

**ANALISIS SENTIMEN ONLINE REVIEW PENGGUNA  
*E-COMMERCE* MENGGUNAKAN METODE *SUPPORT  
VECTOR MACHINE* DAN *MAXIMUM ENTROPY***

(Studi Kasus: Review Bukalapak pada Google Play)

**TUGAS AKHIR**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana  
Jurusan Statistika**



**Disusun Oleh:**

**Ditia Yosmita Praptiwi**

**14 611 220**

**PROGRAM STUDI STATISTIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS ISLAM INDONESIA**

**YOGYAKARTA**

**2018**

**HALAMAN PERSETUJUAN PEMBIMBING**

**TUGAS AKHIR**

Judul : Analisis Sentimen Online Review Pengguna *E-commerce* Menggunakan Metode *Support Vector Machine* dan *Maximum Entropy* (Studi Kasus: Review Bukalapak Pada Google Play)

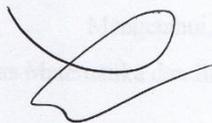
Nama Mahasiswa : Ditia Yosmita Praptiwi

Nomor Mahasiswa : 14 611 220

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

Yogyakarta, 17 April 2018

Pembimbing



**(Tuti Purwaningsih, S.Stat., M.Si.)**

**HALAMAN PENGESAHAN**

**TUGAS AKHIR**

**ANALISIS SENTIMEN ONLINE REVIEW PENGGUNA  
E-COMMERCE MENGGUNAKAN METODE *SUPPORT  
VECTOR MACHINE* DAN *MAXIMUM ENTROPY***

(Studi Kasus: Review Bukalapak pada Google Play)

**Nama Mahasiswa : Ditia Yosmita Praptiwi**

**Nomor Mahasiswa : 14 611 220**

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

**Nama Penguji**

**Tanda tangan**

1. Andrie Pasca Hendradewa, S.T., M.T.

2. Ayundyah Kesumawati, S.Si., M.Si.

3. Tuti Purwaningsih, S.Stat., M.Si.

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



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

## KATA PENGANTAR

*Assalamu alaikum wa rahmatullahi wa barakaatuh*

Puji syukur Kehadirat Allah yang telah melimpahkan Rahmat, Hidayah serta Karunia-Nya sehingga Tugas Akhir dengan judul “*Analisis Sentimen Online Review Pengguna E-commerce Menggunakan Metode Support Vector Machine dan Maximum Entropy (Studi Kasus: Review Aplikasi Bukalapak Pada Google Play)*” ini dapat terselesaikan. Shalawat dan salam semoga senantiasa tercurahkan kepada Nabi Muhammad SAW beserta keluarga, sahabat serta para pengikut beliau hingga akhir zaman yang syafaatnya dinantikan di akhirat kelak.

Tugas akhir ini merupakan salah satu syarat guna memperoleh gelar Sarjana di Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. Selama penyelesaian Tugas Akhir ini, penulis menyadari banyak pihak yang telah memberikan dorongan, bantuan serta bimbingan hingga Tugas Akhir ini dapat terselesaikan. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Bapak Drs. Allwar, M.Sc, Ph.D selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, beserta seluruh jajarannya.
2. Bapak Dr. RB. Fajriya Hakim, S.Si, M.Si, selaku ketua Jurusan Statistika beserta seluruh jajarannya.
3. Tuti Purwaningsih, S.Stat, M.Si, selaku dosen pembimbing yang sangat berjasa dalam penyelesaian Tugas Akhir ini dan selalu memberi bimbingan hingga Tugas Akhir ini dapat terselesaikan.
4. Seluruh dosen pengajar dan staff prodi Statistika, terimakasih atas bekal ilmu dan bantuannya dalam proses belajar, semoga menjadi amal kebaikan Bapak/Ibu sekalian.
5. Kedua orang tua tercinta, Bapak Sutarwi dan Ibu Desmi Yanti yang selalu memberikan semangat, dukungan dan doa setiap saat.

6. Adik tersayang Bella Mudyarti Sulistyorini serta Keluarga Besar yang selalu menyemangati dan mendoakan yang terbaik.
7. Sahabat sekaligus keluarga penulis Cimit, Icha, Erdwika, Een, Reny, Maul, Sari, Ayu, Rabi, Khusnul, Annisa, Dhea, Sita, Ridwan, Rifky, Rian, Oddy yang selalu ada saat dibutuhkan maupun tidak dibutuhkan.
8. Teman-teman KKN Unit 119 Purworejo: Kiki , Idel, Mbak Put, Nadia, Guntur, Adit, Hasan dan Mas Tri yang telah memberikan doa serta hiburan kepada penulis.
9. Seluruh teman-teman Statistika yang telah memberikan semangat dan dukungan.
10. Semua pihak yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari kesempurnaan. Oleh karena itu, segala kritik dan saran yang bersifat membangun selalu penulis harapkan. Semoga Tugas Akhir ini dapat memberikan manfaat. Akhir kata, semoga Allah senantiasa melimpahkan Rahmat dan Hidayah-Nya kepada kita semua.

*Wassalamu alaikum wa rahmatullahi wa barakaatuh.*

Yogyakarta, 17 April 2018

Ditia Yosmita Praptiwi

## DAFTAR ISI

|  |      |
|--|------|
| <b>HALAMAN JUDUL</b> .....                             | i    |
| <b>HALAMAN PERSETUJUAN PEMBIMBING</b> .....            | ii   |
| <b>HALAMAN PENGESAHAN</b> .....                        | iii  |
| <b>KATA PENGANTAR</b> .....                            | iv   |
| <b>DAFTAR ISI</b> .....                                | vi   |
| <b>DAFTAR TABEL</b> .....                              | ix   |
| <b>DAFTAR GAMBAR</b> .....                             | x    |
| <b>DAFTAR LAMPIRAN</b> .....                           | xi   |
| <b>PERNYATAAN</b> .....                                | xii  |
| <b>INTISARI</b> .....                                  | xiii |
| <b>ABSTRACT</b> .....                                  | xiv  |
| <b>BAB I PENDAHULUAN</b> .....                         | 1    |
| 1.1. Latar Belakang .....                              | 1    |
| 1.2. Rumusan Masalah .....                             | 5    |
| 1.3. Batasan Masalah.....                              | 5    |
| 1.4. Tujuan Penelitian.....                            | 6    |
| 1.5. Manfaat Penelitian.....                           | 6    |
| 1.6. Sistematika Penulisan.....                        | 6    |
| <b>BAB II TINJAUAN PUSTAKA</b> .....                   | 8    |
| <b>BAB III LANDASAN TEORI</b> .....                    | 14   |
| 3.1. <i>E-commerce</i> .....                           | 14   |
| 3.1.1. Definisi <i>E-commerce</i> .....                | 14   |
| 3.1.2. Komponen <i>E-commerce</i> .....                | 14   |
| 3.1.3. Model-model <i>E-commerce</i> di Indonesia..... | 16   |
| 3.2. Bukalapak .....                                   | 19   |
| 3.3. <i>Google Play</i> .....                          | 20   |
| 3.4. <i>Online Review</i> .....                        | 20   |
| 3.5. Kepuasan Pelanggan.....                           | 21   |
| 3.6. <i>Data Mining</i> .....                          | 23   |

|  |    |
|--|----|
| 3.7. <i>Machine Learning</i> .....                                     | 24 |
| 3.8. <i>Text Mining</i> .....  | 25 |
| 3.8.1. Pengertian <i>Text Mining</i> .....                             | 25 |
| 3.8.2. <i>Text Preprocessing</i> .....                                 | 26 |
| 3.8.3. <i>Feature Selection</i> .....                                  | 28 |
| 3.9. Pembobotan Kata ( <i>Term Weighting</i> ).....                    | 28 |
| 3.9.1. Simulasi Pembobotan kata dengan <i>TF-IDF</i> .....             | 30 |
| 3.10. Analisis Sentimen.....   | 33 |
| 3.11. Klasifikasi.....   | 34 |
| 3.11.1. Ukuran Evaluasi Model.....                                     | 34 |
| 3.11.2. <i>K-Fold Cross Validation</i> .....                           | 36 |
| 3.12. <i>Support Vector Machine (SVM)</i> .....                        | 37 |
| 3.12.1. <i>SVM</i> Pada Data Terpisah Secara <i>Linear</i> .....       | 38 |
| 3.12.2. <i>SVM</i> Pada Data Tidak Terpisah Secara <i>Linear</i> ..... | 40 |
| 3.12.3. <i>Kernel Trick</i> dan <i>Non Linear Classification</i> ..... | 41 |
| 3.13. <i>Maximum Entropy</i> .....                                     | 42 |
| 3.13.1. Definisi <i>Entropy</i> .....                                  | 44 |
| 3.13.2. Prinsip <i>Maximum Entropy</i> .....                           | 45 |
| 3.13.3. Algoritma Klasifikasi dengan <i>Maximum Entropy</i> .....      | 45 |
| 3.13.4. Simulasi Metode <i>Maximum Entropy</i> .....                   | 46 |
| 3.14. <i>Wordcloud</i> .....   | 48 |
| 3.15. Asosiasi Kata .....  | 48 |
| 3.15.1. Simulasi Perhitungan Asosiasi Pada Data Teks .....             | 49 |
| 3.16. Metode Diagram <i>Fishbone</i> .....                             | 50 |
| <b>BAB IV METODOLOGI PENELITIAN</b> .....                              | 52 |
| 4.1. Populasi dan Sampel.....  | 52 |
| 4.2. Variabel dan Definisi Opsional Variabel .....                     | 52 |
| 4.3. Teknik Pengumpulan data .....                                     | 52 |
| 4.4. Metode Analisis Data .....  | 52 |
| 4.5. Proses Analisis Data .....  | 53 |

|  |           |
|--|-----------|
| <b>BAB V ANALISIS DAN PEMBAHASAN .....</b>                 | <b>54</b> |
| 5.1. Analisis Deskriptif.....                              | 54        |
| 5.2. <i>Preprocessing</i> atau Prapemrosesan Data.....     | 56        |
| 5.2.1. <i>Spelling Normalization</i> .....                 | 57        |
| 5.2.2. <i>Case Folding</i> .....                           | 57        |
| 5.2.3. <i>Tokenizing</i> .....                             | 58        |
| 5.2.4. <i>Filtering</i> .....                              | 59        |
| 5.3. Pelabelan Kelas Sentimen .....                        | 60        |
| 5.3.1. Simulasi Perhitungan Skor Sentimen .....            | 64        |
| 5.4. Pembuatan Data Latih dan Data Uji.....                | 65        |
| 5.5. Klasifikasi dengan <i>SVM</i> dan <i>Maxent</i> ..... | 65        |
| 5.6. Visualisasi dan Asosiasi .....                        | 72        |
| 5.6.1. Ulasan Positif .....                                | 72        |
| 5.6.2. Ulasan Negatif .....                                | 76        |
| 5.7. Diagram <i>Fishbone</i> .....                         | 80        |
| <b>BAB VI KESIMPULAN DAN SARAN .....</b>                   | <b>85</b> |
| 6.1. Kesimpulan.....                                       | 85        |
| 6.2. Saran .....   | 86        |
| <b>DAFTAR PUSTAKA .....</b>                                | <b>87</b> |
| <b>LAMPIRAN.....</b>                                       | <b>92</b> |

## DAFTAR TABEL

|                   |   |    |
|-------------------|---|----|
| <b>Tabel 2.1</b>  | Perbandingan dengan penelitian terdahulu .....                                    | 11 |
| <b>Tabel 3.1</b>  | Hasil perhitungan bobot pada masing-masing dokumen.....                           | 32 |
| <b>Tabel 3.2</b>  | Contoh <i>K-Fold Cross Validation</i> .....                                       | 37 |
| <b>Tabel 3.3</b>  | Pemberian indeks pada masing-masing kata.....                                     | 47 |
| <b>Tabel 4.1</b>  | Definisi Operasional Variabel.....  | 52 |
| <b>Tabel 5.1</b>  | Tahap-tahap pelabelan menggunakan <i>software R</i> .....                         | 61 |
| <b>Tabel 5.2</b>  | Perbandingan jumlah data pada kelas sentimen.....                                 | 63 |
| <b>Tabel 5.3</b>  | Hasil pelabelan kelas sentiment berbasis kamus <i>lexicon</i> .....               | 63 |
| <b>Tabel 5.4</b>  | Simulasi perhitungan skor sentimen.....   | 64 |
| <b>Tabel 5.5</b>  | Perbandingan data latih dan data uji.....   | 65 |
| <b>Tabel 5.6</b>  | Perbandingan penggunaan metode kernel pada klasifikasi <i>SVM</i> ...             | 66 |
| <b>Tabel 5.7</b>  | Klasifikasi <i>Maximum Entropy</i> .....  | 66 |
| <b>Tabel 5.8</b>  | Tahap analisis <i>SVM</i> dan <i>Maximum Entropy</i> dengan <i>software R</i> ... | 67 |
| <b>Tabel 5.9</b>  | Perbandingan nilai akurasi metode <i>SVM</i> dan <i>Maximum Entropy</i> ..        | 69 |
| <b>Tabel 5.10</b> | <i>Confusion matrix</i> .....   | 70 |
| <b>Tabel 5.11</b> | Asosiasi kata pada kelas sentimen positif.....                                    | 74 |
| <b>Tabel 5.12</b> | Asosiasi kata pada kelas sentimen negatif.....                                    | 78 |
| <b>Tabel 5.13</b> | Rencana pemecahan masalah <i>e-commerce</i> Bukalapak.....                        | 81 |

## DAFTAR GAMBAR

|                    |  |    |
|--------------------|--|----|
| <b>Gambar 1.1</b>  | Latar Belakang .....   | 1  |
| <b>Gambar 3.1</b>  | Ilustrasi <i>SVM</i> menemukan <i>hyperplane</i> terbaik yang memisahkan dua kelas -1 dan +1 .....             | 38 |
| <b>Gambar 3.3</b>  | Transformasi dari <i>input space</i> ke <i>feature space</i> .....   | 40 |
| <b>Gambar 3.4</b>  | Permen dan masing-masing labelnya .....  | 45 |
| <b>Gambar 4.1</b>  | <i>Flowchart</i> Penelitian .....  | 53 |
| <b>Gambar 5.1</b>  | Rating pengguna terhadap <i>e-commerce</i> Bukalapak pada situs <i>Google Play</i> .....                       | 54 |
| <b>Gambar 5.2</b>  | Jumlah ulasan Bukalapak berdasarkan waktu.....   | 55 |
| <b>Gambar 5.3</b>  | Jumlah dan persentase ulasan positif, ulasan negatif,dan total ulasan terhadap <i>ecommerce</i> Bukalapak..... | 56 |
| <b>Gambar 5.4</b>  | Proses <i>spelling normalization</i> .....   | 57 |
| <b>Gambar 5.5</b>  | Proses <i>case folding</i> .....   | 58 |
| <b>Gambar 5.6</b>  | Proses <i>tokenizing</i> .....   | 59 |
| <b>Gambar 5.7</b>  | Proses <i>filtering</i> .....  | 60 |
| <b>Gambar 5.8</b>  | Kata yang paling banyak muncul dari kelas positif <i>ecommerce</i> Bukalapak.....                              | 72 |
| <b>Gambar 5.9</b>  | <i>Wordcloud</i> ulasan positif <i>e-commerce</i> Bukalapak.....   | 73 |
| <b>Gambar 5.10</b> | Kata yang paling banyak muncul dari kelas negatif.....   | 76 |
| <b>Gambar 5.11</b> | <i>Wordcloud</i> ulasan negatif.....   | 77 |
| <b>Gambar 5.12</b> | Diagram <i>fishbone</i> komplain pengguna Bukalapak.....   | 81 |

## DAFTAR LAMPIRAN

|                   |  |     |
|-------------------|--|-----|
| <b>Lampiran 1</b> | <i>Script R Preprocessing Data dengan Text Mining</i> .....                          | 92  |
| <b>Lampiran 2</b> | <i>Script R Pelabelan Kelas Sentimen</i> .....                                       | 96  |
| <b>Lampiran 3</b> | <i>Script R Klasifikasi dengan Machine Learning menggunakan SVM dan Maxent</i> ..... | 97  |
| <b>Lampiran 4</b> | <i>Script R Visualisasi dan Asosiasi Kata</i> .....                                  | 98  |
| <b>Lampiran 5</b> | <i>Stopwords Berbahasa Indonesia</i> .....   | 100 |
| <b>Lampiran 6</b> | <i>Outupt SVM dan Maxent</i> .....   | 106 |

ANALISIS SENTIMEN PERNYATAAN VIEW PENGGERA  
E-COMMERCE DENGAN MENGGUNAKAN METODE SUPPORT

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

Universitas Islam Indonesia

INTISARI

Yogyakarta, 17 April 2018

Dukungan merupakan salah satu e-commerce terapan penting bagi sebuah perusahaan atau organisasi untuk mempromosikan produk atau layanan yang mereka tawarkan kepada publik mengenai produk atau layanan yang mereka tawarkan. Opsi yang muncul saat pembeli dapat mempengaruhi organisasi atau perusahaan. Akan tetapi, memonitor dan menganalisis opini masyarakat juga bukanlah hal yang mudah. Opini yang banyak tidak diproses secara manual. Oleh karena itu, diperlukan metode atau teknik khusus yang mampu mengkategorikan review-review tersebut secara otomatis, apakah termasuk positif atau negatif. Salah satu sistem yang menggunakan *fine review* adalah *Google Play*. Data yang diperoleh dari situs *Google Play* dilakukan pelatihan dan diuji dengan menggunakan metode *Support Vector Machine* dan *Maximum Entropy* untuk mengklasifikasi data review. Dari hasil pelatihan yang telah dilakukan diketahui akan dilihat asosiasi teks pada kelas sentimen untuk menentukan sebuah informasi yang dianggap penting dan dapat berguna untuk pengambilan keputusan. Klasifikasi dengan metode SVM diperoleh tingkat akurasi sebesar 91,95%. Sedangkan dengan metode *Maximum Entropy* diperoleh tingkat akurasi yang lebih tinggi yaitu sebesar 92,98%. Selanjutnya, metode asosiasi teks pada kelas sentimen positif diantaranya terdapat barang, layanan, pesanan, pengiriman, respon, pembelian, kualitas, keamanan dan pilihan. Sedangkan pada kelas sentimen negatif yang sering diidentifikasi diantaranya barang, waktu, server, chat, email, transaksi, upload, promo, voucher, pembelian dan upgrade. Hasil analisis regresi tersebut dibuat dalam diagram *fishbone* untuk pemecahan masalah.



Penulis

Kata kunci : Analisis Sentimen, Support Vector Machine (SVM), Maximum Entropy (asosiasi), Asosiasi Teks, Diagram Fishbone, Bukalapak, Google Play

**ANALISIS SENTIMEN PADA ONLINE REVIEW PENGGUNA  
E-COMMERCE DENGAN MENGGUNAKAN METODE *SUPPORT  
VECTOR MACHINE* DAN *MAXIMUM ENTROPY*  
(Studi Kasus: Review Bukalapak pada Google Play)**

Ditia Yosmita Praptiwi

Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Islam Indonesia

**INTISARI**

Bukalapak merupakan salah satu *e-commerce* terkemuka di Indonesia. Penting bagi sebuah perusahaan atau organisasi untuk mengetahui tanggapan publik mengenai produk atau layanan yang mereka tawarkan. Tidak bisa dipungkiri bahwa opini yang muncul dari publik dapat mempengaruhi citra dari sebuah organisasi atau perusahaan. Akan tetapi, memantau dan mengorganisasi opini dari masyarakat juga bukanlah hal yang mudah. Opini yang dimuat jumlahnya terlalu banyak untuk diproses secara manual. Oleh sebab itu, diperlukan sebuah metode atau teknik khusus yang mampu mengkategorikan *review-review* tersebut secara otomatis, apakah termasuk positif atau negatif. Salah satu situs yang menyediakan fitur *review* adalah *Google Play*. Data yang diperoleh dari situs *Google Play* dilakukan pelabelan dan dianalisis dengan menggunakan metode *Support Vector Machine* dan *Maximum Entropy* untuk mengklasifikasikan data *review*. Dari hasil pelabelan yang telah dilakukan kemudian akan dilihat asosiasi teks pada setiap kelas sentimen untuk menemukan sebuah informasi yang dianggap penting dan dapat berguna untuk pengambilan keputusan. Klasifikasi dengan metode *SVM* diperoleh tingkat akurasi sebesar 91,95% . Sedangkan dengan metode *Maxent* memberikan tingkat akurasi yang lebih tinggi yaitu sebesar 92,98%. Selanjutnya, metode asosiasi teks pada kelas sentimen positif diantaranya terkait barang, transaksi, fitur, pelayanan, pesanan, pengiriman, respon, berbelanja, akulaku, kebutuhan dan cicilan. Sedangkan pada kelas sentimen negatif yang sering dikeluhkan diantaranya barang, update, server, chat, email, transaksi, upload, promo, voucher, bukadompet dan upgrade. Hasil ulasan negatif tersebut dibuat dalam diagram *fishbone* untuk pemecahan masalah.

**Kata kunci :** Analisis Sentimen, *Support Vector Machine (SVM)* , *Maximum Entropy (Maxent)*, Asosiasi Teks, Diagram *Fishbone*, Bukalapak, *Google Play*

**SENTIMENT ANALYSIS ONLINE REVIEW E-COMMERCE USER USING  
SUPPORT VECTOR MACHINE AND MAXIMUM ENTROPY**

*(Case Study: Bukalapak Review On Google Play)*

Ditia Yosmita Praptiwi

*Department of Statistics, Faculty of Mathematics and Natural Science*

*Islamic University of Indonesia*

**ABSTRACT**

*Bukalapak is one of the best e-commerce in Indonesia. It is important for Company or organization to know about how public responds about their products or services which they offer. Opinion can affect how the way people make some decision in their life. It is undeniable that opinion which comes from public can affect the image of an organization or company. However, monitoring and organizing public opinion is not an easy work to do. The amount of opinion that expressed in social media is too much to be processed manually. Therefore, a special method or technique is needed to categorize the reviews automatically, whether it is positive or negative. One of the sites that provide a review feature is Google Play. Data obtained from the Google Play site is further labeled and analyzed using the Support Vector Machine (SVM) and Maximum Entropy (Maxent) method to classify data review. From the labeling results that have been done then will be seen text association on each classes of sentiment to find a fact and information that is considered important and can be useful for decision-making. Classification using SVM method showed an accuracy of 91,95% although using Maxent method showed an accuracy of 92,98%. Furthermore, text association method on positive sentiment is about good, transaction, features, service, order, shipping, response, shopping, akulaku, needs, and instalment. Meanwhile on negative sentiment is often complained about good, update, server, chat, email, transaction, promo, voucher, bukadompet, and upgrade. The negative review results are made in the fishbone diagram for troubleshooting.*

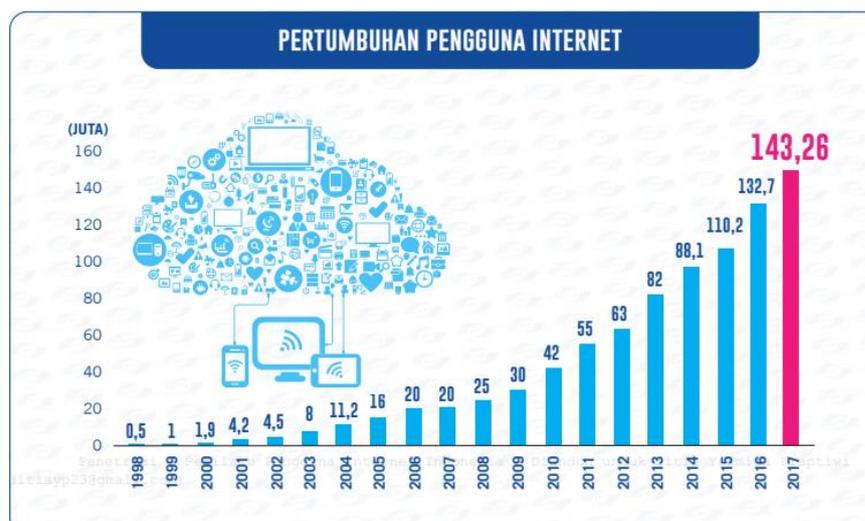
**Keywords :** *Sentiment Analysis, Support Vector Machine (SVM) , Maximum Entropy (Maxent), Association Teks, Fishbone Diagram, Bukalapak, Google Play*

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Dunia teknologi saat ini berkembang semakin pesat menuju ke arah serba digital. Era digital telah membuat manusia memasuki gaya hidup baru yang tidak bisa lepas dari perangkat yang serba elektronik. Teknologi menjadi alat yang membantu kebutuhan manusia, dengan teknologi, apapun dapat dilakukan dengan lebih mudah. Begitu pentingnya peran teknologi inilah yang mulai membawa peradaban memasuki ke era digital. Meningkatnya kebutuhan akan data dan informasi mendorong manusia untuk mengembangkan teknologi baru agar pengolahan data dan informasi dapat dilakukan dengan mudah dan cepat (Josi, 2014). Kemajuan di bidang teknologi, komputer dan telekomunikasi telah mendukung perkembangan teknologi internet. Pengguna internet terus bertambah setiap tahunnya. Berikut ini adalah pertumbuhan pengguna internet dari tahun 1998-2017 berdasarkan *survey* yang dilakukan Asosiasi Penyelenggara Jasa Internet Indonesia.



**Gambar 1.1** *Pertumbuhan Pengguna Internet*  
(sumber: Survey APJII, 2017)

Sebuah *survey* yang diselenggarakan Asosiasi Penyelenggaraan Jasa Internet Indonesia (APJII) menunjukkan bahwa jumlah pengguna internet di Indonesia meningkat tiap tahunnya hingga akhir tahun 2017 mencapai 143,6 juta orang, dari total penduduk sebanyak 262 juta orang. Perkembangan tersebut berdampak pada berbagai bidang. Salah satunya ialah maraknya kegiatan berbelanja melalui media internet.. Menurut survei yang dilakukan oleh APJII 62% dari pengguna internet di Indonesia mengetahui internet sebagai tempat jual beli barang dan jasa, dan 63,5% pernah melakukan transaksi secara online. Selain itu juga 62% sering mengunjungi konten komersial *online shop*.

Indonesia telah menjadi pasar terbesar *e-commerce* di Asia Tenggara. Pada 2014, Euromonitor mencatat, penjualan *online* Indonesia mencapai US\$ 1,1 miliar, lebih tinggi dari Thailand dan Singapura. Namun, jika dibandingkan dengan total perdagangan retail, penjualan *e-commerce* di Indonesia hanya menyumbangkan 0,07 persen. Artinya, pasar *e-commerce* Indonesia berpotensi untuk tumbuh semakin besar. Apalagi dengan jumlah penduduk dan tingkat produk domestik bruto (PDB) terbesar di ASEAN. Euromonitor memperkirakan rata-rata pertumbuhan tahunan (CAGR) penjualan online Indonesia selama 2014-2017 sebesar 38% (DBS Research, 2017).

Salah satu situs *e-commerce* terpopuler yang paling banyak dikunjungi di Indonesia menurut data Alexa per Oktober 2017-Januari 2018 adalah Bukalapak. Berdasarkan kunjungan ke situs Bukalapak berasal dari mesin pencari sebesar 34,30% dan total yang terhubung dengan situs Bukalapak sebesar 31.110.000 (Alexa, 2018). Bukalapak merupakan salah satu model *e-commerce Customer-To Customer* (C2C). Model ini yang paling banyak diterapkan dan di temukan pada situs *e-commerce* di Indonesia. Pasar *Customer-To Customer* (C2C) saat ini masih dominan di pasar ritel *online* Indonesia. Data yang dihimpun dari Euromonitor memperkirakan pasar C2C menyumbangkan 3% dari pasar ritel di Indonesia pada tahun 2017, sedangkan pasar B2C menyumbangkan 1,7% (Euromonitor, 2017). Selain itu juga masyarakat Indonesia lebih banyak memilih berbelanja *online* pada model C2C karena banyak terdapat pilihan produk. Oleh sebab itulah pada penelitian ini akan menganalisis model C2C.

*Google Play* adalah layanan konten digital milik *Google* yang terdiri dari toko produk-produk *online* seperti musik/lagu, buku, aplikasi, permainan, ataupun pemutar media berbasis awan. *Google Play* dapat diakses melalui web, aplikasi *android* (*Play Store*), dan *Google TV*. Dalam *Google Play* dilengkapi dengan adanya fitur berisi ulasan dari para pengguna yang dapat digunakan untuk melihat ulasan dari pengguna aplikasi.

Ulasan dari pengguna sering digunakan sebagai alat yang efektif dan efisien dalam menemukan informasi terhadap suatu produk atau jasa. Menurut Sung (dikutip dalam Fanani, 2017), bahwa penelitian baru-baru ini menemukan hampir 50% dari pengguna internet bergantung pada rekomendasi *word-of-mouth* (opini) sebelum menggunakan suatu produk, karena *review* dari pengguna lain dapat menyediakan informasi terbaru dari produk tersebut berdasarkan perspektif pengguna-pengguna lain yang sudah menggunakan produk tersebut.

Pelanggan atau *klien* yang merasa tidak puas dengan layanan atau produk yang ditawarkan oleh sebuah perusahaan biasanya akan menuliskan keluhannya di media sosial. Di sisi lain, ada juga pelanggan yang merasa puas, yang mengekspresikan sikap positif mereka terhadap sebuah produk di media sosial. Disadari atau tidak, opini-opini pelanggan yang dituliskan di media sosial, sedikit atau pun banyak, akan memberikan pengaruh pada calon pelanggan. Akan tetapi, memantau dan mengorganisasi opini dari masyarakat juga bukanlah hal yang mudah. Opini yang dimuat di media sosial jumlahnya terlalu banyak untuk diproses secara manual. Oleh sebab itulah, diperlukan sebuah metode atau teknik khusus yang mampu mengkategorikan *review-review* tersebut secara otomatis, apakah termasuk positif atau negatif, berdasarkan sebuah property.

Jumlah data ulasan pengguna aplikasi Bukalapak yang masuk ke situs *Google Play* terus bertambah seiring berjalannya waktu, hal ini mengakibatkan sulitnya pihak perusahaan dalam memperoleh informasi secara keseluruhan dari semua ulasan, karena akan membutuhkan waktu yang lama untuk membaca satu persatu setiap ulasan yang masuk pada halaman situs *Google Play*.

Banyak ulasan pengguna di *Google Play* mengenai aplikasi Bukalapak. Citra merek yang baik akan membentuk opini yang baik pula dari konsumen tentang

suatu produk/jasa, dan diharapkan akan mendorong terjadinya proses pembelian oleh konsumen, dan begitu sebaliknya. Berbagai macam tanggapan di situs *Google Play* tentu saja akan mempengaruhi citra dari Bukalapak. Tanggapan negatif maupun positif dari pengguna bisa jadi dipengaruhi oleh beberapa hal yang belum menjadi perhatian dari Bukalapak. Hal ini mungkin terjadi karena adanya beberapa faktor yang harus diperbaiki dan belum diketahui oleh Bukalapak. Dengan menggunakan *text mining* dapat dilihat pembicaraan apa saja yang sering dibahas oleh pengguna.

Salah satu analisis *text mining* yaitu analisis sentimen dapat diaplikasikan pada perusahaan yang mengeluarkan suatu produk dan jasa dan menyediakan layanan untuk menerima pendapat (*feedback*) dari konsumen untuk produk tersebut. Analisis sentiment diaplikasikan untuk mengelompokkan *feedback* positif, negatif, dan netral dari konsumen sehingga mempercepat dan mempermudah tugas perusahaan untuk meninjau kembali kekurangan produk mereka. Apabila ditemukan adanya sentiment negatif, maka perusahaan dapat dengan cepat mengambil tindakan untuk menanggulangnya.

Dari serangkaian latar belakang yang telah diuraikan tersebut, peneliti merasa perlu untuk melakukan analisis lebih lanjut mengenai ulasan pengguna Bukalapak di *Google Play* untuk mengetahui bagaimana opini pengguna terhadap Bukalapak. Peneliti akan mengklasifikasikan ulasan pengguna aplikasi Bukalapak apakah termasuk ulasan positif atau negatif untuk bahan evaluasi dari *e-commerce* Bukalapak biasa disebut dengan *sentiment analysis* dengan metode *Support Vector Machine (SVM)* dan *Maximum Entropy (Maxent)*. Proses klasifikasi dilakukan menggunakan algoritma SVM karena memiliki tingkat akurasi paling tinggi dalam hal klasifikasi teks (Naradhipa & Purwarianti, 2012). Sedangkan metode *Maximum Entropy* mampu mencari distribusi  $p(a|b)$  yang akan memberikan nilai *entropy maksimum* dengan tujuan mendapatkan distribusi probabilitas terbaik yang paling mendekati kenyataan. Berdasarkan alasan tersebut, maka peneliti memilih menggunakan metode *Support Vector Machine* dan *Maximum Entropy* untuk mengklasifikasikan ulasan bahasa Indonesia tentang aplikasi Bukalapak.

Setelah melakukan klasifikasi, penulis mencoba mengekstrak dan mengeksplorasi seluas-luasnya informasi apa yang ada pada setiap klasifikasi sentimen positif maupun sentimen negatif yang sekiranya dianggap penting untuk digunakan pada berbagai keperluan. Pada proses ekstraksi dan eksplorasi informasi, penulis menggunakan statistik deskriptif dan asosiasi antar kata untuk menemukan topik yang sering dibicarakan oleh pengguna. Harapannya, penelitian ini mampu mengklasifikasikan teks ulasan 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 yang telah diuraikan, maka dalam penelitian ini permasalahan yang akan dirumuskan diantaranya adalah:

1. Bagaimana gambaran umum data ulasan tentang Bukalapak berdasarkan situs *Google Play* ?
2. Seberapa tepat *machine learning* dengan menggunakan metode *Support Vector Machine (SVM)* dan *Maximum Entropy (Maxent)* dalam mengklasifikasikan teks berbahasa Indonesia mengenai ulasan pengguna Bukalapak berdasarkan situs *Google Play*?
3. Informasi apa yang didapatkan dalam setiap klasifikasi yang telah dilakukan?
4. Berdasarkan diagram *fishbone*, faktor-faktor apa saja yang harus dilakukan untuk memperbaiki dari hasil ulasan negatif yang didapat?

### **1.3. Batasan Masalah**

Batasan masalah yang ditentukan untuk menghindari perluasan pembahasan dalam penelitian ini adalah sebagai berikut:

1. Data yang akan diklasifikasi hanya data *review* pengguna Bukalapak pada *Google Play*.
2. *Review* pada aplikasi yang akan diklasifikasi adalah yang berbahasa Indonesia.
3. Analisis yang digunakan untuk analisis sentimen adalah metode *SVM* dan *Maxent*.

#### **1.4. Tujuan Penelitian**

Tujuan dari penelitian ini adalah sebagai berikut :

1. Mengetahui gambaran umum data ulasan tentang Bukalapak berdasarkan situs *Google Play*.
2. Mengetahui seberapa tepat *machine learning* dengan menggunakan metode *Support Vector Machine (SVM)* dan *Maximum Entropy (Maxent)* dalam mengklasifikasikan teks berbahasa Indonesia mengenai ulasan pengguna Bukalapak berdasarkan situs *Google Play*.
3. Mendapatkan informasi penting dalam setiap klasifikasi.
4. Mendapatkan informasi faktor-faktor apa saja yang harus dilakukan untuk memperbaiki hasil dari ulasan negatif yang didapat.

#### **1.5. Manfaat Penelitian**

Manfaat dari penelitian ini adalah sebagai berikut:

1. Mengetahui gambaran umum data ulasan tentang aplikasi Bukalapak berdasarkan situs *Google Play*.
2. Pengklasifikasian data ulasan tentang Bukalapak dapat memudahkan pihak perusahaan dalam mengetahui persepsi pengguna jasa dalam bentuk opini negatif dan opini positif, sehingga dapat dijadikan sebagai acuan dalam upaya menjaga kualitas dan memperbaiki kekurangan serta evaluasi ke arah 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 ini akan dibahas tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

##### **BAB II TINJAUAN PUSTAKA**

Bab ini memaparkan penelitian-penelitian terdahulu yang berhubungan dengan permasalahan yang diteliti dan menjadi acuan konseptual.

**BAB III LANDASAN TEORI**

Pada bab ini akan dibahas tentang teori-teori dan konsep yang berhubungan dengan penelitian yang dilakukan dan mendukung dalam pemecahan masalahnya. Selain itu, bab ini juga memuat teori-teori dalam pelaksanaan pengumpulan dan pengolahan data serta saat melakukan penganalisaan.

**BAB IV METODOLOGI PENELITIAN**

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

**BAB V ANALISIS DAN PEMBAHASAN**

Pada bab ini akan dibahas mengenai analisa yang dilakukan terhadap hasil pengumpulan, pengolahan dan analisa data yang diperoleh dari hasil penelitian.

**BAB VI PENUTUP**

Pada bab ini akan dibahas mengenai kesimpulan yang diperoleh dari hasil penelitian dan analisa data yang telah dilakukan serta saran-saran yang dapat diterapkan dari hasil pengolahan data yang dapat menjadi masukan yang berguna kedepannya.

## **BAB II**

### **TINJAUAN PUSTAKA**

Penelitian terdahulu sangatlah penting bagi penulis sebagai kajian untuk mengetahui keterkaitan antara penelitian terdahulu dengan penelitian yang akan dilakukan, untuk menghindari terjadinya tindakan duplikasi yang dilakukan oleh penulis. Tujuan dari tinjauan pustaka ini adalah untuk menunjukkan bahwa penelitian yang dilakukan penulis sangatlah bermanfaat dan mempunyai arti penting sehingga dapat diketahui kontribusi penelitian terhadap ilmu pengetahuan. Berikut beberapa ulasan tentang penelitian-penelitian terdahulu yang pernah dilakukan sebelumnya berkenaan dengan data dan metode yang digunakan. Beberapa jurnal dan penelitian yang penulis jadikan sebagai acuan adalah sebagai berikut.

Dalam penelitian yang dilakukan oleh Pang (2002) tentang klasifikasi sentiment menggunakan teknik *Machine Learning*. Pada penelitian ini menggunakan metode *Naïve Bayes Classification* (NBC), *Maximum Entropy* (ME), dan *Support Vector Machine* (SVM) untuk mengklasifikasikan *review* film ke dalam kelas positif dan negatif. Hasil eksperimen menunjukkan metode SVM memiliki tingkat akurasi yang paling tinggi dibanding dengan metode lainnya, yakni sebesar 82,7%.

Saraswati (2011) mengklasifikasikan opini dengan menggunakan metode NBC dan SVM. Dari hasil penelitiannya, Saraswati menyatakan bahwa metode SVM memiliki tingkat akurasi yang lebih tinggi daripada metode NBC untuk pengujian data opini positif, sedangkan metode NBC menunjukkan hasil yang lebih baik jika digunakan dalam pengujian data opini negatif.

Putranti dan Winarko (2014) tentang analisis sentimen dalam penelitian ini merupakan proses klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen positif dan negatif. Data opini diperoleh dari jejaring sosial *Twitter* berdasarkan *query* dalam Bahasa Indonesia. Penelitian ini bertujuan

untuk menentukan sentimen publik terhadap objek tertentu yang disampaikan di *Twitter* dalam bahasa Indonesia, sehingga membantu usaha untuk melakukan riset pasar atas opini publik. *Algoritma Maxent* digunakan untuk POS tagger dan *SVM* digunakan untuk membangun model klasifikasi. Implementasi klasifikasi diperoleh akurasi 86,81 % pada pengujian *7 fold cross validation* untuk tipe *kernel Sigmoid*. Pelabelan kelas secara manual dengan POS tagger menghasilkan akurasi 81,67%.

Penelitian lain juga dilakukan oleh Gusriani, dkk (2016) tentang analisis sentimen berdasarkan komentar publik terhadap toko online seperti Zalora dan Berry Benka pada media sosial facebook. Metodologi yang digunakan untuk melakukan analisis sentimen dimulai dari data *collecting, preprocessing, feature selection*, klasifikasi dan pengukuran akurasi. Metode klasifikasi *Naive Bayes*, K-NN dan *Decision Tree* digunakan untuk membandingkan hasil prediksi klasifikasi yang terbaik. Hasil analisis pengujian menunjukkan *Naive Bayes*, memiliki kestabilan akurasi setelah diuji dengan beberapa nilai *Frequent Itemset*. *Naive Bayes* memiliki rata-rata akurasi 90,3%.

Ulwan (2016) melakukan penelitian menggunakan *machine learning* dengan metode *Support Vector Machine (SVM)* mengklasifikasikan data teks laporan masyarakat yang diperoleh dari situs LAPOR!. Data teks yang tidak terstruktur (*unstructured data*) tersebut diklasifikasikan menjadi tiga kelas yaitu Aspirasi, Keluhan, dan Pertanyaan. Selanjutnya hasil klasifikasi dianalisis dengan metode *text mining*. Hasil klasifikasi menunjukkan tingkat akurasi sebesar 96,7%. Secara umum metode *text mining* menunjukkan hasil ekstraksi informasi pada kelas aspirasi adalah terkait penertiban terhadap psk, pkl, asap, merokok, busway, dan pembagian bantuan masyarakat. Pada kelas keluhan masyarakat mengeluhkan tentang pembagian BLSM atau KPS yang tidak merata, masalah macet, layanan Telkom yang buruk, serta busway yang sering bermasalah. Sedangkan pada kelas pertanyaan yang menjadi hal yang sering ditanyakan adalah masalah BLSM dan KPS serta seputar informasi mengenai agama, BPJS, beasiswa, sertifikasi dan tunjangan.

Dalam penelitian yang dilakukan oleh Saviera (2017) yang berkaitan dengan analisis sentimen untuk mengetahui bagaimana perbandingan kepuasan pelanggan terhadap tiga situs *e-commerce* yang sering dikunjungi di Indonesia yaitu Bukalapak, Tokopedia dan Elevenia. Data yang didapatkan dari media sosial twitter, dalam penelitiannya *Naïve Bayes Classifier* digunakan sebagai teknik klasifikasi dengan pembobotan TF-IDF , sedangkan untuk melakukan validasi dan evaluasi pada klasifikasi teks *Naïve Bayes* dilakukan menggunakan *K-fold cross validation* dan *confusion matrix*. Hasil penelitian menunjukkan bahwa sentimen negatif terhadap ketiga situs *e-commerce* tersebut lebih mendominasi di media sosial, dan situs *e-commerce* dengan sentimen negatif paling tinggi ialah Bukalapak kemudian Tokopedia dan Elevenia.

Penelitian tentang analisis sentimen juga dilakukan oleh Fanani (2017) dengan mengklasifikasikan *review-review* pengguna aplikasi *mobile banking* yang ada di Indonesia tersebut sebagai *review SPAM* atau bukan *SPAM (HAM)*. Review tersebut diambil dari situs *Google Play* dengan cara *web scraping* dan *Heedzy.com* sebagai pihak ke-3. Analisis sentimen dilakukan dengan Python 2.7. Dalam penelitian ini Algoritma yang digunakan dalam klasifikasi adalah algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes Classifier (NBC)*. Pengujian akhir sistem dilakukan dengan menguji dan membandingkan performa dari algoritma *Support Vector Machine* dan *Naïve Bayes Classifier*. Hasil akhir dari penelitian ini menunjukkan performa sistem klasifikasi dengan algoritma *Support Vector Machine* menghasilkan akurasi lebih baik dibandingkan dengan menggunakan algoritma *Naïve Bayes Classifier*.

Penelitian yang dilakukan oleh Abtohi (2017) dalam Tugas Akhirnya yang berjudul Implementasi Teknik *Web Scraping* dan Klasifikasi *Support Vector Machine* dan Asosiasi (Studi Kasus: Data Ulasan Hotel Royal Ambarrukmo Pada Situs Tripadvisor), menyebutkan bahwa Klasifikasi dengan metode SVM menunjukkan tingkat akurasi sebesar 95,27% pada ulasan berbahasa Inggris dan sebesar 95,00% untuk ulasan berbahasa Indonesia. Hasil klasifikasi menunjukkan perbandingan jumlah yang cukup jauh antara kelas sentimen positif dan kelas sentimen negatif, kelas sentimen positif memiliki jumlah yang lebih besar

dibandingkan dengan kelas sentimen negatif. Secara umum, dengan metode *text mining* diperoleh informasi bahwa lebih banyak pengunjung yang memberikan penilaian positif daripada pengunjung yang memberikan penilaian negatif. Sentimen positif pengunjung diantaranya adalah tentang penilaian kamar hotel yang luas, bersih dan terdapat balkon, staff yang ramah, makanan yang beranekaragam mulai dari kelas tradisional hingga kelas internasional, dan lokasi hotel yang berdekatan dengan mall. Sedangkan sentimen negatif pengunjung diantaranya berupa keluhan tentang kekecewaan dan ketidakpuasan pengunjung terhadap fasilitas kamar yang bermasalah, dan pelayanan *check-in* yang dinilai lambat dan buruk.

Pada **Tabel 2.1** menjadi perbandingan penelitian sebelumnya dan penelitian yang akan dilakukan oleh penulis.

**Tabel 2.1** Perbandingan penelitian sebelumnya dengan penelitian yang penulis lakukan

| <b>Judul Penelitian</b>   | <b>Peneliti, tahun</b> | <b>Metode</b>  | <b>Hasil Penelitian</b>   |
|---|------------------------|--|---|
| <i>Thumbs up? Sentiment Classification Using Machine Learning Techniques.</i> | Pang (2002)            | <i>Naïve Bayes (NBC), Maximum Entropy (ME), dan Support Vector Machine (SVM)</i> | Mengklasifikasikan review film ke dalam kelas positif dan negatif. Hasil eksperimen Pang menunjukkan metode SVM memiliki tingkat akurasi yang paling tinggi dibanding dengan metode lainnya, yakni sebesar 82,7%. |

| <b>Judul Penelitian</b>   | <b>Peneliti, tahun</b>      | <b>Metode</b>  | <b>Penelitian</b>   |
|---|-----------------------------|--|---|
| <i>Text Mining</i> dengan Metode <i>Naive Bayes Classifier</i> dan <i>Support Vector Machines</i> untuk <i>Sentiment Analysis</i> | Saraswati. (2011)           | <i>NBC</i> dan <i>SVM</i>                                | Metode <i>SVM</i> memiliki tingkat akurasi yang lebih tinggi daripada metode <i>NBC</i> untuk pengujian data opini positif.   |
| Analisis Sentimen <i>Twitter</i> untuk Teks Berbahasa Indonesia dengan <i>Maximum Entropy</i> dan <i>Support Vector Machine</i>   | Putranti dan Winarko (2014) | <i>Maximum Entropy</i> dan <i>Support Vector Machine</i> | <i>Algoritma Maxent</i> digunakan untuk POS tagger dan <i>SVM</i> digunakan untuk membangun model klasifikasi. Tingkat akurasi pada pengujian <i>7 fold cross validation</i> untuk tipe <i>kernel Sigmoid</i> sebesar 86,81 % dan pelabelan kelas secara manual dengan POS tagger sebesar 81,67%. |
| Analisis Sentimen Berdasarkan Komentar Publik Terhadap Toko Online Pada Media Sosial Facebook (Studi Kasus:Zalora dan BerryBenka) | Gusriani, dkk (2016)        | <i>Naive Bayes</i> , K-NN dan <i>Decision Tree</i>       | Membandingkan hasil prediksi klasifikasi yang terbaik. Hasil analisis pengujian menunjukkan <i>Naive Bayes</i> , memiliki kestabilan akurasi setelah diuji dengan beberapa nilai <i>Frequent Itemset</i> . <i>Naive Bayes</i> memiliki rata-rata akurasi 90,3%.                                   |

| <b>Judul Penelitian</b>   | <b>Peneliti, tahun</b> | <b>Metode</b>                              | <b>Hasil Penelitian</b>  |
|---|------------------------|--|--|
| <i>Pattern Recognition Pada Unstructured Data Teks Menggunakan Support Vector Machine Dan Association</i>                                 | Ulwan<br>(2016)        | <i>SVM, Asosiasi kata</i>                  | Mengklasifikasikan laporan masyarakat berdasarkan keluhan, aspirasi, dan pertanyaan pada situs LAPOR! dan ekstraksi Informasi  |
| Klasifikasi <i>Review Software Pada Google Play Menggunakan Pendekatan Analisis Sentimen</i>  | Fanani<br>(2017)       | <i>SVM dan NBC</i>                         | Penelitian ini menunjukkan performa sistem klasifikasi <i>Review SPAM dan HAM</i> dengan algoritma <i>Support Vector Machine</i> menghasilkan akurasi lebih baik dibandingkan dengan menggunakan algoritma <i>Naive Bayes Classifier</i> . |
| Implementasi Teknik <i>Web Scraping</i> dan Klasifikasi Sentimen (Studi Kasus: Data Ulasan Hotel Royal Ambarrukmo Pada Situs Tripadvisor) | Abtohi<br>(2017)       | <i>Support Vector Machine dan Asosiasi</i> | Menerapkan teknik <i>web scraping</i> dalam mengumpulkan data ulasan dan melakukan klasifikasi berdasarkan sentimen positif dan negatif.   |

## **BAB III**

### **LANDASAN TEORI**

#### **3.1. *E-commerce***

##### **3.1.1. *Definisi E-commerce***

*Electronic commerce* (disingkat *e-commerce*) sebagai sarana berbisnis menggunakan jaringan komputer, sebenarnya adalah dikenal sejak 20 tahun lalu sejak akhir tahun 70-an dan awal tahun 80-an. Generasi pertama *e-commerce* dilakukan hanya antar perusahaan berupa transaksi jual beli yang difasilitasi oleh *Electronic Data Intechange* (EDI) dalam transaksi jual beli elektronik ini banyak aspek-aspek hukum yang bersentuhan langsung maupun tidak langsung (Firdaus, 2015).

*E-commerce* adalah proses pembelian dan penjualan antara dua belah pihak di dalam suatu perusahaan dengan adanya pertukaran barang, jasa, atau informasi melalui media internet (Indrajit, 2001). Onno (2000) memberikan pengertian tentang *e-commerce* yaitu asset dinamis teknologi, aplikasi, dan proses bisnis yang menghubungkan perusahaan, konsumen dan komunitas melalui elektronik dan perdagangan barang, pelayanan dan informasi yang dilakukan secara elektronik. Sedangkan menurut Berkatulloh dan Prasetyo (2005) menjelaskan bahwa *e-commerce* adalah kegiatan-kegiatan bisnis yang menyangkut konsumen (*consumers*), manufaktur (*manufactures*), *service providers* dan pedagang perantara (*intermediaries*), dengan menggunakan jaringan-jaringan computer (*computer networks*) yaitu internet.

##### **3.1.2. *Komponen E-commerce***

Pada *e-commerce* terdapat mekanisme-mekanisme tertentu yang unik dan berbeda dibandingkan dengan mekanisme-mekanisme yang terdapat pada *traditional commerce*. Dalam mekanisme pasar *e-commerce*, terdapat beberapa komponen yang terlibat, yakni (Turban & King, 2002):

a) *Customer*

*Customer* merupakan para pengguna Internet yang dapat dijadikan sebagai target pasar yang potensial untuk diberikan penawaran berupa produk, jasa, atau informasi oleh para penjual.

## b) Penjual

Penjual merupakan pihak yang menawarkan produk, jasa, atau informasi kepada para *customer* baik individu maupun organisasi. Proses penjualan dapat dilakukan secara langsung melalui *website* yang dimiliki oleh penjual tersebut melalui *marketplace*.

## c) Produk

Salah satu perbedaan antara *e-commerce* dengan *traditional commerce* terletak pada produk yang dijual. Pada dunia maya, penjual dapat menjual produk *digital* yang dapat dikirimkan secara langsung melalui Internet.

## d) Infrastruktur

Infrastruktur pasar yang menggunakan media elektronik meliputi perangkat keras, perangkat lunak dan juga sistem jaringannya.

e) *Front end*

*Front end* merupakan aplikasi web yang dapat berinteraksi dengan pengguna secara langsung. Beberapa proses bisnis pada *front end* ini antara lain: portal penjual, katalog elektronik, *shopping cart*, mesin pencari dan *payment gateway*.

f) *Back end*

*Back end* merupakan aplikasi yang secara tidak langsung mendukung aplikasi *front end*. Semua aktivitas yang berkaitan dengan pemesanan barang, manajemen inventori, proses pembayaran, *packaging* dan pengiriman barang termasuk dalam bisnis proses *back end*.

g) *Intermediary*

*Intermediary* merupakan pihak ketiga yang menjembatani antara produsen dengan konsumen. *Online intermediary* membantu mempertemukan pembeli dan penjual, menyediakan infrastruktur, serta membantu penjual dan pembeli dalam menyelesaikan proses transaksi. *Intermediary* tidak hanya perusahaan

atau organisasi tetapi dapat juga individu. Contoh *intermediary* misalnya broker dan distributor.

h) *Partner* bisnis lain

*Partner* bisnis merupakan pihak selain *intermediary* yang melakukan kolaborasi dengan produsen.

i) *Support services*

Ada banyak *support services* yang saat ini beredar di dunia maya mulai dari sertifikasi dan *trust service*, yang menjamin keamanan sampai pada *knowladge provider*.

### 3.1.3. Model-model E-commerce di Indonesia

Situs *e-commerce* yang ada di Indonesia dapat dikategorikan berdasarkan model bisnisnya. Berikut adalah lima model bisnis yang diusung oleh pelaku bisnis *e-commerce* di Indonesia menurut id.techinasia.com (dikutip dalam Aprilia, 2017):

1) *Classifieds*/listing/iklan baris

Iklan baris adalah model bisnis *e-commerce* paling sederhana yang cocok digunakan di negara-negara berkembang. Dua kriteria yang biasa diusung model bisnis ini:

- a) *Website* yang bersangkutan tidak memfasilitasi kegiatan transaksi online
- b) Penjual individual dapat menjual barang kapan saja, dimana saja secara gratis

Tiga situs iklan baris yang terkenal di Indonesia ialah Tokobagus, Berniaga, dan OLX. Kaskus selaku forum *online* terbesar di Indonesia juga dapat dikatakan masih menggunakan model bisnis iklan baris di forum jual belinya. Ini dikarenakan Kaskus tidak mengharuskan penjualnya untuk menggunakan fasilitas rekening bersama atau *escrow*. Jadi transaksi masih dapat terjadi langsung antara penjual dan pembeli. Metode transaksi yang paling sering digunakan di situs iklan baris ialah metode *cash on delivery* atau COD. Cara model bisnis *e-commerce* ini meraup keuntungan adalah dengan pemberlakuan iklan premium. Situs iklan baris seperti ini cocok bagi penjual

yang hanya ingin menjual sekali-kali saja, seperti barang bekas atau barang yang stoknya sedikit.

2) *Marketplace C2C (Customer-To-Customer)*

*Marketplace C2C* adalah model bisnis dimana *website* yang bersangkutan tidak hanya membantu mempromosikan barang dagangan saja, tapi juga memfasilitasi transaksi uang secara *online*. Berikut ialah indikator utama bagi sebuah *website marketplace*:

- a) Seluruh transaksi *online* harus difasilitasi oleh *website* yang bersangkutan
- b) Bisa digunakan oleh penjual individual

Kegiatan jual beli di *website marketplace* harus menggunakan fasilitas transaksi *online* seperti layanan *escrow* atau rekening pihak ketiga untuk menjamin keamanan transaksi. Penjual hanya akan menerima uang pembayaran setelah barang diterima oleh pembeli. Selama barang belum sampai, uang akan disimpan di rekening pihak ketiga. Apabila transaksi gagal, maka uang akan dikembalikan ke tangan pembeli.

Tiga situs *marketplace* di Indonesia yang memperbolehkan penjual langsung berjualan barang di *website* ialah Tokopedia, Bukalapak, dan Lamido. Ada juga situs *marketplace* lainnya yang mengharuskan penjual menyelesaikan proses verifikasi terlebih dahulu seperti Blanja dan Elevenia. Cara model bisnis *e-commerce* ini meraup keuntungan adalah dengan memberlakukan layanan penjual premium, iklan premium, dan komisi dari setiap transaksi. Situs *marketplace* seperti ini lebih cocok bagi penjual yang lebih serius dalam berjualan online. Biasanya penjual memiliki jumlah stok barang yang cukup besar dan mungkin sudah memiliki toko fisik.

3) *Shopping mall*

Model bisnis ini mirip sekali dengan *marketplace*, tapi penjual yang bisa berjualan di sana haruslah penjual atau brand ternama karena proses verifikasi yang ketat. Satu-satunya situs *online shopping mall* yang beroperasi di Indonesia ialah Blibli. Cara model bisnis *e-commerce* ini meraup keuntungan adalah dengan adanya komisi dari penjual.

#### 4) Toko online B2C (*Business-To-Consumer*)

Model bisnis ini cukup sederhana, yakni sebuah toko *online* dengan alamat *website (domain)* sendiri dimana penjual memiliki stok produk dan menjualnya secara *online* kepada pembeli. Beberapa contohnya di Indonesia ialah Bhinneka, Lazada Indonesia, BerryBenka, dan Bilna 1. Tiket.com yang berfungsi sebagai *platform* jualan tiket secara *online* juga bisa dianggap sebagai toko *online*. Keuntungannya bagi pemilik toko *online* ialah ia memiliki kebebasan penuh disana. Pemilik dapat mengubah jenis tampilan sesuai dengan preferensinya dan dapat membuat blog untuk memperkuat SEO toko onlinenya. Model bisnis *e-commerce* ini mendapatkan profit dari penjualan produk. Model bisnis ini cocok bagi yang serius berjualan *online* dan siap mengalokasikan sumber daya yang dimiliki untuk mengelola situs sendiri.

#### 5) Toko *online* di media sosial

Banyak penjual di Indonesia yang menggunakan situs media sosial seperti Facebook dan Instagram untuk mempromosikan barang dagangan mereka. Uniknyalagi, sudah ada pemain-pemain lokal yang membantu penjual untuk berjualan di situs Facebook yakni Onigi dan LakuBgt. Ada juga *startup* yang mengumpulkan seluruh penjual di Instagram ke dalam satu *website* yakni Shopious. Membuat toko *online* di Facebook atau Instagram sangatlah mudah, sederhana, dan gratis. Namun, penjual tidak dapat membuat templatnya sendiri. Di Indonesia, chanel BBM pun juga sering digunakan sebagai media jual beli barang. Penjual yang menggunakan model bisnis ini biasanya penjual yang ingin memiliki toko *online* sendiri tapi tidak ingin terlalu direpotkan. Ada juga beberapa bisnis *online* yang menggunakan beberapa model bisnis diatas pada saat bersamaan. Dua contohnya ialah Qoo10 dan Rakuten Belanja Online yang memiliki toko *online* B2C mereka sendiri serta *marketplace* yang memverifikasi penjualnya terlebih dahulu.

### 3.2. Bukalapak

Bukalapak merupakan salah satu *online marketplace* terkemuka di Indonesia yang dimiliki dan dijalankan oleh PT. Bukalapak dengan model bisnis *Customer-To-Customer (C2C)*. Bukalapak menyediakan sarana penjualan dari konsumen ke konsumen di mana pun. Siapapun bisa membuka toko *online* untuk kemudian melayani calon pembeli dari seluruh Indonesia baik satuan ataupun dalam jumlah banyak. Pengguna perorangan ataupun perusahaan dapat membeli dan menjual produk, baik baru maupun bekas, seperti sepeda, ponsel, perlengkapan bayi, gawai (gadget), aksesoris gawai, komputer, sabak (*tablet*), perlengkapan rumah tangga, busana, elektronik, dan lain-lain.

Sistem pembayaran transaksi yang dimiliki Bukalapak adalah jaminan keamanan transaksi jual beli dalam sistem pembayaran yang dikenal juga dengan BukaDompet. Berbeda dengan situs yang berkembang pada tahun 2000-an yang umumnya berupa iklan dan memperbolehkan penjual dan pembeli untuk berkomunikasi secara langsung lewat telepon, namun di Bukalapak, penjual dan pembeli tidak diperkenankan untuk berkomunikasi secara langsung karena berpotensi terjadinya penipuan.

Dalam hal ini Bukalapak akan menjadi pihak ketiga yang menengahi transaksi antara penjual dan pembeli. Ketika calon pembeli ingin membeli sebuah barang dari penjual di Bukalapak, maka pembeli harus melakukan transfer pembayaran ke Bukalapak terlebih dahulu. Jika transfer telah berhasil, Bukalapak akan memberi tahu penjual bahwa pembayaran sudah diterima oleh Bukalapak dan penjual bisa melakukan pengiriman barang yang sudah dipesan pembeli melalui pesan sms. Ketika barang tiba di pembeli, pembeli melakukan konfirmasi penerimaan barang kepada Bukalapak, dan Bukalapak akan melakukan transfer uang pembelian kepada penjual. Dengan program jaminan ini, bila pembeli tidak menerima barang sampai batas waktu tertentu, dana pembeli akan dikembalikan 100% (Bukalapak, 2016).

### 3.3. *Google Play*

*Google Play* adalah layanan konten digital milik *Google* yang melingkupi toko *online* untuk produk-produk seperti musik/lagu, buku, aplikasi, permainan, ataupun pemutar media berbasis *cloud*. Layanan ini dapat diakses baik melalui *web*, aplikasi android (*Play Store*) dan *Google TV*. *Google Play* mulai dikenalkan pada bulan Maret 2012 sebagai pengganti dari *Google Play* dan layanan musik *Google*.

*Google Play* milik *Google* saat ini telah menyediakan sekitar 700.000 aplikasi *mobile* menurut AppBrain7. Setelah beberapa bulan, mungkin ada lebih dari sepuluh ribu komentar tekstual dari aplikasi baru yang diluncurkan di *Google Play* (Liu et al, 2013). Hal ini sangat sesuai untuk penerapan analisis sentimen terhadap komentar-komentar pengguna tentang aplikasi pada *Google Play* untuk membantu pengguna dalam memberikan pertimbangan untuk menginstall aplikasi dan *developer* untuk menganalisis aplikasinya.

### 3.4. *Online Review*

Dalam Kamus Besar Bahasa Indonesia, *review* yang diterjemahkan ke dalam bahasa Indonesia menjadi ulasan memiliki pengertian kupasan, tafsiran, komentar, sedangkan kata *online* memiliki arti *controlled by or connected to a computer* sesuai dengan pengertian yang dijelaskan dalam *Oxford Dictionaries online* yang apabila diterjemahkan yaitu dikontrol oleh atau terkoneksi pada sebuah komputer. Menurut Mudambi dan Schuff dalam jurnal yang berjudul “A Study of Customer Review on Amazon.com, “*Online customer reviews can be defined as peer-generated product evaluations posted on company or third party websites*”. Dalam terjemahannya yaitu *online review* dari pelanggan diartikan sebagai hasil evaluasi produk yang diunduh dalam situs perusahaan atau situs pihak ketiga.

Menurut Kotler dan Amstrong (1996) ada empat metode untuk mengukur kepuasan pelanggan yang salah satunya adalah sistem keluhan dan saran melalui berbagai media seperti kartu komentar dan juga internet seperti *online review* yang ada dalam situs-situs *e-commerce*.

### 3.5. Kepuasan Pelanggan

Kepuasan (*Satisfaction*) adalah perasaan senang atau kecewa seseorang yang muncul setelah membandingkan kinerja (hasil) produk yang dipikirkan terhadap kinerja atau hasil yang diharapkan. Jika kinerja berada di bawah harapan maka pelanggan tidak puas, jika kinerja memenuhi harapan maka pelanggan puas, jika kinerja melebihi harapan maka pelanggan amat puas atau senang (Kotler, 2006). Jadi, kepuasan merupakan fungsi dari persepsi atau kesan atas kinerja dan harapan.

Menurut Lovelock dan Wirtz (2011), kepuasan adalah suatu sikap yang diputuskan berdasarkan pengalaman yang didapatkan. Kepuasan merupakan penilaian mengenai ciri atau keistimewaan produk atau jasa pada tingkat kesenangan konsumen berkaitan dengan pemenuhan kebutuhan konsumsi konsumen. Kepuasan konsumen dapat diciptakan melalui kualitas, pelayanan dan nilai. Kunci untuk menghasikan kesetiaan pelanggan adalah memberikan nilai pelanggan yang tinggi.

Secara umum, kepuasan konsumen merupakan kesesuaian antara yang diinginkan dan diharapkan pelanggan dengan kenyataan yang diperoleh. Menurut Tjiptono (2006), terdapat tiga komponen utama dalam definisi kepuasan pelanggan, yaitu:

1. Tipe respon (respon emosional/afektif maupun kognitif) dan intensitas respon (kuat hingga lemah, biasanya dicerminkan melalui istilah-istilah seperti “sangat puas”, “netral”, “sangat senang”, “frustasi”, dan lain-lain).
2. Fokus respon, seperti produk, konsumsi, keputusan pembelian, wiraniaga, toko, dan lain-lain.
3. *Timing responds*, yaitu setelah konsumsi, setelah pilihan pembelian, berdasarkan pengalaman akumulatif, dan lain-lain.

Untuk mengukur kepuasan pelanggan, terdapat empat metode yang dapat digunakan (Tjiptono, 2006), antara lain:

1. Sistem Keluhan dan Saran. Perusahaan menyediakan kesempatan dan akses yang mudah dan nyaman bagi pelanggan untuk menyampaikan saran, kritik,

pendapat, dan keluhan mereka. Media yang digunakan dapat berupa kotak saran, kartu komentar, saluran telepon khusus bebas pulsa, website, dan sebagainya.

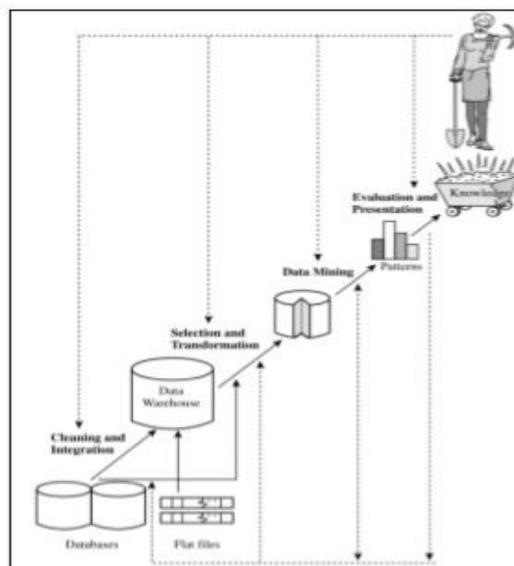
2. *Ghost Shopping (Mystery Shopping)*. Perusahaan mempekerjakan beberapa orang *ghost shopper* untuk berpura-pura sebagai pelanggan potensial produk perusahaan dan pesaing. Mereka diminta untuk berinteraksi secara langsung dengan staf penyedia jasa dan menggunakan produk/jasa perusahaan. Kemudian, mereka harus melaporkan temuannya mengenai kelebihan dan kekurangan produk perusahaan dan pesaing.
3. *Lost Customer Analysis*. Sebisa mungkin perusahaan menghubungi pelanggan-pelanggan yang sudah berhenti membeli atau berpindah pada pemasok lain agar dapat mengetahui penyebab terjadinya hal tersebut dan dapat melakukan perbaikan/penyempurnaan dikemudian hari, namun metode ini cukup sulit untuk dilaksanakan karena tidak semua mantan pelanggan bersedia untuk memberikan masukan dan evaluasi terhadap kinerja perusahaan.
4. Survei Kepuasan Pelanggan. Pengukuran kepuasan pelanggan melalui metode ini dapat dilakukan dengan beberapa cara, antara lain:
  - a. *Directly reported satisfaction*, yaitu menanyakan tingkat kepuasan secara langsung kepada pelanggan.
  - b. *Derived satisfaction*, yaitu memberikan pertanyaan yang mencakup dua hal, yaitu ekspektasi pelanggan terhadap kinerja produk atau perusahaan pada atribut-atribut relevan dan persepsi pelanggan terhadap kinerja aktual produk atau perusahaan yang bersangkutan.
  - c. *Problem analysis*, yaitu meminta kepada responden untuk menyampaikan masalah-masalah yang mereka hadapi terhadap produk atau jasa perusahaan, serta memberikan saran yang dapat membangun perusahaan untuk menjadi lebih baik lagi.
  - d. *Importance-performance analysis*, yaitu meminta kepada responden untuk menilai tingkat kepentingan berbagai atribut relevan dan tingkat kinerja perusahaan pada masing-masing atribut. Nilai rata-rata dari tingkat

kepentingan atribut dan kinerja perusahaan akan dianalisis di *Importance-Performance Matrix*.

### 3.6. Data Mining

*Data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan tiruan, dan *machine learning* untuk mengekstraksi serta mengidentifikasi informasi yang bermanfaat untuk pengetahuan yang terkait dari berbagai database besar (Turban et al, 2005). Menurut Tan et al (2006) *data mining* adalah proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga *knowledge discovery*.

*Data mining* adalah bagian integral dari penemuan pengetahuan dalam *database* (KDD) yang merupakan proses keseluruhan mengubah data mentah menjadi pola-pola data menarik yang merupakan informasi yang dibutuhkan oleh pengguna sebagai pengetahuan. Untuk mengetahui proses *knowledge discovery* dalam *database* bisa dilihat pada **Gambar 3.1** berikut :



**Gambar 3.1** Tahapan proses KDD (Han dan Kamber, 2006)

Han dan Kamber (2006) menyatakan bahwa KDD terdiri dari langkah-langkah sebagai berikut:

1. *Data cleaning* adalah proses menghapus data yang tidak konsisten dan menghilangkan *noise*.
2. *Data integration* adalah proses menggabungkan data apabila memiliki sumber data dalam sistem data mining tersebut.
3. *Data selection* adalah pengambilan data yang relevan yang akan digunakan dalam proses *data mining*.
4. *Data transformation* adalah proses dimana data ditransformasikan menjadi bentuk-bentuk yang sesuai untuk proses dalam *data mining*.
5. *Data mining* adalah suatu proses yang penting dengan melibatkan metode-metode untuk menghasilkan suatu pola data.
6. *Pattern evaluation* adalah proses untuk menguji kebenaran dari pola data yang mewakili *knowledge* yang ada didalam data itu sendiri.
7. *Knowledge representation* adalah proses visualisasi dan teknik menyajikan *knowledge* digunakan untuk menampilkan *knowledge* hasil *mining* kepada pengguna.

### **3.7. Machine Learning**

*Machine Learning* adalah bidang yang mempelajari pengembangan algoritma komputer untuk mengubah data menjadi aksi yang cerdas atau secara singkat dapat juga diartikan sebagai proses mengubah data menjadi informasi (Suyanto, 2017).

Menurut Mohri et al (2012) *machine learning* dapat didefinisikan sebagai metode komputasi berdasarkan pengalaman untuk meningkatkan performa atau membuat prediksi yang akurat. Definisi pengalaman disini ialah informasi sebelumnya yang telah tersedia dan bisa dijadikan data pembelajar. Dalam pembelajaran *machine learning*, terdapat beberapa skenario-skenario. Seperti:

### 1. *Supervised Learning*

Penggunaan skenario *supervised learning*, pembelajaran menggunakan masukan data pembelajaran yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label.

### 2. *Unsupervised Learning*

Penggunaan skenario *Unsupervised Learning*, pembelajaran menggunakan masukan data pembelajaran yang tidak diberi label. Setelah itu mencoba untuk mengelompokan data berdasarkan karakteristik-karakteristik yang ditemui.

### 3. *Reinforcement Learning*

Pada skenario *reinforcement learning* fase pembelajaran dan tes saling dicampur. Untuk mengumpulkan informasi pembelajar secara aktif dengan berinteraksi ke lingkungan sehingga untuk mendapatkan balasan untuk setiap aksi dari pembelajar.

## **3.8. *Text Mining***

### **3.8.1. *Pengertian Text Mining***

Feldman dan Sanger (2007) menyatakan *text mining* adalah sebuah proses pengetahuan intensif dimana pengguna berinteraksi dan bekerja dengan sekumpulan dokumen dengan menggunakan beberapa alat analisis. *Text mining* atau *text analytics* adalah istilah yang mendeskripsikan sebuah teknologi yang mampu menganalisis data teks semi-terstruktur maupun tidak terstruktur, hal inilah yang membedakannya dengan *data mining* dimana data mining mengolah data yang sifatnya terstruktur (Jamil, 2017).

Perbedaan antara *text mining* dengan *data mining* terletak pada sumber data yang digunakan. Dalam *text mining* pola-pola yang diekstrak dari data tekstual yang tidak terstruktur bukan berasal dari suatu *database*. Beberapa kesamaannya adalah data yang digunakan merupakan data besar dan data berdimensi tinggi dengan struktur yang terus berubah. Dalam *data mining* data yang diolah adalah data yang terstruktur dari proses *warehousing* sehingga lebih mudah diproses oleh mesin/komputer. Analisis teks lebih sulit karena teks biasanya hanya digunakan sebagai konsumsi manusia secara langsung bukan

digunakan untuk mesin/komputer. Ditambah struktur teks yang kompleks, struktur yang tidak lengkap, bahasa yang berbeda, dan arti yang tidak standar. Oleh sebab itu pada umumnya digunakan *Natural Language Processing* untuk analisis teks yang tidak berstruktur tersebut. Tahapan-tahapan dalam *text mining* secara umum adalah *text preprocessing* dan *feature selection* (Feldman & Sanger,2007). Dimana penjelasan dari tahap-tahap tersebut adalah sebagai berikut:

### **3.8.2. *Text preprocessing***

Dalam melakukan *text mining*, teks dokumen yang digunakan harus dipersiapkan terlebih dahulu, setelah itu baru dapat digunakan untuk proses utama. Proses mempersiapkan teks dokumen atau dataset mentah disebut juga dengan proses *text preprocessing*. *Text preprocessing* berfungsi untuk mengubah data teks yang tidak terstruktur atau sembarang menjadi data yang terstruktur. Secara umum proses yang dilakukan dalam tahapan *preprocessing* adalah sebagai berikut:

#### **a. *Spelling Normalization***

*Spelling Normalization* merupakan proses perbaikan atau substitusi kata-kata yang salah eja atau disingkat dalam bentuk tertentu. Substitusi kata dilakukan untuk menghindari jumlah perhitungan dimensi kata yang melebar. Perhitungan dimensi kata akan melebar jika kata yang salah eja atau disingkat tidak diubah karena kata tersebut sebenarnya mempunyai maksud dan arti yang sama tetapi akan dianggap sebagai entitas yang berbeda pada saat proses penyusunan matriks.

#### **b. *Case Folding***

*Case folding* adalah proses penyamaan *case* dalam sebuah dokumen. Hal ini dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (dalam hal ini huruf kecil atau *lowercase*).

c. *Tokenizing*

*Tokenizing* adalah proses memecah kalimat menjadi kata-kata yang dilakukan untuk menjadikan sebuah kalimat menjadi lebih bermakna. Tahap pertama yang dilakukan adalah normalisasi kata dengan mengubah semua karakter huruf menjadi huruf kecil atau *toLowerCase*. Proses tokenisasi diawali dengan menghilangkan delimiter-delimiter yaitu simbol dan tanda baca yang ada pada teks tersebut seperti @, \$, &, tanda titik (.), koma (,) tanda tanya (?), tanda seru (!). Tahap tokenisasi selanjutnya yaitu proses penguraian teks yang semula berupa kalimat-kalimat yang berisi kata-kata. Proses pemotongan string berdasarkan tiap kata yang menyusunnya, umumnya setiap kata akan terpisahkan dengan karakter spasi, proses tokenisasi mengandalkan karakter spasi pada dokumen teks untuk melakukan pemisahan. Hasil dari proses ini adalah kumpulan kata saja (Putri, 2016).

d. *Filtering*

Tahap filtrasi adalah tahap mengambil kata-kata penting dari hasil token. Algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata yang penting) dapat digunakan pada tahap ini. *Stopword* adalah kata-kata yang tidak deskriptif dan bukan merupakan kata penting dari suatu dokumen sehingga dapat dibuang. Contoh *stopword* adalah “yang”, “dan”, “di”, “dari” dan seterusnya (Putri, 2016). Dalam filtrasi ini menggunakan *stoplist/stopword* agar kata-kata yang kurang penting dan sering muncul dalam suatu dokumen dibuang sehingga hanya menyisakan kata-kata yang penting dan mempunyai arti yang diproses ke tahap selanjutnya.

### 3.8.3. *Feature Selection*

Tahap ini merupakan tahap lanjutan dari pengurangan dimensi. Walaupun di tahap sebelumnya sudah melakukan penghapusan kata-kata yang tidak deskriptif (*stopwords*), tidak semua kata-kata di dalam dokumen memiliki arti penting. Sehingga untuk mengurangi dimensi, pemilihan hanya dilakukan pada kata-kata yang relevan dan yang benar-benar mempresentasikan isi dari suatu

dokumen. Kata-kata yang dinilai penting dilihat dari intensitas kemunculan dan yang paling informatif dari keseluruhan.

### 3.9. Pembobotan Kata (*Term Weighting*)

Hal yang perlu diperhatikan dalam pencarian informasi dari koleksi dokumen yang heterogen adalah pembobotan *term*. Term dapat berupa kata, frase atau unit hasil indexing lainnya dalam suatu dokumen yang dapat digunakan untuk mengetahui konteks dari dokumen tersebut. Karena setiap kata memiliki tingkat kepentingan yang berbeda dalam dokumen, maka untuk setiap kata tersebut diberikan sebuah indikator, yaitu *term weight* (Zafikri, 2008).

Menurut Zafikri (2008) *term weighting* atau pembobotan *term* sangat dipengaruhi oleh hal-hal berikut ini :

#### 1. *Term Frequency (TF)*

*Term Frequency (TF)* yaitu faktor yang menentukan bobot *term* pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut. Nilai jumlah kemunculan suatu kata (*term frequency*) diperhitungkan dalam pemberian bobot terhadap suatu kata. Semakin besar jumlah kemunculan suatu term (tf tinggi) dalam dokumen, semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuaian yang semakin besar.

#### 2. *Inverse Document Frequency (IDF)*

*Inverse Document Frequency (IDF)* yaitu pengurangan dominansi *term* yang sering muncul di berbagai dokumen. Hal ini diperlukan karena *term* yang banyak muncul di berbagai dokumen, dapat dianggap sebagai *term* umum (*common term*) sehingga tidak penting nilainya. Sebaliknya faktor kejarangmunculan kata (*term scarcity*) dalam koleksi dokumen harus diperhatikan dalam pemberian bobot. Menurut Mandala (dalam Witten, 1999) ‘Kata yang muncul pada sedikit dokumen harus dipandang sebagai kata yang lebih penting (*uncommon terms*) daripada kata yang muncul pada banyak dokumen. Pembobotan akan memperhitungkan faktor kebalikan frekuensi dokumen yang mengandung suatu kata (*Inverse Document Frequency*).

Metode TF-IDF merupakan metode pembobotan *term* yang banyak digunakan sebagai metode pembandingan terhadap metode pembobotan baru. Pada metode ini, perhitungan bobot *term t* dalam sebuah dokumen dilakukan dengan mengalikan nilai *Term Frequency* dengan *Inverse Document Frequency*.

Pada *Term Frequency* (TF), terdapat beberapa jenis formula yang dapat digunakan yaitu (Zafikri, 2008):

1. *TF biner (binary TF)*, hanya memperhatikan apakah suatu kata ada atau tidak dalam dokumen, jika ada diberi nilai satu, jika tidak diberi nilai nol.
2. *TF murni (raw TF)*, nilai TF diberikan berdasarkan jumlah kemunculan suatu kata di dokumen. Contohnya, jika muncul lima kali maka kata tersebut akan bernilai lima.
3. *TF logaritmik*, hal ini untuk menghindari dominansi dokumen yang mengandung sedikit kata dalam *query*, namun mempunyai frekuensi yang tinggi.

$$tf = 1 + \log(tf) \quad (3.1)$$

4. *TF normalisasi*, menggunakan perbandingan antara frekuensi sebuah kata dengan jumlah keseluruhan kata pada dokumen.

$$tf = 0,5 + 0,5x \left( \frac{tf}{\max tf} \right) \quad (3.2)$$

*Inverse Document Frequency* (idf) dihitung dengan menggunakan formula:

$$idf_j = \log \left( \frac{D}{df_j} \right) \quad (3.3)$$

Dimana

$D$  : adalah jumlah semua dokumen dalam koleksi

$df_j$  : adalah jumlah dokumen yang mengandung term  $t_j$

Dengan demikian rumus umum untuk TF-IDF adalah penggabungan dari formula perhitungan *raw TF* dan formula IDF dengan cara mengalikan nilai *Term Frequency* (TF) dengan nilai *Inverse Document Frequency* (idf) :

$$w_{ij} = tf_{ij} \times idf_j$$

$$w_{ij} = tf_{ij} \times \log \frac{D}{df_j} \quad (3.4)$$

Keterangan :

$w_{ij}$  : adalah bobot *term*  $t_j$  terhadap dokumen  $d_i$

$tf_{ij}$  : adalah jumlah kemunculan *term*  $t_j$  dalam dokumen  $d_i$

### 3.9.1. Simulasi Pembobotan Kata dengan TF-IDF

Data teks terlebih dahulu akan diubah ke dalam bentuk vektor menggunakan teknik *Term Frequency–Inverse Document Frequency* (TF-IDF). Untuk simulasi TF-IDF akan dijelaskan perhitungan bobot dokumen terhadap *query* sesuai dengan rumus yang telah dijelaskan. Contoh kata/*query terms* (Q) yang digunakan adalah “membantu”, “aman”, dan “transaksi”.

Misal terdapat tiga buah dokumen yaitu :

Dokumen 1 ( $d_1$ ) = Sangat membantu saya dalam berjualan, suka aplikasi yang sangat efisien, praktis dan aman.

Dokumen 2 ( $d_2$ ) = Bukalapak sangat membantu dalam transaksi jual beli secara aman.

Dokumen 3 ( $d_3$ ) = Aplikasi jual beli yang aman dan fitur lengkap.

Ketika dokumen tersebut akan dimasukkan ke dalam tahap *preprocessing* sehingga kalimatnya mengalami perubahan sebagai berikut.

Dokumen 1 ( $d_1$ ) = membantu berjualan suka aplikasi efisien praktis aman

Dokumen 2 ( $d_2$ ) = membantu transaksi jual beli aman

Dokumen 3 ( $d_3$ ) = aplikasi jual beli aman fitur lengkap

Adapun beberapa *term (documents term)* yang didapatkan dari ketiga dokumen tersebut yaitu :

- |             |             |           |
|-------------|-------------|-----------|
| - membantu  | - efisien   | - beli    |
| - berjualan | - praktis   | - aman    |
| - suka      | - transaksi | - fitur   |
| - aplikasi  | - jual      | - lengkap |

Pada tahap ini tiap dokumen diwujudkan sebagai sebuah vektor dengan elemen sebanyak *term query* yang terdapat dalam tiap dokumen yang berhasil dikenali dari tahap ekstraksi dokumen sebelumnya. Vektor tersebut beranggotakan bobot dari setiap *term query* yang dihitung berdasarkan metode TF-IDF.

Metode TF-IDF berfungsi untuk mencari representasi nilai dari setiap dokumen. Vektor antara dokumen dan *query* yang terbentuk ditentukan oleh nilai bobot *term query* dalam dokumen. Semakin besar nilai bobot yang diperoleh berarti bahwa tingkat similaritas dokumen terhadap *query* juga semakin tinggi. Contohnya untuk menghitung  $w_{ij}$  *term query* kata “transaksi” dalam  $d_2$  diketahui bahwa:

- Jumlah kemunculan kata “transaksi” dalam  $d_2$  adalah sebanyak satu kali sehingga  $tf_{ij} = tf_{transaksi} = 1$
- Total seluruh dokumen yang ada yaitu sebanyak tiga dokumen sehingga  $D = 3$
- Dari ketiga dokumen tersebut, hanya satu dokumen yang memuat kata “transaksi” sehingga  $df_j = df_{transaksi} = 1$

Oleh karena itu, perhitungan nilai bobot *term* “transaksi” pada dokumen pada  $d_2$  menggunakan rumus pada Persamaan 3.4 yaitu :

$$w_{ij} = tf_{ij} \times \log \frac{D}{df_j}$$

$$w(transaksi) = 1 \times \log \frac{3}{1}$$

$$w(transaksi) = 0,477$$

Selanjutnya dapat dihitung nilai bobot untuk setiap *term* pada *query* dalam masing-masing dokumen seperti berikut:

**Tabel 3.1** Hasil perhitungan bobot pada masing-masing dokumen

| $Q$                                      |       |       |       | $df_{transaksi}$ | $IDF$ | $w(transaksi)$ |               |               |
|--|-------|-------|-------|------------------|-------|----------------|---------------|---------------|
|  | $d_1$ | $d_2$ | $d_3$ |                  |       | $d_1$          | $d_2$         | $d_3$         |
| <u>membantu</u>                          | 1     | 1     | 0     | 2                | 0,176 | 0,176          | 0,176         | 0             |
| <u>transaksi</u>                         | 0     | 1     | 0     | 1                | 0,477 | 0              | 0,477         | 0             |
| <u>aman</u>                              | 1     | 1     | 1     | 3                | 0     | 0              | 0             | 0             |
| <b>Nilai Bobot masing-masing dokumen</b> |       |       |       |                  |       | $\sum w(d_1)$  | $\sum w(d_2)$ | $\sum w(d_3)$ |
|  |       |       |       |                  |       | 0,176          | 0,653         | 0             |

Dengan

$w(transaksi)$  = bobot *term* dari kata “transaksi”

$tf_{transaksi}$  = jumlah kata “transaksi” dalam dokumen (*term frequency*)

$D$  = jumlah seluruh dokumen

$df_{transaksi}$  = jumlah dokumen yang memuat kata “transaksi” di dalamnya

$Q$  = kata/*query*

$d_i$  = dokumen ke- $i$  ,  $i = 1, 2, 3, \dots, n$

Pada **Tabel 3.1** diketahui bahwa bobot dengan peringkat pertama (bobot terbesar) terdapat pada  $d_2$  yaitu 0,653 kemudian diikuti oleh bobot dengan peringkat kedua pada  $d_1$  yaitu 0,176. Sedangkan bobot dengan peringkat ketiga (bobot terendah) dimiliki oleh  $d_3$  yaitu 0. Bobot pada masing-masing dokumen menunjukkan besarnya tingkat korelevanan (kesesuaian) antara dokumen dengan *query*. Nilai bobot pada dokumen berbanding lurus dengan tingkat similaritas dokumen terhadap *query* yang dicari. Maka dari itu, di antara ketiga dokumen tersebut ( $d_1, d_2, d_3$ ) yang memiliki tingkat similaritas paling tinggi terhadap *query* “membantu”, “transaksi”, dan “aman” adalah  $d_2$ . Dengan demikian dapat dihasilkan daftar dokumen teranking berdasarkan nilai kesesuaian (*similarity*) antara dokumen dan *query* masukan yang kemudian akan diberikan kepada pengguna. Dengan kata lain, dapat disimpulkan bahwa tingkat relevansi tertinggi dimiliki oleh  $d_2$  dan tingkat relevansi terendah dimiliki oleh  $d_3$ .

### 3.10. Analisis Sentimen

Analisis sentiment atau bisa juga disebut *opinion mining* merupakan sebuah cabang penelitian di domain *text mining* yang mulai banyak dilakukan pada tahun 2013. Lee dan Pang menjelaskan analisis sentiment atau dikenal sebagai *opinion mining* adalah proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi (Lee & Pang, 2008). Secara umum, *opinion mining* diperlukan untuk mengetahui sikap seorang pembicara atau penulis sehubungan dengan beberapa topik atau polaritas kontekstual keseluruhan dokumen. Sikap yang diambil mungkin menjadi pendapat atau penilaian atau evaluasi (teori appraisal), keadaan afektif (keadaan emosional penulis saat menulis) atau komunikasi emosional (efek emosional penulis yang ingin disampaikan pada pembaca) (Saraswati, 2011).

Analisis sentimen dapat digunakan dalam berbagai kemungkinan domain, dari produk konsumen, jasa kesehatan, jasa keuangan, peristiwa sosial dan politik pada pemilu. Kecendrungan penelitian tentang analisis sentimen berfokus pada pendapat yang menyatakan atau menyiratkan suatu sentimen positif atau negatif. Pendapat mewakili hampir semua aktivitas manusia, karena pendapat dapat mempengaruhi terhadap perilaku seseorang. Setiap kali kita perlu membuat keputusan, kita ingin tahu pendapat orang lain. Dalam dunia nyata, bisnis dan organisasi selalu ingin melihat opini publik tentang suatu produk atau jasa (Liu, 2012).

Dengan analisis sentimen, suatu bisnis dapat melacak produk-produk, merek dan orang-orang misalnya dan menentukan apakah dilihat positif atau negatif di web. Hal ini memungkinkan bisnis untuk melacak:

- a. Deteksi Flame (rants buruk)
- b. Persepsi produk baru.
- c. Persepsi Merek.

### 3.11. Klasifikasi

Bagian sangat penting dalam data mining adalah teknik klasifikasi, yaitu bagaimana mempelajari sekumpulan data sehingga dihasilkan aturan yang bias

mengklasifikasi atau mengenali data-data baru yang belum pernah dipelajari. Klasifikasi dapat didefinisikan sebagai proses untuk menyatakan suatu objek data sebagai salah satu kategori (kelas) yang telah didefinisikan sebagai proses untuk menyatakan suatu objek data sebagai salah satu kategori (kelas) yang telah didefinisikan sebelumnya. Klasifikasi banyak digunakan dalam berbagai aplikasi, di antaranya adalah deteksi kecurangan (*fraud detection*), pengelolaan pelanggan, diagnosis medis, prediksi penjualan, dan sebagainya (Suyanto, 2017).

Model klasifikasi dapat dibangun berdasarkan pengetahuan seorang pakar (ahli). Namun, mengingat himpunan data yang sangat besar, model klasifikasi lebih sering dibangun menggunakan teknik pembelajaran dalam bidang *machine learning*. Proses pembelajaran secara otomatis terhadap suatu himpunan data mampu menghasilkan model klasifikasi (fungsi target) yang memetakan objek data  $x$  (*input*) ke salah satu kelas  $y$  yang telah didefinisikan sebelumnya. Jadi, proses pembelajaran memerlukan masukan (*input*) berupa himpunan data latihan (*training set*) yang berlabel (memiliki atribut kelas) dan mengeluarkan *output* yang berupa sebuah model klasifikasi (Suyanto, 2017).

### 3.11.1. Ukuran Evaluasi Model Klasifikasi

Evaluasi terhadap suatu klasifikasi umumnya dilakukan menggunakan sebuah himpunan data uji, yang tidak digunakan dalam pelatihan klasifikasi tersebut, dengan suatu ukuran tertentu. Terdapat sejumlah ukuran yang dapat digunakan untuk menilai atau mengevaluasi model klasifikasi, diantaranya adalah: *accuracy* atau tingkat pengenalan, *error rate* atau tingkat kesalahan atau kekeliruan klasifikasi, *recall* atau *sensitivity* atau *true positive*, *specificity* atau *true negative* dan *precision*. Model klasifikasi yang dibuat adalah pemetaan dari suatu baris data dengan keluaran sebuah hasil prediksi kelas/target dari data tersebut. Klasifikasi yang memiliki dua kelas sebagai keluarannya disebut dengan klasifikasi biner. Kedua kelas tersebut biasa direpresentasikan dalam  $\{0,1\}$ ,  $\{+1,-1\}$  atau  $\{positive; negative\}$ .

Dalam proses evaluasi klasifikasi terdapat empat kemungkinan yang terjadi dari proses pengklasifikasian suatu baris data. Jika data positif dan diprediksi positif akan dihitung sebagai *true positive*, tetapi jika data itu diprediksi

negatif maka akan dihitung sebagai *false negative*. Jika data negatif dan diprediksi negatif akan dihitung sebagai *true negative*, tetapi jika data tersebut diprediksi positif maka akan dihitung sebagai *false positive* (Fawcett, 2006). Hasil klasifikasi biner pada suatu dataset dapat direpresentasikan dengan matriks 2 x 2 yang disebut *confusion matrix*.

|          |                 | Aktual                     |                            |
|----------|-----------------|----------------------------|----------------------------|
|          |                 | <i>Class</i>               | <i>Positive</i>            |
| Prediksi | <i>Positive</i> | <i>True Positive (TP)</i>  | <i>False Negative (FN)</i> |
|          | <i>Negative</i> | <i>False Positive (FP)</i> | <i>True Negative (TN)</i>  |

*Confusion matrix* sangat berguna untuk menganalisis kualitas *classifier* dalam mengenali tuple-tuple dari kelas yang ada. TP dan TN menyatakan bahwa *classifier* mengenali tuple dengan benar, artinya tuple positif dikenali sebagai positif dan tuple negatif dikenali sebagai negative. Sebaliknya, FP dan FN menyatakan bahwa *classifier* salah dalam mengenali tuple, *tuple negative* dikenali sebagai positif dan tuple negatif dikenali sebagai positif. Terdapat beberapa rumus umum yang dapat digunakan untuk menghitung performa klasifikasi. Hasil dari nilai akurasi, presisi, dan *recall* biasa ditampilkan dalam persentase.

a. *Accuracy*

Akurasi adalah jumlah proporsi prediksi yang benar. Adapun rumus penghitungan akurasi dapat dilihat pada persamaan 3.5. (Lim dkk., 2006).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3.5)$$

b. *Precision*

*Precision* adalah proporsi jumlah dokumen teks yang relevan terkenali diantara semua dokumen teks yang terpilih oleh sistem. Rumus *precision* dapat dilihat pada persamaan 3.6. (Lim dkk., 2006).

$$Precision = \frac{True\ Positive}{True\ Positive+False\ Positive} \quad (3.6)$$

c. *Recall*

*Recall* adalah proporsi jumlah dokumen teks yang relevan terkenali diantara semua dokumen teks relevan yang ada pada koleksi. Rumus *recall* dapat dilihat pada persamaan 3.7. (Lim dkk., 2006).

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3.7)$$

### 3.11.2. *K-Fold Cross Validation*

*Cross Validation* merupakan salah satu teknik untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu . Model yang dibuat bertujuan untuk melakukan prediksi atau klasifikasi terhadap suatu data baru yang belum ada di dalam dataset. Data yang digunakan dalam proses pembangunan model disebut data latih, sedangkan data yang akan digunakan untuk memvalidasi model disebut sebagai data tes.

Menurut Refaeilzadeh dkk (2009), *K-fold cross validation* adalah salah satu metode dari *cross validation* yang digunakan untuk menghitung akurasi prediksi suatu sistem. Dalam *K-fold cross validation* data dibagi menjadi k buah segmen yang memiliki ratio yang sama atau hampir sama. Dilakukan pelatihan dan validasi sebanyak k kali dengan tiap perulangannya mengambil satu segmen berbeda sebagai data tes atau validasi dan k-1 segmen sisanya sebagai data latih untuk kemudian diambil nilai rata-rata dari hasil tiap iterasi. Metode *K-fold cross-validation* menggeneralisasi pendekatan ini dengan mensegmentasi data ke dalam k partisi berukuran sama. Selama proses, salah satu dari partisi dipilih untuk *training*, sedangkan sisanya untuk percobaan. Prosedur ini diulangi k kali sedemikian sehingga setiap partisi digunakan untuk *testing* tepat satu kali.

Sebagai gambaran tentang *K-fold cross validation* dapat dilihat pada **Tabel 3.2** dibawah. Berikut contoh dari *K-fold cross validation* dengan k=5.

“Dataset = K1,K2,K3,K4,K5”

**Tabel 3.2** *Contoh K-Fold Cross Validation*

| <b>Eksperimen ke</b> | <b>Data Latih</b> | <b>Data Tes</b> |
|----------------------|-------------------|-----------------|
| 1                    | K2,K3,K4,K5       | K1              |
| 2                    | K1,K3,K4,K5       | K2              |
| 3                    | K1,K2,K4,K5       | K3              |
| 4                    | K1,K2,K3,K5       | K4              |
| 5                    | K1,K2,K3,K4       | K5              |

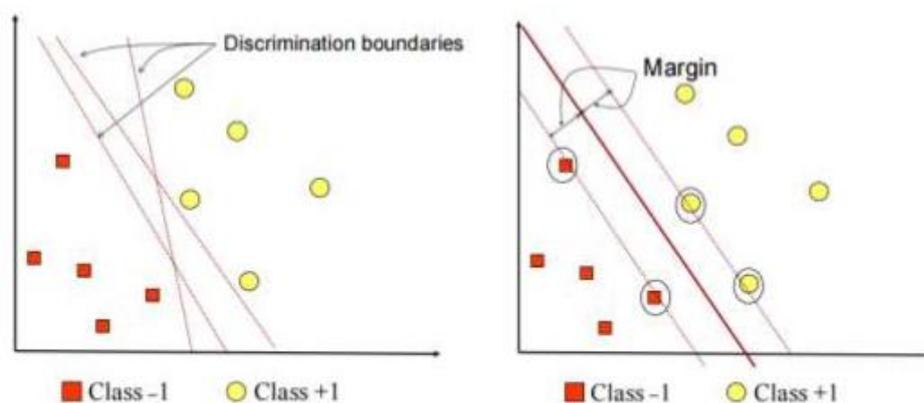
Misalkan dalam penelitian dilakukan 5 dan 10 *K-fold cross validation* sehingga Dilakukan sebanyak 5 dan 10 kali pada tiap dataset, yaitu dengan mengambil data untuk segmen testing (1 segmen) dan segmen sisanya yaitu segmen *training* (k-1 segmen) pada tiap iterasinya (5 kali iterasi). data berjumlah sebanyak 1.923 data, bila menggunakan k=5 maka data latih akan berjumlah 1.538 data latih dan data tes berjumlah 385 data. Dan bila menggunakan k=10 maka data latih akan berjumlah 1.731 data dan data tes berjumlah 192 data.

### **3.12. Support Vector Machine (SVM)**

*Support Vector Machine (SVM)* diperkenalkan oleh Vapnik pada tahun 1992 sebagai suatu teknik klasifikasi yang efisien untuk masalah nonlinear. *Support Vector Machine (SVM)* juga dikenal sebagai teknik pembelajaran mesin (*machine learning*) paling mutakhir setelah pembelajaran mesin sebelumnya yang dikenal sebagai *Neural Network (NN)*. Baik *SVM* maupun *NN* tersebut telah berhasil digunakan dalam pengenalan pola. Pembelajaran dilakukan dengan menggunakan pasangan data input dan data output berupa sasaran yang diinginkan. Konsep *SVM* dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada input *space*. *SVM* berusaha menemukan fungsi pemisah (*hyperplane*) dengan memaksimalkan jarak antar kelas. Dengan cara ini, *SVM* dapat menjamin kemampuan generalisasi yang tinggi untuk data-data yang akan datang (Suyanto, 2017).

### 3.12.1. SVM Pada Data Terpisah Secara Linear

Pada **Gambar 3.2** memperlihatkan beberapa *pattern* yang merupakan anggota dari dua buah kelas : positif (dinotasikan dengan +1) dan negatif (dinotasikan dengan -1). *Pattern* yang tergabung pada kelas negatif disimbolkan dengan kotak, sedangkan *pattern* pada kelas positif, disimbolkan dengan lingkaran. Proses pembelajaran dalam problem klasifikasi diterjemahkan sebagai upaya menemukan garis (*hyperplane*) yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (*discrimination boundaries*) ditunjukkan pada **Gambar 3.2**.



(Sumber : Oktariadi, 2014)

**Gambar 3.2** Ilustrasi SVM menemukan *hyperplane* terbaik yang memisahkan dua kelas -1 dan +1

*Hyperplane* pemisah terbaik antara kedua *class* dapat ditemukan dengan mengukur margin *hyperplane* tersebut dan mencari titik maksimalnya. Margin adalah jarak antara *hyperplane* tersebut dengan data terdekat dari masing-masing kelas. *Subset data training set* yang paling dekat ini disebut sebagai *support vector*. Garis solid pada **Gambar 3.2** sebelah kanan menunjukkan *hyperplane* yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua *class*, sedangkan titik kotak dan lingkaran yang berada dalam lingkaran hitam adalah *support vector*. Upaya mencari lokasi *hyperplane* optimal ini merupakan inti dari proses pembelajaran pada SVM. Data yang tersedia dinotasikan sebagai  $x \in R^d$

sedangkan label masing-masing kelas dinotasikan  $y_i \in \{-1, +1\}$  untuk  $i = 1, 2, 3, \dots, n$ .

Diasumsikan kedua kelas dapat terpisah secara sempurna oleh *hyperplane* berdimensi  $d$ , yang didefinisikan :

$$\vec{w} \cdot \vec{x} + b = 0 \quad (3.8)$$

*Pattern*  $\vec{x}_i$  yang termasuk kelas -1 (sampel negatif) dapat dirumuskan sebagai *pattern* yang memenuhi pertidaksamaan:

$$\vec{w} \cdot \vec{x} + b \leq -1 \quad (3.9)$$

Sedangkan *pattern*  $\vec{x}_i$  yang termasuk kelas +1 (sampel positif)

$$\vec{w} \cdot \vec{x} + b \geq +1 \quad (3.10)$$

dengan :

$R^d$  = ruang vektor ,  $d > 1$

$\vec{w}$  = *vector* bobot

$\vec{x}$  = *vector* data (*input*)

$b$  = bias

*Margin* terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara jarak dan titik terdekatnya, yaitu  $\frac{1}{\|w\|}$ . Hal ini dapat dirumuskan sebagai *Quadratic Programming* (QP) *problem*, yaitu mencari titik minimal persamaan 3.11, dengan memperlihatkan *constraint* persamaan 3.12.

$$\min_w = \tau(w) = \frac{1}{2} \|w\|^2 \quad (3.11)$$

$$y_i(x_i \cdot w + b) - 1 \geq 0 \quad (3.12)$$

dengan  $\|w\|$  adalah vektor normal.

Masalah ini dapat dipecahkan dengan berbagai teknik komputasi, diantaranya *lagrange multiplier* yang dinyatakan pada Persamaan 3.13.

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_i^l \alpha_i (y_i ((\vec{w} \cdot \vec{x} + b) - 1)) \quad (3.13)$$

dengan  $i = 1, 2, \dots, l$

Dimana  $\alpha$  adalah *lagrange multiplier*, yang bernilai 0 atau positif  $\alpha_i \geq 0$ . Nilai optimal dari persamaan 3.13 dapat dihitung dengan meminimalkan L

terhadap  $w$  dan  $b$ , dan memaksimalkan  $L$  terhadap  $\alpha_i$ , dengan memperhatikan sifat bahwa pada titik optimal *gradient*  $L = 0$  persamaan 3.13 dapat dimodifikasi sebagai maksimalisasi *problem* yang hanya mengandung  $\alpha_i$ , sebagaimana terlihat pada persamaan 3.14 dan 3.15 dibawah ini.

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \vec{x}_i \vec{x}_j \quad (3.14)$$

$$\text{Dimana } \alpha_i \geq 0 \ (i = 1, 2, \dots, l) \ \sum_{i=1}^l \alpha_i y_i = 0 \quad (3.15)$$

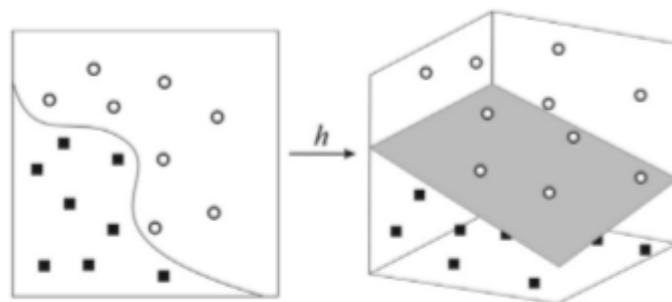
Dengan demikian maka akan diperoleh  $\alpha_i$  yang kebanyakan bernilai positif yang disebut sebagai *support vector* dan juga memperoleh persamaan 3.16 dan Persamaan 3.17 sebagai solusi bidang pemisah.

$$w = \sum \alpha_i y_i x_i \quad (3.16)$$

$$b = y_k - w^T x_k \quad (3.17)$$

### 3.12.2. SVM Pada Data Tidak Terpisah Secara Linear

Dalam beberapa kasus, dapat ditemukan bahwa himpunan data tidak bisa dipisahkan secara linear. SVM mampu menyelesaikan permasalahan tidak linear dengan menggunakan teknik *kernel* (Cortes dan Vapnik, 1995). Pada dasarnya, penggunaan kernel ini memetakan vektor masukan pada ruang berdimensi rendah ke ruang ciri berdimensi lebih tinggi. **Gambar 3.2** menunjukkan bahwa data masukan yang tidak bisa dipisahkan secara linear kemudian ditransformasikan ke dalam ruang berdimensi lebih tinggi (*feature space*). Jika pada data linear *hyperplane* berbentuk sebuah garis yang memisahkan antar kelas, maka pada data non linear *hyperplane* akan berbentuk sebuah bidang yang memisahkan antar kelas.



(Sumber :Moraes dkk, 2013)

**Gambar 3.3** Transformasi dari input space ke feature space

Kasus data yang tidak terpisah secara linear diasumsikan bahwa kelas pada *input space* tidak dapat terpisah secara sempurna. Hal ini menyebabkan *constraint* pada persamaan 3.12 tidak dapat terpenuhi, sehingga optimisasi tidak dapat dilakukan, untuk mengatasi masalah ini SVM dirumuskan ulang dengan memperkenalkan teknik *softmargin*. Dalam *softmargin* persamaan 3.12 dimodifikasi dengan menggunakan *slack variabel* sehingga terlihat pada persamaan 3.18.

$$y_i(x_i \cdot w + b) \geq 1 - \varepsilon_i \quad (3.18)$$

Dengan demikian, persamaan 3.11 diubah menjadi persamaan 3.19.

$$\min_w = \tau(w) = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^l \varepsilon_i \quad (3.19)$$

Fitur  $c$  digunakan untuk mengontrol *tradeoff* antara *margin* dan kesalahan klasifikasi  $\varepsilon$ .

### 3.12.3. Kernel Trick dan Non-Linear Classification pada SVM

Pada umumnya masalah dalam domain dunia nyata (*real world problem*) jarang yang bersifat *linear separable* dan kebanyakan bersifat *non linear*. Untuk menyelesaikan masalah *non linear*, SVM dimodifikasi dengan memasukkan fungsi *Kernel*.

$$\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^d \quad d < q \quad (3.20)$$

Data SVM *non linear*, data  $\vec{x}$  dipetakan oleh fungsi  $\phi(\vec{x})$  ke ruang *vector* yang berdimensi lebih tinggi. Pemetaan ini dilakukan dengan menjaga topologi data, dalam artian dua data yang berjarak dekat pada *input space* akan berjarak dekat juga pada *feature space*, sebaliknya jika dua data yang berjarak jauh pada *input space* maka akan berjarak jauh juga pada *feature space*. Selanjutnya proses pembelajaran pada SVM hanya bergantung pada *dot product* dari data yang sudah ditransformasikan pada ruang baru yang berdimensi lebih tinggi, yaitu  $\phi(x_i) \cdot \phi(x_j)$ . Karena pada umumnya transformasi  $\phi$  tidak diketahui maka perhitungan *dot product* dapat digantikan dengan fungsi *kernel*  $K(x_i, x_j)$  yang mendefinisikan secara implisit fungsi transformasi  $\phi$  tersebut. Inilah yang disebut *Kernel Trick*, yang diformulasikan sebagai

$$K(x_i, x_j) = \phi_i(x_i) \cdot \phi_j(x_j) \quad (3.21)$$

Beberapa *kernel* yang umum dipakai pada SVM adalah:

a. *Polynomial*

*Kernel trick polynomial* cocok digunakan untuk menyelesaikan masalah klasifikasi, dimana dataset pelatihan sudah normal. *Kernel trick* ini dinyatakan dalam persamaan.

$$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i, \vec{x}_j + 1)^d \quad (3.22)$$

b. *Radial Basis Function (RBF)* atau *Gaussian*

*Kernel trick radial basis function* atau *gaussian* merupakan *kernel* yang paling banyak digunakan untuk menyelesaikan masalah klasifikasi untuk dataset yang tidak terpisah secara *linear*, dikarenakan akurasi pelatihan dan akurasi prediksi yang sangat baik pada kernel ini, dimana kernel radial basis *function* dinyatakan dalam persamaan 3.23.

$$K(\vec{x}_i, \vec{x}_j) = \exp\left(-\|\vec{x}_i - \vec{x}_j\|^2\right) \gamma \quad (3.23)$$

c. *Sigmoid Kernel*

*Sigmoid* merupakan *kernel trick* SVM yang merupakan pengembangan dari jaringan saraf tiruan, dimana kernel ini dinyatakan dengan persamaan 3.24.

$$K(\vec{x}_i, \vec{x}_j) = \tanh(\alpha \vec{x}_i + \vec{x}_j + \beta) \quad (3.24)$$

*Kernel trick* memberikan beberapa kemudahan, karena dalam proses pembelajaran SVM, untuk menentukan *support vector*, pengguna hanya cukup mengetahui fungsi *kernel trick* yang dipakai, tanpa perlu mengetahui wujud dari fungsi non-linier. Dari keseluruhan *kernel trick* tersebut, *kernel trick radial basis function* merupakan *kernel trick* yang memberikan hasil terbaik pada proses klasifikasi khususnya untuk data yang tidak bisa dipisahkan secara linear. Selanjutnya klasifikasi sebuah objek data  $x$  dapat diformulasikan sebagai berikut :

$$f(x) = \sum_{i=1, \vec{x}_i \in SV}^n \alpha_i y_i K(\vec{x}_i, \vec{x}_j) + b \quad (3.25)$$

### 3.13. *Maximum Entropy*

Menurut Nigam ( dikutip dalam Masithoh, 2016) *Maximum Entropy* adalah teknik umum yang digunakan untuk mengestimasi probabilitas distribusi data.

Teknik *Maximum Entropy* menyatakan bahwa ketika pada data yang diketahui tidak terdapat informasi, maka data tersebut diusahakan berdistribusi seragam/uniform, yaitu memiliki *Maximum Entropy*. *Maximum Entropy* juga merupakan teknik untuk mengestimasi distribusi dari berbagai *natural language task* seperti *language modeling*, *pelabelan part of speech*, dan segmentasi pada teks.

*Maximum Entropy* merupakan metode yang mencari distribusi dengan  $p(a|b)$  yang akan menghasilkan nilai *entropy maksimum*. Menurut MacKay (dikutip dalam Masithoh, 2016), *Maximum Entropy* didefinisikan sebagai rata-rata nilai informasi yang maksimum untuk suatu himpunan kejadian  $X$  dengan distribusi nilai probabilitas yang seragam. Maksudnya adalah distribusi yang menggunakan faktor ketidakpastian seminimal mungkin atau menggunakan asumsi sesedikit mungkin sehingga didapatkan distribusi yang paling mendekati kenyataan. Pencarian distribusi yang menghasilkan nilai *entropy* yang maksimum bertujuan untuk mendapatkan distribusi probabilitas terbaik yang paling mendekati kenyataan.

Dengan menggunakan metode *Maximum Entropy*, diharapkan bahwa model yang didapat merupakan gambaran yang paling mirip dengan kenyataan yang ada.

### 3.13.1. Definisi *Entropy*

*Entropy* merupakan rata-rata dari himpunan informasi yang terkandung dalam suatu kumpulan kejadian  $X = \{x_1, x_2, \dots, x_n\}$ . Himpunan informasi yang terkandung pada suatu kejadian dinyatakan sebagai:

$$h(x) = \log \frac{1}{p(x)} \quad (3.26)$$

dengan  $h(x)$  merupakan himpunan informasi dari suatu kejadian  $x$  dan  $p(x)$  merupakan probabilitas dari kemunculan kejadian  $x$ . Nilai dari  $h(x)$  dinyatakan dalam ukuran *bit*. Jumlah *bit* pada  $h(x)$  merupakan banyaknya *bit* yang diperlukan untuk merepresentasikan himpunan informasi dari suatu kejadian  $x$ . Semakin besar nilai  $h(x)$  maka semakin besar pula informasi yang dimiliki oleh  $h(x)$ .

Didefinisikan  $A$  adalah himpunan kelas klasifikasi sedangkan  $B$  adalah himpunan dokumen. *Entropy* dari kumpulan dapat dinyatakan sebagai berikut:

$$H(p) = -\sum_{x \in \mathcal{E}} p(x) \log p(x) \quad (3.27)$$

Keterangan:

$a \in A$

$b \in B$

$x = (a, b)$

$\mathcal{E} = A \times B$

$p(x)$ : peluang pada kelas  $a$  dengan terdapat dokumen  $b$

Dengan metode *Maximum Entropy*, hasil yang diinginkan adalah nilai  $H(p)$  yang maksimum. Nilai *entropy* yang maksimum akan tercapai apabila nilai dariseragam sehingga mengakibatkan  $p(x) = \frac{1}{|X|}$ , dengan  $|X|$  merupakan kardinalitas dari  $X$ . Dalam proses klasifikasi, untuk mendapatkan nilai maksimal yang seragam tidaklah sederhana dengan membagi nilai 1 dengan nilai kardinalitas  $X$ . Kardinalitas suatu himpunan adalah ukuran banyaknya elemen yang dikandung oleh himpunan tersebut. Pencarian distribusi probabilitas tersebut juga harus memenuhi batasan-batasan yang ada dengan mengetahui fakta ataupun data yang dimiliki.

### 3.13.2. Prinsip *Maximum Entropy*

Pada teknik *Maximum Entropy* ini menyatakan bahwa ketika tidak ada informasi dari data yang diketahui, maka distribusi diusahakan untuk *uniform*, yaitu mempunyai *entropy maksimal*. Dalam klasifikasi teks, *Maximum Entropy* mengestimasi distribusi label dalam dokumen. Dokumen direpresentasikan oleh seperangkat fitur kemunculan kata. Pada metode *Maximum Entropy*, proses klasifikasi dilakukan dengan hanya menggunakan informasi kemunculan dari suatu fitur dalam suatu dokumen. Hal ini berhubungan dengan pengguna  $f_i \in \{0,1\}$ . Secara garis besar, metode *Maximum Entropy* mencari distribusi probabilitas yang paling seragam.

### 3.13.3. Algoritma Klasifikasi dengan *Maximum Entropy*

Berikut ini adalah algoritma klasifikasi teks menggunakan metode *Maximum Entropy*:

1. Mengidentifikasi kata-kata spesifik yang ada di dalam dokumen (kalimat).
2. Membentuk matriks yang berisi nilai kemunculan kata-kata spesifik tersebut dengan indeks berikut

$$f_j(a, b) = \begin{cases} 1; & \text{jika } f_j \text{ muncul di dokumen } b \text{ pada kelas } a \\ 0; & \text{jika } f_j \text{ tidak muncul di dokumen } b \text{ pada kelas } a \end{cases}$$

3. Membuat model *Maximum Entropy* dengan *data training* yaitu menghitung nilai  $\alpha_j$  untuk setiap kelas dengan prosedur GIS.

$$\alpha_j^{(0)} = 1$$

$$\alpha_j^{(n+1)} = \alpha_j^{(n)} \left[ \frac{E_{\bar{p}} f_j}{E^{(n)} f_j} \right]^{\frac{1}{c}} \quad (3.28)$$

Dimana,

$$E_{\bar{p}} f_j = \sum_{x \in \mathcal{E}} p(x) f_j(x) \quad (3.29)$$

$$E^{(n)} f_j = \sum_{x \in \mathcal{E}} p(x) f_j(x) \quad (3.30)$$

$$p^{(n)}(x) = \pi \prod_{j=1}^k \left( \alpha_j^{(n)} \right)^{f_j(x)} \quad (3.31)$$

$$\forall x \in \sum_{j=1}^k f_1(x) = C$$

4. Mencari *joint probability*  $p(a, b)$  untuk *data testing*

$$a = \{\text{positif}, \text{negatif}\}$$

$$p^*(a, b) = \pi \prod_{j=1}^k \alpha_j^{f_j(a, b)} \quad (3.32)$$

5. Penentuan topik dari dokumen *data testing* dengan melihat nilai  $a^*$  yang paling besar pada suatu kelas

$$a^* = \mathbf{argmax} p(a, b) \quad (3.33)$$

$$a \in (\text{positif}, \text{negatif})$$

### 3.13.4. Simulasi Metode *Maximum Entropy*

#### 1) Simulasi untuk Fitur yang Didasarkan pada Label dan Warna

Simulasi ini diambil dari presentasi yang diunggah oleh Joey Willits yang diinspirasi oleh tutorial Adam Berger tentang *Maximum Entropy*. Misalkan

terdapat sekumpulan permen, masing-masing dengan label terkait ( $A$ ,  $B$ ,  $C$ , atau  $D$ ). Setiap permen memiliki beberapa warna dalam bungkusnya. Setiap permen juga diberi label secara acak berdasarkan beberapa distribusi di atas warna pembungkus.



**Gambar 3.4** Permen dan masing-masing labelnya

Jika yang diinginkan adalah permen merah, maka didapatkan::

$$- P(A|merah) + P(B|merah) + P(C|merah) + P(D|merah) = 1$$

- Distribusi yang sesuai dengan konsep *Maximum Entropy* yaitu :

$$P(A|merah) = 0,25$$

$$P(B|merah) = 0,25$$

$$P(C|merah) = 0,25$$

$$P(D|merah) = 0,25$$

Jika diketahui bahwa 80% dari semua permen dengan label merah diberi label  $A$  atau  $B$ , maka  $P(A|merah) + P(B|merah) = 0,8$ . Dengan demikian, nilai probabilitas terbaru yang mencerminkan keadaan tersebut adalah:

$$P(A|merah) = 0,40$$

$$P(B|merah) = 0,10$$

$$P(C|merah) = 0,40$$

$$P(D|merah) = 0,10$$

## 2) Simulasi untuk *Natural Language Processing* (NLP)

Untuk penggunaan metode *Maximum Entropy* pada kasus NLP didasarkan pada label dan kata-kata yang ada di dalam teks. Misal terdapat tiga buah dokumen yang diambil dari **kelas positif** yaitu:

Dokumen 1 ( $d_1$ ) = **membantu** berjualan **suka** aplikasi **efisien** praktis  
**aman**

Dokumen 2 ( $d_2$ ) = **membantu** transaksi jual beli **aman**

Dokumen 3 ( $d_3$ ) = aplikasi jual beli **aman** fitur **lengkap**

Setelah mengidentifikasi kata-kata spesifik yang ada di dalam dokumen pada fase *training* (yaitu: membantu, suka, efisien, aman, lengkap), maka

terbentuklah matriks yang berisi nilai kemunculan kata-kata spesifik tersebut dengan indeks berikut:

$$s(y, x) = \begin{cases} 1; & \text{jika } (y = \text{positif dan } x = \text{"membantu, suka, efisien, aman, lengkap}) \\ 0; & \text{(y dan x yang lain)} \end{cases}$$

**Tabel 3.3** Pemberian indeks pada masing-masing kata

| Kata      | Indeks | Kata      | Indeks | Kata     | Indeks |
|-----------|--------|-----------|--------|----------|--------|
| membantu  | 1      | aman      | 1      | aplikasi | 0      |
| berjualan | 0      | membantu  | 1      | jual     | 0      |
| suka      | 1      | transaksi | 0      | beli     | 0      |
| aplikasi  | 0      | jual      | 0      | aman     | 1      |
| efisien   | 1      | beli      | 0      | fitur    | 0      |
| praktis   | 0      | aman      | 1      | lengkap  | 1      |

Fungsi ini memberi nilai 1 hanya jika keduanya terjadi dalam kombinasi:

- pada fase *training*, tag dan kata dapat diketahui (tag positif dan kata membantu, suka, efisien, aman, lengkap).
- pada fase *testing*, semua kelas yang ada akan dievaluasi untuk menemukan kelas dengan probabilitas terbesar (kelas positif dan kelas negatif).

### 3.14. Word Cloud

*Word cloud* adalah salah satu hasil dari metode *text mining* yang menampilkan kata-kata populer terkait dengan kata kunci internet dan data teks. Menurut PBC (dikutip dalam Arkhamsiagustinah, 2015) *word cloud* sering digunakan untuk menyoroti istilah populer atau trend berdasarkan frekuensi pengguna. Menurut Shawn Graham, Ian Milligan, dan Scott Weingart (dikutip dalam Arkhamsiagustinah, 2015) *word cloud* merupakan pendekatan yang dapat menjelaskan pertanyaan penelitian dengan sangat cepat dan mudah, dapat menjelajahi *word cloud* secara singkat dan dapat melakukan analisis yang korehensif. Kata yang paling sering muncul di dalam data teks akan memiliki bentuk yang paling besar, begitu pula sebaliknya.

### 3.15. Asosiasi Kata

Isitilah korelasi sering digunakan untuk menyatakan hubungan dua atau lebih variabel yang sifatnya kuantitatif, sedangkan istilah asosiasi sering dimaknai

keeratan hubungan antara dua atau lebih variabel yang sifatnya kualitatif. Penelitian ini menggunakan pendekatan asosiasi untuk menemukan hubungan antar kata pada masing-masing klasifikasi ulasan positif dan ulasan negatif, sehingga mendapatkan informasi yang dapat dijadikan referensi bagi pihak *e-commerce* maupun pengguna/pelanggan untuk mengetahui topik yang paling sering dibicarakan terkait Bukalapak.

$$r_{xy} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{\{n\sum x_i^2 - (\sum x_i)^2\}\{n\sum y_i^2 - (\sum y_i)^2\}}} \quad (3.34)$$

Dengan

- $r_{xy}$  = nilai korelasi antara variabel  $x$  dan variabel  $y$
- $n$  = banyaknya pasangan data  $x$  dan  $y$
- $\sum x_i$  = jumlah nilai pada variabel  $x$   $i = 1, 2, 3, \dots, n$
- $\sum y_i$  = jumlah nilai pada variabel  $y$
- $\sum x_i^2$  = kuadrat dari total nilai variabel  $x$
- $\sum y_i^2$  = kuadrat dari total nilai variabel  $y$
- $\sum x_i \sum y_i$  = jumlah dari hasil perkalian antara nilai variabel  $x$  dan variabel  $y$

### 3.15.1. Simulasi Perhitungan Asosiasi pada Data Teks

Perhitungan asosiasi teks dilakukan dengan menggunakan pendekatan nilai korelasi. Data teks terlebih dahulu harus ditransformasikan ke dalam bentuk *document term matrix (dtm)*. Untuk simulasi ini akan digunakan sebanyak enam data sebagai berikut:

```

kata1
kata1 kata2
kata1 kata2 kata3
kata1 kata2 kata3 kata4
kata1 kata2 kata3 kata4 kata5
kata1 kata2 kata3 kata4 kata5 kata6

```

Keenam kata tersebut diubah menjadi bentuk *document term matrix*.

| Docs | kata1 | kata2 | kata3 | kata4 | kata5 | kata6 |
|------|-------|-------|-------|-------|-------|-------|
| 1    | 1     | 0     | 0     | 0     | 0     | 0     |
| 2    | 1     | 1     | 0     | 0     | 0     | 0     |
| 3    | 1     | 1     | 1     | 0     | 0     | 0     |
| 4    | 1     | 1     | 1     | 1     | 0     | 0     |
| 5    | 1     | 1     | 1     | 1     | 1     | 0     |
| 6    | 1     | 1     | 1     | 1     | 1     | 1     |

Setelah didapatkan nilai dari *documents term matrix* maka dilakukan perhitungan nilai asosiasi dengan menggunakan rumus korelasi. Uji coba menghitung nilai korelasi akan dilakukan pada kata 3 dan kata 5.

| Docs         | kata3    | kata5    | kata3 <sup>2</sup> | kata5 <sup>2</sup> | kata3*kata5 |
|--------------|----------|----------|--------------------|--------------------|-------------|
| 1            | 0        | 0        | 0                  | 0                  | 0           |
| 2            | 0        | 0        | 0                  | 0                  | 0           |
| 3            | 1        | 0        | 1                  | 0                  | 0           |
| 4            | 1        | 0        | 1                  | 0                  | 0           |
| 5            | 1        | 1        | 1                  | 1                  | 1           |
| 6            | 1        | 1        | 1                  | 1                  | 1           |
| <b>Total</b> | <b>4</b> | <b>2</b> | <b>4</b>           | <b>2</b>           | <b>2</b>    |

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{\{n \sum x_i^2 - (\sum x_i)^2\} \{n \sum y_i^2 - (\sum y_i)^2\}}}$$

$$r = \frac{(6 \times 2) - (4 \times 2)}{\sqrt{\{(6 \times 4) - (4^2)\} \{(6 \times 2) - (2^2)\}}}$$

$$r = \frac{4}{\sqrt{64}} = \frac{4}{8} = 0,5$$

Jadi, nilai korelasi kata3 dan kata5 sebesar 0,5. Hal ini menunjukkan bahwa besarnya asosiasi atau hubungan kata3 dengan kata5 sebesar 0,5 atau 50%.

### 3.16. Metode Diagram *Fishbone*

Metode diagram *fishbone* atau sering dikenal dengan “Diagram Ishikawa” karena dikembangkan oleh Ishikawa pada sekitar tahun 1960-an. Diagram ini disebut *fishbone* karena memiliki kerangka yang menyerupai tulang ikan yang

meliputi kepala, sirip dan duri. Diagram *fishbone* merupakan alat visual yang digunakan untuk mengidentifikasi, mengeksplorasi yang secara grafik menggambarkan semua penyebab suatu permasalahan. Konsep dasar dari diagram *fishbone* adalah permasalahan mendasar diletakkan pada bagian kanan dari diagram atau pada bagian kepala dari kerangka tulang ikannya. Sementara sirip dan duri digambarkan sebagai penyebab permasalahannya. Diagram *fishbone* umumnya digunakan untuk mengidentifikasi permasalahan dan menentukan penyebab dari permasalahan tersebut (Fritz, 2016).

Diagram *fishbone* dapat digunakan dalam menganalisis permasalahan baik dalam level individu, tim maupun organisasi. Adapun manfaat *fishbone* dalam analisis masalah, yaitu :

1. Memudahkan dalam mengilustrasikan gambaran singkat dalam tim/organisasi.
2. Lebih gampang dalam pemfokusan individu, tim atau organisasi pada permasalahan utama. Diagram *fishbone* akan membantu tim/organisasi dalam menentukan masalah prioritas.
3. Diagram *fishbone* menghasilkan suatu solusi. Setelah dicari akar penyebab masalah, langkah dalam mengambil solusi akan lebih gampang dicapai.
4. Memudahkan tim dan anggota tim dalam melakukan diskusi dan menjadikan diskusi lebih terarah pada masalah dan penyebabnya.

Elemen-elemen alat pemasaran dapat dikelompokkan menjadi 4P, yaitu: *product*, *place*, *promotion*, dan *price*. Untuk jasa bauran pemasaran terdiri atas 8P, dengan tambahan *people*, *physical evidence*, *process* dan *productivity* (Lovelock dan Wright, 2005).

1. *Process* merupakan suatu metode pengoperasian atau serangkaian tindakan yang diperlukan untuk menyajikan produk dan layanan yang baik kepada pelanggan.
2. *Product* merupakan semua komponen dari kinerja layanan yang menciptakan nilai bagi pelanggan, memberikan manfaat, memenuhi kebutuhan, dan dapat memuaskan konsumen.

3. *People* merupakan semua yang terlibat dalam kegiatan memproduksi produk dan layanan (*service production*).
4. *Place* merupakan keputusan manajemen mengenai kapan, dimana, dan bagaimana menyajikan layanan yang baik kepada pelanggan.
5. *Price* merupakan pengeluaran uang, waktu, dan usaha yang pelanggan korbankan dalam membeli dan mengkonsumsi produk dan layanan yang diberikan perusahaan.
6. *Promotion* merupakan semua aktivitas komunikasi dan perancangan insentif untuk membangun persepsi pelanggan yang dikehendaki perusahaan atas layanan spesifik yang perusahaan berikan.
7. *Physical Evidence* merupakan perangkat-perangkat yang diperlukan dalam menyajikan secara nyata kualitas produk dan layanan.
8. *Productivity & Quality Produktivitas* adalah sejauh mana efisiensi layanan ditransformasikan ke dalam hasil-hasil layanan yang dapat menambah nilai bagi pelanggan, sedangkan kualitas adalah derajat suatu layanan yang dapat memenuhi harapan pelanggan.

## BAB IV

### METODOLOGI PENELITIAN

#### 4.1. Populasi dan Sampel

Populasi dalam penelitian ini adalah database *website Google Play Store*, yaitu semua data *review* pengguna aplikasi Bukalapak sedangkan sampel yang digunakan dalam penelitian ini adalah *review* pengguna aplikasi Bukalapak berbahasa Indonesia pada tanggal 3 Januari 2018 – 18 Maret 2018.

#### 4.2. Variabel dan Definisi Opsional Variabel

Pada **Tabel 4.1** berisi tentang penjelasan dan definisi operasional dari masing-masing variabel penelitian:

**Tabel 4.1** *Definisi Operasional Variabel*

| Variabel      | Definisi Operasional Variabel       |
|---------------|-------------------------------------|
| <i>Rating</i> | Tingkat kepuasan pengguna bukalapak |
| <i>Date</i>   | Tanggal dibuatnya komentar          |
| <i>Review</i> | Isi komentar pengguna Bukalapak     |

#### 4.3. Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data primer. Data tersebut diperoleh dengan *scraping* dari halaman situs *web* Google Play. Data yang diperoleh merupakan data dari *database* website Google Play, yaitu data yang berupa ulasan pengguna Bukalapak sebanyak 1.923 ulasan.

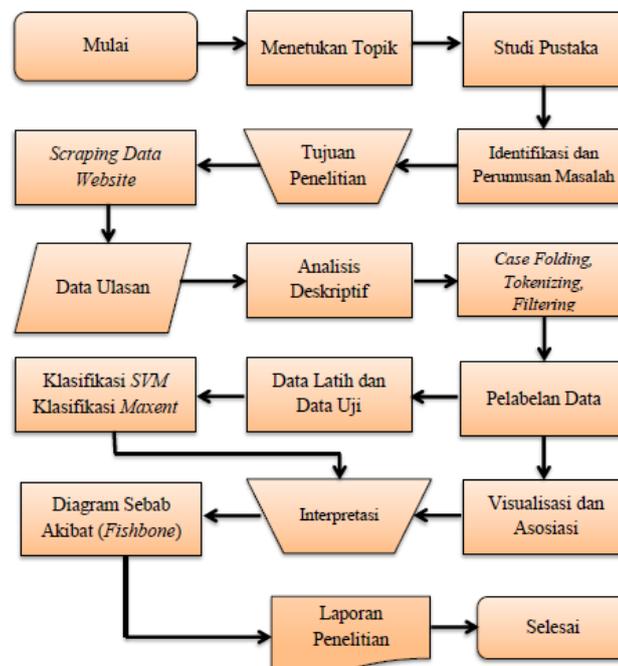
#### 4.4. Metode Analisis Data

Dalam penelitian ini, peneliti menggunakan software *R Studio*, *R 3.3.3* dan *Microsoft Excel 2016*. Adapun metode analisis data yang digunakan dalam penelitian ini yaitu:

1. Analisis Deskriptif, digunakan untuk memberikan gambaran umum ulasan BukaLapak yang ada pada situs *Google Play*.
2. Analisis Sentimen berbasis kamus *lexicon*, digunakan untuk melakukan pelabelan data ke dalam kelas sentimen positif dan negatif.
3. Metode *Machine Learning* yaitu *Support Vector Machine (SVM)* dan *Maximum Entropy (Maxent)* digunakan untuk mengklasifikasikan ulasan yang berbentuk positif dan negatif.
4. *Association*, digunakan untuk mengidentifikasi dan membentuk pola kata yang berasosisasi dengan kata lainnya guna mendapatkan informasi yang dianggap penting dan berguna.
5. Diagram *Fishbone*, digunakan untuk mengidentifikasi faktor-faktor penyebab permasalahan yang didapatkan dari ulasan negatif sehingga dapat dilakukan pemecahan masalah yang dihadapi.

#### 4.5. Proses Analisis Data

Adapun langkah-langkah penelitian divisualisasikan dalam diagram melalui **Gambar 4.1** berikut.



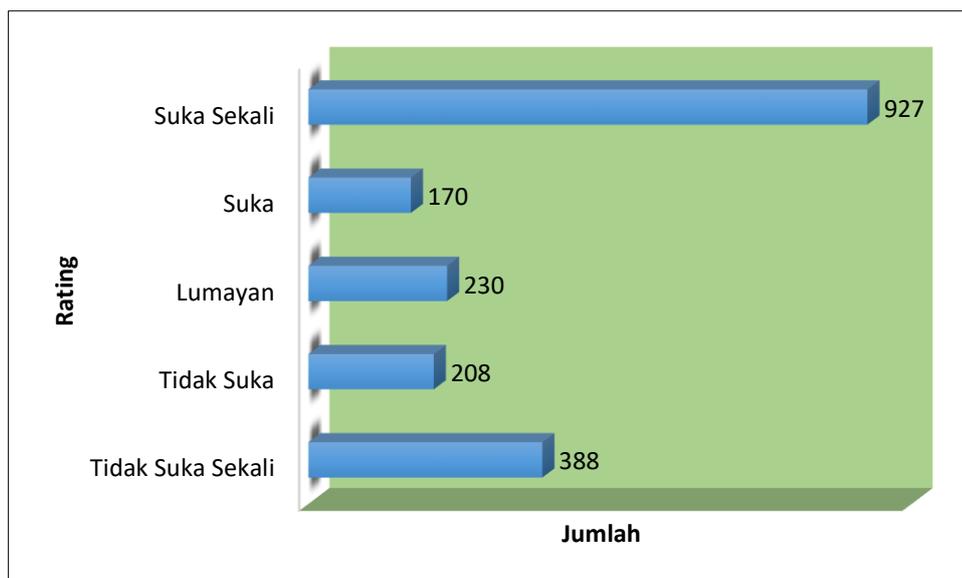
**Gambar 4.1** Flowchart Penelitian

## BAB V

### ANALISIS DAN PEMBAHASAN

#### 5.1. Analisis Deskriptif

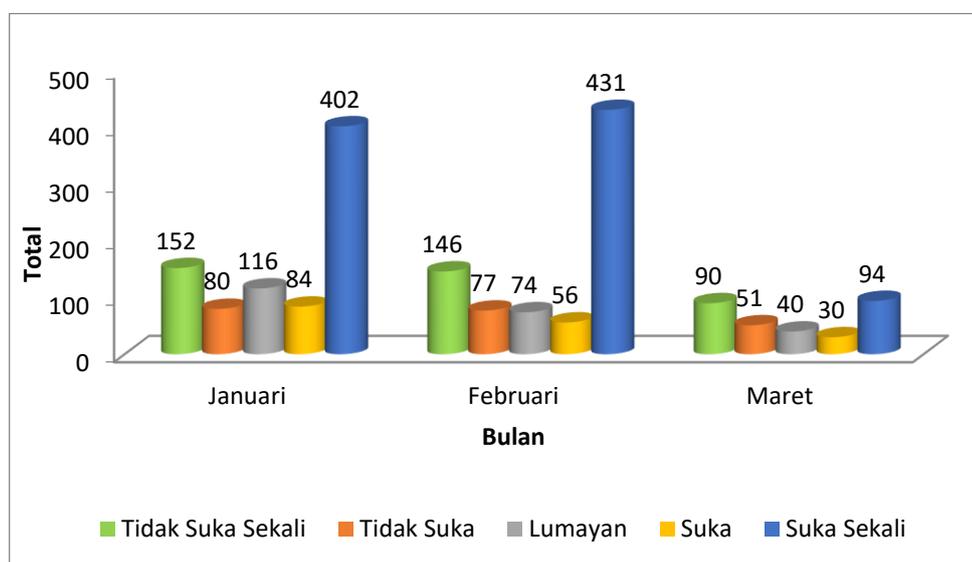
Analisis deskriptif dalam penelitian ini digunakan untuk melihat gambaran secara umum informasi tentang Aplikasi Bukalapak berdasarkan data ulasan pengguna dari situs *Google Play*, yang dilihat dari beberapa aspek diantaranya jumlah ulasan yang masuk berdasarkan urutan waktu, *rating* aplikasi yang diberikan pengguna dan perbandingan jumlah ulasan dari pengguna tersebut dikategorikan menjadi dua kategori yaitu ulasan positif dan ulasan negatif. Berikut ini merupakan *rating* yang menggambarkan penilaian diberikan pengguna terhadap *ecommerce Bukalapak* pada situs *Google Play*.



**Gambar 5.1** *Rating pengguna terhadap e-commerce Bukalapak pada situs Google Play*

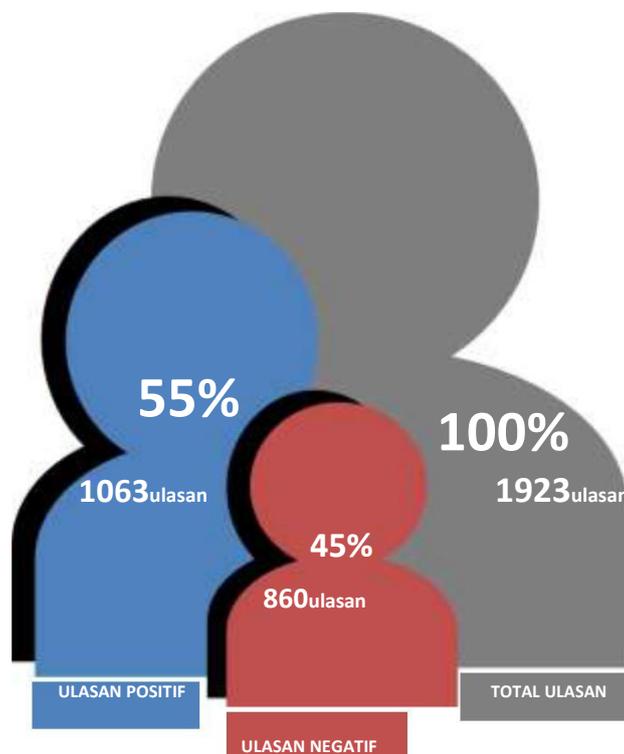
Rating pada situs *Google Play* mempunyai skala 1-5 dengan kategori dari yang paling rendah ialah “Tidak Suka Sekali” yang diberi skor “1”, “Tidak Suka” dengan skor “2”, “Lumayan” dengan skor “3”, “Suka” dengan skor “4”, dan “Suka Sekali” dengan skor “5”. **Dari Gambar 5.1** diatas dapat diketahui bahwa

mayoritas pengguna Bukalapak mempunyai penilaian yang baik terhadap *ecommerce* tersebut. Hal ini terbukti berdasarkan jumlah penilaian pengguna dari 1.923 ulasan, terdapat sebanyak 927 pengguna memberikan penilaian Suka Sekali, 170 pengguna memberikan penilaian Suka dan 230 pengguna memberikan penilaian Lumayan, sedangkan untuk penilaian Tidak Suka berjumlah 208 dan 388 ulasan dengan kategori Tidak Suka Sekali. Akan tetapi banyak pengguna Bukalapak yang tidak sinkron antara isi komentar dengan penilaian rating yang mereka berikan.



**Gambar 5.2** Jumlah ulasan Bukalapak berdasarkan waktu

Pada **Gambar 5.2** menunjukkan perbandingan jumlah ulasan berdasarkan rating yang diberikan oleh pengguna aplikasi Bukalapak. Dapat diketahui bahwa setiap bulannya banyak pengguna yang merasa sangat puas oleh aplikasi milik Bukalapak. Sebanyak pada bulan Januari terdapat 431 pengguna yang suka sekali terhadap Bukalapak, Februari sebanyak 402 pengguna sedangkan bulan maret sebanyak 94 pengguna, karena memang banyak sekali fitur yang ditawarkan oleh Bukalapak kepada penggunanya yang berbeda terhadap *e-commerce* lainnya. Akan tetapi pada bulan januari dan februari banyak yang tidak puas akan aplikasi Bukalapak dikarenakan pembaharuan aplikasi yang membuat kecewa para pengguna dimulai dari akhir bulan januari salah satunya yaitu koneksi keserver sering gangguan.



**Gambar 5.3** Jumlah dan persentase ulasan positif, ulasan negatif dan total ulasan terhadap e-commerce Bukalapak pada bulan Januari hingga Maret 2018

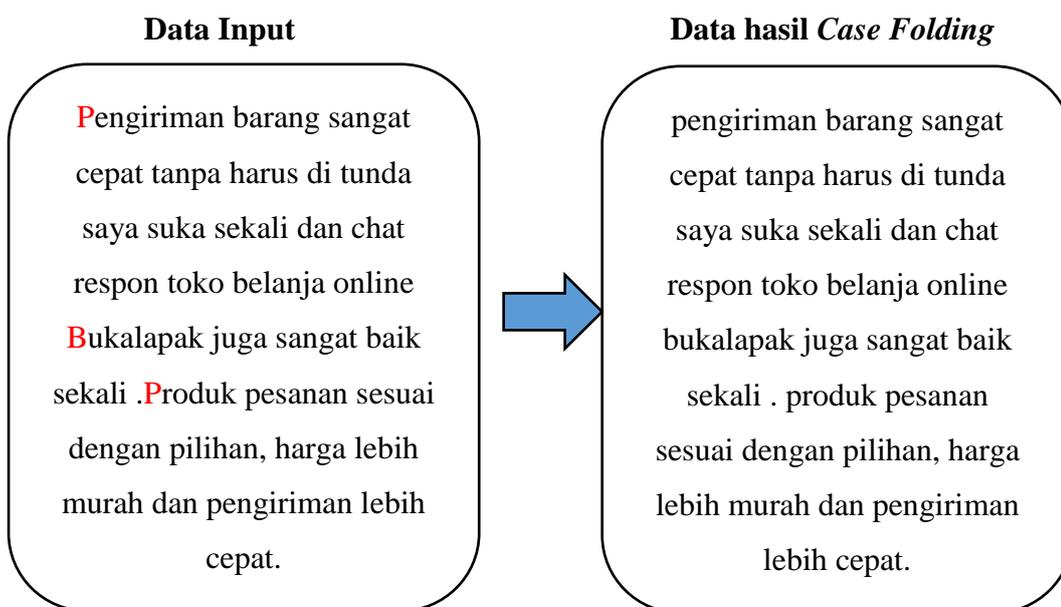
**Gambar 5.3** menjelaskan bahwa sejak bulan Januari hingga bulan Maret 2018, jumlah ulasan pengguna Bukalapak yang diambil dari situs *Google Play* adalah sebanyak 1.923 ulasan. Jika dilihat dari analisis sentimen yang telah dilakukan terdapat jumlah ulasan positif lebih banyak dibandingkan dengan jumlah ulasan negatif. Adapun jumlah ulasan positif yaitu sebanyak 1.063 ulasan atau sebesar 55% sedangkan sisanya merupakan ulasan negatif.

## 5.2. *Preprocessing* atau Prapemrosesan Data

Sebelum melakukan klasifikasi pada dokumen teks, perlu dilakukan *preprocessing*. Data ulasan yang diperoleh belum sepenuhnya siap digunakan untuk proses klasifikasi secara langsung karena data masih tidak terstruktur dengan baik dan terdapat banyak *noise*. Data masih memuat angka, tanda baca, *emoticon*, serta kata-kata lain yang kurang bermakna untuk dijadikan fitur. Maka dari itu, perlu dilakukan *preprocessing* yang bertujuan untuk menyeragamkan



menjadi huruf kecil secara keseluruhan (dari “a” sampai dengan “z”). Hal ini bertujuan agar kata yang ditulis dengan huruf awal kapital dan huruf non kapital tidak terdeteksi memiliki arti yang berbeda. Contoh penggunaan *case folding* dapat dilihat contoh pada **Gambar 5.5** berikut :



**Gambar 5.5** Proses *case folding*

### 5.2.3. *Tokenizing*

*Tokenizing* atau tokenisasi adalah proses memisahkan kata per kata pada sebuah dokumen menjadi kata – kata yang saling independen. *Tokenizing* dilakukan untuk mendapatkan token atau potongan kata yang akan menjadi entitas yang memiliki nilai dalam penyusunan matriks dokumen pada proses selanjutnya. Tokenisasi dapat memudahkan proses perhitungan keberadaan kata tersebut dalam dokumen ataupun untuk menghitung frekuensi kemunculan kata tersebut dalam *corpus*. Contoh proses *tokenizing* ditunjukkan pada **Gambar 5.6** berikut :

pengiriman barang sangat cepat tanpa harus di tunda saya suka sekali dan chat respon toko belanja online bukalapak juga sangat baik sekali . produk pesanan sesuai dengan pilihan, harga lebih murah dan pengiriman lebih cepat.



|            |        |           |         |         |            |
|------------|--------|-----------|---------|---------|------------|
| pengiriman | di     | chat      | juga    | sesuai  | dan        |
| barang     | tunda  | respon    | sangat  | dengan  | pengiriman |
| sangat     | saya   | toko      | baik    | pilihan | lebih      |
| cepat      | suka   | belanja   | sekali  | harga   | cepat      |
| tanpa      | sekali | online    | produk  | lebih   |            |
| harus      | dan    | bukalapak | pesanan | murah   |            |

**Gambar 5.6** 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.

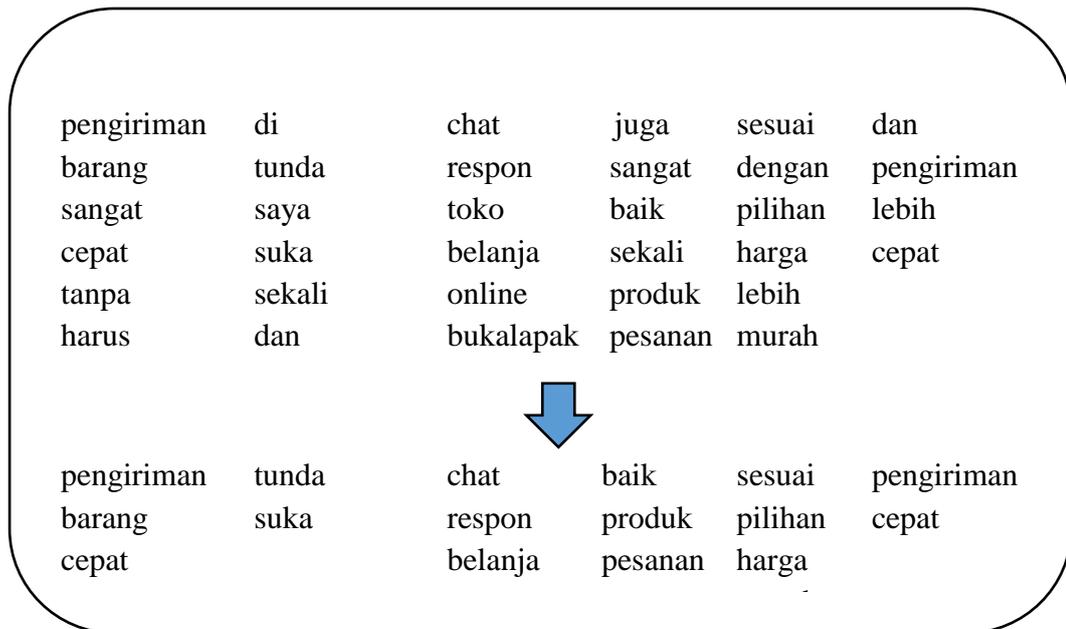
#### **5.2.4. Filtering**

Tahap penyaringan atau *filtering* merupakan tahap dilakukannya pemilihan kata pada dokumen atau pengurangan dimensi kata di dalam *corpus* yang disebut *stopwords*. *Stopwords* merupakan tahap untuk menghilangkan kata-kata yang tidak berpengaruh / tidak informatif namun seringkali muncul dalam dokumen. Kata-kata tersebut seperti kata penghubung, kata ganti orang, kata seruan dan kata lainnya yang tidak begitu memiliki arti dalam penentuan kelas topik suatu dokumen. Adapun kata-kata yang akan dihilangkan yaitu:

- a Penghubung antar kata, contoh: dan, atau, serta
- b Preposisi, contoh: ke, di, pada

- c Kata-kata yang tidak diinginkan (menggunakan perintah *remove words*), yaitu kata “bukalapak”, “toko”, dan “online”.

Contoh proses *filtering* dapat dilihat pada **Gambar 5.7** berikut :



**Gambar 5.7** Proses *filtering*

### 5.3. Pelabelan Kelas Sentimen

Setelah melalui proses *preprocessing*, selanjutnya akan dilakukan analisis sentimen untuk pelabelan data. Proses pelabelan data dilakukan secara otomatis oleh *kamus lexicon* dengan cara menghitung skor sentimen. Pembobotan kata dilakukan dengan menghitung frekuensi kemunculan kata pada sebuah dokumen teks. Semakin sering sebuah kata muncul pada sebuah dokumen teks, maka bobot kata tersebut semakin besar dan kata tersebut dianggap sebagai kata yang sangat merepresentasikan dokumen teks tersebut (Yates dan Neto, 1999 dikutip dalam Basnur, 2009).

Pada umumnya, analisis sentimen digunakan untuk melakukan klasifikasi (pelabelan) dokumen teks ke dalam tiga kelas sentimen, yaitu sentimen positif, negatif dan netral. Cara menentukan kelas sentimen adalah dengan menghitung skor jumlah kata positif dikurangi skor jumlah kata negatif dalam setiap kalimat

ulasan (Susanti, 2016). Kalimat yang memiliki skor  $> 0$  akan diklasifikasikan ke dalam kelas positif, kalimat yang memiliki skor  $= 0$  akan diklasifikasikan ke dalam kelas netral, sedangkan kalimat yang memiliki skor  $< 0$  diklasifikasikan ke dalam kelas negatif. Akan tetapi pada penelitian ini digunakan dua pelabelan kelas sentimen, yaitu sentimen positif dan sentimen negatif. Hal ini dilakukan karena menimbang bahwa kelas sentimen netral dianggap kurang memberikan manfaat bagi pihak Bukalapak. Berikut ini adalah tahapan melakukan pelabelan dengan menggunakan *Software R*.

**Tabel 5.1** Tahap-tahap pelabelan menggunakan software R

| Script R   | Fungsi   |
|--|--|
| <pre>library(tm) setwd("E:/BISMILLAH TUGAS AKHIR") kalimat2&lt;-read.csv   ("cleaning.csv",header=TRUE)</pre>  | <ol style="list-style-type: none"> <li>1. Menjalankan <i>packages</i> "tm" yang telah terinstal pada program R</li> <li>2. Mengatur direktori kerja dalam program R</li> <li>3. Membuka <i>file csv</i> yang akan diberi label</li> </ol>                      |
| <pre>positif &lt;- scan("E:/BISMILLAH TUGAS AKHIR/Bismillah TA Ditia/Analisis/s- pos.txt",what="character",comment.char=";" ) negatif &lt;- scan("E:/BISMILLAH TUGAS AKHIR/Bismillah TA Ditia/Analisis/s- neg.txt",what="character",comment.char=";" ) kata.positif = c(positif, "senang") kata.negatif = c(negatif, "kecewa")</pre> | <ol style="list-style-type: none"> <li>4. Melakukan <i>scanning file</i> daftar kata positif dan kata negatif yang tersimpan dalam format <i>.txt file</i></li> </ol>  |
| <pre>score.sentiment = function(kalimat2, kata.positif, kata.negatif, .progress='none') { require(plyr) require(stringr) scores = laply(kalimat2, function(kalimat, kata.positif, kata.negatif) { kalimat = gsub('[:punct:]', '', kalimat) kalimat = gsub('[:cntrl:]', '', kalimat) kalimat = gsub('\\d+', '', kalimat)</pre>        | <ol style="list-style-type: none"> <li>5. Melakukan proses skoring menggunakan <i>function</i> dengan tahapan : <ol style="list-style-type: none"> <li>a. Menjalankan <i>packages plyr</i> dan <i>stringr</i></li> <li>b. Menggabungkan</li> </ol> </li> </ol> |

| Script R   | Fungsi   |
|--|--|
| <pre> kalimat = tolower(kalimat) list.kata = str_split(kalimat, '\\s+') kata2 = unlist(list.kata) positif.matches = match(kata2, kata.positif) negatif.matches = match(kata2, kata.negatif) positif.matches = !is.na(positif.matches)  negatif.matches = !is.na(negatif.matches) score = sum(positif.matches) - (sum(negatif.matches)) return(score) }, kata.positif, kata.negatif, .progress=.progress ) scores.df = data.frame(score=scores, text=kalimat2) return(scores.df) } </pre> | <p>setiap daftar inisial menjadi sebuah <i>array</i></p> <ul style="list-style-type: none"> <li>c. Menghapus <i>noisedan</i> melakukan <i>case folding</i></li> <li>d. Merubah kalimat menjadi potongan kata (<i>tokenizing</i>) dan menyederhanakan daftar kata</li> <li>e. Mengidentifikasi kata positif dan kata negatif pada setiap potongan kata</li> <li>f. Mengindikasi kata positif dan kata negatif ke dalam bentuk logika</li> <li>g. Menghitung jumlah skor sentiment Menyimpan skor dan kalimat ke dalam bentuk tabel</li> </ul> |
| <pre> hasil = score.sentiment(kalimat2\$text, kata.positif, kata.negatif)  hasil\$klasifikasi&lt;- ifelse(hasil\$score&lt;0, "Negatif","Positif") data &lt;- hasil[c(3,1,2)]  write.csv(data, file = "hasil_pelabelan.csv") </pre>   | <ul style="list-style-type: none"> <li>6. Memanggil <i>function</i> hasil skoring yang telah dibuat</li> <li>7. Melakukan konversi nilai skor ke dalam kelas positif dan negatif</li> <li>8. Menyimpan file hasil pelabelan ke dalam format <i>csv</i>.</li> </ul>   |

Adapun hasil pelabelan kelas sentimen diperoleh jumlah data seperti berikut :

**Tabel 5.2** *Perbandingan jumlah data pada kelas sentiment*

| Sentimen | Jumlah Ulasan |
|----------|---------------|
| Positif  | 1.063         |
| Negatif  | 860           |

Berdasarkan **Tabel 5.2**, hasil pelabelan kelas sentimen menunjukkan bahwa jumlah ulasan positif memiliki frekuensi yang lebih tinggi dibandingkan dengan jumlah ulasan negatif. Jumlah ulasan positif sebanyak 1.063 ulasan, dan ulasan negatif sebanyak 860 ulasan.

Klasifikasi yang akan digunakan pada penelitian ini hanyalah data dengan sentimen positif dan negatif. Suatu ulasan diklasifikasikan sebagai sentimen positif bila mengandung pernyataan positif seperti pujian, ungkapan terima kasih, atau testimoni positif tentang *e-commerce* Bukalapak . Suatu ulasan diklasifikasikan sebagai sentimen negatif bila mengandung pernyataan-pernyataan negatif seperti ketidakpuasan, penghinaan, laporan kegagalan layanan, dan sebagainya. Hasil pelabelan data ulasan dapat dilihat pada **Tabel 5.3** berikut:

**Tabel 5.3** *Hasil pelabelan kelas sentiment berbasis kamus lexicon dan proses manual*

| Kelas Sentimen | Skor | Ulasan   |
|----------------|------|--|
| Positif        | 3    | aplikasi membantu barang laku cepat tanggap menangani transaksi aman pembeli penjual                                     |
| Positif        | 5    | pengiriman barang cepat tunda suka chat respon belanja baik produk pesananan sesuai pilihan harga murah pengiriman cepat |

| Kelas Sentimen | Skor | Ulasan  |
|----------------|------|---|
| Negatif        | -3   | kecewa notifikasi pesan akun email telat respon pembeli kesal   |
| Negatif        | -2   | kecewa pesan barang paket pengiriman express kirim stok barang konfirmasi telepon email rugi uang membayar pengiriman express |

### 5.3.1. Simulasi Perhitungan Skor Sentimen

Berdasarkan teks ulasan “pengiriman barang cepat tunda suka chat respon belanja baik produk pesananan sesuai pilihan harga murah pengiriman cepat”, terdapat 6 kata positif dan 1 kata negatif yang terdeteksi pada kamus *lexicon*, yakni “cepat”, “suka”, “baik”, “esuai”, “murah” sebagai kata positif, dan “tunda-tunda” sebagai kata negatif. Adapun rumus perhitungan skor sentiment yang digunakan dalam proses pelabelan adalah sebagai berikut:

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

**Tabel 5.4** Simulasi perhitungan skor sentiment

| Teks Ulasan  | Kata Positif                                      | Kata Negatif |
|--|---|--------------|
| pengiriman barang<br>cepat tunda suka<br>chat respon belanja<br>baik produk<br>pesananan sesuai<br>pilihan harga murah<br>pengiriman cepat | cepat<br>suka<br>baik<br>sesuai<br>murah<br>cepat | tunda        |
| <b>Jumlah</b>  | <b>6</b>  | <b>1</b>     |

Sehingga dengan demikian diperoleh perhitungan sebagai berikut :

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

$$\text{Skor} = 6 - 1 = 5$$

Skor akhir yang diperoleh dari simulasi perhitungan bernilai  $> 0$ , sehingga hasil dari klasifikasi ulasan tersebut positif.

#### 5.4. 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%. Suthaharan (2015) menyatakan bahwa meskipun penelitian ekstensif belum dilakukan dalam pemilihan rasio yang optimal antara kumpulan data ini, ada beberapa praktik umum dalam memilih ukuran kumpulan data ini. 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      | 1.063        | 850,4 $\approx$ 850 | 212,6 $\approx$ 213 |
| Negatif      | 860          | 688                 | 172                 |
| <b>Total</b> | <b>1.923</b> | <b>1.538</b>        | <b>385</b>          |

Berdasarkan **Tabel 5.5**, dengan perbandingan data latih dan data uji sebesar 80%:20% dari total 1.923 ulasan berbahasa Indonesia, digunakan sebanyak 1.538 ulasan sebagai data latih dan 385 ulasan sebagai data uji.

#### 5.5. Klasifikasi dengan *Support Vector Machine* dan *Maximum Entropy*

Dari proses pelatihan data tersebut akan didapatkan model klasifikasi dari masing-masing *machine learning*. Model tersebut selanjutnya akan diuji untuk mengetahui tingkat akurasi model atau sejauh mana model tersebut dapat mengklasifikasikan data uji, proses inilah yang disebut sebagai *meachine learning*. Data latih positif dan data latih negatif digunakan oleh algoritma *SVM* dan algoritma *Maximum Entropy* dalam mempelajari pola data berdasarkan ciri-ciri data pada masing-masing kelas.

Dalam penelitian dilakukan percobaan menggunakan beberapa *kernel* yaitu *kernel Linear*, *Polynomial*, *Radial Basis Function (RBF)*, dan *Sigmoid* untuk memperoleh klasifikasi dengan hasil akurasi terbaik pada SVM. Berikut ini adalah hasil perbandingan dari keempat *kernel* yang telah di uji coba dapat dilihat pada **Tabel 5.6** berikut :

**Tabel 5.6** Perbandingan penggunaan metode *kernel* pada klasifikasi SVM

| <b>Kernel</b>     | <b>Akurasi</b> |
|-------------------|----------------|
| <i>Linear</i>     | 90,39%         |
| <i>Polynomial</i> | 55,32%         |
| <i>RBF</i>        | 91,95%         |
| <i>Sigmoid</i>    | 91,17%         |

Pada **Tabel 5.6** dapat diketahui bahwa dari keempat *kernel* yang telah diuji, *kernel Radial Basis Function (RBF)* memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan metode *kernel* lainnya, dengan demikian pada penelitian ini digunakan *kernel RBF* dalam melakukan proses klasifikasi. Selain menggunakan algoritma SVM, penulis juga menggunakan algoritma *Maximum Entropy* untuk memperoleh nilai akurasi dari klasifikasi. Adapun hasil dari algoritma *Maximum Entropy* adalah sebagai berikut :

**Tabel 5.7** Klasifikasi *Maximum Entropy*

| <b>Metode</b>          | <b>Akurasi</b> |
|------------------------|----------------|
| <i>Maximum Entropy</i> | 92,98%         |

Adapun tahap-tahap melakukan klasifikasi ditampilkan dalam **Tabel 5.8** berikut :

**Tabel 5.8** Tahap melakukan analisis SVM dan Maximum Entropy dengan software R

| Script R   | Fungsi   |
|--|--|
| <pre>setwd("E:/BISMILLAH TUGAS AKHIR") positifL = readLines("PL.csv") negatifL = readLines("NL.csv") positifT = readLines("PT.csv") negatifT = readLines("NT.csv")  reviewL = c(positifL, negatifL) reviewT= c(positifT, negatifT) review_all = c(reviewL,reviewT) sentiment_trainingc(rep("positifL",   length(positifL) ),   rep("negatifL", length(negatifL))) sentiment_test = c(rep("positifL",   length(positifT) ),   rep("negatifL",   length(negatifT))) sentiment_all = as.factor(c(sentiment_training, sentiment_test))</pre> | <ol style="list-style-type: none"> <li>1. Mengatur direktori kerja pada program R</li> <li>2. Membuka <i>file</i> data latih dan data uji dalam format <i>csv</i></li> <li>3. Mendefinisikan masing-masing data latih dan data uji</li> <li>4. Menggabungkan data latih dan data uji yang telah terdefinisi</li> <li>5. Mendefinisikan label kelas pada data latih dan data uji</li> <li>6. Menggabungkan label kelas dan mengubah tipe data menjadi tipe data faktor</li> </ol> |
| <pre>library(RTextTools) library(e1071)</pre>  | <ol style="list-style-type: none"> <li>7. Menjalankan <i>packages</i> <i>RTextTools</i> dan <i>e1071</i></li> </ol>  |
| <pre>mat = create_matrix(review_all, language = "indonesian", removeStopwords = FALSE, removeNumbers = TRUE, stemWords = FALSE, tm::weightTfIdf)  mat = as.matrix(mat)</pre>   | <ol style="list-style-type: none"> <li>8. Membuat objek kelas <i>DocumentTermMatrix</i></li> <li>9. Mengubah data ke dalam bentuk <i>matrix</i></li> </ol>   |

| Script R   | Fungsi  |
|--|---|
| <pre>container&lt;-create_container(mat, sentiment_all,  trainSize=1:1538,testSize=1539:1923, virgin=FALSE)</pre>  | 10. Membuat wadah untuk proses <i>training</i> dan <i>testing</i> data  |
| <pre>model &lt;- train_model(container, 'SVM',kernel='radial') results &lt;- classify_model(container, model) table(as.character(sentiment_all[1539:1923]) , as.character(results[,"SVM_LABEL"]))</pre>          | 11. Melakukan <i>training</i> untuk mendapatkan model dengan algoritma <i>SVM</i><br>12. Menggunakan model data <i>training</i> untuk mengklasifikasikan data baru<br>13. Membuat tabel <i>confusion matrix</i>             |
| <pre>recall_accuracy(sentiment_all[1539:1923], results[,"SVM_LABEL"]) create_precisionRecallSummary(container, results)</pre>  | 14. Menghitung nilai akurasi<br>15. Menghitung nilai <i>precision</i> dan <i>recall</i>   |
| <pre>models &lt;- train_models(container, algorithms="MAXENT") results &lt;- classify_models(container, models) table(as.character(sentiment_all[1539:1923]) , as.character(results[,"MAXENTROPY_LABEL"]))</pre> | 16. Melakukan <i>training</i> untuk mendapatkan model dengan algoritma <i>Maximum Entropy</i><br>17. Menggunakan model data <i>training</i> untuk mengklasifikasikan data baru<br>18. Membuat tabel <i>confusion matrix</i> |
| <pre>recall_accuracy(sentiment_all[1539:1923], results[,"MAXENTROPY_LABEL"]) create_precisionRecallSummary(container, results)</pre>   | 19. Menghitung nilai akurasi<br>20. Menghitung nilai <i>precision</i> dan <i>recall</i>   |

Proses klasifikasi dilakukan dengan cara membuat *machine learning* menggunakan data latih dan data uji secara acak. Penelitian ini menggunakan

metode *confusion matrix* dalam proses evaluasi. *Confusion matrix* merupakan salah satu *tools* penting dalam metode evaluasi yang digunakan pada *machine learning* yang biasanya memuat dua kategori atau lebih (Manning, dkk, 2009). Setiap unsur matiks menunjukkan jumlah contoh data uji untuk kelas sebenarnya yang digambarkan dalam bentuk baris sedangkan kolom menggambarkan kelas yang diprediksi. Dalam melakukan evaluasi model, pada percobaan ini dilakukan dengan membuat 5 buah *machine learning* untuk menemukan nilai akurasi prediksi terbaik. Adapun hasil masing-masing percobaan *machine learning* menggunakan metode *Support Vector Machine* dan *Maximum Entropy* adalah sebagai berikut :

**Tabel 5.9** Perbandingan nilai akurasi *machine learning* dengan metode *SVM* dan *Maximum Entropy*

| <i>Machine Learning</i>   | <b>Akurasi Model</b> |               |
|---------------------------|----------------------|---------------|
|                           | <i>SVM</i>           | <i>Maxent</i> |
| <i>Machine Learning 1</i> | 91,95%               | 92,98%        |
| <i>Machine Learning 2</i> | 87,53%               | 91,43%        |
| <i>Machine Learning 3</i> | 82,08%               | 84,93%        |
| <i>Machine Learning 4</i> | 82,86%               | 82,60%        |
| <i>Machine Learning 5</i> | 77,92%               | 75,06%        |

Berdasarkan **Tabel 5.9** diatas, dari 5 percobaan *meachine learning* yang dilakukan menggunakan metode *SVM* dan *Maxent*, *machine learning* 1 menghasilkan tingkat akurasi tertinggi yakni untuk metode *SVM* sebesar 91,95% dan untuk metode *Maximum Entropy* sebesar 92,98%. Hasil perhitungan tingkat akurasi diperoleh dari jumlah data uji yang terklasifikasi dengan benar dibandingkan dengan total semua data yang di uji.

Untuk menguji performa *machine* dalam melakukan klasifikasi, maka dilakukan *cross validation* menggunakan *5-fold cross validation* dengan hasil rata-rata akurasi diperoleh sebesar 84,47% untuk metode *SVM* sedangkan 85,40% untuk metode *Maximum Entropy*.

*Confusion matrix* digunakan untuk memudahkan dalam proses perhitungan akurasi dengan mengetahui jumlah data uji yang terklasifikasi dengan benar dan jumlah data uji yang salah pengklasifikasiannya. Adapun perbandingan *confusion matrix* kedua metode yang diperoleh pada *machine learning 1* dapat dilihat pada **Tabel 5.10** berikut :

**Tabel 5.10** *Confusion matrix*

| Prediksi       | SVM     |         | Maxent         |         |
|----------------|---------|---------|----------------|---------|
|                | Positif | Negatif | Positif        | Negatif |
| Positif        | 202     | 20      | 203            | 17      |
| Negatif        | 11      | 152     | 10             | 155     |
| <b>Akurasi</b> |         |         | <b>Akurasi</b> |         |
| 91,95%         |         |         | 92,98%         |         |

Berdasarkan **Tabel 5.10**, menjelaskan seberapa besar persentase data yang dapat diprediksi dengan benar oleh masing-masing *machine learning*. Banyaknya data observasi berkategori negatif yang mampu diprediksi negatif (diprediksi dengan tepat) oleh *machine learning* disebut dengan *true negatif*. Banyaknya data observasi berkategori positif yang mampu diprediksi positif (diprediksi dengan tepat) oleh *machine learning* disebut dengan *true positif*. Banyaknya data observasi yang berkategori positif akan tetapi terdapat kesalahan prediksi disebut dengan *false positif*. Banyaknya data observasi yang berkategori negatif akan tetapi terdapat kesalahan prediksi disebut dengan *false negatif*.

Pada metode *SVM* diperoleh hasil prediksi bahwa pada kelas positif, dari 213 ulasan positif yang diuji, terdapat 202 ulasan yang sudah terklasifikasi dengan benar dan terdapat kesalahan prediksi sebesar 11 ulasan yang masuk kedalam ulasan negatif. Sedangkan pada ulasan negatif yang diuji, dari total 172 ulasan terdapat 152 ulasan yang sudah terklasifikasi dengan benar sebagai ulasan negatif dan terdapat kesalahan prediksi sebanyak 20 ulasan yang masuk ke dalam ulasan positif. Kemudian dari nilai *confusion matrix* tersebut diperoleh tingkat akurasi

sebesar 91,95%, artinya dari 385 data ulasan yang diujikan, terdapat 354 ulasan yang benar pengklasifikasiannya oleh model *Support Vector Machine (SVM)*.

Sedangkan dengan menggunakan metode *maxent* diperoleh hasil prediksi bahwa pada kelas positif, dari 213 ulasan positif terdapat 203 ulasan yang sudah terklasifikasi dengan benar dan terdapat kesalahan prediksi sebesar 10 ulasan yang masuk kedalam ulasan negatif. Sedangkan pada ulasan negatif, dari total 172 ulasan terdapat 155 ulasan yang sudah terklasifikasi dengan benar sebagai ulasan negatif dan terdapat kesalahan prediksi sebanyak 17 ulasan yang masuk ke dalam ulasan positif. Kemudian dari nilai *confusion matrix* tersebut diperoleh tingkat akurasi sebesar 92,98%, artinya dari 385 data ulasan yang diujikan, terdapat 358 ulasan yang benar pengklasifikasiannya oleh model *Maximum Entropy (Maxent)*. Jika dibandingkan dengan metode *Support Vector Machine (SVM)*, metode *Maximum Entropy (Maxent)* memiliki tingkat akurasi yang lebih tinggi.

Adapun proses perhitungan nilai akurasi dilakukan dengan menggunakan rumus berikut :

$$Akurasi = \frac{\text{jumlah data yang diprediksi benar}}{\text{jumlah semua data yang di uji}} \times 100\% \quad (5.2)$$

Sehingga, untuk metode *SVM* nilai akurasi diperoleh dari perhitungan berikut:

$$Akurasi = \frac{202 + 152}{202 + 11 + 20 + 152} \times 100\%$$

$$Akurasi = \frac{354}{385} \times 100\%$$

$$Akurasi = 91,95\%$$

Adapun nilai akurasi metode *Maxent* diperoleh dari perhitungan berikut:

$$Akurasi = \frac{203 + 155}{203 + 10 + 17 + 155} \times 100\%$$

$$Akurasi = \frac{358}{385} \times 100\%$$

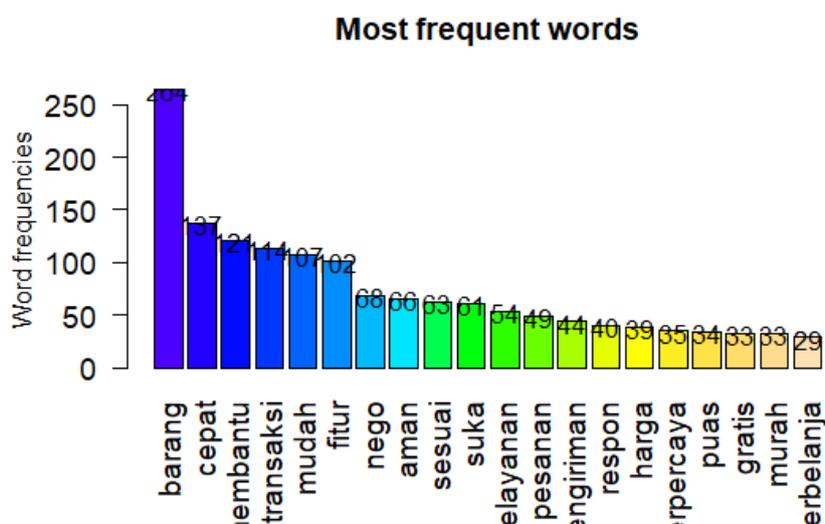
$$Akurasi = 92,98$$

## 5.6. Visualisasi dan Asosiasi

Visualisasi dilakukan terhadap masing-masing klasifikasi kelas sentimen. Adapun tujuan visualisasi adalah untuk mengekstraksi informasi berupa topik yang paling sering di bicarakan / diulas oleh pengguna Bukalapak, sehingga dari sekian banyak teks ulasan yang ada, dapat diambil informasi yang dianggap penting serta dicari asosiasi antar kata yang paling sering muncul secara bersamaan, sehingga mampu memperkuat pencarian informasi tersebut. Berikut penjelasan hasil visualisasi dan asosiasi kata dari setiap klasifikasi kelas sentimen.

### 5.6.1. Ulasan Positif

Data ulasan positif yang digunakan adalah data hasil pelabelan yang dilakukan baik menggunakan kamus *lexicon* maupun secara manual. Ekstraksi informasi pada ulasan positif dilakukan secara berulang-ulang hingga mendapatkan informasi tentang ulasan positif pengguna Bukalapak yang paling sering diulas / dibicarakan. Ulasan positif tersebut diidentifikasi berdasarkan frekuensi kata dalam ulasan, berikut adalah visualisasi hasil ekstraksi informasi yang didapatkan dari ulasan pengunjung dengan klasifikasi ulasan positif.



**Gambar 5.8** Kata yang paling banyak muncul dari kelas positif *e-commerce Bukalapak*



positif yang didalamnya mengandung kata “barang”, sedangkan **Gambar 5.9 (b)** merupakan *wordcloud* pada ulasan positif tanpa menggunakan kata “barang”. Selanjutnya, dilakukan pencarian asosiasi antar kata yang sering muncul secara bersamaan dan diperoleh hasil sebagai berikut :

**Tabel 5.11** Asosiasi kata pada akelas sentimen positif

| Barang           |      | Transaksi   |      | Fitur      |      |
|------------------|------|-------------|------|------------|------|
| ketersediaan     | 0,30 | mekanisme   | 0,29 | diskusi    | 0,33 |
| ready            | 0,22 | keamanan    | 0,25 | membantu   | 0,18 |
| sesuai           | 0,20 | menjamin    | 0,20 | fleksibel  | 0,18 |
| favorit          | 0,15 | berkualitas | 0,20 | mudah      | 0,18 |
| Pelayanan        |      | Pesanan     |      | Pengiriman |      |
| memuaskan        | 0,61 | sesuai      | 0,33 | jasa       | 0,47 |
| bermutu          | 0,50 | packing     | 0,19 | pickup     | 0,26 |
| memperioritaskan | 0,24 | cepat       | 0,16 | cepat      | 0,19 |
| menerima         | 0,24 | Berbelanja  |      | gosend     | 0,19 |
| responsive       | 0,24 | kooperatif  | 0,18 | Kebutuhan  |      |
| Respon           |      | mempermudah | 0,16 | terpenuhi  | 0,30 |
| cepat            | 0,31 | rewards     | 0,18 | membantu   | 0,17 |
| konfirmasi       | 0,19 | Akulaku     |      | Cicilan    |      |
| professional     | 0,15 | merchant    | 0,30 | ringan     | 0,71 |
| mempermudahkan   | 0,15 | cicilannya  | 0,30 |            |      |

Berdasarkan **Tabel 5.11** diperoleh asosiasi kata pada klasifikasi kelas positif. Proses ekstraksi informasi dengan asosiasi dilakukan secara berulang-ulang dengan cara menyaring kata-kata yang memiliki hubungan dengan kata lain dan didasarkan pada relevansi kata dengan topik yang di ulas. Dari **Tabel 5.11** jika dilihat asosiasi kata yang berkaitan dengan kata “barang”, dapat diperoleh informasi tentang barang yang dijual di Bukalapak dinilai ketersediaan dan barang ready, menjaga loyalitas pembeli dengan menjaga stok ketersediaan barang. Pembeli di Bukalapak menilai bahwa barang yang dijual di Bukalapak sesuai dengan apa yang diinginkan dan barang yang dijual juga menjadi favorit.

Kata-kata yang berasosiasi dengan kata “transaksi” memberikan informasi tentang transaksi di Bukalapak yang mekanisme pembayarannya terjamin keamanannya dan berkualitas.

Kata-kata yang berasosiasi dengan kata “fitur” memberikan informasi tentang fitur diskusi yang dapat membantu calon pembeli untuk bertanya atau berdiskusi kepada penjual terkait barang yang akan dibelinya. Diskusi dalam bentuk tanya jawab yang dilakukan oleh calon pembeli dan pelapak akan ditampilkan pada halaman detail barang, sehingga calon pembeli lain juga dapat melihat diskusi yang sudah ada. Selain itu jual beli barang atau jasa lebih fleksibel karena ada fitur Beli Instan yang memudahkan *user* Bukalapak baik pembeli atau Pelapak, dapat melakukan transaksi jual beli barang atau jasa yang tidak ada di Lapak dengan harga dan ongkos kirim sesuai kesepakatan.

Kata-kata yang berasosiasi dengan “pelayanan” memberikan informasi bahwa pelayanan yang telah diberikan oleh Bukalapak kepada pengguna sudah memuaskan, bermutu, memprioritaskan pembeli, menerima komplain dengan responsif.

Kata-kata yang berasosiasi dengan “pesanan” memberikan informasi tentang pesanan yang diterima oleh pembeli sesuai dengan yang diinginkan, selain itu pesanan juga dinilai packingnya atau cara pengemasan pesanan dan pesanan sampai ke pembeli dengan cepat tanpa menunggu waktu yang lama.

Kata-kata yang berasosiasi dengan “pengiriman” memberikan informasi tentang jasa *pickup* yang cepat dalam mengirim barang seperti Go-Send untuk pilihan layanan pengiriman dari Go-jek, barang akan sampai dalam hitungan jam pada hari yang sama.

Kata-kata yang berasosiasi dengan “respon” memberikan informasi tentang pelapak maupun *customer service* yang membalas cepat konfirmasi kepada pembeli.

Kata-kata yang berasosiasi dengan “berbelanja” memberikan informasi tentang kooperatif, mempermudah dalam berbelanja dan mendapatkan rewards.

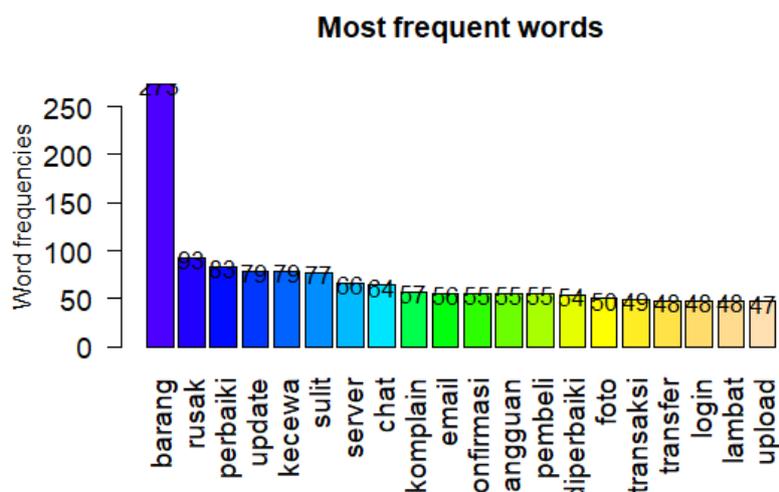
Kata-kata yang berasosiasi dengan “kebutuhan” memberikan informasi tentang pembeli yang merasa belanja di Bukalapak membantu dalam memenuhi kebutuhan, karena mulai dari peralatan rumah tangga, produk fashion hingga perlengkapan otomotif, semuanya tersedia di Bukalapak.

Kata-kata yang berasosiasi dengan “akulaku” memberikan informasi bahwa salah satu cara pembayaran dengan akulaku dinilai memudahkan pembeli berbelanja secara kredit, dapat membeli produk apapun di *merchant partner* Akulaku dan membayar cicilannya 1 bulan kemudian atau cicilan 2 bulan, 3 bulan dan 6 bulan dengan uang muka untuk semua barang.

Kata-kata yang berasosiasi dengan “cicilan” memberikan informasi bahwa cicilan belanja yang ditawarkan oleh Bukalapak termasuk ringan sehingga pembeli dapat mengangsur pembayaran dalam kurun waktu tertentu hingga lunas.

### 5.6.2. Ulasan Negatif

Ekstraksi informasi pada ulasan negatif dilakukan secara berulang-ulang hingga mendapatkan informasi tentang ulasan negatif pengguna Bukalapak yang paling sering di ulas/dibicarakan. Dari total ulasan sebanyak 1.923 ulasan, teridentifikasi sebanyak 860 ulasan negatif. Hasil ekstraksi informasi berupa ulasan negatif diidentifikasi berdasarkan frekuensi kata dalam ulasan, selain itu juga didasarkan pada relevansi kata dengan topik yang mengacu pada sentimen negatif. Berikut adalah visualisasi hasil ekstraksi informasi yang didapatkan dari ulasan pengunjung dengan klasifikasi negatif.



**Gambar 5.10** Kata yang paling banyak muncul dari kelas negatif

Pada hasil klasifikasi ulasan negatif diperoleh beberapa kata yang paling banyak muncul dengan topik yang dianggap relevan sebagai sentimen negatif diantaranya adalah kata “barang” dengan frekuensi sebanyak 273 kali, “aplikasi” dan “rusak” sebanyak 93 kali, “perbaiki” sebanyak 83 kali, “update” sebanyak 79 kali, dan seterusnya. Kata-kata yang muncul seperti pada **Gambar 5.10** merupakan kata yang memiliki sentiment negatif dan merupakan topik pembicaraan yang paling banyak di ulas oleh pengguna Bukalapak. Kata-kata tersebut selanjutnya digunakan sebagai dasar untuk menemukan asosiasi lainnya, sehingga dapat diperoleh informasi berupa sentiment negatif yang lebih akurat. Kumpulan kata-kata yang sering muncul tersebut ditampilkan dalam bentuk *wordcloud* seperti terlihat pada **Gambar 5.11**.



**Gambar 5.11** *Wordcloud ulasan negatif*

Visualisasi *wordcloud* pada **Gambar 5.11** memberikan gambaran yang lebih jelas tentang topik dan kata-kata negatif yang sering digunakan pengguna Bukalapak dalam memberikan ulasan. Beberapa topik yang sering dibahas pengguna diantaranya adalah tentang barang, update, server, chat dan sebagainya. Selanjutnya, dilakukan pencarian asosiasi antar kata yang sering muncul secara bersamaan dan diperoleh hasil sebagai berikut :

**Tabel 5.12** Asosiasi kata pada kelas sentimen negatif

| Barang        |      | Update           |      | Server      |      |
|---------------|------|------------------|------|-------------|------|
| keterangan    | 0,21 | history          | 0,24 | gangguan    | 0,63 |
| belum         | 0,19 | resah            | 0,20 | koneksi     | 0,39 |
| diterima      | 0,19 | berimbas         | 0,20 | login       | 0,27 |
| return        | 0,15 | ketidakpercayaan | 0,20 | facebook    | 0,16 |
| rusak         | 0,15 | penanganan       | 0,20 | overload    | 0,16 |
|               |      | web              | 0,17 |             |      |
|               |      | parah            | 0,16 |             |      |
| Chat          |      | Email            |      | Transaksi   |      |
| terlambat     | 0,39 | mengirim         | 0,36 | kadaluarsa  | 0,31 |
| pemberitahuan | 0,28 | verifikasi       | 0,30 | dibatalkan  | 0,27 |
|               |      | riset            | 0,26 | status      | 0,15 |
|               |      | password         | 0,25 |             |      |
|               |      | bantuannya       | 0,24 |             |      |
|               |      | gagal            | 0,16 |             |      |
|               |      | dipersulit       | 0,16 |             |      |
| Upload        |      | Promo            |      | Voucher     |      |
| foto          | 0,49 | diperbanyak      | 0,34 | kode        | 0,54 |
| gagal         | 0,28 | digabung         | 0,37 | iming-iming | 0,51 |
| pembaruan     | 0,26 | voucher          | 0,32 | perbanyak   | 0,36 |
| produk        | 0,20 | terbatas         | 0,26 | kehabisan   | 0,36 |
|               |      | ongkir           | 0,24 | konsisten   | 0,25 |
| Bukadompet    |      | Upgrade          |      | dibatalkan  | 0,25 |
| topup         | 0,46 | keseringan       | 0,29 | perpanjang  | 0,25 |
| dibekukan     | 0,37 | maintanance      | 0,29 | membatasi   | 0,24 |
| isi           | 0,37 | performanya      | 0,29 | penipuan    | 0,23 |
| registrasi    | 0,37 | jelek            | 0,17 |             |      |

**Tabel 5.12** 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 “barang” pada ulasan negatif memberikan informasi tentang keluhan pengguna Bukalapak terkhusus pembeli terhadap keterangan barang yang kurang detail, barang yang sudah lama dipesan tetapi belum diterima dan ada pengguna bukalapak yang mengajukan *return* dikarenakan barang rusak.

Kata-kata yang berasosiasi dengan kata “update” pada ulasan negatif memberikan informasi tentang keluhan pengguna Bukalapak terhadap history

pengiriman pesanan yang membuat pembeli merasa resah dan menurunkan loyalitas pembeli seperti munculnya rasa ketidakpercayaan pembeli untuk berbelanja di Bukalapak. Oleh sebab itu dibutuhkan penanganan web untuk mengatasi update history pengiriman supaya tidak mengecewakan pembeli.

Kata-kata yang berasosiasi dengan kata “server” pada ulasan negatif memberikan informasi tentang keluhan pengguna Bukalapak terhadap koneksi server gangguan yang bisa terjadi karena upgrade aplikasi, membuat pengguna sulit untuk login ke aplikasi meskipun sudah login melalui facebook karena server overload. Istilah overload digunakan untuk menggambarkan kondisi ketika usaha suatu objek tidak mampu lagi untuk mengakomodasi beban atau jumlah yang melebihi kemampuannya.

Kata-kata yang berasosiasi dengan kata “chat” pada ulasan negatif memberikan informasi tentang keluhan pengguna Bukalapak terhadap keterlambatan pemberitahuan chat kepada pengguna Bukalapak.

Kata-kata yang berasosiasi dengan kata “email” pada ulasan negatif memberikan informasi tentang keluhan pengguna Bukalapak terhadap gagalnya dan dipersulit dalam verifikasi ganti password email dan diharapkan pihak Bukalapak dapat membantu menangani permasalahan tersebut.

Kata-kata yang berasosiasi dengan kata “transaksi” pada ulasan negatif memberikan informasi tentang keluhan pengguna Bukalapak terhadap status tagihan berubah menjadi kadaluarsa dan status pembelian berubah menjadi dibatalkan.

Kata-kata yang berasosiasi dengan kata “upload” pada ulasan negatif memberikan informasi tentang keluhan pengguna Bukalapak terkhusus pelapak terhadap kegagalan dalam mengupload foto produk yang mereka jual dikarenakan pembaruan aplikasi padahal pelapak ingin mengupload produk yang akan mereka jual. Oleh karena itu pihak Bukalapak seharusnya bisa mengatasi permasalahan tersebut dengan memperbaiki sistem aplikasi.

Kata-kata yang berasosiasi dengan kata “promo” pada ulasan negatif memberikan informasi tentang keluhan pengguna Bukalapak terhadap promo yang ditawarkan oleh pihak Bukalapak agar promo diperbanyak, promo yang

tidak dapat digabung dengan promo lainnya seperti promo voucher, promo ongkir yang terbatas karena ada minimum transaksinya.

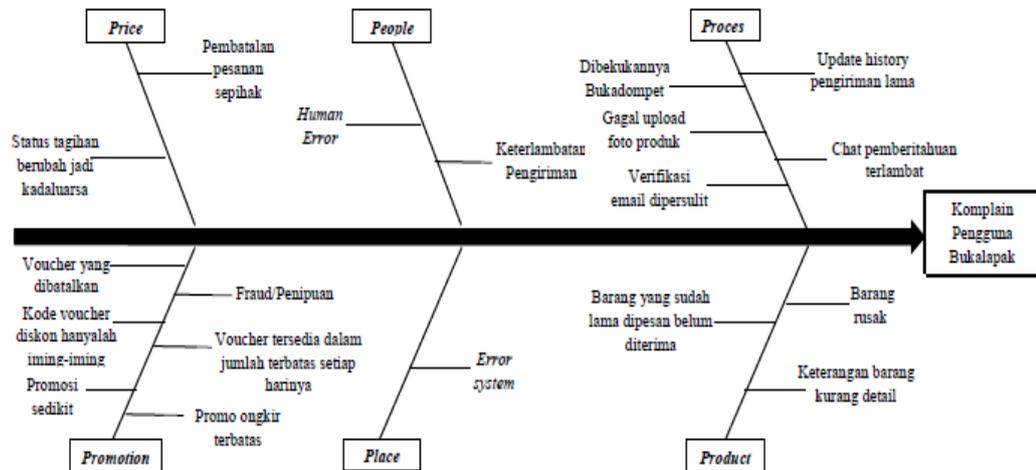
Kata-kata yang berasosiasi dengan kata “voucher” pada ulasan negatif memberikan informasi tentang keluhan pengguna Bukalapak terhadap kode voucher yang dianggap hanya sebagai iming-iming untuk menarik pembeli. Vouchernya diperbanyak juga karena ada yang kehabisan serta tidak konsisten dalam memberikan promo voucher adanya voucher yang dibatalkan. Pembeli juga mengharapkan kode voucher diperpanjang. Sebagian pembeli merasa adanya penipuan terhadap promo voucher.

Kata-kata yang berasosiasi dengan kata “bukadompot” pada ulasan negatif memberikan informasi tentang keluhan pengguna Bukalapak terhadap *topup* bukadompot yang dibekukan. Bukadompot adalah sebuah dompet virtual (*e-wallet*) yang dimiliki setiap pengguna Bukalapak guna menyimpan dana hasil penjualan (*remit*) dan dana hasil pengembalian (*refund*) transaksi. Ada juga yang sudah isi saldo bukadompot akan tetapi tidak bisa registrasi.

Kata-kata yang berasosiasi dengan kata “upgrade” pada ulasan negatif memberikan informasi tentang keluhan pengguna Bukalapak terhadap keseringan aplikasi dalam mengupgrade kadang membuat para pengguna Bukalapak kecewa, serta masalah *maintenance* juga ikut menjadi keluhan para pengguna Bukalapak. Oleh karena itu pihak Bukalapak lebih meningkatkan performanya. Kata upgrade juga berasosiasi dengan kata jelek.

### **5.7. Diagram Fishbone**

Berdasarkan hasil ulasan negatif yang didapatkan dari data, maka dapat diperoleh informasi mengenai masalah yang terjadi terkait dengan ulasan negatif pada *e-commerce* Bukalapak berdasarkan **Gambar 5.12**.



**Gambar 5.12** Diagram fishbone komplain pengguna Bukalapak

Pada **Gambar 5.12.** dapat diketahui informasi faktor-faktor yang menyebabkan *e-commerce* Bukalapak 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 *e-commerce* Bukalapak dapat dilihat pada **Tabel 5.13** berikut.

**Tabel 5.13** Rencana pemecahan masalah *e-commerce* bukalapak

| No | Faktor  | Permasalahan                   |                         | Pemecahan Masalah   |
|----|---------|--------------------------------|-------------------------|---|
|    |         | Penjual                        | Pembeli                 |   |
| 1  | Process | Update history pengiriman lama | -                       | Pihak <i>developer</i> segera memperbaiki sistem cek resi agar history pengiriman cepat <i>update</i>   |
|    |         | Chat pemberitahuan terlambat   | -                       | Perbaiki pada sistem chat   |
|    |         | Dibekukannya Bukadompet        | Dibekukannya Bukadompet | Pihak <i>Customer Service</i> segera membantu mengatasi dibekukannya bukadompet dengan memberikan penjelasan terkait dengan proses pencairan dana yang gagal. Dilakukan lagi pengecekan terhadap nomor tiket yang |

| No | Faktor         | Permasalahan                            |   | Pemecahan Masalah   |
|----|----------------|---|---|---|
|    |                | Penjual                                 | Pembeli                                       |   |
|    |                |   |   | bermasalah. Oleh sebab itulah perlunya transparansi terhadap konsumen di Bukalapak  |
|    |                | Gagal <i>update</i> foto produk terbaru | -   | Mengevaluasi sistem pembaruan aplikasi  |
|    |                | -                                       | Verifikasi email dipersulit                   | Menjelaskan secara detail langkah-langkah untuk verifikasi email di Bukalapak beserta kegunaanya  |
| 2  | <i>Product</i> | -                                       | Keterangan barang kurang detail               | Memberikan SOP untuk penjual memberikan spesifikasi terhadap produk yang dijual sehingga ada batas keharusan informasi minimal seperti apa yang harus ada dalam setiap produk.  |
|    |                | -                                       | Barang yang sudah lama dipesan belum diterima | Bekerja sama kepada pihak ekspedisi dan pihak <i>customer service</i> mengkonfirmasi penyebab keterlambatan pengiriman barang   |
|    |                | -                                       | Barang rusak                                  | Lebih dijelaskan lagi di <i>website</i> atau <i>customer service</i> bahwa Bukalapak tidak bertanggung jawab terkait barang yang rusak, akan tetapi untuk permasalahan barang rusak bisa mengajukan pengembalian pesanan ( <i>return</i> ) kepada pelapak di bagian “Diskusi Barang Komplain” yang telah disediakan. Disini admin Bukalapak akan menengahi jalannya diskusi dengan adil |

| No | Faktor | Permasalahan |   | Pemecahan Masalah  |
|----|--------|--------------|---|--|
|    |        | Penjual      | Pembeli                                   |  |
| 3  | People | -            | Human Error                               | Membuat kebijakan kepada penjual yang sudah sering menjual barang rusak/cacat tidak diperbolehkan menjualkan barangnya di Bukalapak ( <i>blacklist</i> )   |
|    |        | -            | Keterlambatan Pengiriman                  | Penyebab keterlambatan pengiriman sebenarnya biasanya terletak pada kesalahan pihak jasa pengiriman. Oleh sebab itu dimuat dalam blog Bukalapak tips untuk mengatasi masalah/keterlambatan, dijelaskan juga mengenai masalah klasik yang sering terjadi dalam pengiriman |
| 4  | Place  | Error System | Error System                              | Error system biasanya terjadi ketika ada pembaruan aplikasi. Oleh sebab itu, sebelum diluncurkan aplikasi versi terbaru harus teliti dalam melakukan pengujian sampai aplikasi tidak terdeteksi <i>error</i> .   |
| 5  | Price  | -            | Pembatalan pesanan sepihak                | Membuat SOP kepada penjual agar tidak melakukan pembatalan sepihak kepada pembeli  |
|    |        | -            | Status tagihan berubah menjadi kadaluarsa | Pihak <i>customer service</i> segera membantu dengan meminta pembeli untuk mengisi nomor transaksi, detail masalah dan mengupload bukti transfer agar ditindaklanjuti  |

| No | Faktor           | Permasalahan |                                    | Pemecahan Masalah  |
|----|------------------|--------------|------------------------------------|--|
|    |                  | Penjual      | Pembeli                            |  |
| 6  | <i>Promotion</i> | -            | Promosi sedikit                    | Menambah promosi dalam bentuk <i>sale stock</i> perharinya   |
|    |                  | -            | Promo ongkir terbatas              | Menambah kerja sama dengan pihak jasa pengiriman yang belum melakukan kerja sama   |
|    |                  | -            | Kode Voucher Hanyalah Iming- iming | Sebaiknya jika memang voucher tidak berlaku lagi maka tidak usah ditampilkan di bagian promosi.  |
|    |                  | -            | Voucher yang dibatalkan            | Mengevaluasi lagi promosi voucher yang telah diberikan kepada pengguna Bukalapak   |
|    |                  | -            | Penipuan                           | Menjelaskan lagi kepada pengguna bukalapak agar berhati-hati dalam menggunakan akun dan pihak <i>developer</i> Bukalapak lebih memperketat sistem keamanannya. |

## BAB VI

### KESIMPULAN DAN SARAN

#### 3.1. Kesimpulan

Berdasarkan hasil analisis dan pembahasan pada bab sebelumnya, penulis dapat menarik beberapa kesimpulan sebagai berikut :

1. Berdasarkan *rating* dapat diketahui bahwa mayoritas pengguna Bukalapak mempunyai penilaian ataupun persepsi yang baik terhadap *e-commerce* tersebut. Diketahui bahwa setiap bulannya banyak pengguna yang merasa sangat puas oleh aplikasi milik Bukalapak, karena memang banyak sekali fitur yang ditawarkan oleh Bukalapak kepada penggunanya yang berbeda terhadap *e-commerce* lainnya. Akan tetapi pada bulan Januari dan Februari banyak yang tidak puas akan aplikasi Bukalapak dikarenakan pembaharuan aplikasi yang membuat kecewa para pengguna dimulai dari akhir bulan Januari koneksi keserver sering gangguan. Sedangkan jika dilihat dari pelabelan sentimen jumlah ulasan positif lebih banyak dibandingkan dengan jumlah ulasan negatif. Adapun jumlah ulasan positif yaitu sebanyak 1.063 ulasan atau sebesar 55% sedangkan sisanya merupakan ulasan negatif.
2. Dengan menggunakan perbandingan data latih dan data uji sebesar 80% : 20% diperoleh hasil klasifikasi sentimen menggunakan metode *Support Vector Machine (SVM)* diperoleh tingkat akurasi sebesar 91,95% artinya dari 385 data ulasan yang diujikan, terdapat 354 ulasan yang benar pengklasifikasiannya oleh metode *SVM*. Sedangkan dengan menggunakan metode *Maximum Entropy (Maxent)* memberikan tingkat akurasi yang lebih tinggi daripada metode *SVM* yaitu sebesar 92,98% artinya dari 385 data ulasan yang diujikan, terdapat 358 ulasan yang benar pengklasifikasiannya oleh metode *Maxent*.
3. Berdasarkan hasil klasifikasi dan asosiasi teks yang dilakukan, secara umum dapat diketahui bahwa pengguna aplikasi Bukalapak mayoritas membicarakan

mengenai barang dan transaksi karena selalu muncul baik pada kelas sentimen positif maupun negatif. Secara umum metode asosiasi teks yang digunakan menunjukkan hasil ekstraksi informasi pada kelas positif diantaranya terkait barang, transaksi, fitur, pelayanan, pesanan, pengiriman, respon, berbelanja, akulaku, kebutuhan dan cicilan. Sedangkan pada kelas negatif yang sering dikeluhkan diantaranya barang, update, server, chat, email, transaksi, upload, promo, voucher, bukadompot dan upgrade.

4. Berdasarkan hasil analisis diagram sebab-akibat (diagram *fishbone*) terdapat 18 permasalahan hasil identifikasi dari ulasan negatif pengguna Bukalapak dan diklasifikasikan menjadi 6P faktor yang harus diperhatikan yaitu *Process*, *Product*, *People*, *Place*, *Price* dan *Promotion*.

### 3.2. Saran

Berdasarkan hasil analisis dan kesimpulan, dapat diberikan saran sebagai berikut:

1. Bagi pihak Bukalapak, 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. Data yang digunakan pada penelitian ini hanya satu periode aplikasi tersebut berjalan pada system *Google Play*, sehingga perlu ditambahkan data agar hasil klasifikasi sentimen lebih baik.
3. Penelitian ini hanya menganalisis satu *e-commerce* yang merupakan model C2C, untuk penelitian selanjutnya dapat membandingkan lebih dari satu *e-commerce* atau menganalisis model *e-commerce* lain seperti B2C.
4. Sistem pelabelan kelas sentimen yang digunakan dalam penelitian ini hanya sebatas pada pendeteksian sentimen antar kata menggunakan kamus *lexicon*, sehingga kata-kata negasi belum dapat teridentifikasi dengan baik, untuk penelitian selanjutnya sebaiknya dapat menggunakan sistem pelabelan yang memiliki tingkatan lebih tinggi, yakni mampu mendeteksi sentimen pada frasa dan kalimat.

## DAFTAR PUSTAKA

- Abtohi, S. 2017. *Implementasi Teknik Web Scraping dan Klasifikasi Sentimen Menggunakan Metode Support Vector Machine dan Asosiasi*. Skripsi. Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam UII Yogyakarta.
- Alexa. 2018. *Traffic Statistic Bukalapak.com*. <https://www.alexacom/siteinfo/bukalapak.com>. Akses, 3 Januari 2018.
- APJII. 2017. *Penetrasi & Perilaku Pengguna Internet Indonesia*. <https://apjii.or.id/survei2017>. Akses, 2 Januari 2018.
- Aprilia. 2017. *Analisis Faktor-Faktor yang Mempengaruhi Perilaku Pembelian Online Pada Mahasiswa Yogyakarta*. Skripsi. Program Studi Manajemen Fakultas Ekonomi UNY Yogyakarta.
- Arkhamsiagustinah. 2015. *Perbandingan Metode Second-Order Fuzzy Time Series Dari HSU dan CHEN Dalam Peramalan Produk Domestik Regional Bruto (PDRB)*. Skripsi. Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam UII Yogyakarta.
- Basnur, P.W. 2009. *Pengklasifikasian Artikel Berita Berbahasa Indonesia Secara Otomatis Menggunakan Ontologi*. Skripsi. Program Ilmu Komputer Fakultas Ilmu Komputer UI Depok.
- Berkatullah, A.,H., & Prasetyo, T. 2005. *Bisnis E-commerce Studi Sistem Keamanan dan Hukum di Indonesia*. Yogyakarta:Pustaka Pelajar.
- Berry, M.W., & Kogan, J. 2010. *Text Mining Application and theory*. United Kingdom: WILEY.
- Bukalapak. 2016. *Tentang Bukalapak*. <https://www.bukalapak.com/about>. Akses 10 Januari 2018.
- Cortes, C., & V. Vapnik. 1995. *Support Vector Networks*. Dalam Machine Learning Vol. 20, 273-297.
- DBS. 2017. *Indonesia Pusat E-commerce ASEAN*. <https://www.dbsinsights.com/id/digital-dan-inovasi/indonesia-pusat-e-commerce-asean>. Akses, 10 Januari 2018.

- Euromonitor. 2017. *Market Research Indonesia*.  
<http://www.euromonitor.com/indonesia>. Akses, 12 Mei 2018.
- Fanani, F. 2017. *Klasifikasi Review Software Pada Google Play Menggunakan Pendekatan Analisis Sentimen*. Skripsi. Program Studi Teknologi Informasi Fakultas Teknik UGM Yogyakarta.
- Fawcett, T. 2006. *An introduction to ROC analysis*. Pattern Recognition Letters 27.8, pp. 861–874.
- Feldman, R., & Sanger, J. 2007. *The Text Mining Handbook Advanced Approaches In Analyzing Unstructured Data*. New York : Cambridge University Press.
- Firdaus, M. 2015. *Analisis Faktor-Faktor Yang Mempengaruhi Pengguna Belanja Secara Online (E-Commerce)*. Skripsi. Program Studi Akuntansi Fakultas Ekonomi UII Yogyakarta.
- Fritz, G. 2016. *Analisa Bad Hike Pada Kran Lavatory Tipe S11234R Menggunakan Metode Nominal Group Technique dan Metode Fishbone di PT Surya Toto Indonesia Tbk*. Skripsi. Program Diploma Teknik Mesin Sekolah Vokasi UGM Yogyakarta.
- Gusriani, S., Wardhani, K.D.K., & Zul, I.M. 2016. *Analisis Sentimen Berdasarkan Komentar Publik Terhadap Toko Online Pada Media Sosial Facebook (Studi Kasus:Zalora dan BerryBenka)*. Jurnal Aksara Komputer Terapan Vol 5, No 2.
- Han, J., & Kamber, M. 2006. *Data Mining : Concepts and Techniques Second Edition*. San Francisco : Morgan Kauffman.
- Ha, S.H., Bae, S., & Son, L.K. 2015. *Impact of Online Consumer Reviews on Product Sales:Quantitative Analysis of the Source Effect*. Journal of Applied Mathematics & Information Sciences.Vol 9. No 2: 373-387.
- Indrajit, R., E. 2011. *Manajemen Sistem Informasi dan Teknologi Informasi*. Jakarta: Elex Media Komputindo.

- Jamil, H.N. 2017. *Analisis Sentimen Pada Online Review Menggunakan Kombinasi Metode Lexicon Based dan Naïve Bayes Classifier*. Skripsi. Program Studi Statistika FMIPA UII Yogyakarta.
- Josi, A., Abdillah, L.A., & Suryayusra. 2014. *Penerapan Teknik Web Scraping Pada Mesin Pencari Artikel Ilmiah*. Jurnal Sistem Informasi, Volume 5, Nomor 2, September 2014, hlm. 159-164.
- Kaestner, C. 2013. *Support Vector Machines and Kernel Functions for Text Processing*. RITA Volume 20 Number 3 201.
- Kotler, P., & Armstrong, G. 1996. *Dasar-Dasar Pemasaran, Edisi V, jilid 2*. Jakarta: Intermedia.
- Kotler, P. 2006. *Manajemen Pemasaran Edisi 11*. Jakarta: PT. Indeks.
- Lee, L., & Pang, B. 2008. *Opinion Mining and Sentiment Analysis*. Foundation and Trends in Information Retrieval, 2(1-2): 1-135.
- Lim, S. Y., Song, M.H., & Lee, S.J. 2006. *Ontology-based automatic classification of web documents*. Springer-Verlag, 690-700.
- Liu, B. 2012. *Sentiment Analysis and Subjectivity*. Synthesis Lectures on Human Language Technologies. USA: Morgan & Claypool Publishers.
- Liu, L., Chunfang & Zhouyang. 2013. *Analysis of Customer Satisfaction from Chinese Reviews using Opinion Mining*. 6th IEEE International Conference on Software Engineering and Service Science (ICSESS).
- Lovelock, C., & Wirtz, J. 2005. *Manajemen Pemasaran Jasa*. Indonesia: Kelompok Gramedia Indeks.
- Lovelock, C., & Wirtz, J. 2011. *Service Marketing, People, Technology, Strategy*. New Jersey: Prentice Hall Upper Sadle River.
- Mahatma, R. 2016. *Data Statistik Mengenai Pertumbuhan Pangsa Pasar Ecommerce di Indonesia*. <https://buattokoonline.id/data-statistik-mengenai-pertumbuhan-pangsa-pasar-e-commerce-di-indonesia-saat-ini/>. Akses, 2 Januari 2018.
- Manning, C. D., Raghavan, P., & Schütze, H. 2009. *An Introduction to Information Retrieval – Online Edition*. Cambridge: Cambridge University Press.

- Masithoh, N. 2016. *Analisis Klasifikasi Topik Menggunakan Metode Naïve Bayes Classifier, Naïve Bayes Multinomial Classifier, Dan Maximum Entropy Pada Artikel Berita*. Skripsi. Program Studi Statistika FMIPA UGM Yogyakarta.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. 2012. *Foundations of Machine Learning*. MIT Press.
- Moraes, R., Joan, V.F., & Wilson, P.G. 2013. *Document-level Sentiment Classification: An Empirical Comparison between SVM and ANN*. Expert Systems with Applications, 40 (2), 621–633.
- Naradhipa, R.A., & Purwarianti, A. 2012. *Sentiment Classification for Indonesian Message in Social Media*. International Conference on Electrical Engineering and Informatics: Bandung, 17-19 July 2011.
- Oktariadi, B.C. 2014. *Perbandingan Verifikasi Tanda Tangan Dengan Menggunakan Jaringan Saraf Tiruan Backpropagation dan Support Vector Machine*. Tesis. Program Studi S2 Ilmu Komputer FMIPA UGM Yogyakarta.
- Onno, P. 2000. *Mengenal E-commerce*. Jakarta: PT Elek Media Komputindo.
- Pang, B. 2002. "Thumbs up? Sentiment Classification Using Machine Learning Techniques." Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 79 - 86.
- Putranti, N.D., & Winarko, E. 2014. *Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine*. Indonesian Journal Of Computing and Cybernetics Systems Vol 8 No 1, January 2014, pp.91-100.
- Putri, D.U.K. 2016. *Implementasi Inferensi Fuzzy Mamdani Untuk Keperluan Sistem Rekomendasi Berita Berbasis Konten*. Skripsi. Program Studi Ilmu Komputer FMIPA UGM Yogyakarta.
- Refaeilzadeh, P., Tang, L., & Liu, H. 2009. *Cross Validation*. Editors: M. Tamer dan Ling Liu. Encyclopedia of Database Systems, Springer. New York.

- Saraswati, N.S. 2011. *Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis*. Skripsi. Program Studi Teknologi Informasi Fakultas Teknik UGM Yogyakarta.
- Susanti, A.R. 2016. *Analisis Klasifikasi Sentimen Twitter Terhadap Kinerja Layanan Provider Telekomunikasi Menggunakan Varian Naive Bayes*. Tesis. Institut Pertanian Bogor.
- Suthaharan, S. 2015. *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*. Springer.p. 10.ISBN 9781489976413.
- Suyanto. 2017. *Data Mining Untuk Klasifikasi dan Klusterisasi Data*. Bandung:Informatika Bandung.
- Tan, P., Steinbach, M., & Karpatne, A. 2006. *Introduction To Data Mining*. USA: Addison-Wesley.
- Tjiptono, F. 2006. *Manajemen Pelayanan Jasa*. Yogyakarta: Andi.
- Turban, E., & King, D. 2002. *Electronic Commerce 2002 – A Managerial Perspective (Second edition)*. New York: Prentice Hall.
- Turban, E., Aronson, J., & Liang, T. 2005. *Decision Support System And Intelligent System*. Upper Saddle River, New Jersey USA: Prentice Hall.
- , 2005. *Service Quality Satisfaction*. Yogyakarta: Andi.
- Ulwan, M.N. 2016. *Pattern Recognition Pada Unstructured Data Teks Menggunakan Support Vector Machine Dan Association*. Skripsi. Program Studi Statistika FMIPA UII Yogyakarta.
- 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.
- Zafikri, A. 2008. *Implementasi Metode Term Frequency Inverse Document Frequency (TF-IDF) Pada Sistem Temu Kembali Informasi*. Skripsi. Program Studi S-1 Ilmu Komputer FMIPA USU.

## 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("E:/BISMILLAH TUGAS AKHIR")
docs<-readLines("bukalapak.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("jam", "minggu", "full", "jaya", "ambil",
"nambah", "tokopedia", "shopee", "dateng", "gabisa", "kota", "dikasih",
"kemana", "yah", "yaa", "jga", "kena", "back", "lazada", "negatif", "emang",
"bagusup", "indo", "hrs", "urus", "rumah", "seminggu", "batas", "tulis",
"nulis", "dapatkan", "gede", "dipakai", "setia", "semenjak", "namanya",
"rating", "lakukan", "job", "smpai", "pdahal", "merah", "kta", "bintangnya",
"tulisan", "tanda", "edit", "butuh", "mah"))

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

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

#Replace words
docs<- tm_map(docs, gsub, pattern="ribet", replacement="rumit")
docs<- tm_map(docs, gsub, pattern="ngawur", replacement="asal-
asalan")
docs<- tm_map(docs, gsub, pattern="ngaco", replacement="asal-
asalan")
docs<- tm_map(docs, gsub, pattern="top", replacement="bagus")
docs<- tm_map(docs, gsub, pattern="keren", replacement="bagus")
docs<- tm_map(docs, gsub, pattern="belom", replacement="belum")
docs<- tm_map(docs, gsub, pattern="system", replacement="sistem")
docs<- tm_map(docs, gsub, pattern="histori",
replacement="history")
docs<- tm_map(docs, gsub, pattern="sgen", replacement="agen")
docs<- tm_map(docs, gsub, pattern="handphone", replacement="hp")
docs<- tm_map(docs, gsub, pattern="kadaluarsa",
replacement="kadaluwarsa")
docs<- tm_map(docs, gsub, pattern="tlp", replacement="telepon")
docs<- tm_map(docs, gsub, pattern="telpon", replacement="telepon")
docs<- tm_map(docs, gsub, pattern="lelet", replacement="lambat")
docs<- tm_map(docs, gsub, pattern="lemot", replacement="lambat")
docs<- tm_map(docs, gsub, pattern="tf", replacement="transfer")
docs<- tm_map(docs, gsub, pattern="cepat", replacement="cepat")
docs<- tm_map(docs, gsub, pattern="duit", replacement="uang")

```

```

docs<- tm_map(docs, gsub, pattern="nunggu",
replacement="menunggu")
docs<- tm_map(docs, gsub, pattern="ilang", replacement="hilang")
docs<- tm_map(docs, gsub, pattern="males", replacement="malas")
docs<- tm_map(docs, gsub,
pattern="notif",replacement="notifikasi")
docs<- tm_map(docs, gsub, pattern="ngasih", replacement="memberi")
docs<- tm_map(docs, gsub, pattern="brg", replacement="barang")
docs<- tm_map(docs, gsub, pattern="brang", replacement="barang")
docs<- tm_map(docs, gsub, pattern="hhilang", replacement="hilang")
docs<- tm_map(docs, gsub, pattern="nyesel",
replacement="menyesal")
docs<- tm_map(docs, gsub, pattern="komplen",
replacement="komplain")
docs<- tm_map(docs, gsub, pattern="nomer", replacement="nomor")
docs<- tm_map(docs, gsub, pattern="voucer", replacement="voucher")
docs<- tm_map(docs, gsub, pattern="apps", replacement="aplikasi")
docs<- tm_map(docs, gsub, pattern="credits", replacement="kredit")
docs<- tm_map(docs, gsub, pattern="blum", replacement="belum")
docs<- tm_map(docs, gsub, pattern="pocher", replacement="voucher")
docs<- tm_map(docs, gsub, pattern="vouchernya",
replacement="voucher")
docs<- tm_map(docs, gsub, pattern="ongkirnya",
replacement="ongkir")
docs<- tm_map(docs, gsub, pattern="negonya", replacement="nego")
docs<- tm_map(docs, gsub, pattern="cpt", replacement="cepat")
docs<- tm_map(docs, gsub, pattern="resinya", replacement="resi")
docs<- tm_map(docs, gsub, pattern="pesen", replacement="pesan")
docs<- tm_map(docs, gsub, pattern="cairin", replacement="cair")
docs<- tm_map(docs, gsub, pattern="kesel", replacement="kesal")
docs<- tm_map(docs, gsub, pattern="nyari", replacement="mencari")
docs<- tm_map(docs, gsub, pattern="sempet", replacement="sempat")
docs<- tm_map(docs, gsub, pattern="seneng", replacement="senang")
docs<- tm_map(docs, gsub, pattern="tqut", replacement="takut")
docs<- tm_map(docs, gsub, pattern="seneng", replacement="ketipu")
docs<- tm_map(docs, gsub, pattern="menungguin",
replacement="menunggu")
docs<- tm_map(docs, gsub, pattern="ktipu", replacement="ketipu")
docs<- tm_map(docs, gsub, pattern="bgus", replacement="bagus")
docs<- tm_map(docs, gsub, pattern="baguss", replacement="bagus")
docs<- tm_map(docs, gsub, pattern="mantap", replacement="bagus")
docs<- tm_map(docs, gsub, pattern="lemooottt", replacement="lama")
docs<- tm_map(docs, gsub,
pattern="menuaskan",replacement="memuaskan")
docs<- tm_map(docs, gsub, pattern="top",replacement="bagus")
docs<- tm_map(docs, gsub, pattern="cape",replacement="lelah")
docs<- tm_map(docs, gsub, pattern="nipu",replacement="tipu")
docs<- tm_map(docs, gsub, pattern="tunggu",replacement="menunggu")

```

```
docs<- tm_map(docs, gsub,
pattern="tranfer",replacement="transfer")
docs<- tm_map(docs, gsub, pattern="simpl",replacement="mudah")
docs<- tm_map(docs, gsub, pattern="photo",replacement="foto")
docs<- tm_map(docs, gsub, pattern="gambar",replacement="foto")
docs<- tm_map(docs, gsub, pattern="blanja",replacement="belanja")
docs<- tm_map(docs, gsub,
pattern="belanjanya",replacement="belanja")
docs<- tm_map(docs, gsub, pattern="sdah",replacement="sudah")
docs<- tm_map(docs, gsub, pattern="repot",replacement="rumit")

#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, 25)

dataframe<-data.frame(text=unlist(sapply(docs,
`[`) ),stringsAsFactors=F)
write.csv(dataframe,"E:/BISMILLAH TUGAS AKHIR/Bismillah TA
Ditia/Analisis/Analisis2/cleaning.csv")
save.image()
```

## Lampiran 2 Script R Pelabelan Kelas Sentimen

```

library(tm)
setwd("E:/BISMILLAH TUGAS AKHIR")
kalimat2<-read.csv("cleaning.csv",header=TRUE)

#skoring
positif<- scan("E:/BISMILLAH TUGAS AKHIR/Bismillah TA
Ditia/Analisis/s-pos.txt",what="character",comment.char=";")
negatif<- scan("E:/BISMILLAH TUGAS AKHIR/Bismillah TA
Ditia/Analisis/s-neg.txt",what="character",comment.char=";")
kata.positif = c(positif, "senang")
kata.negatif = c(negatif, "kecewa")
score.sentiment = function(kalimat2, kata.positif, kata.negatif,
.progress='none')
{
require(plyr)
require(stringr)
scores = laply(kalimat2, function(kalimat, kata.positif,
kata.negatif)
{
kalimat = gsub('[[:punct:]]', '', kalimat)
kalimat = gsub('[[:cntrl:]]', '', kalimat)
kalimat = gsub('\\d+', '', kalimat)
kalimat = tolower(kalimat)
list.kata = str_split(kalimat, '\\s+')
kata2 = unlist(list.kata)
positif.matches = match(kata2, kata.positif)
negatif.matches = match(kata2, kata.negatif)
positif.matches = !is.na(positif.matches)
negatif.matches = !is.na(negatif.matches)
score = sum(positif.matches) - (sum(negatif.matches))
return(score)
}, kata.positif, kata.negatif, .progress=.progress )
scores.df = data.frame(score=scores, text=kalimat2)
return(scores.df)
}

hasil = score.sentiment(kalimat2$text, kata.positif, kata.negatif)
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 = "hasil_pelabelan.csv")

```

### Lampiran 3 Script R Klasifikasi dengan Machine Learning menggunakan SVM dan Maxent

```

setwd("E:/BISMILLAH TUGAS AKHIR")
positifL = readLines("PL.csv")
negatifL = readLines("NL.csv")
positifT = readLines("PT.csv")
negatifT = readLines("NT.csv")
reviewL = c(positifL, negatifL)
reviewT= c(positifT, negatifT)
review_all = c(reviewL,reviewT)
sentiment_training = c(rep("positifL", length(positifL) ),
rep("negatifL", length(negatifL)))
sentiment_test = c(rep("positifL", length(positifT) ),
rep("negatifL", length(negatifT)))
sentiment_all = as.factor(c(sentiment_training, sentiment_test))

library(RTextTools)
library(e1071)

mat = create_matrix(review_all, language = "indonesian",
removeStopwords =
                        FALSE,
removeNumbers = TRUE, stemWords = FALSE, tm::weightTfIdf)
mat = as.matrix(mat)

#SVM
container<-create_container(mat, sentiment_all,
trainSize=1:1538,testSize=1539:1923, virgin=FALSE)
model<- train_model(container, 'SVM',kernel='radial')
results<- classify_model(container, model)
table(as.character(sentiment_all[1539:1923]),
as.character(results[, "SVM_LABEL"]))

recall_accuracy(sentiment_all[1539:1923], results[, "SVM_LABEL"])
create_precisionRecallSummary(container, results)

# train a MAXENT Model
models<- train_models(container, algorithms="MAXENT")
results<- classify_models(container, models)
table(as.character(sentiment_all[1539:1923]),
as.character(results[, "MAXENTROPY_LABEL"]))
recall_accuracy(sentiment_all[1539:1923],
results[, "MAXENTROPY_LABEL"])
create_precisionRecallSummary(container, results)

```

#### Lampiran 4 Script R Visualisasi dan Asosiasi Kata

```

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

setwd("E:/BISMILLAH TUGAS AKHIR")
docs<-readLines("bismillahulasannegatif.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("dipake","olshop","belah","gram","pcs",
"rame","tetangga","boss","alhasil","buku","serius","pelaya","ujung",
","gara",
"ada","mati","dasar"))

# Eliminate extra white spaces
docs<- tm_map(docs, stripWhitespace)
#Replace words
docs<- tm_map(docs, gsub, pattern="didiperbanyak",
replacement="diperbanyak")
docs<- tm_map(docs, gsub, pattern="konpirmasi",
replacement="konfirmasi")
docs<- tm_map(docs, gsub, pattern="keabisan",
replacement="kehabisan")
docs<- tm_map(docs, gsub, pattern="ketepu", replacement="ketipu")
docs<- tm_map(docs, gsub, pattern="ketipu",
replacement="penipuan")
docs<- tm_map(docs, gsub, pattern="susah", replacement="sulit")
docs<- tm_map(docs, gsub, pattern="pemengiriman",
replacement="pengiriman")

#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, 25)

#Generate the Word cloud
set.seed(860)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=45, random.order=FALSE, rot.per=0.25,

```

```
colors=brewer.pal(8, "Dark2"))

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

#asosiasi kata
v<-as.list(findAssocs(dtm, terms
=c("barang","update","server","chat","email","transaksi","upload",
"promo","voucher","bukadompet","upgrade"),
      corlimit =
c(0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15)))
v
#barplot
k<-barplot(d[1:20,]$freq, las = 2, names.arg =
d[1:20,]$word,cex.axis=1.2,cex.names=1.2,
main = "Most frequent words",
ylab = "Word frequencies",col =topo.colors(20))

termFrequency<- rowSums(as.matrix(dtm))
termFrequency<- subset(termFrequency, termFrequency>=48)

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

**Lampiran 5** *Stopwords Berbahasa Indonesia*

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

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

|               |                |                  |              |
|---------------|----------------|------------------|--------------|
| kini          | mau            | mengakhiri       | minta        |
| kinilah       | maupun         | mengapa          | mirip        |
| kira          | melainkan      | mengatakan       | misal        |
| kira-kira     | melakukan      | mengatakannya    | misalkan     |
| kiranya       | melalui        | mengenai         | misalnya     |
| kita          | melihat        | mengerjakan      | mula         |
| kitalah       | melihatnya     | mengetahui       | mulai        |
| kok           | memang         | menggunakan      | mulailah     |
| kurang        | memastikan     | menghendaki      | mulanya      |
| lagi          | memberi        | mengibaratkan    | mungkin      |
| lagian        | memberikan     | mengibaratkannya | mungkinkah   |
| lah           | membuat        | mengingat        | nah          |
| lain          | memerlukan     | mengingatkan     | naik         |
| lainnya       | memihak        | menginginkan     | namun        |
| lalu          | meminta        | mengira          | nanti        |
| lama          | memintakan     | mengucapkan      | nantinya     |
| lamanya       | memisalkan     | mengucapkannya   | nyaris       |
| lanjut        | memperbuat     | mengungkapkan    | nyatanya     |
| lanjutnya     | mempergunakan  | menjadi          | oleh         |
| lebih         | memperkirakan  | menjawab         | olehnya      |
| lewat         | memperlihatkan | menjelaskan      | pada         |
| lima          | mempersiapkan  | menuju           | padahal      |
| luar          | mempersoalkan  | menunjuk         | padanya      |
| macam         | mempertanyakan | menunjuki        | pak          |
| maka          | mempunyai      | menunjukkan      | paling       |
| makanya       | memulai        | menunjuknya      | panjang      |
| makanan       | memungkinkan   | menurut          | pantas       |
| makin         | menaiki        | menuturkan       | para         |
| malah         | menambahkan    | menyampaikan     | pasti        |
| malahan       | menandaskan    | menyangkut       | pastilah     |
| mampu         | menanti        | menyatakan       | penting      |
| mampukah      | menanti-nanti  | menyebutkan      | pentingnya   |
| mana          | menantikan     | menyeluruh       | per          |
| manakala      | menanya        | menyiapkan       | percuma      |
| manalagi      | menanyai       | merasa           | perlu        |
| masa          | menanyakan     | mereka           | perlukah     |
| masalah       | mendapat       | merekalah        | perlunya     |
| masalahnya    | mendapatkan    | merupakan        | pernah       |
| masih         | mendatang      | meski            | persoalan    |
| masihkah      | mendatangi     | meskipun         | pertama      |
| masing        | mendatangkan   | meyakini         | pertama-tama |
| masing-masing | menegaskan     | meyakinkan       | pertanyaan   |

|                |             |                    |                  |
|----------------|-------------|--------------------|------------------|
| pertanyakan    | sebelum     | sekurang-kurangnya | sering           |
| pihak          | sebelumnya  | sekurangnya        | seringnya        |
| pihaknya       | sebenarnya  | sela               | serta            |
| pukul          | seberapa    | selagi             | serupa           |
| pula           | sebesar     | selain             | sesaat           |
| pun            | sebetulnya  | selaku             | sesama           |
| punya          | sebisanya   | selalu             | sesampai         |
| rasa           | sebuah      | selama             | sesegera         |
| rasanya        | sebut       | selama-lamanya     | sesekali         |
| rata           | sebutlah    | selamanya          | seseorang        |
| rupanya        | sebutnya    | selanjutnya        | sesuatu          |
| saat           | secara      | seluruh            | sesuatunya       |
| saatnya        | secukupnya  | seluruhnya         | sesudah          |
| saja           | sedang      | semacam            | sesudahnya       |
| sajalah        | sedangkan   | semakin            | setelah          |
| saling         | sedemikian  | semampu            | setempat         |
| sama           | sedikit     | semampunya         | setengah         |
| sama-sama      | sedikitnya  | semasa             | seterusnya       |
| sambil         | seenaknya   | semasih            | setiap           |
| sampai         | segala      | semata             | setiba           |
| sampai-sampai  | segalanya   | semata-mata        | setibanya        |
| sampaikan      | segera      | semaunya           | setidak-tidaknya |
| sana           | seharusnya  | sementara          | setidaknya       |
| sangat         | sehingga    | semisal            | setinggi         |
| sangatlah      | seingat     | semisalnya         | seusai           |
| satu           | sejak       | sempat             | sewaktu          |
| saya           | sejauh      | semua              | siap             |
| sayalah        | sejenak     | semuanya           | siapa            |
| se             | sejumlah    | semula             | siapakah         |
| sebab          | sekadar     | sendiri            | siapapun         |
| sebabnya       | sekadarnya  | sendirian          | sini             |
| sebagai        | sekali      | sendirinya         | sinilah          |
| sebagaimana    | sekali-kali | seolah             | soal             |
| sebagainya     | sekalian    | seolah-olah        | soalnya          |
| sebagian       | sekaligus   | seorang            | suatu            |
| sebaik         | sekalipun   | sepanjang          | sudah            |
| sebaik-baiknya | sekarang    | sepantasnya        | sudahkah         |
| sebaiknya      | sekecil     | sepantasnyalah     | sudahlah         |
| sebaliknya     | seketika    | seperlunya         | supaya           |
| sebanyak       | sekiranya   | seperti            | tadi             |
| sebegini       | sekitar     | sepertinya         | tadinya          |
| sebegitu       | sekitarnya  | sepihak            | tahu             |

|                |              |           |         |
|----------------|--------------|-----------|---------|
| tahun          | tersampaikan | yakin     | tp      |
| tak            | tersebut     | yakni     | udh     |
| tambah         | tersebutlah  | yang      | lalu    |
| tambahnya      | tertentu     | ok        | make    |
| tampak         | tertuju      | oke       | nya     |
| tampaknya      | terus        | jos       | tgl     |
| tandas         | terutama     | deh       | bos     |
| tandasnya      | tetap        | siiiiiiip | ga      |
| tanpa          | tetapi       | sip       | gak     |
| tanya          | tiap         | hmm       | ngga    |
| tanyakan       | tiba         | shiip     | nggak   |
| tanyanya       | tiba-tiba    | waw       | nah     |
| tapi           | tidak        | okayyyy   | na      |
| tegas          | tidakkah     | byk       | kok     |
| tegasnya       | tidaklah     | bukalapak | jg      |
| telah          | tiga         | klo       | yg      |
| tempat         | tinggi       | min       | ol      |
| tengah         | toh          | sy        | cara    |
| tentang        | tunjuk       | saya      | stlh    |
| tentu          | turut        | tolong    | doank   |
| tentulah       | tutur        | tdk       | doang   |
| tentunya       | tuturnya     | pake      | sdh     |
| tepat          | ucap         | dlm       | admin   |
| terakhir       | ucapnya      | baru      | buka    |
| terasa         | ujar         | akhir     | pelapak |
| terbanyak      | ujarnya      | pas       | lapak   |
| terdahulu      | umum         | uhuy      | beli    |
| terdapat       | umumnya      | oce       | ulasan  |
| terdiri        | ungkap       | gw        | lengkap |
| terhadap       | ungkapnya    | cihuy     | beli    |
| terhadapnya    | untuk        | pdhl      | belanja |
| teringat       | usah         | bintang   | aja     |
| teringat-ingat | usai         | moga      | udah    |
| terjadi        | waduh        | dong      | kasih   |
| terjadilah     | wah          | oc        | mohon   |
| terjadinya     | wahai        | krn       | banget  |
| terkira        | waktu        | malah     | kalo    |
| terlalu        | waktunya     | biar      | toko    |
| terlebih       | walau        | kaya      | good    |
| terlihat       | walaupun     | trima     | terima  |
| termasuk       | wong         | cumn      | masuk   |
| ternyata       | yaitu        | trims     | jual    |

|             |          |           |            |
|-------------|----------|-----------|------------|
| semoga      | tuh      | liat      | bnyk       |
| kali        | ane      | dapet     | yah        |
| coba        | Hahahhaa | hati      | kaga       |
| terimakasih | kpd      | dibuka    | abis       |
| gimana      | yth      | trs       | ribu       |
| sih         | ehh      | gini      | pilih      |
| blm         | sma      | bilang    | abis       |
| aplikasinya | tq       | indonesia | sengaja    |
| nih         | klw      | the       | belum      |
| sebelah     | hny      | gmn       | pokoknya   |
| isi         | hemmmm   | nama      | smoga      |
| bikin       | lbh      | tetep     | bnyk       |
| cari        | sich     | sayang    | adakan     |
| maju        | cingcay  | dri       | sehari     |
| best        | suwun    | lihat     | mending    |
| utk         | ðy       | blom      | kedepannya |
| sya         | bgt      | males     | trimakasih |
| habis       | tpi      | cuman     | joss       |
| karna       | jdi      | ama       | tkz        |
| dah         | bsa      | ngak      | ayo        |
| dgn         | gan      | harap     | tertera    |
| isi         | knp      | org       | malam      |
| sampe       | skrg     | dlu       | msh        |
| ganti       | gitu     | kemaren   | kebanyakan |
| trus        | donk     | smpe      | ampun      |
| blm         | mulu     | dll       | smua       |
| app         | gitu     | pke       | pdhal      |
| pulsa       | suruh    | tanggal   | krna       |
| sukses      | orang    | pagi      | knpa       |
| gua         | klik     | masak     | kya        |
| kl          | jgn      | makasih   | tlg        |
| kan         | maaf     | apk       | tgl        |
| spy         | tau      | kayak     | ajah       |
| y           | kadang   | lgi       | selamat    |
| sdangkan    | bener    | thx       | nya        |
| slalu       | muncul   | bantu     | pun        |
| deh         | thanks   | uda       | abis       |
| lgsg        | nyampe   | push      | koq        |
| ah          | yah      | apapun    | an         |
| apl         | gue      | tingkat   | pastinya   |
| hehe        | dpt      | kagak     | gtu        |
| loh         | gmna     | banyakan  | br         |

bl  
ko  
ttg  
ya  
des  
januari  
rupiah  
dg  
nmr  
hp  
aje  
nyusul  
paling  
sh  
gmna  
ni  
cincai  
suda  
tuju  
mao  
v  
nye  
gni  
kapok  
haha  
ols  
dst  
x  
ya  
malh  
g  
sekat  
huhhhh  
sekarnag  
manapun  
cape  
kemarin  
yakali  
tlong  
gann  
alhamdulillah  
hadeh  
mesen

### Lampiran 6 *Output SVM dan Maxent*

```
> mat = as.matrix(mat)
> container<-create_container(mat, sentiment_all,
+                             trainSize=1:1538,testSize=1539:1923, virgin=FALSE)
> model <- train_model(container, 'SVM',kernel='radial')
> results <- classify_model(container, model)
> table(as.character(sentiment_all[1539:1923]),
+       as.character(results["SVM_LABEL"]))

      negatifL positifL
negatifL      152       20
positifL       11      202
> recall_accuracy(sentiment_all[1539:1923], results["SVM_LABEL"])
[1] 0.9194805
```

```
> models <- train_models(container, algorithms="MAXENT")
> results <- classify_models(container, models)
> table(as.character(sentiment_all[1539:1923]),
+       as.character(results["MAXENTROPY_LABEL"]))

      negatifL positifL
negatifL      155       17
positifL       10      203
> recall_accuracy(sentiment_all[1539:1923], results["MAXENTROPY_LABEL"])
[1] 0.9298701
```