata positif pada bulan Maret 2018.

## 5.8 Diagram *Fishbone*

Berdasarkan hasil ulasan negatif yang didapatkan dari data ulasan, maka dapat diperoleh informasi mengenai masalah yang terjadi terkait dengan ulasan negatif pada *e-commerce* Shopee berdasarkan **Gambar 5.23**.



**Gambar 5.23** Diagram *fishbone* komplain pengguna Shopee

Pada **Gambar 5.23**, dapat diketahui informasi faktor-faktor yang menyebabkan *e-commerce* Shopee memiliki ulasan negatif yaitu dari segi *process*, *product*, *people*, *place*, *price* dan *promotion*. Langkah selanjutnya yaitu menentukan pemecahan masalah yang ada. Adapun rencana pemecahan masalah di Shopee dapat dilihat pada **Tabel 5.9** berikut.

**Tabel 5.9** Rencana pemecahan masalah *e-commerce* Shopee

No	Faktor	Permasalahan	Pemecahan Masalah
1	<i>Process</i>	Status pengiriman barang lama <i>update</i>	Pihak developer segera memperbaiki sistem cek resi agar <i>history</i> pengiriman cepat <i>update</i>
		Gagal <i>upload</i> gambar produk	Pengurangan kategori/ketentuan yang dirasa tidak penting pada sistem <i>upload</i> gambar untuk mempermudah penjual
		Pengembalian dana susah	Pihak <i>Help Center</i> harus lebih cekatan dan tidak merumitkan untuk

			pengembalian dana penjual maupun pembeli
2	<i>Product</i>	Barang rusak/Tidak sesuai gambar	Melakukan penjelasan bahwa pihak Shopee tidak bertanggung jawab atas kerusakan barang atau barang tidak sesuai gambar. Untuk pemecahan masalah barang rusak pembeli bisa meminta pengembalian pesanan dan dana kepada tokonya dengan menghubungi pihak Help Center yang akan menjadi penengah antara penjual dan pembeli dalam pengembalian barang dan pengembalian dana
		Barang lambat dikirim	Melakukan penjelasan bahwa pihak Shopee tidak bertanggung jawab atas pengiriman barang yang lambat, dari pihak shopee sendiri telah memberikan batas waktu untuk penjual mengirimkan barangnya ke ekspedisi terkait. Penyebab lainnya bisa karena pada ekspedisi pengiriman belum mengirimkan barang dari penjual
3	<i>People</i>	Pelayanan buruk	Memberikan penegasan/evaluasi kinerja bagian <i>customer service</i> untuk melayani pertanyaan dan komplain konsumen dengan baik
		Penjual Penipu	Melakukan selektif pemilihan penjual yang jelas dan melakukan filter untuk akun yang tidak aktif lagi dan juga



			melakukan pemblokiran akun yang telah diketahui penipu
4	<i>Place</i>	<i>System Error</i>	Sistem eror biasanya karena aplikasi yang lambat, berat, gambar tidak muncul dan sebagainya. Untuk perbaikan sendiri dengan melakukan filtering/hapus akun yang tidak aktif lagi dan melakukan <i>press</i> (pengecilan ukuran aplikasi yang dianggap terlalu membutuhkan ruang penyimpanan yang besar
5	<i>Price</i>	Ongkir mahal	Pihak Shopee harus melakukan <i>update</i> biaya ongkos kirim sesuai dengan ekspedisi
6	<i>Promotion</i>	Promo <i>daily prize</i> hanya bohongan	Pihak Shopee harus lebih jujur dan konsisten dalam pembagian <i>puzzle</i> yang seimbang pada promo <i>daily prize</i>
		Subsidi ongkir dikurangi dan dibatasi pembelian	Pihak Shopee harus lebih konsisten dalam pemberian subsidi ongkir dengan melakukan pengembalian subsidi awal dalam promonya yang gratis ongkir seluruh Indonesia dan tidak melakukan pembatasan belanja untuk mendapatkan subsidi ongkir

## BAB VI

### PENUTUP

#### 5.1. Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan, diperoleh beberapa kesimpulan sebagai berikut :

1. Gambaran umum mengenai persepsi pengguna *Shopee* pada *website Google Play* berdasarkan *Rating* dari bulan Januari 2018 – Maret 2018 adalah paling banyak memilih nilai 5 dengan persentase 44% atau sebanyak 1.675 pengguna, selanjutnya kedua terbanyak memilih nilai 1 dengan persentase 21% atau sebanyak 815 pengguna, untuk yang ketiga dan keempat tertinggi memilih nilai 4 dan 3 pada persentase 12% dengan 475 pengguna dan 454 pengguna, untuk terakhir paling sedikit memilih nilai 2 dengan persentase 11% atau sebanyak 421 pengguna.
2. Hasil dari penerapan metode *Naïve Bayes Classifier* dalam mengklasifikasikan data ulasan pengguna *Shopee* menjadi kelas positif dan negatif dengan perbandingan data latih dan data uji sebesar 80% : 20% diperoleh hasil klasifikasi sentimen dengan tingkat akurasi sebesar 97,4%.
3. Informasi yang didapatkan dari hasil klasifikasi dan asosiasi teks yang dilakukan, secara umum dapat diketahui bahwa pengguna *Shopee* banyak membicarakan mengenai aplikasi, lambat, ongkir, bagus dan mudah untuk 5 kata yang teratas dalam keseluruhan ulasan. Pada kelas sentimen *positif* pengguna *Shopee* banyak membicarakan mengenai aplikasi, bagus, mudah, murah dan gratis untuk 5 kata teratas dengan ekstrasi informasi yang didapatkan dari asosiasi positif diantaranya yaitu pada kata aplikasi, ongkir, harga dan puas. Pada kelas sentimen negatif pengguna *Shopee* banyak membicarakan mengenai aplikasi, lambat, ongkir, penjual dan susah untuk 5 kata teratas dengan ekstrasi informasi yang didapatkan dari asosiasi negatif

diantaranya yaitu pada kata aplikasi, ongkir, buruk, susah, gambar, pengiriman, dan penjual.

4. Berdasarkan diagram fishbone (sebab-akibat) terdapat 6 faktor dengan 11 permasalahan yang harus pada aplikasi *e-commerce* Shopee dimana untuk faktor-faktornya yaitu *Process, Product, People, Place, Price* dan *Promotion*.

## 5.2. Saran

Berdasarkan kesimpulan di atas, dapat diberikan beberapa saran sebagai berikut.

1. Bagi pihak Shopee, hasil ekstraksi informasi dari ulasan-ulasan yang telah diberikan oleh pengguna khususnya ulasan yang berbentuk negatif dapat dijadikan bahan evaluasi dalam peningkatan kepuasan pengguna dan memberikan pelayanan semaksimal mungkin, serta untuk pengembangan pembaharuan aplikasi serta bisnis selanjutnya.
2. Penanganan negasi belum menjadi fokus utama dalam analisis sentimen pada penelitian ini, sehingga kalimat yang memiliki kata negasi belum dapat ditentukan polaritasnya secara optimal. Untuk penelitian selanjutnya diharapkan dapat melakukan penanganan khusus terhadap kata negasi agar hasil yang didapatkan lebih akurat.
3. Dalam penelitian ini data ulasan yang digunakan masih dibatasi untuk ulasan yang berbahasa Indonesia saja, sehingga perlu dikembangkan pada penelitian selanjutnya dengan menggunakan ulasan berbagai bahasa.
4. Bagi peneliti selanjutnya, dapat menggunakan pendekatan *machine learning* lain sebagai pembanding performa algoritma *Naïve Bayes Classifier* untuk mengklasifikasikan ulasan aplikasi *e-commerce* Shopee pada situs *Google Play*.

## DAFTAR PUSTAKA

- APJII. 2017. .Penetrasi & Perilaku Pengguna Internet Indonesia. <https://apjii.or.id/survei2017>. Diakses pada tanggal 10 Maret 2018 pukul 19:20 WIB
- Arifin, Oki. 2016. *Penentuan Prioritas Pemasangan Internet Untuk Pelanggan Baru Perusahaan Menggunakan Naive Bayes (Studi Kasus : Pt. Time Excelindo)*. Tesis. Program Studi Magister Ilmu Komputer FMIPA UGM Yogyakarta.
- Bishop. C. M., 2005. *Neural Network for Pattern Recognition*. United States: Oxford University Press.
- Deepa, V. K. dan Geetha, J. R. R. 2013, Rapid Development of Applications in Data Mining, International Conference on Green High Performance Computing, India.
- Fadlisyah, B. D. A. 2014. *Statistika : Terapannya di Informatika, 1st ed.* Yogyakarta: Graha Ilmu.
- Fayyad, U., Piatetsky-Shapiro, G., dan Smyth, P. 1996, From Data Mining to Knowledge Discovery in Databases, American Association for Artificial Intelligence, pp. 37–54.
- Hamzah, A., 2012, *Klasifikasi Teks Dengan Naïve Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita dan Akademis*, ISSN:1979-9111, vol. 3, 2012.
- Han, J. and Kamber, M. 2006. *Data Mining : Concepts and Techniques Second Edition*. San Francisco : Morgan Kauffman.
- Hayatin, N., Mentari, M., & Izzah, A., (2014). *Opinion Extraction of Public Figure based on Sentiment Analysis in Twitter*, IPTEK, Journal of Engineering ITS, vol. 1, no. 1, hal. 9-14, Surabaya.

- Iprice. 2017. .Kilas Balik Persaingan E-Commerce Indonesia Tahun 2017. <https://iprice.co.id/trend/insights/kilas-balik-e-commerce-indonesia-2017/>. Diakses pada tanggal 13 Maret 2018 pukul 16.54 WIB
- Kristiyanti, D.A.2015. *Analisis Sentimen Review Produk Kosmetik Melalui Komparasi Feature Selection*. Bekasi. ISBN 978-602-72850-0-2.74.
- Manning, C. D., Prabhakar, R., dan Hinrich, S. 2009. *An Introduction to Information Retrieval – Online Edition*. Cambridge: Cambridge University Press.
- Margaretha, Fitri.2017. *Analisis Hubungan Antara Motif dengan Tingkat Kepuasan Pengguna Aplikasi Shopee sebagai Media Berbelanja Online pada Shopeeholics di Kota Samarinda*. eJournal Ilmu Komunikasi.
- Marres, R., Joan, V.,F, dan Wilson, P., G. 2013. *Document-level Sentiment Classification: An Empirical Comparison between SVM and ANN*
- Mulajati, Muhammad. 2017. *Implementasi Teknik Web Scraping dan Klasifikasi Sentimen Menggunakan Metode Naïve Bayes Classifier dan Asosiasi Teks (Studi Kasus : Data Ulasan Penumpang Maskapai Penerbangan Garuda Indonesia Pada Situs TripAdvisor*. Skripsi. Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam UII Yogyakarta
- O’keefe, T. Koprinska, I. 2009. *Feature selection and weighting methods in sentiment analysis*. In: *Proceedings of the 14th Australasian document computing symposium*. Sydney, pp 67–74.
- Pramudiono, 2007.Text Mining Infrastructure in R, Journal of Statistical Software, 25(5), 1-54.
- Prasetyo, E., 2012. *Data Mining Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Andi.
- Purwanto, D.D., dan Santoso, J., 2015, *Multinomial Naïve Bayes Classifier untuk Menentukan Review Positif atau Negatif pelanggan Website Penjualan*, Seminar Nasional “Inovasi dalam Desain dan Teknologi” 2015.

- Putri, D.U.K. 2016. *Implementasi Inferensi Fuzzy Mamdani Untuk Keperluan Sistem Rekomendasi Berita Berbasis Konten*. Skripsi. Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Gadjah Mada, Yogyakarta.
- R.Kohavi, 1995. *A study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*, [Online]. Available: <http://frostiebek.free.fr/docs/Machine%20>
- Reza, J. I.2015. *Ini yang Bikin Shopee Beda dengan Aplikasi Belanja Online lain*. Diunduh dari <http://www.liputan6.com/>.
- Ramadhani, Tiara G. 2015. *Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Model Dokumen Bernoulli Dan Support Vector Machine*. Program Studi Statistika FMIPA UGM Yogyakarta.
- Rianto, Bagus.2016. *Implementasi dan Perbandingan Metode Prapemrosesan Pada Analisis Sentimen Gubernur DKI Jakarta Menggunakan Metode Support Vector Machine dan Naïve Bayes*. Skripsi. Program Studi Ilmu Komputer FMIPA UGM Yogyakarta.
- Rish, I., 2001, An Empirical Study of The Naive Bayes Classifier, In *Proceedings of IJCAI-01 workshop on Empirical Methods in AI*, New York.
- Sadiku, M. N. O., Shadare, A. E., dan Musa, S. M. 2015, A Brief Introduction to Data Mining, 11.21.
- Ramdhani dan Saragih, H, R.2012. *Pengaruh Intensi Pelanggan Dalam Berbelanja Online Kembali Melalui Media Teknologi Informasi forum Jual Beli (FJB) Kaskus*. Journal of Information System, 8(2), 100-112.
- Sekaran, U. 2011. *Research Methods for business Edisi I and 2*. Jakarta: Salemba Empat.
- Sugiyono. 2011. *Metode Penelitian Kualitataif dan R&D*. Bandung: Alfabeta.

- Suryadi.2017. *Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis*, Tesis, Program Pascasarjana, Universitas Udayana, Denpasar.
- Sutabri, Tata. 2012. Analisis Sistem Infomasi. Yogyakarta: Andi.
- Tan, P.N., Michael, S., dan Vipin, K. 2006. *Introduction To Data Mining*. USA: Addison-Wesley.
- Turban, E., King, D., Lee, J.K., Liang, T.-P., Turban, D.C. 2012. *Electronic Commerce: A Managerial and Social Networks. Perspective 8 th ed. Revised Edition*. USA: Sringer. Wong, CK.
- Turban, E., Aronson, J. E., Liang, T.-p., dan McCarthy, R. V., 2006, *Decision Support Systems*, Pearson Education.
- Ulwan, M. N. 2016. *Pattern Recognition Pada Unstructured Data Teks Menggunakan Support Vector Machine Dan Association*. Tugas Akhir. Program Studi Statistika FMIPA UII Yogyakarta.
- Vargiu, E., dan Urru, M. 2012. *Exploiting Web scraping in a collaborative filtering-based approach to web advertising*. Artificial Intelligence Research, 2(1): 44-54.
- Wijayanti, W,N. 2014.*Analisis Sentimen Pada Review Pengguna Sistem Operasi Windows Phone dengan Menggunakan Metode Support Vector Machine (SVM)*. Skripsi. Program Studi Teknologi Informasi Fakultas Teknik UGM Yogyakarta
- Witten, I. H., Frank, E., dan Hall, M. a., 2011, *Data Mining: Practical Machine Learning Tools and Techniques*, Third Edition.
- Xhemali, D., Hinde Christopher J., Stone, Roger J. 2009. *Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages*.

## LAMPIRAN

### **Lampiran 1** *Script R Preprocessing Data dengan Text Mining*

```
# Install
install.packages("tm") # for text mining
install.packages("SnowballC") # for text stemming
install.packages("wordcloud") # word-cloud generator
install.packages("RColorBrewer") # color palettes

# Load
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
library(stringr)

setwd("D://TA/semoga")
docs<-readLines("terbaru.csv")

# Load the data as a corpus
docs <- Corpus(VectorSource(docs))

#Inspect the content of the document
inspect(docs)

#Replacing ?/?, ?@? and ?|? with space:
toSpace <- content_transformer(function (x , pattern )
gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/" )
docs <- tm_map(docs, toSpace, "@")
docs <- tm_map(docs, toSpace, "\\|")

#Cleaning the text
# Convert the text to lower case
docs <- tm_map(docs, content_transformer(tolower))

#Remove punctuation
docs <- tm_map(docs, toSpace, "[[:punct:]]")

#Remove numbers
docs <- tm_map(docs, toSpace, "[[:digit:]]")

# add two extra stop words: "available" and "via"
myStopwords = readLines("stopwordsid.csv")

# remove stopwords from corpus
```



```

docs <- tm_map(docs, removeWords, myStopwords)

# Remove your own stop word
# specify your stopwords as a character vector
docs <- tm_map(docs, removeWords,
c("saya", "kamu", "tolong", "shopee", "mohon", "shoppe", "bulan", "tanggal"))

# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)

# Remove URL
removeURL <- function(x) gsub("http[[:alnum:]]*", " ", x)
docs <- tm_map(docs, removeURL)

#Build a term-document matrix
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m), decreasing=TRUE)
d <- data.frame(word = names(v), freq=v)
head(d, 10)

dataframe<-data.frame(text=unlist(sapply(docs, `[`)),
stringsAsFactors=F)

write.csv(dataframe, "D://TA/semoga/data_cleaning.csv")
save.image()

```

## Lampiran 2 Script R Pelabelan dan Pembobotan

```
library(tm)
setwd("D://TA/semoga/")
docs<-read.csv("data_cleaning.csv",header=TRUE)

#skoring
positif <- scan("D://TA/semoga/s-
pos.txt",what="character",comment.char=";")
negatif <- scan("D://TA/semoga/s-
neg.txt",what="character",comment.char=";")
score.sentiment = function(docs, kata.positif, kata.negatif,
.progress='none')
{
  require(plyr)
  require(stringr)
  scores = laply(docs, 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) - (1*sum(negatif.matches))
return(score)
}, kata.positif, kata.negatif, .progress=.progress )
scores.df = data.frame(score=scores, text=docs)
return(scores.df)
}

hasil = score.sentiment(docs$text, kata.positif, kata.negatif)
View(hasil)

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

#EXCHANGE ROW SEQUENCE
data <- hasil[c(3,1,2)]
View(data)
write.csv(data, file = "Label-indo.csv")
```

### **Lampiran 3 Script R Klasifikasi dengan menggunakan metode Naïve Bayes Classifier (NBC)**

```
#=====Preproses=====#
setwd("D:/TA/semoga")
# Load required libraries
library(tm)
library(RTextTools)
library(e1071)
library(dplyr)
library(caret)
# Library for parallel processing

#Input Data
df<- read.delim("clipboard", stringsAsFactors = TRUE)
glimpse(df)
View(df)
write.csv(df,file = "df10.csv")

#Randomize Dataset
set.seed(10)
df <- df[sample(nrow(df)),]
df <- df[sample(nrow(df)),]
glimpse(df)
write.csv(df,file = "df10.csv")

#Tokenization
corpus <- Corpus(VectorSource(df$text))
corpus
inspect(corpus[1:5])

#Clean Up
corpus.clean <- corpus %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removePunctuation) %>%
  tm_map(removeNumbers) %>%
  tm_map(removeWords, stopwords(kind="en")) %>%
  tm_map(removeWords,
c("mengapa","apasih","dan","please","cucok","dari","membuat","oleh",
",
"dalam","adalah","ini","untuk","sekarang","dulu","cobadeh","jos",
"brengsek","siapa","aseek","hehe","kenapasih","ayo","lagi",
"tapi","karena","saat","ada","sudah","atas","kamu","bloomberg","ai
ng",
"soal","tinggal","dong","minta","please","", "masih","semua")) %>%
  tm_map(stripWhitespace)

#Matriks
```

```

dtm <- DocumentTermMatrix(corpus.clean)
inspect(dtm[40:50, 10:15])

#Partitioning
df.train <- df[1:3072,]
df.test <- df[3073:3840,]

dtm.train <- dtm[1:3072,]
dtm.test <- dtm[3073:3840,]

corpus.clean.train <- corpus.clean[1:3072]
corpus.clean.test <- corpus.clean[3073:3840]
write.csv(df.train,file = "TrainNB10.csv")
write.csv(df.test,file = "TestNB10.csv")

#Featured Selection
dim(dtm.train)

fivefreq <- findFreqTerms(dtm.train,1)
length((fivefreq))
dtm.train.nb <- DocumentTermMatrix(corpus.clean.train,
control=list(dictionary = fivefreq))
dim(dtm.train.nb)
dtm.test.nb <- DocumentTermMatrix(corpus.clean.test,
control=list(dictionary = fivefreq))
dim(dtm.train.nb)

#Boolean Naive Bayes
convert_countNB <- function(x) {
  y <- ifelse(x > 0, 1,0)
  y <- factor(y, levels=c(0,1), labels=c("No", "Yes"))
  y
}

#====Naive Bayes=====#
df=read.csv2("df10.csv")
#Naive Bayes
TrainNB=read.csv2("TrainNB10.csv")
TestNB=read.csv2("TestNB10.csv")
trainNB <- apply(dtm.train.nb, 2, convert_countNB)
testNB <- apply(dtm.test.nb, 2, convert_countNB)

#Training
classifier <- naiveBayes(trainNB, df.train$class, laplace = 1)
#Testing
pred <- predict(classifier, testNB)

#Tabel Prediksi vs Aasli
table("Predictions"= pred, "Actual" = df.test$class )

```

```
#confusion matriks
# Prepare the confusion matrix
conf.mat <- confusionMatrix(pred, df.test$class)
conf.mat

conf.mat$byClass
conf.mat$overall
conf.mat$overall['Accuracy']*100
```

#### Lampiran 4 Script R Visualisasi dan Asosiasi Teks (Untuk Ulasan Positif)

```
# Install
install.packages("tm") # for text mining
install.packages("SnowballC") # for text stemming
install.packages("wordcloud") # word-cloud generator
install.packages("RColorBrewer") # color palettes
# Load
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
library(stringr)

setwd("D://TA/semoga")
docs<-readLines("positif.csv")

# Load the data as a corpus
docs <- Corpus(VectorSource(docs))

# Remove your own stop word
# specify your stopwords as a character vector
docs <- tm_map(docs, removeWords,
c("mesin","dan","mendapatkan","sekarang","syarat","tinggi",
"kali","ketentuan","subsidi","mahal","checkout","berbayar","status",
,"update","produk","kurang",
"kurir","saya","disini","lama","lambat","lemot","ini","dibatasi",
"lambatt","ada","sedikit",
"dalam","terdapat","bbisa","semua","dengan","online","shopee","tid
ak","noror","bismillah",
"jalan","kerena","diatas","super","pemula","pusat","bisnis","downl
oad","tiki","buat","diam",
"terakhir","ragiu","help","diitindak","santai","hialang","perbaika
n"))

# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)

#Replace words
docs <- tm_map(docs, gsub, pattern="banayk",replacement="banyak")
docs <- tm_map(docs, gsub, pattern="stoko",replacement="toko")
docs <- tm_map(docs, gsub, pattern="tiak",replacement="tidak")
docs <- tm_map(docs, gsub, pattern="uupdate",replacement="update")
docs <- tm_map(docs, gsub, pattern="isa",replacement="bisa")
docs <- tm_map(docs, gsub, pattern="uploada",replacement="upload")
docs <- tm_map(docs, gsub, pattern="gratris",replacement="gratis")
docs <- tm_map(docs, gsub,
pattern="kosmetidak",replacement="kosmetik")
docs <- tm_map(docs, gsub, pattern="tidk",replacement="tidak")
docs <- tm_map(docs, gsub, pattern="tdak",replacement="tidak")
docs <- tm_map(docs, gsub,
pattern="sendiir",replacement="sendiri")
docs <- tm_map(docs, gsub, pattern="apps",replacement="aplikasi")
docs <- tm_map(docs, gsub, pattern="grastis",replacement="gratis")
```

```

docs <- tm_map(docs, gsub, pattern="selalau",replacement="selalu")
docs <- tm_map(docs, gsub, pattern="jni",replacement="ini")
docs <- tm_map(docs, gsub, pattern="app",replacement="aplikasi")
docs <- tm_map(docs, gsub, pattern="udh",replacement="sudah")
docs <- tm_map(docs, gsub,
pattern="komplan",replacement="komplain")
docs <- tm_map(docs, gsub,
pattern="serrvice",replacement="service")
docs <- tm_map(docs, gsub,
pattern="digunkam",replacement="digunakan")
docs <- tm_map(docs, gsub, pattern="sayrat",replacement="syarat")
docs <- tm_map(docs, gsub,
pattern="membantyu",replacement="membantu")
docs <- tm_map(docs, gsub, pattern="ongkir",replacement="ongkir")
docs <- tm_map(docs, gsub, pattern="krediat",replacement="kredit")
docs <- tm_map(docs, gsub, pattern="muah",replacement="mudah")

#Build a term-document matrix
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 15)

#Generate the Word cloud
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=50, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))

#Explore frequent terms and their associations
findFreqTerms(dtm, lowfreq = 4)

#Find related words
v<-as.list(findAssocs(dtm, terms
=c("aplikasi","ongkir","pelayanan",
"pengiriman","harga","penjual"),
corlimit = c(0.15,0.15,0.15,0.15,0.15,0.15)))
v
#Find related words (one by one)
v<-as.list(findAssocs(dtm, terms =c("aplikasi"),
corlimit = c(0.15)))
View(v$aplikasi)

#barplot
k<-barplot(d[1:15,]$freq, las = 2, names.arg =
d[1:15,]$word,cex.axis=1.2,cex.names=1.2,
          main ="Most frequent words",
          ylab = "Word frequencies",col = terrain.colors(20))

termFrequency <- rowSums(as.matrix(dtm))
termFrequency <- subset(termFrequency, termFrequency>=115)

text(k,sort(termFrequency, decreasing = T)-
2,labels=sort(termFrequency, decreasing = T),pch = 6, cex =
1)

```

## Lampiran 5 Script R Visualisasi dan Asosiasi Teks (Untuk Ulasan Negatif)

```
# Install
install.packages("tm") # for text mining
install.packages("SnowballC") # for text stemming
install.packages("wordcloud") # word-cloud generator
install.packages("RColorBrewer") # color palettes

setwd("D://TA/semoga")
docs<-readLines("negatif.csv")

# Load the data as a corpus
docs <- Corpus(VectorSource(docs))

# Remove your own stop word
# specify your stopwords as a character vector
docs <- tm_map(docs, removeWords,
c("mesin","dan","bbisa","bisa","tak","tersedia"))

# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)

#Replace words
docs <- tm_map(docs, gsub, pattern="banayk",replacement="banyak")
docs <- tm_map(docs, gsub, pattern="stoko",replacement="toko")
docs <- tm_map(docs, gsub, pattern="lamba",replacement="lambat")
docs <- tm_map(docs, gsub, pattern="uupdate",replacement="update")
docs <- tm_map(docs, gsub, pattern="isa",replacement="bisa")
docs <- tm_map(docs, gsub, pattern="uploada",replacement="upload")
docs <- tm_map(docs, gsub, pattern="gratris",replacement="gratis")
docs <- tm_map(docs, gsub,
pattern="kosmetidak",replacement="kosmetik")
docs <- tm_map(docs, gsub, pattern="lambatt",replacement="lambat")
docs <- tm_map(docs, gsub, pattern="tdak",replacement="tidak")
docs <- tm_map(docs, gsub,
pattern="sendiir",replacement="sendiri")
docs <- tm_map(docs, gsub, pattern="apps",replacement="aplikasi")
docs <- tm_map(docs, gsub, pattern="grastis",replacement="gratis")
docs <- tm_map(docs, gsub, pattern="selalau",replacement="selalu")
docs <- tm_map(docs, gsub, pattern="jni",replacement="ini")
docs <- tm_map(docs, gsub, pattern="app",replacement="aplikasi")
docs <- tm_map(docs, gsub, pattern="udh",replacement="sudah")
docs <- tm_map(docs, gsub,
pattern="komplan",replacement="komplain")
docs <- tm_map(docs, gsub,
pattern="serrvice",replacement="service")
docs <- tm_map(docs, gsub,
pattern="digunkam",replacement="digunakan")
docs <- tm_map(docs, gsub, pattern="sayrat",replacement="syarat")
docs <- tm_map(docs, gsub,
pattern="membantyu",replacement="membantu")
docs <- tm_map(docs, gsub, pattern="ongir",replacement="ongkir")
docs <- tm_map(docs, gsub, pattern="krediat",replacement="kredit")
docs <- tm_map(docs, gsub, pattern="muah",replacement="mudah")
```



```

#Build a term-document matrix
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 15)

#Generate the Word cloud
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
           max.words=50, random.order=FALSE, rot.per=0.35,
           colors=brewer.pal(8, "Dark2"))

#Explore frequent terms and their associations
findFreqTerms(dtm, lowfreq = 4)

#Find related words
v<-as.list(findAssocs(dtm, terms
=c("aplikasi","ongkir","pelayanan",
"pengiriman","harga","penjual"),
corlimit = c(0.15,0.15,0.15,0.15,0.15,0.15)))
v
#Find related words (one by one)
v<-as.list(findAssocs(dtm, terms =c("aplikasi"),
corlimit = c(0.15)))
View(v$aplikasi)

#barplot
k<-barplot(d[1:15,]$freq, las = 2, names.arg =
d[1:15,]$word,cex.axis=1.2,cex.names=1.2,
           main ="Most frequent words",
           ylab = "Word frequencies",col = terrain.colors(20))

termFrequency <- rowSums(as.matrix(dtm))
termFrequency <- subset(termFrequency, termFrequency>=115)

text(k,sort(termFrequency, decreasing = T)-
2,labels=sort(termFrequency, decreasing = T),pch = 6, cex =
1)

```

## Lampiran 6 *Stopwords Berbahasa Indonesia*

ada	antaranya	bagaimanapun	begitukah
adalah	apa	bagi	begitulah
adanya	apaan	bagian	begitupun
adapun	apabila	bahkan	bekerja
agak	apakah	bahwa	belakang
agaknya	apalagi	bahwasanya	belakangan
agar	apatah	baik	belum
akan	artinya	bakal	belumah
akankah	asal	bakalan	benar
akhir	asalkan	balik	benarkah
akhiri	atas	banyak	benarlah
akhirnya	atau	bapak	berada
aku	ataukah	baru	berakhir
akulah	ataupun	bawah	berakhirilah
amat	awal	beberapa	berakhirnya
amatlah	awalnya	begini	berapa
anda	bagai	beginian	berapakah
andalah	bagaikan	beginikah	berapalah
antar	bagaimana	beginilah	berapapun
antara	bagaimanakah	begitu	berarti
berawal	bermula	bisakah	dapat
berbagai	bersama	boleh	dari
berdatangan	bersama-sama	bolehkah	daripada
beri	bersiap	bolehlah	datang
berikan	bersiap-siap	buat	dekat
berikut	bertanya	bukan	demi
berikutnya	bertanya-tanya	bukankah	demikian
berjumlah	berturut	bukanlah	demikianlah
berkali-kali	berturut-turut	bukannya	dengan
berkata	bertutur	bulan	depan
berkehendak	berujar	bung	di
berkeinginan	berupa	cara	dia
berkenaan	besar	caranya	diakhiri
berlainan	betul	cukup	diakhirinya
berlalu	betulkah	cukupkah	dialah
berlangsung	biasa	cukuplah	diantara
berlebihan	biasanya	cuma	diantaranya
bermacam	bila	dahulu	diberi
bermacam-	bilakah	dalam	diberikan

macam			
bermaksud	bisa	dan	diberikannya
dibuat	dilakukan	diperlihatkan	ditujukan
dibuatnya	dilalui	diperlukan	ditunjuk
didapat	dilihat	diperlukannya	ditunjuki
didatangkan	dimaksud	dipersoalkan	ditunjukkan
digunakan	dimaksudkan	dipertanyakan	ditunjukkannya
diibaratkan	dimaksudkannya	dipunyai	ditunjuknya
diibaratkannya	dimaksudnya	diri	dituturkan
diingat	diminta	dirinya	dituturkannya
diingatkan	dimintai	disampaikan	diucapkan
diinginkan	dimisalkan	disebut	diucapkannya
dijawab	dimulai	disebutkan	diungkapkan
dijelaskan	dimulailah	disebutkannya	dong
dijelaskannya	dimulainya	disini	dua
dikarenakan	dimungkinkan	disinilah	dulu
dikatakan	dini	ditambahkan	empat
dikatakannya	dipastikan	ditandaskan	enggak
dikerjakan	diperbuat	ditanya	enggaknya
diketahui	diperbuatnya	ditanyai	entah
diketuainya	dipergunakan	ditanyakan	entahlah
dikira	diperkirakan	ditegaskan	guna
hal	ingat-ingat	jelas	kapan
hampir	ingin	jelaskan	kapankah
hanya	inginkan	jelaslah	kapanpun
hanyalah	inginkan	jelasnya	karena
hari	ini	jika	karenanya
harus	inikah	jikalau	kasus
haruslah	inilah	juga	kata
harusnya	itu	jumlah	katakan
hendak	itukah	jumlahnya	katakanlah
hendaklah	itulah	justru	katanya
hendaknya	jadi	kala	ke
hingga	jadilah	kalau	keadaan
ia	jadinya	kalaulah	kebetulan
ialah	jangan	kalaupun	kecil
ibarat	jangan	kalian	kedua
ibaratkan	janganlah	kami	keduanya
ibaratnya	jauh	kamilah	keinginan
ibu	jawab	kamu	kelamaan
ikut	jawaban	kamulah	kelihatan
ingat	jawabnya	kan	kelihatannya

kelima	kita	makanya	melakukan
keluar	kitalah	makanan	melalui
kembali	kok	makin	melihat
kemudian	kurang	malah	melihatnya
kemungkinan	lagi	malahan	memang
kemungkinannya	lagian	mampu	memastikan
kenapa	lah	mampukah	memberi
kepada	lain	mana	memberikan
kepadanya	lainnya	manakala	membuat
kesampaian	lalu	manalagi	memerlukan
keseluruhan	lama	masa	memihak
keseluruhannya	lamanya	masalah	meminta
keterlaluhan	lanjut	masalahnya	memintakan
ketika	lanjutnya	masih	memisalkan
khususnya	lebih	masihkah	memperbuat
kini	lewat	masing	mempergunakan
kinilah	lima	masing-masing	memperkirakan
kira	luar	mau	memperlihatkan
kira-kira	macam	maupun	mempersiapkan
kiranya	maka	melainkan	mempersoalkan
mempertanyakan	mengapa	menuju	meyakinkan
mempunyai	mengatakan	menunjuk	minta
memulai	mengatakannya	menunjuki	mirip
memungkinkan	mengenai	menunjukkan	misal
menaiki	mengerjakan	menunjuknya	misalkan
menambahkan	mengetahui	menurut	misalnya
menandaskan	menggunakan	menuturkan	mula
menanti	menghendaki	menyampaikan	mulai
menanti-nanti	mengibaratkan	menyangkut	mulailah
menantikan	mengibaratkannya	menyatakan	mulanya
menanya	mengingat	menyebutkan	mungkin
menanyai	mengingatkan	menyeluruh	mungkinkah
menanyakan	menginginkan	menyiapkan	nah
mendapat	mengira	merasa	naik
mendapatkan	mengucapkan	mereka	namun
mendatang	mengucapkannya	merekalah	nanti
mendatangi	mengungkapkan	merupakan	nantinya
mendatangkan	menjadi	meski	nyaris
menegaskan	menjawab	meskipun	nyatanya
mengakhiri	menjelaskan	meyakini	oleh

olehnya	pertama	sama-sama	kapok
pada	pertama-tama	sambil	haha
padahal	pertanyaan	sampai	ols
padanya	pertanyakan	sampai-sampai	dst
pak	pihak	sampaikan	x
paling	pihaknya	sana	ya
panjang	pukul	sangat	malh
pantas	pula	sangatlah	g
para	pun	satu	sekat
pasti	punya	saya	huhhhh
pastilah	rasa	sayalah	sekarnag
penting	rasanya	se	manapun
pentingnya	rata	sebab	cape
per	rupanya	sebabnya	kemarin
percuma	saat	sebagai	yakali
perlu	saatnya	sebagaimana	tlong
perlukah	saja	sebagainya	gannn
perlunya	sajalah	sebagian	alhamdulillah
pernah	saling	sebaik	hadeh
persoalan	sama	sebaik-baiknya	mesen
admin	mohon		
buka	banget		
pelapak	kalo		
lapak	toko		
beli	good		
ulasan	terima		
lengkap	masuk		
beli	Jual		
belanja	kasih		
aja			
udah			

Atau untuk *stopword* berbahasa Indonesia dapat diunduh pada situs <https://sites.google.com/site/kevinbouge/stopwords-lists>

Untuk kamus *lexicon* berbahasa Indonesia dapat diunduh pada situs <https://github.com/masdevi/ID-OpinionWords>

**Lampiran 7** Hasil iterasi pada proses klasifikasi dengan *Naïve Bayes Clasifier*

Iterasi ke-	Akurasi	Presisi	Recall
1	0.974	0.982	0.97
2	0.957	0.965	0.953
3	0.953	0.965	0.945
4	0.97	0.974	0.972
5	0.96	0.973	0.951
6	0.965	0.98	0.955
7	0.953	0.964	0.944
8	0.958	0.96	0.96
9	0.965	0.966	0.968
10	0.969	0.973	0.969
Rata-Rata	0.9624	0.9702	0.9587